



UNIVERSIDAD DE OVIEDO

Departamento de Informática

Programa de Doctorado en Informática

**RECONOCIMIENTO DE ATRIBUTOS FACIALES
MEDIANTE VISIÓN POR COMPUTADOR PARA LA
DETECCIÓN DE DISTRACCIÓN Y SOMNOLENCIA
EN CONDUCTORES**

TESIS DOCTORAL POR

Alberto Fernández Villán

8 de marzo de 2017



AUTORIZACIÓN PARA LA PRESENTACIÓN DE TESIS DOCTORAL

Año Académico: 2016/2017

1.- Datos personales del autor de la Tesis		
Apellidos: Fernández Villán	Nombre: Alberto	
DNI/Pasaporte/NIE: 10.902.715W	Teléfono: 678507523	Correo electrónico: fernandezvillan.alberto@gmail.com

2.- Datos académicos	
Programa de Doctorado cursado: Informática	
Órgano responsable: Centro Internacional de Postgrado	
Departamento/Instituto en el que presenta la Tesis Doctoral: Informática	
Título definitivo de la Tesis	
Español/Otro Idioma: Reconocimiento de atributos faciales mediante visión por computador para la detección de distracción y somnolencia en conductores	Inglés: Facial attributes recognition using computer vision to detect drowsiness and distraction in drivers
Rama de conocimiento: Ingeniería y Arquitectura	

3.- Autorización del Director/es y Tutor de la tesis	
D/D ^a : Rubén Usamentiaga Fernández	DNI/Pasaporte/NIE: 13.942.310D
Departamento/Instituto: Informática	
D/D ^a : Rubén Casado Tejedor	DNI/Pasaporte/NIE: 53.540.545H
Departamento/Instituto/Institución:	
Autorización del Tutor de la tesis	
D/D ^a : Daniel F. García Martínez	DNI/Pasaporte/NIE: 10.820.663Z
Departamento/Instituto: Informática	

Autoriza la presentación de la tesis doctoral en cumplimiento de lo establecido en el Art.30.1 del Reglamento de los Estudios de Doctorado, aprobado por el Consejo de Gobierno, en su sesión del día 17 de junio de 2013 (BOPA del 25 de junio de 2013)

Gijón, a 26 de enero de 2017

Director/es de la Tesis

Fdo.: Rubén Usamentiaga

Fdo.: Rubén Casado

Tutor de la Tesis

Fdo.: Daniel F. García



RESOLUCIÓN DE PRESENTACIÓN DE TESIS DOCTORAL

Año Académico: 2016/2017

1.- Datos personales del autor de la Tesis		
Apellidos: Fernández Villán	Nombre: Alberto	
DNI/Pasaporte/NIE: 10.902.715W	Teléfono: 678507523	Correo electrónico: fernandezvillan.alberto@gmail.com

2.- Datos académicos	
Programa de Doctorado cursado: Informática	
Órgano responsable: Centro Internacional de Postgrado	
Departamento/Instituto en el que presenta la Tesis Doctoral: Informática	
Título definitivo de la Tesis	
Español/Otro Idioma: Reconocimiento de atributos faciales mediante visión por computador para la detección de distracción y somnolencia en conductores	Inglés: Facial attributes recognition using computer vision to detect drowsiness and distraction in drivers
Rama de conocimiento: Ingeniería y Arquitectura	
Señale si procede: ☐ Mención Internacional ☐ Idioma de presentación de la Tesis distinto al español ☒ X Presentación como compendio de publicaciones	

3.- Autorización del Presidente de la Comisión Académica / Director del Departamento	
D/Dª: Luciano Sánchez Ramos	DNI/Pasaporte/NIE: 11413974V
Departamento/Instituto: Informática	

Resolución

La Comisión Académica del Programa de Doctorado Informática en su reunión de fecha 6/2/2017, acordó la presentación de la tesis doctoral a la Comisión de Doctorado, previa comprobación de que la tesis presentada y la documentación que la acompaña cumplen con la normativa vigente, según lo establecido en el Art.30.7 del Reglamento de los Estudios de Doctorado, aprobado por el Consejo de Gobierno, en su sesión del día 17 de junio de 2013 (BOPA del 25 de junio de 2013)

Además, informa:

	Favorable	Desfavorable
Mención Internacional	☐	☐
Idioma	☐	☐
Presentación como compendio de publicaciones	X	☐



Justificación

La tesis cumple con todos los criterios acordados por la comisión académica del programa

Gijón, 6/2/2017

Presidente de la Comisión Académica del Programa de Doctorado

Fdo.: Luciano Sánchez Ramos

Contra la presente resolución podrá interponer recurso de alzada ante el Excmo. Sr. Rector Magfco. de esta Universidad en el plazo de un mes a contar desde el siguiente a la recepción de la presente resolución, de acuerdo con lo previsto en el artículo 114 de la Ley 30/92, de 26 de noviembre, del Régimen Jurídico de las Administraciones Públicas y Procedimiento Administrativo Común (B.O.E. de 27 de noviembre), modificada por la Ley 4/1999, de 13 de enero (B.O.E. de 14 de enero)

SR.

(Doctorando, Director de la Tesis Doctoral y Sra. Directora del Centro Internacional de Postgrado)



RESUMEN DEL CONTENIDO DE TESIS DOCTORAL

1.- Título de la Tesis	
Español/Otro Idioma: Reconocimiento de atributos faciales mediante visión por computador para la detección de distracción y somnolencia en conductores	Inglés: Facial attributes recognition using computer vision to detect drowsiness and distraction in drivers
2.- Autor	
Nombre: Alberto Fernández Villán	DNI/Pasaporte/NIE: 10.902.715W
Programa de Doctorado: Informática	
Órgano responsable: Centro Internacional de Postgrado	

RESUMEN (en español)

La conducción es una actividad que requiere un alto grado de concentración por parte de la persona que la realiza, ya que el más pequeño descuido es suficiente para sufrir un accidente con las consiguientes pérdidas materiales y/o humanas. De acuerdo con el estudio más reciente publicado por la Organización Mundial de la Salud (OMS) en 2013, se estimó que 1.25 millones de personas mueren como resultado de accidentes de tráfico y entre 20 y 50 millones más sufren accidentes sin perder la vida pero pudiendo derivar en dolencias crónicas. Todas estas muertes y accidentes no sólo afectan de manera directa a los familiares de las víctimas sino que, además, tienen un alto coste sobre los presupuestos de los gobiernos, que se estima entre un 3 y un 5% del producto interior bruto.

De todos estos accidentes, muchos son provocados por lo que se conoce como inatención. Este término engloba diferentes estados del conductor, como pueden ser la distracción y la somnolencia, siendo precisamente éstos los que más fatalidades ocasionan. Existen muchas publicaciones e investigaciones que intentan poner cifras que indiquen las consecuencias producidas por la inatención (y sus subtipos), pero no existe una figura exacta sobre los accidentes causados por la inatención puesto que todos estos estudios están realizados en diferentes lugares, diferentes marcos temporales, y por tanto, en diferentes condiciones. En líneas generales, se calcula que la inatención ocasiona entre el 25% y el 75% de los accidentes y casi-accidentes.

En un estudio realizado en 10 países europeos acerca de somnolencia y conducción se determinó que dicha conjunción incrementa el tiempo de reacción en un 86% y es la cuarta causa de muerte en las carreteras españolas. Además, cabe destacar que el 75% de los conductores españoles han sufrido episodios de somnolencia mientras conducían, porcentaje muy superior a la media del 47% que han admitido este hecho. Además, otro factor importante a tener en cuenta es que, aunque los accidentes producidos por la somnolencia suelen ser muy graves (vistas las estadísticas anteriores de mortalidad), muchos conductores infravaloran esta situación y conducen aunque noten la presencia de sus síntomas. Bostezos frecuentes, cabeceos, visión borrosa, caída de párpados y esfuerzos por mantener tanto la atención como los ojos abiertos son signos habituales de somnolencia. Respecto a la distracción, éste es uno de los factores que más fatalidades ocasiona en España. De acuerdo con la Dirección General de Tráfico (DGT), la distracción es la primera causa detectada en los accidentes con víctimas, un 13.15% de los casos. Los últimos datos y estadísticas arrojados por inatención en los conductores españoles son dramáticos, siendo la causa que más víctimas mortales ocasiona (36%), muy por encima de la velocidad inadecuada (21%) o el nivel de alcoholemia (11%).

A causa de estas cifras y sus consecuencias, la inatención se ha convertido en un campo ampliamente estudiado por la comunidad investigadora, donde los estudios y



soluciones para luchar contra la distracción y la somnolencia, en particular, y la inatención, en general, se pueden clasificar en tres grandes grupos: 1) métodos basados en el análisis del comportamiento del vehículo, 2) métodos basados en el análisis de variables fisiológicas del conductor capturadas por diferentes sensores y 3) métodos basados en el análisis de características visuales del conductor mediante la captura de imágenes aplicando métodos de visión por computador, los cuales, por sus características, se han posicionado como una forma no intrusiva y eficaz para la detección tanto de la distracción como de la somnolencia.

La motivación de la presente tesis doctoral es consecuencia directa de las cifras anteriores y radica en ofrecer mecanismos para detectar y reducir los efectos ocasionados por la inatención en los conductores. Es por ello que el objetivo principal del presente trabajo sea alcanzar un mayor grado de conocimiento en todo lo relacionado con la inatención en los conductores, teniendo como fin último la reducción del número de accidentes y víctimas mortales ocasionados por esta causa haciendo uso de herramientas de la información y la comunicación (TIC).

A tal efecto, la investigación se ha centrado en el reconocimiento de atributos faciales, que puedan ser empleados a posteriori para una detección robusta de indicadores, que permitan una clasificación de episodios de distracción y somnolencia en conductores. En segundo lugar, la investigación ha intentado establecer un marco de referencia para caracterizar la distracción en los conductores con lo que futuras investigaciones tengan un punto de partida como referente. En tercer lugar, se ha propuesto, construido y validado una arquitectura basada en el análisis de características visuales mediante el empleo de técnicas de visión por computador y aprendizaje automático para la detección tanto de la distracción como de la somnolencia. En concreto, se propone una arquitectura de procesamiento especialmente diseñada para operar en entornos vehiculares, con una carga computacional muy baja y fácilmente integrable en dispositivos con reducidas capacidades de cómputo, capaz de lidiar con imágenes en distintas condiciones muy presentes en este tipo de entornos. El sistema de control propuesto integra varios elementos innovadores permitiendo que pueda operar de forma completamente autónoma para la detección robusta de los principales indicadores visuales, caracterizando tanto la distracción como la somnolencia del conductor. La arquitectura se ha validado, en primer lugar, con bases de datos de referencia validando los diferentes módulos que la componen y, en segundo lugar, con usuarios en entornos reales obteniendo, en ambos casos, unos resultados prometedores con una carga computacional adecuada para los dispositivos embebidos habituales en entornos vehiculares.

RESUMEN (en Inglés)

Driving is an activity that requires a high degree of concentration on the part of the person who performs it since the slightest negligence is sufficient to provoke an accident with the consequent material and/or human losses. According to the most recent study published by the World Health Organization (WHO) in 2013, it was estimated that 1.25 million people die as a result of traffic accidents, whereas between 20 and 50 million do not die but consequences may result in chronic conditions. All these deaths and accidents not only have a direct impact on victims and families, but they also mean a high cost for government budgets, estimated at between 3% and 5% of their Gross Domestic Product (GDP).

Many of these accidents are caused by what is known as inattention. This term encloses a driver's different conditions such as distraction and drowsiness, which are, precisely, the ones that cause more fatalities. Many publications and research have tried to set figures that indicate the consequences of inattention (and its subtypes), but there is no exact number of the accidents caused by inattention since all these studies have been carried out in different places, different time frames and, therefore, under different conditions. Overall, it has been estimated that inattention causes between 25% and 75% of accidents and near-accidents. A study on drowsiness while driving in ten European



countries found that fatigue risks increasing reaction time by 86% and it is the fourth leading cause of death on Spanish roads. In addition, it is noted that 75% of Spanish drivers have suffered episodes of sleepiness while driving, a much higher percentage than the average of 47% who admitted this fact. In addition, another important factor to consider is that, although accidents caused by drowsiness are usually very serious (having regard to the abovementioned fatal statistics), many drivers underestimate this situation and drive even if they notice the presence of symptoms. Frequent yawning, pitching movements, blurred vision, drooping upper eyelids and efforts to keep both attention and eyes open are common signs of drowsiness. With respect to distraction, this is a major contributor to fatal accidents in Spain. According to the Directorate General of Traffic (DGT), distraction is the first violation found in fatal accidents, 13.15% of the cases. The latest statistics on inattentive driving in Spanish drivers are alarming, appearing as the leading cause of fatalities (36%), well above excessive speed (21%) or alcohol consumption (11%).

Because of these figures and their consequences, inattention has become a widely studied field by the research community, whose studies and solutions to combat distraction and sleepiness, in particular, and inattention, in general, can be divided into three broad groups: 1) methods based on the analysis of the behavior of the vehicle, 2) methods based on the analysis of the driver's physiological variables captured by different sensors and 3) methods based on the analysis of the driver's visual characteristics by capturing images using computer vision methods, which, for their non-intrusive and effective characteristics, have become a leading way to detect both distraction and drowsiness.

The reason for this PhD thesis is the direct consequences of the abovementioned figures and its purpose is to provide mechanisms to help reduce driver inattention effects. That is why the main objective is to reach a better knowledge regarding driving inattention in order to reduce the number of accidents and fatalities. The main focus is that the extraction of facial attributes in a solid way could be used for a robust detection of indicators, which enables a classification of distraction and drowsiness episodes in drivers. Secondly, research has attempted to establish a frame of reference to characterize distraction in drivers in order to provide solid foundations for future research. Thirdly, an architecture based on the analysis of visual characteristics has been proposed, constructed and validated by using techniques of computer vision and automatic learning for the detection of both distraction and drowsiness. In particular, a processing architecture specially designed to operate in vehicular environments is proposed, with a very low computational load and easily embeddable into devices with reduced computational capacities in order to deal with images in the different conditions prevailing in this type of environments. The proposed control system integrates several innovative elements in order to operate in a completely autonomous way for the robust detection of the main visual indicators characterizing the driver's both distraction and drowsiness. The architecture has been validated, firstly, with reference databases testing the different modules that compose it, and, secondly, with users in real environments, obtaining in both cases, promising results with a suitable computational load for the embedded devices in vehicle environments.

SR. DIRECTOR DE DEPARTAMENTO DE INFORMÁTICA/

SR. PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO EN INFORMÁTICA



INFORME PARA LA PRESENTACIÓN DE TESIS DOCTORAL COMO COMPENDIO DE PUBLICACIONES

Año Académico: 2016/2017

1.- Datos personales del autor de la Tesis		
Apellidos: Fernández Villán	Nombre: Alberto	
DNI/Pasaporte/NIE: 10.902.715W	Teléfono: 678507523	Correo electrónico: fernandezvillan.alberto@gmail.com

2.- Datos académicos	
Programa de Doctorado cursado: Informática	
Órgano responsable: Centro Internacional de Postgrado	
Departamento/Instituto en el que presenta la Tesis Doctoral: Informática	
Título definitivo de la Tesis	
Español/Otro Idioma: Reconocimiento de atributos faciales mediante visión por computador para la detección de distracción y somnolencia en conductores	Inglés: Facial attributes recognition using computer vision to detect drowsiness and distraction in drivers
Rama de conocimiento: Ingeniería y Arquitectura	

3.- Director/es de la Tesis	
D/D ^a : Rubén Usamentiaga Fernández	DNI/Pasaporte/NIE: 13.942.310D
Departamento/Instituto: Informática	
D/D ^a : Rubén Casado Tejedor	DNI/Pasaporte/NIE: 53.540.545H
Departamento/Instituto/Institución:	

4.- Informe
La tesis doctoral se presenta como compendio de publicaciones, incluyendo cuatro publicaciones en revistas indexadas en el <i>Journal Citation Reports</i> (cuartiles Q1, Q2, Q4 y Q4), otra publicación en revista de menor relevancia y tres publicaciones en congresos internacionales. Una de estas publicaciones obtuvo el premio como mejor artículo del congreso. Todo ello justifica la presentación de la tesis como compendio de publicaciones.

Gijón, 26 de enero de 2017

Director/es de la Tesis Doctoral

Fdo.: Rubén Usamentiaga

Fdo.: Rubén Casado

FOR-MAT-VOA-034-2



ACEPTACIÓN COAUTORES PRESENTACIÓN TRABAJOS FORMANDO PARTE DE TESIS DOCTORAL COMO COMPENDIO DE PUBLICACIONES

1.- Datos personales del coautor

Apellidos: Casado Tejedor	Nombre: Rubén	
DNI/Pasaporte/NIE: 53.540.545-H	Teléfono: 667948053	Correo electrónico ryden13@gmail.com

2.- Publicaciones que formarán parte de la tesis y de las que es coautor

- Fernández, A., García, R., Usamentiaga, R., & Casado, R. (2015). Glasses detection on real images based on robust alignment. *Machine Vision and Applications*, 26(4), 519-531.
- Fernández, A., Carús, J. L., Usamentiaga, R., Alvarez, E., & Casado, R. (2015, July). Unobtrusive health monitoring system using video-based physiological information and activity measurements. In *Computer, Information and Telecommunication Systems (CITS), 2015 International Conference on* (pp. 1-5). IEEE.
- Fernández, A., Casado, R., & Usamentiaga, R. (2015, August). A Real-Time Big Data Architecture for Glasses Detection Using Computer Vision Techniques. In *Future Internet of Things and Cloud (FiCloud), 2015 3rd International Conference on* (pp. 591-596). IEEE.
- Fernández, A., Usamentiaga, R., Carús, J. L., & Casado, R. (2016). Driver distraction using visual-based sensors and algorithms. *Sensors*, 16(11), 1805.
- Villan, A. F., Candas, J. L. C., Fernandez, R. U., & Tejedor, R. C. (2016). Face recognition and spoofing detection system adapted to visually-impaired people. *IEEE Latin America Transactions*, 14(2), 913-921.
- Fernández, A., Carús, J. L., Usamentiaga, R., Alvarez, E., & Casado, R. (2017). Wearable and ambient sensors to health monitoring using computer vision and signal processing techniques. in *Journal of Networks*.
- Fernández, A., Usamentiaga, R. & Casado, R. (2017). Automatic system to detect both distraction and drowsiness in drivers using robust visual features. *Revista Iberoamericana de Automatica e Informatica industrial*.

ACEPTACIÓN:

Acepto que las publicaciones anteriores formen parte de la tesis doctoral titulada Reconocimiento de atributos faciales mediante visión por computador para la detección de distracción y somnolencia en conductores

Y elaborada por D. Alberto Fernández Villán

Gijón, 25 de Enero de 2017

Firma



ACEPTACIÓN COAUTORES PRESENTACIÓN TRABAJOS FORMANDO PARTE DE TESIS DOCTORAL COMO COMPENDIO DE PUBLICACIONES

1.- Datos personales del coautor

Apellidos: González Losada	Nombre: Daniel	
DNI/Pasaporte/NIE: 71773296W	Teléfono: 650148839	Correo electrónico: danigonzaalezlosada@gmail.com

2.- Publicaciones que formarán parte de la tesis y de las que es coautor

- Losada, D. G., Lopez, G. A. R., Acevedo, R. G., & Villan, A. F. (2013, December). AVIUE—Artificial vision to improve the user experience. In *New Concepts in Smart Cities: Fostering Public and Private Alliances (SmartMILE), 2013 International Conference on* (pp. 1-6). IEEE.

ACEPTACIÓN:

Acepto que las publicaciones anteriores formen parte de la tesis doctoral titulada Reconocimiento de atributos faciales mediante visión por computador para la detección de distracción y somnolencia en conductores

Y elaborada por D. Alberto Fernández Villán

Gijón, 25 de Enero de 2017

Firma





ACEPTACIÓN COAUTORES PRESENTACIÓN TRABAJOS FORMANDO PARTE DE TESIS DOCTORAL COMO COMPENDIO DE PUBLICACIONES

1.- Datos personales del coautor

Apellidos: Álvarez Álvarez	Nombre: Eduardo	
DNI/Pasaporte/NIE: 10872640B	Teléfono: 667155554	Correo electrónico: edualvarez@uniovi.es

2.- Publicaciones que formarán parte de la tesis y de las que es coautor

- Fernández, A., Carús, J. L., Usamentiaga, R., Alvarez, E., & Casado, R. (2015, July). Unobtrusive health monitoring system using video-based physiological information and activity measurements. In *Computer, Information and Telecommunication Systems (CITS), 2015 International Conference on* (pp. 1-5). IEEE.
- Fernández, A., Carús, J. L., Usamentiaga, R., Alvarez, E., & Casado, R. (2017). Wearable and ambient sensors to health monitoring using computer vision and signal processing techniques. in *Journal of Networks*.

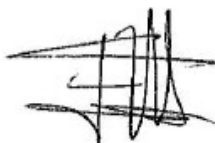
ACEPTACIÓN:

Acepto que las publicaciones anteriores formen parte de la tesis doctoral titulada Reconocimiento de atributos faciales mediante visión por computador para la detección de distracción y somnolencia en conductores

Y elaborada por D. Alberto Fernández Villán

Gijón, 25 de Enero de 2017

Firma





ACEPTACIÓN COAUTORES PRESENTACIÓN TRABAJOS FORMANDO PARTE DE TESIS DOCTORAL COMO COMPENDIO DE PUBLICACIONES

1.- Datos personales del coautor

Apellidos: Rosal Lopez	Nombre: Gustavo	
DNI/Pasaporte/NIE: 10.865.944-P	Teléfono: 660180538	Correo electrónico: gusavoole@gmail.com

2.- Publicaciones que formarán parte de la tesis y de las que es coautor

- Losada, D. G., Lopez, G. A. R., Acevedo, R. G., & Villan, A. F. (2013, December). AVIUE—Artificial vision to improve the user experience. In *New Concepts in Smart Cities: Fostering Public and Private Alliances (SmartMILE), 2013 International Conference on* (pp. 1-6). IEEE.

ACEPTACIÓN:

Acepto que las publicaciones anteriores formen parte de la tesis doctoral titulada Reconocimiento de atributos faciales mediante visión por computador para la detección de distracción y somnolencia en conductores

Y elaborada por D. Alberto Fernández Villán

Gijón, 25 de Enero de 2017

Firma





ACEPTACIÓN COAUTORES PRESENTACIÓN TRABAJOS FORMANDO PARTE DE TESIS DOCTORAL COMO COMPENDIO DE PUBLICACIONES

1.- Datos personales del coautor

Apellidos: Carús Candás	Nombre: Juan Luis	
DNI/Pasaporte/NIE: 71.664.340-C	Teléfono: 609035645	Correo electrónico: juanluis.carus@gmail.com

2.- Publicaciones que formarán parte de la tesis y de las que es coautor

- Fernández, A., Carús, J. L., Usamentiaga, R., Alvarez, E., & Casado, R. (2015, July). Unobtrusive health monitoring system using video-based physiological information and activity measurements. In *Computer, Information and Telecommunication Systems (CITS), 2015 International Conference on* (pp. 1-5). IEEE.
- Fernández, A., Usamentiaga, R., Carús, J. L., & Casado, R. (2016). Driver distraction using visual-based sensors and algorithms. *Sensors*, 16(11), 1805.
- Villan, A. F., Candás, J. L. C., Fernandez, R. U., & Tejedor, R. C. (2016). Face recognition and spoofing detection system adapted to visually-impaired people. *IEEE Latin America Transactions*, 14(2), 913-921.
- Fernández, A., Carús, J. L., Usamentiaga, R., Alvarez, E., & Casado, R. (2017). Wearable and ambient sensors to health monitoring using computer vision and signal processing techniques. in *Journal of Networks*.

ACEPTACIÓN:

Acepto que las publicaciones anteriores formen parte de la tesis doctoral titulada Reconocimiento de atributos faciales mediante visión por computador para la detección de distracción y somnolencia en conductores

Y elaborada por D. Alberto Fernández Villán

Gijón, 25 de Enero de 2017

Firma





ACEPTACIÓN COAUTORES PRESENTACIÓN TRABAJOS FORMANDO PARTE DE TESIS DOCTORAL COMO COMPENDIO DE PUBLICACIONES

1.- Datos personales del coautor

Apellidos: García Acevedo	Nombre: Rodrigo	
DNI/Pasaporte/NIE: 53543129A	Teléfono: 625869737	Correo electrónico: garciaacevedorodrigo@gmail.com

2.- Publicaciones que formarán parte de la tesis y de las que es coautor

- Fernández, A., García, R., Usamentiaga, R., & Casado, R. (2015). Glasses detection on real images based on robust alignment. *Machine Vision and Applications*, 26(4), 519-531.

ACEPTACIÓN:

Acepto que las publicaciones anteriores formen parte de la tesis doctoral titulada Reconocimiento de atributos faciales mediante visión por computador para la detección de distracción y somnolencia en conductores

Y elaborada por D. Alberto Fernández Villán

Gijón, 25 de Enero de 2017

Firma



ACEPTACIÓN COAUTORES PRESENTACIÓN TRABAJOS FORMANDO PARTE DE TESIS DOCTORAL COMO COMPENDIO DE PUBLICACIONES

1.- Datos personales del coautor

Apellidos: Usamentiaga Fernández	Nombre: Rubén	
DNI/Pasaporte/NIE: 13942310D	Teléfono: 985182626	Correo electrónico rusamentiaga@uniovi.es

2.- Publicaciones que formarán parte de la tesis y de las que es coautor

- Fernández, A., García, R., Usamentiaga, R., & Casado, R. (2015). Glasses detection on real images based on robust alignment. *Machine Vision and Applications*, 26(4), 519-531.
- Fernández, A., Carús, J. L., Usamentiaga, R., Alvarez, E., & Casado, R. (2015, July). Unobtrusive health monitoring system using video-based physiological information and activity measurements. In *Computer, Information and Telecommunication Systems (CITS), 2015 International Conference on* (pp. 1-5). IEEE.
- Fernández, A., Casado, R., & Usamentiaga, R. (2015, August). A Real-Time Big Data Architecture for Glasses Detection Using Computer Vision Techniques. In *Future Internet of Things and Cloud (FiCloud), 2015 3rd International Conference on* (pp. 591-596). IEEE.
- Fernández, A., Usamentiaga, R., Carús, J. L., & Casado, R. (2016). Driver distraction using visual-based sensors and algorithms. *Sensors*, 16(11), 1805.
- Villan, A. F., Candas, J. L. C., Fernandez, R. U., & Tejedor, R. C. (2016). Face recognition and spoofing detection system adapted to visually-impaired people. *IEEE Latin America Transactions*, 14(2), 913-921.
- Fernández, A., Carús, J. L., Usamentiaga, R., Alvarez, E., & Casado, R. (2017). Wearable and ambient sensors to health monitoring using computer vision and signal processing techniques. in *Journal of Networks*.
- Fernández, A., Usamentiaga, R. & Casado, R. (2017). Automatic system to detect both distraction and drowsiness in drivers using robust visual features. *Revista Iberoamericana de Automatica e Informatica industrial*.

ACEPTACIÓN:

Acepto que las publicaciones anteriores formen parte de la tesis doctoral titulada Reconocimiento de atributos faciales mediante visión por computador para la detección de distracción y somnolencia en conductores

Y elaborada por D. Alberto Fernández Villán

Gijón, 25 de Enero de 2017

Firma



RENUNCIA COAUTORES PRESENTACIÓN TRABAJOS FORMANDO PARTE DE TESIS DOCTORAL

1.- Datos personales del coautor

Apellidos: González Losada	Nombre: Daniel	
DNI/Pasaporte/NIE: 71773296W	Teléfono: 650148839	Correo electrónico: danigonzalezlosada@gmail.com

2.- Tesis Doctoral

Título: Reconocimiento de atributos faciales mediante visión por computador para la detección de distracción y somnolencia en conductores

Autor: Alberto Fernández Villán

Programa de doctorado: Programa de Doctorado en Ingeniería Informática

2.- Publicaciones que formarán parte de la tesis y de las que es coautor

- Villan, A. F, Losada, D. G., Lopez, G. A. R. & Acevedo, R. G. (2013, December). AVIUE—Artificial vision to improve the user experience. In *New Concepts in Smart Cities: Fostering Public and Private Alliances (SmartMILE), 2013 International Conference on* (pp. 1-6). IEEE.

RENUNCIA:

Renuncio a que las publicaciones anteriores sean presentadas como parte de otra tesis doctoral presentada como compendio de publicaciones.

Firma

Gijón, 25 de Enero de 2017

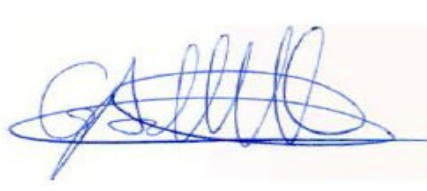


RENUNCIA COAUTORES PRESENTACIÓN TRABAJOS FORMANDO PARTE DE TESIS DOCTORAL

1.- Datos personales del coautor		
Apellidos: Rosal Lopez	Nombre: Gustavo	
DNI/Pasaporte/NIE: 10.865.944-P	Teléfono: 660180538	Correo electrónico: gusavoole@gmail.com

2.- Tesis Doctoral
Título: Reconocimiento de atributos faciales mediante visión por computador para la detección de distracción y somnolencia en conductores
Autor: Alberto Fernández Villán
Programa de doctorado: Programa de Doctorado en Ingeniería Informática

2.- Publicaciones que formarán parte de la tesis y de las que es coautor
Villan, A. F, Losada, D. G., Lopez, G. A. R. & Acevedo, R. G. (2013, December). AVIUE—Artificial vision to improve the user experience. In <i>New Concepts in Smart Cities: Fostering Public and Private Alliances (SmartMILE)</i>, 2013 International Conference on (pp. 1-6). IEEE.

RENUNCIA:
Renuncio a que las publicaciones anteriores sean presentadas como parte de otra tesis doctoral presentada como compendio de publicaciones.
Firma  Gijón, 25 de Enero de 2017



UNIVERSIDAD DE OVIEDO

Departamento de Informática

Programa de Doctorado en Informática

**RECONOCIMIENTO DE ATRIBUTOS FACIALES
MEDIANTE VISIÓN POR COMPUTADOR PARA LA
DETECCIÓN DE DISTRACCIÓN Y SOMNOLENCIA
EN CONDUCTORES**

TESIS DOCTORAL POR

Alberto Fernández Villán

6 de mayo de 2017

*De nuestros miedos nacen nuestros corajes, y en nuestras dudas
viven nuestras certezas. Los sueños anuncian otra realidad posible, y
los delirios otra razón. En los extravíos nos esperan los hallazgos
porque es preciso perderse para volver a encontrarse*

Eduardo Galeano

Agradecimientos

Me gustaría empezar agradeciendo el esfuerzo y dedicación empleados por mis directores de tesis, Rubén Usamentiaga y Rubén Casado, que me han sabido orientar, conducir y ayudar en todo momento. Quiero agradecer también a mi familia su apoyo y comprensión durante este tiempo. Sin ellos no hubiera sido posible. Al equipo de I+D+i de Fundación CTIC por permitir mi desarrollo profesional en un clima envidiable, dirigido con maestría por Antonio Campos, Vanesa Lobato y Eduardo Álvarez. Mis agradecimientos a la empresa TSK Electrónica y Electricidad S.A., donde actualmente realizo mi actividad investigadora, por posibilitar que ampliara mis miras profesionales accediendo a otros campos de investigación, relacionados o no con la temática de mi tesis. Agradecer también a Juan Luis Carús su dedicación y contribución. Finalmente a Sonia, por sus innumerables horas empleadas.

Resumen

La conducción es una actividad que requiere un alto grado de concentración por parte de la persona que la realiza, ya que el más pequeño descuido es suficiente para sufrir un accidente con las consiguientes pérdidas materiales y/o humanas. De acuerdo con el estudio más reciente publicado por la Organización Mundial de la Salud (OMS) en 2013, se estimó que 1.25 millones de personas mueren como resultado de accidentes de tráfico y entre 20 y 50 millones más sufren accidentes sin perder la vida pero pudiendo derivar en dolencias crónicas. Todas estas muertes y accidentes no sólo afectan de manera directa a los familiares de las víctimas sino que, además, tienen un alto coste sobre los presupuestos de los gobiernos, que se estima entre un 3 y un 5 % del producto interior bruto.

De todos estos accidentes, muchos son provocados por lo que se conoce como inatención. Este término engloba diferentes estados del conductor, como pueden ser la distracción y la somnolencia, siendo precisamente éstos los que más fatalidades ocasionan. Existen muchas publicaciones e investigaciones que intentan poner cifras que indiquen las consecuencias producidas por la inatención (y sus subtipos), pero no existe una figura exacta sobre los accidentes causados por la inatención puesto que todos estos estudios están realizados en diferentes lugares, diferentes marcos temporales, y por tanto, en diferentes condiciones. En líneas generales, se calcula que la inatención ocasiona entre el 25 % y el 75 % de los accidentes y casi-accidentes.

En un estudio realizado en 10 países europeos acerca de somnolencia y conducción se determinó que dicha conjunción incrementa el tiempo de reacción en un 86 % y es la cuarta causa de muerte en las carreteras españolas. Además, cabe destacar que el 75 % de los conductores españoles han sufrido episodios de somnolencia mientras conducían, porcentaje muy superior a la media del 47 % que han admitido este hecho. Además, otro factor importante a tener en cuenta es que, aunque los accidentes producidos por la somnolencia suelen ser muy graves (vistas las estadísticas anteriores de mortalidad), muchos conductores infravaloran esta situación y conducen aunque noten la presencia de sus síntomas. Bostezos frecuentes, cabeceos, visión borrosa, caída de párpados y esfuerzos por mantener tanto la atención como los ojos abiertos son signos habituales de somnolencia. Respecto a la distracción, éste es uno de los factores que más fatalidades ocasiona en España. De acuerdo con la Dirección General de Tráfico (DGT), la distracción es la primera causa detectada en los accidentes con víctimas, un 13.15 % de los casos. Los últimos datos y estadísticas arrojados por

inatención en los conductores españoles son dramáticos, siendo la causa que más víctimas mortales ocasiona (36 %), muy por encima de la velocidad inadecuada (21 %) o el nivel de alcoholemia (11 %).

A causa de estas cifras y sus consecuencias, la inatención se ha convertido en un campo ampliamente estudiado por la comunidad investigadora, donde los estudios y soluciones para luchar contra la distracción y la somnolencia, en particular, y la inatención, en general, se pueden clasificar en tres grandes grupos: 1) métodos basados en el análisis del comportamiento del vehículo, 2) métodos basados en el análisis de variables fisiológicas del conductor capturadas por diferentes sensores y 3) métodos basados en el análisis de características visuales del conductor mediante la captura de imágenes aplicando métodos de visión por computador, los cuales, por sus características, se han posicionado como una forma no intrusiva y eficaz para la detección tanto de la distracción como de la somnolencia.

La motivación de la presente tesis doctoral es consecuencia directa de las cifras anteriores y radica en ofrecer mecanismos para detectar y reducir los efectos ocasionados por la inatención en los conductores. Es por ello que el objetivo principal del presente trabajo sea alcanzar un mayor grado de conocimiento en todo lo relacionado con la inatención en los conductores, teniendo como fin último la reducción del número de accidentes y víctimas mortales ocasionados por esta causa haciendo uso de herramientas de la información y la comunicación (TIC).

A tal efecto, la investigación se ha centrado en el reconocimiento de atributos faciales, que puedan ser empleados a posteriori para una detección robusta de indicadores, que permitan una clasificación de episodios de distracción y somnolencia en conductores. En segundo lugar, la investigación ha intentado establecer un marco de referencia para caracterizar la distracción en los conductores con lo que futuras investigaciones tengan un punto de partida como referente. En tercer lugar, se ha propuesto, construido y validado una arquitectura basada en el análisis de características visuales mediante el empleo de técnicas de visión por computador y aprendizaje automático para la detección tanto de la distracción como de la somnolencia. En concreto, se propone una arquitectura de procesamiento especialmente diseñada para operar en entornos vehiculares, con una carga computacional muy baja y fácilmente integrable en dispositivos con reducidas capacidades de cómputo, capaz de lidiar con imágenes en distintas condiciones muy presentes en este tipo de entornos. El sistema de control propuesto integra varios elementos innovadores permitiendo que pueda operar de forma completamente autónoma para la detección robusta de los principales indicadores visuales, caracterizando tanto la distracción como la somnolencia del conductor. La arquitectura se ha validado, en primer lugar, con bases de datos de referencia validando los diferentes módulos que la componen y, en segundo lugar, con usuarios en entornos reales obteniendo, en ambos casos, unos resulta-

dos prometedores con una carga computacional adecuada para los dispositivos embebidos habituales en entornos vehiculares.

Abstract

Driving is an activity that requires a high degree of concentration on the part of the person who performs it since the slightest negligence is sufficient to provoke an accident with the consequent material and/or human losses. According to the most recent study published by the World Health Organization (WHO) in 2013, it was estimated that 1.25 million people die as a result of traffic accidents, whereas between 20 and 50 million do not die but consequences may result in chronic conditions. All these deaths and accidents not only have a direct impact on victims and families, but they also mean a high cost for government budgets, estimated at between 3 % and 5 % of their Gross Domestic Product (GDP).

Many of these accidents are caused by what is known as inattention. This term encloses a driver's different conditions such as distraction and drowsiness, which are, precisely, the ones that cause more fatalities. Many publications and research have tried to set figures that indicate the consequences of inattention (and its subtypes), but there is no exact number of the accidents caused by inattention since all these studies have been carried out in different places, different time frames and, therefore, under different conditions. Overall, it has been estimated that inattention causes between 25 % and 75 % of accidents and near-accidents. A study on drowsiness while driving in ten European countries found that fatigue risks increasing reaction time by 86 % and it is the fourth leading cause of death on Spanish roads. In addition, it is noted that 75 % of Spanish drivers have suffered episodes of sleepiness while driving, a much higher percentage than the average of 47 % who admitted this fact. In addition, another important factor to consider is that, although accidents caused by drowsiness are usually very serious (having regard to the abovementioned fatal statistics), many drivers underestimate this situation and drive even if they notice the presence of symptoms. Frequent yawning, pitching movements, blurred vision, drooping upper eyelids and efforts to keep both attention and eyes open are common signs of drowsiness. With respect to distraction, this is a major contributor to fatal accidents in Spain. According to the Directorate General of Traffic (DGT), distraction is the first violation found in fatal accidents, 13.15 % of the cases. The latest statistics on inattentive driving in Spanish drivers are alarming, appearing as the leading cause of fatalities (36 %), well above excessive speed (21 %) or alcohol consumption (11 %).

Because of these figures and their consequences, inattention has become a widely studied field by the research community, whose studies and solutions

to combat distraction and sleepiness, in particular, and inattention, in general, can be divided into three broad groups: 1) methods based on the analysis of the behavior of the vehicle, 2) methods based on the analysis of the driver's physiological variables captured by different sensors and 3) methods based on the analysis of the driver's visual characteristics by capturing images using computer vision methods, which, for their non-intrusive and effective characteristics, have become a leading way to detect both distraction and drowsiness.

The reason for this PhD thesis is the direct consequences of the abovementioned figures and its purpose is to provide mechanisms to help reduce driver inattention effects. That is why the main objective is to reach a better knowledge regarding driving inattention in order to reduce the number of accidents and fatalities. The main focus is that the extraction of facial attributes in a solid way could be used for a robust detection of indicators, which enables a classification of distraction and drowsiness episodes in drivers. Secondly, research has attempted to establish a frame of reference to characterize distraction in drivers in order to provide solid foundations for future research. Thirdly, an architecture based on the analysis of visual characteristics has been proposed, constructed and validated by using techniques of computer vision and automatic learning for the detection of both distraction and drowsiness. In particular, a processing architecture specially designed to operate in vehicular environments is proposed, with a very low computational load and easily embeddable into devices with reduced computational capacities in order to deal with images in the different conditions prevailing in this type of environments. The proposed control system integrates several innovative elements in order to operate in a completely autonomous way for the robust detection of the main visual indicators characterizing the driver's both distraction and drowsiness. The architecture has been validated, firstly, with reference databases testing the different modules that compose it, and, secondly, with users in real environments, obtaining in both cases, promising results with a suitable computational load for the embedded devices in vehicle environments.

Índice general

Índice de figuras	xiii
Índice de tablas	xv
1. Introducción	1
1.1. Motivación y objetivos	8
1.2. Contexto	10
1.3. Publicaciones	11
2. Discusión de resultados	13
3. Conclusiones y trabajo futuro	41
3.1. Aspectos futuros a considerar	44
4. Difusión de la tesis doctoral	45
4.1. Artículos	48
4.1.1. AVIUE Artificial Vision to Improve the User Experience .	48
4.1.2. Glasses detection on real images based on robust alignment	55
4.1.3. Unobtrusive Health Monitoring System Using Video-Based	
Physiological Information and Activity Measurements . .	69
4.1.4. Wearable and ambient sensors to health monitoring using	
computer vision and signal processing techniques	76
4.1.5. A Real-Time Big Data Architecture for Glasses Detection	
Using Computer Vision Techniques	87
4.1.6. Face Recognition and Spoofing Detection System Adapted	
To Visually-Impaired People	94
4.1.7. Driver Distraction Using Visual-Based Sensors and Algo-	
rithms	104
4.1.8. Sistema Automático Para la Detección de Distracción y	
Somnolencia en Conductores por Medio de Características	
Visuales Robustas	149
Bibliografía	173

Índice de figuras

1.1. Víctimas mortales en accidentes de tráfico en España	3
1.2. Principales factores que contribuyen a los accidentes en España en el año 2016	4
1.3. Aproximaciones para la detección de la inatención en los conductores	6
1.4. Entidades y organismos de financiación habilitadores de la presente tesis doctoral	10
1.5. Resumen de las publicaciones	11
2.1. Ejemplo de la extracción automática de atributos utilizando la aproximación LBP + SVM	15
2.2. Alineamiento de la región facial	17
2.3. Tasa de acierto en función del número de divisiones sin llevar a cabo el proceso de rotar las imágenes para lidiar con la pose (izquierda) y llevándolo a cabo (derecha)	19
2.4. Región normalizada para la detección de gafas	20
2.5. Representación visual de cómo afecta la resolución a las imágenes	20
2.6. Relación entre el tamaño de la imagen y la tasa de reconocimiento para diferentes cantidades de regiones en la construcción del histograma	21
2.7. Relación entre el tamaño de la imagen y la tasa de reconocimiento para diferentes cantidades de regiones en la construcción del histograma	22
2.8. Visualización de la región de la frente utilizada para la extracción de las variables fisiológicas	23
2.9. Procedimiento para la validación experimental de los algoritmos de visión por computador para la extracción de variables fisiológicas	24
2.10. Arquitectura Big Data para el procesamiento en <i>streaming</i> de imágenes faciales para la detección de gafas	27
2.11. Arquitectura de procesamiento para la detección de otro atributo facial (sonrisa) utilizando únicamente un <i>Bolt</i> de normalización diferente	27
2.12. Principales pasos en el procesamiento de las imágenes para la identificación facial y la detección de <i>spoofing</i>	30

2.13. Para las pruebas se selecciona un Lenovo ThinkPad X200 Tablet que se posiciona sobre el videoportero con la webcam en la parte inferior del mismo. Se puede ver un detalle de la imagen capturada	31
2.14. Ejemplos visuales del funcionamiento de los algoritmos de identificación facial y detección de spoofing realizados en el laboratorio	33
2.15. La de-identificación permite preservar detalles en la imagen manteniendo el anonimato del individuo	33
2.16. Principales puntos tratados en la séptima Contribución de la tesis	35
2.17. Principales puntos del sistema propuesto para la detección tanto de la somnolencia como de la distracción en conductores	38
4.1. Resumen de las publicaciones	45
4.2. Citas realizadas a los artículos de la presente tesis doctoral. Fuente: Google scholar actualizado con publicaciones todavía no indexadas	46

Índice de tablas

1.1. Objetivos concretos de la presente tesis doctoral	9
2.1. Principales algoritmos para la extracción automática de atributos faciales utilizando la aproximación LBP + SVM/Adaboost . . .	14
2.2. Principales trabajos para la detección de gafas	16
2.3. Resumen de los resultados para la extracción de las variables fi- siológicas	25
2.4. Resumen de los resultados de las pruebas realizadas con 25 usua- rios tanto para el reconocimiento facial como para la detección de <i>spoofing</i> durante 5 días consecutivos	31
2.5. Tiempos de ejecución para varios dispositivos de gama media/baja	32
2.6. Aproximaciones para la detección de la distracción en los conduc- tores	36

Capítulo 1

Introducción

La conducción es una actividad que requiere un alto grado de concentración por parte de la persona que la realiza, ya que hasta el más pequeño descuido es suficiente para sufrir un accidente con las consiguientes pérdidas materiales y/o humanas. De acuerdo con el último estudio publicado por la Organización Mundial de la Salud (OMS), que data del 2013, se estimó que 1.25 millones de personas mueren como resultado de accidentes de tráfico y entre 20 y 50 millones sufren accidentes sin perder la vida pero pudiendo derivar en dolencias crónicas [1]. Todas estas muertes y accidentes no sólo afectan de manera directa a los familiares de las víctimas, sino que, además, tienen un alto coste sobre los presupuestos de los gobiernos, que se estima entre un 3 y un 5 % del producto interior bruto [2]. Los accidentes de tráfico representan actualmente la novena causa de muerte entre todos los grupos de edad a nivel mundial y está previsto que, en 2030, escale hasta la séptima posición [1].

De todos estos accidentes, muchos son provocados por lo que se conoce como inatención. Este término engloba diferentes estados del conductor, como pueden ser la distracción y la somnolencia, siendo precisamente éstos los que más fatalidades ocasionan.

Existen muchas publicaciones e investigaciones que intentan poner cifras que indiquen la cantidad de accidentes y casi-accidentes producida por la inatención (y sus subtipos), pero no existe una cifra exacta, puesto que todos estos estudios están realizados en diferentes lugares, diferentes marcos temporales, y por tanto, en diferentes condiciones. En líneas generales, se calcula que la inatención ocasiona entre el 25 % y el 75 % de los accidentes y casi-accidentes [3–12, 12–14, 14–17].

Tanto la distracción como la somnolencia y, en líneas generales, la inatención han sido definidas de manera inconsistente y la relación entre ellas está poco clara [18], lo que puede ocasionar diversos problemas tanto a los investigadores como a los profesionales de la seguridad [19]. Dichas inconsistencias pueden llevar a que la comparación de los resultados no tenga sentido, e incluso, a la obtención de diferentes interpretaciones de los datos de los accidentes, obteniendo diferentes estimaciones ocasionadas por la inatención [20]. Uno de los trabajos que mejor define estos conceptos es el propuesto por [20], el cual des-

cribe una taxonomía para la inatención con sus diferentes subtipos y la define como *‘insuficiente o falta completa de atención a las actividades críticas para una conducción segura’*. En esta taxonomía: a) la somnolencia se engloba en el tipo *‘Conductor con Atención Restringida’* y es definida como *‘insuficiente o falta completa de atención a las actividades críticas para una conducción segura debido a factores biológicos en los que el conductor no es capaz de procesar la información crítica, como por ejemplo, en eventos de micro-sueños, parpadeos, o procesos de somnolencia’*, y b) la distracción se engloba en el tipo *‘Conductor con Atención Desviada’*, y es definida como *‘desviación de la atención de las actividades críticas para una conducción segura’*. Sin embargo, esta taxonomía presenta alguna deficiencia que se debe poner de manifiesto con vistas a futuros trabajos.

A pesar de la diversidad de datos acerca de las consecuencias de la inatención en conductores, se concluye que ésta es una de las causas que más mortandad ocasiona. En un estudio realizado en 10 países europeos sobre somnolencia y conducción se consideró que la fatiga incrementa el tiempo de reacción en un 86 % y es la cuarta causa de muerte en las carreteras españolas [21]. Además, cabe destacar que el 75 % de los conductores españoles han admitido haber sufrido episodios de somnolencia mientras conducían, porcentaje muy superior a la media del 47 % del resto de europeos, que también han admitido este hecho. Además, otro factor importante a tener en cuenta es que, aunque los accidentes producidos por somnolencia suelen ser muy graves (según las estadísticas anteriores de mortalidad), muchos conductores infravaloran esta situación y conducen aunque noten la presencia de sus síntomas. Bostezos frecuentes, cabeceos, visión borrosa, caída de párpados y esfuerzos por mantener tanto los ojos abiertos como la atención son signos habituales de somnolencia [21].

Respecto a la distracción, éste es uno de los factores que más fatalidades ocasiona en España. De acuerdo con la Dirección General de Tráfico (DGT), la distracción es la primera infracción detectada en los accidentes con víctimas mortales, con un 13,15 % de los casos [22]. Así mismo, la distracción es un problema que se ha visto altamente incrementado debido al uso de dispositivos móviles durante la conducción [23]. De manera más específica, Gregorio Serrano, director general de Tráfico, ha presentado el balance de siniestralidad vial del 2016 con estos datos generales: 1.038 accidentes mortales en vías interurbanas en los que 1.160 personas perdieron la vida y 5.067 necesitaron hospitalización a consecuencia de sus heridas. Estas cifras suponen aumentos del 1,4 % (+15) en accidentes mortales; 2,6 % (+29) en el número de fallecidos y 4,3 % (+209) en heridos hospitalizados [24]. Estos datos pueden verse de manera gráfica en la Figura 1.1. Supone la cifra más elevada desde 2012 y, además, el fin de 12 años de descenso continuado de la mortalidad, puesto que entre 2003 (3.993 muertos) y 2015 (1.131) se produjo una reducción de 2.862 víctimas mortales en el periodo de mayor caída de la siniestralidad desde que la Dirección General de Tráfico

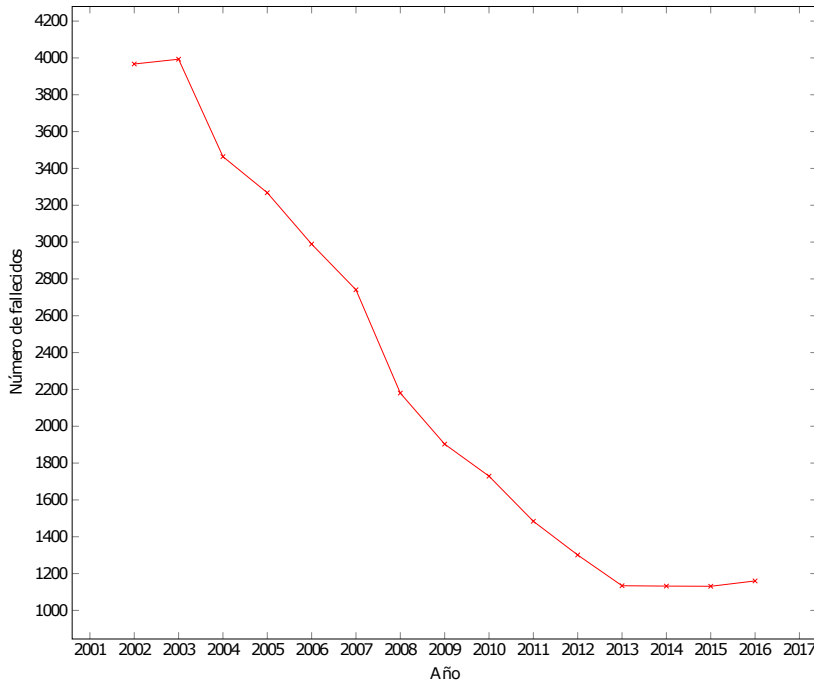


Figura 1.1: Víctimas mortales en accidentes de tráfico en España

(DGT) comenzara en 1960 a registrar datos de accidentes.

Los principales factores que contribuyeron a los accidentes mortales o graves en 2016 fueron la distracción (28 %), la velocidad inadecuada (21 %), el no respetar la prioridad (9 %), el cansancio o el sueño (8 %), el consumo de alcohol (11 %) y el mismo porcentaje en drogas. Por esto, se puede concluir que en España, la inatención fue la causante del 36 % de las víctimas mortales en accidentes de tráfico en el año 2016. Estos datos se pueden ver de manera más clara en la Figura 1.2.

La comunidad investigadora ha explorado vastamente este campo clasificando los diferentes estudios y soluciones para luchar contra la inatención en tres grandes grupos.

El primero de ellos se corresponde con los métodos basados en el comportamiento vehicular. Estos métodos detectan el estado del conductor analizando constantemente ciertas métricas, como pueden ser la posición del coche, los movimientos del volante, la presión del acelerador o del freno y el cambio de marchas (entre otros), y si en alguno se sobrepasa un determinado umbral, es probable que el conductor esté somnoliento o distraído [25–27]. En líneas generales, el principal

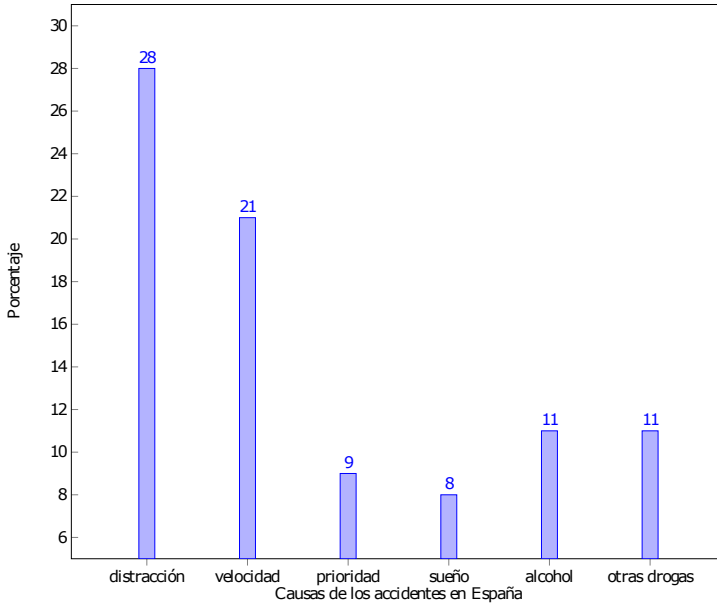


Figura 1.2: Principales factores que contribuyen a los accidentes en España en el año 2016

inconveniente de estos métodos es que su eficacia depende principalmente de las características individuales del vehículo, conductor y carretera [27–29]. Dentro de los métodos basados en el comportamiento vehicular, empiezan a desarrollarse alternativas que requieren la comunicación entre vehículos para operar correctamente [30].

El segundo de los grupos se basa en el análisis de variables fisiológicas, principalmente para la detección de la somnolencia. Son métodos muy robustos, pues permiten la detección de la somnolencia en sus fases tempranas con una baja tasa de falsos positivos [27]. En este grupo destacan los métodos basados en: a) electroencefalograma (EEG), b) electromiograma (EMG), c) electrocardiograma (ECG), d) electrooculograma (EOG). De entre todos ellos, el más común para la detección de la somnolencia es el EEG, donde se analizan diferentes bandas de frecuencia [27], incluyendo la banda delta (0.5-4Hz), que se corresponde con la actividad del sueño, la banda theta (4-8Hz), correspondiente a la somnolencia, la banda alpha (8-13Hz), que representa la relajación y la creatividad, y la banda beta (13-25Hz), que se corresponde con la alerta. Un decremento en la potencia de la señal en la frecuencia alpha y un incremento en la frecuencia theta indican la presencia de somnolencia [27]. Sin embargo, estos métodos requieren

contacto con el conductor para la realización de las medidas, lo que ocasiona que su implementación en entornos reales no sea ni lo más adecuado ni lo más práctico [27,31].

Finalmente, el tercero de los grupos se basa en el análisis de las características visuales que presenta un conductor distraído o bajo un estado somnoliento. Un conductor distraído se caracteriza por no mantener la atención puesta en la carretera, por lo que son continuos los movimientos de cabeza hacia ambos lados sin mantener fija la mirada. En cuanto a la somnolencia, las características visuales que la describen son muy variadas, incluyendo movimientos faciales, parpadeos rápidos y constantes, cabeceos y bostezos frecuentes. Los métodos basados en visión por computador para la caracterización del estado del conductor se han posicionado como uno de los más prometedores por proporcionar resultados bastante robustos y por su baja intrusión en la conducción.

En la mayoría de las ocasiones, estas características visuales de la somnolencia aparecen en espacios temporales diferentes y normalmente bien definidos [29]. De manera específica, los bostezos ocurren generalmente antes de que el conductor entre en somnolencia, mientras que, normalmente, los cabeceos ocurren cuando el conductor está empezando a dormirse. Es por ello que los métodos basados tanto en los bostezos como en los cabeceos no son capaces de detectar con exactitud cuando un conductor está empezando a estar somnoliento. Sin embargo, los métodos basados en obtener información de los ojos pueden detectar con precisión este punto, es decir, son los métodos visuales más adecuados para la detección de la somnolencia [32]. No obstante, puesto que existen esas diferencias temporales entre los distintos signos visuales, un punto importante a considerar puede ser la combinación de varias de estas características para aumentar la robustez final de la solución [27,29].

Recientemente, se han propuesto algoritmos de visión por computador para la obtención de variables fisiológicas, con lo que se abre un gran potencial en el campo de la detección de la inatención al combinarse las bondades de ambos tipos de procesamiento, ya que proporcionaría las bondades de los sistemas basados en la obtención de variables fisiológicas sin la necesidad de incluir sensores conectados al conductor. Sin embargo, es necesario profundizar en este nuevo campo para obtener su viabilidad en entornos reales.

En la Figura 1.3 se puede ver una tabla donde se recogen las aproximaciones antes comentadas para la detección tanto de la somnolencia, como de los principales tipos de distracción que existen (visual, manual y cognitiva). Así mismo, se recogen tanto las características de intrusividad que proporcionan las diferentes aproximaciones, como su efectividad para detectar los diversos episodios de los principales tipos de inatención.

La visión por computador es una disciplina científica que incluye métodos para adquirir, procesar, analizar y comprender las imágenes del mundo real con el fin de producir información numérica o simbólica y puedan éstas ser tratadas

Aproximación	Baja Intrusión	Somnolencia	Distracción		
			Cognitiva	Visual	Manual
Comportamiento vehicular	●	◐	◐	○	○
Variables fisiológicas (sensores)	○	●	◐	○	○
Variables fisiológicas (visión)	●	◐	◐	○	○
Análisis características visuales	●	◐	◐	●	◐

Figura 1.3: Aproximaciones para la detección de la inatención en los conductores

por un computador. Así pues, las imágenes juegan un papel fundamental, siendo la entrada que alimenta todo el proceso. La resolución de una imagen, que es el grado de detalle o calidad de una imagen digital, ya sea escaneada, fotografiada o impresa, es uno de los factores clave, puesto que indica la cantidad de detalles que puede observarse en ella. Sin embargo, a mayor resolución, más exigentes serán las capacidades de cómputo necesarias para realizar el procesamiento, ya que la cantidad de píxeles a analizar será mayor. En entornos vehiculares, se cuenta con espacio reducido para incluir elementos de cómputo, por lo que se requiere de dispositivos de procesamiento compactos que puedan ser ubicados en lugares estratégicos para interferir lo mínimo posible en la conducción. Además, no siempre se dispone de calidad adecuada en las imágenes, pues muchas son obtenidas por cámaras de seguridad con una calidad mediocre. Todo esto hace necesario profundizar en la relación existente entre los algoritmos de visión por computador y la resolución de la imagen de entrada.

Además de la resolución, existen aspectos más específicos que tienen influencia en el procesamiento de imágenes faciales en entornos vehiculares. En los sistemas para la detección de inatención en los conductores, uno de los puntos clave radica tanto en la detección de la cabeza del conductor como de la cara y su seguimiento a lo largo del tiempo para realizar un procesamiento posterior más avanzado, que permita extraer los síntomas que caracterizan a un conductor bajo un estado somnoliento o distraído. Esto hace que los algoritmos de *tracking* y detección facial sean considerados clave para una extracción robusta de estos síntomas.

Otro de los puntos a tener en cuenta es lo que se conoce como pose, que es la orientación y posición que un objeto (en este caso, una cara) presenta en relación a la cámara. Así pues, dependiendo de la pose, la apariencia que muestra la cara puede cambiar dramáticamente, reduciendo el rendimiento de los algoritmos de visión por computador. La pose es uno de los aspectos que mayor impacto causa, por lo que es necesario profundizar para mejorar el rendimiento global de los algoritmos.

Otro aspecto que decrementa sustancialmente el resultado de los algoritmos de procesamiento facial es la oclusión, que ocurre cuando un objeto se interpone

entre la región facial y la cámara. La consecuencia producida es que existirán zonas de la región facial tapadas por dicho objeto, imposibilitando su procesamiento posterior. En entornos vehiculares, uno de los objetos que mayor oclusión ocasiona son las gafas, pues, en muchos casos, imposibilitan la detección de los ojos, los cuales proporcionan información muy relevante para la detección tanto de la somnolencia como de la distracción. La detección de estos objetos, por tanto, es clave para un buen desempeño de los algoritmos.

Para que un sistema de ayuda a la conducción tenga una mayor aceptación por parte de los conductores, es aspecto fundamental que dicho sistema funcione de manera completamente automática, sin necesidad de cooperación alguna. Muchos de los sistemas propuestos para la detección de la inatención requieren en algún punto de su ejecución (calibración, obtención de imágenes de entrenamiento o validación) cooperación por parte del conductor para completar su operación. Es necesario, por tanto, automatizar al máximo estos sistemas para lograr una buena aceptación, proporcionando una baja intrusión a los usuarios.

Una vez los algoritmos de visión por computador han sido diseñados, el último paso es portarlos a un sistema embebido para ser incluidos dentro del habitáculo de conducción. Esto no es una tarea sencilla, pues una serie de aspectos claves debería ser tenida en cuenta; procesamiento en tiempo real, bajo coste, tamaño compacto o consumo eléctrico contenido son sólo algunos de estos factores. Todo esto hace necesario que se recojan y analicen para ser tenidos en cuenta a la hora de llevar a cabo dicha implementación en un entorno embebido.

El hecho de procesar un volumen de imágenes relativamente grande para cumplir con los requisitos de procesamiento en tiempo real pone de manifiesto una conexión entre el análisis de imágenes y el procesamiento en *streaming*, con lo que se abre una nueva ventana de investigación en todo lo relacionado con el procesamiento Big Data de imágenes. Proponer y validar una arquitectura de procesamiento streaming de imágenes es clave para poder analizar un volumen de imágenes en tiempo real.

La implementación de un algoritmo de visión por computador sólo representa una parte de todo el ciclo de vida del diseño de un producto. Una de las tareas más arduas consiste en validar el sistema completo en una amplia variedad de escenarios [33]. Esto hace necesario validar los algoritmos en diversas condiciones, teniendo en cuenta que los escenarios simulados tienen unas características muy diferentes a los escenarios reales, mostrando en ambos casos diferentes comportamientos en los conductores. Por todo esto, el hecho de validar los algoritmos para la detección de la inatención en los conductores es una tarea sobre la que es necesario profundizar.

En la presente introducción se han mostrado los aspectos más importantes para el desarrollo e implementación de un sistema para la detección de la inatención en los conductores, haciendo hincapié en aspectos y conceptos claves para la consecución de esta tarea. Este compendio de aspectos recoge desde los más

generales, como la categorización de la inatención, las cifras arrojadas o las carencias y deficiencias en las taxonomías propuestas, hasta los más específicos, como la resolución de la imagen, la pose facial o las oclusiones. Todos ellos serán abordados y serán objeto de análisis en la presente tesis, pues se corresponden con los objetivos que la vertebran y sustentan. Además, se comentan e identifican aspectos futuros, que podrían suponer una clara continuación en la temática del presente trabajo de investigación. Esta tesis se presenta como un compendio de publicaciones, organizada como sigue: resultados obtenidos y contribuciones aportadas en el capítulo 2, conclusiones y trabajo futuro en el capítulo 3. Finalmente, el capítulo 4 contiene los artículos que forman el compendio de publicaciones así como información sobre su impacto.

1.1. Motivación y objetivos

Los últimos datos y estadísticas arrojados por la inatención en los conductores españoles son importantes, siendo la causa que más víctimas mortales ocasiona (36 %), muy por encima de la velocidad inadecuada (21 %) o el consumo de alcohol (11 %) [23,24]. La motivación de la presente tesis doctoral es consecuencia directa de las cifras anteriores y radica en **ofrecer mecanismos para ayudar a contribuir a la disminución de los efectos ocasionados por la inatención en los conductores**. Es por ello, también, que su principal objetivo es **alcanzar un mayor grado de conocimiento en todo lo relacionado con la inatención en los conductores** para que tenga como fin último la reducción del número de accidentes y víctimas mortales en accidentes producidos por esta causa, haciendo uso de herramientas de la información y la comunicación (TIC). Para ello, **se propondrá, construirá y validará una arquitectura basada en la captación de atributos faciales y características visuales robustos mediante el empleo de técnicas de visión por computador y aprendizaje automático para la detección tanto de la distracción como de la somnolencia en los conductores**.

Para llevar a cabo el objetivo anterior, es necesario la consecución de objetivos más concretos, concisos y medibles, que abarquen todo el ciclo de desarrollo de una herramienta TIC para la detección de la inatención, desde el marco teórico que posibilite la base de conocimiento necesaria hasta los aspectos más concretos que deba cumplir la ya mencionada herramienta informática. Dichos objetivos concretos se exponen a continuación en la Tabla 1.1.

Objetivos de la presente tesis doctoral
1. Analizar la clasificación actual de la inatención en los conductores
2. Proponer mejoras en lo que a la clasificación de la inatención se refiere para concluir una taxonomía que palie las deficiencias encontradas
3. Estudiar y comparar los datos producidos por la inatención en los conductores a nivel mundial
4. Analizar los diferentes métodos que existen para la detección de la inatención en los conductores
5. Analizar y clasificar los métodos basados en visión por computador para la detección de la inatención
6. Profundizar en los algoritmos de visión por computador y aprendizaje automático para la extracción automática de atributos
7. Valorar el potencial de los algoritmos para la extracción automática de atributos
8. Proponer mejoras en el estado del arte para la detección de la inatención por medio de nuevos algoritmos en el campo de la visión por computador
9. Analizar el potencial de las variables fisiológicas obtenidas por medio de algoritmos de visión por computador para su aplicabilidad en entornos vehiculares
10. Estudiar la relación entre la resolución de una imagen y la precisión obtenida por un algoritmo de visión artificial en el caso concreto de atributos faciales
11. Profundizar en los algoritmos de <i>tracking</i> y detección facial en entornos vehiculares, pues son la base de los sistemas para la detección de la inatención en los conductores mediante técnicas de visión por computador
12. Proponer mejoras para reducir el impacto de la pose de la cabeza del conductor en los algoritmos de visión por computador
13. Detectar la presencia de gafas, pues es un punto muy importante a incorporar en los algoritmos para el procesamiento de la inatención por la oclusión que ocasionan
14. Automatizar los sistemas de ayuda a la conducción, evitando la necesidad de colaboración por parte del conductor
15. Analizar los aspectos a considerar para la implementación de los algoritmos de visión por computador en entornos vehiculares
16. Investigar en arquitecturas de <i>streaming</i> para el procesamiento de grandes volúmenes de imágenes faciales en tiempo real
17. Investigar en métodos comerciales para la detección de la inatención en conductores mediante técnicas de visión por computdor
18. Investigar en la comparativa entre métodos comerciales y no comerciales para la detección de la inatención en conductores mediante visión por computaor
19. Aumentar la seguridad en los sistemas para la detección de la inatención (reconocimiento facial, detección de <i>spoofing</i> , bases de datos)

Tabla 1.1: Objetivos concretos de la presente tesis doctoral



Figura 1.4: Entidades y organismos de financiación habilitadores de la presente tesis doctoral

1.2. Contexto

El contexto en el que se han desarrollado las actividades de investigación que han dado lugar a la presente tesis doctoral han sido llevadas a cabo en: 1) Fundación CTIC - Centro Tecnológico y 2) TSK Electrónica y Electricidad, S.A., siempre bajo el paraguas de las diversas convocatorias, tanto regionales como nacionales financiadas parcialmente por diversos organismos y entidades (Figura 1.4), facilitando la ejecución de proyectos de investigación. En el caso de la Fundación CTIC, se han llevado a cabo varios proyectos de investigación (algunos finalizados y otros todavía en ejecución). Estos proyectos están muy relacionados con la presente tesis doctoral, los cuales se comentan a continuación:

- **NINE** (Natural Language Interaction Engine). Aumentar la capacidad de interacción de las máquinas interpretando el lenguaje oral y corporal de las personas. Referencia: IE09-092C1. Entidad financiadora: PCTI. Plazo de ejecución: 2010 - 2012.
- **CAVIAR** (Creacion de algoritmos de visión artificial). Desarrollo de algoritmos para la obtención automática de atributos faciales en dispositivos móviles. Referencia IE09-511. Entidad financiadora: FICYT. Plazo ejecución: 2012 - 2013.
- **VISOR** (VISion Object Recognition). Framework de reconocimiento de objetos adaptado a los dispositivos de realidad aumentada. Referencia: CT14-04-2-01. Entidad financiadora: PCTI. Plazo ejecución: 2014 - 2015.
- **emoPLAY**. Solución tecnológica para el tratamiento del Trastorno del Espectro Autista mediante la detección de la emoción del usuario. Referencia: Soluciones tecnológicas aplicadas al autismo 2015 Fundación Orange. Entidad financiadora: Fundación Orange. Plazo de ejecución: 2015 - 2016.

Además, en la empresa TSK se han desarrollado proyectos que han permitido generar el conocimiento necesario para abordar ciertas temáticas también

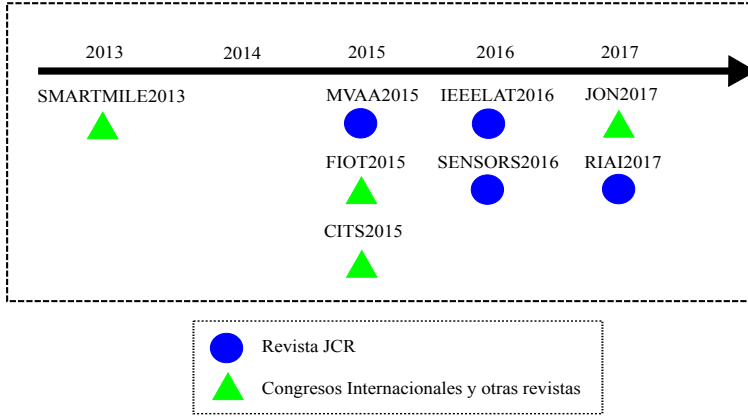


Figura 1.5: Resumen de las publicaciones

importantes en la tesis, como puede ser el procesamiento en *streaming* de la información:

- **SISFOTÓN**. Monitoriza de forma remota numerosos procesos industriales a través de una innovadora plataforma de tratamiento de la información basada en técnicas estadísticas y de inteligencia artificial, mediante la incorporación de tecnologías de tratamiento masivo de datos (Big Data), almacenamiento NoSQL (bases de datos no relacionales) o procesamiento distribuido, entre otros. Referencia: IDI-20130200. Entidad financiadora: CDTI. Plazo ejecución: 2013 - 2014.

1.3. Publicaciones

Como se ha comentado, esta tesis se presenta como compendio de publicaciones. Se incluyen, por tanto, revistas indexadas en el JCR. También se incluyen artículos presentados en congresos internacionales y en otras revistas. Estos últimos permiten ampliar la visión general de la labor investigadora llevada a cabo durante todos estos años. En la Figura 1.5 se puede ver la clasificación de las publicaciones según la división antes comentada. Esta Figura puede aportar una idea del alcance de la presente tesis doctoral. A continuación, se adjuntan en forma de lista las publicaciones referenciadas en la Figura anterior y que se comentarán en capítulos sucesivos:

- **SMARTMILE2013**: Losada, D. G., Lopez, G. A. R., Acevedo, R. G., & Villan, A. F. (2013, December). AVIUE Artificial vision to improve the

- user experience. In *New Concepts in Smart Cities: Fostering Public and Private Alliances (SmartMILE)*, 2013 International Conference on (pp. 1-6). IEEE.
- **MVAA2015**: Fernández, A., García, R., Usamentiaga, R., & Casado, R. (2015). Glasses detection on real images based on robust alignment. *Machine Vision and Applications*, 26(4), 519-531.
 - **FIOT2015**: Fernández, A., Casado, R., & Usamentiaga, R. (2015, August). A Real-Time Big Data Architecture for Glasses Detection Using Computer Vision Techniques. In *Future Internet of Things and Cloud (Fi-Cloud)*, 2015 3rd International Conference on (pp. 591-596). IEEE.
 - **CITS2015**: Fernández, A., Carús, J. L., Usamentiaga, R., Alvarez, E., & Casado, R. (2015, July). Unobtrusive health monitoring system using video-based physiological information and activity measurements. In *Computer, Information and Telecommunication Systems (CITS)*, 2015 International Conference on (pp. 1-5). IEEE.
 - **IEEEELAT2016**: Villan, A. F., Candas, J. L. C., Fernandez, R. U., & Tejedor, R. C. (2016). Face recognition and spoofing detection system adapted to visually-impaired people. *IEEE Latin America Transactions*, 14(2), 913-921.
 - **SENSORS2016**: Fernández, A., Usamentiaga, R., Carús, J. L., & Casado, R. (2016). Driver distraction using visual-based sensors and algorithms. *Sensors*, 16(11), 1805.
 - **JON2017**: Fernández, A., Carús, J. L., Usamentiaga, R., Alvarez, E., & Casado, R. Wearable and ambient sensors to health monitoring using computer vision and signal processing techniques. *Journal of Networks*. 2017.
 - **RIAI2017**: Fernández, A., Usamentiaga, R. & Casado, R. Sistema Automático Para la Detección de Distracción y Somnolencia en Conductores por Medio de Características Visuales Robustas. *Revista Iberoamericana de Automática e Informática Industrial RIAI*. 2017.

Capítulo 2

Discusión de resultados

La visión por computador y las técnicas de aprendizaje automático son claves para la detección de la inatención en los conductores y, al mismo tiempo, son la base en la presente tesis. En este punto, la obtención automática de atributos juega un papel fundamental en la visión por computador, pues la información que se puede obtener de manera no intrusiva para el usuario presenta un gran potencial aplicable a multitud de campos. Atributos como el género, edad, etnia, emoción e identificación facial, entre otros, se pueden obtener usando técnicas de visión por computador de manera no intrusiva. Además, el reconocimiento automático de atributos permite la extracción de los principales signos de inatención [34]. La base para la extracción automática de atributos radica en una obtención robusta de la información de la cara del conductor. Para llevar a cabo este paso existen diferentes alternativas, entre las que destaca el operador Local Binary Pattern (LBP) [35], un tipo de descriptor visual usado ampliamente en la visión por computador, pues se caracteriza por su robustez, por su facilidad en la implementación y por su baja carga computacional. Es habitual utilizar conjuntamente Support Vector Machines (SVM) o Adaboost en la etapa de clasificación, debido a que en muchos casos los descriptores visuales obtienen vectores con alta dimensionalidad que las arriba mencionadas herramientas de aprendizaje automático son capaces de clasificar. En la Tabla 2.1 se puede observar una colección de algoritmos, considerados referentes, que hacen uso de esta aproximación (LBP + SVM/Adaboost) para la extracción automática de atributos.

Con el objetivo de validar la aproximación anterior, se realizó una implementación del operador LBP para la etapa de extracción de características y en la etapa de clasificación se hizo uso de SVM, mediante la librería LIBSVM [45]. La idea de utilizar el operador LBP está motivada por el hecho de que una cara puede ser asociada a una composición de micro-patrones, los cuales son descriptores con robustez por dicho operador [44]. Los detalles de dicha implementación pueden verse en la primera Contribución de la tesis doctoral [46]. Para valorar el posible potencial de dichos algoritmos, el caso de uso identificado radica en su aplicación para la personalización de servicios utilizando datos abiertos del

Atributo	Aproximación
Género	(LBP + SVM): Makinen et al. [36] investigan en la extracción automática de género incluyendo el pelo de los usuarios además del resto de la región facial
Género	(LBP + SVM): Lian et al. [37] llevan a cabo la extracción de género incluyendo imágenes con una pose acusada
Género	(LBP + SVM): Shan et al. [38] proponen la extracción de género utilizando imágenes reales. Utilizan Adaboost para la selección de los patrones LBP más significativos
Género	(LBP + Adaboost): Yang et al. [39] hacen uso de Adaboost en la etapa de clasificación para la clasificación de género
Género	(LBP + SVM): Lyle et al. [40] realizan una clasificación de género basada en la región periocular
Edad	(LBP + Adaboost): Yang et al. [39] emplean Adaboost para la obtención de diferentes grupos de edades (niño/adulto/anciano)
Etnia	(LBP + Adaboost): Yang et al. [39] realizan una clasificación binaria de la etnia (asiática/no asiática)
Etnia	(LBP + SVM): Lyle et al. [40] realizan una clasificación de la etnia (asiática/no asiática) basada en la región periocular
Emoción	(LBP + SVM): Shan et al. [41] proponen una clasificación de las emociones prototípicas
Emoción	(LBP + SVM): Shan et al. [42] llevan a cabo la implementación de una solución muy similar al expuesto anteriormente
Identificación	(LBP + comparación de histogramas): Woodard et al. [43] realizan la identificación facial a partir de la región periocular
Identificación	(LBP + SVM): Ahonen et al. [44] realizan la identificación facial considerando toda la región

Tabla 2.1: Principales algoritmos para la extracción automática de atributos faciales utilizando la aproximación LBP + SVM/Adaboost

Ayuntamiento de Gijón. Una representación gráfica puede verse en la Figura 2.1.

Al tratarse de una implementación relativamente sencilla y tener margen de mejora en muchas etapas del procesamiento, se sigue investigando en este punto para obtener una información más precisa. Se empieza, por tanto, a profundizar en determinadas etapas de los ya mencionados algoritmos que se determinaron claves para, por un lado, aumentar su robustez, y por otra, para obtener un mayor conocimiento de todo el procesamiento. Se detecta una carencia en el estado del arte en lo relacionado con la detección de gafas, cuyos campos de aplicación son bastante numerosos e importantes: marketing y publicidad, sistemas de seguridad o identificación basada en *soft biometrics*, entre otros. Éste, es un punto importante, pues la obtención de signos de fatiga en los ojos del conductor puede verse dificultada si el individuo lleva gafas [34]. Por esta razón, un método para la detección de gafas puede ayudar a los sistemas de monitorización de los

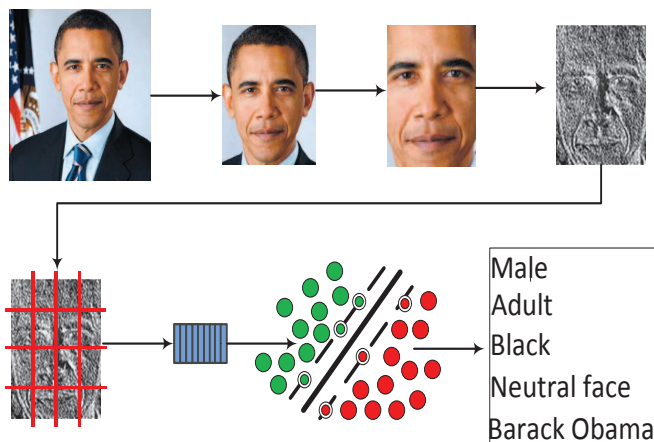


Figura 2.1: Ejemplo de la extracción automática de atributos utilizando la aproximación LBP + SVM

conductores.

Los principales trabajos para la detección de gafas se pueden ver en la Tabla 2.2. Como parte de la presente tesis, se presenta en [47] la segunda Contribución un método para este fin. Como principales innovaciones, caben destacar varios aspectos.

El primero de ellos es la implementación de un algoritmo robusto para la detección de gafas. Además, se realiza un alineamiento de la cara con el objetivo de lidiar con la pose, pues es uno de los factores que decrementan el rendimiento de los sistemas de procesamiento facial. También se lleva a cabo la implementación y validación de una mejora del operador LBP, el cual proporciona una tasa de reconocimiento más alta y se presenta más robusto frente al ruido, habitual en imágenes con baja resolución. Por último, también se investiga en varios factores que pueden influir en la obtención de un operador más eficiente computacionalmente sin decrementar la tasa de reconocimiento, como son el tamaño de la imagen y el número de divisiones del operador LBP.

A continuación, se comentan con mayor detalle los aspectos introducidos con anterioridad.

En primer lugar, se propone un algoritmo de procesamiento que localiza la región facial por medio del método propuesto por Viola y Jones [56]. Luego, se realiza un pre-procesamiento de la región facial para lidiar con aspectos como la pose, la rotación, el escalado y otras inexactitudes en la detección facial. A continuación, se emplea un algoritmo para establecer la región de la cara

Autor	Aproximación
Jiang et al. [48]	Primer algoritmo propuesto para la detección de las gafas en imágenes faciales. Los resultados se reportan de manera similar en [49]
Dantcheva et al. [49]	Entrenamiento: base de datos FERET. El testing está realizado con un subconjunto de imágenes de la base de datos FERET. Tasa de detección: 87.7 %. Tasa de falsos positivos: 7.17 %. Tasa de faltos negativos: 5.66 %
Jing et al. [50]	Testing con un conjunto de 419 personas, 151 de ellas portando gafas. Tasa de detección: 99.52 %, presentando 2 falsos positivos
Vaquero et al. [51]	1044 imágenes obtenidas de Internet para el entrenamiento (308 gafas de sol, 318 gafas, 418 sin gafas). Validado con un video de videovigilancia. Gafas de sol (precisión: 85 %, exhaustividad: 35 %), Gafas (precisión: 90 %, exhaustividad 8.9 %), Sin gafas (precisión: 77 %, exhaustividad: 99.5 %)
Wu et al. [52]	3000 imágenes (1500 con gafas y 1500 sin ellas) obtenidas de FERET y de Internet utilizando validación cruzada. Los mejores resultados fueron obtenidos mediante características de Gabor. Precisión: 98.9 % (98.1 % con gafas y 99.7 % sin ellas). Utilizando imágenes reales para el testeo, obtuvieron una tasa de acierto del 94.0 %
Wu et al. [53]	Para entrenar el sistema emplearon un total de 513 imágenes faciales de 19 personas utilizando 3 tipos de gafas diferentes y 9 poses faciales diferentes. Tasa de detección: ≈ 90 %
Shan et al. [54]	Para entrenar utilizaron 100 imágenes de la base de datos Bern. Para validación, utilizaron 200 imágenes de la misma base de datos, obteniendo una tasa de acierto del 77 %
Wu et al. [55]	Entrenaron un detector para la región de los ojos basado en el detector facial propuesto por Viola y Jones [56] empleando 12710 ejemplos. Validaron el detector con 1386 imágenes faciales. Tasa de detección: 96 % y una falsa alarma de 10^{-4}
Heo et al. [57]	Emplearon una base de datos de imágenes termográficas (NIST), obteniendo una tasa de acierto del 86.6 % para los usuarios que llevaban gafas. Para las personas que no llevaban gafas, obtuvieron un tasa de acierto del 97.1 %

Tabla 2.2: Principales trabajos para la detección de gafas

alrededor de los ojos a partir de la cual se aplica el operador LBP mejorado (RLBP) [58] para obtener el vector de características. En último lugar, se aplica la etapa de clasificación, la cual se realiza por medio de una SVM, clasificando el histograma en dos clases. Como se ha comentado con anterioridad, las SVMs son una herramienta muy útil para la clasificación de datos, en general, y para tareas de clasificación facial, en particular [59].

A continuación, se resumen los resultados recogidos en la mencionada publicación [47].

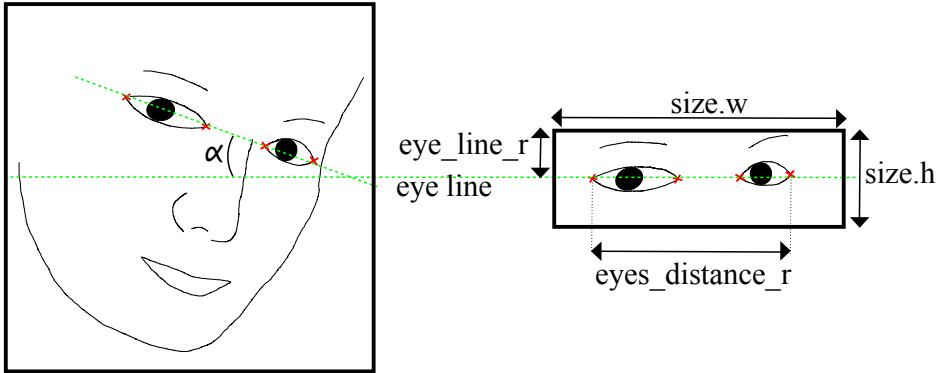


Figura 2.2: Alineamiento de la región facial

Una de las carencias detectadas en muchos algoritmos de visión por computador es la falta de claridad de las imágenes utilizadas tanto para entrenar como para validar el sistema. En este caso, se emplea la base de datos LFW [60]. Esta base de datos fue creada para estudiar el problema del reconocimiento facial de imágenes reales, la cual contiene 13 233 imágenes faciales recogidas de 5 749 personas diferentes recogidas de la Web. Esta base de datos ha sido utilizada recientemente para la comparación de atributos faciales [38, 61]. En este trabajo de la tesis [47] se emplean 3 000 imágenes faciales, de las cuales 1 500 presentan gafas y 1 500 no. Todos los resultados que se muestran a continuación se obtuvieron mediante validación cruzada en 5 subconjuntos (*5-fold cross-validation*) con el objetivo de usar prácticas estandarizadas [38].

El algoritmo de preprocesamiento permite aumentar la robustez en la detección de las gafas de media en un 1.55 %, tal y como se puede ver de manera gráfica en la Figura 2.2. Este algoritmo permite enfrentarse en gran medida con la pose de la cara del usuario.

Una vez corregida la pose, el siguiente paso es el de obtener automáticamente la región facial de los ojos. Para ello se propone el Algoritmo 1, que extrae de las imágenes faciales dicha región de manera totalmente automática. El mencionado algoritmo está basado en el propuesto por Makinen et al. [36], pero eliminando la componente manual que requería éste. De esta manera, se consigue extraer la región facial normalizada alrededor de los ojos de manera totalmente automatizada para posteriormente extraer el vector de características mediante el operador LBP. Para crear dicho vector, la región facial normalizada se divide en m regiones más pequeñas, no superpuestas $\{R_0, \dots, R_{m-1}\}$. Para cada una de estas regiones, se crea el histograma LBP, que se corresponde con las concatenaciones de los histogramas para cada una de las regiones, obteniendo

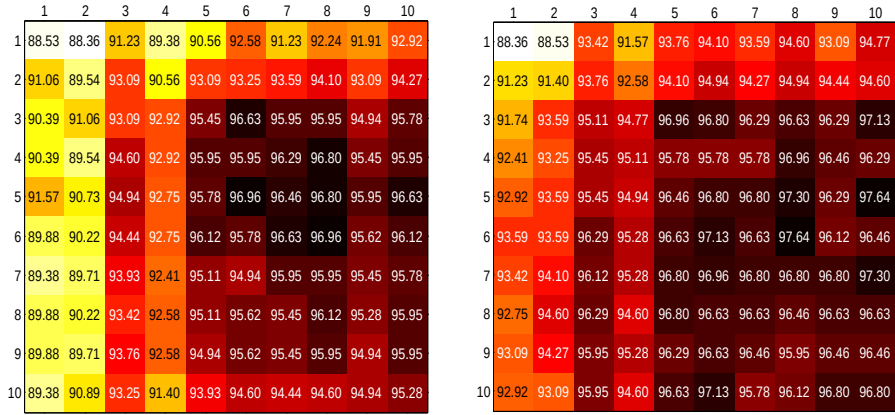
Algorithm 1 Normalización facial del ROI

eyes_distance_r, eye_line_r, size

- 1: Se localizan 8 puntos faciales. Se calcula recta de regresión basada en los 4 puntos de los ojos. El ángulo α de desviación es calculado y la imagen se rota teniendo en cuenta dicho ángulo para alinear las imágenes
 - 2: La distancia Euclídea d_0 se calcula entre los ojos en la imagen rotada
 - 3: La distancia entre los ojos en la imagen redimensionada se calcula según la fórmula $d_t = size.w * eyes_distance_r$
 - 4: El ratio r se calcula como $r = d_0/d_t$
 - 5: La anchura w_0 and la altura h_0 de la región alrededor de los ojos se calcula como $w_0 = r * size.w$ y $h_0 = r * size.h$
 - 6: Las coordenadas de las esquinas de la región facial en la imagen rotada se calculan como $x_l = x_e - w_0/2$, $y_t = y_e - h_0/eye_line_r$, $x_r = x_l + w_0$ y $y_b = y_t + h_0$, donde x_l es la coordenada x del borde izquierdo, x_e es la coordenada x del punto medio entre los ojos, y_t es la coordenada y del borde superior, y_e es la coordenada y de los ojos, x_r es la coordenada x del borde derecho e y_b es la coordenada y del borde inferior. Este *ROI* es extraído de la imagen
 - 7: **return** *ROI*
-

un *histograma espacial extendido* [44]. La longitud total del histograma es de $B = 59 * m$, puesto que el operador LBP genera 59 valores diferentes en función de los diferentes micro-patrones encontrados en la imagen. Por esto, el número m de regiones en las que se divide la región facial juega un papel importante que conviene estudiar en detalle. Un número de regiones pequeño permite un histograma más compacto pero, por el contrario, se pierde información espacial. Un número m de regiones más elevado, permite obtener mayor información espacial pero, por contra, el histograma generado presenta un número de características mayor. Esto puede ocasionar que la etapa de clasificación (mediante SVM) no pueda mapear con precisión los vectores a las clases correspondientes. En la Figura 2.3 se puede ver de manera gráfica un estudio exhaustivo donde se puede apreciar como varía la tasa de acierto en la clasificación de las gafas en función de las diferentes regiones. En la Tabla de la izquierda se parte de la región sin alinear (sin rotar las imágenes en el ángulo α de desviación). En la Tabla de la derecha, por el contrario, la imagen se rota en función del ángulo α de desviación.

En función de la gama de colores se puede apreciar que se producen mejores resultados si las imágenes han sido alineadas previamente (de media un 1.55 %). Sin embargo, llega un momento en que no se producen mejores reconocimientos a pesar de aumentar el número de regiones. Además, cuanto más compacto sea el vector de características, menos recursos computacionales y memoria serán



(a) Tasa de acierto en función del número de divisiones (imagen sin rotar) (b) Tasa de acierto en función del número de divisiones (imagen rotada)

Figura 2.3: Tasa de acierto en función del número de divisiones sin llevar a cabo el proceso de rotar las imágenes para lidiar con la pose (izquierda) y llevándolo a cabo (derecha)

necesarios.

Otro punto importante es la resolución de las imágenes de entrada al algoritmo. En aplicaciones de videovigilancia y sistemas de monitorización de usuarios las imágenes suelen presentar una baja resolución [62], así que es necesario realizar pruebas para establecer una resolución mínima. Para llevar a cabo este procedimiento, se tienen en cuenta las 3000 imágenes de la base de datos LFW y se obtiene el tamaño medio de la región facial. La distancia media entre los puntos exteriores de los ojos fue de 63.86 pixels, con una desviación estándar de 5.39. Para establecer la región facial se tiene en cuenta que las gafas habitualmente se colocan a la misma altura tanto arriba y abajo, como de derecha a izquierda del centro de los ojos. La altura, por tanto, se establece teniendo en cuenta el tamaño de las gafas. El ancho está determinado teniendo en cuenta la distancia entre los ojos y la forma de los lados de las gafas. Tras realizar varios experimentos, se establece que el ancho de la región normalizada de las gafas es un 122 % más grande que la altura. En la Figura 2.4 se pueden ver varios ejemplos de la zona que se emplea para la detección de las gafas.

Teniendo en cuenta los datos anteriores, las pruebas realizadas para establecer la influencia de la resolución presentan una anchura máxima de 100 píxeles y 45 píxeles de altura. Así, el tamaño máximo de la región normalizada de las gafas es de 4500 píxeles. El tamaño mínimo considerado es de 16 píxeles y 7 píxeles

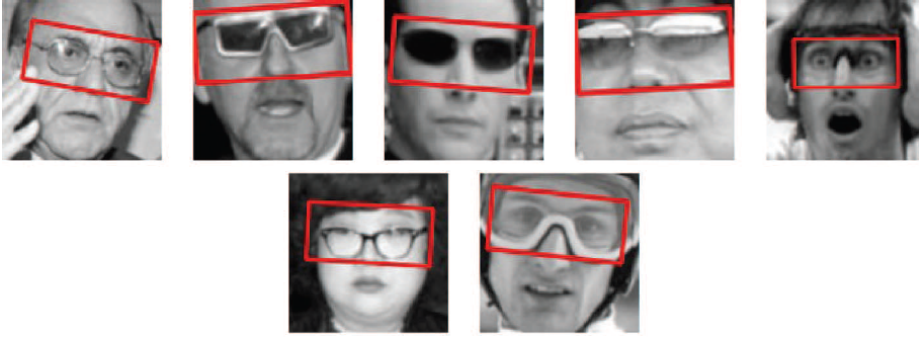


Figura 2.4: Región normalizada para la detección de gafas



Figura 2.5: Representación visual de cómo afecta la resolución a las imágenes

de anchura y altura, respectivamente. En la Figura 2.5 se puede ver, para una misma región normalizada de las gafas, cómo influye la resolución.

En la Figura 2.6 se puede ver la tasa de reconocimiento en función del tamaño de la región facial de las gafas para diferentes configuraciones en el número de divisiones del operador LBP. Como se puede apreciar, existe una relación logarítmica entre el tamaño y la tasa de reconocimiento.

Esta relación logarítmica depende del número de divisiones del operador LBP en la región facial. En la Figura 2.7 se pueden ver dos ejemplos. La siguiente lista de valores $R^2 = \{0.9880, 0.9875, 0.9826, 0.9812, 0.9897, 0.9880, 0.9876, 0.9891, 0.9915, 0.9898\}$ se corresponde con los valores de los coeficientes de correlación obtenidos de las diferentes series de valores de la Figura 2.6.

Es la primera vez que se emplea la base de datos LFW para la comparación de los algoritmos de detección de gafas. Por lo tanto, la comparación con el algoritmo aquí presentado es imposible. Pero con el objetivo de mostrar que el algoritmo propuesto presenta un claro avance en el estado del arte, el algoritmo también fue comparado haciendo uso de la base de datos FERET [63], pues alguno de los algoritmos para la detección de gafas utilizaron dicha base de datos [49,52]. Dantcheva et al. [49] obtienen una tasa de detección del 87.7 %. Wu et al. [52], por su parte, obtienen una tasa del 98.9 %, utilizando no sólo imágenes de la base de datos FERET sino también imágenes de Internet. El algoritmo

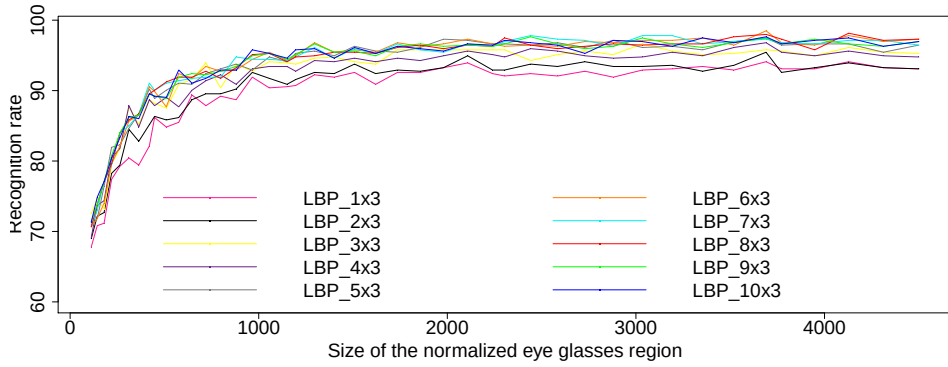


Figura 2.6: Relación entre el tamaño de la imagen y la tasa de reconocimiento para diferentes cantidades de regiones en la construcción del histograma

aquí propuesto obtiene unas tasas de reconocimiento del 99.89 % y del 99.83 % utilizando RLBP y LBP respectivamente, mejorando los resultados de algoritmos anteriores. Por otro lado, los estudios realizados para comprender como influyen tanto el número de divisiones en el operador LBP, como el alineamiento previo de la región facial y la resolución de la imagen de entrada, permiten obtener un mayor conocimiento y llegar a un *trade-off* de los diversos parámetros antes comentados. En lo que a la resolución de la imagen de entrada se refiere, se ha detectado un estudio realizado por Shan et al. [41, 42] para ver cómo influye el tamaño de la imagen de entrada y el reconocimiento automático de emociones. Sin embargo, este estudio se limita a únicamente emplear 6 tamaños diferentes, mientras que en la Contribución anteriormente comentada, se evalúan un total de 43 resoluciones diferentes.

A medida que se va profundizando en la temática de la detección de inatención en los conductores mediante técnicas de visión por computador y aprendizaje automático, se evidencia que ciertas variables fisiológicas obtenidas del conductor pueden ser muy valiosas para la detección de la inatención y caracterización de su estado. Por ejemplo, el pulso cardíaco tiende a incrementarse cuando la carga cognitiva aumenta [64]. Esto puede ayudar a caracterizar mejor la situación del conductor ante situaciones de estrés [65–67]. Por medio de la información obtenida por el pulso cardíaco, el estado cognitivo del conductor puede ser monitoreado [68]. Apparies et al. [65] evidenciaron que el ritmo cardíaco y su variabilidad pueden servir como indicadores tempranos de fatiga en el conductor. Zhao et al. [69] comprobaron en un simulador que durante una situación de estrés el pulso cardíaco fluctúa violentamente.

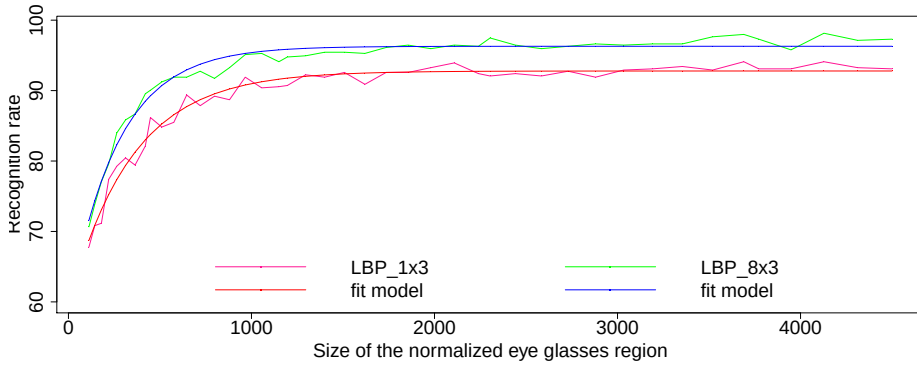


Figura 2.7: Relación entre el tamaño de la imagen y la tasa de reconocimiento para diferentes cantidades de regiones en la construcción del histograma

Con el avance de la visión por computador, empiezan a proliferar algoritmos, los cuales intentan extraer las variables fisiológicas antes comentadas mediante análisis de imágenes faciales en tiempo real [70, 71]. En este sentido, se dedican recursos y esfuerzos para profundizar con el objetivo de validar la potencial aplicabilidad de este tipo de algoritmos para un sistema de detección de inatención en los conductores.

La tercera Contribución de la tesis [72] tiene por objetivo la creación de un sistema de monitorización del estado de salud de un usuario mediante técnicas de visión por computador. Además, el sistema se completa con información de su actividad por medio de un reloj de pulsera. Usando técnicas de visión por computador se extraen las diferentes variables fisiológicas: a) ritmo cardíaco, b) variabilidad del ritmo cardíaco y c) ritmo respiratorio. Previamente, Lewandowska et al. [73] llevan a cabo una serie de experimentos con la conclusión de que la frente de los usuarios es una zona representativa de toda la región facial. Sin embargo, no proponen ningún método para detectar dicha región de forma automática. Alexander et al. [74], por un lado, y Rouast et al. [75], por otro, llegan a las mismas conclusiones, es decir, la región de la frente es perfectamente válida, pero tampoco emplean métodos robustos para su detección, pues sus algoritmos se basan en operaciones matemáticas básicas a partir de la información de la posición devuelta por el algoritmo de Viola y Jones para la detección facial. En esta Contribución [72], se propone un método de detección y localización de la región de la frente de forma totalmente automática para después proceder a procesarla y extraer las variables fisiológicas finalmente. En la Figura 2.8 se pueden observar varios ejemplos de la mencionada zona. Para la localización

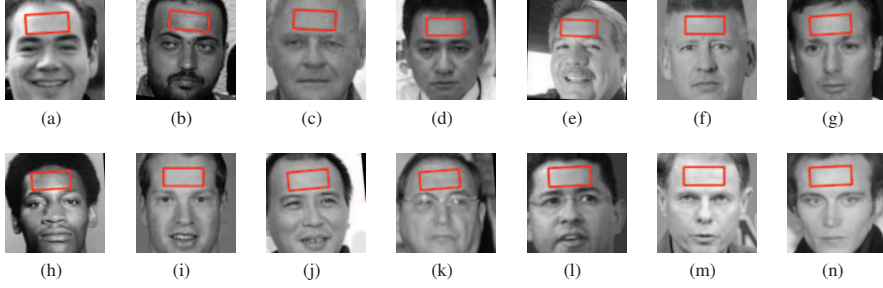


Figura 2.8: Visualización de la región de la frente utilizada para la extracción de las variables fisiológicas

de la región de interés (ROI, por sus siglas en inglés) de la frente se utiliza el método previamente aplicado para ubicar el ROI de la región de las gafas, pero modificado para posicionar el ROI correctamente (ver Figura 2.8). Las imágenes con las que se realizan pruebas para validar el método de extracción de la región de la frente están obtenidas de la base de datos LFW. Como se puede observar, el método la localiza de forma robusta. El siguiente paso radica en la extracción de las variables fisiológicas a partir del ROI de la frente.

En esta publicación sólo se lleva a cabo la validación del ritmo cardíaco utilizando las métricas y procedimientos comunes a publicaciones previas [70, 73, 76]. Para llevar a cabo dicha validación, una *webcam* se coloca en la parte superior del espejo y está conectada a una unidad de procesamiento. Todos los vídeos están registrados en color (24-bit RGB con 3 canales x 8 bits/canal) a 20 *fps* con una resolución de 640x480. Doce participantes (6 hombres y 6 mujeres) con una edad entre 26 y 35 años (media = 29, SD = 2.30) fueron involucrados en el experimento. Al mismo tiempo, un sensor para la medición del pulso por oximetría fue colocado en el dedo del usuario para la comparación de los resultados (Nonin Onyx II Model 9560 Finger Pulse Oximeter). Los experimentos fueron realizados en un laboratorio con la luz solar como única fuente de iluminación. Durante la grabación de los vídeos los participantes siguieron las instrucciones de permanecer de pie, respirar de manera espontánea y mirar hacia la *webcam*. En la Figura 2.9 se puede ver el lugar donde se realizaron los experimentos y la disposición de los diversos elementos.

De manera general, se obtuvieron resultados muy interesantes al compararse con el sensor (media del error = 1.43, SD = 1.20), muy similar a otros trabajos [70, 73, 76], empleando únicamente la región de la frente en vez de un procesamiento de toda la región facial. Además, es de esperar que mediante el procedimiento propuesto para la detección y seguimiento del ROI de la frente, el sistema sea más robusto que los sistemas antes comentados. Para ejemplificar

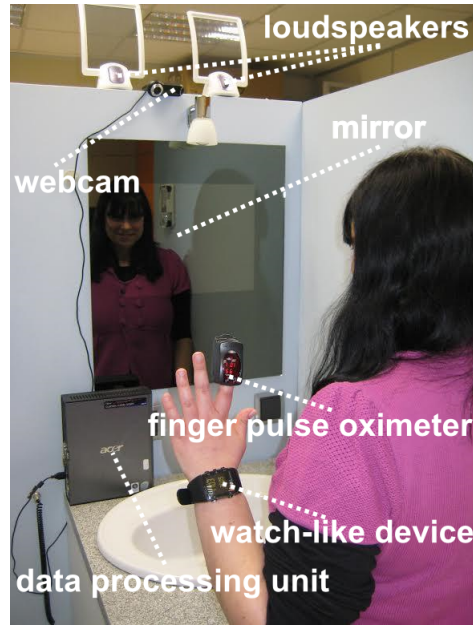


Figura 2.9: Procedimiento para la validación experimental de los algoritmos de visión por computadora para la extracción de variables fisiológicas

este punto, en el sistema propuesto en [70] se puede observar que, incluso sin un movimiento aparente de la región facial, el ROI localizado fluctúa debido a irregularidades en el *tracking* facial. Este problema radica en la base de utilizar el algoritmo de detección facial propuesto por Viola y Jones [56] como algoritmo de *tracking* facial unido al hecho de realizar operaciones matemáticas sencillas para restringir el ROI. Este problema es común al resto de publicaciones analizadas ([71, 76–79]).

Por otro lado, en la presente publicación únicamente se ha llevado a cabo la validación del pulso cardíaco, con lo que se requiere trabajo adicional para llevar a cabo una validación del resto de variables fisiológicas. Para ello, se extiende el trabajo presentado aquí [72] para incluir la validación de la solución completa, es decir, validar el potencial de las variables fisiológicas para su posible inclusión en un sistema de inatención en los conductores. Esta extensión se corresponde con la cuarta Contribución de la presente tesis [80], donde se intenta reproducir el mismo experimento que se presenta en [72], pero incluyendo un sensor de tipo cinturón sobre el pecho para validar el ritmo respiratorio. Los resultados obtenidos se pueden ver en la Tabla 2.3.

Parámetro	Error medio	SD	Índice R
HR	1.43	1.20	0.98
HRV LF	9.26	9.56	0.91
HRV HF	9.16	9.62	0.91
HRV LF/HF	1.11	1.47	0.86
RR	0.84	1.83	0.93

Tabla 2.3: Resumen de los resultados para la extracción de las variables fisiológicas

Los resultados obtenidos se consideran, como en el caso anterior, muy interesantes, similares a otras publicaciones analizadas. Como se ha comentado previamente, es de esperar que la solución aquí mostrada aparezca más robusta cuando los usuarios presenten movimientos más pronunciados. Se puede concluir, por tanto, que los algoritmos de extracción de variables fisiológicas a partir de la región facial del usuario presentan un resultado, a priori válido, para poder ser incluidos y validados en entornos vehiculares. Bien es cierto que se requerirá de esfuerzos adicionales para enfrentarse con las características de las imágenes en este tipo de entornos, como pueden ser las condiciones cambiantes de iluminación [81, 82].

Una vez comprobado experimentalmente el potencial de las variables fisiológicas, otro punto importante es el hecho de incluir mecanismos, que permitan procesar en tiempo real un volumen grande de imágenes. En este sentido se investigó en arquitecturas en *streaming* de procesamiento que pudieran ser aplicadas al procesamiento de imágenes en tiempo real, cuyo resultado se muestra en la quinta Contribución de la tesis [83].

En los últimos años, con el rápido crecimiento de Internet, Internet of Things (IoT) y el Cloud Computing, se ha llegado a un incremento exponencial de los datos en prácticamente cualquier sector de la industria y negocios. El Big Data se ha convertido en un tópico muy activo tanto para la comunidad investigadora, como para la industria y los gobiernos a nivel mundial. Se caracteriza por las tres V's del procesamiento: velocidad, variedad y volumen [84, 85]. Además, dentro de él, han surgido diferentes paradigmas de procesamiento, principalmente: a) procesamiento en *batch*, b) procesamiento en *streaming*, y c) procesamiento híbrido.

El procesamiento en *batch* es una manera eficiente de procesar grandes volúmenes de datos, que se recogen durante un período de tiempo determinado y, después, se procesan en lotes. Sin embargo, algunas aplicaciones requieren un procesamiento en tiempo real de los datos (*streams*). El procesamiento en *streaming* requiere una fuente de datos continua, así como un procesamiento de los datos también continuo [85]. Está especialmente diseñado para operar en tiempo

real, con lo que en este tipo de arquitecturas prima la velocidad sobre la rigurosidad en el análisis. El procesamiento híbrido combina las bondades de ambos tipos. Sin embargo, se requiere una etapa final donde los resultados procedentes del procesamiento en *batch* y en *streaming* se combinan [86].

En la última década, el procesamiento de imágenes y vídeo generado por multitud de sensores y dispositivos se ha convertido en una de las mayores fuentes de datos [87]. Un ejemplo es el procesamiento de vídeos de vigilancia, donde se requiere un proceso en tiempo real para cumplir con los requisitos, entre los que se pueden destacar el reconocimiento facial como clave para la detección de intrusión. En este sentido, el hecho de proponer una arquitectura de *streaming* de datos para el procesamiento de imágenes faciales es clave. En esta quinta Contribución se propone una arquitectura Big Data de procesamiento aportando dos contribuciones principales: 1) una arquitectura en *streaming* para el procesamiento de imágenes faciales por medio de tecnologías Big Data, y 2) paralelización de una innovadora estrategia para la detección de gafas. En lo que al estado del arte se refiere, existen escasas publicaciones relacionadas con el procesamiento facial de imágenes y el Big Data, destacando la extracción de atributos faciales [88], o el reconocimiento facial [89,90].

En la quinta Contribución, se parte del algoritmo para la detección de las gafas propuesto en [47] para adaptarlo a una estrategia de procesamiento en *streaming* capaz de procesar las imágenes en tiempo real. Para ello, se propone una topología de procesamiento haciendo uso de Apache Storm [91,92]. Storm es una herramienta de computación distribuida en tiempo real liberada como código libre por Twitter. En Apache Storm existen dos piezas principales en el procesamiento:

- *Spouts*: Se consideran las fuentes de datos y los emiten en forma de tuplas.
- *Bolts*: Se consideran las unidades de procesamiento.

Combinando estas piezas el objetivo es conseguir procesar la información de la manera adecuada cumpliendo con los objetivos establecidos.

Al tratarse de una arquitectura modular y hacer uso de diferentes nodos de procesamiento, es relativamente fácil adaptar la arquitectura para procesar otro tipo de atributos. Por ejemplo, el hecho de clasificar si un individuo presenta los ojos abiertos o cerrados utilizaría exactamente la misma arquitectura de procesamiento. Únicamente sería necesario modificar la base de datos utilizada para el entrenamiento del sistema. Si por el contrario se quiere saber si una persona está sonriendo, simplemente se tendría que cambiar el *Bolt* de normalización, para que recogiera la región de la boca, como se puede ver en la Figura 2.11. Es decir, la arquitectura se puede modificar para detectar otros atributos faciales y procesarlos en tiempo real: emoción, género, edad o reconocimiento facial entre otros.

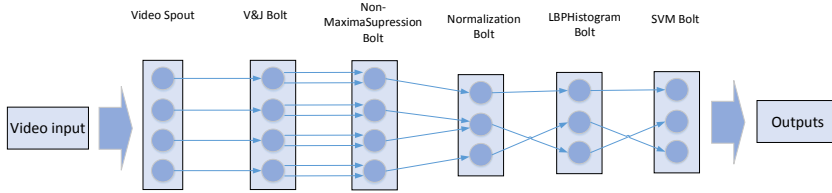


Figura 2.10: Arquitectura Big Data para el procesamiento en *streaming* de imágenes faciales para la detección de gafas

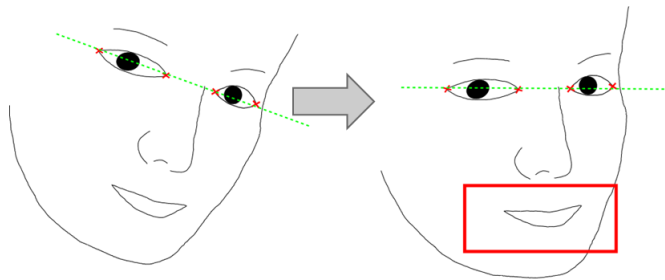


Figura 2.11: Arquitectura de procesamiento para la detección de otro atributo facial (sonrisa) utilizando únicamente un *Bolt* de normalización diferente

Como conclusión, tras la realización de las pruebas con la arquitectura propuesta (mediante inyectores sintéticos para la generación de imágenes), se determina experimentalmente que la mencionada arquitectura es adecuada para el procesamiento en tiempo real de imágenes para la obtención y clasificación automática de atributos faciales. Además, como se muestra en la Figura 2.11 y como se ha comentado previamente, es una arquitectura que se puede adaptar para la extracción de cualquier atributo facial de forma automática.

Otro aspecto que no puede pasar desapercibido en el procesamiento de imágenes es todo lo relacionado con la seguridad. Para esto, son varios aspectos los que se deben tener en cuenta para que sea lo más seguro posible.

El primero de estos aspectos es la identificación facial. Además de proporcionar mecanismos de seguridad a una aplicación de monitorización de usuarios, permite una personalización de la aplicación en función de los gustos y preferencias de un terminado usuario. Además, los sistemas de procesamiento de imágenes presentan ciertas vulnerabilidades de seguridad. Una de las más importantes es lo que se conoce como *spoofing*. En términos de seguridad de redes, este término

hace referencia al uso de técnicas a través de las cuales, un atacante, generalmente con usos maliciosos, se hace pasar por una entidad distinta a través de la falsificación de los datos en una comunicación. Los sistemas de procesamiento facial son vulnerables a los ataques de tipo *spoofing*. Esto puede ocurrir cuando una persona presenta una fotografía de la persona deseada a la cámara en lugar de la suya propia, presentado un gran problema, pues es relativamente fácil y sencillo hacerse con una fotografía para ser usada a posteriori. Basado en las premisas anteriores, como siguiente trabajo de la tesis, se propone construir y validar una arquitectura basada en reconocimiento facial y sistema *anti-spoofing* que pueda ser integrado tanto en un videoportero como en una aplicación móvil adaptado a personas con discapacidad visual [93], que se corresponde con la sexta Contribución de la tesis. Si bien el caso de uso seleccionado es tremendamente interesante, el objetivo de integrar tanto el sistema de reconocimiento facial como el sistema *anti-spoofing* en un videoportero y en una aplicación móvil es el de conseguir unos algoritmos que conjuguen tanto la robustez como una capacidad de cómputo contenida para su aplicación posterior en entornos vehiculares. A continuación, se procede a comentar la publicación antes mencionada.

El reconocimiento facial ha recibido muchas mejoras en los últimos años y hoy en día se acerca a la perfección. Los avances en el reconocimiento facial no han sido ajenos a las personas con discapacidad. Por ejemplo, recientemente se ha presentado un bastón inteligente para ciegos que utiliza reconocimiento facial [94]. Sin embargo, además de realizar la tarea del reconocimiento, actualmente, los sistemas biométricos tienen que lidiar con otro tipo de problemas, como el *spoofing*. La arquitectura propuesta se ha validado con usuarios reales y en un entorno real simulando las mismas condiciones que se podrían dar tanto en las imágenes capturadas por un videoportero como las imágenes capturadas por una persona con discapacidad visual por medio de su dispositivo móvil. El problema del reconocimiento facial adaptado a las personas con discapacidad visual ha sido investigado en sus diferentes formas. A continuación, se resumen los trabajos más importantes, indicando para cada uno de ellos las características más importantes que han ido influyendo en el desarrollo de la arquitectura aquí propuesta. En [95] se presenta un sistema de reconocimiento facial en dispositivos móviles para discapacitados visuales, pero se centra principalmente en reuniones, con lo que aspectos como el campo visual capturado por el dispositivo móvil centran gran parte de la temática. En [96] se desarrolló un sistema de reconocimiento facial basado en LBP. Los autores comparan este descriptor con otras alternativas (Local Ternary Pattern [97] o Histogram of Gradients [98]) y llegan a la conclusión de que el rendimiento de LBP es un poco superior, su coste computacional es menor y la representación de la información es más compacta. En ninguno de estos métodos se lleva a cabo la detección de *spoofing*, haciendo que el sistema tenga una vulnerabilidad total ante este tipo de ataques.

Además, ninguna de las alternativas antes comentadas está orientada a los

videoporteros. Como en ninguno de los casos anteriores se ha estudiado la detección de *spoofing* para ayudar a personas con discapacidad visual, se procede a comentar los resultados más significativos en lo que a la detección de *spoofing* se refiere.

Existen diferentes aproximaciones para su detección. Sin embargo, uno de los factores clave en una aplicación que debe funcionar en tiempo real y en un dispositivo embebido radica en que el método sea computacionalmente ligero. La mayoría de los algoritmos propuestos, o bien son muy complejos y, por tanto, no son aptos para entornos reales, o bien no usan imágenes convencionales (por ejemplo multi-espectrales o termográficas) [99]. Los algoritmos basados en el análisis de micro-texturas ofrecen un buen resultado con un coste computacional relativamente bajo. En [99], se ha aplicado el algoritmo LBP a la detección de *spoofing* aplicando dicho operador a diferentes escalas. En [100], se analizan también diversas variantes del operador LBP para su detección. En [101], también aplican el operador LBP para la extracción de características y SVM en la etapa de clasificación para la detección de *spoofing*. En [102], se propone una aproximación muy similar a la anterior, pero incorporan un método para reducir la longitud del histograma generado por el operador LBP antes de la etapa de clasificación por SVM. Tanto en [100] como en [101] utilizan una división de la región facial en 9 regiones (3x3 regiones), pues comentan que es el número de divisiones que mejores resultados ha generado. Después aplican el operador LBP en cada una de las regiones y concatenan el histograma generado. Este factor es tenido en cuenta en la generación del histograma LBP en la Contribución [93].

De manera sucinta, los principales puntos del procesamiento tanto para la identificación facial, como para la detección de *spoofing* se pueden ver en la Figura 2.12 y se comentan a continuación.

La imagen original (a) se convierte a escala de grises y se detectan los puntos característicos (b). A continuación, se rota para alinearla en función del ángulo (c). Después, se obtiene la región normalizada (64x64) de la cara (d). Se aplican 3x3 divisiones a la región facial (e) y se construye la imagen LBP (f). En el siguiente paso, se construye el histograma dadas estas regiones para formar el vector de características (g). Por último, se clasifica este vector usando ambos clasificadores (h) para obtener las respuestas finales (i).

Para seleccionar los parámetros del algoritmo, como el número de divisiones del operador (3x3) y otros adicionales que se pueden ver en la publicación [93], se realizan unas pruebas preliminares. Los valores seleccionados de estos parámetros están en consonancia con otras investigaciones realizadas [100, 101, 103]. Una vez seleccionados los parámetros del algoritmo, se realizan pruebas durante 5 días consecutivos y con diferentes condiciones de iluminación (por la mañana y por la tarde). Las pruebas se realizan a la entrada del edificio de la Fundación CTIC, como puede verse en la Figura 2.13.

Aunque la entrada del edificio no recibe luz directa, es un sitio bastante ilumi-

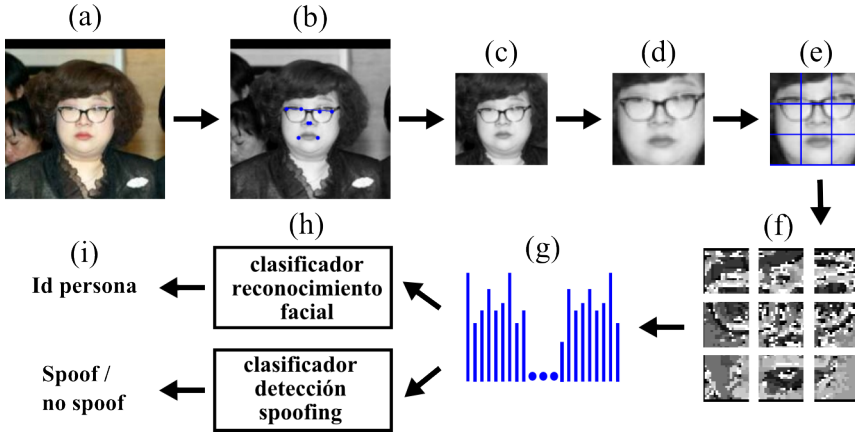


Figura 2.12: Principales pasos en el procesamiento de las imágenes para la identificación facial y la detección de *spoofing*

nado (ver Figura 2.13). En total se realizan pruebas con 25 usuarios. Se considera este número como un número suficiente de personas con las que más a menudo interaccione una persona con discapacidad visual (familiares, amigos más próximos, etc). Para validar el sistema de detección de *spoofing*, se imprimen fotografías de cada uno de los 25 usuarios. Para cada usuario se validan ambos algoritmos (reconocimiento facial y detección de *spoofing*). Para ello, el usuario se debe situar delante del videoportero. Con el objetivo de que el algoritmo obtenga una buena detección facial, el sistema de *tracking* facial procesa sucesivos *frames* hasta que localiza uno donde las características faciales detectadas estén simétricas (dentro de unos umbrales). De esta manera, se garantiza que la detección es frontal y en condiciones para que los algoritmos operen con propiedad. En caso de que el algoritmo de *tracking* facial no detecte una detección frontal por el mecanismo antes comentado, se le proporciona al usuario un comando de voz para que se sitúe frontalmente al videoportero. A continuación, se resume en forma de Tabla (ver Tabla 2.4) los principales resultados a los que se ha llegado.

Como se desprende de los resultados, el reconocimiento facial presenta unos resultados muy robustos. El sistema de detección de *spoofing* presenta unos resultados menos robustos, pero aún así son unos resultados bastante buenos. Cabe destacar que en el cuarto día de pruebas (Día = 4) el tiempo fue bastante soleado. Se cree que pudo afectar al rendimiento de ambos algoritmos.

Con el objetivo de sacar el máximo partido a la arquitectura, se decidió portar los algoritmos a la plataforma *Android*. De esta manera, el usuario con discapacidad visual dispone de una herramienta portable y usable para sus interacciones.



Figura 2.13: Para las pruebas se selecciona un Lenovo ThinkPad X200 Tablet que se posiciona encima del videoportero con la webcam en la parte inferior del mismo. Se puede ver un detalle de la imagen capturada

Algoritmos	Día 1	Día 2	Día 3	Día 4	Día 5
Reconocimiento facial	97 %	98 %	97 %	94 %	97 %
Detección de spoofing	86 %	86 %	89 %	84 %	90 %

Tabla 2.4: Resumen de los resultados de las pruebas realizadas con 25 usuarios tanto para el reconocimiento facial como para la detección de *spoofing* durante 5 días consecutivos

En el caso de uso de la aplicación móvil, se tienen en cuenta factores adicionales que no necesitaron de estudio en el caso del videoportero, los cuales se comentan a continuación. Es necesario tener en cuenta que las imágenes capturadas son tomadas por personas con discapacidad visual y, como consecuencia, muchas imágenes pueden presentar ruido, desenfoque, borrosidad y diferentes condiciones de iluminación. En segundo lugar, es necesario aportar al usuario información auditiva acerca del estado de la aplicación: si ha detectado a una persona, si ha perdido el *tracking* de dicha persona, si la ha conseguido identificar y, por último, si la identificación ha sido real con el objetivo de proporcionarle un *feedback* con el estado de la aplicación, pero sin abrumarle con un exceso de información innecesaria. El principal punto a tratar es que la persona con discapacidad visual no sabe donde está enfocando su dispositivo. Para solucionar este factor, se proponen dos medidas:

1. Guía del estado del *tracking* por medio de pitidos sonoros
2. Analizar varios *frames* antes de predecir un resultado.

Dispositivo móvil	fps	Tpo ejecución (N = 5)	Tpo de ejecución aceptable
HTC DESIRE X	1.82	2.75 s	SI
LG OPTIMUS L2	1.86	2.69 s	SI
WOXTER (TABLET)	0.88	5.68 s	NO
SAMSUNG GALAXY Y	0.76	6.58s	NO

Tabla 2.5: Tiempos de ejecución para varios dispositivos de gama media/baja

En la primera medida, en caso de que el usuario de la aplicación detecte una cara, se producirá un pitido. De esta manera, el usuario mantendrá el móvil en esa posición para ‘mantener’ el *tracking* y que el algoritmo pueda funcionar correctamente. Si se pierde el *tracking* antes de que el algoritmo haya terminado, se producirán dos pitidos que indicarán al usuario que el algoritmo ha finalizado sin concluir una respuesta. En cambio, si el algoritmo consigue dar una respuesta al usuario es que ha procesado un número suficiente de *frames* N, que se corresponde con el segundo de los puntos indicados antes. Tras realizar varias pruebas y consultar la bibliografía relacionada, se ha establecido que el número de *frames* N que se debe procesar antes de proporcionar un resultado es de $N = 5$. Se ha establecido este valor con el objetivo de: (1) proporcionar unos resultados más robustos; y (2) proporcionar un *feedback* rápido para mejorar la fluidez en la interacción. Esto mejorará los resultados en los casos en los que las imágenes estén borrosas o presenten ruido. Por lo tanto, para cada uno de los clasificadores (identificador facial y *spoofing*) se obtienen tanto el identificador de la clase predicha como el valor de fiabilidad de ésta. Es por ello, que para un determinado número de *frames* N, antes de obtener el valor final, se calcula la clase ganadora, y por tanto, el valor a predecir en función de estos valores de confianza. En la Tabla 2.5 se recogen los tiempos de ejecución del algoritmo. Como se puede observar, se han seleccionado dispositivos móviles y *tablets* de gama media/baja disponibles en el laboratorio, pues el objetivo del sistema es que los algoritmos se ejecuten en dispositivos con baja potencia computacional.

A la vista de los resultados mostrados para la publicación anterior [93], se puede concluir que éstos han sido muy satisfactorios. Además, comentar que se han tenido en cuenta conclusiones y elementos claves obtenidos de otras publicaciones relevantes con el objetivo de lograr unos resultados más robustos, de manera que la arquitectura final fuera lo más usable y adaptable a las personas con discapacidad visual. En la Figura 2.14 se pueden ver dos ejemplos realizados en el laboratorio tras la ejecución de los algoritmos.

Otro aspecto de seguridad a tener en cuenta es el tema de mantener el anonimato de los usuarios que aparecen en las bases de datos tanto para entrenar como para validar los sistemas de monitorización de usuarios en entornos vehiculares. El estudio del comportamiento del conductor para evaluar los algoritmos



Figura 2.14: Ejemplos visuales del funcionamiento de los algoritmos de identificación facial y detección de spoofing realizados en el laboratorio

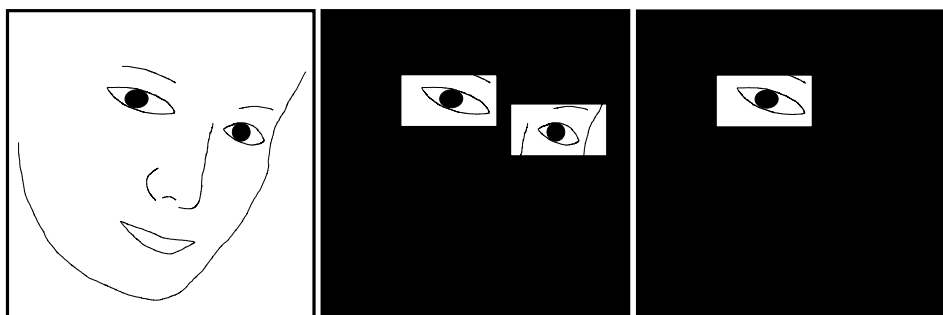


Figura 2.15: La de-identificación permite preservar detalles en la imagen manteniendo el anonimato del individuo

de monitorización requiere de grandes cantidades de vídeos tomadas en entornos reales [104]. Sin embargo, en la literatura actual existe una falta de bases de datos disponibles principalmente debido a temas de privacidad y seguridad. En este punto, en la séptima Contribución [17] se lleva a cabo una revisión de los métodos más importantes para proteger la identidad del usuario. Estos métodos están basados en lo que se conoce como *de-identification* [105]. Un algoritmo ideal de de-identificación debería, por un lado, mantener el anonimato del usuario, pero, por el otro, debería preservar suficientes detalles en la imagen para concluir el estado del conductor [104]. En la Figura 2.15 se puede ver la idea que hay detrás de estos algoritmos.

Una vez completados varios aspectos que se consideran claves para la monitorización de una persona a través de técnicas de visión por computador (detección automática de atributos faciales, identificación facial, detección de *spoofing*, desarrollo e implementación de arquitecturas en entornos embebidos o seguridad), el siguiente punto radica en profundizar en aspectos específicos en la detección de

la inatención en los conductores y en los algoritmos de visión por computador para establecer y obtener una base de conocimiento cubriendo los aspectos más importantes para el desarrollo e implementación de sistemas de monitorización de conductores en entornos embebidos. Todos estos detalles se recogen en la séptima Contribución de la tesis [17].

En concreto, los puntos sobre los que trata esta Contribución son los siguientes:

- Introducción a la detección de la distracción en los conductores
- Detección facial y detección de puntos característicos en entornos vehiculares
- Distracción visual
- Distracción biomecánica
- Distracción cognitiva
- Combinación de varios tipos de distracción
- Relación entre expresiones faciales y distracción
- Principales pasos para portar un algoritmo de visión por computador a un entorno embebido
- Entornos simulados *versus* entornos reales para validar los sistemas de monitorización de usuarios
- Aspectos de privacidad y seguridad en aplicaciones de monitorización de conductores
- Discusión general de resultados
- Aspectos futuros a considerar

El propósito de la publicación es el de realizar un análisis del estado del arte relacionado con la detección de la distracción. El ámbito de la publicación, que se comentó anteriormente de forma resumida, se puede ver también en la Figura 2.16, en la que se puede apreciar que se lleva a cabo una categorización de los diferentes tipos de distracción en los conductores, principalmente de tipo biomecánico, visual y cognitivo. Además, se realiza una revisión a la taxonomía existente, destacando los aspectos positivos y haciendo hincapié en varias carencias detectadas, que es necesario considerar. Además, se realiza una revisión de las publicaciones existentes que realizan contribuciones importantes para portar un algoritmo de visión por computador en un entorno vehicular. El hecho de entrenar y validar un algoritmo de este tipo es otro punto importante

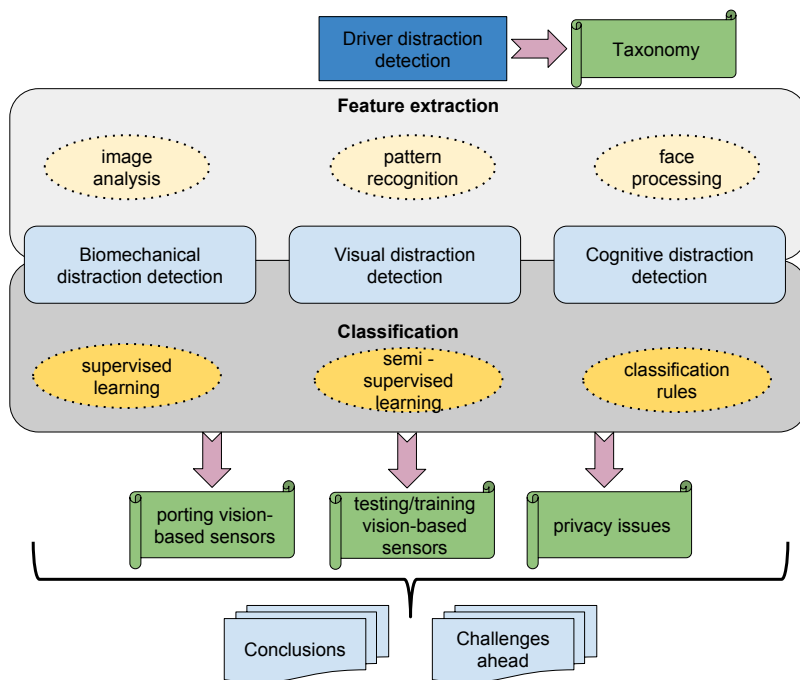


Figura 2.16: Principales puntos tratados en la séptima Contribución de la tesis

que es tratado. Los problemas derivados de la seguridad, que fueron comentados previamente, también son contemplados. También se recogen los principales sistemas comerciales que realizan la detección de la inatención en los conductores comentando sus características más destacadas. Cabe decir que no se han encontrado publicaciones que realicen una comparación entre sistemas comerciales y sistemas basados en publicaciones existentes. Todos estos aspectos se tratan de manera detallada en la publicación anterior.

Puesto que uno de los principales objetivos de la presente tesis es la detección de la inatención en los conductores utilizando algoritmos de visión por computador, en la Tabla 2.6 se recogen los principales trabajos para la detección de la distracción, en la que se tiene en cuenta si el sistema implementado detecta alguno de los tres tipos de distracción (biomecánica, visual o cognitiva), además de si el sistema está preparado para operar en condiciones reales y si es capaz de operar tanto de día como de noche.

Una vez establecida la base teórica, el siguiente punto radica en el de proponer un sistema que detecte tanto la distracción como la somnolencia en los

Ref	Tipos de distracción			Real	Operación	
	Manual	Visual	Cognitiva		Día	Noche
Zhao et al. [106]	✓	✗	✗	✗	✓	✗
Zhao et al. [107]	✓	✗	✗	✗	✓	✗
Zhao et al. [108]	✓	✗	✗	✗	✓	✗
Zhao et al. [109]	✓	✗	✗	✗	✓	✗
Bosch et al. [110]	✓	✗	✗	✗	✓	✗
Lowe et al. [111]	✓	✗	✗	✗	✓	✗
Yan et al. [112]	✓	✗	✗	✗	✓	✗
Yan et al. [113]	✓	✗	✗	✗	✓	✗
Zhang et al. [114]	✓	✗	✗	✗	✓	✗
Artan et al. [115]	✓	✗	✗	✓	✓	✓
Berri et al. [116]	✓	✗	✗	✓	✓	✗
Xu et al. [117]	✓	✗	✗	✓	✓	✓
Seshadri et al. [118]	✓	✗	✗	✓	✓	✗
Ohn et al. [119]	✓	✗	✗	✓	✓	✗
Ohn et al. [119]	✓	✗	✗	✓	✓	✗
Martin et al. [120]	✓	✗	✗	✓	✓	✗
Martin et al. [120]	✓	✗	✗	✓	✓	✗
Ohn et al. [121]	✓	✗	✗	✓	✓	✗
Ohn et al. [121]	✓	✗	✗	✓	✓	✗
Bergasa et al. [122]	✗	✓	✗	✓	✓	✓
Lee et al. [123]	✗	✓	✗	✓	✓	✓
Vicente et al. [124]	✗	✓	✗	✓	✓	✓
Cyganek et al. [125]	✗	✓	✗	✓	✓	✓
Donmez et al. [126]	✗	✓	✗	✗	✓	✗
Klauer et al. [11]	✗	✓	✗	✗	✓	✗
Kircher et al. [127–129]	✗	✓	✗	✗	✓	✗
Victor et al. [130]	✗	✓	✗	✗	✓	✗
Zhang et al. [131]	✗	✗	✓	✗	✓	✗
Zhang et al. [131]	✗	✗	✓	✗	✓	✗
Zhang et al. [131]	✗	✗	✓	✗	✓	✗
Zhang et al. [131]	✗	✗	✓	✗	✓	✗
Liang et al. [132]	✗	✗	✓	✗	✓	✗
Liang et al. [132]	✗	✗	✓	✗	✓	✗
Liang et al. [132]	✗	✗	✓	✗	✓	✗
Liang et al. [133]	✗	✗	✓	✗	✓	✗
Miyaji et al. [134]	✗	✗	✓	✗	✓	✗
Miyaji et al. [134]	✗	✗	✓	✗	✓	✗
Miyaji et al. [134]	✗	✗	✓	✗	✓	✗
Yang et al. [135]	✗	✗	✓	✗	✓	✗
Yang et al. [135]	✗	✗	✓	✗	✓	✗
Liu et al. [136]	✗	✗	✓	✗	✓	✗
Li et al. [137]	✗	✓	✓	✗	✓	✗
Craye et al. [138]	✓	✓	✗	✗	✓	✗
Liu et al. [139]	✗	✓	✓	✗	✓	✗
Ragab et al. [140]	✓	✓	✗	✗	✓	✗

Tabla 2.6: Aproximaciones para la detección de la distracción en los conductores

conductores, por ser los tipos de inatención que más mortalidad ocasionan. Todo esto se aborda en la octava (y última) Contribución de la tesis doctoral [141]. En ella, se propone una arquitectura de procesamiento especialmente diseñada para operar en entornos vehiculares, con una carga computacional muy baja y fácilmente integrable en dispositivos con reducidas capacidades de cómputo, capaz de lidiar con distintas condiciones de imágenes muy presentes en este tipo de entornos.

El sistema se validó, en primer lugar, con bases de datos de referencia testeando los diferentes módulos que la componen. En concreto, se detecta la presencia o ausencia del conductor con una precisión del 100 %, 90.56 %, 88.96 % por medio de un marcador ubicado en el reposacabezas del conductor, por medio del operador LBP, o por medio del operador CS-LBP, respectivamente. En lo que respecta a la validación mediante la base de datos CEW [142] para la detección del estado de los ojos, se obtiene una precisión de 93.39 % y de 91.84 % utilizando una nueva aproximación basada en LBP (LBP_RO) y otra basada en el operador CS-LBP (CS-LBP_RO). Tras la realización de varios experimentos para ubicar la cámara en el lugar más adecuado, se posicionó la misma en el salpicadero, pudiendo aumentar la precisión en la detección de la región facial de un 86.88 % a un 96.46 %. Las pruebas en entornos reales se realizaron durante varios días recogiendo condiciones lumínicas muy diferentes durante las horas diurnas involucrando a 16 conductores, los cuales realizaron diversas actividades para reproducir síntomas de distracción y somnolencia. Dependiendo del tipo de actividad y su duración, se obtuvieron diferentes resultados. De manera general y considerando de forma conjunta todas las actividades se obtiene una tasa media de detección del 93.11 %. En resumen, la principal contribución es el hecho de presentar una solución completamente autónoma para detección de distracción y somnolencia pues son dos de los tipos en la inatención que más accidentes ocasionan. Además, dicha solución presenta las siguientes características:

1. detección automática de la presencia del conductor en el entorno vehicular para dirigir el flujo del algoritmo
2. detección facial adaptada al entorno vehicular
3. normalización facial para enfrentarse a características de imagen difíciles
4. detección rápida y robusta de distracción
5. detección rápida y robusta de somnolencia

De manera gráfica, en la Figura 2.17, se incluyen los principales puntos del sistema propuesto en esta octava Contribución de la tesis [141]. En ella se puede ver cómo interactúan los diferentes módulos para concluir el estado del conductor.

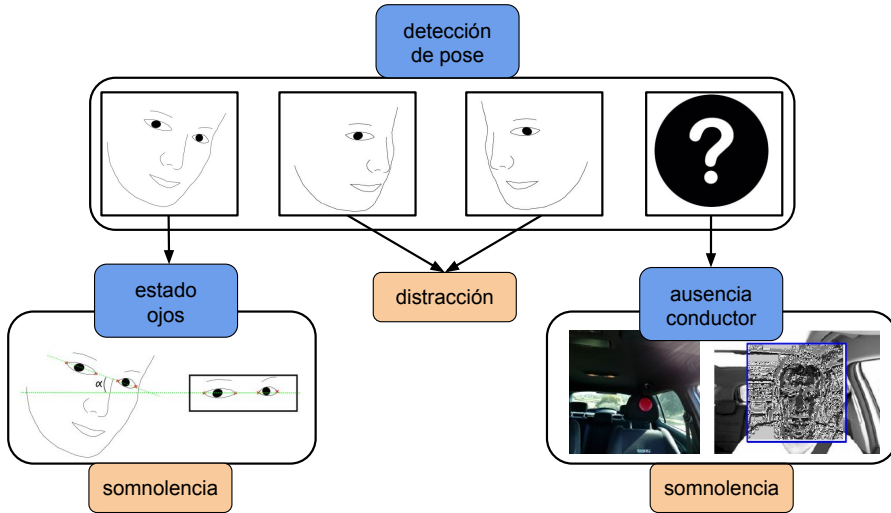


Figura 2.17: Principales puntos del sistema propuesto para la detección tanto de la somnolencia como de la distracción en conductores

La comparación mediante datasets y bases de datos disponibles para la comparación de los diferentes algoritmos, como puede ser la base de datos CEW en este caso, permite una comparación ‘justa’ entre las diferentes aproximaciones. En este sentido, de los resultados desprendidos tras la realización de la validación de los módulos que componen el sistema con diferentes bases de datos, se puede decir que los diferentes módulos se comportan de forma robusta. La detección de ausencia del conductor por medio de un marcador es un método simple y efectivo para detectar si el conductor está presente en el puesto de conducción. Hasta donde se sabe, se trata de la primera publicación que propone la detección del conductor por medio de un marcador ubicado en el asiento. De los resultados obtenidos tras la ejecución de las pruebas en entorno real controlado, se puede deducir que se detecta con bastante precisión los eventos de distracción y somnolencia, obteniéndose los resultados mas elevados para la detección del estado de los ojos. El hecho de que el conductor lleve gafas en el momento de las pruebas ocasionó que la tasa de reconocimiento se decrementara aproximadamente en un 2%. Por otro lado, la posición de la cámara juega un papel fundamental, pues la detección facial es el punto principal sobre el que pivota el resto del sistema. Esto es así porque los algoritmos de detección facial están entrenados con caras con una determinada pose para poder ser detectadas de frente o de perfil. Así pues, las caras capturadas en la imagen desde el interior vehicular deben tener una apariencia similar a aquellas con las que fueron entrenadas. Si la cámara se

sitúa en el salpicadero en frente del conductor, se puede incrementar la detección facial en, aproximadamente, 8 puntos porcentuales respecto a si la cámara está en el espejo retrovisor (de un 86.88 % se puede pasar a un 96.46 %), lo cual está en consonancia con otros trabajos, que establecen que el salpicadero es el mejor lugar para la ubicación de la cámara ([143–145]).

El grado de detección facial logrado con la ubicación final de la cámara y utilizando el algoritmo de detección facial antes comentado, es superior al de todos los trabajos analizados ([145–151]). Puesto que el resto de la arquitectura de estos sistemas trabaja sobre la detección de la cara del conductor, este es un factor determinante.

Además, el Algoritmo de Viola & Jones es frecuentemente usado por su facilidad de uso y por sus resultados más o menos robustos. En los trabajos analizados ([145–151]) es usado como detector facial, lo que puede conllevar alguna restricción en el procesamiento. Por ejemplo, en [148], con el objetivo de satisfacer los requisitos de ‘tiempo real’, se vieron obligados a bajar la resolución a 320×240 píxeles, lo que no ocurre en el sistema aquí propuesto.

Como conclusiones a esta publicación, existen básicamente dos limitaciones en el sistema propuesto. El primero de ellos es que está orientado como sistema para la detección tanto de la distracción como de la somnolencia en condiciones diurnas. El sistema se debería completar con los algoritmos adecuados para contemplar condiciones nocturnas para ofrecer un funcionamiento ininterrumpido.

Por otro lado, y relacionado con la detección de la distracción visual, el sistema utiliza lo que se conoce como una aproximación *coarse*, lo que podría suponer alguna limitación en alguna situación muy concreta. Por ejemplo, cuando el conductor esté realizando una tarea secundaria que involucre intervención visual, en la que presente una orientación en la cabeza detectada por el algoritmo de *tracking* como de distracción, pero pudiendo alternar de manera continua y constante la mirada entre la carretera y el foco de atención de dicha tarea secundaria.

Capítulo 3

Conclusiones y trabajo futuro

La conducción requiere de una gran concentración y cualquier descuido puede provocar un accidente. Los accidentes de tráfico representan actualmente la novena causa de muerte, de los cuales, muchos son provocados por la inatención, pues se calcula que ocasiona entre el 25 % y el 75 % de los accidentes y casi-accidentes. La inatención se ha convertido en un campo muy estudiado, donde los sistemas de visión por computador se han posicionado como una forma no intrusiva y eficaz para la detección de la inatención.

La presente tesis doctoral se ha centrado en la detección de la inatención de los conductores mediante el reconocimiento de atributos faciales haciendo uso de la visión por computador. Para llevar a cabo el objetivo principal antes comentado, se identifican, en primer lugar, un conjunto de objetivos particulares los cuales es necesario abordar a lo largo de todo el Plan de Investigación. A tal efecto, se han presentado ocho Contribuciones, que van planteando soluciones y respuestas a los mencionados objetivos. Los resultados obtenidos de las Contribuciones concluyen que es posible plantear una arquitectura de procesamiento capaz de recoger los principales signos que caracterizan a un conductor inatento y procesarlos mediante algoritmos de visión por computador para ser embebidos en un dispositivo con bajas capacidades de cómputo. En concreto, los principales resultados se mencionan a continuación de manera concisa haciendo referencia a la publicación que los aborda con mayor profundidad:

- Tras analizar la diversa bibliografía, se determina que la inatención es la causa que más muertes ocasiona en accidentes de tráfico. Además, con el auge de los dispositivos móviles, el problema se ha visto agravado en los últimos años. En España, el número de fallecidos en 2016 en carretera acaba con 13 años de descenso de víctimas [SENSORS2016].
- La distracción, somnolencia y, en general, la inatención han sido incorrectamente definidas en muchas publicaciones. Actualmente, se proponen taxonomías donde se empiezan a vislumbrar reglas claras para la categorización de estos conceptos. Sin embargo, existen deficiencias en la clasificación de la inatención y es necesario contemplar y recoger varios aspectos,

tales como el prejuicio de retrospectiva (*hindsight bias*) y la dinámica de la distracción [SENSORS2016].

- Se han analizado y recogido, en función de diversos criterios, los principales métodos para la detección de la inatención en los conductores. De forma resumida (se puede consultar la Figura 1.3), existen cuatro tipos: 1) métodos basados en el análisis del comportamiento vehicular, 2) métodos basados en el análisis de variables fisiológicas por medio de sensores, 3) métodos basados en el análisis de variables fisiológicas por medio de visión por computador, y 4) métodos basados en el análisis de características visuales [RIAI2017].
- En función de las publicaciones analizadas, los métodos de visión por computador se han posicionado como uno de los métodos más adecuados para caracterizar el estado del conductor debido a su baja intrusión en la conducción, por el potencial que presentan y por el auge del *hardware* necesario para llevarlos a cabo [SENSORS2016].
- Se ha llevado a cabo un estudio exhaustivo donde se analizan los principales métodos basados en visión por computador para la detección de la distracción, clasificados en función de diversos criterios: tipos de distracción (manual, visual o cognitiva), si han sido validados en entornos reales, o si son capaces de operar tanto de día como de noche [SENSORS2016].
- Se han clasificado y analizado los principales algoritmos de *tracking* facial para entornos vehiculares, pues son el primer paso en los sistemas de monitorización de los conductores. Dichos algoritmos presentan características diferentes, pudiendo requerir en su operación de una o varias cámaras colocadas a lo largo del habitáculo para poder capturar con mayor robustez el rango de movimientos de la cabeza del conductor [SENSORS2016].
- Se han clasificado y analizado los principales métodos para la detección automática de los principales atributos faciales (género, edad, emoción, identificación, etnia) basados en el operador LBP para la etapa de extracción de características y de SVM en la etapa de clasificación [SMARTMILE2013].
- Se han clasificado y analizado los principales métodos para la detección automática de gafas en imágenes faciales [MVAA2015].
- Se han propuesto y validado mejoras en la clasificación automática de atributos faciales para lidiar con aspectos como la pose, el escalado o la rotación en imágenes faciales, lo que permite aumentar la robustez final de los algoritmos. Estas mejoras están aplicadas a la clasificación automática de las gafas [MVAA2015].

-
- Se ha analizado como influye la resolución de la imagen en la detección final del algoritmo para un atributo en concreto (detección de gafas), encontrándose una relación entre ambos factores [MVAA2015].
 - Se ha analizado como influye el número de divisiones del operador LBP en la creación del *vector espacial extendido* y su posterior influencia en la clasificación final del algoritmo para un atributo en concreto (detección de gafas)[MVAA2015].
 - Se ha validado el potencial de los algoritmos para la extracción automática de variables fisiológicas a partir del procesamiento de imágenes para su aplicabilidad en entornos vehiculares [CITS2015,JON2017].
 - Se ha propuesto un método para la automatización de los sistemas de monitorización de los conductores, mediante la inclusión de un marcador, o bien mediante algoritmos de aprendizaje automático [RIAI2017].
 - Se han analizado los principales factores a tener en cuenta para la implementación de los algoritmos de visión por computador en entornos vehiculares [SENSORS2016].
 - Se ha investigado en arquitecturas de *streaming* de datos para el procesamiento de grandes volúmenes de imágenes en tiempo real [FIOT2015].
 - Se han recogido los principales requerimientos en cuanto a la seguridad, investigando y proponiendo algoritmos para la identificación facial, la detección de *spoofing* y un análisis de los principales métodos en la identificación de imágenes a incluir en las bases de datos para la comparación de los algoritmos [IEEEELAT2016,SENSORS2016].

La presente tesis, por tanto, intenta recoger los signos de inatención del conductor. Sin embargo, la visión por computador también se podría usar para obtener, no sólo la información de dentro del vehículo, sino también la información de lo que ocurre fuera del habitáculo. El paso final radicaría en fusionar la información "out" y la información "in" para una caracterización completa del estado del conductor. Además, esta tesis se centra en la obtención de información "puramente" visual, pero se podría completar con la información procedente de los sensores existentes tanto en los dispositivos móviles como en los vehículos. Estos puntos antes comentados se consideran los principales aspectos futuros a tratar y, comentados en detalle en la Contribución de la tesis [17], se exponen a continuación en aras de la exhaustividad.

3.1. Aspectos futuros a considerar

Como se ha comentado en la Introducción, el hecho de usar sensores ya disponibles en los vehículos para analizar el comportamiento del conductor es una alternativa interesante a los sistemas basados en visión por computador [27,152]. Sin embargo, estos sistemas no se deberían tratar como diferentes alternativas, sino que la información debería ser fusionada para obtener unos resultados finales más robustos [153].

Además, las técnicas de visión por computador pueden ser usadas para obtener, no sólo información del interior del habitáculo, sino también para extraer información de fuera del mismo, como por ejemplo: caracterización del tráfico, peligros en la carretera, peatones, condiciones externas, intersecciones, o posición relativa de otros vehículos. En este sentido, existen algunos trabajos que intentan recoger ambos tipos de información, pero en muchos casos se restringen a aspectos muy concretos [154–159], con lo que todavía hay margen de mejora debido, por un lado, al avance de los algoritmos de visión por computador, y por otro, al avance en el hardware disponible, principalmente cámaras con mejores prestaciones y unidades de procesamiento más potentes.

Capítulo 4

Difusión de la tesis doctoral

En este capítulo se adjuntan las publicaciones que conforman la tesis doctoral. Conviene destacar que se incluyen tanto los artículos publicados en medios con impacto, como son revistas indexadas en el JCR o similares, así como artículos presentados en congresos de menor relevancia. No obstante, estos últimos permiten ampliar la visión general de la labor investigadora llevada a cabo durante la realización de la tesis doctoral, por lo que su inclusión puede considerarse obligada para entender apropiadamente el alcance de la misma. Adicionalmente, se incluyen tanto la información referente a la calidad como el factor de impacto de las publicaciones que componen la tesis doctoral. En la Figura 4.1 se puede ver la clasificación de las publicaciones.

Además, se incluyen las citas realizadas a los artículos incluidos en la Figura anterior y mostrados en la Figura 4.2.

A continuación se adjuntan en forma de lista las publicaciones referenciadas

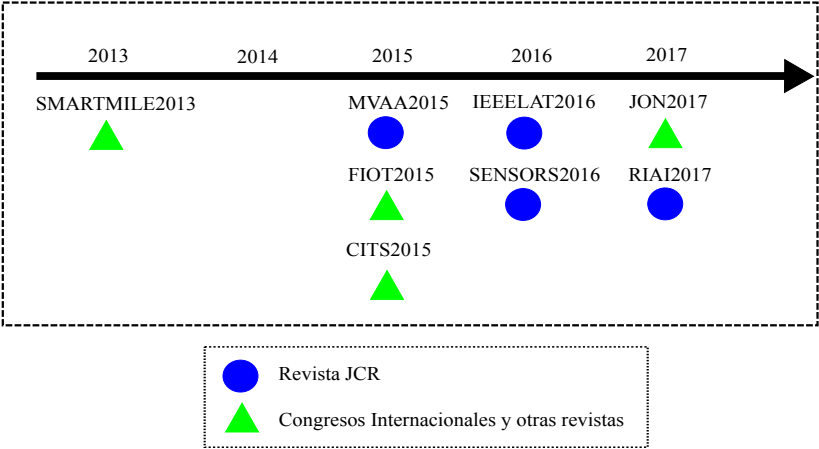


Figura 4.1: Resumen de las publicaciones

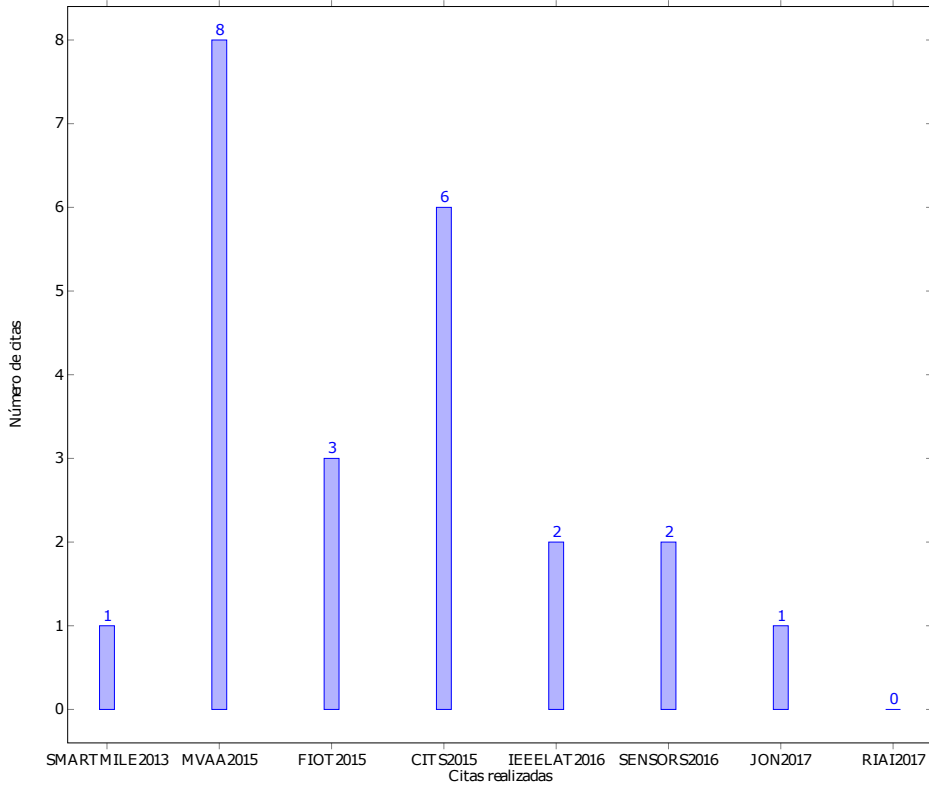


Figura 4.2: Citas realizadas a los artículos de la presente tesis doctoral. Fuente: Google scholar actualizado con publicaciones todavía no indexadas

en la Figura anterior y que se comentarán a continuación:

- **SMARTMILE2013:** Losada, D. G., Lopez, G. A. R., Acevedo, R. G., & Villan, A. F. (2013, December). AVIUE Artificial vision to improve the user experience. In *New Concepts in Smart Cities: Fostering Public and Private Alliances (SmartMILE)*, 2013 International Conference on (pp. 1-6). IEEE.
- **MVAA2015:** Fernández, A., García, R., Usamentiaga, R., & Casado, R. (2015). Glasses detection on real images based on robust alignment. *Machine Vision and Applications*, 26(4), 519-531.
- **FIOT2015:** Fernández, A., Casado, R., & Usamentiaga, R. (2015, August). A Real-Time Big Data Architecture for Glasses Detection Using

-
- Computer Vision Techniques. In Future Internet of Things and Cloud (Fi-Cloud), 2015 3rd International Conference on (pp. 591-596). IEEE.
- **CITS2015:** Fernández, A., Carús, J. L., Usamentiaga, R., Alvarez, E., & Casado, R. (2015, July). Unobtrusive health monitoring system using video-based physiological information and activity measurements. In Computer, Information and Telecommunication Systems (CITS), 2015 International Conference on (pp. 1-5). IEEE.
 - **IEEEELAT2016:** Villan, A. F., Candas, J. L. C., Fernandez, R. U., & Tejedor, R. C. (2016). Face recognition and spoofing detection system adapted to visually-impaired people. IEEE Latin America Transactions, 14(2), 913-921.
 - **SENSORS2016:** Fernández, A., Usamentiaga, R., Carús, J. L., & Casado, R. (2016). Driver distraction using visual-based sensors and algorithms. Sensors, 16(11), 1805.
 - **JON2017:** Fernández, A., Carús, J. L., Usamentiaga, R., Alvarez, E., & Casado, R. Wearable and ambient sensors to health monitoring using computer vision and signal processing techniques. Journal of Networks. 2017.
 - **RIAI2017:** Fernández, A., Usamentiaga, R. & Casado, R. Sistema Automático Para la Detección de Distracción y Somnolencia en Conductores por Medio de Características Visuales Robustas. Revista Iberoamericana de Automática e Informática Industrial RIAI. 2017.

4.1. Artículos

4.1.1. AVIUE Artificial Vision to Improve the User Experience

- Losada, D. G., Lopez, G. A. R., Acevedo, R. G., & Villan, A. F. (2013, December). AVIUE Artificial vision to improve the user experience. In *New Concepts in Smart Cities: Fostering Public and Private Alliances (SmartMILE)*, 2013 International Conference on (pp. 1-6). IEEE.
- DOI: 10.1109/SmartMILE.2013.6708188
- URL: <http://ieeexplore.ieee.org/abstract/document/6708188/>

AVIUE - Artificial Vision to Improve the User Experience

Daniel González Losada
and Gustavo Adolfo Rosal López
SINERCO
Rendueles Llanos 8
33205 Gijón
Spain

Rodrigo García Acevedo
and Alberto Fernández Villán
Fundación CTIC (Technological Center)
Technological Scientific Park of Gijón
33204 Gijón
Spain

Abstract—Users require specialized and specific services increasingly. In this paper a vision-based system to personalize and specialize services is proposed. Attributes such as gender, age, ethnicity, emotion and identity can be acquired using computer vision techniques by employing state-of-the-art algorithms. In this way, user attributes are captured in an automated and non-intrusive way. In order to test the presented solution, current open data available for cities is used. In particular, data provided by the city of Gijón is taken into account. Thereby, personalized services can be developed in order to provide better information to the user. Data is captured in strategic locations with no impact on the urban furniture with the aim of collecting users' information on a non-intrusive way and develop new services tailored to the users' profile. The proposed system is complemented with a gestural user interface in order to achieve a more relaxed and natural man-machine interaction. **Index Terms**—Personalization, Natural Interaction, User attributes, Emotional Design, Senseable City

I. INTRODUCTION

Personalization involves a process of gathering user-information during interaction with the user. This information is then used to deliver appropriate content and services, tailor-made to the users' needs. So the aim of personalization is to improve the users' experience of a service. Personalization can be described as being about "building customer loyalty by building a meaningful one-to-one relationship" [1]. Users' satisfaction is the ultimate aim of personalization. It is motivated by the recognition that a user has needs, and meeting them successfully is likely to lead to a satisfying relationship and re-use of the services offered. Beyond the common goal, however, there is great diversity in how personalization can be achieved. Information about the user can be obtained in many different ways [2]:

- Directly (survey). This is an explicit method where the system asks the user to express preferences or choices explicitly and uses this feedback to build a user profile and make recommendations accordingly.
- From a history of previous sessions (implicit method).

This data is used to build consumer profiles. The profiles should include the information describing who the consumers

are and how they behave in a way that is more relevant for the particular personalization application(s)/service(s).

However, due to the static nature of gathering this information, dynamic information cannot be obtained. For example, the emotion of one user is an important aspect to take into account. A personalization system sending too much feedback to the user is intrusive and often irritating. Therefore, minimizing intrusiveness is usually desirable. Another key point to take into account is that although consumers want personalized services, they may be unwilling to provide some of the personal information needed to improve the system ability to make recommendations.

In this research, a way of obtaining users' attributes avoiding intrusiveness and using dynamic information is proposed. Attributes such as gender, age, ethnicity, emotion and identity can be acquired using computer vision techniques.

The proposed system is complemented with a complete gestural user interface in order to make the interaction efficient, natural and satisfying for the user.

In order to validate the current approach, it is expected to use the open data provided by Gijón Town Council for the customization of services. Gijón keeps on with the idea of becoming a smart city developing smart services for citizens, improving environmental sustainability, energy efficiency in the use of facilities and infrastructures, and in general, life quality. Nowadays there are 423 datasets published by the Gijón Town Council which can be used by the citizens.

It is believed that the benefits taken by the current approach will be:

- The promotion of innovative products and services by contributing to the social and economical development and employment.
- The contribution to improve citizens' life quality by means of more efficient and personalized services.
- The improvement of citizens' involvement and participation in the town affairs.
- The promotion to open new datasets by companies and individuals that want to expose their data.
- The acceleration to become a smart city.

II. RELATED WORK

A. Similar initiatives

Open data is a philosophy and practice which aims to make certain data available to all, using free, open, accessible and standard formats. New services can be created fostering innovation by exposing and reusing this data.

Moreover it improves transparency on governance and promotes interoperability and interaction between administrations.

There are multiple initiatives both at national and international levels focusing on generating added value through reuse of open data. Basque Government became the first Spanish authority to make the public data in its possession available on a specific Internet portal. This resulted in the development of applications such as mini-dictionaries from translation memories, or lists of grants and subsidies offered by the Basque government itself. Aporta project, born in October 2008 in the framework of the Avanza Plan, has provided, among other services, customized weather forecasts and collections of laws with national and international jurisprudence ready to be consumed by citizens. Open Data in Zaragoza. Providing initiatives in Facebook to geotag public schools. This action has been awarded the Idezar (EUROGI Award / eSDI-Net). Open Data in Catalonia is another example. Very useful services based on geolocation functions are developed. Weather forecast, real time traffic incident information in Catalonia and tourist accommodation are examples of mobile services based on Generalitat open data. Similarly, there are many international initiatives on reuse of open data.

One of the prominent examples in the world is the Open Data San Francisco initiative. Data is offered in several formats: drawings, tables, graphs, links to external sources, etc. This data is provided by public and private organizations, and is enriched with opinions or comments of portal users. The data can be also be queried using smartphones applications (apps). Some of this apps include information about criminology ("Score Neighborhood", "alerted", "Are You Safe San Francisco", "CrimeReports", or "CrimeMapping SpotCrime"), restoration and what to eat ("Landmarks: San Francisco", "SF Way", "SparkleDine"), education such as the aforementioned "Score Neighborhood" offering different data about neighborhoods of the city, surroundings, environment and recycling ("Open City Agora", "Ecofinder"), maps and transport ("Parking" to locate parking, "Smart Ride" to locate urban transport, "SFFILMSPOTS" with offers for recreation and culture), news like "Everyblock" and politics like "HowSFVotes" that shows historical information of San Francisco election results since 2004, etc. Other international initiatives like "Toilet Map" app developed in the UK to locate public toilets, apps about bicycle accidents in CA, crime map in UK, Buses in Real Time in Gijon, "Where do you live?" in the UK, "Health" in the US, can be cited.

Regarding the use of computer vision techniques to develop new services for the citizen, the initiative of the Zaragoza City Council can be cited. There is an implementation of a system for the detection of queues at traffic lights. Cameras are

installed to use computer vision techniques to obtain the queue measure of cars that are waiting. Each time a traffic lights is closed, it calculates the accumulated cars, and this allows adapting the traffic lights to avoid long traffic jams. Computer vision systems are also used in exhibitors for advertisement purposes.

However, as far as authors know, there is no initiative that combines the reuse of open data with artificial vision techniques in order to provide enhanced services to users.

B. User Experience and emotions

The user experience is any aspect of a person's interaction with a system, which results in the generation of a positive or negative perception of the service it offers. This experience not only depends on traditional design factors (hardware, software, graphic design, content quality ...), it also depends on desirable or undesirable aspects induced in the user.

The emotional interface design is oriented to create a bond between the user and the system or product so the interaction creates a pleasant experience. The user interface must be useful, usable, desirable, valuable, believable and accessible [3].

Besides, the user is the main element related to the design of the system. Designers have to work trying to make products with a clear idea about what the user needs and likes, instead of making products without a specific final user. It is well known that ergonomics is the origin of usability, because its goal is effectiveness and efficiency, as well as the welfare and health of the user. The effort is to adapt work to people, so that systems and devices can adopt their way of thinking, behaving and working. Thus, an experiential and emotional imprint in people with meaning to them can be generated. The content must be intelligible at a cognitive level, otherwise it can generate frustration in the subject creating negative emotions and thereby reducing the chances of future interaction.

When the subject cognitively understands the product interface, he can learn to use the system easily and effectively. The desire to achieve a better experience to the user ends up in more efficient interfaces.

Considering the emotional factor in the design, makes possible that products labeled as "smarts" reach a higher degree of humanization and achieve a higher level as a "senseable" product.

C. Open Data Catalog

The goal of the action plan promoted by the City Council of Gijon is to make the information generated and managed by the City Council available to the citizens. This project will show transparency and will provide the basis to improving citizen's participation. This initiative is also aimed at promoting innovation, facilitating the creation of new services or adding value to existing ones and enabling new business models for companies. This will benefit both businesses and users, as users will have new high value useful services. Gijon City Council is committed to an effective aperture of this public data provided under open property licenses

that allow redistribution, reuse and its use for commercial purposes. The data catalog provided by the city council currently has a total of 423 datasets organized by sectors, among which are: science and technology, law enforcement, culture and leisure, demographics, sport, economy, education, employment, energy, finance, information, administration information, environment, transport, tourism, health, public sector, security, urban planning and infrastructure. Society of new generation has become more selective, more critical and with the ability to access a greater volume of information. However, the information is globalized and each individual looks for something in particular, according to his likes and emotions. This information could be enriched and customized to each person using attributes and emotional state extracted by means of a computer vision system. Some of the datasets are susceptible to be improved if they are reused in conjunction with computer vision techniques and thus provide greater value to the user.

With this initiative, there are some data sets prone to be enriched with artificial vision techniques to generate added value to the final services such as datasets related to tourism and information sectors. Thus, within the tourism sector should be noted the following datasets: wine and tapas, gourmet shops, bakery, cafes, shops, casinos, lodging, accommodation, village houses, hostels, hotels and rural hotels. These datasets are the most susceptible to improve in terms of customizing because they are related to tastes, preferences and dynamic information.

D. Vision-based algorithms to obtain users' attributes

A high degree of accuracy is needed in order to get users' information. Attributes such as gender, age, ethnicity, emotion and identity can be acquired using computer vision techniques by employing state-of-the-art algorithms in a non-intrusive way. Local Binary Pattern (LBP) operator [4] is a type of feature that is commonly used to get the face information. The face image is divided into several regions from which the LBP feature distributions are extracted and concatenated into an enhanced texture vector to be used as a face descriptor. The classification is usually performed by using Support Vector Machines (SVMs), which had been shown to be superior to traditional pattern classifiers in these classification problems.

This simple but efficient approach is widely being used for attribute recognition. Reference algorithms based on Local Binary Pattern and Support Vector Machine (most of them) are listed on Table I. Therefore, these approaches are used to build the classifiers in order to get the users' attributes. Information about how these classifiers are built, classes of each one, images used for training and another additional details can be seen on next section.

E. Gesture recognition for human-computer interface interaction

An intelligent and efficient human-computer interface must recognize motion by a human, involving hands, arms, face, head and/or body in real time [14]. A gesture can be seen as

TABLE I: Attributes and related work

Attribute	Results
Gender	Makinen et al [5] FERET images without normalization: LBP + SVM: 75.56% (without hair) 72.19% (with hair) FERET images with normalization: LBP + SVM 80.56% (without hair) 92.00% (with hair) WWW images with normalization: LBP + SVM 75.32% (without hair) 76.71% (with hair) Lian et al [6] CAS-PEAL face database: LBP + SVM: 96.75% Shan et al [7] LFW database: LBP + SVM: 89.78% (female) 95.73% (male) LBP + Adaboost: 91.98% (female) 95.98% (male) Boosted LBP + SVM : 92.02% (female) 96.64% (male) Yang et al [8] Real AdaBoost(LBPH) 3.68% (error on 5 fold-cross correlation on snapshot dataset) 6.7% (error on FERET for testing) 8.9% (error on PIE for testing)
Age	Yang et al [8] LBP + Adaboost: 6.82% (error on 5 fold-cross correlation on snapshot dataset) 7.88% (error on FERET for testing) 12.5% (error on PIE for testing)
Ethnicity	Yang et al [8] 2-class: Asian/Non Asian LBP + Adabppst: 3.01% (error on 5 fold-cross correlation on FERET and snapshot datasets) 6.8% (error on PIE for testing) Lyle et al [9] 2-class: Asian/Non Asian LBP + SVM: 91% on FRGC face dataset on 5-fold cross validation using periocular images
Emotion	Shan et al [10] 6-class / 7-class (including neutral face) LBP + Template Matching (10-fold cross-validation): 84.5% (6-class), 79.1% (7-class) LBP+SVM (10-fold cross-validation): 87.2% (linear kernel) 88.4% (polynomial kernel) 87.6% (RBF kernel) Shan et al [11] LBP+SVM: Train with CohnKanade and Test with MMI 50.8% (linear kernel) 50.8% (polynomial kernel) 51.1% (RBF kernel) LBP+SVM: Train with CohnKanade Test with JAFFE 40.4% (linear kernel) 40.4% (polynomial kernel) 41.3% (RBF kernel)
Identification	Woodard et al [12] LBP + histogram comparison: left eye 0.9%, right eye 0.9% (using periocular region) LBP + histogram comparison: 0.92% (using full face images) Ahonen et al [13] LBP + histogram comparison: 0.76% (mean of different tests with FERET database)

Attributes and references

a piece of raw or processed sense data that corresponds to an instance of human motion [15].

For the sensing technology there are different approaches: camera, accelerometer [16], [17], or touch screen. With Microsofts launch of Kinect in 2010, and release of Kinect SDK

in 2011, camera sensing technology has been widely used exploring new ways in human-computer interaction. There are also different approaches for the gesture recognizer. The most commonly used are: Dynamic time warping (DTW) and Hidden Markov Models (HMM).

DTW can significantly outperform HMMs when the number of pre-classified gestures is very low (5 or less) [18]. With a sufficient number of pre-classified gestures, the accuracy of both methods is approximately equal [15]. DTW has been widely used due to its simplicity and facility of modification in order to increase its robustness [19], [20].

DTW is a template matching algorithm that warps a time sequence of joints to reference time sequences and is able to cope with different speeds. This advantage is suitable for dynamic gestures, but its a limitation regarding static gestures. If a user is required to hold a certain pose for a given time (2 sec), DTW is not the right method to do it due to its intrinsic design concerning different speeds. Therefore, a static gesture recognizer is added to alleviate this limitation. A complete but simple gesture recognizer is proposed to deal with both static and dynamic gestures for a human-computer interaction. Details on next section.

III. SYSTEM

A. Architecture of the application

The ultimate goal regarding the hardware design of the prototype that combines artificial vision and open data reuse is the non intrusion of the device in the ecosystem of the city. From this point of view, the system will not require a specific place within the Smart City. This prototype is intended to be implanted in everyday systems, leveraging the existing infrastructure of the Smart City. Therefore, it is not required to design a specific structure that could disturb the aesthetics of the city. Thus, the only hardware requirements are a graphic display with suitable dimensions, a camera to capture a constant stream of video to recognize gestures and user attributes and a computer system able to process all information received from the previous system. The screen is the link between the citizen, urban environment and open data. Through it, the user could have a completely satisfying experience.

B. Attribute recognition

Local Binary Pattern operator (LBP) [4] is one of the best texture descriptors and it can be used for face description. The idea of using LBP for face description is motivated by the fact that faces can be seen as a composition of micro-patterns which are well described by such operator [13]. The facial image is divided into local regions and texture descriptors are extracted from each region independently. The descriptors are then concatenated to form a global description of the face. The classification is performed by using Support Vector Machines (SVMs), which had been shown to be superior to traditional pattern classifiers in these classification problems [21]. This approach can be seen in Figure 1.

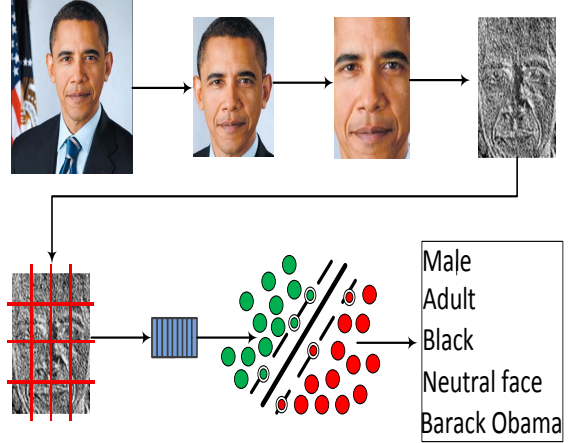


Fig. 1: LBP operator is applied on the normalized face region and a vector descriptor is obtained. SVM is used to classify the descriptor

TABLE II: Databases and attributes

Database	Attributes
AT&T face Database [22]	Gender, Emotion
The Extended Cohn-Kanade Dataset (CK+) [23]	Gender, Emotion
Extended Yale B Database [24]	Gender, Ethnicity
FEI face Database [25]	Gender
FERET Database [26]	Gender, Ethnicity, Age
FG-NET Aging Database [27]	Age
Morph Database [28]	Age
MPLab GENKI Database [29]	Gender, Ethnicity
Indian Face Database [30]	Gender, Ethnicity, Emotion
The Japanese Female Facial Expression Database (JAFFE) [31]	Gender, Ethnicity, Emotion
Labeled Faces in the Wild Database [32]	Gender, Age, Ethnicity, Emotion
Mug facial expression Database [33]	Gender, Age, Emotion
The Images of Groups Dataset [34]	Gender, Age, Emotion

Classifiers are trained with face images obtained from reference databases. Details about what databases are used and images obtained to train every classifier can be seen on Table II. Classes and number of images for every classifier can be seen on Table III.

C. Complete gesture recognizer

Microsoft Kinect sensor is used to obtain joint positions. Kinect SDK tracks 3D coordinates of 20 body joints in real time. Common gestures used in human-computer interaction (Right hand push up, Left hand push up, Right hand pull down, Left hand pull down, Right hand swipe left, Left hand swipe right, Right hand wave and Left hand wave) can be effectively identified using only 6 joints [19], [20]. These 6 joints (left

TABLE III: Attributes, classes and number of images

Attribute Id	Classes	Total of images
Gender	2 (Male/Female)	3000
Age	3 (Young/Adult/Old)	5000
Ethnicity	3 (White/Black/Asian)	5000
Emotion	7 (Seven basic emotions)	7000
Identity	100 (100 different people)	2000

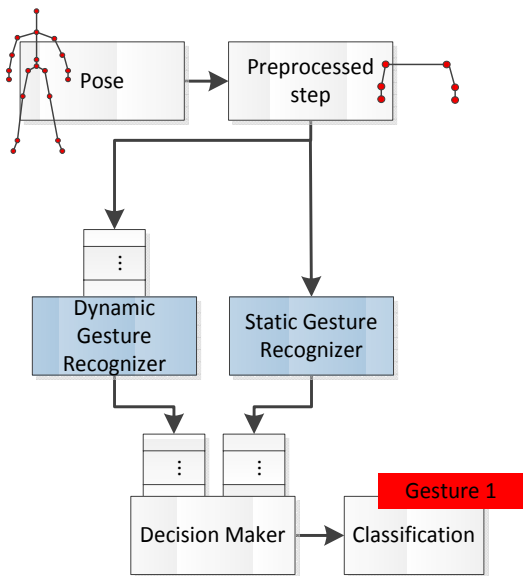


Fig. 2: Static and dynamic gesture recognizer

hand, right hand, left wrist, right wrist, left elbow and right elbow) generate a feature vector of 18 components.

Feature vectors are preprocessed in order to eliminate variations due to the size of the user or its position in the camera field of view. Firstly, a subtraction of the shoulder center from all elements which account for cases where the user is not in the center of the depth image is performed. Secondly, feature vector is normalized with the euclidean distance between the shoulders to account for the variations due to the size of the user.

After preprocessed step, feature vectors are concatenated to create the gesture sequence. The gesture sequence is used to recognize gestures by matching them with pre-stored reference sequences. Both static and dynamic recognizers are fed with this sequence. After, a decision maker is applied in order to get the final gesture. DTW using a weighted distance in the cost computation is applied because this approach outperforms classical DTW [19], [20]. A static gesture recognizer is proposed in order to cope with static gestures over time. Euclidean distance is used to measure the difference between the two sequences at a given time. The decision maker classifies the

final gesture based on which of the two recognizers has detected a gesture for a longer time. It can be seen in Figure 2.

D. System location and conclusions

As already mentioned in previous points, the system will attempt not to have an impact on the environment and ecosystem in the Smart City. There is no need to create a space exclusively for it; it could be integrated into other existing urban elements of the environment. Besides having the goal of not impact the existing ecosystem of Smart City, another goal is that the user can feel more comfortable when using the system in well known locations of their daily lives, thus preventing rejection to the intrusion of a new system in their space. Given the minimum requirements level of system hardware to run the service, it's only needed a graphical display with suitable dimensions and a camera to capture gestures and user attributes. With these requirements, and analyzing the typical environments of cities and observing where users often have long waits on a usual place within the urban space, the objective is to locate the system initially in bus stops, where advertising is usually shown. This location allows users to interact with the system to carry out the experience, without introducing new architectural elements and in a place where future users usually have long waits. As a new element in a known space, it will not cause an initial rejection and will allow users to practice the personalized experience of this service that reuses open data of the Smart City.

Through open data and computer vision techniques new services can be developed to enrich existing ones. This is an important source of potential growth of innovative online services. We believe that introducing computer vision techniques open data will be transformed into valuable tools and enable collaborations across sectors in both public and private settings. Therefore, governments, companies and individuals will take benefit of this approach leading to greater transparency and innovative reuse of public sector information.

REFERENCES

- [1] Doug Riecken et al. Personalized views of personalization. *Communications of the ACM*, 43(8):27–28, 2000.
- [2] Gediminas Adomavicius and Alexander Tuzhilin. Personalization technologies: a process-oriented perspective. *Communications of the ACM*, 48(10):83–90, 2005.
- [3] Jaíre Ederson Passos and Tânia L Koltermann da Silva. La evolución tecnológica y su impacto en el diseño de la interfaz.
- [4] Timo Ojala, Matti Pietikainen, and Topi Maenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987, 2002.
- [5] Erno Mäkinen and Roope Raisamo. An experimental comparison of gender classification methods. *Pattern Recognition Letters*, 29(10):1544–1556, 2008.

- [6] Hui-Cheng Lian and Bao-Liang Lu. Multi-view gender classification using local binary patterns and support vector machines. In *Advances in Neural Networks-ISNN 2006*, pages 202–209. Springer, 2006.
- [7] Caifeng Shan. Learning local binary patterns for gender classification on real-world face images. *Pattern Recognition Letters*, 33(4):431–437, 2012.
- [8] Zhiguang Yang and Haizhou Ai. Demographic classification with local binary patterns. In *Advances in Biometrics*, pages 464–473. Springer, 2007.
- [9] Jamie R Lyle, Philip E Miller, Shrinivas J Pundlik, and Damon L Woodard. Soft biometric classification using periocular region features. In *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, pages 1–7. IEEE, 2010.
- [10] Caifeng Shan, Shaogang Gong, and Peter W McOwan. Robust facial expression recognition using local binary patterns. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 2, pages II–370. IEEE, 2005.
- [11] Caifeng Shan, Shaogang Gong, and Peter W McOwan. Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*, 27(6):803–816, 2009.
- [12] Damon L Woodard, Shrinivas J Pundlik, Jamie R Lyle, and Philip E Miller. Periocular region appearance cues for biometric identification. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 162–169. IEEE, 2010.
- [13] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen. Face description with local binary patterns: Application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12):2037–2041, 2006.
- [14] Sushmita Mitra and Tinku Acharya. Gesture recognition: A survey. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(3):311–324, 2007.
- [15] Scott Steffes. Modern approaches to gesture recognition. 2012.
- [16] Thomas Schlömer, Benjamin Poppinga, Niels Henze, and Susanne Boll. Gesture recognition with a wii controller. In *Proceedings of the 2nd international conference on Tangible and embedded interaction*, pages 11–14. ACM, 2008.
- [17] Jiahui Wu, Gang Pan, Daqing Zhang, Guande Qi, and Shijian Li. Gesture recognition with a 3-d accelerometer. In *Ubiquitous intelligence and computing*, pages 25–38. Springer, 2009.
- [18] Joseph A Kogan and Daniel Margoliash. Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden markov models: A comparative study. *The Journal of the Acoustical Society of America*, 103:2185, 1998.
- [19] Sait Celebi, Ali S Aydin, Talha T Temiz, and Tarik Arici. Gesture recognition using skeleton data with weighted dynamic time warping. *Computer Vision Theory and Applications. Visapp*, 2013.
- [20] Tarik Arici, Sait Celebi, Ali S Aydin, and Talha T Temiz. Robust gesture recognition using feature pre-processing and weighted dynamic time warping. *Multimedia Tools and Applications*, pages 1–18, 2013.
- [21] Vladimir Vapnik. *The nature of statistical learning theory*. springer, 2000.
- [22] AT&T Laboratories Cambridge. The at&t database of faces. 2002.
- [23] Patrick Lucey, Jeffrey F Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 94–101. IEEE, 2010.
- [24] A.S. Georgiades, P.N. Belhumeur, and D.J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 23(6):643–660, 2001.
- [25] Carlos Eduardo Thomaz and Gilson Antonio Giralaldi. A new ranking method for principal components analysis and its application to face image analysis. *Image and Vision Computing*, 28(6):902–913, 2010.
- [26] P Jonathon Phillips, Harry Wechsler, Jeffery Huang, and Patrick J Rauss. The feret database and evaluation procedure for face-recognition algorithms. *Image and vision computing*, 16(5):295–306, 1998.
- [27] The FG-NET Aging Database. <http://www.fgnet.rsunit.com/>. 2002.
- [28] Karl Ricanek and Tamirat Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, pages 341–345. IEEE, 2006.
- [29] <http://mplab.ucsd.edu>. The MPLab GENKI Database.
- [30] Vedit Jain and Amitabha Mukherjee. The indian face database, 2002.
- [31] Michael J Lyons, Julien Budynek, and Shigeru Akamatsu. Automatic classification of single facial images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(12):1357–1362, 1999.
- [32] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.
- [33] Niki Aifanti, Christos Papachristou, and Anastasios Delopoulos. The mug facial expression database. In *Image Analysis for Multimedia Interactive Services (WIAMIS), 2010 11th International Workshop on*, pages 1–4. IEEE, 2010.
- [34] A. Gallagher and T. Chen. Understanding images of groups of people. In *Proc. CVPR*, 2009.

4.1.2. Glasses detection on real images based on robust alignment

- Fernández, A., García, R., Usamentiaga, R., & Casado, R. (2015). Glasses detection on real images based on robust alignment. *Machine Vision and Applications*, 26(4), 519-531.
- DOI: 10.1007/s00138-015-0674-1
- URL: <http://link.springer.com/article/10.1007/s00138-015-0674-1>
- Para el año 2015, la revista MACHINE VISION AND APPLICATIONS (ISSN: 0932-8092) tuvo un índice de impacto de 1.272 (Q2) y un índice de impacto a 5 años de 1.401.

Glasses detection on real images based on robust alignment

Alberto Fernández¹ · Rodrigo García¹ · Rubén Usamentiaga² · Rubén Casado³

Received: 28 May 2014 / Revised: 27 February 2015 / Accepted: 13 March 2015 / Published online: 31 March 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Automatic glasses detection on real face images is a challenging problem due to different appearance variations. Nevertheless, glasses detection on face images has not been thoroughly investigated. In this paper, an innovative algorithm for automatic glasses detection based on Robust Local Binary Pattern and robust alignment is proposed. Firstly, images are preprocessed and normalized in order to deal with scale and rotation. Secondly, eye glasses region is detected considering that the nosepiece of the glasses is usually placed at the same level as the center of the eyes in both height and width. Thirdly, Robust Local Binary Pattern is built to describe the eyes region, and finally, support vector machine is used to classify the LBP features. This algorithm can be applied as the first step of a glasses removal algorithm due to its robustness and speed. The proposed algorithm has been tested over the Labeled Faces in the Wild database showing a 98.65 % recognition rate. Influences of the resolution, the alignment of the normalized images and the number of divisions in the LBP operator are also investigated.

Keywords Glasses detection · Face alignment · Robust Local Binary Pattern · Face image processing

1 Introduction

Automatic face recognition is an attractive research issue due to its great potential in real-life applications. However, robust face recognition is not a trivial task and occlusions (like wearing glasses) can decrease the performance of an automatic face recognition system. In some cases, certain glasses can identify a person or in other cases, act like soft biometrics for person identification (matrix glasses, U2 glasses). Automatic gender recognition suffers from the same problems as face recognition. However, there is certain appearance difference between men glasses and women glasses, which could help to distinguish a person's gender. Even, glasses appearance between a young person and an older one could improve an automatic age classification algorithm rather than decrease its performance. Many security systems require the use of glasses, their use is mandatory, and intelligent advertising and marketing could also take advantage of this. Last but not least, glasses can provide valuable information. For instance, glasses could reveal the activity a person is doing at a given time (safety glasses, sport glasses, snow glasses, sunglasses, underwater glasses, diving mask, reading glasses, eyeglasses, Google glasses, etc).

Automatic recognition of eyeglasses on facial images needs a thorough study. Different types of glasses on real images can be seen in Fig. 1. The inherent appearance variation on glasses frames, which are even not clearly distinguishable (Fig. 1d), makes it a challenging problem. Further challenges are the different types of glasses: sport glasses (Fig. 1a, b, h) or sunglasses (Fig. 1g, i), different sizes (Fig. 1e, f), reflections and glare (Fig. 1c) or rotation of the face image (Fig. 1d).

✉ Alberto Fernández
alberto.fernandez@fundacionctic.org

Rodrigo García
rodrigo.garcia@fundacionctic.org

Rubén Usamentiaga
rusamentiaga@uniovi.es

Rubén Casado
ruben.casado@treelogic.com

¹ Fundación CTIC (Technological Center), Technological Scientific Park of Gijón, 33204 Gijón, Asturias, Spain

² Department of Computer Science and Engineering, University of Oviedo, Campus de Viesques, 33204 Gijón, Asturias, Spain

³ Treelogic, Technological Scientific Park of Asturias, 33428 Lanera, Asturias, Spain



Fig. 1 Glasses can convey information about the user's context. Different variations in glasses appearance can be seen. These images are obtained from LFW database

Summarizing, the main areas that can benefit from an automatic eyeglasses recognition on facial images would be the use of safety glasses in security systems, intelligent advertising and marketing, soft biometrics for person identification, face recognition, gender classification, age classification, hints of activity detection based on type of glasses, Google glasses recognition and first step of automatic glasses removal.

The main contributions are summarized as follows. Firstly, the problem of eyeglasses detection needs an up-to-date review. Main studies on glasses detection are summarized and analyzed. Secondly, a LBP-based algorithm is developed to cope with the glasses detection problem and a simple yet robust algorithm is proposed in order to deal with face alignment normalization. Lastly, (1) influences of resolution of the image, (2) alignment of faces, (3) number of divisions of the LBP operator and (4) a performance comparison between LBP and Robust LBP are analyzed.

2 Related work

2.1 Glasses detection, extraction and removal

Facial image analysis has stimulated increasing interest, including tasks as face detection, face recognition, facial

expression analysis, demographic classification and other related topics. Although there are less research papers focusing on glasses detection, there are some studies in this regard (a short description of the results of main studies on glasses detection can be seen in Table 1).

In [1], six measures for the likelihood of glasses in different regions for detecting the presence of glasses are introduced. Furthermore, a combination of them turns out to consistently improve the performance of the individual measures. Experiments showed that measure from the region containing the nosepiece of the glasses is the most powerful criterion for glasses recognition. In another work [3], glasses detection and extraction are proposed. Detection is carried out using edge information within a small area defined between the eyes. Detection is similar to the method proposed in [1]. As mentioned in [1], the nosepiece is one of the most common features existing on all of the glasses. Extraction is achieved with a deformable contour, combining edge and geometrical features. In [11], glasses detection and removal for face recognition using Bayes rules are proposed. Experiments within a face image database showed that this method is effective and has a better performance than that proposed in [3]. Extraction is realized using Bayes rules that incorporate the features around each pixel and the prior knowledge on glasses features that were learnt and stored in a database. Glasses removal is achieved with an adaptive median filter conducted in the points classified as glasses.

A method that makes use of the 3D features obtained by a trinocular stereo vision system is proposed in [6] to perform glasses frame detection with 3D Hough transform. It is based on the fact that the rims of a pair of glasses lie on the same plane in 3D space. This method requires more cameras and computational time than other 2D image-based methods. In [7], a framework for classifying facial attributes is presented. Similarity is measured by the reconstruction error motivated by the Eigenface-based methods, and thus, this paper extends the Eigenface method by representing each attribute by using a 'facial attribute-specific subspace (FASS)'.

Wavelet feature-based boosting methods can also be applied to glasses recognition [5]. This method was originated from the work on face detection [9]. The glasses detectors use a variation of the original boosting algorithm called real AdaBoost [12]. Authors suggest that the nose-piece of the glasses frame is an important clue for glasses detection. An algorithm for eyeglasses detection, localization and removal is proposed in [8]. An eye region detector, trained offline, is used to approximately locate the region of the eyes and thus the region of the eyeglasses. The location of the eyeglasses pattern is defined as twice as big as the detected eye area. Afterward, eyeglasses localization is done with a Markov chain Monte Carlo (MCMC) method. A Delaunay triangulation-based method is proposed to locate glasses from a front-up face image [13]. They decompose

Table 1 Results of main studies on glasses detection from face images

Jiang et al. [1]	Forerunner algorithm on detection of glasses in facial images. Results are reported similar in [2]
Dantcheva et al. [2]	Train: FERET database. Test performed on a testing set of images on FERET database. Detection rate: 87.7 %. True-positive rate: 7.17 %. False- positive rate: 5.66 %
Jing et al. [3]	Tested on a set of 419 people, 151 wearing glasses under different imaging conditions. Detection rate: 99.52 %, falsely detecting the presence of glasses in two facial images
Vaquero et al. [4]	1044 internet images were used for training (308 sunglasses, 318 eyeglasses, 418 no glasses). Tested on a surveillance video. Sunglasses (precision: 85 %, recall: 35 %), eyeglasses (precision: 90 %, recall: 8.9 %), no glasses (precision: 77 %, recall: 99.5 %)
Wu et al. [5]	3000 images (1500 wearing glasses and 1500 with no glasses) obtained from FERET database and world wide web were used for fivefold cross-validation. Best results were obtained using gabor boost features. Accuracy: 98.9 % (98.1 % with glasses and 99.7 % without glasses). They implemented an automatic glasses detection system whose input is real-life photographs and the overall correctness is 94.0 %
Wu et al. [6]	513 facial images of 19 people wearing three kinds of glasses and nine kinds of pose were used for training. Detection rate: $\approx 90\%$
Shan et al. [7]	100 face images from Bern face database were used for training: 50 examples from five subjects wearing glasses were used to learn glasses-FASS, and 50 examples from five subjects no wearing glasses were used to learn non-glasses-FASS. Tested on 200 images from Bern face database with an average correct rate of 77 %
Wu et al. [8]	They train a detector for the eye region based on the face detector [9] with 12710 samples. They tested the detector with 1386 face images. Detection rate: 96 % and false alarm at 10^{-4}
Heo et al. [10]	Correct detection rate was 86.6 % for the subjects wearing eyeglasses. For the face images with no eyeglasses, 97.1 % true negative accuracy was achieved. False-positive and false-negative errors were 2.9 and 13.4 %, respectively. The database used in this experiment is comprised of thermal images from the database developed by the National Institute of Standards and Technology (NIST) and Equinox Corporation

the face shape using the Delaunay triangulation. In the first step of the algorithm, they perform a binarization of the face. This method can have some problems if glass frames have

very low contrast. In addition, they do not obtain very good results if the glasses do not cover eyes. Data-fused image composed of visual and thermal ones is used in [10] to produce an integrated image for eyeglass removal in order to perform a robust face recognition. The eyeglass region in thermal face images can be represented by two ellipses.

A method for removing glasses from a human frontal face image is proposed [14]. Glasses region is automatically extracted using color and shape information. They perform glasses removal, but not glasses detection. Firstly, glasses region is automatically extracted using color and shape information, and then, a natural looking facial image without glasses is generated by means of recursive error compensation using PCA reconstruction. They carried out some experiments, and the recognition performance is improved by using glasses removal methods. In [4], authors trained nine Viola and Jones detectors, one for each facial attribute. Glasses type is performed. They distinguish three types: sunglasses, eyeglasses and the absence of them.

Soft biometrics (eyeglasses, for example) can be used for person identification [2]. They implement an algorithm deduced from [1], which performs edge detection on a pre-processed gray-level image.

2.2 Face alignment and normalization

Alignment step of a face recognition algorithm is often ignored or not detailed. In some cases, the alignment is done manually, where the positions of the eyes are manually labeled. In other cases, this step is ignored, under the assumption that the face detection algorithm will perform some kind of alignment. Many recognition algorithms depend on an accurate positioning of the face region into a canonical pose before the recognition step takes place. This alignment process can lead to improve recognition accuracy on real images, even for algorithms robust to misalignments [16]. In [17], authors propose a method for aligning images using poorly aligned examples of a class with no additional data generating a alignment machine for that object class. In [18], authors use a combination of unsupervised joint alignment with unsupervised feature learning in order to align images.

Some studies are done in relation to the alignment of the images. It was found that automatic alignment methods did not increase gender classification rate, while manual alignment increased the classification rate a little [19]. Therefore, they concluded that automatic alignment needs improving. This statement, valid at that time, should be reviewed, because automatic alignment methods have improved in both accuracy and robustness.

A face normalization algorithm based on eyes detection is proposed [20]. The algorithm detects the position of the pupils in the face image. After that, the algorithm normalizes the orientation, the scale and the gray scale of the face

image. Face region is adjusted based on the position of the mouth. This algorithm has the drawback it needs to locate the position of the pupils, and thereby, glasses and partial occlusions can alter the result of the normalization step. Another face normalization algorithm is proposed [21]. The eyes are located manually in the faces, and faces are rotated and aligned in the images so that each face has eyes in the same location. After the normalization step, images are scaled to a fixed size.

In this work, a face alignment and normalization algorithm is proposed. In order to get the angle of misalignment, a regression line is calculated based on four points of the eyes. There are some automatic facial feature points detector [22], [23], but the detector of facial landmarks used in this paper [15] takes advantage of relationship among some facial features of the face (eyes, nose and mouth) in terms of appearance and structure distribution. In this way, this algorithm can estimate the position of the eyes under occlusion (like wearing glasses or sunglasses).

3 Methodology

In Fig. 2, the main steps of the algorithm, which are discussed below, can be seen. First of all, this work starts with location of the face using the algorithm by Viola and Jones [9]. After the face is located in the image, some preprocessing is necessary in order to deal with pose, rotation, scale and inaccuracies of the located face. A face normalization algorithm is applied to get the region around the eyes. Afterward, LBP and Robust LBP are applied in order to get the feature sets. Finally, support vector machine (SVM) is applied on the classification step. SVM is applied to classify the extracted feature histograms over the normalized eyes glasses regions. The output of the SVM classifier is a two-class classification problem, i.e., glasses vs no glasses. Therefore, different type of glasses (sport glasses, sunglasses, safety glasses, reading glasses, etc.) belong to the same category. SVMs are a useful technique for data classification and have been proven useful in a number of pattern recognition tasks including face and facial recognition [24]. LIBSVM was used for the training and testing of SVMs [25].

3.1 Face alignment and normalization

3.1.1 Landmarks relationship to estimate eyes position

When the shape or intensity characteristics of the eyes cannot be reliably measured due to occlusion (like wearing glasses or sunglasses), the context characteristics are very useful for eye localization. This is because eyes in the face context usually have stable relationship with other facial features in terms of both appearance and structure distribution [26]. In this way,

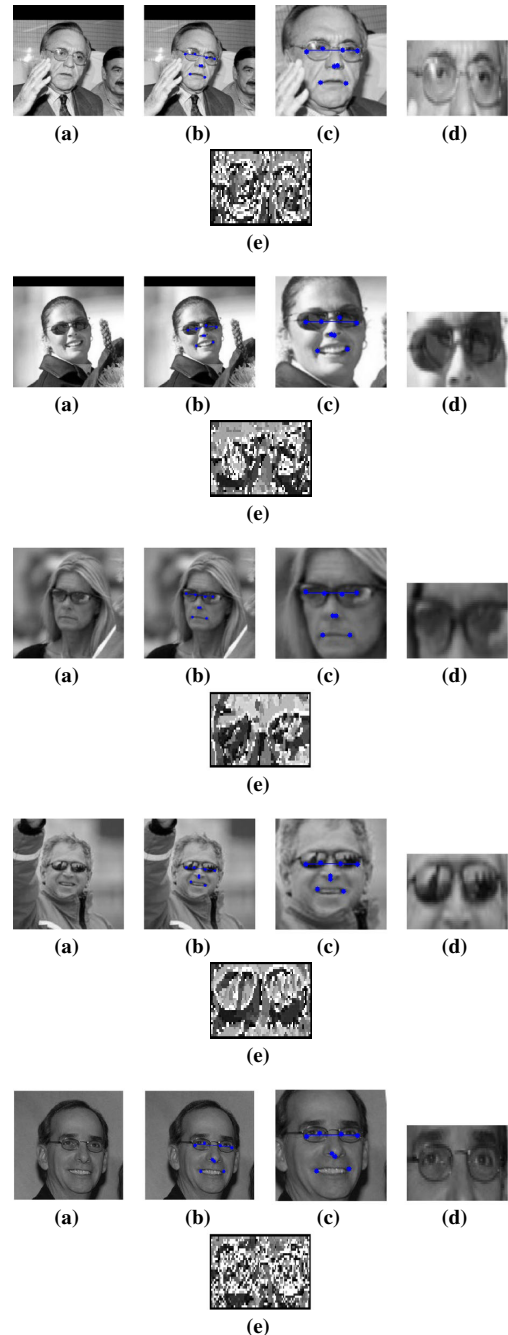


Fig. 2 Pipeline of the algorithm for glasses detection on facial images: **a** input image obtained from LFW database, **b** landmarks detection and regression line, **c** rotated face, **d** normalized eye glasses region and **e** Local Binary Pattern

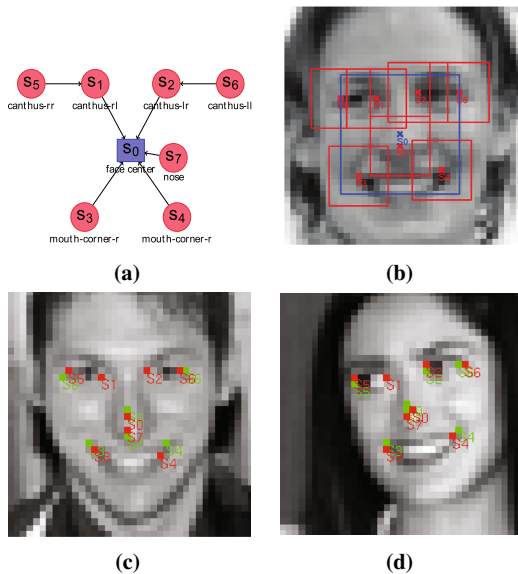


Fig. 3 Definition of **a** underlying graph for the landmark configuration and **b** the components of the detector. Sample images **c** and **d** where the estimated landmark positions have the mean normalized deviation equal to 10 %. The *green* and *red* points denote the manually annotated and estimated landmarks, respectively. These images are obtained from [15] (color figure online)

a detector of facial landmarks learned by structured output SVM [15] has been applied in order to detect the positions of the main points of the face and hence of the eyes.

The input of the classifier is a still image of fixed size containing a single face. The output of the classifier is estimated locations of a set of facial landmarks. A graph constraint with eight components is taken into account see Fig. 3a. Each facial feature is computed on a rectangular window see Fig. 3b. The accuracy of the classifier has been tested on LFW database [27]. The detector estimates around 97 % of the images with the mean normalized deviation between the estimated and the ground truth positions < 10 %. In order to clarify this measure, two sample images (see Fig. 3c, d) are shown, where the estimated landmark positions have the mean normalized deviation equal to 10 %.

3.1.2 Procedure

In order to perform a robust, fully automatic ROI normalization a new algorithm is proposed in Algorithm 1. The input of the normalization algorithm is a still image. This image is resized to a certain size. Different tests are performed in order to know how this size affects the recognition rate (see Sect. 4). This algorithm is based on [21] but fully automatic, where:

- *eyes_distance_r* is the ratio of the distance from left outer eye to right outer eye in the resized image.
- *eye_line_r* is the ratio of the height above and below eyes.
- *size* is the size of the resized face image
 - *size.w* is the width of the resized image
 - *size.h* is the height of the resized image

Main steps are summarized as follows. Eight facial landmarks are obtained [15]. In order to get the angle of misalignment, a regression line is calculated based on four points: the canthi of the left and right eye, i.e., left inner eye, left outer eye, right inner eye and right outer eye (see Fig. 4). After that, faces are rotated and aligned in the images so that eyes are located in the same coordinates for all the images. The area around the eyes is calculated in source image. Once the coordinates for the corners of the eye glasses region in the rotated image are calculated, the resulting area is cropped from the rotated image.

Algorithm 1 Face ROI normalization

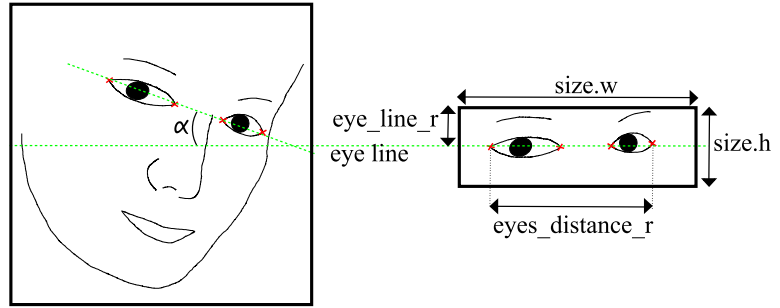
eyes_distance_r, *eye_line_r*, *size*

- 1: Eight facial landmarks are located. Linear regression is calculated based on the four landmarks of the eyes. Angle α of misalignment is calculated, and image is rotated based on α in order to align the source image
- 2: Euclidean distance d_0 between the eyes is calculated in the rotated image
- 3: Distance of the eyes in the resized image is calculated by $d_r = size.w * eyes_distance_r$
- 4: Ratio r is calculated by $r = d_0/d_r$
- 5: Width w_0 and height h_0 of the area around the eyes are calculated by $w_0 = r * size.w$ and $h_0 = r * size.h$
- 6: Coordinates for the corners of the face area in the rotated image are calculated by $x_l = x_c - w_0/2$, $y_l = y_e - h_0/eye_line_r$, $x_r = x_l + w_0$ and $y_b = y_l + h_0$, where x_l is x-coordinate of the left border, x_e is x-coordinate of the point in the halfway between the eyes, y_l is y-coordinate of the top border, y_e is y-coordinate of the eyes, x_r is x-coordinate of the right border, and y_b is y-coordinate of the bottom border. This *ROI* is cropped from the image
- 7: **return** *ROI*

3.2 Local Binary Pattern

The Local Binary Pattern operator is a type of feature used for classification. It has been found to be a powerful feature for texture classification. It was introduced in 1996 as a means of summarizing local gray-level structure [28]. The operator takes a local neighborhood around each pixel, thresholds the pixels of the neighborhood at the value of the central pixel and uses the resulting binary-valued image patch as a local image descriptor. It was originally defined for 3×3 neighborhoods, giving eight bit codes based on the eight pixels around the central one. Formally, the LBP operator takes the form:

Fig. 4 Linear regression based on four landmarks of the eyes



$$\text{LBP}(x_c, y_c) = \sum_{p=0}^7 2^p s(g_p - g_c) \quad (1)$$

where in this case, p runs over the eight neighbors of the central pixel c , g_c and g_p are the gray-level values at c and p and

$$s(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

The LBP encoding process is illustrated in Fig. 5. The operator was extended to use neighborhoods of different sizes [29], thus making it feasible to deal with textures at different scales. This is denoted by (P, R) where P represents the number of sampling points and R represents the radius of the neighborhood. When the sampling points do not fall at integer positions, the intensity values are bilinearly interpolated. See Fig. 6 for an example of the circular $(8, 2)$ neighborhood. This implementation is called circular LBP ($\text{LBP}_{P,R}$):

$$\text{LBP}_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} 2^p s(g_p - g_c) \quad (3)$$

Another extension to the original operator [29] defined the so-called uniform patterns: An LBP is 'uniform' if it contains at most two bitwise transitions from 0 to 1 or vice versa when viewed as a circular bit string. For example, 00000000, 00011110 and 10000011 are uniform patterns. Uniformity is an important concept in the LBP methodology, representing primitive structural information such as edges and corners. Although only 58 of the 256 8-bit patterns are uniform, nearly 90% of all observed image neighborhoods are uniform [29]. The following notation for the uniform LBP operator is used: $\text{LBP}_{P,R}^u$. Because the gray value of the central pixel is used as threshold, LBP is sensitive to noise, especially in the near-uniform image regions. Another demerit of LBP is that many different structural patterns may have the same LBP code. In order to enhance the discriminative capability of the local structure, Completed LBP (CLBP)

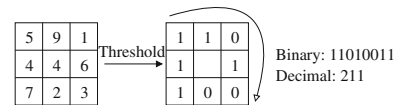


Fig. 5 The basic LBP operator

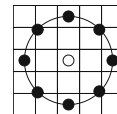


Fig. 6 The circular $(8, 2)$ neighborhood. The pixel values are bilinearly interpolated whenever the sampling point is not in the center of a pixel

was proposed [30]. Although CLBP solves some confusion of different patterns, not all of these patterns can be differentiated perfectly. Besides, CLBP is sensitive to noise since the value of a pixel is still used as a threshold directly.

In order to solve these difficulties, Completed Robust Local Binary Pattern (CRLBP) was proposed [31]. This descriptor is used to extract the region glasses information. Firstly, Average Local Gray Level (ALG) is calculated as follows:

$$\text{ALG} = \frac{\sum_{i=1}^8 g_i + g}{9} \quad (4)$$

where g represents the gray value of the center pixel and g_i denotes the gray value of the neighbor pixel. ALG represents the average gray level of local texture, which is obviously more robust to noise than the gray value of the central pixel. ALG ignores the specific value of an individual pixel, while sometimes the specific information of the central pixel is needed. To make a balance between anti-noise and information of individual pixel, a Weighted Local Gray Level (WLG) is defined as follows:

$$\text{WLG} = \frac{\sum_{i=1}^8 g_i + \alpha g}{8 + \alpha} \quad (5)$$

where α is a parameter set by user. Now the RLBP can be calculated as follows:

$$\text{RLBP}_{P,R} = \sum_{p=0}^{P-1} 2^p s(g_p - \text{WLG}_c) \quad (6)$$

That is the same as:

$$\text{RLBP}_{P,R} = \sum_{p=0}^{P-1} 2^p s\left(g_p - \frac{\sum_{i=1}^8 g_i + \alpha g_c}{8 + \alpha}\right) \quad (7)$$

After labeling an image with the different LBP operators, a histogram of the labeled image $f_i(x, y)$ can be defined as:

$$H_i = \sum_{x,y} I\{f_i(x, y) = i\}, \quad i = 0, \dots, n-1 \quad (8)$$

where n is the number of different labels produced by the LBP operator and:

$$I\{A\} = \begin{cases} 1, & \text{if } A = \text{true} \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

For efficient face representation, feature extracted should retain also spatial information. Hence, the facial image is divided into m regions $\{R_0, \dots, R_{m-1}\}$. In this way, the basic histogram defined above can be extended into a *spatially enhanced histogram* [32], which encodes both the appearance and the spatial relations of facial regions. The spatially enhanced histogram is defined as:

$$H_{i,j} = \sum_{x,y} I\{f_i(x, y) = i\} I\{(x, y) \in R_j\} \quad (10)$$

where $i = 0, \dots, n-1$, $j = 0, \dots, m-1$. The histogram has a description of the normalized eye glasses region on three different levels of locality: The labels for the histogram contain information about the patterns on a pixel level, the labels are summed over a small region to produce information on a regional level, and the regional histograms are concatenated to build a global description [32]. In conclusion, the extracted feature histogram represents the local texture and global shape of the normalized eye glasses region.

4 Results

A series of experiments are conducted on LFW database [27]. LFW database is a database originally created to study the problem of unconstrained face recognition, which contains 13,233 face images of 5749 distinct subjects collected from the web. This database has also recently been used as benchmark for another attribute recognition algorithms [33], [34].

In this paper, the use of this database as benchmark for glasses recognition in real images is carried out. Three thousand face images (1500 wearing glasses, 1500 not wearing glasses) are considered in the experiment; see Fig. 1 for some examples. All experimental results were obtained using the commonly used fivefold cross-validation in order to use standardized practices [33].

In order to get a robust and computationally lightweight (efficient) algorithm, a good trade-off between recognition performance and feature vector length should be addressed. There are some parameters that can be chosen to optimize this algorithm: LBP operator, radius of the operator, number of neighbors, multiresolution analysis and number of divisions.

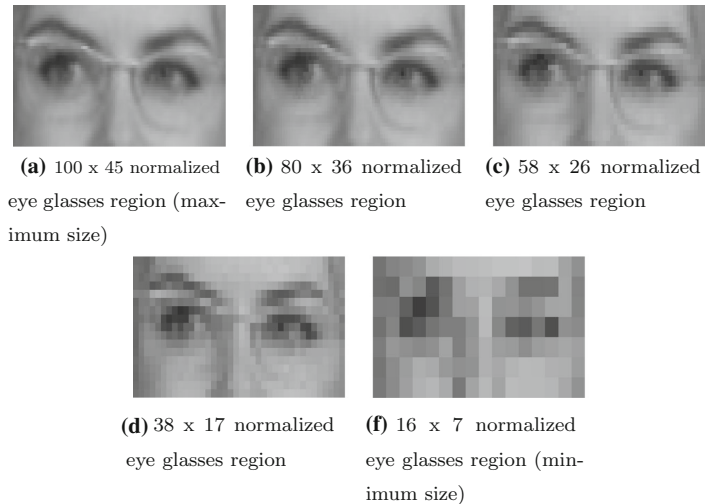
Uniform patterns produces 59 labels for a neighborhood of eight pixels and produces 256 labels for standard LBP. Additionally, for the 16 neighborhoods, the numbers are 243 and 65,536, respectively. In experiments with facial images [32], it was found that 90.6 % of the patterns in the (8, 1) neighborhood and 85.2 % of the patterns in the (8, 2) neighborhood are uniform. Using uniform patterns instead of all the possible patterns has produced better recognition results in many applications [35]. By varying sampling radius R , LBP of different resolutions can be obtained, and thus, multiresolution analysis can be accomplished by combining the information provided by multiple operators varying (P, R) [36]. However, for most of the existing work, LBP with a fixed radius ($R = 2$) was applied. A small radius of the operator makes the information encoded in the histogram more local [37].

Preliminary tests carried out in this work showed that $\text{LBP}_{8,2}^{u2}$ gives better results than other combinations. Mean recognition rates achieved on these preliminary tests for the LBP indicated that multiresolution analysis is not necessary for glasses detection.

4.1 Resolution and size of the normalized region

In many practical video surveillance applications, the faces acquired by outdoor cameras are of low resolution [38], so tests are carried out in order to establish the minimum resolution of the images. First of all, taken into account the set of selected images from LFW database, the average size of the facial regions is calculated to establish an approximate size of the facial region in the algorithm. The average distance from left outer eye to right outer eye is 63.86 pixels and a standard deviation of 5.39 pixels $\left(5.39 = \sqrt{\frac{\sum (63.86 - x)^2}{3000 - 1}}\right)$. In order to decrease the amount of data to process and to avoid false-positive recognition, eye glasses region is selected taken into account only the eyes zone. A symmetrical region around the center of the eyes is selected considering that the nosepiece of the glasses is usually placed at the same level as the center of the eyes in both height and width. The height of the eye glasses region is heavily determined by the size of the

Fig. 7 Normalized eye glasses regions at different resolutions. Original image corresponds with Fig. 1e and normalized eyes glasses region matches with Fig. 13k



glasses. The width is determined taking into account the distance between the eyes and the sides of the glasses. In these experiments, the width of the normalized eye glasses region is 122 % bigger than the height of the normalized region (see Fig. 13 for some examples).

Since the average distance from left outer eye to right outer eye is 63.86 pixels, the maximum tested width for the normalized eye glasses region is 100 pixels and a height of 45 pixels. So, the size of the maximum normalized eye glasses region is 4500 pixels. The minimum tested width is 16 pixels and a height of seven pixels. So, the size of the minimum normalized eye glasses region is 112 pixels. See Fig. 7 for some examples. In Fig. 8, different tests are performed varying from the maximum to the minimum size of the normalized eye glasses region. It can be appreciated that the relationship between the size of normalized eye glasses region and the recognition rate follows a logarithmic distribution. This logarithmic distribution depends on the number of division of the LBP operator. Two examples can be seen in Fig. 9. $R^2 = \{0.9880, 0.9875, 0.9826, 0.9812, 0.9897, 0.9880, 0.9876, 0.9891, 0.9915, 0.9898\}$ correspond to the correlation coefficients obtained from the different LBP series from Fig. 8.

4.2 Number of divisions of the LBP operator

Since the LBP histogram is calculated over each rectangular region of the image, using a small number of divisions makes the feature vector shorter, but also means losing spatial information. In order to encode the spatial information, face image is usually divided into a grid of non-overlapping regions $\{R_0, \dots, R_{m-1}\}$. So, the length of the feature vector becomes $B = 59 \times m$ using uniform patterns, in which m is the number of regions. The mean recognition rates for the

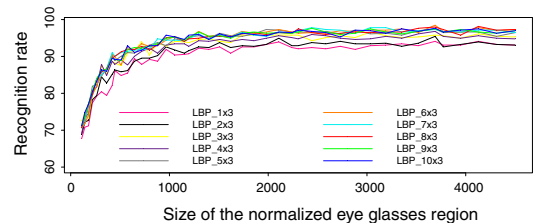


Fig. 8 Results varying from the maximum to the minimum size of the normalized eye glasses region

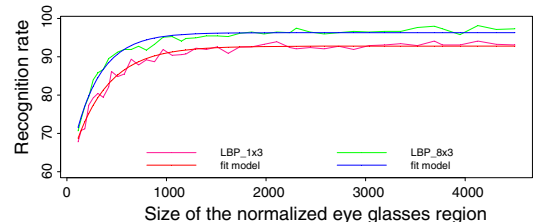


Fig. 9 Relationship between size of normalized eye glasses region and recognition rate follows a logarithmic distribution. Two examples are shown

$LBP_{8,2}^{u2}$ as a function of the number of regions is plotted in Figs. 10 and 11. When looking for the optimal number of regions, it is observed the changes in the number of regions may cause big difference in the length of the feature vector, but the performance is not necessarily affected significantly. However, it is noticed the divisions in the width of the image have a greater impact on the performance than the divisions in the height of the image. It makes sense due to the characteristics of the normalized eyes region.

	1	2	3	4	5	6	7	8	9	10
1	88.53	88.36	91.23	89.38	90.56	92.58	91.23	92.24	91.91	92.92
2	91.06	89.54	93.09	90.56	93.09	93.25	93.59	94.10	93.09	94.27
3	90.39	91.06	93.09	92.92	95.45	96.63	95.95	95.95	94.94	95.78
4	90.39	89.54	94.60	92.92	95.95	95.95	96.29	96.80	95.45	95.95
5	91.57	90.73	94.94	92.75	95.78	96.96	96.46	96.80	95.95	96.63
6	89.88	90.22	94.44	92.75	96.12	95.78	96.63	96.96	95.62	96.12
7	89.38	89.71	93.93	92.41	95.11	94.94	95.95	95.95	95.45	95.78
8	89.88	90.22	93.42	92.58	95.11	95.62	95.45	96.12	95.28	95.95
9	89.88	89.71	93.76	92.58	94.94	95.62	95.45	95.95	94.94	95.95
10	89.38	90.89	93.25	91.40	93.93	94.60	94.44	94.60	94.94	95.28

Fig. 10 Recognition rates as a function of the number of regions. No alignment of the images is carried out

	1	2	3	4	5	6	7	8	9	10
1	88.36	88.53	93.42	91.57	93.76	94.10	93.59	94.60	93.09	94.77
2	91.23	91.40	93.76	92.58	94.10	94.94	94.27	94.94	94.44	94.60
3	91.74	93.59	95.11	94.77	96.96	96.80	96.29	96.63	96.29	97.13
4	92.41	93.25	95.45	95.11	95.78	95.78	95.78	96.96	96.46	96.29
5	92.92	93.59	95.45	94.94	96.46	96.80	96.80	97.30	96.29	97.64
6	93.59	93.59	96.29	95.28	96.63	97.13	96.63	97.64	96.12	96.46
7	93.42	94.10	96.12	95.28	96.80	96.96	96.80	96.80	96.80	97.30
8	92.75	94.60	96.29	94.60	96.80	96.63	96.63	96.46	96.63	96.63
9	93.09	94.27	95.95	95.28	96.29	96.63	96.46	95.95	96.46	96.46
10	92.92	93.09	95.95	94.60	96.63	97.13	95.78	96.12	96.80	96.80

Fig. 11 Recognition rates as a function of the number of regions. Alignment of the images is carried out

4.3 Alignment of the images

A performance comparison between alignment of the faces and non alignment can be seen in Figs. 10 and 11. It can be seen that the recognition rate improves if the alignment step is performed, which is increased by an average of 1.55 %.

4.4 Varying α on the Robust Local Binary Pattern algorithm

A set of experiments is carried out to select the optimal parameter α (see Fig. 12). RLBP introduced a parameter α in order to make a balance of robustness and stability. It should be pointed out that RLBP performs more stably under complex illumination and viewpoint variations, since it extracts

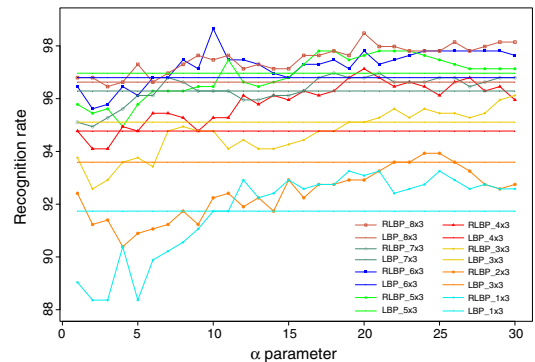


Fig. 12 LBP and RLBP comparison varying α between [1–30]

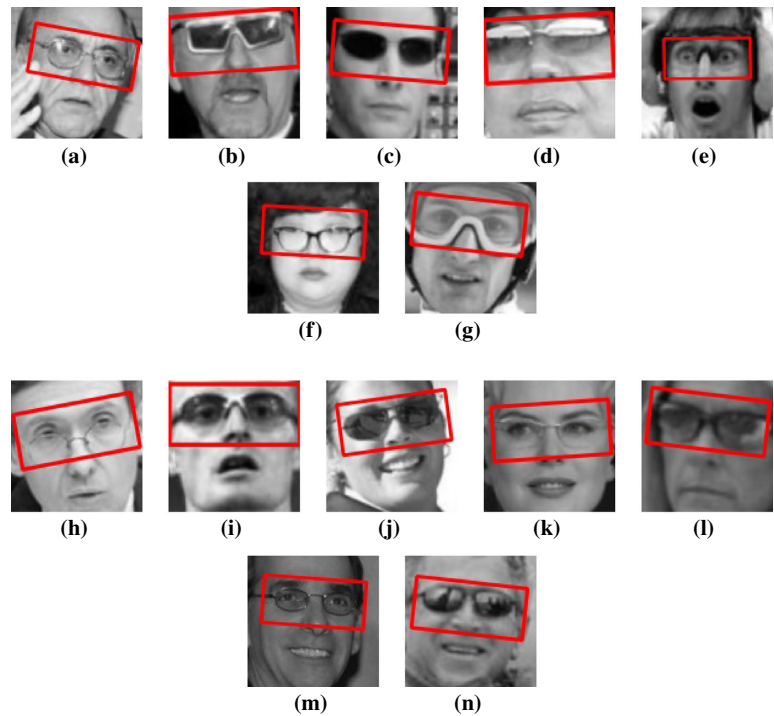
the gray-level information of both local neighbor set and individual pixel. RLBP outperforms standard LBP.

5 Discussion

The proposed algorithm has been tested on real images over the Labeled Faces in the Wild database achieving a 98.65 % recognition rate (Fig. 13). Most algorithms are not tested with real-life photographs. In [5], the overall correctness is 94.0 % on real-life photographs. In [4], the system was tested on a surveillance video with real conditions and the best results are detecting eyeglasses with a precision of 90 % and a recall of 8.9 %.

As far as authors know, this is the first time this database is proposed for comparison of glasses detection. So, comparisons with our algorithm using this database are not possible. But in order to see that the proposed method advances the state of the art, the algorithm has also been tested on FERET database [39] because some of the previous methods were trained/tested with this database [2], [5]. A detection rate of 87.7 % was obtained by the algorithm proposed by Dantcheva et al. [2]. Wu et al. [5] achieved a remarkable accuracy of 98.9 % using wavelet features. It is worth mentioning that the algorithm proposed in [5] was trained/tested not only with FERET database images but also with other World Wide Web images. The results of our algorithm on FERET database achieved 99.89 and 99.83 % recognition rate with RLBP and LBP, respectively. Some examples can be seen in Figs. 14 and 15. In most images, the results are like the ones shown in Fig. 14, that is, the normalized eye glasses region are calculated on FERET face images based on the accurately position of the facial landmarks. In Fig. 15, even when some of the four landmarks of the eyes are not accurately detected, the normalized eye glasses region algorithm can operate robustly. This fact can be appreciated even better in Fig. 15g, h. The algorithm used for detecting the landmarks did not perform

Fig. 13 Normalized eye glasses region for all images from LFW database contained in this work



well probably because these images exhibit a profile view pose. Even so, due to the fact that the relationship among facial features could approximate the location of the eyes, the eye glasses normalization algorithm can operate quite well. The failure case from FERET database can be seen in Fig. 15i. So, in conclusion, the proposed glasses region normalization algorithm has the capabilities of detecting glasses region with different poses and under different conditions.

Furthermore, since a low-resolution face contains limited information [38], several tests are performed in order to know how the resolution of the face images affects recognition rate. A total of forty-three different resolutions of the normalized eye glasses region were studied in this research. The lower-resolution images were down-sampled from original images. It should also be noted that the relationship between the size of normalized eye glasses region and the recognition rate follows a logarithmic distribution. In other words, the LBP features perform robustly and stably over a useful range of low resolutions.

In case of alignment of the face images, an average recognition rate of 95.19 % is achieved (average value from Fig. 11). The best average performance is achieved with ten division in the width (average performance of 96.40 %) and seven division in the height (96.03 %). Additionally, if the choice is not to align the faces, an average recognition rate of 93.64 % is performed (average value from Fig. 10). The

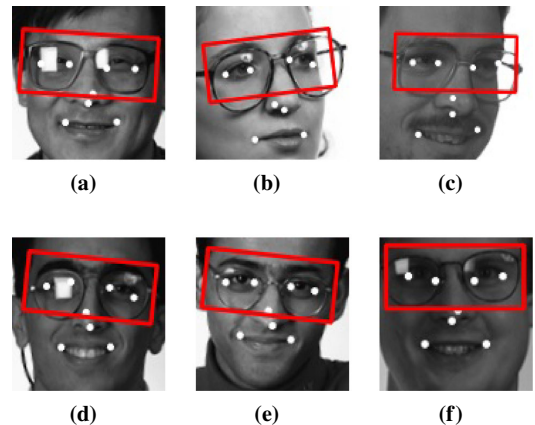


Fig. 14 The normalized eye glasses region is calculated on FERET face images based on the accurate position of the facial landmarks

best average performance is achieved with eight divisions in the width (average performance of 95.54 %) and five divisions in the height (average performance of 94.85 %). It can therefore be concluded that a large number of divisions can decrease the performance due to the nonalignment of the face images.

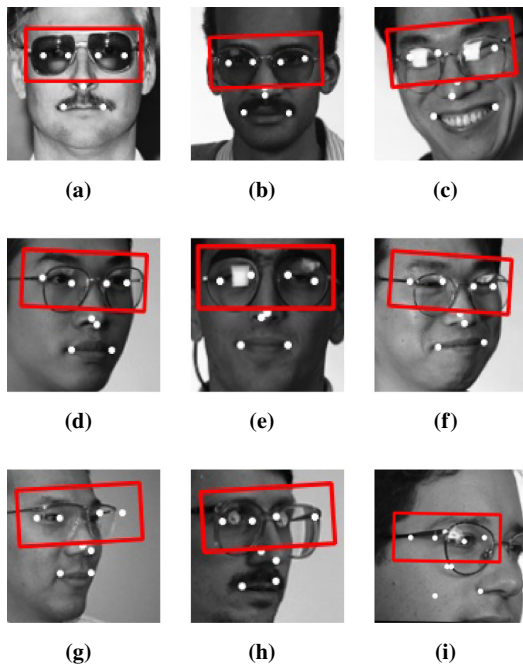


Fig. 15 The normalized eye glasses region algorithm can operate robustly under different image conditions from FERET database. The failure case is included (Fig. 15i)

A large number of LBP variations are designed to expand the scope of application, which offers better performance as well as improvement in the robustness of one or more aspects of the original LBP [40]. In this study, Robust Local Binary Pattern is applied and compared with Local Binary Pattern. RLBP outperforms LBP for almost all values of α in the range of (1–30). Small values of α did not increase the performance compared with original Local Binary Pattern. However, for bigger values, the performance is improved considerably. For instance, varying α from 10 to 30, the average performance can increase by 0.42%.

Finally, it should also be noted that the relationship between size of the normalized eye glasses region and the execution time for both RLBP and LBP follows a linear distribution. $R^2 = \{0.9993, 0.9981\}$ values are obtained for RLBP and LBP, respectively, see Fig. 16. LBP is in average, 2.31% times faster than RLBP. All steps and processes involved in the eye glasses region normalization operation (face detection, landmark detection, face rotation, eye glasses region cropping and resize to desired output) take an average of 30 ms. All experiments in this study are carried out on a 2500-MHz Intel running Windows.

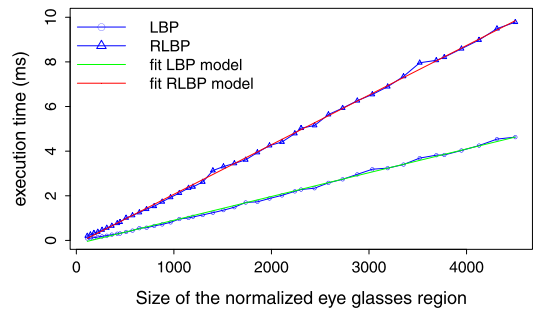


Fig. 16 Relationship between size of the normalized eye glasses region and the execution time for both RLBP and LBP follows a liner distribution

6 Conclusions

Automatic glasses detection on real face images has many potential uses in security systems, intelligent advertising and marketing. In this study, glasses detection on real face images has been investigated. A normalized eye glasses region is proposed, and a comparison between the original LBP and RLBP is carried out varying α parameter. Influences of the resolution, the alignment of the normalized images and the number of divisions in the LBP operator are also carried out. Experimental results demonstrate that a simple yet efficient algorithm can obtain impressive classification accuracy achieving 98.65% recognition rate on LFW database. This algorithm has been tested on FERET database too, achieving 99.89% recognition rate. Experimental results also show that the proposed algorithm is robust under a wide range of lighting conditions, different poses and can deal with occlusion, that are very common with sunglasses, for example. Even when main features from the face are not correctly extracted, the algorithm can estimate glasses region quite robust.

Normalized eye glasses region described in this study can also be used to extract relevant information to deal with the problem of eye closeness detection. Eyes closeness detection from face images has wide applications like facial expression recognition, driver fatigue detection and so on. Closed eyes in the Wild (CEW) database has recently been released [41]. Future work includes the development of an architecture able to face with both problems simultaneously using LFW and CEW databases and new models and descriptors in order to get better image features to achieve better recognition performance without sacrificing simplicity. For example, Completed Hybrid Local Binary Pattern (CHLBP) has recently been proposed [42] to avoid tuning α parameter. A new comprehensive database to benchmark glasses detection and recognition will be proposed. To improve generalization, different databases will be combined and some checks will be performed [43] in order to build a realistic

and, as far as possible, not biased database. Different types of glasses (sport glasses, sunglasses, safety glasses, reading glasses, etc) will be included in this database in order to test both detection and recognition.

Acknowledgments Authors are grateful to anonymous reviewers for constructive feedback and insightful suggestions that greatly improved this article.

References

- Jiang, X., Binkert, M., Bernard, A., Bunke, H.: Towards detection of glasses in facial images. *Pattern Anal. Appl.* **3**(1), 9–18 (2000)
- Dantcheva, A., Velardo, C., Dangelo, A., Dugelay, J.-L.: Bag of soft biometrics for person identification. *Multimed. Tool. Appl.* **51**(2), 739–777 (2011)
- Jing, Z., Mariani, R.: Glasses detection and extraction by deformable contour. In: *Proceedings 15th International Conference on Pattern Recognition*, 2000, vol. 2, pp. 933–936. IEEE (2000)
- Vaquero, D.A., Feris, R.S., Tran, D., Brown, L., Hampapur, A., Turk, M.: Attribute-based people search in surveillance environments. In: *Workshop on Applications of Computer Vision (WACV)*, 2009, pp. 1–8. IEEE (2009)
- Wu, B., Ai, H., Liu, R.: Glasses detection by boosting simple wavelet features. In: *Proceedings of the 17th International Conference on Pattern Recognition*, 2004. *ICPR 2004*, vol. 1, pp. 292–295. IEEE (2004)
- Wu, H., Yoshikawa, G., Shioyama, T., Lao, T., Kawade, T.: Glasses frame detection with 3d hough transform. In: *Proceedings 16th International Conference on Pattern Recognition*, 2002, vol. 2, pp. 346–349. IEEE (2002)
- Shan, S., Gao, W., Lu, Y., Cao, B., Chen, X., Zhao, D., Zeng, W.: Unified framework for classifying facial images based on facial attribute-specific subspaces and minimum reconstruction error. In: *Proc. of ACCV*, vol. 2. Citeseer (2002)
- ChenYu, W., Liu, C., Shum, H.-Y., Xy, Y.-Q., Zhang, Z.: Automatic eyeglasses removal from face images. *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(3), 322–336 (2004)
- Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001, *CVPR 2001*, vol. 1, pp. 1–511. IEEE (2001)
- Heo, J., Kong, S.G., Abidi, B.R., Abidi, M.A.: Fusion of visual and thermal signatures with eyeglass removal for robust face recognition. In: *Conference on Computer Vision and Pattern Recognition Workshop*, 2004, *CVPRW'04*, pp. 122–122. IEEE (2004)
- Jing, Z., Mariani, R., Wang, J.: Glasses detection for face recognition using bayes rules. In: *Advances in Multimodal Interfaces/ICMI 2000*, pp. 127–134. Springer, Berlin (2000)
- Schapire, R.E., Singer, Y.: Improved boosting algorithms using confidence-rated predictions. *Mach. Learn.* **37**(3), 297–336 (1999)
- Xiao, Y., Yan, H.: Extraction of glasses in human face images. In: *Biometric Authentication*, pp. 214–220. Springer, Berlin (2004)
- Park, J.-S., Oh, Y.H., Ahn, S.C., Lee, S.-W.: Glasses removal from facial image using recursive error compensation. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(5), 805–811 (2005)
- Uricár, M., Franc, V., Hlavác, V.: Detector of facial landmarks learned by the structured output svm. In: *VISAPP12: Proceedings of the 7th International Conference on Computer Vision Theory and Applications*, vol. 1, pp. 547–556 (2012)
- Wolf, L., Hassner, T., Taigman, Y.: Similarity scores based on background samples. In: *Computer Vision-ACCV 2009*, pp. 88–97. Springer, Berlin (2010)
- Huang, G.B., Jain, V., Learned-Miller, E.: Unsupervised joint alignment of complex images. In: *IEEE 11th International Conference on Computer Vision*, 2007, *ICCV 2007*, pp. 1–8. IEEE (2007)
- Huang, G.B., Mattar, M., Lee, H., Learned-Miller, E.G.: Learning to align from scratch. In: *Neural Information Processing Systems. Neural Information Processing Systems (NIPS)*, Lake Tahoe (2012)
- Mäkinen, E., Raisamo, R.: Evaluation of gender classification methods with automatically detected and aligned faces. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(3), 541–547 (2008)
- Li, G., Cai, X., Li, X., Liu, Y.: An efficient face normalization algorithm based on eyes detection. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006, pp. 3843–3848. IEEE (2006)
- Mäkinen, E., Raisamo, R.: An experimental comparison of gender classification methods. *Pattern Recogn. Lett.* **29**(10), 1544–1556 (2008)
- Wu, H., Yokoyama, T., Pramadihanto, D., Yachida, M.: Face and facial feature extraction from color image. In: *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition*, 1996, pp. 345–350. IEEE (1996)
- Wu, H., Inada, J., Shioyama, T., Chen, Q., Simada, T.: Automatic facial feature points detection with susan operator. In: *Proceedings of the Scandinavian Conference on Image Analysis*, pp. 257–263 (2001)
- Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The extended cohn-kanade dataset (ck+): a complete dataset for action unit and emotion-specified expression. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2010, pp. 94–101. IEEE (2010)
- Chang, C.-C., Lin, C.-J.: LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**, 27:1–27:27 (2011)
- Song, F., Tan, X., Chen, S., Zhou, Z.-H.: A literature survey on robust and efficient eye localization in real-life scenarios. *Pattern Recogn.* **46**(12), 3157–3173 (2013)
- Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: a database for studying face recognition in unconstrained environments. In: *Technical Report 07–49*. University of Massachusetts, Amherst (2007)
- Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. *Pattern Recogn.* **29**(1), 51–59 (1996)
- Ojala, T., Pietikäinen, M., Maenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(7), 971–987 (2002)
- Guo, Z., Zhang, L., Zhang, D.: A completed modeling of local binary pattern operator for texture classification. *IEEE Trans. Image Proc.* **19**(6), 1657–1663 (2010)
- Zhao, Y., Jia, W., Hu, R.-X., Min, H.: Completed robust local binary pattern for texture classification. *Neurocomputing* **106**, 68–76 (2013)
- Ahonen, T., Hadid, A., Pietikäinen, M.: Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(12), 2037–2041 (2006)
- Shan, C.: Learning local binary patterns for gender classification on real-world face images. *Pattern Recogn. Lett.* **33**(4), 431–437 (2012)
- Kumar, N., Berg, A.C., Belhumeur, P.N., Nayar, S.K.: Attribute and simile classifiers for face verification. In: *IEEE 12th International Conference on Computer Vision*, 2009, pp. 365–372. IEEE (2009)
- Pietikäinen, M., Hadid, A., Zhao, G., Ahonen, T.: Local binary patterns for still images. In: *Computer Vision Using Local Binary Patterns*, pp. 13–47. Springer, Berlin (2011)

36. Huang, D., Shan, C., Ardabilian, M., Wang, Y., Chen, L.: Facial image analysis based on local binary patterns: a survey. *IEEE Trans. Syst. Man Cyber.-Part C* **41**(6), 765–781 (2011)
37. Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with local binary patterns. In: *Computer Vision-eccv 2004*, pp. 469–481. Springer, Berlin (2004)
38. Zou, W.W., Yuen, P.C., Chellappa, R.: Low-resolution face tracker robust to illumination variations. *IEEE Trans. Image Proc.* **22**(5), 1726–1739 (2013)
39. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The feret evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(10), 1090–1104 (2000)
40. Huang, D., Shan, C., Ardabilian, M., Wang, Y., Chen, L.: Local binary patterns and its application to facial image analysis: a survey. *IEEE Trans. Syst. Man Cyber. Part C Appl. Rev.* **41**(6), 765–781 (2011)
41. Song, F., Tan, X., Liu, X., Chen, S.: Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients. *Pattern Recogn.* **47**(9), 2825–2838 (2014)
42. Yuan, J.-H., Huang, D.-S., Zhu, H.-D., Gan, Y.: Completed hybrid local binary pattern for texture classification. In: *International Joint Conference on Neural Networks (IJCNN)*, 2014, pp. 2050–2057. IEEE (2014)
43. Torralba, A., Efros, A.A.: Unbiased look at dataset bias. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 1521–1528. IEEE (2011)



Alberto Fernández received his M.S. degree in Computer Science from University of Oviedo in 2009. Since 2007 he has been developing his professional career at the CTIC Foundation (Centre for the Development of Information and Communication Technologies in Asturias). He has been working in R&D projects. His research interests include artificial vision systems for both industrial and medical processes.



Rodrigo García received the MS degree in computer science from the University of Oviedo, Spain, in 2008. He worked for the R&D Department of CTIC Foundation until October 2014, when he founded his own company, UbiqWare. He also collaborates with the computer engineering area at the University of Oviedo where he is pursuing his PhD in the topic of QoS in distributed systems.



Rubén Usamentiaga Associate Professor in the Department of Computer Science and Engineering at the University of Oviedo. He received his M.S. and Ph.D. degrees in Computer Science from the University of Oviedo in 1999 and 2005, respectively. In recent years he has been working on several projects related to computer vision and industrial systems. His research interests include real-time imaging systems and thermographic applications for industrial processes.



Rubén Casado received a B.Sc. degree in Computer Science in 2005, a M.Sc. in Computing in 2008 and a PhD in Software Systems in 2013 from University of Oviedo, Spain. He has worked as a researcher and teaching assistant at the University of Oviedo, where he is currently a member of the Software Engineering Research Group. Currently he is the leader of the Big Data research program at Treelogic.

4.1.3. Unobtrusive Health Monitoring System Using Video-Based Physiological Information and Activity Measurements

- Fernández, A., Carús, J. L., Usamentiaga, R., Alvarez, E., & Casado, R. (2015, July). Unobtrusive health monitoring system using video-based physiological information and activity measurements. In *Computer, Information and Telecommunication Systems (CITS), 2015 International Conference on* (pp. 1-5). IEEE.
- DOI: 10.1109/CITS.2015.7297728
- URL: <http://ieeexplore.ieee.org/abstract/document/7297728/>
- Ratio de aceptación: 40 % [160]
- Premio al mejor artículo del congreso



IEEE



Universidad de Oviedo



2015 IEEE International Conference on Computer, Information and
Telecommunication Systems, CITS 2015

CITS 2015 Best Paper Award

Paper entitled:

Unobtrusive Health Monitoring System Using Video-Based Physiological
Information and Activity Measurements

By:

Alberto Fernandez, CTIC, Spain;

Juan Luis Carús, CTIC, Spain;

Ruben Usamentiaga, University of Oviedo, Spain;

Eduardo Alvarez, University of Oviedo, Spain;

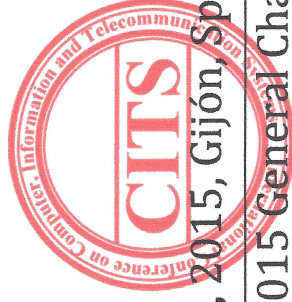
Ruben Casado, University of Oviedo, Spain

M. S. Obaidat

Daniel Garcia

July 15-17, 2015, Gijón, Spain

Mohammad S. Obaidat and Daniel F. Garcia, CITS 2015 General Chairs



Unobtrusive Health Monitoring System Using Video-Based Physiological Information and Activity Measurements

Alberto Fernández*, Juan Luis Carús†
Fundación CTIC (Technological Center)
Technological Scientific Park of Gijón
33204 Gijón, Asturias, Spain
Email: *alberto.fernandez@fundacionctic.org
†juanluis.carus@fundacionctic.org

Rubén Usamentiaga‡, Eduardo Alvarez§, Rubén Casado||
University of Oviedo, Campus de Viesques
33204 Gijón, Asturias, Spain
Email: ‡rusamentiaga@uniovi.es
§edualvarez@uniovi.es
||rcasado@uniovi.es

Abstract—The progressive aging population and the prevalence of chronic diseases are some of the major challenges of current society. This paper aims to illustrate an innovative approach to pervasive health monitoring based on state of the art technology. The proposed solution is based on two main agents: a non-contact health monitoring system embedded in a mirror-like device and a non-invasive and automatic activity monitoring system embedded in a watch-like device. Heart Rate (HR), Respiratory Rate (RR), Heart Rate Variability (HRV) are extracted from the user's face by the mirror-like device. Physical Activity (PA) is extracted from the user's activity by the wearable watch-like device. The health monitoring system proposed in this paper works in an unobtrusive way tracking the user's health status without interaction.

I. INTRODUCTION

The progressive aging population and the prevalence of chronic diseases have emerged as two of the main challenges of the global health care systems. In recent years, the number of people over 65 has increased considerably and it is expected to triple between 2008 and 2060. Nowadays, the chronic diseases are the leading cause of death and disability in the world. These two factors have caused an increase in costs both in intervention and health care.

In response to these challenges, researchers have been actively seeking for innovative and technological-based solutions to improve the quality of health care as well as to reduce the cost of care through early detection and more effective disease management.

Physiological and activity monitoring are a good and common strategy to detect, control and prevent health diseases through the use of sensors. However, people's reluctance to use and interact with sensors avoids wide application. Video technology can be used in this context to recognize and monitor physiological data in a unobtrusive and transparent way for the user. Thus, wearable sensors integrated in a personal item can measure users' activity automatically and reduce invasiveness.

This paper illustrates an innovative approach to pervasive health monitoring based on the state of the art technology. The proposed solution is based on a non-contact health monitoring system embedded in a mirror-

like device and an activity sensor embedded in a watch-like device. Heart Rate (HR), Respiratory Rate (RR) and Heart Rate Variability (HRV) are extracted from the users' face by the mirror and the physical activity (PA) by the analysis of the data collected by the watch. The methods used to collect and analyzed both data are presented with a experimental stage involving real users.

Therefore, an unobtrusive health monitoring system is proposed and validated in this work through the non-invasive collection of medical information (physiological data) and contextual information (activity data). Both data can be checked and analyzed by caregivers using a web page. Alerts and recommendations will be automatically generated according to the gathered information. The main barriers of state-of-the art solutions are broken in this work using a wearable watch-like activity sensor and video-based technologies in a mirror-like device.

II. BACKGROUND

A. Physiological parameter monitoring

In 2008, Verkruysse et al. found that in an ambient light condition, the PPG-signal has different relative strength in three colour channels of an RGB camera that senses human skin [1]. Based on this finding, Poh et al. [2] proposed a linear combination of RGB channels defining three independent signals with Independent Component Analysis (ICA) using non-Gaussianity as the criterion to separate independent resource signals. In [3], they extended their prior work in order to detect multiple physiological parameters. They extract the Blood Volume Pulse (BVP) for computation of Heart Rate (HR), Respiratory Rate (RR), as well as HR Variability (HRV). They improved their method by adding several temporal filters both before and after applying ICA. Lewandowska et al. [4] varied this concept defining three independent linear combinations of the color channels with Principal Component Analysis (PCA).

In 2012, MIT developed a method called "Eulerian video magnification" to amplify the subtle colour changes through band-pass filtering the temporal pyramidal image differences [5]. However, any motion-induced colour distortions within the same frequency band as that of the pulse are unfortunately amplified. More recently, de

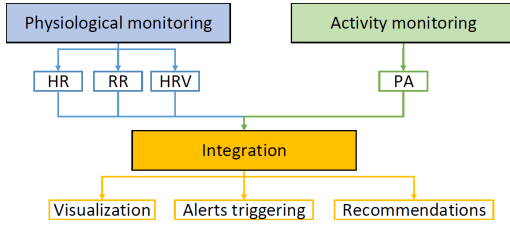


Fig. 1: Architecture of the system

Haan et al. [6] has introduced the chrominance-based rPPG method (CHROM) to consider the pulse as a linear combination of three-colour channels under a standardised skin-tone assumption.

Another approach is to detect the periodic head motions caused by the blood pulsing from heart to head via the abdominal aorta and carotid arteries [7].

It is worth mentioning that on most of these previous methods ([5], [7], [2], [3]) no subjects' motions were involved since no task was assigned and subjects were asked to keep still during video recording. The performance of these algorithms degrades when subjects' motions are involved [8].

B. Activity monitoring

Two main alternatives can be used for activity monitoring: ambient sensors and wearable sensors.

Ambient sensors are embedded in users' environments creating smart homes. They are usually integrated in common objects or in common locations providing pervasive sensing in free-living environments. This information can be used to predict and to prevent health risks by health care platforms or alarm-triggering systems [9]. However, ambient sensors tend to be intrusive in some scenarios and cannot be used in an easy way in outdoor environments.

Wearable sensors refer to sensors that can be worn by users. Although they are more accurate than ambient sensors to measure the human body, some works report their invasiveness and the reluctance of some users to wear them [10]. The solution is to integrate sensors in clothes or clothing accessories (personal items) to achieve an accurate wearable measurement method without invading their privacy. Accelerometers are the most common wearable sensor used for activity monitoring through physical activity. They are small, lightweight and portable, and motion can be recorded to provide an indication of the frequency, duration, and intensity of activities [11]. Their drawback is the lack of standards for conversion of the raw output of an accelerometer to activity magnitudes [12]. Several works use accelerometers for activity monitoring [13],[14]. They differ on the number and position of the accelerometers used; wrist, waist and chest are the most common locations for sensors on the human body.

III. PROPOSED SYSTEM AND ARCHITECTURE

The main components of the proposed solution can be seen in Fig. 1. This solution is composed mainly by three subsystems: physiological monitoring subsystem, activity monitoring subsystem and integration subsystem.

The physiological monitoring subsystem collects physiological data (HR, RR and HRV) using computer vision techniques through a mirror-like device. The activity monitoring subsystem is based on a wearable watch-like measuring the physical activity of the user through a novel magnitude called JIM. All the information collected (physiological and activity) are sent to the integration subsystem.

The integration subsystem works in a remote device processing all the information collected by the physiological and the activity monitoring subsystems. Based on the data collected, alerts and recommendations will be given to the users and their caregivers. All the data collected can also be visualized by the caregivers through a web page.

A. Physiological monitoring subsystem

The physiological monitoring subsystem is mainly divided in three steps: face detection, ROI detection and tracking and quantification of physiological parameters on forehead ROI.

1) *Face detection*: A common first step of these algorithms is to detect the face to operate properly. The process of determining the location of the patient's face is challenging for video recordings. The method proposed by Viola and Jones [15] is usually used to track the face over time ([16],[2],[3],[7],[6],[8]). This detector is most effective only on frontal faces. It can hardly cope with 45° face rotation both around the vertical and horizontal axis. Moreover, multiple detections of the same face can be get due to overlapping sub-windows. Therefore, a robust algorithm is needed to detect the face rectangle on the first frame and track it over time.

In order to solve the aforementioned difficulties, three different face detectors are combined to detect the face on the first frame. Each face detector is based on the Viola and Jones algorithm [15]. Therefore, three face detectors are trained: a frontal detector that handles yaw angles in a range of -40° to +40°, a left half-profile detector that handles yaw angles in a range of 30° to 60°, and a right half-profile detector that handles yaw angles in a range of -30° to -60°. Each of these also handles pitch angles in a range of -30° to +30°. If two or more detectors detect faces, a bounding box containing all detected rectangles is returned. This bounding box is treated in the following step of the algorithm.

2) *ROI detection and Tracking*: A common second step of these algorithms is to detect a ROI inside the face region. Some methods are manual and non-automatic [4]. Most methods rely on the rectangle output by Viola and Jones algorithm. For example, in ([2],[3]), they selected the center 60% width and full height of the rectangle output by Viola and Jones as the ROI. In [7], they opt to use 50% of the rectangle widthwise and 90% heightwise from top in order to ensure that the entire rectangle is within the facial region. In [16], they opt to use 60% width and the full height of the box. These methods are not very accurate. Viola and Jones algorithm only finds coarse face locations as rectangles, which is not precise enough since non-face pixels on the corners of rectangles are always included. Therefore, further processing of the

image is needed. In [6], a skin selection process is applied, which produces a skin-map to be used later on. However, obtaining robust color representations against varied illumination conditions is a major problem. In [17], the Active Appearance Model (AAM) technique is used to detect facial landmark locations. However, the AAM based detectors are not very robust, particularly in low resolution images [18].

In this research work, in order to track the face over time and detect the ROI robustly, (1) a face normalization algorithm and (2) a tracking algorithm are proposed.

The face normalization algorithm is necessary in order to deal with pose, rotation, scale and inaccuracies of the located face. This algorithm is based on [19] but modified in order to get the forehead ROI. The main steps are summarized as follows. Once the face is detected on the first frame, a robust detector of facial landmarks based on the Deformable Parts Models is applied [18]. The output of the classifier is estimated locations of a set of facial landmarks. In order to get the angle of misalignment, a regression line is calculated based on four points: the canthi of the left and right eye, i.e., left inner eye, left outer eye, right inner eye and right outer eye. After that, faces are rotated and aligned in the images so that eyes are located in the same coordinates for all the images. The desired area above the eyes is calculated in the source image. Once the coordinates of the corners of the desired region in the rotated image are calculated, the resulting area is cropped from the rotated image.

A bounding box containing the estimated locations of the set of facial landmarks given by the detector is provided to a robust tracker [20]. This tracker will follow the face over time. The output of this tracker will be used for the next frames. The normalized region above the eyes is used as ROI in order to measure the users' physiological parameters.

3) *Quantification of Physiological Parameters on Forehead ROI*: Two main factors influenced the forehead selection as the ROI: (1) the first one is to exclude the eye region since blinking may interfere with calculations; and (2) the mouth region is also excluded since talking or facial expressions may also interfere with calculations. For the quantification of physiological parameters, a similar algorithm proposed by [3] is used. ROI pixels are separated into the three RGB channels and spatially averaged over all pixels in the ROI to yield a red, blue and green measurement point for each frame and form the raw traces. Afterwards, these raw traces are smoothed and normalized. The normalized raw traces are then decomposed into three independent source signals using Independent Component Analysis (ICA). Then, a temporal filter to exclude frequencies outside the range of interest is applied. Temporal filters have been demonstrated to be helpful for HR measurement [8], [3]. For the HR measurements, inter-beat intervals (IBIs) are calculated. Analysis of HRV was performed by Power Spectral Density (PSD) estimation using the Lomb periodogram. The RR can be estimated from the HRV power spectrum.

B. Activity Monitoring Subsystem

Physical activity (PA) is measured using a novel and efficient magnitude. This subsystem is based on [21], but this method has been modified in order to only monitor physical activity.

Motion wearable sensors (mainly accelerometers) are proposed in this work to measure PA. They are the oldest and most used technique in physical activity measurement [22]. Accelerometers are light, non-invasive, easy to use, adequate for general free-living assessments, and widely integrated in common devices (e. g. mobile phones, camera and watches). There are many accelerometer features [12] that can be used to measure and analyze physical activity. The most widely used and accepted is the signal magnitude area (SMA) [13], which requires an acceleration sample rate of 50 Hz and several filtering stages.

In this work a jerk-based inactivity magnitude (JIM) [23] is used to measure physical activity in a more computationally efficient way than SMA. JIM is calculated each minute ($N = 60$) with an acceleration sample rate of 1Hz. Eq. 1 shows the JIM definition for N samples. It can be interpreted for each acceleration sample as the rate of change of the acceleration (i.e. the jerk). The JIM calculation is implemented in a watch-like wearable device, which fulfills the features required in [24] for a wearable activity monitor. The use of JIM to measure PA increases the battery life (one of the critical factors of wearable sensors [13]) by about 5 times.

$$JIM = \frac{1}{N} \sum_{i=1}^N |Jerk_x[i]| + |Jerk_y[i]| + |Jerk_z[i]| \quad (1)$$

C. Integration Subsystem

The information from the Physiological Monitoring Subsystem and the Activity Monitoring Subsystem is remotely processed in the Integration Subsystem. This subsystem receives the HR, RR and HRV from the mirror-like device and the PA from the watch-like device. The data collected will be processed with three main objectives: visualization, alerts triggering and recommendations.

1) *Visualization*: Through a web page, the data collected can be checked by caregivers. Selecting a time span, the PA is shown for each minute. The measured HR, RR and HRV are also shown in the same graph indicating the time of acquisition. Therefore, the physiological measured information (HR, RR and HRV) could be analyzed according to the contextual information obtained from the PA.

2) *Alerts Triggering*: The measured HR, RR and HRV are remotely processed to generate alerts related to critical conditions. Medical experts fixed the following thresholds to trigger alerts:

- HR must be below 65% of HR_{max} being:
 - 1) $HR_{max} = (220 - \text{age})$ bpm, for males
 - 2) $HR_{max} = (210 - \text{age})$ bpm, for females.
- RR must be in the range of 12-20 breaths per minute.
- HRV is evaluated by the relation between the LF (lower frequency) and HF (high frequency) power of the Lomb periodogram. The relation must be in the range of 1.5-2.

3) *Recommendations*: Using the PA, the activity profile of the user is identified. Users are divided into: sedentary, low, medium and high activity profile depending on the amount of PA measured throughout the day. Users with sedentary and low activity profile will be recommended to change their lifestyle. Physiological measures will be used to give recommendations depending on the health status. Users with a high HR or a critical RR will receive fitness recommendations of lower intensity.

IV. EXPERIMENTAL RESULTS

The experimental results shown in this section are based on the expected performance according to the initial validation of the two main subsystems of the solution: physical monitoring subsystem and activity monitoring subsystem. Further experiments should be done to do a validation of the whole solution.

A. Physical monitoring subsystem results

1) *Experimental results for the Forehead ROI validation*: First of all, the selected forehead ROI is presented on real images from the Labeled Faces in the Wild database¹. LFW database has been created to study the problem of unconstrained face recognition, which contains 13,233 face images of 5,749 distinct subjects collected from the web. This database has also recently been used as benchmark for another attribute recognition algorithms. The only constraint on these faces is that they were detected by the Viola-Jones face detector. The selected region corresponding to the forehead ROI for the quantification of the physiological parameters can be seen in Fig. 2. Further experiments and metrics will be provided in order to give quantitative results. In any case, experimental results on LWF database, showed that the proposed ROI normalization algorithm is quite robust.

2) *Experimental setup, measurement procedure and results*: In this paper, only the HR is tested. Future work will include testing the other physiological parameters. The experimental setup (see Fig. 3) is very similar to other works in order to allow comparisons ([2], [4], [16]). A basic webcam² is located on the top of a mirror and is connected to a laptop (see Fig. 3). All videos are recorded in color (24-bit RGB with 3 channels x 8 bits/channel) by means of a webcam at 20 fps with pixel resolution of 640x480. Twelve participants (6 male, 6 female) ranging from 26 to 35 years old (mean = 29, SD = 2.30) were involved in the experiment. At the same time, a finger pulse oximeter sensor³ was used to measure the participants' pulse rate. The experiments were conducted indoors and with a varying amount of sunlight as the only source of illumination. During the video recording, participants were asked to stand still, breathe spontaneously and stare at the webcam.

Overall, the proposed algorithm showed very good results compared with the finger pulse oximeter (mean error = 1.43, SD = 1.20), very similar to other works

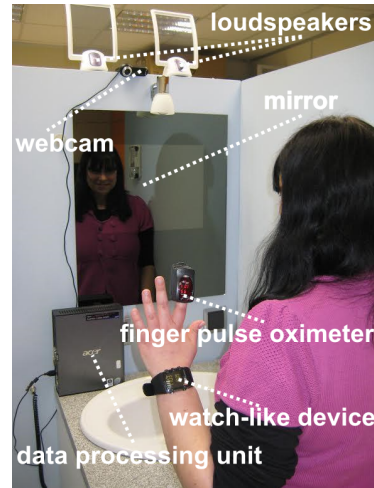


Fig. 3: Experimental setup

([2],[4],[16]) but it is expected that with the use of the proposed robust face tracking algorithm results will improve. In [2] for example, it can be seen that even in the lack of face movement, the localized ROI fluctuates due to tracking irregularities. This benefit is expected to be seen in presence of moving testing videos (see conclusions and future work), where people's motions are involved.

B. Activity monitoring subsystem results

The Activity Monitoring Subsystem is based on the JIM magnitude to measure PA. The proposed subsystem was implemented in a commercial watch-like device⁴ with accelerometer integrated (see Fig. 3). The device was configured to calculate the PA using the JIM and the SMA magnitudes. The experimental setup was focused on the JIM validation as PA estimator.

The same twelve users presented in Section IV.A (6 males, 6 females) ranging from 26 to 35 years old (mean=29, SD=2.30), with different activity profiles and a mean BMI (body mass index) of 22.71 were involved in the experiment. Users were required to wear the watch-like device 24 hours, follow their usual routine and take off the device at bath and sleep times.

The validation of JIM as a PA estimator was done analyzing the relation between JIM and SMA using the Spearman correlation index. The correlation study proves the overall strength of association between the two magnitudes (mean = 0.95, SD = 0.03, p-value < 0.05).

The computational performance features of SMA and JIM were also analyzed during the experiment. The main advantage of the JIM calculation against SMA is the lower accelerometer sample frequency. This characteristic decreases the current consumption of the accelerometer and increases battery life by about 5 times in the used watch-like device. Therefore, JIM can be considered as good activity estimator as SMA with better computational features.

¹Labeled Faces in the Wild <http://vis-www.cs.umass.edu/lfw/>

²Logitech Webcam Pro 9000. USB connection and usb cable (6 feet). <http://support.logitech.com/product/quickcam-pro-9000>

³Nonin Onyx II Model 9560 Finger Pulse Oximeter with wireless technology <http://www.nonin.com/Onyx9560>

⁴TI eZ430 Chronos <http://www.ti.com/tool/eZ430-chronos>

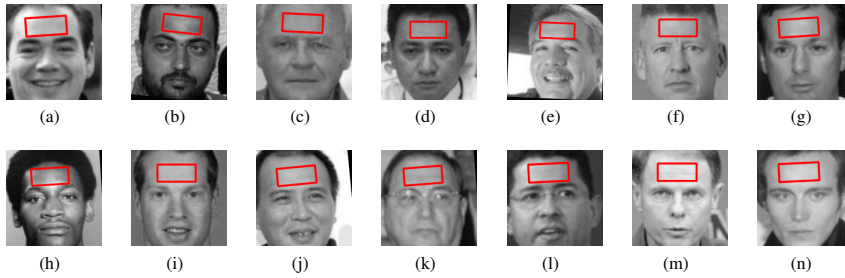


Fig. 2: Normalized forehead region for some images from LFW database

V. CONCLUSION AND FUTURE WORK

In this paper a health monitoring system using video technology and a wearable activity device is proposed. The system measures Heart Rate (HR), Respiration Rate (RR) and Heart Rate Variability (HRV) based on facial analysis through video technology and Physical Activity (PA) based on users' activity through a wearable device.

The proposed solution combines physiological data with activity data, giving contextual information (activity data) to the medical information (physiological data). All data can be visualized using a web page and are used to give recommendations and trigger alerts based on the combination of the medical and contextual information.

The section of the experimental results only includes results of the HR parameter. Therefore, this work should be extended to include more detailed results of the three physiological parameters. Despite the fact that the system is on a preliminary stage, interesting results were found, so the system is promising.

Future work will include: (1) the testing of the whole solution and (2) the testing of the system using elderly people and (3) extended results.

REFERENCES

- [1] W. Verkruijsse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Optics express*, vol. 16, no. 26, pp. 21 434–21 445, 2008.
- [2] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Optics express*, vol. 18, no. 10, pp. 10 762–10 774, 2010.
- [3] —, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 1, pp. 7–11, 2011.
- [4] M. Lewandowska, J. Ruminski, T. Koceljko, and J. Nowak, "Measuring pulse rate with a webcam non-contact method for evaluating cardiac activity," in *Computer Science and Information Systems (FedCSIS), 2011 Federated Conference on*. IEEE, 2011, pp. 405–410.
- [5] H.-Y. Wu, M. Rubinstein, E. Shih, J. V. Guttag, F. Durand, and W. T. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Trans. Graph.*, vol. 31, no. 4, p. 65, 2012.
- [6] G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rppg," *Biomedical Engineering, IEEE Transactions on*, vol. 60, no. 10, pp. 2878–2886, 2013.
- [7] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 3430–3437.
- [8] X. Li, J. Chen, G. Zhao, and M. Pietikainen, "Remote heart rate measurement from face videos under realistic situations," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 4264–4271.
- [9] M. Chan, E. Campo, D. Estève, and J.-Y. Fourniols, "Smart homes - current features and future perspectives," *Maturitas*, vol. 64, no. 2, pp. 90–7, Oct. 2009.
- [10] A. Dias, L. Gorzelniak, R. A. Jörres, R. Fischer, G. Hartvigsen, and A. Horsch, "Assessing physical activity in the daily life of cystic fibrosis patients," *Pervasive and Mobile Computing*, vol. 8, no. 6, pp. 837–844, Dec. 2012.
- [11] N. Butte, U. Ekelund, and K. Westerterp, "Assessing physical activity using wearable monitors: measures of physical activity," *Med Sci Sports Exerc*, vol. 44, no. 1 Suppl 1, pp. S5–12, Jan. 2012.
- [12] D. S. Ward, K. R. Evenson, A. Vaughn, A. B. Rodgers, and R. P. Troiano, "Accelerometer Use in Physical Activity: Best Practices and Research Recommendations," *Medicine & Science in Sports & Exercise*, vol. 37, no. Supplement, pp. S582–S588, Nov. 2005.
- [13] D. Curone, G. M. Bertolotti, A. Cristiani, E. L. Secco, and G. Magenes, "A real-time and self-calibrating algorithm based on triaxial accelerometer signals for the detection of human posture and activity," *IEEE transactions on information technology in biomedicine*, vol. 14, no. 4, pp. 1098–105, Jul. 2010.
- [14] F. Albinali, M. S. Goodwin, and S. Intille, "Detecting stereotypical motor movements in the classroom using accelerometry and pattern recognition algorithms," *Pervasive and Mobile Computing*, vol. 8, no. 1, pp. 103–114, Feb. 2012.
- [15] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1. IEEE, 2001, pp. I–511.
- [16] H. Xia, Z. Bao, and H. Jiang, "Mobile cardiac pulse measurements," *Electrical Engineering*.
- [17] H. E. Tasli, A. Gudi, and M. den Uyl, "Remote ppg based vital sign measurement using adaptive facial regions," in *International Conference on Image Processing*, 2014.
- [18] M. Uříčář, V. Franc, and V. Hlaváč, "Detector of facial landmarks learned by the structured output svm," *VISAPP*, vol. 12, pp. 547–556, 2012.
- [19] A. Fernández, R. García, R. Usamentiaga, and R. Casado, "Glasses detection on real images based on robust alignment," *Machine Vision and Applications*, 2015. [Online]. Available: <http://dx.doi.org/10.1007/s00138-015-0674-1>
- [20] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," in *Proceedings of the British Machine Vision Conference BMVC*, 2014.
- [21] J. L. C. Candás, V. Peláez, G. López, M. Á. Fernández, E. Álvarez, and G. Díaz, "An automatic data mining method to detect abnormal human behaviour using physical activity measurements," *Pervasive and Mobile Computing*, vol. 15, pp. 228–241, 2014.
- [22] C. V. Bouten, K. T. Koekkoek, M. Verduin, R. Kodde, and J. D. Janssen, "A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity," *IEEE transactions on bio-medical engineering*, vol. 44, no. 3, pp. 136–47, Mar. 1997.
- [23] J. L. Carús, V. Peláez, G. López, and V. Lobato, "JIM: a novel and efficient accelerometric magnitude to measure physical activity," *Studies in health technology and informatics*, vol. 177, pp. 283–8, Jan. 2012.
- [24] K. Y. Chen and D. R. Bassett, "The Technology of Accelerometry-Based Activity Monitors: Current and Future," *Medicine & Science in Sports & Exercise*, vol. 37, no. Supplement, pp. S490–S500, Nov. 2005.

4.1.4. Wearable and ambient sensors to health monitoring using computer vision and signal processing techniques

- Fernández, A., Carús, J. L., Usamentiaga, R., Alvarez, E., & Casado, R. Wearable and ambient sensors to health monitoring using computer vision and signal processing techniques. *Journal of Networks*. 2017.
- Aceptado, pero sin asignar DOI
- La revista *Journal of Networks* (ISSN 1796-2056) ha sido aceptada para su inclusión en Emerging Source Citation Index (ESCI), ya que cumple los criterios básicos de indexación: a) tener revisión por pares, b) cumplimiento de las prácticas de publicación éticas, c) contenidos de interés elevado para la comunidad científica, d) contenidos que cumplan con los requisitos técnicos de Thomson Reuters. Además, está indexado en las siguientes bases de datos: Academic Journals Database; Academy Search; BASE; BibSonomy; Cabell - Computer Science/Business Information Systems; CAID - Computer Abstracts International Database; CNKI; CrossRef; CSA; DBLP; Directory of Open Access Journals (DOAJ) - Computer Science; DOI; EBS-CO; EI Compendex (On Hold); Emerging Sources Citation Index (ESCI); Engineering Village 2; EZB; GALE; Genamics JournalSeek; GetCITED; Gold Rush; Google Scholar; INSPEC; io-port.net; iThenticate; J4F; JAL; JournalTOCs; NewJour; OAI-PMH Registered Data Providers; OAJSE; OAKList; OCLC; Open J-Gate - Engineering & Technology (JET); Ovid LinkSolver; PASCAL; PKP Open Archives Harvester; ProQuest; PubZone; QCAT; ResearchBib; SCImago; Scirus; SCOPUS; SHERPA/RoMEO; Socolar; The Index of Information Systems Journals; Trove; True Serials; Udini; UIUC OAI-PMH Data Provider Registry; ULRICH's Periodicals Directory; WorldCat; WorldWideScience; ZDB.



ACADEMY PUBLISHER

Journal of Networks

Dear Mr. Alberto Fernández Villán:

Your paper titled “Wearable and ambient sensors to health monitoring using computer vision and signal processing techniques” has been accepted for publication in the Special Issue of Selected Best Papers of CITS 2015, that will be published on Journal of Networks (JNW, ISSN 1796-2056).

On behalf of the editors, we thank you for your paper. Your contribution is greatly appreciated.

Sincerely,



UNIVERSIDAD DE SEVILLA
DEPARTAMENTO DE ARQUITECTURA
Y TECNOLOGÍA DE COMPUTADORES

Prof. Daniel Cascado Caballero
Journal of Networks, Invited editor.

Wearable and ambient sensors to health monitoring using computer vision and signal processing techniques

Alberto Fernández^a, Juan Luis Carús^a, Rubén Usamentiaga^b, Eduardo Alvarez^b, Rubén Casado^b

^a TSK, Technological Scientific Park of Gijón, 33203, Gijón, Asturias, Spain

Email: {alberto.fernandez,juanluis.carus}@grupotsk.com

^b University of Oviedo, Campus de Viesques, 33204, Gijón, Asturias, Spain

Email: {rusamentiaga,edualvarez,rcasado}@uniovi.es

Abstract—The progressive aging population and the prevalence of chronic diseases are some of the major challenges of current society. To address these unfulfilled healthcare needs, especially for the early prediction and treatment of major diseases, health informatics has emerged as an active area of interdisciplinary research. It is possible to improve the quality of life maintaining the independence of the vulnerable elderly people using novel technological solutions. This paper aims to illustrate an innovative approach to pervasive health monitoring based on state of the art technology. This approach could be used by caregivers to act over the users' behaviour improving it and helping them in their daily routine. The proposed solution is based on two main agents: a non-contact health monitoring system embedded in a mirror-like device and a non-invasive and automatic activity monitoring system embedded in a watch-like device. Heart Rate (HR), Respiratory Rate (RR) and Heart Rate Variability (HRV) are extracted from the user's face by the mirror-like device. Physical Activity (PA) is extracted from the user's activity by the wearable watch-like device. Experimental results involving 12 participants (6 male, 6 female) aged between 26 and 35 have shown that the system demonstrates expectations. The proposed solution combines both physiological and activity data, giving contextual information (activity data) to the medical information (physiological data). The health monitoring system proposed in this paper works in an unobtrusive way tracking the user's health status without interaction. Future work will include the testing of the system using elderly people.

Index Terms—pervasive health monitoring, healthcare, physiological and activity data, computer vision, health monitoring system

I. INTRODUCTION

The progressive aging population and the prevalence of chronic diseases have emerged as two of the main challenges of the global health care systems. In recent years, the number of people over 65 has increased considerably and it is expected to triple between 2008 and 2060 [1]. Nowadays, the chronic diseases are the leading cause of death and disability in the world. These two factors have caused an increase in costs both in intervention and health care.

In response to these challenges, researchers have been actively seeking for innovative and technological-based solutions to improve the quality of health care, as well

as to reduce the cost of care through early detection and more effective disease management.

One of the main aims of health informatics is to provide personal safety to the elderly people. People with long-term health conditions often experience a reduced quality of life and dependence upon health and social care services [2]. It is expected that the future healthcare system should be preventive, predictive, preemptive, personalized, pervasive, participatory, patient-centered, and precise, i.e., p-health system [3]. Physiological and activity monitoring are a good and common strategy to detect, control and prevent health diseases through the use of sensors. However, people's reluctance to use and interact with sensors avoids wide application. Video technology can be used in this context to recognize and monitor physiological data in a unobtrusive and transparent way for the user. Thus, wearable sensors integrated in a personal item can measure users' activity automatically and reduce invasiveness.

This paper illustrates an innovative approach to pervasive health monitoring based on the state of the art technology shown in Fig. 1. The proposed solution is based on a non-contact health monitoring system embedded in a mirror-like device and an activity sensor embedded in a watch-like device. Heart Rate (HR), Respiratory Rate (RR) and Heart Rate Variability (HRV) are extracted from the users' face by the mirror and their physical activity (PA) by the analysis of the data collected by the watch. The methods used to collect and analyzed both data are presented with an experimental stage involving real users.

It should be noted the present work is an extension of our previous one presented here [4]. Therefore, with the purpose of clarifying the differences between these two versions, the improvements of this work are commented as follows. Our previous work proposed an unobtrusive health monitoring system using video-based physiological information and activity measurements. More specifically, it proposed to extract HR, RR and HRV from the users' face by the mirror and the PA by the analysis of the data collected by the watch. However, only the HR was tested and a finger pulse oximeter sensor with wireless technology was used to measure the participants' pulse

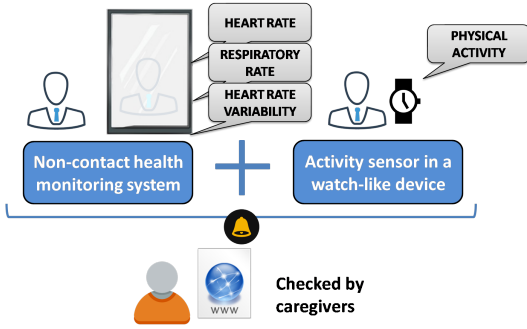


Figure 1: General overview of the proposed approach

rate. Moreover, due to space constraints, some of the sections of the paper needed further explanations. In the present work, these weaknesses are resolved, a more detailed solution is presented and the other physiological parameters are extracted and tested with the required sensors to be used as benchmarks.

Therefore, an unobtrusive health monitoring system is proposed and validated in this work through the non-invasive collection of medical information (physiological data) and contextual information (activity data). Both data can be checked and analyzed by caregivers using a web page. Alerts and recommendations will be automatically generated according to the gathered information. The main barriers of state-of-the art solutions are broken in this work using a wearable watch-like activity sensor and video-based technologies in a mirror-like device.

The rest of the paper is organized as follows. Section II introduces main research work about both physiological parameters and activity monitoring approaches. Section III describes the stages of the proposed system. The experimental results to validate the proposed approach are detailed in Section IV. Finally, in Section V, conclusions and future work are given.

II. BACKGROUND

The main objective of unobtrusive sensing is to enable continuous monitoring of physical activities and behaviors, as well as physiological and biochemical parameters during the subject's daily routine. Unobtrusive sensing can be implemented in two ways: 1) sensors are worn by the subject, e.g., in the form of shoes, eyeglasses, earring, clothing, gloves and watch, or 2) sensors are embedded into the ambient environment or as smart objects interacting with the subjects (see Zheng et al. [3] for references). As in this research work these two main methods are combined, a review of both methods is carried out in this section.

A. Physiological parameter monitoring

In 2008, Verkruijsse et al. found that in an ambient light condition, the PPG-signal (photoplethysmography: light-based technology to sense the rate of blood flow

as controlled by the hearts pumping action) has different relative strength in three colour channels of an RGB camera that senses human skin [5]. Based on this finding, Poh et al. [6] proposed a linear combination of RGB channels defining three independent signals with Independent Component Analysis (ICA) using non-Gaussianity as the criterion to separate independent resource signals. In [7], they extended their prior work in order to detect multiple physiological parameters. They extract the Blood Volume Pulse (BVP) for computation of Heart Rate (HR), Respiratory Rate (RR), as well as HR Variability (HRV). They improved their method by adding several temporal filters both before and after applying ICA. Lewandowska et al. [8] varied this concept defining three independent linear combinations of the color channels with Principal Component Analysis (PCA). In [4], a health monitoring system using video-based physiological information and activity measurements is proposed. Regarding video-based physiological information, only the HR is tested.

In 2012, MIT developed a method called "Eulerian video magnification" to amplify the subtle colour changes through band-pass filtering the temporal pyramidal image differences [9]. However, any motion-induced colour distortions within the same frequency band as that of the pulse are unfortunately amplified. More recently, de Haan et al. [10] has introduced the chrominance-based rPPG method (CHROM) to consider the pulse as a linear combination of three-colour channels under a standardised skin-tone assumption.

Another approach is to detect the periodic head motions caused by the blood pulsing from heart to head via the abdominal aorta and carotid arteries [11]. A review of the main methods for the measurements of both non-contact heart rate and heart rate variability can be seen here [12]. It is worth mentioning that in most of these previous methods ([9], [11], [6], [7]) no subjects' motions were involved since no task was assigned and subjects were asked to keep still during video recording. The performance of these algorithms degrades when subjects' motions are involved [13].

B. Activity monitoring

Two main alternatives can be used for activity monitoring: ambient sensors and wearable sensors.

Ambient sensors are embedded in users' environments creating smart homes. They are usually integrated in common objects or in common locations providing pervasive sensing in free-living environments. This information can be used to predict and to prevent health risks by health care platforms or alarm-triggering systems [14]. However, ambient sensors tend to be intrusive in some scenarios and cannot be used in an easy way in outdoor environments.

Wearable sensors refer to sensors that can be worn by users. Although they are more accurate than ambient sensors to measure the human body, some works report their invasiveness and the reluctance of some users to wear them [15]. The solution is to integrate sensors in clothes or clothing accessories (personal items) to achieve

an accurate wearable measurement method without invading their privacy. Accelerometers are the most common wearable sensor used for activity monitoring through physical activity. They are small, lightweight and portable, and motion can be recorded to provide an indication of the frequency, duration, and intensity of activities [16]. Their drawback is the lack of standards for conversion of the raw output of an accelerometer to activity magnitudes [17]. Several works use accelerometers for activity monitoring [18], [19]. They differ on the number and position of the accelerometers used; wrist, waist and chest are the most common locations for sensors on the human body.

III. PROPOSED SYSTEM AND ARCHITECTURE

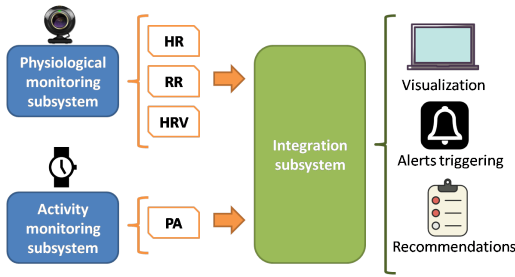


Figure 2: Architecture of the system

The main components of the proposed solution can be seen in Fig. 2. This solution is composed mainly by three subsystems: physiological monitoring subsystem, activity monitoring subsystem and integration subsystem.

The physiological monitoring subsystem collects physiological data (HR, RR and HRV) using computer vision techniques through a mirror-like device. The activity monitoring subsystem is based on a wearable watch-like measuring the physical activity of the user through a novel magnitude called JIM (jerk-based inactivity magnitude). All the information collected (physiological and activity) are sent to the integration subsystem.

The integration subsystem works in a remote device processing all the information collected by the physiological and the activity monitoring subsystems. Based on the data collected, alerts and recommendations will be given to the users and their caregivers. All the data collected can also be visualized by the caregivers through a web page.

A. Physiological monitoring subsystem

The physiological monitoring subsystem is mainly divided in three steps: face detection, region of interest (ROI) detection and tracking and quantification of physiological parameters on forehead ROI, which can be seen in Figure 3. The novelty of this approach is based on automatic face tracking and localization of ROI measurement.

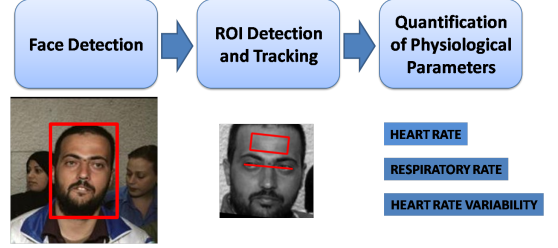


Figure 3: The physiological monitoring subsystem: a) face detection, b) ROI detection and tracking and c) quantification of physiological parameters

1) *Face detection*: A common first step of these algorithms is to detect the face to operate properly. The process of determining the location of the patient's face is challenging for video recordings. The method proposed by Viola and Jones [20] is usually used to track the face over time ([21], [6], [7], [11], [10], [13]). This detector is most effective only on frontal faces. It can hardly cope with 45° face rotation both around the vertical and horizontal axis. Moreover, multiple detections of the same face can be taken due to overlapping sub-windows. Therefore, a robust algorithm is needed to detect the face rectangle on the first frame and track it over time.

In order to solve the aforementioned difficulties, three different face detectors are combined based on the idea proposed by Asthana et al. [22] to detect the face on the first frame. Each face detector is based on the Viola and Jones algorithm [20]. Therefore, three face detectors are trained: a frontal detector that handles yaw angles in a range of -40° to +40°, a left half-profile detector that handles yaw angles in a range of 30° to 60°, and a right half-profile detector that handles yaw angles in a range of -30° to -60°. Each of these also handles pitch angles in a range of -30° to +30°. If two or more detectors detect faces, a bounding box containing all detected rectangles is returned. This bounding box is treated in the following step of the algorithm.

2) *ROI detection and Tracking*: A common second step of these algorithms is to detect a ROI inside the face region. Some methods are manual and non-automatic [8]. Most methods rely on the rectangle output by Viola and Jones algorithm. For example, in ([6], [7]), they selected the center 60% width and full height of the rectangle output by Viola and Jones as the ROI. In [11], they opt to use 50% of the rectangle widthwise and 90% heightwise from top in order to ensure that the entire rectangle is within the facial region. They also removed eyes from the region so that blinking artifacts do not affect calculations. To achieve this, they removed the subrectangle spanning 20% to 55% heightwise. For videos where the face is not visible, they marked the region manually.

In [21], they opt to use 60% width and the full height of the box. These methods are not very accurate. Viola and Jones algorithm only finds coarse face locations as rectangles, which is not precise enough since non-face pixels on

the corners of rectangles are always included. Therefore, further processing of the image is needed. In [10], a skin selection process is applied, which produces a skin-map to be used later on. However, obtaining robust color representations against varied illumination conditions is a major problem [23]. In [24], the Active Appearance Model (AAM) technique [25] is used to detect facial landmark locations. However, the AAM based detectors are not very robust, particularly in low resolution images [26].

In this research work, in order to track the face over time and detect the ROI robustly, (1) a face normalization algorithm and (2) a tracking algorithm are proposed.

The face normalization algorithm is necessary in order to deal with pose, rotation, scale and inaccuracies of the located face. This algorithm is based on [27] but modified in order to get the forehead ROI. We have previously used this face normalization algorithm achieving state of the art results. For example, in [27], [28] we used this algorithm to automatically detect glasses on real face images. In [29], we also performed this face normalization algorithm in order to build a system for face recognition adapted to the visually-impaired people.

The main steps of this algorithm are summarized as follows. Once the face is detected on the first frame, a robust detector of facial landmarks based on the Deformable Parts Models is applied [26]. The output of the classifier are the locations (x,y coordinates) of a set of facial landmarks. In order to get the angle of misalignment, a regression line is calculated based on four points: the canthi of the left and right eye, i.e., left inner eye, left outer eye, right inner eye and right outer eye. After that, faces are rotated and aligned in the images so that eyes are located in the same coordinates for all the images. The desired area above the eyes is calculated in the source image. Once the coordinates of the corners of the desired region in the rotated image are calculated, the resulting area is cropped from the rotated image.

A bounding box containing the estimated locations of the set of facial landmarks given by the detector is provided to a robust tracker [30]. This tracker will follow the face over time. The output of this tracker will be used for the next frames. The normalized region above the eyes is used as ROI in order to measure the users' physiological parameters.

3) *Quantification of Physiological Parameters on Forehead ROI*: Two main factors influenced the forehead selection as the ROI: (1) the first one is to exclude the eye region since blinking may interfere with calculations; and (2) the mouth region is also excluded since talking or facial expressions may also interfere with calculations.

For the quantification of physiological parameters, a similar algorithm proposed by [7] is used, but further calculations of the physiological parameters are extracted from the forehead ROI. For the sake of completeness, these steps are explained as follows.

ROI pixels are separated into the three RGB channels and spatially averaged over all pixels in the ROI to yield

a red, blue and green measurement point for each frame and form the raw traces. Afterwards, these raw traces are smoothed and normalized. The normalized raw traces are then decomposed into three independent source signals using Independent Component Analysis (ICA).

The component whose power spectrum contained the highest peak is selected [6], [7].

Temporal filtering to exclude frequencies outside the range of interest is applied. Temporal filters have been demonstrated to be helpful for HR measurement [13], [7].

Specifically, a moving-average filter, which removes random noise using temporal average of adjacent frames, is firstly applied. Secondly, a Hamming window-based finite impulse response bandpass filter (cutoff frequency of [0.7, 4] Hz) is applied. For the HR measurements, interbeat intervals (IBIs) are calculated.

Analysis of HRV was performed by Power Spectral Density (PSD) estimation using the Lomb periodogram. Frequency domain methods assign bands of frequency and then, count the number of normal-to-normal (NN) intervals that match each band. The bands are typically high frequency (HF) from 0.15 to 0.4 Hz and low frequency (LF) from 0.04 to 0.15 Hz. HRV signals contain well-defined rhythms including physiological information [31]:

- HF generally represents parasympathetic activity and is therefore generally considered to be a marker of vagal activity [32].
- LF is influenced by both sympathetic and parasympathetic activity [33].
- The ratio of HF to LF (HF/LF) represents the balance between parasympathetic and sympathetic activity [34].

Therefore, LF and HF powers were measured as the area under the PSD curve corresponding to 0.04-0.15 and 0.15-0.4 Hz, respectively, and quantified in normalized units (n.u.) to minimize the effect on the values of the changes in total power [7].

Finally, since the HF component is connected with breathing, the RR can be estimated from the HRV power spectrum. It should be noted that when the frequency of respiration changes, the center frequency of the HF peak shifts in accordance with RR [35]. RR is calculated by locating the center frequency of the HF peak in the HRV PSD.

B. Activity Monitoring Subsystem

Main steps of the Activity Monitoring Subsystem are shown in Fig. 4. Physical activity (PA) is measured using a novel and efficient magnitude called JIM and based on the derivative of the acceleration. This subsystem is based on [36], but it has been modified in order to only monitor physical activity.

Motion wearable sensors (mainly accelerometers) are proposed in this work to measure PA. They are the oldest and most used technique in physical activity measurement [37]. Accelerometers are light, non-invasive, easy to use, adequate for general free-living assessments, and widely integrated in common devices (e. g. mobile phones,

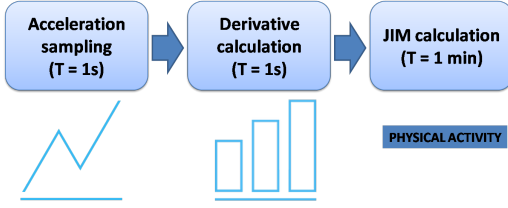


Figure 4: The Activity Monitoring Subsystem: a) acceleration sampling, b) derivative calculation and c) JIM calculation

camera and watches). There are many accelerometer features [17] that can be used to measure and analyze physical activity. The most widely used and accepted is the signal magnitude area (SMA) [18], which requires an acceleration sample rate of 50 Hz and several filtering stages.

In this work a jerk-based inactivity magnitude (JIM) [38] is used to measure physical activity in a more computationally efficient way than SMA. JIM is calculated each minute ($N = 60$) with an acceleration sample rate of 1Hz. Eq. 1 shows the JIM definition for N samples. It can be interpreted for each acceleration sample as the rate of change of the acceleration (i.e. the jerk). The JIM calculation is implemented in a watch-like wearable device, which fulfills the features required in [39] for a wearable activity monitor. The use of JIM to measure PA increases the battery life (one of the critical factors of wearable sensors [18]) by about 5 times.

$$JIM = \frac{1}{N} \sum_{i=1}^N |Jerk_x[i]| + |Jerk_y[i]| + |Jerk_z[i]| \quad (1)$$

C. Integration Subsystem

The information from the Physiological Monitoring Subsystem and the Activity Monitoring Subsystem is remotely processed in the Integration Subsystem. This subsystem receives the HR, RR and HRV from the mirror-like device and the PA from the watch-like device. The data collected will be processed with three main objectives: visualization, alerts triggering and recommendations.

1) *Visualization*: Through a web page, the data collected can be checked by caregivers. Selecting a time span, the PA is shown for each minute. The measured HR, RR and HRV are also shown in the same graph indicating the time of acquisition. Therefore, the physiological measured information (HR, RR and HRV) could be analyzed according to the contextual information obtained from the PA.

2) *Alerts Triggering*: The measured HR, RR and HRV are remotely processed to generate alerts related to critical conditions. Medical experts fixed the following thresholds to trigger alerts:

- HR must be below 65% of HR_{max} being:
 - 1) $HR_{max} = (220 - \text{age})$ bpm, for males

2) $HR_{max} = (210 - \text{age})$ bpm, for females.

- RR must be in the range of 12-20 breaths per minute.
- HRV is evaluated by the relation between the LF (lower frequency) and HF (high frequency) power of the Lomb periodogram. The relation must be in the range of 1.5-2.

3) *Recommendations*: Using the PA, the activity profile of the user is identified. Users are divided into: sedentary, low, medium and high activity profile depending on the amount of PA measured throughout the day. Users with sedentary and low activity profile will be recommended to change their lifestyle. Physiological measures will be used to give recommendations depending on the health status. Users with a high HR or a critical RR will receive fitness recommendations of lower intensity.

IV. EXPERIMENTAL RESULTS

The experimental results shown in this section are based on the expected performance according to the initial validation of the two main subsystems of the solution: physical monitoring subsystem and activity monitoring subsystem. Further experiments should be done to do a validation of the whole solution.

A. Physical monitoring subsystem results

1) *Experimental results for the Forehead ROI validation*: First of all, the selected forehead ROI is presented on real images from the Labeled Faces in the Wild database [40] LFW database has been created to study the problem of unconstrained face recognition, which contains 13,233 face images of 5,749 distinct subjects collected from the web. This database has also recently been used as a benchmark for another attribute recognition algorithms. The only constraint on these faces is that they were detected by the Viola-Jones face detector. The selected region corresponding to the forehead ROI for the quantification of the physiological parameters can be seen in Fig. 5. Further experiments and metrics will be provided in order to give quantitative results. In any case, experimental results on LFW database, showed that the proposed ROI normalization algorithm is quite robust.

2) *Experimental setup, measurement procedure and results*: The experimental setup (see Fig. 6) is very similar to other works in order to allow comparisons ([6], [8], [21]). A basic webcam is located on the top of a mirror and is connected to a laptop. All videos are recorded in color (24-bit RGB with 3 channels x 8 bits/channel) by means of a webcam at 20 fps with pixel resolution of 640x480. Twelve participants (6 male, 6 female) ranging from 26 to 35 years old (mean = 29, SD = 2.30) were involved in the experiment. At the same time, a finger BVP sensor and a chest belt respiration sensor where used to record their BVP and spontaneous breathing.

The experiments were conducted indoors and with a varying amount of sunlight as the only source of illumination. During the video recording, participants were asked to stand still, breathe spontaneously and stare at

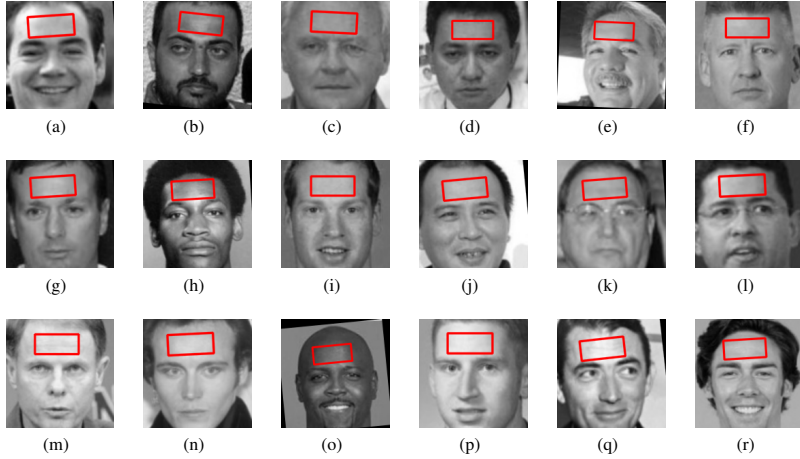


Figure 5: Normalized forehead region for some images from LFW database



Figure 6: Experimental setup

the webcam. All analyses and results are performed on one-minute recording from all the participants.

Concerning HR analysis, the proposed algorithm has shown very good results compared with the finger BVP (mean error = 1.43, SD = 1.20), which are very similar to other works ([6], [8], [21]). Results derived from the HRV calculations are commented as follows. LF n.u. (mean error = 9.26, SD = 9.56), HF n.u. (mean error = 9.16, SD = 9.62), LF/HF (mean error = 1.11 SD = 1.47)

are obtained compared with the finger BVP. Results from the RR measurements (mean error = 0.84, SD = 1.83) are compared with a chest belt respiration sensor. These physiological measurements were obviously correlated across all parameters with $r = 0.98$ for HR, $r = 0.91$ for HF and LF, $r = 0.86$ for LF/HF, and $r = 0.93$ for RR.

It is expected that with the use of the proposed robust face tracking algorithm results will improve. In [6] for example, it can be seen that even in the lack of face movement, the localized ROI fluctuates due to tracking irregularities. This benefit is expected to be seen in presence of moving testing videos (see conclusions and future work), where people's motions are involved.

All steps and processes involved in the physiological monitoring subsystem (face detection, landmark detection, face rotation, forehead region detection and tracking, and the quantification of the physiological parameters) take an average of 50 ms. All experiments in this study are carried out on a 3600-MHz Intel running Windows. In other words, the system is capable of running at 20 fps. This frame rate is the same one that was introduced before by the webcam, which is capable of recording frames at 20fps with pixel resolution of 640x480.

B. Activity monitoring subsystem results

The Activity Monitoring Subsystem is based on the JIM magnitude to measure PA. The proposed subsystem was implemented in a commercial watch-like device with accelerometer integrated (see Fig. 6). The device was configured to calculate the PA using the JIM and the SMA magnitudes. The experimental setup was focused on the JIM validation as PA estimator.

The same twelve users presented in Section IV.A (6 males, 6 females) ranging from 26 to 35 years old (mean=29, SD=2.30), with different activity profiles and a mean BMI (body mass index) of 22.71 were involved in the experiment. Users were required to wear the watch-

TABLE I.: Summary of experimental results

Parameter	Mean error	SD	R index
HR	1.43	1.20	0.98
HRV LF	9.26	9.56	0.91
HRV HF	9.16	9.62	0.91
HRV LF/HF	1.11	1.47	0.86
RR	0.84	1.83	0.93
PA	-	-	0.95

like device 24 hours, follow their usual routine and take off the device at bath and sleep times.

The validation of JIM as a PA estimator was done analyzing the relation between JIM and SMA using the Spearman correlation index. The correlation study proves the overall strength of association between the two magnitudes (mean = 0.95, SD = 0.03, p-value < 0.05).

The computational performance features of SMA and JIM were also analyzed during the experiment. The main advantage of the JIM calculation against SMA is the lower accelerometer sample frequency. This characteristic decreases the current consumption of the accelerometer and increases battery life by about 5 times in the used watch-like device. Therefore, JIM can be considered as good activity estimator as SMA with better computational features.

Table I shows the main results of this section describing the mean and SD of the error produced in the estimation by the proposed system and the R index.

V. CONCLUSION AND FUTURE WORK

In this paper a health monitoring system using video technology and a wearable activity device is proposed. The system measures Heart Rate (HR), Respiration Rate (RR) and Heart Rate Variability (HRV) based on facial analysis through video technology and Physical Activity (PA) based on users' activity through a wearable device.

The proposed solution combines physiological data with activity data, giving contextual information (activity data) to the medical information (physiological data). All data can be visualized using a web page and are used to give recommendations and trigger alerts based on the combination of medical and contextual information. For the sake of argument, a user with high heart rate or high respiratory rate related to high activity will not rise an alarm. On the other hand, high values of both heart rate or respiratory rate but low activity will trigger an alarm.

The activity monitoring subsystem proposed in this paper will be able to measure the activity of the user in an unobtrusive way using a wearable watch-like device. The proposed system has a very good performance with a high correlation index, so it can be considered as a reliable method to measure physical activity. The non-contact health monitoring subsystem proposed in this paper will also be able to measure some key physiological parameters extracted from the face by using face tracking and a normalization algorithm to locate the ROI measurement. The integration subsystem proposed in this paper has three main objectives: visualization, alerts triggering

and recommendations. Based on the reliability of the physiological and activity monitoring subsystems, this subsystem can be used by the caregiver to act over the users behavior improving it and helping them in their daily routine.

Experimental results involving 12 participants (6 male, 6 female) aged between 26 and 35 have shown that the system is promising. It should be noted that respiratory rate estimation from HRV worked well in healthy young volunteers, but it would be more unlikely to give accurate results in elderly people, especially those with chronic diseases, most of whom lack autonomic function [41]. This aspect makes it almost impossible to implement these methods accurately in unhealthy elderly patients [42]. Therefore, future work will include the testing of the system in elderly people and further investigation will be performed in this matter.

ACKNOWLEDGMENT

The authors are grateful to the anonymous referees for their valuable comments and suggestions to improve the presentation of this paper.

REFERENCES

- [1] B. Rechel, E. Grundy, J.-M. Robine, J. Cylus, J. P. Mackenbach, C. Knai, and M. McKee, "Ageing in the european union," *The Lancet*, vol. 381, no. 9874, pp. 1312–1322, 2013.
- [2] F. Cardinaux, D. Bhowmik, C. Abhayaratne, and M. S. Hawley, "Video based technology for ambient assisted living: A review of the literature," *Journal of Ambient Intelligence and Smart Environments*, vol. 3, no. 3, pp. 253–269, 2011.
- [3] Y. Zheng, X. Ding, C. Poon, B. Lo, H. Zhang, X. Zhou, G. Yang, N. Zhao, and Y. Zhang, "Unobtrusive sensing and wearable devices for health informatics," 2014.
- [4] A. Fernandez, J. L. Carus, R. Usamentiaga, E. Alvarez, and R. Casado, "Unobtrusive health monitoring system using video-based physiological information and activity measurements," in *Computer, Information and Telecommunication Systems (CITS), 2015 International Conference on*. IEEE, 2015, pp. 1–5.
- [5] W. Verkrusse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Optics express*, vol. 16, no. 26, pp. 21 434–21 445, 2008.
- [6] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Optics express*, vol. 18, no. 10, pp. 10 762–10 774, 2010.
- [7] —, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 1, pp. 7–11, 2011.
- [8] M. Lewandowska, J. Ruminski, T. Kociejko, and J. Nowak, "Measuring pulse rate with a webcamera non-contact method for evaluating cardiac activity," in *Computer Science and Information Systems (FedCSIS), 2011 Federated Conference on*. IEEE, 2011, pp. 405–410.
- [9] H.-Y. Wu, M. Rubinstein, E. Shih, J. V. Guttag, F. Durand, and W. T. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Trans. Graph.*, vol. 31, no. 4, p. 65, 2012.
- [10] G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rppg," *Biomedical Engineering, IEEE Transactions on*, vol. 60, no. 10, pp. 2878–2886, 2013.

- [11] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 3430–3437.
- [12] J. Kranjec, S. Beguš, G. Geršak, and J. Drnovšek, "Non-contact heart rate and heart rate variability measurements: A review," *Biomedical Signal Processing and Control*, vol. 13, pp. 102–112, 2014.
- [13] X. Li, J. Chen, G. Zhao, and M. Pietikainen, "Remote heart rate measurement from face videos under realistic situations," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 4264–4271.
- [14] M. Chan, E. Campo, D. Estève, and J.-Y. Fourmiols, "Smart homes - current features and future perspectives," *Maturitas*, vol. 64, no. 2, pp. 90–7, Oct. 2009.
- [15] A. Dias, L. Gorzelniak, R. A. Jörres, R. Fischer, G. Hartvigsen, and A. Horsch, "Assessing physical activity in the daily life of cystic fibrosis patients," *Pervasive and Mobile Computing*, vol. 8, no. 6, pp. 837–844, Dec. 2012.
- [16] N. Butte, U. Ekelund, and K. Westerterp, "Assessing physical activity using wearable monitors: measures of physical activity," *Med Sci Sports Exerc*, vol. 44, no. 1 Suppl 1, pp. S5–12, Jan. 2012.
- [17] D. S. Ward, K. R. Evenson, A. Vaughn, A. B. Rodgers, and R. P. Troiano, "Accelerometer Use in Physical Activity: Best Practices and Research Recommendations," *Medicine & Science in Sports & Exercise*, vol. 37, no. Supplement, pp. S582–S588, Nov. 2005.
- [18] D. Curone, G. M. Bertolotti, A. Cristiani, E. L. Secco, and G. Magenes, "A real-time and self-calibrating algorithm based on triaxial accelerometer signals for the detection of human posture and activity," *IEEE transactions on information technology in biomedicine*, vol. 14, no. 4, pp. 1098–105, July 2010.
- [19] F. Albinali, M. S. Goodwin, and S. Intille, "Detecting stereotypical motor movements in the classroom using accelerometry and pattern recognition algorithms," *Pervasive and Mobile Computing*, vol. 8, no. 1, pp. 103–114, Feb. 2012.
- [20] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1. IEEE, 2001, pp. 1–511.
- [21] H. Xia, Z. Bao, and H. Jiang, "Mobile cardiac pulse measurements," *Electrical Engineering*.
- [22] A. Asthana, T. K. Marks, M. J. Jones, K. H. Tieu, and M. Rohith, "Fully automatic pose-invariant face recognition via 3d pose normalization," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 937–944.
- [23] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods," *Pattern recognition*, vol. 40, no. 3, pp. 1106–1122, 2007.
- [24] H. E. Tasli, A. Gudi, and M. den Uyl, "Remote ppg based vital sign measurement using adaptive facial regions," in *International Conference on Image Processing*, 2014.
- [25] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 6, pp. 681–685, 2001.
- [26] M. Uříčák, V. Franc, and V. Hlaváč, "Detector of facial landmarks learned by the structured output svm," *VISAPP*, vol. 12, pp. 547–556, 2012.
- [27] A. Fernández, R. García, R. Usamentiaga, and R. Casado, "Glasses detection on real images based on robust alignment," *Machine Vision and Applications*, 2015. [Online]. Available: <http://dx.doi.org/10.1007/s00138-015-0674-1>
- [28] A. Fernandez, R. Casado, and R. Usamentiaga, "A real-time big data architecture for glasses detection using computer vision techniques," in *Future Internet of Things and Cloud (FiCloud), 2015 3rd International Conference on*. IEEE, 2015, pp. 591–596.
- [29] A. Fernndez, J. Cars, R. Usamentiaga, and R. Casado, "Face recognition and spoofing detection system adapted to visually-impaired people," *IEEE Latin America Transactions*, vol. 14, no. 2, pp. 1–9, 2016.
- [30] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," in *Proceedings of the British Machine Vision Conference BMVC*, 2014.
- [31] L. Keselbrener and S. Akselrod, "Selective discrete fourier transform algorithm for time-frequency analysis: method and application on simulated and cardiovascular signals," *Biomedical Engineering, IEEE Transactions on*, vol. 43, no. 8, pp. 789–802, 1996.
- [32] S. Akselrod, D. Gordon, F. A. Ubel, D. C. Shannon, A. Berger, and R. J. Cohen, "Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat-to-beat cardiovascular control," *science*, vol. 213, no. 4504, pp. 220–222, 1981.
- [33] T. F. of the European Society of Cardiology *et al.*, "Heart rate variability standards of measurement, physiological interpretation, and clinical use," *Eur Heart J*, vol. 17, pp. 354–381, 1996.
- [34] G. E. Billman, "Heart rate variability—a historical perspective," 2011.
- [35] T. E. Brown, L. A. Beightol, J. Koh, and D. L. Eckberg, "Important influence of respiration on human rr interval power spectra is largely ignored," *Journal of Applied Physiology*, vol. 75, no. 5, pp. 2310–2317, 1993.
- [36] J. L. C. Candás, V. Peláez, G. López, M. Á. Fernández, E. Álvarez, and G. Díaz, "An automatic data mining method to detect abnormal human behaviour using physical activity measurements," *Pervasive and Mobile Computing*, vol. 15, pp. 228–241, 2014.
- [37] C. V. Bouten, K. T. Koekkoek, M. Verduin, R. Kodde, and J. D. Janssen, "A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity," *IEEE transactions on bio-medical engineering*, vol. 44, no. 3, pp. 136–47, Mar. 1997.
- [38] J. L. Carús, V. Peláez, G. López, and V. Lobato, "JIM: a novel and efficient accelerometric magnitude to measure physical activity," *Studies in health technology and informatics*, vol. 177, pp. 283–8, Jan. 2012.
- [39] K. Y. Chen and D. R. Bassett, "The Technology of Accelerometry-Based Activity Monitors: Current and Future," *Medicine & Science in Sports & Exercise*, vol. 37, no. Supplement, pp. S490–S500, Nov. 2005.
- [40] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Technical Report 07-49, University of Massachusetts, Amherst, Tech. Rep., 2007.
- [41] D. F. Dietrich, C. Schindler, J. Schwartz, J.-C. Barthélémy, J.-M. Tschopp, F. Roche, A. von Eckardstein, O. Brändli, P. Leuenberger, D. R. Gold, *et al.*, "Heart rate variability in an ageing population and its association with lifestyle and cardiovascular risk factors: results of the sapaldia study," *Europace*, vol. 8, no. 7, pp. 521–529, 2006.
- [42] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. Clifton, and C. Pugh, "Non-contact video-based vital sign monitoring using ambient light and auto-regressive models," *Physiological measurement*, vol. 35, no. 5, p. 807, 2014.

Alberto Fernández received his M.S. degree in Computer Science from University of Oviedo in 2009. He developed his professional career at the CTIC Foundation (Centre for the Development of Information and Communication Technologies in Asturias). He has been working in R&D projects. His research

interests include artificial vision systems for both industrial and medical processes. He is now working at TSK on several projects related to computer vision, industrial systems and energy efficiency.

Juan Luis Carús received his PhD in Industrial Engineering in 2015. He was working at the CTIC Foundation in projects focused on advanced electronic technologies for health (e-Health), ICT solutions for wellbeing, sensor networks, wearable computing, ambient intelligence and signal processing techniques. He is now working at TSK as R&D&i project manager involved in national and international projects related to energy efficiency.

Ruben Usamentiaga Fernández Associate Professor in the Department of Computer Science and Engineering at the University of Oviedo. He received his M.S. and Ph.D. degrees in Computer Science from the University of Oviedo in 1999 and 2005, respectively. In recent years he has been working on several projects related to computer vision and industrial systems. His research interests include real-time imaging systems and thermographic applications for industrial processes.

Eduardo Alvarez received the M.S. degree in electrical engineering from the Industrial Engineering School, University of Oviedo, Gijón, Spain. He has worked on several R&D projects for Spanish electrical companies. His research is focused on microgrids and the impact of distributed generation on distribution systems. He was with the R&D Department, Fundación Centro Tecnológico de la Información y Comunicación, Gijón, where he was in charge of R&D project management and was involved with IT research projects at national and European levels. He is also a Professor of electrical technology at the University of Oviedo.

Rubén Casado received a B.Sc. degree in Computer Science in 2005, a M.Sc. in Computing in 2008 and a PhD in Software Systems in 2013 from University of Oviedo, Spain. He has worked as a researcher and teaching assistant at the University of Oviedo, where he is currently a member of the Software Engineering Research Group. From 2012 to 2016 he was the Head of the Big Data unit at Treelogic. He is now working at Accenture as Big Data Manager. His research interests include Big Data, Distributed Data Management, Software Testing and Cloud and Services based systems.

4.1.5. A Real-Time Big Data Architecture for Glasses Detection Using Computer Vision Techniques

- Fernández, A., Casado, R., & Usamentiaga, R. (2015, August). A Real-Time Big Data Architecture for Glasses Detection Using Computer Vision Techniques. In Future Internet of Things and Cloud (FiCloud), 2015 3rd International Conference on (pp. 591-596). IEEE.
- DOI: 10.1109/FiCloud.2015.78
- URL: <http://ieeexplore.ieee.org/abstract/document/7300872/>

A Real-Time Big Data Architecture For Glasses Detection Using Computer Vision Techniques

Alberto Fernández*

Fundación CTIC (Technological Center)

Technological Scientific Park of Gijón

33204 Gijón, Asturias, Spain

Email: *alberto.fernandez@fundacionctic.org

Rubén Casado[†], Rubén Usamentiaga[‡]
University of Oviedo, Campus de Viesques

33204 Gijón, Asturias, Spain

[†]rcasado@uniovi.es

Email: [‡]rusamentiaga@uniovi.es

Abstract—Automatic glasses detection is a hot topic within the large-scale face images classification domain, which has impact on face recognition or soft biometrics for person identification. In many practical video surveillance applications, the faces acquired by cameras are low resolution. Therefore, this type of applications requires processing of a large number of relatively small-sized images. However, continuous stream of image and video data processing is a big data challenge. This need fits with the goals of Big Data streaming processing systems. In this paper, we propose a real-time Big Data architecture in order to collect, maintain and analyze massive volumes of images related with the problem of automatic glasses detection. This architecture can be used as an automatic image tagging related with glasses detection on face images.

I. INTRODUCTION

In recent years, the rapid development of Internet, Internet of Things, and Cloud Computing have led to an impressive growth of data in almost every industry and business area. Big data has rapidly developed into a hot topic, which attracts thorough attention from academia, industry, and governments around the world.

Big Data has become important as many organizations, both public and private, have been collecting massive amounts of domain-specific information, which can contain useful information about problems such as national intelligence, cyber security, fraud detection, marketing, and medical informatics. Companies such as Google and Microsoft are analyzing large volumes of data for business analysis and decisions, influencing existing and future technology [1].

Big Data is a collection of data sets, which is enormously large and complex so that conventional database systems cannot process within desired time. For instance, storing and processing of daily tweets at Twitter demand significant data storage, processing and data analytics capabilities e.g., finding correlations among millions of tweets or analyze the demographics of users. Though conventional SQL-based databases have proved to be highly efficient, reliable and consistent in terms of storing and processing structured (or relational) data, they fall short of processing Big Data, which is characterized by large volume, variety, velocity, openness, inappropriate structure, and visualization among others [2].

Big Data is set to play a major role in various domains such as science, research, engineering, medicine, healthcare, finance, business, and, eventually, society itself [3]. It can be used to analyze and forecast business trends,

profit and loss, identify real-time road traffic conditions, healthcare, weather information and so on.

Big Data is generally characterized by the 3Vs: variety, volume and velocity.

Volume refers to the size of the data to be processed. Volume of Big Data goes far beyond the conventional limits of megabytes or gigabytes and reaches the terabytes or even petabyte. Velocity refers not only to frequency of the data generation but also to the dynamic aspects of the data as well as the need of generating the results in real-time. Variety refers to the multimodal nature of data such as different data schemas of data sources, structured data like ontologies and unstructured data like sensor signals [4], [5].

Furthermore, the processing, availability or acquisition of Big Data can be classified into different categories, including batch processing, real-time processing and hybrid processing. Batch processing is an efficient way of processing high volumes of data, which are collected during a period of time. In this scheme, data are collected and stored into the data sources and processed. The batch results are then produced. However, several applications require real-time processing of data (streams), which is acquired from heterogeneous data sources. Streaming processing involves continuous input, processing and output of data [4]. Low latency is the main goal of this processing paradigm, that is, data must be processed in a small (or near real) time period. Application domains include smart cities, entertainment, and disaster management.

Note that batch processing provides rigorous results since more data can be used and it performs better training of predictive models. But it is not feasible for domains which need low-response time. Real-time processing generally ensures low response time. However, low-response time can be achieved at the expense of less rigorous analysis of data. The hybrid approach is, therefore, required so that application domains (using Big Data) can benefit from both batch and real-time processing. To obtain desired results under this approach, both batch and real-time results are queried. The results are then merged together, synchronized or composed. Data acquisition and analysis become more complicated under the approach [6].

Big Data has grown exponentially in the last decade and video/image generated by sensors and devices has become the largest source [7]. Big Data on image analytics is best applicable on processing photos and surveillance videos. Processing surveillance videos for information

extraction requires real-time stream processing. The video data will be continuous in motion, which can reach massive proportions over time. The video data requires to get processed on time to extract the full benefit of surveillance.

Face processing and Big Data are two facets that had scarcely anything in common, but recent advances in several related domains have led them to converge in impactful ways. These domains include Social Networking along with headway in the nationalization of law enforcement facial Databases.

These unprecedented amounts of video/image data have created great opportunities uncovering new insights since Big Data is a top business priority and it has also posed many fundamental technological challenges. One of the most critical challenges is how to intelligently analyze and understand the visual information in such large-scale video/image data.

In this work we propose a Big Data architecture for streaming processing of large amounts of images. Our objective is to provide a real-time scalable system for automatic glasses detection using video images. Main contributions of this paper are summarized as twofold:

- A scalable low-latency architecture for image analysis using Big Data technologies.
- Parallelization of an innovative glasses detection strategy.

The rest of the paper is organized as follows. Section 2 some approaches related with Big Data processing are resumed. Section 3 is composed by two subsections. Firstly, the algorithm used for glasses detection is described. Afterwards, the Big Data architecture for streaming processing of large amounts of images is proposed. Finally, in Section 4 conclusions are given.

II. BACKGROUND

Big Data has been speeding up its development to trend as the most important topic attracting considerable attention from researchers, academicians, industries, and governments worldwide.

Jin et al. [8] present, briefly, the concept of Big Data combined with its features and challenges. They put forward a few necessary conditions for the success of a Big Data project. Since Big Data consists of a large amount of information, it is necessary to identify the specific requirements regardless of its nature. Also, it is said that the kernel data/structure is to be explored efficiently. The most important point noted by them is the application of a top-down approach to handle Big Data. This allows the isolated solutions to be put together to come to a complete solution. They also support the conclusion of a project by integrating the solution.

Gandomi et al. [9] present the various analytics methods used for Big Data. Since a majority of the available Big Data is unstructured, their focus on this dataset provides a wider understanding on the applications on Big Data. They have stressed the importance of real-time analytics, which are bound to become the major field of research in the future. This is highly due to the growth of social networking and mobile apps. They have identified that predictive analysis has been dominating all fields of

analytics and present the case for new methods to address the differences of Big Data.

The primary application of video analytics in recent years has been in automated security and surveillance systems.

Video analytics, also known as video content analysis (VCA), involves a variety of techniques to monitor, analyze, and extract meaningful information from video streams. Although video analytics is still in its infancy compared to other types of data mining [10], various techniques have already been developed for processing real-time as well as pre-recorded videos. The increasing prevalence of closed-circuit television (CCTV) cameras and the booming popularity of video-sharing websites are the two leading contributors to the growth of computerized video analysis. A key challenge, however, is the sheer size of video data. To put this into perspective, one second of a high-definition video, in terms of size, is equivalent to over 2000 pages of text [11].

Big data technologies turn this challenge into opportunity. Obviating the need for cost-intensive and risk-prone manual processing, big data technologies can be leveraged to automatically sift through and draw intelligence from thousands of hours of video. As a result, the big data technology is the third factor contributing to the development of video analytics [9].

Video analytics can efficiently and effectively perform surveillance functions such as detecting breaches of restricted zones, identifying objects removed or left unattended, detecting loitering in a specific area, recognizing suspicious activities, and detecting camera tampering, to name a few. Upon detection of a threat, the surveillance system may notify security personnel in real time or trigger an automatic action (e.g., sound alarm, lock doors, or turn on lights) [9].

The data generated by CCTV cameras in retail outlets can be extracted for business intelligence. Marketing and operations management are the primary application areas. For instance, smart algorithms can collect demographic information about customers, such as age, gender, and ethnicity [9].

Soyata et al. [12] have presented a mobile-cloudlet-cloud architecture called MOCHA as a platform for our target face recognition application. This architecture is designed to minimize the overall response time of the face detection and face recognition algorithms under the circumstances of having heterogeneous communication latencies and compute powers of cloud servers at diverse geographical placements. This architecture reduces overall processing time when face recognition applications run on mobile devices using the cloud as the backend servers.

Vinay et al. [13] have proposed a face recognition approach operating on Big Data applied for the task of Face Tagging in the context of social networks. They employ Extreme Learning Machines (ELM) to conduct facial extraction. The architecture proposed by Vinay et al. is mainly focused on Face Tagging.

III. PROPOSED SYSTEM AND ARCHITECTURE

In this Section, the algorithm used for glasses detection is described. Afterwards, the Big Data architecture

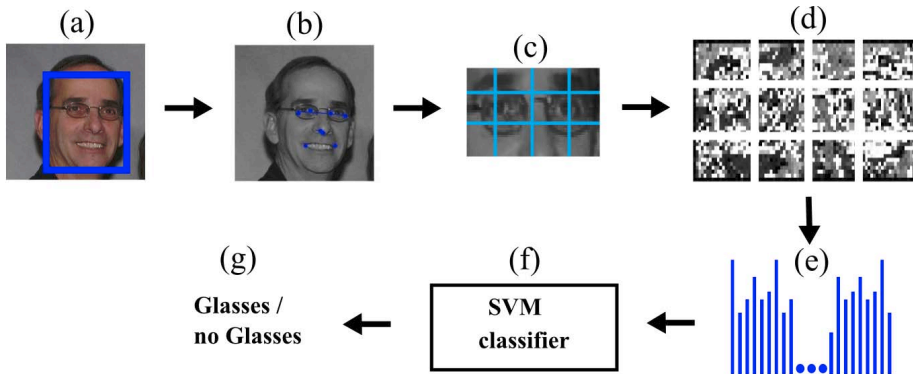


Fig. 1: Standard approach for glasses detection on face images: (a) input image, (b) location of the face, landmark detection and regression line, (c) aligned and normalized eye glasses region, (d) LBP descriptors are built by partitioning the LBP face image into a grid, (e) LBP histograms are built over each grid cell and concatenated into a vector, (f) This vector is used for SVM classification, and (g) final output

for streaming processing of large amounts of images is proposed.

A. Glasses detection on face images

Our original glasses detection algorithm [14] is explained hereafter. It can be seen in Figure 1.

First of all, this work starts with the face location using the algorithm by Viola & Jones [15].

After the face is located in the image, some preprocessing is necessary in order to deal with pose, rotation, scale and inaccuracies of the located face.

A face normalization algorithm is applied to get the region around the eyes (see Figure 2). The face normalization algorithm is explained as follows. Once the face is detected on the first frame, a robust detector of facial landmarks based on the Deformable Parts Models is applied [16]. The output of the classifier is estimated locations of a set of facial landmarks. In order to get the angle of misalignment, a regression line is calculated based on four points: the canthi of the left and right eye, i.e., left inner eye, left outer eye, right inner eye and right outer eye. After that, faces are rotated and aligned in the images so that eyes are located in the same coordinates for all the images. The desired area above the eyes is calculated in the source image. Once the coordinates of the corners of the desired region in the rotated image are calculated, the resulting area is cropped from the rotated image.

Afterwards, Local Binary Pattern (LBP) [17] is applied in order to get the feature sets.

Finally, Support Vector Machine (SVM) is applied on the classification step. SVM is applied to classify the extracted feature histograms over the normalized eyes glasses regions. The output of the SVM classifier is a two-class classification problem i.e. glasses vs no glasses. LIBSVM was used for the training and testing of SVMs [18].

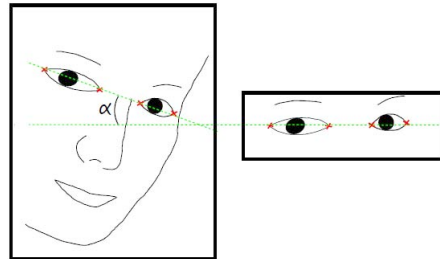


Fig. 2: Face normalization algorithm to get the region around the eyes

1) *Local Binary Pattern*: The Local Binary Pattern operator is a type of feature used for classification. It has been found to be a powerful feature for texture classification. It was introduced in 1996 as a means of summarizing local gray-level structure [17]. The operator takes a local neighborhood around each pixel, thresholds the pixels of the neighborhood at the value of the central pixel and uses the resulting binary-valued image patch as a local image descriptor. It was originally defined for 3×3 neighborhoods, giving 8 bit codes based on the 8 pixels around the central one.

The operator was extended to use neighborhoods of different sizes [19], thus making it feasible to deal with textures at different scales. Another extension to the original operator [19] defined the so-called uniform patterns: an LBP is 'uniform' if it contains at most two bitwise transitions from 0 to 1 or vice versa when viewed as a circular bit string. For example, 00000000, 00011110 and 10000011 are uniform patterns.

2) *Support Vector Machine*: SVMs are supervised learning models with associated learning algorithms, which analyze data and recognize patterns, used for classification and regression analysis. Given a set of

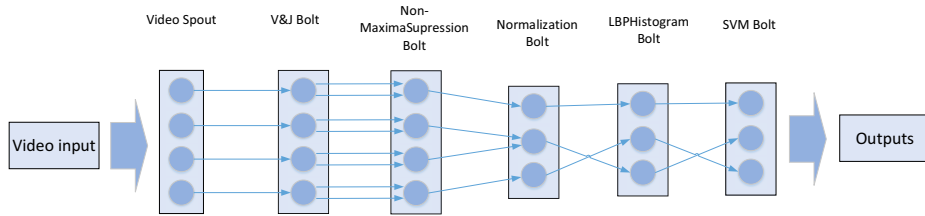


Fig. 3: Big Data architecture for streaming processing of large amounts of images

training examples, each marked for belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier.

An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap, which is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. More formally, a support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane with the largest distance to the nearest training-data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.

B. Big data architecture for glasses detection on face images

The main objective of the proposed architecture is to parallelize the different steps of the glasses detection workflows. In order to do that, we propose a topology implemented with a streaming technology like Apache Storm. Storm is distributed real-time computation system released as open source by Twitter. The architecture is simple. There are two types of elements:

- 1) Spouts reading information from the source and emitting the data as K-V tuples.
- 2) Bolts processing information coming from the spouts or other bolts.

By connecting bolts, Storm defines topologies similar to Jobs in MapReduce. The main difference is that a topology never ends, because the data to process is in streaming. Thus, there is always new information to process. Following the principles of the Apache Storm architecture, Figure 3 depicts the proposed system:

1) *VideoSpout*: These software components split the video streaming into a sequence of images (frames). These frames are the input information for the glasses detection algorithm. This Spout uses a shuffle grouping. So, frames are randomly distributed across the next bolts in the topology in such a way that each bolt is guaranteed to get an equal number of frames.

2) *V&JBolt*: The basic face detection method is Adaboost algorithm with a cascading Haar-like feature classifiers based on the framework proposed by Viola & Jones

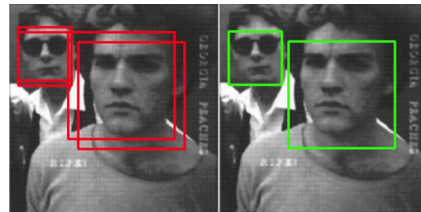


Fig. 4: The classifiers initially detected four bounding boxes, but by applying non-maximum suppression, the correct faces are obtained

[15]. Different face sizes are detected by repeating the classification at different scales. Good results are obtained by trying different scales a factor of 1.25 apart [15].

The proposed algorithm is implemented using a bolt node to process frames in parallel. The computational complexity of the Viola & Jones algorithm is independent of scale [20]. If the complexity were not scale-independent, the distribution would have to balance the computational load by randomly assigning scales to each bolt, mixing face scales evenly to the bolts. Anyway, this bolt also uses shuffle grouping. Frames are randomly distributed across the bolt tasks in such a way that each bolt is guaranteed to get an equal number of frames. Parallelizing, the face detection task helps to reduce overall processing times.

3) *Non-Maximum SupressionBolt*: Since the Viola & Jones algorithm is invariant to minor position and scale changes, there will often be a set of overlapping faces being detected around the vicinity of an actual face. These overlapping results should be processed in order to determine the final valid face locations. This process is usually called Non-Maximum Supression (NMS), which is a key post-processing step in many computer vision applications. In the context of object detection, it is used to transform a smooth response map, which triggers many imprecise object window hypotheses in, ideally, a single bounding-box for each detected object. Therefore, two key points on NMS include:

- A real face may result in multiple nearby detections
- Postprocess detected subwindows to combine overlapping detections into a single detection



Fig. 5: Normalized eye glasses region for some images from LFW database

Many approaches have been recently proposed to prune the set of responses for object detection [21], [22], [23]. However, with the purpose of keeping the original NMS proposed by Viola & Jones algorithm, this procedure is taken into account and it is described as follows (see Figure 4).

The Viola & Jones detector [15] partitions those responses in disjoint sets, grouping together responses as soon as they overlap, and propose, for each group with enough windows, a window whose coordinates are the group average. The NMS stage involves grouping the 'raw' bounding boxes into equivalence classes based on closeness, deleting groups containing fewer than a threshold number of boxes, and removing any groups lying too close within another group.

4) *NormalizationBolt*: In this bolt, the face normalization algorithm explained before is applied to get the region around the eyes (some examples can be seen in Figure 5). This bolt also uses shuffle grouping. Afterwards, LBP is applied on the normalized eye glasses region in order to get the feature sets.

5) *HistogramBolt*: After labelling an image with the LBP operator, a histogram of the labelled image $f_l(x, y)$ can be defined as:

$$H_i = \sum_{x,y} I\{f_l(x, y) = i\}, i = 0, \dots, n - 1 \quad (1)$$

where n is the number of different labels produced by the LBP operator and:

$$I\{A\} = \begin{cases} 1, & \text{if } A = \text{true} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

For efficient face representation, LBP features should retain spatial information. Hence, the facial image is divided into m regions $\{R_0, \dots, R_{m-1}\}$. In this way, the basic histogram defined above can be extended into a *spatially enhanced histogram* [24], which encodes both the appearance and the spatial relations of facial regions. This bolt also uses shuffle grouping.

6) *SVMBolt*: Support Vector Machine (SVM) is applied on the classification step. SVM is applied to classify the extracted feature histograms over the normalized eyes glasses regions. The output of the SVM classifier

is a two-class classification problem i.e. glasses vs no glasses. Therefore, different type of glasses (sport glasses, sunglasses, safety glasses, reading glasses, etc) belong to the same category. LIBSVM was used for the training and testing of SVMs [18].

This bolt uses an offline SVM model created using the Labeled Faces in The Wild (LFW) Database [25]. LFW database is a database originally created to study the problem of unconstrained face recognition, which contains 13,233 face images of 5,749 distinct subjects collected from the web. This database has also recently been used as benchmark for another attribute recognition algorithms [26], [27]. In our previous work [14], 3,000 face images (1,500 wearing glasses, 1,500 not wearing glasses) were used to train a SVM classifier achieving a 98.65% recognition rate. Therefore, this generated model is used in this bolt.

IV. CONCLUSION AND FUTURE WORK

Big data has made a strong impact in almost every sector and industry today. In this paper, a real-time Big Data architecture in order to collect, maintain and analyze massive volumes of images related with the problem of automatic glasses detection is proposed.

As commented before, Viola & Jones face detector typically produces a number of positive responses close by to the correct detection, and this leads to the need to have a further non-maximum suppression (NMS) stage to thin out the multiple responses, and to suppress spurious responses. Different alternatives are proposed [21], [22], [23] and should be checked in order to increase both performance and accuracy of the detections. Therefore, several tests and comparisons will be performed among these alternatives.

Deep Learning algorithms extract high-level, complex abstractions as data representations through a hierarchical learning process. Complex abstractions are learnt at a given level based on relatively simpler abstractions formulated in the preceding level in the hierarchy. A key benefit of Deep Learning is the analysis and learning of massive amounts of unsupervised data, making it a valuable tool for Big Data Analytics, where raw data are largely unlabeled and un-categorized [28]. Therefore, future work will include Deep Learning algorithms in our pipeline detection architecture.

From the architecture point of view, new approaches will be study. New streaming technologies like Apache Spark Streaming [29] and Samza [30] provides different strategies (once-at-time vs micro-batched) for streaming processing. In addition, hybrid computation models like Lambda [31] and Kappa [32] architectures could improve the system by generating dynamic models using unsupervised data.

REFERENCES

- [1] N. Council, "Frontiers in massive data analysis," 2013.
- [2] S. A. Schneider, "big data: big challenge, big opportunity," 2012.
- [3] O. Media, "Big data now," *O'Reilly Media*, 2012.
- [4] B. Taube, "Leveraging big data and real-time analytics to achieve situational awareness for smart grids,"
- [5] E. Dumbill, "What is big data? an introduction to the big data landscape," *oreilly.com*, <http://radar.oreilly.com/2012/01/what-is-big-data.html>, 2012.
- [6] R. Casado and M. Younas, "Emerging trends and technologies in big data processing," *Concurrency and Computation: Practice and Experience*, vol. 27, no. 8, pp. 2078–2091, 2015.
- [7] G. Bebis, R. Boyle, B. Parvin, D. Koracin, R. McMahan, J. Jerald, H. Zhang, S. Drucker, K. Chandra, Z. Deng et al., *Advances in Visual Computing: 10th International Symposium, ISVC 2014, Las Vegas, NV, USA, December 8-10, 2014, Proceedings*. Springer, 2014, vol. 8887.
- [8] X. Jin, B. W. Wah, X. Cheng, and Y. Wang, "Significance and challenges of big data research," *Big Data Research*, vol. 2, no. 2, pp. 59–64, 2015.
- [9] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *International Journal of Information Management*, vol. 35, no. 2, pp. 137–144, 2015.
- [10] B. K. Panigrahi, A. Abraham, and S. Das, *Computational intelligence in power engineering*. Springer Science & Business Media, 2010, vol. 302.
- [11] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A. H. Byers, and M. G. Institute, "Big data: The next frontier for innovation, competition, and productivity," 2011.
- [12] T. Soyata, R. Muralidharan, C. Funai, M. Kwon, and W. Heinzelman, "Cloud-vision: Real-time face recognition using a mobile-cloudlet-cloud acceleration architecture," in *Computers and Communications (ISCC), 2012 IEEE Symposium on*. IEEE, 2012, pp. 000 059–000 066.
- [13] A. Vinay, V. S. Shekhar, J. Rituparna, T. Aggrawal, K. B. Murthy, and S. Natarajan, "Cloud based big data analytics framework for face recognition in social networks using machine learning," *Procedia Computer Science*, vol. 50, pp. 623–630, 2015.
- [14] A. Fernández, R. García, R. Usamentiaga, and R. Casado, "Glasses detection on real images based on robust alignment," *Machine Vision and Applications*, vol. 26, no. 4, pp. 519–531, 2015.
- [15] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1. IEEE, 2001, pp. I–511.
- [16] M. Uřířák, V. Franc, and V. Hlaváč, "Detector of facial landmarks learned by the structured output svm," *VISAPP*, vol. 12, pp. 547–556, 2012.
- [17] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition*, vol. 29, no. 1, pp. 51–59, 1996.
- [18] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.
- [19] T. Ojala, M. Pietikäinen, and T. Maenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002.
- [20] A. Ranjan and S. Malik, "Parallelizing a face detection and tracking system for multi-core processors," in *Computer and Robot Vision (CRV), 2012 Ninth Conference on*. IEEE, 2012, pp. 290–297.
- [21] R. Rothe, M. Guillaumin, and L. Van Gool, *Non-maximum suppression for object detection by passing messages between windows*. Springer, 2015.
- [22] C. K. Williams and W. Wojcikiewicz, "A principled approach to non-maximum suppression for object detection."
- [23] M. B. Blaschko, "Branch and bound strategies for non-maximal suppression in object detection," in *Energy Minimization Methods in Computer Vision and Pattern Recognition*. Springer, 2011, pp. 385–398.
- [24] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 12, pp. 2037–2041, 2006.
- [25] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Technical Report 07-49, University of Massachusetts, Amherst, Tech. Rep., 2007.
- [26] C. Shan, "Learning local binary patterns for gender classification on real-world face images," *Pattern Recognition Letters*, vol. 33, no. 4, pp. 431–437, 2012.
- [27] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 365–372.
- [28] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics," *Journal of Big Data*, vol. 2, no. 1, pp. 1–21, 2015.
- [29] V. Hidalgo Santamaría, "Nuevas herramientas para el análisis de opinión en flujos de texto," 2015.
- [30] M. Gorawski, A. Gorawska, and K. Pasterak, "A survey of data stream processing tools," in *Information Sciences and Systems 2014*. Springer, 2014, pp. 295–303.
- [31] B. Data, "Principles and best practices of scalable realtime data systems," *N. Marz J. Warren. Henning*, 2014.
- [32] M. Kleppmann, "Designing data-intensive applications," 2015.

4.1.6. Face Recognition and Spoofing Detection System Adapted To Visually-Impaired People

- Villan, A. F., Candas, J. L. C., Fernandez, R. U., & Tejedor, R. C. (2016). Face recognition and spoofing detection system adapted to visually-impaired people. *IEEE Latin America Transactions*, 14(2), 913-921.
- DOI: 10.1109/TLA.2016.7437240
- URL: <http://ieeexplore.ieee.org/document/7437240/>
- Para el año 2015 la revista IEEE LATIN AMERICAN TRANSACTIONS (ISSN: 1548-0992) tuvo un índice de impacto de 0.436 (Q4) y índice de impacto a 5 años de 0.372.

Face Recognition and Spoofing Detection System Adapted To Visually-Impaired People

A. Fernández, J. L. Carús, R. Usamentiaga and R. Casado

Abstract— According to estimates by the World Health Organization, about 285 million people suffer from some kind of visual disability, of whom 39 million are blind, resulting in 0.7% of the world population. Computer vision techniques and image analysis can help improve visually-impaired people. In this project, a system that allows for facial recognition and detection of spoofing adapted to the needs of disabled people is proposed, implemented and validated. The architecture has been carefully selected and subsequently implemented following an innovative facial normalization algorithm in order to increase both the recognition rate of facial identification and spoofing detection. The information provided to the user is composed by the name of the person identified and whether it is real or fake image (photograph). This information is provided by means of a text-to-speech tool. This architecture can be integrated into video door-phone installations (videointercom installations), devices with reduced computing capabilities or the users' mobile phones. The architecture has been validated in a real environment with both real users and printed images achieving very good results.

Keywords— face recognition, spoofing detection, visually-impaired, system architecture.

I. INTRODUCCION

LA CEGUERA es una discapacidad visual que afecta a un 0.7% de la población mundial. Según las últimas estimaciones, casi un millón de personas en España padecen algún tipo de discapacidad visual y debido a las enfermedades retinianas mencionadas, alrededor de 70.000 personas presentan ceguera total. Según estimaciones de la Organización Mundial de la Salud (OMS), alrededor de 285 millones de personas padecen algún tipo de discapacidad visual de las cuales 39 millones son ciegas, lo que supone un 0.7% de la población mundial [1].

La discapacidad visual afecta de manera desigual a los distintos grupos de edad siendo más incisiva en personas mayores de 50 años representando el 65% del total (a pesar de que este grupo sólo representa el 20% del total de la población) [2]. Entre los cambios que se producen en la visión a consecuencia de la edad podemos destacar [1]:

- (1) Pérdida de la sensibilidad de la retina a la iluminación que origina una necesidad de utilizar iluminación más brillante.
- (2) Opacidad del cristalino que ocasiona menor visión y

reflejos molestos.

- (3) Elasticidad del cristalino y pérdida de la capacidad para enfocar.
- (4) Degeneración del vítreo que provoca la visión de manchas.
- (5) Disminución de la capacidad de las conjuntivas y glándulas lagrimales para lubricar adecuadamente los ojos.

Todo ello provoca que con la edad se pierda parte de la capacidad visual y se desarrollen patologías como pueden ser las cataratas, el glaucoma, la degeneración macular, afecciones parpebrales o la sequedad de los ojos [1].

Además, las estimaciones apuntan a un mayor envejecimiento de la población en Europa. En la Unión Europea (UE), se estima que la población mayor de 65 años y susceptible de padecer algunas de las principales patologías de ceguera aumente del 17.4% actual al 29% en 2050 [1]. La situación de la ceguera en España es muy parecida a la que encontramos en Europa o en otros países desarrollados; se prevé que en el futuro ésta aumente como consecuencia del incremento de distintos factores de riesgo como el envejecimiento de la población o el aumento en prevalencia de la diabetes.

Actualmente la tasa de personas mayores de 65 años en España se sitúa en torno a 17%, muy parecida a la tasa media registrada en Europa que es del 17.4% [3]. Sin embargo, se estima que en el futuro España será uno de los países con mayores retos para enfrentar el envejecimiento de su población, ya que está previsto que para el año 2050, el 33% de las personas serán mayores de 65 años, 4 puntos porcentuales por arriba de la media de la Unión Europea que se situará en el 29% [3]. La prevalencia de diabetes en España se sitúa también por encima de la UE [4].

Las tecnologías de la información y la comunicación (TIC) suponen una gran oportunidad en el desarrollo de nuevos sistemas y soluciones que permitan en líneas generales aumentar la calidad de vida de las personas con discapacidad visual. En este sentido, la visión por computador puede ser de gran ayuda para mejorar el día a día de estas personas. En concreto, el análisis facial puede servir para extraer información muy útil y relevante con el objetivo de ayudar a las personas con discapacidad visual en varias de sus tareas diarias dotándoles de un mayor grado de autonomía y seguridad.

El reconocimiento facial ha recibido muchas mejoras en los últimos años y hoy en día se acerca a la perfección. Los avances en el reconocimiento facial no han sido ajenos a las personas con discapacidad. Por ejemplo, recientemente se ha

A. Fernández, Fundación CTIC Centro Tecnológico, Asturias, España, alberto.fernandez@fundacionctic.org

J. L. Carús, Fundación CTIC Centro Tecnológico, Asturias, España, juanluis.carus@fundacionctic.org

R. Usamentiaga, Universidad de Oviedo, Asturias, España rusamentiaga@uniovi.es

R. Casado, Treelogic Centro Tecnológico, Asturias, España, ruben.casado@treelogic.com

presentado un bastón inteligente para ciegos que utiliza reconocimiento facial [5]. El bastón viene equipado con un sistema de reconocimiento facial, GPS y Bluetooth. Al divisar la cara de cualquier conocido o amigo cuya foto está almacenada en la tarjeta SD del bastón, este vibrará y dará, a través de un auricular Bluetooth, las instrucciones necesarias para llegar hasta esta persona. El sistema funciona con cualquier persona que se encuentre a 10 metros o menos. Además, gracias al GPS, el usuario recibirá instrucciones para llegar a donde quiera, como con cualquier navegador GPS.

Sin embargo, además de realizar la tarea de reconocimiento, hoy en día los sistemas biométricos tienen que lidiar con otro tipo de problemas, como el spoofing. En términos de seguridad de redes, este término hace referencia al uso de técnicas a través de las cuales un atacante, generalmente con usos maliciosos, se hace pasar por una entidad distinta a través de la falsificación de los datos en una comunicación.

Los sistemas de reconocimiento facial son vulnerables a los ataques de tipo spoofing. Esto puede ocurrir cuando una persona presenta una fotografía de la persona deseada a la cámara en lugar de la suya propia. Esto presenta un gran problema, pues es relativamente fácil y sencillo hacerse con una fotografía para ser usado a posteriori. Creemos que este es un tema realmente importante en las personas que presentan discapacidad visual.

Basado en las premisas anteriores, el objetivo del presente artículo es el de proponer, construir y validar una arquitectura basada en reconocimiento facial y sistema anti-spoofing que pueda ser integrado tanto en un videoportero como una aplicación móvil. De esta manera, se quiere dotar a los ciegos y disminuidos visuales de un instrumento o herramienta que le permita en un último fin mejorar la calidad de vida y aumente tanto su seguridad como la sensación de la misma en su hogar o cuando tenga interacciones con otras personas. La arquitectura propuesta se ha validado con usuarios reales y en un entorno real simulando las mismas condiciones que se podrían dar tanto en las imágenes capturadas por un videoportero como las imágenes capturadas por una persona con discapacidad visual por medio de su dispositivo móvil. Las contribuciones se comentan a continuación:

En primer lugar se propone un algoritmo para la normalización de la cara del usuario robusto en cuanto a rotaciones y desajustes en el algoritmo de detección facial. Está demostrado que un algoritmo robusto de normalización puede aumentar considerablemente la tasa de acierto en un algoritmo de detección facial.

En segundo lugar se propone usar un algoritmo de detección de spoofing con el objetivo de aumentar la seguridad del sistema facial a reconocer. Para ello el algoritmo se basa en el análisis de texturas pues este tipo de algoritmos han presentado una gran alta tasa de acierto con la ventaja de haber aplicado el algoritmo de normalización de la cara en el paso previo.

En tercer lugar se ha diseñado e implementado una arquitectura que está especialmente pensada para ser ejecutada en dispositivos con reducidas capacidades de cómputo. El

algoritmo ha sido diseñado e implementado en C++ y ha sido portado a la plataforma Android mediante JNI para la comunicación con los algoritmos de visión artificial.

Por último comentar que la arquitectura se complementa con una herramienta de "text to speech" para que la persona con discapacidad obtenga la información final del sistema: de quien se trata y si se trata efectivamente de una persona real o por el contrario se trata de una falsificación.

II. ESTADO DEL ARTE

A. Reconocimiento facial orientado a la discapacidad visual

El problema del reconocimiento facial adaptado a las personas con discapacidad visual ha sido investigado en sus diferentes formas. A continuación se resumen los trabajos más importantes, indicando para cada uno de ellos las características más importantes que han ido motivando el desarrollo de la arquitectura aquí propuesta.

En [6] se presenta un sistema de reconocimiento facial en dispositivos móviles para discapacitados visuales, pero se centra principalmente en reuniones con lo que aspectos como el campo visual capturado por el dispositivo móvil centran gran parte de la temática. En [7] se desarrolló un sistema de reconocimiento facial basado en Local Binary Pattern (LBP) [8]. Compararon este descriptor con otras alternativas (Local Ternary Pattern [9] o Histogram of Gradients [10]) y llegaron a la conclusión que el rendimiento de LBP es un poco superior, su coste computacional es menor y la representación de la información es más compacta. Y como se ha comentado anteriormente, en [5] se ha desarrollado un sistema de reconocimiento facial integrado en un bastón.

En ninguno de estos métodos se lleva a cabo la detección de spoofing, haciendo que el sistema tenga una vulnerabilidad alta ante este tipo de ataques. Creemos que es un punto muy importante sobre todo en personas con discapacidad visual. Además, ninguna de las alternativas antes comentadas está orientada a los videoporteros.

B. Detección de Spoofing

Como en ninguno de los casos anteriores se ha estudiado la detección de spoofing para ayudar a personas con discapacidad visual, procederemos a comentar los resultados más significativos en lo que a la detección de spoofing se refiere.

Existen muchos métodos diferentes para la detección de spoofing. Sin embargo, uno de los factores clave en una aplicación que debe funcionar en tiempo real y en un dispositivo embebido es que el método sea computacionalmente ligero. La mayoría de los algoritmos propuestos o bien son muy complejos y por tanto no son aptos para entornos reales, o bien no usan imágenes convencionales (por ejemplo multi-espectrales o termográficas) [11]. Los algoritmos basados en el análisis de micro-texturas ofrecen un buen resultado con un coste computacional relativamente bajo. En [11], se ha aplicado el algoritmo LBP a la detección de spoofing aplicando dicho operador a diferentes escalas. En [12], se analizan también diversas variantes del operador LBP para la detección de spoofing. En [13], también aplican el operador LBP junto con Máquinas de Soporte Vectorial (SVM) para la detección de spoofing. Tanto en [12] como en

[13], utilizan una división de la región facial en 9 regiones (3x3 regiones), pues comentan que es el número de divisiones que mejores resultados ha generado. Después aplican el operador LBP en cada una de las regiones y concatenan el histograma generado. Este factor será tenido en cuenta en nuestra arquitectura a la hora de generar el histograma LBP.

III. METODOLOGÍA

A continuación se comentan las principales fases del algoritmo. En los siguientes puntos de esta sección se irán detallando. Los principales pasos del algoritmo se pueden ver en la Fig. 1.

En primer lugar, este trabajo empieza con la localización de la cara de la persona en la imagen. Para esta tarea se usa un algoritmo basado en Viola & Jones [14], pero ampliado y modificado para lograr un algoritmo más robusto.

En segundo lugar y una vez localizada la cara en la imagen, un pre-procesamiento es necesario para lidiar con la pose, rotación e imprecisiones que provienen de la detección facial previa. Es por ello que se aplica un algoritmo de normalización facial. Nosotros hemos propuesto previamente este algoritmo de normalización facial para la detección robusta de gafas en personas en imágenes reales [15]. También lo hemos usado para la obtención de variables fisiológicas (frecuencia cardiaca y frecuencia respiratoria) obteniendo en ambos casos muy buenos resultados [16].

Una vez la región facial ha sido normalizada, LBP se emplea para obtener el conjunto de características que describirán la cara. Con el foco puesto en la complejidad computacional del algoritmo, el conjunto de características extraídas mediante el operador LBP ha sido cuidadosamente seleccionado, pues está produciendo muy buenos resultados tanto en la detección facial como en la detección de spoofing. Es por ello, que una vez extraído el conjunto de características faciales mediante el operador LBP, éste sirve como entrada en la etapa de clasificación a dos máquinas de soporte vectorial (SVM). SVM se aplica para clasificar el histograma obtenido de la región normalizada de la cara. La salida de los dos clasificadores SVM es por un lado la identificación facial y si dicha detección se corresponde con una falsificación o por el contrario es una cara real. SVMs son una técnica muy usada para la clasificación de los datos y ha sido propuesta en muchas ocasiones con temas relacionados con las tareas de reconocimiento de patrones, como por ejemplo reconocimiento facial [17]. Se usó LIBSVM para las tareas de entrenamiento y testeo de SVMs [18].

El sistema ha sido validado en un entorno real por medio de una aplicación desarrollada para la plataforma Android mediante llamadas JNI a los algoritmos de visión por computador que están desarrollados enteramente en C++. El motivo de desarrollar los algoritmos de visión por computador en C++ es debido a que es un lenguaje altamente compatible y recomendado para ser ejecutado en dispositivos con reducidas capacidades de cómputo.

La arquitectura ha sido desarrollada y validada en un videoportero y también se han portado los algoritmos al dispositivo móvil de una persona con discapacidad visual. Para validar la arquitectura en un videoportero ésta ha sido desarrollada meticulosamente teniendo en cuenta las posiciones, orientaciones y condiciones que presentan los

videoporteros con el objetivo de que la captura de imágenes para validar el sistema fueran lo más fidedignas posibles a un entorno real. El entorno seleccionado para validar el sistema se encuentra a la puerta del Centro Tecnológico donde realizamos nuestra actividad investigadora.

Por otro lado, para validar la arquitectura en una aplicación móvil, la aplicación fue desarrollada teniendo en cuenta las características y recomendaciones propuestas por otros autores, además de características que faciliten el encontrar la cara del usuario para permitir tomar las fotografías de la persona con la que está interactuando la persona con discapacidad visual.

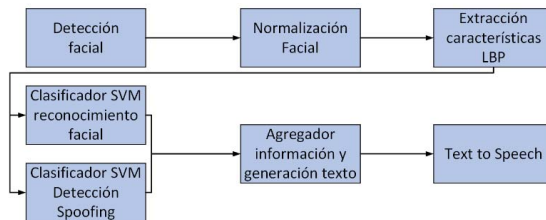


Figura 1. Principales pasos del algoritmo propuesto para la identificación facial, detección de spoofing y conversión a voz de la información agregada.

A. Detección facial

El detector facial propuesto por Viola & Jones [14] es comúnmente usado para realizar el seguimiento de la cara a lo largo del tiempo. Este detector presenta buenos resultados cuando la cara está prácticamente frontal, sin embargo, no puede lidiar cuando se presentan rotaciones a partir de 45 grados tanto en el eje vertical como en el eje horizontal. Es además un detector facial, es decir, no realiza el seguimiento de la cara a lo largo del tiempo (no es el algoritmo más adecuado para hacer el seguimiento o tracking facial). Además, es común que se encuentren falsos positivos o múltiples detecciones solapadas ante una misma cara por la forma que tiene de proceder dicho algoritmo. Por lo tanto, es necesario un algoritmo robusto para realizar un seguimiento de la cara a lo largo del tiempo. Con el objetivo de solucionar las dificultades anteriormente comentadas, se propone combinar tres detectores faciales en el primer frame o cuando se pierde una cara y se está buscando una nueva aparición. Cada detector está basado en el algoritmo de Viola & Jones antes comentado. De esta manera, se entrenan tres detectores: un detector frontal, un detector del perfil izquierdo y un detector del perfil derecho.

El resultado de aplicar el algoritmo de Viola & Jones a una imagen se corresponde con detecciones faciales en forma de rectángulos. En caso de que varios de los detectores faciales detecten una cara, la mínima región que recoge a ambos rectángulos es creada. El siguiente paso del algoritmo es el que trata en profundidad estas posibles detecciones.

B. Seguimiento y normalización facial

El siguiente paso se corresponde con el seguimiento de la cara (tracking) y la normalización de la región facial. El algoritmo de Viola & Jones solamente encuentra las posiciones de las caras en la imagen que se corresponden con rectángulos. Esto no es suficientemente preciso, pues se pueden incluir en el rectángulo de detección píxeles que no se corresponden con

región facial, sobre todo en las esquinas del rectángulo. Por lo tanto, un algoritmo de normalización es necesario.

Es bastante común en algoritmos de procesamiento facial usar un detector de piel para construir un mapa de piel que después servirá para normalizar la región facial [19]. Sin embargo, estos detectores no son muy robustos a los cambios de iluminación. También es muy común el uso de la técnica Active Appearance Model (AAM) [20] para detectar los principales puntos característicos de la cara. Sin embargo, un detector basado en AAM no es muy robusto, sobre todo en imágenes con baja resolución [21].

Es por ello que en la presente publicación, se propone: (1) un algoritmo de normalización facial y (2) un algoritmo de seguimiento facial para solucionar los problemas antes comentados.

El algoritmo de normalización facial es necesario para lidiar con la pose, rotación, escala e inexactitudes de la cara localizada. Este algoritmo está basado en el algoritmo propuesto en [15] pero modificado para obtener toda la región facial necesaria para después proceder tanto al reconocimiento facial como a la detección de spoofing. A continuación se resumen los principales puntos del algoritmo de normalización.

1) Algoritmo de normalización facial

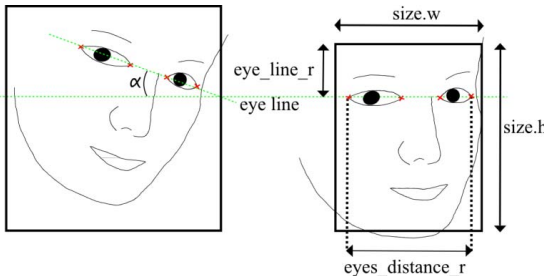


Figura 2. El algoritmo de normalización tiene en cuenta la rotación y realiza una corrección y normalización de la región facial.

A continuación se comentan los principales pasos del algoritmo. Una vez que la cara ha sido detectada, se aplica un detector robusto de características faciales basado en “Deformable Parts Models” (DPM) [21]. La salida del detector se corresponde con estimaciones de localizaciones para un conjunto de puntos característicos en la imagen: esquinas de los ojos, esquinas de la boca y nariz. Con el objetivo de calcular el ángulo de desviación de la cara, se calcula una recta de regresión que utiliza los cuatro puntos de los ojos. Esto puede verse en la Fig. 2. A continuación las caras son rotadas y alineadas de manera que los ojos siempre se encuentran en las mismas coordenadas en la imagen final. A continuación se calcula la región facial por encima y por debajo de los ojos, para que únicamente información relevante se procese en las etapas siguientes del algoritmo. Mediante este algoritmo de normalización hemos comprobado que la tasa de reconocimiento puede incrementarse significativamente. Aplicándolo al caso de reconocimiento de gafas por ejemplo, hemos obtenido una mejora del 1.55%. Como se ha comentado anteriormente, todos los detalles del algoritmo pueden verse en [15], aquí únicamente se muestran de manera resumida.

2) Algoritmo de seguimiento facial

Para realizar el seguimiento de las caras a lo largo del tiempo, se emplea un algoritmo de tracking muy robusto y recién publicado que está obteniendo muy buenos resultados [22]. Para el primer frame se emplea un rectángulo que envuelve el conjunto de características faciales devuelto por el detector. En caso de haber varias caras, se realiza el seguimiento de la cara que ocupe mayor espacio en la imagen, ya que se supone que es la principal. Dicho algoritmo de tracking realiza el seguimiento a lo largo del tiempo. La salida del algoritmo de tracking se utiliza después en el resto de frames y se aplica el algoritmo de normalización antes comentado.

C. Extracción de características mediante LBP

En primer lugar comentaremos en qué consiste el operador LBP y en segundo lugar comentaremos como se ha aplicado dicho operador para la extracción de las características faciales tanto para la identificación facial como para la detección de spoofing.

1) Operador LBP y mejoras incorporadas a dicho operador

El operador Local Binary Pattern (LBP) [23] es un tipo de operador que se suele utilizar para temas de clasificación. Es un operador muy potente en todo lo relacionado con el tema de clasificación de texturas. Dicho operador fue introducido en 1996 como un método para sintetizar la estructura del nivel de grises en imágenes. Dicho operador tiene en cuenta un vecindario local de píxeles alrededor de un píxel central. A continuación umbraliza los píxeles del vecindario con el valor del píxel central y usa el resultado como un número en binario como descriptor para ese vecindario y así sucesivamente para toda la imagen. Fue originalmente propuesto para un vecindario de 3x3, con 8 bits para codificar los valores binarios, puesto que son precisamente 8 los píxeles vecinos. Formalmente el operador LBP presenta la siguiente forma:

$$LBP(x_c, y_c) = \sum_{p=0}^7 2^p s(g_p - g_c) \quad (1)$$

donde en este caso p recorre los 8 vecinos con respecto al píxel central c, g_c y g_p son los valores del nivel de gris en c y p y:

$$s(x) = \begin{cases} 1, & \text{si } x \geq 0 \\ 0, & \text{en otro caso} \end{cases} \quad (2)$$

El proceso de codificación del operador original se ilustra en la Fig. 3.

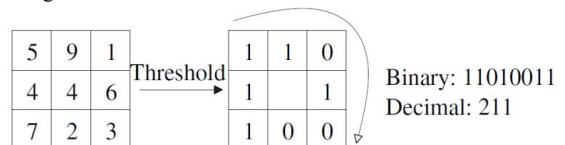


Figura 3. Operador LBP original (básico).

El operador fue posteriormente extendido para incorporar vecindarios de píxeles de diferentes tamaños, haciendo por tanto posible lidiar con las texturas a diferentes escalas [24].

Este hecho se denota por (P,R) donde P representa el número de puntos de muestreo (es decir el número de vecinos equiespaciados alrededor del píxel central) y R representa el radio del vecindario. Cuando las posiciones de los puntos de muestreo no se corresponden con posiciones enteras en la imagen, se utiliza la técnica de interpolación bilineal. En la Fig. 4 se puede ver un ejemplo de un radio circular del tipo $(8,2)$. La implementación de este LBP circular ($LBP_{P,R}$) toma la siguiente forma:

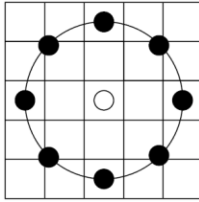


Figura 4. El operador circular $(8,2)$. Los valores de los píxeles son bilinealmente interpolados siempre que el punto de muestreo no coincida en el centro de un píxel.

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} 2^p s(g_p - g_c) \quad (3)$$

Otra extensión al operador original define como los patrones uniformes [24]. Un patrón LBP es uniforme cuando contiene como mucho dos transiciones a nivel de bit de 0 a 1 o viceversa visto como una cadena circular de bits cara uno de los patrones. Por ejemplo, los patrones 00000000, 00011110 and 10000011 son uniformes. El concepto de uniformidad es un concepto muy importante en la metodología de LBP, pues representa información estructural de primitivas como pueden ser los bordes o las esquinas en la imagen. A pesar de que únicamente hay 58 patrones uniformes de los 256 posibles patrones considerando un vecindario de 8 píxeles, cerca del 90% de los patrones en la región facial son uniformes. Es por ello que los patrones uniformes se pueden usar para reducir considerablemente la dimensionalidad de los datos manejados sin perder excesiva información. Para referirse a los patrones uniformes, se usa la siguiente notación: $LBP_{P,R}^{u2}$.

2) Extracción de características mediante el operador LBP para identificación facial y detección de spoofing

Con el objetivo de conseguir un algoritmo eficiente, robusto y computacionalmente ligero, se investigaron qué diferentes descriptores y operadores se podrían aplicar para representar de manera eficiente la región facial. Tras realizar una profunda investigación, analizar el estado del arte y hacer unas pruebas preliminares, se llegó a la conclusión que el operador LBP produce resultados excelentes tanto en el reconocimiento facial como en la detección de spoofing. Además, es un operador computacionalmente ligero. Por último y no por ello menos importante, sólo se computa una vez el operador, pues la información extraída de la región facial por medio del operador LBP se comparte por ambos módulos.

Después de etiquetar la imagen aplicándole el operador LBP, un histograma de esta imagen etiquetada $f_i(x,y)$ se puede definir como:

$$H_i = \sum_{x,y} I\{f_i(x,y) = i\}, i = 0, \dots, n-1 \quad (4)$$

donde n representa el número de valores diferentes producidos por el operador LBP y:

$$I\{A\} = \begin{cases} 1, & \text{si } A = \text{verdadero} \\ 0, & \text{en otro caso} \end{cases} \quad (5)$$

Para una representación eficiente de la información facial, las características extraídas mediante el operador LBP deberían disponer de información espacial. Es por ello que la imagen se ha dividido en m regiones $\{R_0, R_1, \dots, R_{m-1}\}$. De esta manera, el histograma básico descrito arriba puede extenderse en lo que se conoce como un "histograma mejorado espacialmente" [8], el cual es capaz de codificar tanto la apariencia y las relaciones espaciales de las distintas regiones faciales. Este nuevo histograma se define como:

$$H_{i,j} = \sum_{x,y} I\{f_i(x,y) = i\} I\{(x,y) \in R_j\} \quad (6)$$

donde $i=0, \dots, n-1$, $j=0, \dots, m-1$. El histograma por tanto sirve para describir la región normalizada de la cara teniendo en cuenta tres niveles de localidad: las etiquetas del histograma contienen información sobre los patrones a nivel de píxel, estas etiquetas son tenidas en cuenta para crear histogramas en regiones y en un último nivel, todos estos histogramas se concatenan para conseguir un histograma global.

D. Agregador información y generación de audio

Con el objetivo de minimizar la cantidad de información que se le proporciona al usuario y que ésta sea lo más fidedigna posible, los resultados se proporcionan al usuario al analizar un conjunto N de frames determinados. Por lo tanto, el sistema analiza los últimos N frames y en de que el algoritmo de tracking detecte la presencia de una cara, el sistema proporciona al usuario la información correspondiente a la detección facial y al sistema de spoofing.

Tras realizar varias pruebas y también tras analizar varias publicaciones [7], se ha llegado a la conclusión que un número reducido de frames ($N = 5$) es suficiente para minimizar la cantidad de información proporcionada al usuario y además, proporcionarle una información rápida, precisa y fiable al usuario.

Para el caso de uso de la aplicación móvil (ver Sección V), se proporciona además, información acústica acerca del estado del tracking con el objetivo de ayudar al usuario a "encontrar" a la persona que está hablando.

Para el caso del videoportero (ver Sección V) esto no es necesario, pues la detección facial del sistema no depende de la persona con discapacidad visual y no suele haber problemas para localizar y hacer el seguimiento de la cara en las imágenes capturadas por el videoportero.

IV. IMPLEMENTACIÓN

El sistema aquí propuesto ha sido desarrollado en C++ y haciendo uso de la librería de OpenCV. OpenCV (Open Source Computer Vision Library) es una biblioteca de visión por computador multiplataforma, publicada bajo la licencia BSD que permite ser usada tanto para uso académico como comercial. Incluye más de 500 algoritmos. La última versión estable es la 2.4.11 que es la que ha sido usada en la implementación del sistema.

Como herramienta para entrenar los modelos para el reconocimiento facial y detección de spoofing basados en SVM se hizo uso de la librería LibSVM [18].

El sistema ha sido construido de manera modular. Cada uno de los módulos que componen el sistema son los descritos anteriormente (ver Sección Metodología). Tanto la modularidad como el hecho de que el sistema estuviera desarrollado enteramente en C++ permitieron adaptar la el sistema a dos casos de uso diferentes.

Por un lado el sistema se adaptó para el caso de uso de los videoporteros.

El segundo caso de uso fue el de portar el sistema y los algoritmos a la plataforma Android para desarrollar una aplicación móvil que sirviera a las personas con discapacidad visual en su interacción diaria con otras personas.

A continuación comentamos de manera sucinta ambas implementaciones y en la Sección Casos de Estudio se comentan los resultados y detalles para cada uno de los dos casos de estudio.

A. Implementación en caso de uso de los videoporteros

En un primer lugar, el sistema se implementó bajo el sistema operativo Windows, haciendo uso de la librería OpenCV antes comentada y enteramente en C++. El objetivo es el de desarrollar una aplicación que pueda ser incluida en un videoportero. Puesto que el lenguaje C++ es uno de los más portables y con mejores rendimientos en cuanto a capacidad y velocidad de computación, la implementación desarrollada es perfectamente portable a un dispositivo embebido sin realizar apenas modificaciones al sistema.

B. Implementación en caso de uso de aplicación móvil

Dada la modularidad del sistema, el siguiente paso fue portar los algoritmos de visión por computador a la plataforma Android usando el framework JNI. Es un framework que permite que partes de la aplicación en Android se comuniquen con los algoritmos de visión artificial cuya implementación seguiría estando en C++. De esta manera el sistema no perdería excesivamente en rendimiento.

El objetivo de portar los algoritmos a la plataforma Android fue el de construir una aplicación para los dispositivos móviles que pudiera realizar la autenticación facial y la detección de spoofing orientada a gente con discapacidad visual. De esta manera, la aplicación ayudaría a estas personas a sus interacciones diarias con otra gente.

V. CASOS DE ESTUDIO Y RESULTADOS OBTENIDOS

En esta Sección se muestran dos casos de estudio, donde la herramienta desarrollada sirve de soporte para la creación de dos aplicaciones diferentes.

En la primera de ellas, se comentan los detalles para adaptar el sistema a la identificación facial y detección de spoofing en un videoportero, teniendo en cuenta las características que esto implica.

En segundo lugar, se portaron los algoritmos construidos a la plataforma Android con el objetivo de construir una aplicación móvil que sirviera como soporte a las personas con discapacidad visual en su interacción con otras personas.

A. Caso de estudio 1: videoportero

En primer lugar, con el objetivo de validar el sistema en un entorno lo más realista posible, se decidió realizar las pruebas a la entrada del centro tecnológico donde realizamos nuestra actividad investigadora. Además, otro factor que ha sido tenido en cuenta para una correcta validación del sistema, fue la ubicación y tipo de la cámara a instalar. Las imágenes capturadas tienen que ser lo más realistas posibles pues son las que sirven tanto para entrenar como para validar el sistema. Para ello, se realizó un estudio en el que se recogieron dos factores principales: altura del videoportero, distancia a la que se suelen ubicar las personas (ver Fig. 5).

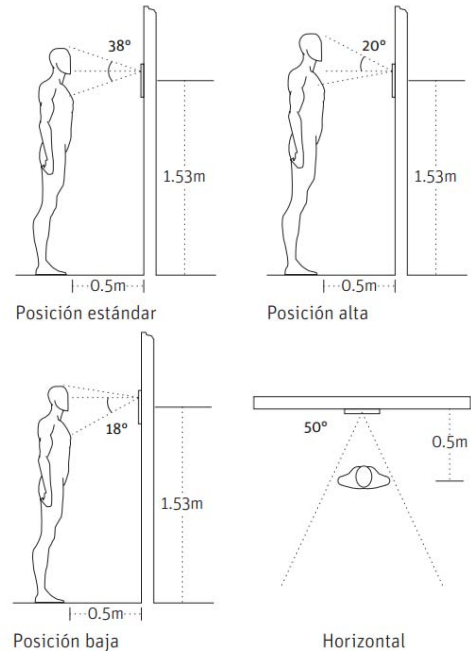


Figura 5. Alcance visual (posición y distancia) habituales de los usuarios respecto del videoportero

Como se puede apreciar, los usuarios se suelen colocar a una distancia de unos 0.5 metros. El videoportero ha sido colocado a una altura de unos 1.53m.

Puesto que el histograma LBP es calculado sobre cada división en la imagen y posteriormente concatenado, un número de divisiones pequeño consigue un histograma más pequeño pero tiene el inconveniente de que se pierde información espacial.

Es por ello que se realizaron diversas pruebas con el objetivo de ver el número de divisiones que mejores resultados aportaba tanto al reconocimiento facial como a la detección de spoofing así como otros parámetros del operador LBP. La mejor configuración de parámetros fue $LBP_{8,1}^{u2}$ (patrones uniformes, $R = 1$, $P = 8$) con SVM usando un kernel del tipo RBF, que está en consonancia con otras investigaciones relacionadas [12],[13],[25]. Un diagrama donde se muestra el proceso puede verse en la Fig. 6.

Para establecer los parámetros del algoritmo se hicieron uso

de las bases de datos de imágenes más comunes para la comparación de los algoritmos correspondientes al reconocimiento facial y a la detección de spoofing a partir de imágenes: Labeled Faces in the Wild (LFW) [26] y NUAU Photograph Imposter Database [27].

Con el objetivo de validar la arquitectura desarrollada, se instaló el prototipo a la entrada del Centro Tecnológico donde realizamos nuestra labor investigadora. Para ello se tuvieron en cuenta los detalles llevados a cabo por otros estudios, el alcance visual (posición y distancia) habituales de los usuarios respecto del videoportero.

Con esta configuración, lo habitual es que la cara detectada en la imagen tenga una resolución de unos 64x64 píxeles aproximadamente. Es por ello que las pruebas realizadas con las imágenes de las bases de datos antes comentadas se redimensionen a dicha resolución una vez la cara ha sido detectada (ver Fig. 6).

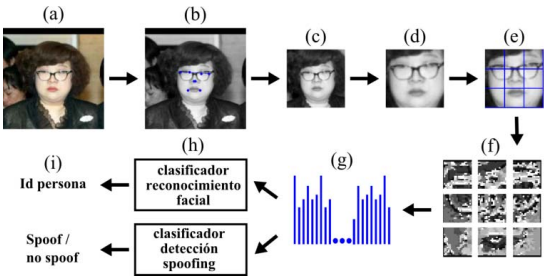


Figura 6. La imagen original (a) se convierte a escala de grises y se detectan los puntos característicos (b). A continuación se rota para alinearla en función del ángulo (c). Después obtenemos la región normalizada (64x64) de la cara (d). Aplicamos 3x3 divisiones a la región facial (e) y construimos la imagen $LBP_{8,1}^{u2}$ (f). Se construye el histograma dadas estas regiones para formar el vector de características (g). Por último se clasifica este vector usando ambos clasificadores (h) para obtener las respuestas finales (i).

Una vez seleccionados los parámetros del algoritmo, se hicieron pruebas durante 5 días consecutivos y con diferentes condiciones de iluminación (por la mañana y por la tarde). Cabe decir, que aunque la entrada del edificio no recibe luz directa es un sitio bastante iluminado. En total se realizaron pruebas con 25 usuarios.

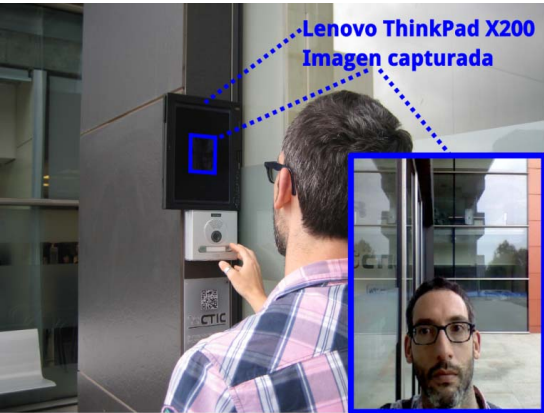


Figura 7. Para las pruebas se seleccionó un Lenovo ThinkPad X200 Tablet que se posicionó encima del videoportero con la webcam en la parte inferior del mismo. Se puede ver un detalle de la imagen capturada.

Consideramos este número como un número suficiente de personas con las que más a menudo interaccione una persona con discapacidad visual (familiares, amigos más próximos, etc). La disposición del equipo y su ubicación puede verse en la imagen de la Fig. 7.

Para llevar a cabo las pruebas se seleccionó un Lenovo ThinkPad X200. Para la captura de imágenes se utilizó la webcam incorporada al equipo cuya resolución es de 640x480 píxeles.

Para probar el sistema de spoofing, se imprimieron fotografías de cada uno de los 25 usuarios. Para cada usuario se validaron ambos algoritmos (reconocimiento facial y detección de spoofing). Para ello el usuario se debe situar delante del videoportero. Con el objetivo de que el algoritmo obtenga una buena detección facial, el sistema de tracking facial procesa frames hasta que localiza uno donde las características faciales detectadas estén simétricas (dentro de unos umbrales). De esta manera garantizamos que la detección es frontal y en condiciones para que los algoritmos operen con propiedad.

En caso de que el algoritmo de tracking facial no detecte una detección frontal por el mecanismo antes comentado, se le proporciona al usuario un comando de voz para que se sitúe frontalmente al videoportero. A continuación se resume en forma de tabla los principales resultados a los que se ha llegado.

TABLA I. RESULTADOS VIDEOPORTERO EN ENTORNO REAL.

ALGORITMOS	D = 1	D = 2	D = 3	D = 4	D = 5
RECONOCIMIENTO FACIAL	97%	98%	97%	94%	97%
DETECCIÓN SPOOFING	86%	86%	89%	84%	90%

Como se desprende de los resultados, el reconocimiento facial presenta unos resultados bastante buenos. El sistema de detección de spoofing presenta unos resultados algo peores, pero aún así son unos resultados bastante buenos. Cabe destacar en el cuarto día de pruebas (D = 4) el tiempo fue bastante soleado. Creemos pudo afectar al rendimiento de los algoritmos.

B. Caso de estudio 2: Aplicación móvil

Con el objetivo de sacar el máximo partido a la arquitectura, se decidió portar los algoritmos a la plataforma Android. De esta manera, el usuario con discapacidad visual dispone de una herramienta portable y usable para sus interacciones.

En caso de la aplicación móvil se han tenido en cuenta diversos factores que no ha sido necesario contemplar en el caso de uso expuesto anteriormente y que se comentan a continuación. En este caso, es necesario tener en cuenta que el dispositivo móvil y por tanto las imágenes capturadas son tomadas por personas con discapacidad y como consecuencia, muchas imágenes pueden presentar ruido, desenfoque, borrosidad y diferentes condiciones de iluminación. En

segundo lugar, es necesario aportar al usuario información auditiva acerca del estado de la aplicación: si ha detectado a una persona, si ha perdido el tracking de dicha persona, si la ha conseguido identificar y por último si la identificación ha sido real con el objetivo de proporcionarle un “feedback” con el estado de la aplicación, pero sin abrumarle con un exceso de información innecesaria.

El principal punto a tratar es que la persona con discapacidad visual no sabe a donde está enfocando su dispositivo. Para solucionar este factor, proponemos dos medidas: (1) guía del estado del tracking por medio de pitidos sonoros; y (2) analizar varios frames antes de predecir un resultado. A continuación comentamos los dos puntos.

En el primero de ellos, en caso de que el usuario de la aplicación detecte una cara, se producirá un pitido. De esta manera, el usuario mantendrá el móvil en esa posición para “mantener” el tracking y que el algoritmo pueda funcionar correctamente.

Si se pierde el tracking antes de que el algoritmo haya terminado, se producirán dos pitidos que indicarán al usuario que el algoritmo ha finalizado sin concluir una respuesta. En cambio, si el algoritmo consigue dar una respuesta al usuario es que ha procesado un número suficiente de frames N , que se corresponde con el segundo de los puntos indicado antes. Tras realizar varias pruebas y consultar la bibliografía relacionada, se ha establecido que el número de frames N que se deben procesar antes de proporcionar un resultado es de $N = 5$. Se ha establecido este valor con el objetivo de: (1) proporcionar unos resultados más robustos; y (2) proporcionar un feedback rápido para mejorar la fluidez en la interacción. Esto mejorará los resultados en los casos en los que las imágenes estén borrosas o presenten ruido.

Por lo tanto para cada uno de los clasificadores (identificador facial y spoofing) se obtienen, para cada imagen procesada tanto el identificador de la clase predicha como el valor de confianza para esa clase.

Es por ello que para un determinado número de frames N , antes de obtener el valor final, se calcula la clase ganadora, y por tanto el valor a predecir en función de éstos valores de confianza.

A continuación se adjunta una tabla donde se recogen los tiempos de ejecución del algoritmo. Para ello se han seleccionado dispositivos móviles y tablets de gama media/baja que disponemos en nuestro laboratorio, pues el objetivo del sistema es que los algoritmos se ejecuten en dispositivos con baja potencia computacional.

TABLA II. TIEMPOS DE EJECUCIÓN PARA VARIOS DISPOSITIVOS DE GAMA MEDIA/BAJA.

DISPOSITIVO MOVIL	FPS	TPO EJECUCIÓN SEC PARA N=5	TPO DE EJECUCIÓN ACEPTABLE
HTC DESIRE X	1.82	2.75	SI
LG OPTIMUS L2	1.86	2.69	SI
WOXTER (TABLET)	0.88	5.68	NO
SAMSUNG GALAXY Y	0.76	658	NO

Actualmente se está usando el valor promedio de los valores

de confianza, aunque el sistema podría contemplar otras alternativas. Entendemos que un tiempo superior a 3 segundos representaría un tiempo inaceptable para un correcto funcionamiento de la aplicación móvil.

Hay que recordar que la aplicación tiene como objetivo ayudar a mejorar a las personas con discapacidad visual en su interacción. Mantener durante más de 3 segundos el móvil en una posición más o menos fija es complicado para un usuario con estas características. Los dos modelos que cumplen estas restricciones son dos modelos que tienen unas especificaciones similares. Son móviles con un procesador de más o menos 1Ghz de doble núcleo y unos 768MB de memoria RAM. En la actualidad prácticamente cualquier móvil duplica en prestaciones a los dispositivos antes comentados. Es por ello que creemos que la aplicación puede ser perfectamente usable en cuanto a características y requerimientos técnicos.

A modo de ejemplo, la aplicación fue instalada en una tablet Edison 2 3G Quad Core y tiene un rendimiento de unos 4.33fps. Esto significa que en poco más de 1 segundo, el usuario tiene el resultado requerido.

En la imagen 8 se puede ver la captura de pantalla de varias ejecuciones que se corresponden con imágenes reales (parte izquierda de la imagen) y con fotografías (parte derecha de la imagen).



Figura 8. Capturas de pantalla de la aplicación móvil. Parte izquierda: imágenes reales. Parte derecha: imágenes impresas (no reales).

VI. CONCLUSIONES

En el presente trabajo hemos presentado una arquitectura para la identificación facial y detección de spoofing orientado a las personas con discapacidad visual. La arquitectura ha sido diseñada y desarrollada con el objetivo de conseguir un resultado robusto y computacionalmente ligero que pudiera ser embebido en elementos con una capacidad de cómputo moderada, como pudiera ser un videoportero o el dispositivo móvil ambos de personas con discapacidad visual.

El sistema ha sido probado y validado en entornos y condiciones reales obteniendo unos resultados muy satisfactorios. Además comentar que se han tenido en cuenta conclusiones y elementos clave obtenidos de otras publicaciones relevantes con el objetivo de obtener unos resultados más robustos y que la arquitectura final fuera lo más usable y adaptada a las personas con discapacidad visual.

Como trabajo futuro comentar dos aspectos principalmente. Uno de ellos es mejorar el sistema para hacerlo más robusto ante diferentes condiciones de iluminación. En [9] se comenta un algoritmo de pre-procesamiento para mejorar el reconocimiento facial cuando las condiciones de iluminación son adversas. Habría que ver cómo influye dicha etapa de pre-procesamiento en el algoritmo de detección de spoofing, pues podría disminuir las texturas en las imágenes faciales que hacen diferenciar una imagen real de otra falsa. El segundo de los factores a mejorar radica en mejorar la tasa de reconocimiento en lo que al algoritmo de detección de spoofing se refiere.

En conclusión, creemos que las personas con discapacidad visual pueden verse beneficiadas por la presente solución con el objetivo final de mejorar su calidad de vida.

REFERENCIAS

- [1] Gómez-Ulla de Irazábal, F., & Ondategui-Parra, S. (2012). Informe sobre la ceguera en España.
- [2] Vision 2020: The right to sight. World Health Organization (WHO) and International Agency for Blindness Prevention (IAPB). Action Plan (2006-2011).
- [3] Portal web del Eurostat, 2011.
- [4] Diabetes Atlas. International Diabetes Federation. 2010.
- [5] Maidenbaum, S., Hanassy, S., Abboud, S., Buchs, G., Chebat, D. R., Levy-Tzedek, S., & Amedi, A. (2014). The "EyeCane", a new electronic travel aid for the blind: Technology, behavior & swift learning. *Restorative neurology and neuroscience*, 32(6), 813-824.
- [6] C. Kramer, K. M., Hedin, D. S., & Rolkosky, D. J. (2010, August). Smartphone based face recognition tool for the blind. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE* (pp. 4538-4541). IEEE.
- [7] Balduzzi, L., Fusco, G., Odone, F., Dini, S., Mesiti, M., Destrero, A., & Lovato, A. (2010, September). Low-cost face biometry for visually impaired users. In *Biometric Measurements and Systems for Security and Medical Applications (BIOMS), 2010 IEEE Workshop on* (pp. 45-52). IEEE.
- [8] Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12), 2037-2041.
- [9] Tan, X., & Triggs, B. (2010). Enhanced local texture feature sets for face recognition under difficult lighting conditions. *Image Processing, IEEE Transactions on*, 19(6), 1635-1650.
- [10] Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* (Vol. 1, pp. 886-893). IEEE.
- [11] Maatta, J., Hadid, A., & Pietikainen, M. (2011, October). Face spoofing detection from single images using micro-texture analysis. In *Biometrics (ICB), 2011 international joint conference on* (pp. 1-7). IEEE.
- [12] Chingovska, I., Anjos, A., & Marcel, S. (2012, September). On the effectiveness of local binary patterns in face anti-spoofing. In *Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG-Proceedings of the International Conference of the* (pp. 1-7). IEEE.
- [13] Benlamoudi, A., Samai, D., Ouafi, A., Taleb-Ahmed, A., Bekhouche, S. E., & Hadid, A. Face Spoofing Detection From Single Images Using Active Shape Models with Stasm And LBP.
- [14] Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (Vol. 1, pp. I-511). IEEE.
- [15] Fernández, A., García, R., Usamentiaga, R., & Casado, R. (2015). Glasses detection on real images based on robust alignment. *Machine Vision and Applications*, 26(4), 519-531.
- [16] A. Fernandez, J.L. Carus, R. Usamentiaga, E. Alvarez, R. Casado, "Unobtrusive Health Monitoring System Using Video-Based Physiological Information and Activity Measurements", In *IEEE International Conference on Computer, Information, and Telecommunication Systems, CITS 2015, IEEE, vol. 1, no. 1, pp. 100-104, Gijón (Spain), 2015.*
- [17] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010, June). The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on* (pp. 94-101). IEEE.
- [18] Chang, C. C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3), 27.
- [19] de Haan, G., & Jeanne, V. (2013). Robust pulse rate from chrominance-based rPPG. *Biomedical Engineering, IEEE Transactions on*, 60(10), 2878-2886.
- [20] Coates, T. F., Edwards, G. J., & Taylor, C. J. (2001). Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, 23(6), 681-685.
- [21] Urfičar, M., Franc, V., & Hlaváč, V. (2012). Detector of facial landmarks learned by the structured output SVM. *VISAPP*, 12, 547-556.
- [22] Danelljan, M., Häger, G., Khan, F., & Felsberg, M. (2014). Accurate scale estimation for robust visual tracking. In *British Machine Vision Conference, Nottingham, September 1-5, 2014. BMVA Press.*
- [23] Ojala, T., Pietikainen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 29(1), 51-59.
- [24] Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7), 971-987.
- [25] de Freitas Pereira, T., Anjos, A., De Martino, J. M., & Marcel, S. (2013, June). Can face anti-spoofing countermeasures work in a real world scenario?. In *Biometrics (ICB), 2013 International Conference on* (pp. 1-8). IEEE.
- [26] Huang, G. B., Ramesh, M., Berg, T., & Learned-Miller, E. (2007). Labeled faces in the wild: A database for studying face recognition in unconstrained environments (Vol. 1, No. 2, p. 3). Technical Report 07-49, University of Massachusetts, Amherst.
- [27] Tan, X., Li, Y., Liu, J., & Jiang, L. (2010). Face liveness detection from a single image with sparse low rank bilinear discriminative model. In *Computer Vision-ECCV 2010* (pp. 504-517). Springer Berlin Heidelberg.



Alberto Fernández received his M.S. degree in Computer Science from University of Oviedo in 2009. Since 2007 he has been developing his professional career at the CTIC Foundation (Centre for the Development of Information and Communication Technologies in Asturias). He has been working in R&D projects. His research interests include both industrial and medical processes.



Juan Luis Carús PhD in Industrial Engineering (2015) from the Spanish University for Distance Education (UNED) and MSc in Telecommunications Engineering (2009) from the University of Oviedo. In 2007, he received the Thesis Award "Telefónica España" for the best record in engineering. Since July 2008, he is researcher with the R&D&I Area of CTIC – Technology Centre. His interests include ambient intelligence research projects focusing on signal processing, wearable computing and eHealth.



Rubén Usamentiaga Associate Professor in the Department of Computer Science and Engineering at the University of Oviedo. He received his M.S. and Ph.D. degrees in Computer Science from University of Oviedo in 1999 and 2005, respectively. In recent years he has been working on several projects related to computer vision and industrial systems. His research interests include real-time imaging systems and thermographic applications for industrial processes.



Rubén Casado received a B.Sc. Degree in Computer Science in 2005, a M.Sc. in Computing in 2008 and a PhD in Software Systems in 2013 from University of Oviedo, Spain. He has worked as a researcher and teaching assistant at the University of Oviedo, where he is currently a member of the Software Engineering Research Group. Currently he is the leader of the Big Data research program at Treelogic.

4.1.7. Driver Distraction Using Visual-Based Sensors and Algorithms

- Fernández, A., Usamentiaga, R., Carús, J. L., & Casado, R. (2016). Driver distraction using visual-based sensors and algorithms. *Sensors*, 16(11), 1805.
- DOI: 10.3390/s16111805
- URL: <http://www.mdpi.com/1424-8220/16/11/1805>
- Para el año 2015 la revista SENSORS-BASEL (ISSN: 1424-8220) tuvo un índice de impacto de 2.033 (Q1) y un índice de impacto a 5 años de 2.437.

Review

Driver Distraction Using Visual-Based Sensors and Algorithms

Alberto Fernández ^{1,*}, Rubén Usamentiaga ², Juan Luis Carús ¹ and Rubén Casado ²

¹ Grupo TSK, Technological Scientific Park of Gijón, 33203 Gijón, Asturias, Spain; juanluis.carus@grupotsk.com

² Department of Computer Science and Engineering, University of Oviedo, Campus de Viesques, 33204 Gijón, Asturias, Spain; rusamentiaga@uniovi.es (R.U.); rcasado@lsi.uniovi.es (R.C.)

* Correspondence: alberto.fernandez@grupotsk.com; Tel.: +34-984-29-12-12; Fax: +34-984-39-06-12

Academic Editor: Gonzalo Pajares Martinsanz

Received: 14 July 2016; Accepted: 24 October 2016; Published: 28 October 2016

Abstract: Driver distraction, defined as the diversion of attention away from activities critical for safe driving toward a competing activity, is increasingly recognized as a significant source of injuries and fatalities on the roadway. Additionally, the trend towards increasing the use of in-vehicle information systems is critical because they induce visual, biomechanical and cognitive distraction and may affect driving performance in qualitatively different ways. Non-intrusive methods are strongly preferred for monitoring distraction, and vision-based systems have appeared to be attractive for both drivers and researchers. Biomechanical, visual and cognitive distractions are the most commonly detected types in video-based algorithms. Many distraction detection systems only use a single visual cue and therefore, they may be easily disturbed when occlusion or illumination changes appear. Moreover, the combination of these visual cues is a key and challenging aspect in the development of robust distraction detection systems. These visual cues can be extracted mainly by using face monitoring systems but they should be completed with more visual cues (e.g., hands or body information) or even, distraction detection from specific actions (e.g., phone usage). Additionally, these algorithms should be included in an embedded device or system inside a car. This is not a trivial task and several requirements must be taken into account: reliability, real-time performance, low cost, small size, low power consumption, flexibility and short time-to-market. The key points for the development and implementation of sensors to carry out the detection of distraction will also be reviewed. This paper shows a review of the role of computer vision technology applied to the development of monitoring systems to detect distraction. Some key points considered as both future work and challenges ahead yet to be solved will also be addressed.

Keywords: driver distraction detection; visual-based sensors; image processing

1. Introduction

According to the most recent published World Health Organization (WHO) report, it was estimated that, in 2013, 1.25 million people were killed on the roads worldwide, making road traffic injuries a leading cause of death globally [1]. Most of these deaths happened in low- and middle-income countries, where rapid economic growth has been accompanied by an increased motorization and therefore, road traffic injuries. In addition to deaths on the roads, up to 50 million people incur non-fatal injuries each year as a result of road traffic crashes, while there are additional indirect health consequences associated with this growing epidemic. Road traffic injuries are currently estimated to be the ninth leading cause of death across all age groups globally, and are predicted to become the seventh leading cause of death by 2030 [1].

Distracted driving is a serious and growing threat to road safety [1]. Collisions caused by distracted driving have captured the attention of the US Government and professional medical organizations during the last years [2]. The prevalence and identification as a contributing factor in crashes is seen as an *epidemic* of American roadways, in words of Ray LaHood, when he was US Secretary of Transportation in 2012 [3]. There is not an exact figure regarding statistics about accidents caused by inattention (and its subtypes) since studies are made in different places, different time frames and therefore, different conditions. The studies referenced below show both the different statistics about inattention in general and those recorded when produced by distraction and fatigue in particular. These authors have estimated that distraction and inattention account for somewhere between 25% and 75% of all crashes and near crashes [4–8].

The trend towards increasing the use of in-vehicle information systems (IVISs) is critical [9] because they induce visual, manual and cognitive distraction [10] and may affect driving performance in qualitatively different ways [11]. Additionally, the advancement and prevalence of personal communication devices has exacerbated the problem during these last years [12]. All these factors can lead to the increment of the number of tasks subordinate to driving activity. These tasks, namely secondary tasks, which may lead to distraction [13], include eating, drinking, the act of taking something or tuning the radio or the use of cell phones and other technologies. The secondary tasks that take drivers' eyes off the forward roadway [14,15] reduce visual scan [16] and increase cognitive load may be particularly dangerous [13]. For example, the use of cell phones while driving, according to naturalistic studies [17], causes thousands of fatalities in the United States every year [18,19].

The purpose of this paper is the analysis of the state-of-the-art regarding the detection of drivers' distraction. The scope of the paper can be seen in Figure 1 and is commented as follows. The main methods for face detection, face tracking and detection of facial landmarks are summarized in Section 2 because they are a key component in many of the video-based inattention monitoring systems. In Sections 3–5, the main algorithms for biomechanical, visual and cognitive distraction detection are reviewed, respectively. Additionally, in Section 6, there are some algorithms detecting mixed types of distraction and, hence, are also reviewed. The relationship between facial expressions and distraction is also explored in Section 7. The key points for the development and implementation of sensors to carry out the detection of distraction will be considered in Section 8. In Section 9, the key ones to test and train driving monitoring systems are summarized. Privacy issues related to camera sensors are commented in Section 10. Lastly, conclusions, future aspects and challenges ahead will be considered in Section 11.

With the objective of introducing the scope and limitations of this review, some key aspects have been briefly introduced as follows. Driver distraction is just one form of inattention, which occurs when drivers divert their attention away from the driving task to focus on another activity. Therefore, a “complete” solution should consider all aspects of inattention. At least, the system should detect both distraction and drowsiness as the main contributing factors in crashes and near-crashes. As stated before, in this work, only distraction algorithms are summarized but one must not forget that other forms of inattention should be taken into account. Moreover, the use of on-board sensors already available in the vehicle to analyze driver behaviour is a low-cost and powerful alternative to the vision-based monitoring systems [20,21]. However, these systems should not be treated like different alternatives, because they can be used together (fused) in order to obtain indicators for monitoring [22]. Hence, for the sake of completeness, in this paper review only “purely” vision-based monitoring systems have been reviewed.

One of the challenges in decreasing the prevalence of distracted drivers is that many of them report that they believe they can drive safely while distracted [23]. However, for example, in connection with the use of mobile phones while driving, there is a great deal of evidence interacting with mobile devices, such as sending messages or engaging in conversations, which can impair driving performance because this interaction can create distraction. Moreover, a recent research showed that phone notifications alone significantly disrupted performance, even when drivers did not directly

interact with a mobile device during the task [24]. Another study suggests that people in general can reduce both inattention and hyperactivity symptoms simply by silencing the smartphones and avoiding notifications [25]. Therefore, it is clear that drivers should not use and notice the presence of the smartphones inside the car while driving. It should be pointed out that distraction generation is a very complex process and is scarcely addressed here. We recommend some research papers that focused on driver distraction generation: Angell et al. [26] focused on the process of cognitive load in naturalistic driving; Liang et al. [27] addressed the adaptive behaviour of the driver under task engagement and their results on visual, cognitive and combined distraction; Caird analyzed the effects of texting on driving [28]. In the context of intelligent vehicles, Ohn et al. [29] highlights the role of humans by means of computer vision techniques.

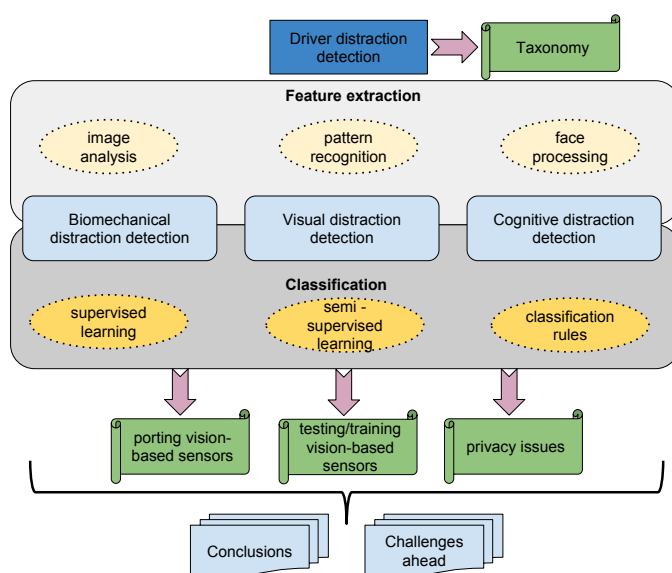


Figure 1. Scope of the present work.

1.1. Taxonomy

Both distraction and inattention have been inconsistently defined and the relationship between them remains unclear [30]. The use of different, and sometimes inconsistent, definitions of driver distraction can create a number of problems for researchers and road safety professionals [31]. Inconsistent definitions across studies can make the comparison of research findings difficult or impossible, can also lead to different interpretations of crash data and, therefore, to conclude different estimates of the role of distraction in crashes. This problem can be further seen in these recent works [32–35]. Many definitions have been proposed in order to define distraction [5,7,8,31]. Regan et al. [35] proposed a taxonomy of both driver distraction and inattention in which distraction is conceptualized as just one of several factors that may give rise to inattention. They concluded that driver inattention means “insufficient or no attention to activities critical for safe driving”. They defined driver distraction as “the diversion of attention away from activities critical for safe driving toward a competing activity, which may result in insufficient or no attention to activities critical for safe driving”. The definition proposed here is almost identical to that coined for driver distraction by Lee et al. [31].

It is acknowledged that the taxonomy proposed by Regan et al. [35] suffers from “hindsight bias”, that is, the forms of driver inattention proposed are derived from studies of crashes and critical incidents in which judgements have been made after the fact about whether or not a driver was

attentive to an activity critical for safe driving [35]. Driving consists of a variety of sub-tasks and it may not be possible to attend to all at the same time. Determining which sub-task is more important (and the driver, thus, should attend to) can often only be determined after the fact (i.e., after a crash or incident occurred) and, hence, this attribution of inattention is somewhat arbitrary [36]. Additionally, the dynamics of distraction [37], which identifies breakdowns on interruption as an important contributor to distraction should also be considered as part of this taxonomy, and hence, timing and context have implications on the algorithm design that should be taken into account.

1.2. Methodology

Papers addressed in this review are within the topic of distraction detection using vision-based systems. The search and review strategy is described below. A comprehensive review of the English language scientific literature was performed. It encompassed the period from 1 January 1980 to 31 August 2016. The following databases were used: EBSCO, ResearchGate, ScienceDirect, Scopus, Pubmed, Google Scholar and Web of Knowledge. Search terms related to driver distraction were employed combining all of them: driver, visual, cognitive, manual, biomechanical, vision, vision-based, impairment, distraction, distractions, review, task, tasks, inattention, performance, phone, sms, vehicle, problem, looking, face, head, pose, glasses, illumination, self-driving, tracking, sensors, image, traffic, safety, facts, privacy, issues, porting, taxonomy. Many items were returned from the search criteria shown before. These were, then, reviewed using the following criteria. Exclusion criteria were obviously non-relevant papers or from medical, electronic, networking, marketing and patent topics. Only publications from peer-reviewed English language journals were considered for inclusion. Additionally, reviewed papers were ordered by the number of references in order to include all relevant papers. Finally, in order to get the latest published papers, search filters were applied for this purpose. Search filters were applied to get publications only from years 2015 and 2016. References and bibliographies from the selected papers identified were examined to determine potentially additional papers. A total of approximately 1500 publications were revised in the review process.

2. Face and Facial Landmarks Detection

A common face processing scheme in many inattention monitoring systems, which can be seen in Figure 2, includes the following steps:

- Face detection and head tracking. In many cases a face detection algorithm is used as a face tracking one. In other cases, a face detection algorithm is used as an input for a more robust face tracking algorithm. When the tracking is lost, a face detection call is usually involved (that is why in Figure 2 these steps are placed inside the same external orange box).
- Localization of facial features (e.g., eyes). Facial landmarks localization is usually performed, but it should be noted that, in some cases, no specific landmarks are localized. So, in such cases, estimation of specific cues are extracted based on anthropometric measures from both face and head.

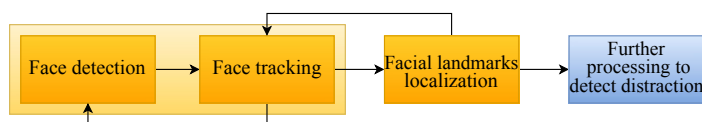


Figure 2. Common steps in most distraction monitoring systems.

2.1. Face Detection

Viola-Jones [38] have made object detection practically feasible in real world applications, which contains three main ideas that make possible to build and run in real time: the integral image,

classifier learning with AdaBoost, and the attentional cascade structure [39]. This framework is used to create state-of-the-art detectors (e.g., face detector [40]), available, for example, in OpenCV library. However, this framework turned out to be really time-consuming [41]. Moreover, cascade detectors work well on frontal faces but sometimes, they fail to detect profile or partially occluded faces.

One possible solution is to use the standard approach for human detection [42], which can also be used for face detection [39]. This approach is based on the Histogram of Oriented Gradients (HOG), which is a feature descriptor used in computer vision and image processing for the purpose of object detection. This approach can be trained with less images and faster [43]. Deep Learning approaches can also be used for face detection. For example, in [44], a deep learning approach, called DP2MFD, is used. DP2MFD detects faces at multiples scales, poses and occlusion by integrating deep pyramid features with Deformable Parts Models (DPMs). Experiments were carried out on four publicly available unconstrained face detection datasets, which demonstrated the effectiveness of the approach. However, this face detector was tested on a machine with 4 cores, 12 GB RAM, 1.6 GHz processing speed and it took about 26 s. Consequently, complex features may provide better discrimination power than Haar-like features for the face detection task. However, they generally increase the computational cost [44].

Some modifications to the Viola-Jones algorithm have been proposed [45,46] to speed up the algorithm. For example, in [45], different optimization techniques to speed up the Viola-Jones detector for embedded smart camera applications have been discussed. In their paper, skin colour information is integrated with the Viola-Jones detector in order to reduce the computation time. PICO (Pixel Intensity Comparison-based Object detection) is another modification of the standard Viola-Jones object detection framework, which scans the image with a cascade of binary classifiers at all reasonable positions and scales [46]. This algorithm can achieve competitive results at high processing speed. This is especially evident on devices with limited hardware support for floating point operations. PICO outperforms the other two OpenCV detectors in terms of accuracy and processing speed.

Since driver face monitoring system should work in all light conditions, lighting and camera selection is one of the most important stage in the design of the system. Lighting devices not only should provide enough light in environment, but they also should not hurt his/her eyes. For example, learning-based methods (e.g., Viola-Jones algorithm or PICO) can also be used for face detection in Infrared (IR) images [47].

Position of the camera inside the car is another key factor in the detection rate. For example, in [48], if the camera is installed under the front mirror of the car, face detection has 85% accuracy. But if it is installed on the dashboard, face detection reaches up to 93%. This is because they used the Viola-Jones face detector, which is trained to distinguish faces that are tilted up to about 45° out of plane (towards a profile view) and up to about 15° in plane. Therefore, if the camera is installed on the dashboard, the captured images will contain frontal or near-frontal faces. In [49], the camera was placed over the steering wheel column for two reasons: a) it facilitates the estimation of gaze angles, such as pitch, which is relevant for detecting distraction, and b) from a production point of view, it is convenient to integrate a camera into the dashboard. On the downside, when the wheel is turning, there will be some frames in which the drivers face may be occluded by the steering wheel. However, the driver is seldom very sleepy or inattentive to traffic while turning the steering wheel.

2.2. Face Tracking

Head pose estimation can be defined as the ability to infer the orientation of a person's head relative to the view of a camera and different studies have reported statistics showing consistent range of head motion [50], which (see Figure 3) can be decomposed in:

- Sagittal flexion/extension, i.e., forward to backward movement of the neck usually from -60° to 70° , which can be characterized by pitch angle.
- Axial rotation, i.e., right to left rotation of the head usually from -80° to 75° , which can be characterized by yaw angle.

- Lateral bending, i.e., right to left bending of the neck usually from -41° to 36° , which can be characterized by roll angle.

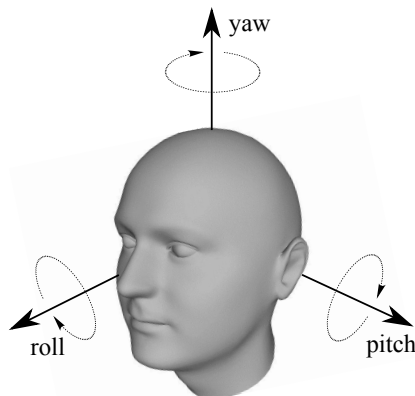


Figure 3. Head pose can be decomposed in pitch, yaw and roll angles.

Many vision-based algorithms for pose estimation have shown good performance when the head is near frontal, which is 95% of the time. But it is during those 5% of the time when interesting events, which are critical for safety, will occur [51]. Furthermore, as face orientation while driving is normally frontal, if the driver faces in other directions for a long period of time, this is probably due to fatigue or inattention [52]. Hence, a key component for a monitoring system based on face tracking is the ability to robustly and continuously operate even during large head movements. However, face tracking remains a challenging vision problem and, hence, a system for a continuous estimation of head movement is needed. On the other hand, as many head tracking algorithms have shown good performance when the head is near frontal, it can be concluded that the driver is looking away when tracking is unavailable. This information could be an alternative approach instead of adding more cameras to increase the range of the tracker.

Accordingly, numerous research works and publications have been trying to perform face tracking using a single camera and they are discussed as follows. Head pose estimation methods based on geometric approach using facial landmark and its 3D correspondences [49,53–56] can provide a good estimation and operate in real-time. For example, in [53], 3D pose estimation is achieved based on the position of the eyes and the mouth. A similar approach is proposed in [54], where only three points (eye centers and the middle point between the nostrils) are used to estimate continuous head orientation and gaze direction. Very closed to this approach, in [55], at least four prominent facial features are extracted from the face. After that, their correspondence on a 3D generic-face model is used to estimate head orientation. Oyini et al. [56] proposed the visual analysis of head position using a single camera aligning and scaling the 3D head model of the face according to the position and distance between the two eyes of the face in the 2D image. Another interesting approach recently published is [49], where a 3D head pose estimation system is proposed. This system is based on the 49 tracked 2D facial landmarks from Supervised Descent Method (SDM) tracker [57].

Other options include the combination of information [58–60], using for example, several classifiers [58,59] or combining 2D and 3D algorithms [60]. Asthana et al. [58] developed a system able to handle 3D pose variations up to $\pm 45^\circ$ in yaw and $\pm 30^\circ$ in pitch angles combining four different face detectors based on Viola-Jones framework. The drawback of this approach is that it requires four classifiers in order to track the face so it increases the execution time and memory requirements. In [59], the system consists of three interconnected modules, which detects drivers'

head, provides initial estimates of head pose, and continuously tracks its position and orientation in six degrees of freedom. Pelaez et al. [60], combined 2D and 3D algorithms to provide head pose estimation and regions of interest identification based on 3D information from a range imaging camera.

Alternatively, more than a camera can be used to implement the tracking [51,61–63], that is, a distributed camera system is commonly used, where two or more cameras can be located inside the car cockpit. Following this line of research, in [61], they proposed a distributed camera framework for gaze estimation using head pose dynamics based on the algorithm proposed in [51]. They predict three gaze zones: right, front and left. In [51], a continuous head movement estimator (CoHMEt) is proposed, which independently tracks the head in each camera, and their outputs are further analyzed to choose the best perspective and corresponding head pose. When tracking is lost, due to either the loss of facial point detection or the rejection of the estimated points, reinitialization is performed using a scoring criterion. In [62], they also used a two-camera system to overcome challenges in head pose estimation, which allows for continuous tracking even under large head movements, as proposed in [51]. Therefore, following the setup of [51], a two-camera system can provide a simple solution in order to improve tracking during large head movements. Two cameras are also used in [63] for head pose estimation. Head pose is tracked over a wide operational range in the yaw rotation angle using both camera perspectives.

For a quantitative evaluation over the algorithms the Mean Absolute Error (MAE) is commonly used. Best results for the different algorithms can be seen in Table 1, where different databases are used. For example, in [49,56] the public database of Boston University (BU) is used to evaluate the performance of the proposed head pose estimation scheme. Some other algorithms used naturalistic on-road data set [59]. Moreover, some algorithms achieving good performance did not indicate any database [60]. LISA-P Head Pose database [55] introduces head pose data from on-road daytime and nighttime drivers of different age, race and gender, with continuous ground truth measurements and manual annotation of facial features. Therefore, this database can be used to compare head pose algorithms and head behaviour studies. The LISA-P Head Pose Database consists of 14 video sequences of drivers in on-road driving environment in natural and spontaneous conditions. The video sequences were collected at a frame rate of 30 frames per second, with a 640×480 pixel resolution.

Table 1. Mean Absolute Error (MAE) (in degrees) of face tracking algorithms comparison working in an automobile environment.

Algorithm	Roll(°)	Yaw(°)	Pitch(°)
La Cascia et al. [64]	9.8	4.7	2.4
Oyini et al. [56] average results 1 camera	5.3	3.9	5.2
Oyini et al. [56] uniform illumination 1 camera	4.8	3.8	3.9
Oyini et al. [56] varying illumination 1 camera	5.3	5.1	6.3
Vicente et al. [49] 1 camera	3.2	4.3	6.2
Pelaez et al. [60] 1 Kinect device	2.7	3.8	2.5
Murphy et al. [59] 1 camera	2.4	4.7	3.4
Tawari et al. [51] (MPS + POS) 1 camera	3.0	8.2	7.6
Tawari et al. [51] (MPS + POS) 2 cameras	3.8	7.0	8.6
Tawari et al. [51] (MPS + POS) 3 cameras	3.5	5.9	9.0
Tawari et al. [51] (CLM + POS) 1 camera	3.4	6.9	9.3
Tawari et al. [51] (CLM + POS) 2 cameras	3.6	5.7	8.8
Tawari et al. [51] (CLM + POS) 3 cameras	2.7	5.5	8.5

Based on the results from Table 1, in [56], the MAE decreased by an average of 1.3° due to illumination variations. In [51], the best performance of 3.9% failure rate, which is the percentage of the time that the system output is unreliable, is achieved with the three-camera view compared with that of over 15% for the single view, which is a significant improvement.

2.3. Location of Facial Features

The detection of facial features (also called landmarks) is an essential part of many face monitoring systems. The problem of the precise and robust detection of facial landmarks has drawn a lot of attention during this decade. State-of-the-art methods include tree models [65,66], DPM [67], SDM [57], explicit shape regression [68] or learning local binary features [69]. A comprehensive survey of facial feature point detection can be seen here [70]. All the above listed research suffers more or less from a lack of verification and performance analysis with a realistic variation in lighting conditions. Therefore, further research should be performed in order to adapt these algorithms to the traffic research in general and to the drivers' monitoring systems in particular. Difficulties for proper detection of drivers' facial features are mainly due to the non-uniformity of light sources, asymmetric shades on their face and eye regions, or rapid changes in light intensity during real-world driving due to shadows caused by buildings, bridges, trees, or, for example, when entering or leaving a tunnel [71].

Eyes, as one of the most salient facial features reflecting individuals' affective states and focus of attention [72], have become one of the most remarkable information sources in face analysis. Eye tracking serves as the first step in order to get glance behaviour, which is of most interest because it is a good indicator of the direction of the driver's attention [73]. Glance behaviour can be used to detect both visual and cognitive distraction [74]. It has also been used by many studies as an indicator of distraction while driving [75] and has been evaluated in numerous ways [73]. Therefore, both eye detection and tracking form the basis for further analysis to get glance behaviour, which can be used for both cognitive and visual distraction.

Eye tracking data is typically captured through the use of a vehicle instrumented with an in-vehicle eye tracker system. On one hand, complex systems consist of single or multiple cameras directed at the driver's face. As the number of face cameras increases, so does the ability of the system to capture larger and more dramatic head movements of the driver. On the other hand, simpler systems consisting of one or two cameras are usually less expensive and easier to install than more complex systems. For example, in [76], a comparison of eye tracking systems with one and three cameras using Smart Eye technology [77] is performed. The system uses a single standard camera of VGA resolution together with IR flash illuminators. The three-camera system used is the Smart Eye Pro [77], which has similar properties as the one-camera system, but it also facilitates gaze direction in full 3D.

Eye detection is required before eye region processing. Eye detection methods can be divided into two general categories: (1) methods based on imaging in IR spectrum; and (2) feature-based methods. A literature survey on robust and efficient eye localization in real-life scenarios can be seen in [72], and a review on eye localization in car environment can be seen in [78].

Methods based on imaging in IR spectrum, which are commonly called "hardware-based" approaches, rely on IR illuminators to generate the bright pupil effect to driver head pose and gaze estimation. These methods use two ring-type IR light-emitting diodes: one located near the camera optical axis and the other located far from it. This approach is often used to detect visual distraction. In contrast to these methods, in [79], the authors use a progressive camera and only one on-axis lighting source [80]. In this situation, the camera always produces images with bright pupils and image processing techniques are applied to detect pupils. Based on thresholding techniques, the possible pupils can be selected. An appearance model, trained using Principal Component Analysis (PCA) and Support Vector Machine (SVM), is exploited to verify the final pupils. To increase the robustness against eyeglasses, the Generalized Symmetry Transform (GST) is incorporated achieving a recognition rate of 99.4% and 88.3% for users not wearing and wearing eyeglasses, respectively.

Regarding feature-based methods, different techniques are commonly applied. Image binarization [81], projection [82,83], face anthropometric properties of the face [84], individual classifiers [85] or particle filtering [86] can be used to detect driver's eyes. For example, in [86], an algorithm for eyes tracking based on particle filtering is proposed. Their method works with a low-cost IR camera device at a low frame rate. They used a single particle filter to track both eyes at the same time. Evaluation was carried out in a driving simulator with five users achieving an average accuracy of 93.25%. In [85],

two individual classifiers based on Haar-like features, one for the head and another for both eyes, were used. They tested face and eye detection in their research vehicle in daylight conditions achieving a hit rate of 97.2% for eye detection and a false alarm of 4.6%.

All in all, the task of accurate eye localization is challenging due to the high degree of eyes appearance variability: facial expression variations, occlusion, pose, lighting and other imaging conditions and quality [72], are frequently encountered in car environments. Another problem that is scarcely addressed in the literature is that, in strong sunlight, the driver tends to squint, which makes, even more difficult to track the eyes. To mitigate these deficiencies, different approaches can be adopted. Sigari et al. [82] proposed to extract symptoms of hypo-vigilance based on eye-region processing but without explicit eye detection stage. Flores et al. [84] proposed a combination of algorithms in order to deal with illumination conditions for both day and night. Rezaei et al. [71] used a methodology to enhance the accuracy, performance and effectiveness of Haar-like classifiers, especially for complicated lighting conditions. These authors also proposed ASSAM [87], which is based on the asymmetric properties of the driver's face due to illumination variations. A good solution is also to use a "divide and conquer" strategy to handle different variations at different stages [72].

3. Biomechanical Distraction

In connection with biomechanical detection and recognition using computer vision techniques, we can find two approaches. The first one involves hands secondary tasks recognition involving hands action, while the second one is based on hands tracking and information.

3.1. Secondary Tasks Involving Biomechanical Distraction

Zhao et al. [88–91] proposed different machine learning approaches to detect predefined driving postures, where four predefined postures were considered: (1) grasping the steering wheel; (2) operating the shift lever; (3) eating; and (4) talking on a cellular phone, which are recorded from the passenger seat, that is, from the right profile view of the driver. Yan et al. [92] proposed a combination of the Motion History Image (MHI) and POHG, and the application of Random Forest (RF) classifier for driving actions recognition. Trying to improve the accuracy of the aforementioned approach, the same authors included a Convolutional Neural Network (CNN) [93], which was tested over three datasets covering four driving postures: (1) normal driving; (2) responding to a cell phone call; (3) eating; and (4) smoking. For fair comparison, Yan et al. [93] re-implemented aforementioned state-of-the-art approaches [88–91] and carried out experiments on other two popular vision descriptor approaches (PHOG [94] and SIFT [95]). Classification accuracy of all of these methods can be seen in Table 2 evaluated on the Southeast University (SEU) driving posture dataset [88].

In connection with secondary tasks recognition, different computer vision algorithms have been proposed in order to detect cell phone usage of the driver while driving [96–100]. High recognition rates are usually obtained (from 86.19% to 95%) using very different approaches. Computer vision techniques seem to be the best approach for this task, whose results can be seen in Table 3, compared to other non-computer vision algorithms relying on inertial sensors of the mobile phone [101]. Best results are obtained by the algorithm proposed by Xu et al. [99], which consists of two stages: first, the frontal windshield region localization using DPM; next, they utilized Fisher vectors (FV) representation to classify the driver's side of the windshield into cell phone usage violation and non-violation classes. The proposed method achieved about 95% accuracy with a dataset of more than 100 images with drivers in a variety of challenging poses with or without cell phones.

It can be concluded that many different computer vision and machine learning techniques can be used to recognize predefined postures involving hand gestures. The CNN model offered a better performance than other approaches but with some limitations. The algorithm needs high computational resources making difficult to be applied in some conditions with common hardware architecture (e.g., embedded systems). Moreover, training a CNN needs a large amount of data, which is also difficult to obtain in some scenarios.

Table 2. Classification accuracy evaluated on the Southeast University (SEU) driving posture dataset [88].

Algorithm	Features	Classifier	Average Accuracy (%)
Zhao et al. [88]	Homomorphic filtering, skin-like regions segmentation and Contourlet Transform (CT)	RF	90.63
Zhao et al. [89]	Geronimo-Hardin-Massopust (GHM) multiwavelet transform	Multiwavelet Transform	89.23
Zhao et al. [90]	Histogram-based feature description by Pyramid Histogram of Oriented Gradients (PHOG) and spatial scale-based feature description	Perceptron classifiers	94.20
Zhao et al. [91]	Homomorphic filter, skin-like regions segmentation, canny edge detection, connected regions detection, small connected regions deletion and spatial scale ratio calculation	Bayes classifier	95.11
Bosch et al. approach [94]	PHOG	SVM	91.56
Lowe et al. approach [95]	SIFT	SVM	96.12
Yan et al. [93]		CNN	99.78

Table 3. Computer vision algorithms to detect cell phone usage. High recognition rates are usually obtained using very different approaches.

Algorithm	Features	Classifier	Recognition Rate (%)
Zhang et al. [96]	Features from the driver's face, mouth and hand	Hidden Conditional Random Fields (HCRF)	91.20
Artan et al. [97]	Image descriptors extracted from a region of interest around the face	SVM	86.19
Berri et al. [98]	Percentage of the Hand and Moment of Inertia	FV	91.57
Xu et al. [99]	DPM	FV	95
Seshadri et al. [100]	Raw pixels and HOG features	Real AdaBoost, SVM, RF	93.86

3.2. Hands Information

Hand detection is a challenging problem as human hands are highly deformable and are also exposed to different illumination conditions [102]. One approach for object detection relies on a sliding-window, where a model is learned based on positive samples (i.e., hands in different poses) of fixed size and negative samples with no hands. A classifier is then used to learn a classification rule. In order to detect hands at different scales, this scheme can be applied on hand images at different sizes. But a sliding window-based approach trained on hand instances was shown to be prone to false positive detection rates [103]. A recent common approach to improve the results is the assumption that hands can only be found in a small and predefined set of regions [103,104].

As opposed to training a model for hand shape or appearance and running a sliding window detector, two different approaches are analyzed in [103] taking into account three activity classes:

(1) two hands on the wheel; (2) hands on the instrument panel and (3) hand on the gear shift. The motion-cue-based hand approach uses temporal accumulated edges in order to maintain the most reliable and relevant information motion and then, it is fitted with ellipses in order to produce the location of the hands. The static-cue-based approach uses features in each frame in order to learn a hand presence model for each of the three regions and a second-stage classifier (SVM) produces the final activity classification. Martin et al. [104] also constraint the problem of hands detection to a number of regions of interest. They used HOG at different scales. Afterwards, a SVM is used to learn a hand presence in each of the three regions and ‘two hands on the wheel’ model for the wheel region. A similar approach is proposed in [102], training a linear SVM model for each region using a different set of descriptors. Ohn et al. [62] incorporated hand gestures in order to study preparatory motions before a maneuver had been performed, training a hand detector using fast feature pyramids. Gradient and colour channels are extracted for each patch image. They used CIE-LUV colour channels because they worked better compared to RGB and HSV. Afterwards, an AdaBoost classifier was applied in order to learn the features from the hands and finally, they trained a SVM-based detector using HOG features to differentiate the left hand from the right one. Later on, Ohn et al. [105] also explored the use of a pyramidal representation for each region of interest using HOG finding that edge features are particularly successful in the task of hands detection.

In order to compare these algorithms, a dataset of synchronized RGB and depth videos collected in an operating vehicle was proposed [106]. The CVRR-HANDS 3D dataset was designed in order to study natural human activity under difficult settings (background, illumination, occlusion) containing three subsets: (1) hand localization; (2) hand and objects localization; and (3) 19 hand gestures for occupant-vehicle interaction. Five regions of interest were considered: (1) wheel; (2) lap; (3) hand rest; (4) gear; and (5) instrument panel. Recognition rates from some of these previous algorithms using this database can be seen in Table 4.

Table 4. Hands recognition in different regions inside the car using CVRR-HANDS 3D dataset [106].

Algorithm	Features	Classifier	Regions	Recognition Rate (%)
Ohn et al. [106]	RGB data	SVM	5	52.1
Ohn et al. [106]	RGB combined with depth data	SVM	5	69.4
Martin et al. [104]	Hands cues	SVM	3	83
Martin et al. [104]	Hands and head cues	SVM	3	91
Ohn et al. [105]	Hands cues	SVM	3	90
Ohn et al. [105]	Hands and head cues	SVM	3	94

Summarizing, a common approach is to recognize if the hands are positioned in one of the established areas (wheel, gearbox and so on) and to track them over time. It could be considered that the steering wheel is the critical area because it is where hands should remain most of the time while driving. If hands remained in a non-critical zone for a certain period of time, which could be different for each of the non-critical areas, an alarm would be created to warn drivers to lay their hands in the correct position.

Hand Disambiguation

There is another interesting problem to solve related to hands detection that needs further research: hand disambiguation [107]. Once hands are detected, it is crucial to ensure that the hands belong to the driver. Both hand disambiguation and hand activity detection should be studied and considered together in order to infer final, clear and unambiguous results.

4. Visual Distraction

Visual distraction is often related to the on-board presence of electronic devices such as mobile phones, navigation or multimedia systems, requiring active control from the driver. It can also be

related to the presence of salient visual information away from the road causing spontaneous off-road eye glances and momentary rotation of the head. A 2006 report on the results of a 100-car field experiment [4] showed that almost 80% of all crashes and 65% of all near-crashes involved drivers looking away from the forward roadway just prior to the incident.

Engagements in visually distracting activities divert drivers' attention from the road and cause occasional lapses, such as imprecise control of the vehicle [108], missed events [28], and increasing reaction times [108]. Visual time sharing between the driving task and a secondary task reveals that the glance frequency to in-car devices is correlated to the task duration, but the average glance duration does not change with task time or glance frequency [109]. Drivers do not usually increase the glance duration for more difficult or longer tasks but rather increase the accumulated visual time sharing duration by increasing the number of glances away from the road [110]. As both single long glances and accumulated glance duration have been found to be detrimental for safety [110–112], a driver distraction detection algorithm based on visual behaviour should take both glance duration and repeated glances into account [113].

On the one hand, high-resolution cameras placed throughout the cabin are needed to view the driver's eyes from all head positions and at all times. Several economic and technical challenges of integrating and calibrating multiple cameras should be tackled to achieve this. Technically, eye orientation cannot always be measured in vehicular environments because eye region can be occluded by (1) sunlight reflections on eyeglasses; (2) the eye blink of the driver; (3) a large head rotation; (4) sunglasses; (5) wearing some kind of mascaras; (6) direct sunlight; (7) hats, caps, scarves; or (8) varying real world illumination conditions.

On the other hand, many security systems do not require such detailed gaze direction but they need coarse gaze direction to reduce false warnings [114,115]. For example, forward collision warning (FCW) systems need not only exterior observations but interior observations of the driver's attention as well to reduce false warnings (distracting and bothering the driver), that is, coarse gaze direction can be used in order to control the timing of warning emission when the system detects that the driver is not facing forwards.

Taking into account that errors in facial feature detection greatly affect gaze estimation [116], many researchers have measured coarse gaze direction by using only head orientation with the assumption that coarse gaze direction can be approximated by head orientation [117]. Head pose is a strong indicator of a driver's field of view and his/her focus of attention [59]. It is intrinsically linked to visual gaze estimation, which is the ability to characterize the direction in which a person is looking [118]. However, it also should be noted that drivers use a time-sharing strategy when engaged in a visual-manual task where the gaze is constantly shifted between the secondary task and the driving scene for short intervals of time [119] and often position the head in between the two involved gaze targets and only uses the eyes to quickly move between the two targets. In this situation, a face tracking algorithm would recognize this as a distracted situation based on head position, but the driver is constantly looking the road ahead. Therefore, in an ideal situation, both driver gaze tracking and eyes-off-road should be detected together [49].

In short, visual distraction can be categorized into two main approaches as it can be seen in Figure 4. In the first approach, which can be called "coarse", researchers measured the coarse gaze direction and the focus of attention by using only head orientation with the assumption that the coarse gaze direction can be approximated by the head orientation. In the second approach, which can be called "fine", researchers considered both head and eye orientation in order to estimate detailed and local gaze direction.

Moreover, considering its operating principles, visual distraction systems can be grouped in two main categories: hardware- and software-based methods. Additionally, some systems can combine these two approaches and therefore, a third category can also be considered, as seen in Figure 4.

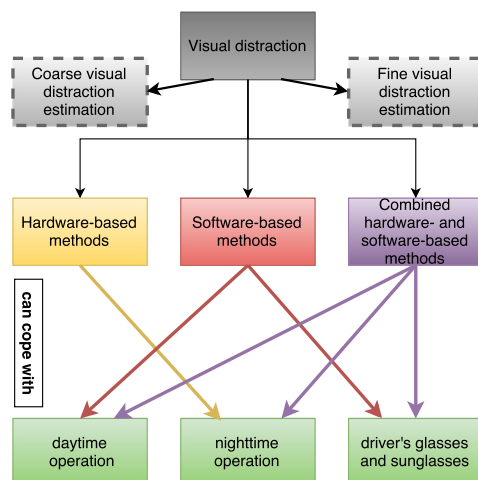


Figure 4. Visual distraction algorithms categorization.

4.1. Hardware-Based Methods to Extract Gaze Direction

Hardware-based approaches to head pose and gaze estimation rely on Near Infrared (NIR) illuminators to generate the bright pupil effect. These methods use two ring-type IR light-emitting diodes: one located near the camera's optical axis and the other located far from it [120–126]. The light source near the camera optical axis makes a bright pupil image caused by the red-eye effect, and the other light source makes a normal dark pupil image. The pupil was, then, easily localized by using the difference between bright and dark pupil images. Ji et al. used the size, shape, and intensity of pupils, as well as the distance between the left and right pupil, to estimate the head orientation. Specifically, the authors used the pupil-glint displacement to estimate nine discrete gaze zones [121,122], a geometric disposition of the IR LEDs similar to that of Morimoto et al. [120] and two Charge Coupled Device (CCD) cameras embedded on the dashboard of the vehicle. In connection with the CCD cameras, the first one is a narrow angle camera, focusing on the driver's eyes to monitor eyelid movement while the second one is a wide angle camera focusing on his/her head to track and monitor head movement. Based on this work, Gu et al. [124] proposed a combination of the Kalman filtering with the head motion to predict the features localization and used Gabor wavelet in order to detect the eyes constrained to the vicinity of predicted location. Another existent approach proposed by Batista et al. used dual Purkinje images to estimate a driver's discrete gaze direction [125]. A rough estimation of the head-eye gaze was described based on the position of the pupils. The shape of the face is modeled with an ellipse and the 3D face pose is recovered from a single image assuming a ratio of the major and minor axes obtained through anthropometric face statistics. In this method, further research is necessary in order to improve the accuracy of the face orientation estimation, which is highly dependent on the image face ellipse detection.

The aforementioned NIR illumination systems work particularly well at night. The major advantage of these methods is the exact and rapid localization of the pupil. However, performance can drop dramatically due to the contamination introduced by external light sources [126,127]. In addition, during daytime, sunlight is usually far stronger than NIR light sources and hence, the red-eye effect may not occur. Moreover, these methods could not work with drivers wearing glasses because the lenses create large specular reflections and scatter NIR illumination [127–129]. While the contamination due to artificial lights can easily be filtered with a narrow band pass filter, sunlight contamination will still exist [126]. Furthermore, such systems are vulnerable to eye occlusion caused by head rotation and blinking [114].

4.2. Software-Based Methods to Extract Gaze Direction

Combining facial feature locations with statistical elliptical face modelling, Batista et al. [83] presented a framework to determine the gaze of a driver. To determine the gaze of the face, an elliptical face modelling was used taking the eye's pupil locations to constraint the shape, size and location of the ellipse. The proposed solution can measure yaw head rotation over $[-30^\circ, +30^\circ]$ interval and pitch head rotation over $[-20^\circ, +20^\circ]$ interval.

Furthermore, despite the technical challenges of integrating multiple cameras, Bergasa et al. [130] proposed a subspace-based tracker based on head pose estimation using two cameras. More specifically, the initialization phase was performed using the Viola and Jones algorithm [40] and a 3D model of the face was constructed and tracked. In this work, head pose algorithm, which was the base for visual distraction estimation, could track the face correctly up to $[-40^\circ, +40^\circ]$.

A limitation of the software-based methods is the fact that they cannot often be applied at night [126,131]. This has motivated some researchers to use active illumination based on IR LEDs, exploiting the bright pupil effect, which constitutes the basis of these systems [126,131] (explained in previous section), or combine both methods, which can be seen in the next section.

4.3. Hardware- and Software-Based Methods to Extract Gaze Direction

Lee et al. [114] proposed a system for both day and night conditions. A vision-based real-time gaze zone estimator based on a driver's head orientation composed of yaw and pitch is proposed. The authors focused on estimating a driver's gaze zone on the basis of his/her head orientation, which is essential in determining a driver's inattention level. For night conditions, additional illumination to capture the driver's facial image was provided. The face detection rate was higher than 99% for both daytime and nighttime.

The use of face salient points to track the head was introduced by Jimenez et al. [132], instead of attempting to directly find the eyes using object recognition methods or the analysis of image intensities around the eyes. The camera was modified to include an 850 nm band-pass filter lens covering both the image sensor and the IR LEDs in order: (a) to improve the rejection of external sources of IR radiation and reduce changes in illumination and (b) to facilitate the detection of the pupils, because the retina is highly reflective of the NIR illumination of the LEDs. An advantage of salient points tracking is that the approach is more robust to the eyes occlusion whenever they occur, due to the driver's head or body motion.

Later on, the same authors extended their prior work in order to improve non-invasive systems for sensing a driver's state of alert [133]. They used a kinematic model of the driver's motion and a grid of salient points tracked using the Lukas-Kanade optical flow method [132]. The advantage of this approach is that it does not require one to directly detect the eyes, and therefore, if the eyes are occluded or not visible from the camera when the head turns, the system does not lose the tracking of the eyes or the face, because it relies on the grid of salient points and the knowledge of the driver's motion model. Experiments involving fifteen people showed the effectiveness of the approach with a correct eyes detection rate of 99.41% on average. It should be noted that this work is focused on sensing the drivers' state of alert, which is calculated measuring the percentage of eyelid closure over time (PERCLOS), and it is not focused on distraction detection.

Eyes Off the Road (EOR) detection system is proposed in [49]. The system collects videos from a CCD camera installed on the steering wheel column and tracks facial features. Using a 3D head model, the system estimates the head pose and gaze direction. For night time operation, the system requires an IR illumination. The proposed system does not suffer from the common drawbacks of NIR based systems [121,122,125], because it does not rely on the bright pupil effect. The system works reliably with drivers of different ethnicities wearing different types of glasses. However, if the driver is wearing sunglasses, it is not possible to robustly detect the pupil. Thus, to produce a reliable EOR estimation in this situation, only head pose angles are taken into account.

Cyganek et al. [134] proposed a setup of two cameras operating in the visible and near infra-red spectra for monitoring inattention. In each case (visible and IR) two cascade of classifiers are used. The first one is used for the detection of the eye regions and the other for the verification stage.

Murphy-Chutorian et al. used Local Gradient Orientation (LGO) and Support Vector Regression (SVR) to estimate the driver's continuous yaw and pitch [135]. They used head pose information extracted from a LGO and SVR to recognize drivers' awareness. The algorithm was further developed in [59] by introducing a head tracking module built upon 3D motion estimation and a mesh model of the driver's head. There is a general weakness here as the tracking module may easily diverge from face shapes that are highly different to the given mesh model.

4.4. Driver Distraction Algorithms Based on Gaze Direction

In these previous Sections 4.1–4.3, gaze direction is extracted using different methods. The next step is to detect distraction using gaze direction regardless of the type of method used to extract this information, and hence, is commented as follows.

Many software-based methods have been proposed in order to detect visual distraction, many of which, rely on “course” information extracted from visual cues [114,136–139]. Hattori et al. [136] introduced a FCW system using drivers' behavioural information. Their system determines distraction when it detects that the driver is not looking straight ahead. Following this approximation, an Android app [137] has been developed to detect and alert drivers of dangerous driving conditions and behaviour. Images from the front camera of the mobile phone are scanned to find the relative position of the driver's face. By means of a trained model [38] four face related categories were detected: (1) no face is present; (2) facing forwards, towards the road; (3) facing to the left and (4) facing to the right. Another related system is proposed by Flores et al. [138] where, in order to detect distraction, if the system detects that the face position is not frontal, an alarm cue is issued to alert the driver of a danger situation. Lee et al. [114] proposed a vision-based real-time gaze zone estimator based on a driver's head orientation composed of yaw and pitch. This algorithm is based on normalized histograms of horizontal and vertical edge projections combined with an ellipsoidal face model and a SVM classifier for gaze estimation. In the same research line but in a more elaborated fashion, Yüging et al. [139] used machine vision techniques to monitor the driver's state. The face detection algorithm is based on detection of facial parts. Afterwards, the facial rotation angle is calculated based on the analysis of the driver's head rotation angles. When the angle of facial orientation is not in a reasonable range and lasts for a relatively long time, it can be thought that the driver is distracted and warning information will be provided.

Additionally, other software-based approaches rely on “fine” information considering both head and eye orientation in order to estimate distraction [83,130,140,141]. Pohl et al. [140] focused on estimating the driver's visual distraction level using head pose and eye gaze information with the assumption that the visual distraction level is non-linear: visual distraction increased with time (the driver looked away from the road scene) but nearly instantaneously decreased (the driver re-focused on the road scene). Based on the pose and eye signals, they established their algorithm for visual distraction detection. Firstly, they used a Distraction Calculation (DC) to compute the instantaneous distraction level. Secondly, a Distraction Decision-Maker (DDM) determined whether the current distraction level represented a potentially distracted driver. However, to increase the robustness of the method, also the robustness of the eye and head tracking device to adverse lighting conditions has to be improved.

Bergasa et al. [126] presented a hardware- and software-based approach for monitoring driver vigilance. It is based on a hardware system, for real time acquisition of driver's images using an active IR illuminator and a software implementation for real time pupil tracking, ocular measures and face pose estimation is proposed. Finally, driver's vigilance level is determined from the fusion of the measured parameters into a fuzzy system. The authors yielded an accuracy percentage close to 100% both at night and for users not wearing glasses. However, the performance of the system decreases

during daytime, especially in bright days, and at the moment, the system does not work with drivers wearing glasses [126].

Recently, Lee et al. [141] evaluated four different vision-based algorithms for distraction under different driving conditions. These algorithms were chosen for their ability to distinguish between distracted and non-distracted states using eye-tracking data [141]. The resulting four algorithms, summarized in Table 5, are commented next:

1. Eyes off forward roadway (EOFR) estimates distraction based on the cumulative glances away from the road within a 6-s window [7].
2. Risky Visual Scanning Pattern (RVSP) estimates distraction by combining the current glance and the cumulative glance durations [142].
3. “AttenD” estimates distraction associated with three categories of glances (glances to the forward roadway, glances necessary for safe driving (i.e., at the speedometer or mirrors), and glances not related to driving), and it uses a buffer to represent the amount of road information the driver possesses [143–145].
4. Multi distraction detection (MDD) estimates both visual distraction using the percent of glances to the middle of the road and long glances away from the road, and cognitive distraction by means of the concentration of the gaze on the middle of the road. The implemented algorithm was modified from Victor et al. [146] to include additional sensor inputs (head and seat sensors) and adjust the thresholds for the algorithm variables to improve robustness with potential loss of tracking.

Table 5. AUC comparisons by algorithm across tasks.

Task	Algorithm			
	RVSP	EOFR	AttenD	MDD
Arrows	0.67	0.75	0.71	0.87
Bug	0.78	0.87	0.80	0.86

Considering the results of the ROC curves, AUC values, accuracy and precision, it is apparent that a trade-off exists between ensuring distraction detection and avoiding false alarms, which complicates determining the most promising algorithm. More specifically, the MDD algorithm showed the best performance across all evaluation metrics (accuracy, precision, AUC). Although the EOFR algorithm had promising AUC values, the AttenD algorithm often yielded better accuracy and precision. Additionally, the RVSP algorithm consistently yielded the lowest values for both accuracy and precision, but yielded a slightly higher AUC value than AttenD. All of the algorithms succeeded in detecting distraction well above chance detection ($AUC = 0.5$). The performance of the algorithms varied by task, with little difference in performance for the looking and reaching task (bug) but more stark differences for the looking and touching (arrows). The AUC for each task for each algorithm is provided in Table 5.

5. Cognitive Distraction

Cognitive distraction is a critical area of concern with regard to driver distraction, particularly as related to tasks of listening and conversing, but also, as related to spontaneously occurring processes like daydreaming or becoming lost in thought, which may occur really often on long drives. The term “cognitive load” can be defined as any workload imposed on a driver’s cognitive processes [26]. There are several types (and subtypes) of scenarios where cognitive load may occur during (see Figure 5), and therefore, affect driving. For further information, the reader may refer to [26]. These include:

1. Cognitive load imposed by secondary tasks undertaken while driving.
2. Cognitive load associated with the driver's internal activity.
3. Cognitive load arising from the driving task itself.

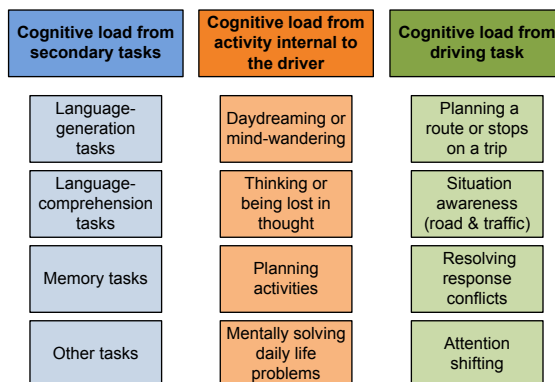


Figure 5. Classification of main types and subtypes of cognitive load while driving.

5.1. Behavioral and Physiological Indicators of Cognitive Load

The research literature documents several types of measures associated with periods of cognitive load. Secondary tasks imposing cognitive load lead to: (1) a high percentage of glances on the forward road and; (2) unusually long glances on the forward road. These two metrics together have been found to be uniquely indicative of cognitive loads [110,147]. Moreover, a narrowing of the spatial extent of scanning [147] is also produced, which is reflected in slightly fewer glances to locations where the mirrors, the speedometer and the areas peripheral to the road centre are located [26,148]. As a result, cognitive load may cause an increasing gaze concentration towards the middle of the road [11].

An eye-gaze pattern could be used to differentiate the action of only driving from driving under the influence of any cognitive task [147]. Drivers under cognitive distraction had fewer saccades per unit time, which was consistent with less exploration of the driving environment [149]. Saccades may be a valuable index of mental workload [150]. In fact, the standard deviations of both eye and head movement could be suitable for detecting cognitive distraction causing gaze concentration and slow saccades when drivers are looking forward [151]. A higher blink rate and a shrink in visual searching range were observed when the driver was cognitively distracted [152]. Kircher et al. [144] indicated the percentage of time the driver spent observing the road ahead, which is called the percentage road center (PRC) of gaze direction, was more than 92% under cognitive distraction in a field study.

Therefore, both glance and blink measures can be used to detect cognitive distraction. For example, He et al. [153] have observed that mind-wandering has effects on glance patterns and blink rates similar to those observed in periods of cognitive secondary task load. During mind-wandering, there is also an increasing concentration of gaze on the forward road with concomitant narrowing of scanning, longer glances on the forward road, and changes in blink rate [26]. Results from [154] suggested that performance data and oculomotor scanning behavior may allow the detection of drivers' mind wandering episodes before they are recognized by the driver himself/herself, potentially providing interventions to detect inattentiveness and alert drivers. Blink rate seems to be a promising indicator of cognitive processing [27]. However, there are measurement issues that may affect how successfully it can be applied in discriminating different types of task loading during driving. For example, there are some questions left about whether it offers sufficient sensitivity when extracted from real world data acquired from a complex task like driving wherein there are inherent temporal variations in driving task load [26,27].

Physiological measures can also be used to detect cognitive load. The average value of pupil diameter is suitably used as a physiological feature for detection of cognitive distraction [155]. When cognitive loads such as arithmetic or conversation were imposed to the subjects, dilation of pupils occurred by acceleration of the sympathetic nerve, resulting in an increase of diameter of pupils [156]. The average value of pupil diameter caused by cognitive loads, such as arithmetic, increased by 13.1% compared with ordinary driving [156]. The tests were performed in a driver simulator in controlled settings. Further experiments are required in a naturalistic setting. Moreover, additional works highlighted the difficulty in estimating cognitive load using pupil diameter during a dialogue task [157] or in different lighting conditions [158].

The same limitation applies to other physiological measures, as Heart Rate (HR), which tends to increase as cognitive task load raises [159]. The traditional method to quantify these physiological measures is by wearing physiological sensors. However, HR measurements can be acquired using computer vision techniques, and consequently, special care has been taken reviewing HR information. Additionally, it is considered a good indicator of fatigue, stress and cognitive load.

By means of the use of HR information the cognitive state of a driver can be monitored [160] in controlled settings. Changes in HR have been noted during certain driving tasks [161]. Similarly, Apparies et al. [162] showed that HR and Heart Rate Variability (HRV) may serve as early indicators of fatigue. In general, HRV specifically measures mental workload, while HR measures physical one [163]. HRV analysis is a strong indicator of mental stress or workload caused by driving tasks [162,164,165]. Experiments carried out in a driving simulator by Zhao et al. [166] found that human heart rates violently fluctuate during a mental stress situation. Ostlund et al. [167] and Miller et al. [165] identified both HR and HRV as promising indicators of the driver's stress level, by increasing HR and decreasing HRV [165,167]. Physiological measures, such as HR and skin conductance level, tend to increase as cognitive task load increases [159].

There are some research works able to extract HR and HRV from face video images in real time from human faces [168–170]. Eulerian Video Magnification framework [171] can be also used to obtain human pulse from a video sequence [172]. In [173], the authors described an approach offering a non-invasive, non-contact means of cardiac monitoring. Once the HRV time series are extracted, feature generation, feature selection and classification should be performed. The conventional method that uses Fast Fourier Transform (FFT) analysis on HRV is 2-min long. In [174], a new method developed by using wavelet-based feature and SVM for classification uses only 1-min HRV signals. Moreover, this method increases accuracy, sensitivity and specificity compared to FFT-based results.

Therefore, concerning cardiovascular measures, they have been reported to be sensitive to mental workload changes and both HR and HRV are widely adopted mental workload measures because they are easy to use and provide fundamental information about the autonomic nervous system [175]. Most methods [168–170] enable low-cost, non-contact cardiovascular activity monitoring using regular RGB cameras by analyzing minute skin color changes caused by periodic blood flow. Nevertheless, for automotive applications, these methods can encounter difficulties under different illumination conditions [176]. In [176], the authors proposed an artifact reduction method, which is caused by lighting variation. Another option is to use an IR-based camera system suitable for automotive applications [177].

To conclude this section, the use of physiological parameters can be used to monitor the cognitive state of the driver. Many of these parameters have been described in controlled settings, but further experiments are required to validate their capability in naturalistic conditions. The main algorithms in this matter are included in the next section.

5.2. Algorithms

Zhang et al. [178] used a decision tree approach to estimate drivers' cognitive workload from eye gaze-related features and driving performance. Liang, Reyes, et al. [179] showed that the SVM models can also detect cognitive distraction. The model's accuracy and sensitivity increased with

window size, suggesting that using longer periods to summarize the data made the distraction signal easier for the models to detect. The conclusion was that the best models were obtained using 40-s window size. Additionally, Liang, Lee, et al. [180] also used Bayesian Network (BN) models and found that they could identify cognitive load reliably for simulator data, and also found that Dynamic Bayesian Networks (DBNs), which considered time dependencies of driver's behaviour, gave a better performance than static BN models. This fact suggests that time-dependent relationship is critical in estimating the cognitive state of the driver. However, to train DBN models, longer training sequences are necessary to obtain more accurate and sensitive models. The results obtained in [180] using BNs, which stated that window size did not affect model performance, clearly conflict with those of Liang et al. [179], which found that larger window sizes improved the detection of cognitive distraction, although another data mining method, SVM, was applied in that study. An additional work from Liang et al. [181] compared SVMs, SBNs, and DBNs in detecting cognitive distraction using the best parameter settings from the same dataset used in the previous two studies [179,180]. DBNs produced the most accurate and sensitive models compared to SBN and SVM. Based on the comparisons of SVMs and BNs, Liang et al. [27,182,183] used a hierarchical layered algorithm, which incorporated both a DBN and a supervised clustering algorithm, to identify feature behaviors when drivers were in different cognitive states. This layered algorithm includes a DBN algorithm at the higher level to model the time-dependent relationship of driver behavior and a supervised clustering algorithm at the lower level to identify feature behaviors. The layered algorithm overcomes the disadvantages of DBNs and significantly improves computational efficiency in training and prediction. Miyaji et al. [184] proposed an approach to detect eye and head movement tracked via standard deviation and categorized features for pattern recognition by using AdaBoost method to detect distraction. The authors compared the performance achieved by both SVM and AdaBoost in estimating cognitive workload, finding that AdaBoost could achieve higher accuracy. Additionally, Miyaji et al. [156] introduced a mixed method by applying a SVM and an AdaBoost classifier for three parameters: (1) heart rate; (2) visual information (standard deviation of both gaze and head rotation angles) and (3) pupil diameter to assess the level of the driver's vigilance. Recently, a new machine learning tool, Extreme Learning Machine (ELM) [185,186]), has gained much attention due to its simple structure, high generalization capability, and fast computational speed. For example, in [187], ELM and SVM were applied to detect drivers' workload using eye movement, as well as eye movement combined with driving performance data. The results suggested that both methods can detect drivers' workload at high accuracy, but ELM outperformed SVM in most cases.

The results of all the works mentioned so far can be summarized in Table 6. Common features include the use of eye gaze-related features, driving performance, pupil diameter features and HR. It should also be noted that very good results can be obtained using only eye gaze-related features. Additionally, many supervised machine learning techniques have been proposed so far: decision trees, SVM, BN, DBN, AdaBoost or ELM.

All these distraction detection systems are based on supervised learning, meaning that the training of such systems need to be "supervised" by human experts by providing a target set for training data containing distraction status. The supervised learning paradigm is only suitable for early stage research and may not be suitable for implementation in real driving cases because of the huge cost and difficulty of creating a target distraction status set, which would require additional subjective ratings by the driver [115], post-processing by the experimentalists [56], or additional computation based on data from other sources [179]. For example, in a recent study [188], labeling drivers' distraction state involves the development of Graphical User Interface (GUI), the training of external evaluators, and the actual labeling time, which is approximately 21.5 h of manpower (43 min per one of the 30 evaluator) to label the entire video segments. For naturalistic driving, where the driver voluntarily decides which tasks to perform at any time, the labeling process can become infeasible. On the other hand, data without known distraction states (unlabeled data) can be collected readily, e.g., from drivers' naturalistic driving records.

Table 6. Supervised algorithms for cognitive distraction detection.

Algorithm	Features	Classifier	Accuracy (%)
Zhang et al. [178]	Eye gaze-related features and driving performance	Decision Tree	81
Zhang et al. [178]	Eye gaze-related features	Decision Tree	80
Zhang et al. [178]	Pupil-diameter features	Decision Tree	61
Zhang et al. [178]	Driving performance	Decision Tree	60
Liang, Reyes, et al. [179]	Eye gaze-related features and driving performance	SVM	83.15
Liang, Reyes, et al. [179]	Eye gaze-related features	SVM	81.38
Liang, Reyes, et al. [179]	driving performance	SVM	54.37
Liang, Lee, et al. [180]	Eye gaze-related features and driving performance data	DBNs	80.1
Miyaji et al. [156]	Heart rate, Eye gaze-related features and pupil diameter	AdaBoost	91.5
Miyaji et al. [156]	Eye gaze-related features	SVM	77.1 (arithmetic task)
Miyaji et al. [156]	Eye gaze-related features	SVM	84.2 (conversation task)
Miyaji et al. [156]	Eye gaze-related features	AdaBoost	81.6 (arithmetic task)
Miyaji et al. [156]	Eye gaze-related features	AdaBoost	86.1 (conversation task)
Yang et al. [187]	Eye gaze-related features and driving performance data	ELM	87.0
Yang et al. [187]	Eye gaze-related features and driving performance data	SVM	82.9

With the purpose of tackle these deficiencies, Unsupervised and Semi-Supervised algorithms can be used. For example, in [12], Semi-Supervised Extreme Learning Machine (SS-ELM) is proposed for drivers' distraction detection. SS-ELM outperformed supervised ELM in both accuracy and model sensitivity, suggesting that the proposed semi-supervised detection system can extract information from unlabeled data effectively to improve the performance. SS-ELM based detection system has the potential of improving accuracy and alleviating the cost of adapting distraction detection systems to new drivers, and thus, more promising for real world applications. However, several points are unclear from these preliminary results [12] further explored in [189], where the Semi-Supervised Learning (SSL) paradigm is introduced to real time detection of distraction based on eye and head movements.

In [189], two graph-based SSL methods were compared with supervised learning methods. These algorithms are detailed as follows. Laplacian Support Vector Machine (LapSVM), which is an extension of SVMs to SSL under manifold regularization framework [190], and SS-ELM were compared with three supervised learning methods (static BN with Supervised Clustering (SBN-SC) [180,183], ELM and SVM) and one low-density-separation-based method (Transductive SVM (TSVM) [191]). To capture realistic eye and head movements patterns, data from an on-road experiment were used. By utilizing unlabeled data, the graph-based semi-supervised methods reduced the labeling cost and improved the detection accuracy. The highest accuracy of 97.2% and G-mean of 0.959 were achieved by SS-ELM. The benefits of using SSL methods increased with the size of unlabeled data set showing that by exploring the data structure without actually labeling them, extra information to improve models performance can be obtained.

It is worth noting that cognitive distraction detection is only performed in "laboratory settings" and not in real conditions. In real life situations, when the driver is under cognitive load (e.g., mind wandering): (1) he is alone and does not interact with anybody; (2) he is also the only one who can decide whether or not to activate the attentional processing of distractive thoughts [192];

and (3) drivers are likely to be performing multiple tasks at the same time (e.g., talking on the mobile phone and listening to music). Moreover, there are two main limitations intrinsic to laboratory-based studies. First of all, most of these studies require that the execution of predefined tasks last for no more than some minutes. In our opinion, such experiments make it very difficult, if not impossible, to infer, for instance, the long-term effectiveness of for example, warning signals, monotonous driving (in general, real driving), on the basis of the results of experiments that are typically so short; And secondly, the drivers are abnormally vigilant to the driving task because they are being observed [193]. In connection with this point, the use of physiological parameters, which form the basis for cognitive distraction detection, have also been extracted in controlled settings and not in real conditions.

6. Mixing Types of Distraction

There are some algorithms trying to detect mixing types of distraction, whose results can be seen in Table 7. In [194], facial features are extracted to detect both visual and cognitive distractions. Binary classifiers (normal vs distracted) are built for visual and cognitive distraction detection. Gaze and Action Units (AU) features are useful in order to detect visual distractions, while AU features are particularly important for cognitive distractions. It should be pointed out that the cognitive tasks considered in this study are closely related to talking activities.

Table 7. Mixing types of distraction detection algorithms.

Algorithm	Features	Classifier	Average Accuracy (%)
Li et al. [194]	AU and head pose	LDC (visual distraction) and SVM (cognitive distraction)	80.8 (LDC), 73.8 (SVM)
Craye et al. [195]	eye behaviour, arm position, head orientation and facial expressions using both color and depth images	Adaboost and HMM	89.84 (Adaboost), 89.64 (HMM)
Liu et al. [196]	Head and eye movements	SVM, ELM and CR-ELM	85.65 (SVM), 85.98 (ELM), 86.95 (CR-ELM)
Ragab et al. [197]	arm position, eye closure, eye gaze, facial expressions and head orientation using depth images	Adaboost, HMM, RF, SVM, CRF, NN	82.9 (RF—type of distraction detection), 90 (RF—distraction detection)

Liu et al. [196] applied Cluster Regularized Extreme Learning Machine (CR-ELM) for detecting mixing types of distraction. Compared with the traditional ELM, CR-ELM introduces an additional regularization term penalizing large covariance of training data within the same clusters in the output space. CR-ELM, ELM and SVM were compared to predict mixing types of distraction. They simulated the mixing types of distraction by combining two types of distracting driving activities (a visual task and a cognitive one). CR-ELM showed lower error rate on most of the 11 subjects (see Table 7).

There are other approaches trying to merge both RGB and depth images to get the features to be used by the algorithms [195,197]. Craye et al. [195] extracted features from face and body using both color and depth images in order to build a distraction system, which is composed of four sub-modules: eye behaviour (gaze and blinking), arm position, head orientation and facial expressions. The information from these modules are merged together using two different classification techniques: Adaboost classifier and Hidden Markov Model (HMM). A set of video sequences was collected to test the system. Five distractive tasks were recorded and manually labelled for training and evaluation. HMM outperforms Adaboost for most drivers. Finally, a vision-based driver distraction is investigated using several machine learning techniques in [197]. IR and Kinect cameras were used in this system, where five visual cues were calculated: arm position, eye closure, eye gaze, facial expressions and

head orientation. These cues were fed into a classifier, such as Adaboost, HMM, RF, SVM, Conditional Random Field (CRF) or NN, in order to detect and recognize the type of distraction.

7. The Relationship between Facial Expressions and Distraction

Facial expressions can be described at different levels [198]. A widely used description is Facial Action Coding System (FACS) [199], which is a human-observer-based system developed to capture subtle changes in facial expressions. With FACS, these expressions are decomposed into one or more AUs [200]. AU recognition and detection have attracted much attention recently [201]. Meanwhile, psychophysical studies indicate that basic emotions have corresponding universal facial expressions across all cultures [202]. This is reflected by most current facial expression recognition systems attempting to recognize a set of prototypic emotional expressions including disgust, fear, joy, surprise, sadness and anger [201], which can be helpful in predicting driving behaviour [203].

Therefore, in this work, main facial expression works in the driving environment are described in accordance with the two aforementioned levels (FACS and prototypic emotional expressions) and how they are related with distraction.

On one hand, in connection with FACS and distraction while driving, the reference work is the one proposed by Li et al. [194]. The authors performed the analysis of driver's facial features under cognitive and visual distractions. In addition to the obvious facial movement associated with secondary tasks such as talking, they hypothesized that facial expression can play an important role in cognitive distraction detection. They studied the top five features (from a total of 186 features) to predict both cognitive and visual distraction. For cognitive distraction, the most important features to consider are: (1) head yaw; (2) Lip Corner Depressor (AU15); (3) Lip Puckerer (AU18); (4) Lip Tightener (AU23) and (5) head roll. For visual distraction, the most important features to consider are: (1) Lip Tightener (AU23); (2) jaw drop (AU26); (3) head yaw; (4) Lip Suck (AU28) and (5) Blink (AU45). The results indicated that gaze and AU features are useful for detecting visual distractions, while AU features are particularly important for cognitive distractions. It should be pointed out that since the cognitive tasks considered in this study are closely related to talking activities, their future work will include the analysis of other cognitive tasks (e.g., thinking or solving math problems).

On the other hand, in connection with prototypic emotional expressions, there are some works trying to study how these emotions affect behaviour.

The relationship between emotion and cognition is complex, but it is widely accepted that human performance is altered when a person is in any emotional state. It is really important to fully understand the impact of emotion on driving performance because, for example, roadways are lined with billboard advertisements and messages containing a lot of different emotional information. Moreover, the distracting effects of emotions may come in other forms, such as cell phone, passenger conversations, radio information or texting information [204]. For example, Chan et al. [204] conducted a study to examine the potential for distraction from emotional information presented on roadside billboards. The findings in this study showed that emotional distraction: (a) can seriously modulate attention and decision-making abilities and have adverse impacts on driving behavior for several reasons and (b) can impact driving performance by reorienting attention away from the primary driving task towards the emotional content and negatively influence the decision-making process. In another study with a similar line of work, Chan et al. [205] showed that emotion-related auditory distraction can modulate attention to differentially influence driving performance. Specifically, negative distractions reduced lateral control and slowed driving speeds compared to positive and neutral distractions.

Some studies have shown that drivers who are more likely to become angry (e.g., those with high trait anger rates) tend to engage in more aggressive behavior on the road, which can result in negative outcomes such as crashes [206]. Moreover, anger negatively influences several driving performance and risky behaviors such as infractions, lane deviations, speed, and collisions [207].

In conclusion, aggressiveness and anger are emotional states that extremely influence driving behaviour and increase the risk of accident. However, a too low level of activation (e.g., resulting from emotional states like sadness) also leads to reduced attention and distraction as well as prolonged reaction time and, therefore, lowers driving performance [208]. On this basis, research and experience have demonstrated that being in a good mood is the best precondition for safe driving and that happy drivers produce fewer accidents [209]. In other words, *happy drivers are better drivers* [208,210]. Facial expression and emotion recognition can be used in advanced car safety systems, which, on one hand, can identify hazardous emotional drivers' states that can lead to distraction and, on the other, can provide tailored (according to each state and associated hazards) suggestions and warnings to the driver [211].

8. Sensors

Once the algorithms for distraction detection have been designed and implemented, the next step is to port them to an embedded device or system to be executed inside the car. However, porting a vision-based algorithm is not a straightforward step and some key factors should be taken into account. Furthermore, there is not a standard implementation platform, so different alternatives have been proposed by both the scientific community and the industry.

8.1. Porting a Vision Algorithm to an Embedded Automotive System

The implementation of computer vision applications in automotive environments is not straightforward because several requirements must be taken into account: reliability [212,213], real-time performance [213–215], low-cost [216–219], spatial constraints [217,219], low power consumption [220], flexibility [219], rapid prototyping [215,221], design requirements [217] and short time to market [217]. Therefore, there must be a trade-off among these design requisites [217]. Moreover, there is not a commonly accepted hardware and software platform, so different solutions have been proposed by the industry and the scientific community. Last but not least, some driver distraction guidelines and test procedures for all applications to be used while driving should be considered [222], and so should ADAs.

One approach can rely on the use of microprocessors, which incorporates the functions of a computer's central processing unit (CPU) on a single integrated circuit (IC). For example, in [223], a vision-based system for monitoring the loss of attention, tested under day and night driving conditions, is proposed. The algorithm was cross-validated using brain signals and finally, implemented on a Single Board Computer (SBC). Another example is presented in [224], where a vehicle was equipped with a USB camera connected to the system in order to track the driver's eyes for fatigue detection.

A similar approach is the use of digital signal processors (DSPs) [225], which can perform multiplications and additions in a single cycle and have parallel processing capabilities. DSPs have been used in image and audio signal processing when the use of microcontrollers was not enough. These processors were used in [215], where an optimized vision library approach for embedded systems was presented. VLIB is a software library that accelerates computer vision applications for high-performance embedded systems. By significantly speeding up pixel-intensive operations, the library provides more headroom for innovative algorithms, and enables processing of more channels at higher resolutions. Authors optimized the library for the Texas Instruments C64x/C64x+ DSP cores. Karuppusamy et al. [226] proposed an embedded implementation of facial landmarks detection based on both Viola-Jones face detector and facial landmarks detection using extended Active Shape Model (ASM) [227]. However, DSPs imply a much higher cost compared with other options such as field-programmable gate arrays (FPGAs) [228].

Another option is to use hardware implementation, since it can achieve a much better computational performance, where two types are commonly used namely, FPGA and ASIC. A FPGA is an integrated circuit designed to be configured by a customer or a designer after manufacture.

FPGAs take advantage of high speed operations, especially for parallelizable operations achieving good performance in face monitoring applications [229–232]. For example, several well-known algorithms have been used and optimized for this field of application, such as: (a) spatial and temporal filtering, motion detection and optical flow analysis [229] or (b) gray scale projection, edge detection with Prewitt operator and complexity functions [230]. Additionally, the use of Application-Specific Integrated Circuits (ASIC), which is an IC customized for a particular use rather than intended for general-purpose use, has also been considered [233]. FPGAs have an important advantage over ASICs: they are reconfigurable, which gives them some of the flexibility of software. ASICs are only used for high volume manufacturing and long series due to higher initial engineering cost.

Developing the whole application in hardware is a cumbersome task, so hybrid solutions have appeared combining both software and hardware implementations. The work in [234] describes a System on a Chip (SOC) designed to support a family of vision algorithms. However, this system uses an ASIC, so it cannot be completely reconfigured. This important drawback makes impossible to update the device. A generic embedded hardware and software architecture was proposed to design and evaluate ADAS vision applications [221]. Although the system is useful to test some vision applications, the performance obtained in the case study showed that the system is not powerful enough to run more complex applications including I/Os management, vehicle communications or more demanding vision applications. In [219], a reconfigurable embedded vision system reaching the requirements of ADAS applications is presented. A SOC, which is formed by an FPGA with a dual core ARM, is prepared to be easily reconfigured. A lane departure warning system was implemented in the case study obtaining a good computational performance. The obtained computational time allows the system to include another more complex vision algorithm running in parallel. In [235], they proposed an approach to predict performances of image processing algorithms on different computing units of a given heterogeneous SOC.

Despite the fact that in recent years some authors have been trying to propose some architectures in order to achieve some key factors in embedded ADAS systems inside a vehicle [219,221,235], these efforts do not seem to be sufficient to reach the requirements stated before. The use of microprocessors in embedded computer vision-based systems has experienced a significant growth in recent years. Moreover, current premium cars implement more than 90 Electronic Control Units (ECU) with close to 1 Gigabyte embedded software code [236]. In 2018, 30% of the overall vehicle cost is predicted to stem from vehicle electronics [237]. The independence of different applications (with different criticality levels) running on the same platform must be made evident. Therefore, the development of embedded automotive systems has become quite complex. To that end, the use of standards and frameworks is indispensable.

8.2. Commercial Sensors

8.2.1. Smart Eye

The Smart Eye [77] system is a well-suited head and gaze tracking method for the demanding environment of a vehicle cabin and flexible to cope with most research projects. It consists of a multi-camera system running on a single PC and on a single algorithm. The system is scalable from 2 up to 8 cameras allowing 360° head and eye tracking. A typical configuration inside a vehicle cabin is composed of four cameras with two IR lightings, located on the dashboard on either side of the steering wheel. Smart Eye offers a sampling rate of 60 Hz (up to 8 cameras) or 120 Hz (up to 4 cameras). The field of view, depending on the number of cameras, is in the range of 90°–360°. The data output includes over 145 values covering, among others, gaze, eyelid, pupilometry and head tracking, raw and filtered gaze, blinks, fixations and saccades. Smart Eye has been used in several driver assistance and inattention systems, such as [76,143–145,238].

8.2.2. EyeAlert

EyeAlert [239], cited in several publications [128,240], has been conceived to detect driver inattention using computer vision and to generate a warning signal in case of dangerous situation. The EyeAlert system focuses entirely on the driver's alertness levels or inattention to the road ahead, regardless of the time of the day or the weather conditions. Three models are available:

- EyeAlert EA410 detects both distracted and fatigue driving. The EA410 has a highly integrated IR camera, a computer, an image processing unit and an alarm. The EA410 technology is protected by over ten patents. The system will also respond in case the driver does not focus on driving.
- EyeAlert EA430 with GPS detects both distracted and fatigue driving. Moreover, a minimum speed threshold is programmed into the internal GPS to prevent false alarms in urban environments.
- EyeAlert EA450 with Data detects both distracted and fatigue driving. Additionally, minimum speed threshold, sensitivity, volume and data can be remotely programmed. The minimum speed and sensitivity controls allow the reduction of false alarms in highway and urban environments.

8.2.3. Seeing Machines

Seeing Machines [241] builds image-processing technology that tracks the movement of a person's eyes, face, head, and facial expressions. It monitors fatigue and distraction events in real-time and uses IR technology to provide fatigue and distraction monitoring at any time of the day. The system can also combine multiple camera sensors to detect a wider range of movements. The Seeing Machines' system continuously measures operator eye and eyelid behaviour to determine the onset of fatigue and micro sleeps and delivers real-time detection and alerts. The system has been used in many different driver assistance and inattention systems [11,142,151,156,242–244].

8.2.4. Visage Technologies AB

Visage Technologies AB [245] provides a state-of-the-art commercial head tracker based on feature-point detection and tracking of the nose boundary and eye regions. Visage SDK finds and tracks the face and facial features, including gaze direction, in video sequences in real time. It provides pupil coordinates, 3D gaze direction as well as (with a calibration step) screen-space gaze point. Visage Technologies also features support for embedded systems like FPGA and IR light tracking for poor lighting conditions.

8.2.5. Delphi Electronics Driver Status Monitor

Delphi Electronics, a major automotive electronics manufacturer, developed a single camera Driver Status Monitor (DSM) [246]. By detecting and tracking the driver's facial features, the system analyzes eye-closures and head pose to infer his/her fatigue or distraction. This information is used to warn the driver and to modulate the actions of other safety systems. The system includes the use of NIR illumination, an embedded processing unit, as well as the camera (resolution of 640×480 pixels).

8.2.6. Tobii Technologies

Tobii Technologies develops Tobii's eye-tracking technology for integration into volume products such as computers, computer games, virtual reality and cars. The Tobii platform consists of two-camera sensors, placed at different angles, and operating at IR frequencies to eliminate interference from external light. The system can distinguish whether the driver's eyes are open or closed or if the driver has turned his/her head. The sensors work even when the driver is wearing glasses or sunglasses. By observing the specifics of eyelid closure, in combination with eye gaze patterns, an active safety system powered by Tobii's eye tracking sensor can reliably detect if a driver is falling asleep and warns him/her properly. Moreover, Tobii Technologies provides the Tobii EyeChip, which is a dedicated eye tracking SOC ASIC.

8.2.7. SensoMotoric Instruments

SensoMotoric Instruments GmbH (SMI) [247] is a German company, whose eye tracking solutions can measure head position and orientation, gaze direction, eyelid opening, and pupil position and diameter. Eye trackers use a sampling rate of 120 Hz for head pose and gaze measurement, 120 Hz for eyelid closure and blink measurement, and 60 Hz for combined gaze, head pose, and eyelid measurement. It also provides PERCLOS information for drowsiness detection. It is a computer-based system and needs user calibration. In [248], SensoMotoric was used to recognize the pupil in each image in order to measure horizontal and vertical eye movements.

8.2.8. Automobile Manufacturers

Nissan introduces its new Driver Attention Alert system with the 2016 Nissan Maxima, which was unveiled at the New York International Auto Show [249]. The automaker has highlighted that the new system will be able to detect drowsy and inattentive driving and it will alert the driver about the situation by giving visual and audible warning. Ford's Driver Alert [250] seems only to detect drowsiness but not distraction. The Driver Alert system comprises a small forward-facing camera connected to an on-board computer. The camera is mounted on the back of the rear view mirror and is trained to identify lane markings on both sides of the vehicle. When the vehicle is on the move, the computer looks at the road ahead and predicts where the car should be positioned relative to the lane markings. Then, it measures where the vehicle actually is and, if the difference is significant, the system issues a warning. The Saab Driver Attention Warning System [251] detects visual inattention and drowsy driving. The system uses two miniature IR cameras integrated with Smart Eye technology [77] to accurately estimate head pose, gaze, and eyelid status. When a driver's gaze is not located inside the primary attention zone (which covers the central part of the frontal windshield) for a predefined period, an alarm is triggered. Toyota has equipped their luxury Lexus models with their Driver Monitoring System [252]. The system permanently monitors the movement of the driver's head when looking from side to side using a NIR camera installed at the top of the steering wheel column. The system is integrated into Toyota's pre-crash system, which warns the driver when a collision is likely to happen. In 2007, Volvo Cars introduced Driver Alert Control to alert tired and non-concentrating drivers [253,254]. Based on the idea that the technology for monitoring eyes is not yet sufficiently mature and human behavior varies from one person to another, Volvo Cars developed the system based on car progress on the road. It is reported that Driver Alert Control monitors the car movements and assesses whether the vehicle is driven in a controlled or uncontrolled way. More recently, a Hyundai concept car (the Hyundai HCD-14) incorporates Tobii Technologies to track the eyes [255].

9. Simulated vs. Real Environment to Test and Train Driving Monitoring Systems

The development of the computer vision algorithm only represents one part of all the cycle of the product design. One of the hardest tasks is to validate the whole system with the wide variety of driving scenarios [256]. In order to complete the whole "process development" of the vision-based ADAS, some key points are presented.

In order to monitor both the driver and his/her driving behaviour, several hardware and software algorithms are being developed, but they are tested mostly in simulated environments instead of in real driving ones. This is due to the danger of testing inattention in real driving environments [21]. Experimental control, efficiency, safety, and ease of data collection are the main advantages of using simulators [257,258]. Some researches have validated that driving simulators can create driving environment relatively similar to road experiments [259–261]. However, some considerations should be taken into account since simulators can produce inconsistent, contradictory and conflicting results. For example, low-fidelity simulators may evoke unrealistic driving behavior and, therefore, produce invalid research outcomes. One common issue is that real danger and the real consequences

of actions do not occur in a driving simulator, giving rise to a false sense of safety, responsibility, or competence [262]. Moreover, simulator sickness symptoms may undermine training effectiveness and negatively affect the usability of simulators [262].

A study on distraction in both simulated and real environment was conducted in [11] and it was found out that the driver's physiological activity showed significant difference. Engstorm et al. [11] stated that physiological workload and steering activity were both higher under real driving conditions compared to simulated environments. In [257], the authors compared the impact of a narrower lane using both a simulator and real data, showing that the speed was higher in the simulated roads, consistent with other studies. In [263], controlled driving yielded more frequent and longer eye glances than the simulated driving setting, while driving errors were more common in simulated driving. In [167], the driver's heart rate changed significantly while performing the visual task in real-world driving relative to a baseline condition, suggesting that visual task performance in real driving was more stressful.

After the system is properly validated in a driver simulator, it should be validated in real conditions as well, because various factors including light variations and noise can also affect the driver's attention. The application on a real moving vehicle presents new challenges like changing backgrounds and sudden variations of lighting [264]. Moreover, a useful system should guarantee real time performance and quick adaptability to a variable set of users and to natural movements performed during driving [264]. Thus, it is necessary to make simulated environments appear more realistic [203].

To conclude, in most previous studies, independent evaluations using different equipment and conditions (mainly simulated environments) resulted in time-consuming and redundant efforts. Moreover, inconsistency in the algorithm performance metrics makes it difficult to compare algorithms. Hence, the only way to compare most algorithms and systems is the metrics provided by each author when comparing their values, but with scarce information about the used images and conditions. Public data sets covering simulated and real driving environments should be released in the near future, as stated by some authors previously [203].

10. Privacy Issues Related to Camera Sensors

Although there is a widespread agreement for intelligent vehicles to improve safety, the study of driver behaviour to design and evaluate intelligent vehicles requires large amounts of naturalistic driving data [265]. However, in current literature, there is a lack of publicly available naturalistic driving data largely due to concerns over individual privacy. It also should be noted that a real-time visual-based distraction detection system does not have to save the video stream. Therefore, privacy issues are mostly relevant in research works where video feed is collected and stored to be studied at a later stage, for example in the large naturalistic studies conducted in the US.

Typical protection of the individuals' privacy in a video sequence is commonly referred as "de-identification" [266]. Although this fact will help protect the identities of individual drivers, it impedes the purpose of sensorizing vehicles to control both drivers and their behaviour. In an ideal situation, a de-identification algorithm would protect the identity of drivers while preserving sufficient details to infer their behaviour (e.g., eye gaze, head pose or hand activity) [265].

Martin et al. [265,267] proposed the use of de-identification filters to protect the privacy of drivers while preserving sufficient details to infer their behaviour. Following this idea, a de-identification filter preserving only the mouth region can be used for monitoring yawning or talking and a de-identification filter preserving eye regions can be used for detecting fatigue or gaze direction, which is precisely proposed by Martin et al. [265,267]. More specifically, the authors implemented and compared de-identification filters made up of a combination of preserving eye regions for fine gaze estimation, superimposing head pose encoded face masks for providing spatial context and replacing background with black pixels for ensuring privacy protection. A two-part study revealed that human facial

recognition experiment had a success rate well below the chance while gaze zone estimation accuracy disclosed 65%, 71% and 85% for One-Eye, Two-Eyes and Mask with Two-Eyes, respectively.

Fernando et al. [268] proposed to use video de-identification in the automobile environment using personalized Facial Action Transfer (FAT), which has recently attracted a lot of attention in computer vision due to its diverse applications in the movie industry, computer games, and privacy protection. The goal of FAT is to “clone” the facial actions from the videos of a person (source) to another one (target) following a two-step approach. In the first step, their method transfers the shape of the source person to the target subject using the triangle-based deformation transfer method. In the second step, it generates the appearance of the target person using a personalized mapping from shape changes to appearance changes. In this approach video de-identification is used to pursue two objectives: (1) to remove person-specific facial features and (2) to preserve head pose, gaze and facial expression.

11. General Discussion and Challenges Ahead

The main visual-based approaches reviewed in this paper are summarized in Table 8 according to some key factors.

A major finding emerging from two recent research works reveals that just-driving baselines may, in fact, not be “just driving” [26,269], containing a considerable amount of cognitive activity in the form of daydreaming and lost-in-thought activity. Moreover, eye-gaze patterns are somewhat idiosyncratic when visual scanning is disrupted by cognitive workload [27]. Additionally, “look-but-failed-to-see” impairment under cognitive workload is an obvious detriment to traffic safety. For example, Strayer et al. [270] found that recognition memory for objects in the driving environment was reduced by 50% when the driver was talking on a handsfree cell phone, inducing failures of visual attention during driving. Indeed, visual, manual and cognitive distraction often occur simultaneously while driving (e.g., texting while driving and other cell-phone reading and writing activities). Therefore, the estimates of crash risk based on comparisons of activities to just-driving baselines may need to be reconsidered in light of the possible finding that just-driving baselines may contain the aforementioned frequent cognitive activity. As a result, for example, secondary tasks effects while driving should be revised [269]. Accordingly, as detecting driver distraction depends on how distraction changes his/her behavior compared to normal driving without distraction, periods with minimal or no cognitive activity should be identified in order to train the distraction detection algorithms.

Additionally, computer vision techniques can be used, not only for extracting information inside the car, but also for extracting information outside the car, such as traffic, road hazards, external conditions of the road ahead, intersections, or even position regarding other cars. The final step should be the correlation between the driver’s understanding and the traffic context. One of the first works trying to fuse “out” information (visual lane analysis) and “in” information (driver monitoring) is the one proposed by Apostoloff et al. [271], pointing out the benefits of this approach. Indeed, visual lane analysis can be used for “higher-order tasks”, which are defined by interacting with other modules in a complex driver assistance system (e.g., understanding the driver’s attentiveness—distraction—to the lane-keeping task [272]). Hirayama et al [273] focused on temporal relationships between the driver’s eye gaze and the peripheral vehicles behaviour. In particular, they concluded that the timing when a driver gazes towards the overtaking event under cognitive distraction is later than that under the neutral state. Therefore, they showed that the temporal factor, that is, timing, of a reaction is important for understanding the state by focusing on cognitive distraction in a car-driving situation. Additionally, Rezaei et al. [87] proposed a system correlating the driver’s head pose to road hazards (vehicle detection and distance estimation) by analyzing both simultaneously. Ohn et al. [274] proposed a framework for early detection of driving maneuvers using cues from the driver, the environment and the vehicle. Tawari et al. [275] provided early detection of driver distraction by continuously monitoring driver and surround traffic situation. Martin et al. [276] focused on intersections and studied the interaction of head, eyes and hands as the driver approaches

a stop-controlled intersection. In this line work of research, Jain et al. [277] deal with the problem of anticipating driving maneuvers a few seconds before the driver performs them.

Table 8. Summary of visual-based approaches to detect different types of driver distraction.

Approach	Distraction Detection Approaches			Real Conditions	Operation	
	Manual	Visual	Cognitive		Daytime	Nighttime
Zhao et al. [88]	✓	✗	✗	✗	✓	✗
Zhao et al. [89]	✓	✗	✗	✗	✓	✗
Zhao et al. [90]	✓	✗	✗	✗	✓	✗
Zhao et al. [91]	✓	✗	✗	✗	✓	✗
Bosch et al. [94]	✓	✗	✗	✗	✓	✗
Lowe et al. [95]	✓	✗	✗	✗	✓	✗
Yan et al. [92]	✓	✗	✗	✗	✓	✗
Yan et al. [93]	✓	✗	✗	✗	✓	✗
Zhang et al. [96]	✓	✗	✗	✗	✓	✗
Artan et al. [97]	✓	✗	✗	✓	✓	✓
Berri et al. [98]	✓	✗	✗	✓	✓	✗
Xu et al. [99]	✓	✗	✗	✓	✓	✓
Seshadri et al. [100]	✓	✗	✗	✓	✓	✗
Ohn et al. [106]	✓	✗	✗	✓	✓	✗
Martin et al. [104]	✓	✗	✗	✓	✓	✗
Ohn et al. [105]	✓	✗	✗	✓	✓	✗
Morimoto et al. [120]	✗	✓	✗	✓	✗	✓
Ji et al. [121]	✗	✓	✗	✓	✗	✓
Ji et al. [122]	✗	✓	✗	✓	✗	✓
Ji et al. [123]	✗	✓	✗	✓	✗	✓
Gu et al. [124]	✗	✓	✗	✓	✗	✓
Batista et al. [125]	✗	✓	✗	✓	✗	✓
Bergasa et al. [126]	✗	✓	✗	✓	✓	✓
Lee et al. [114]	✗	✓	✗	✓	✓	✓
Vicente et al. [49]	✗	✓	✗	✓	✓	✓
Cyganek et al. [134]	✗	✓	✗	✓	✓	✓
Donmez et al. [142]	✗	✓	✗	✗	✓	✗
Klauer et al. [7]	✗	✓	✗	✓	✓	✓
Kircher et al. [143]	✗	✓	✗	✓	✓	✓
Kircher et al. [144]	✗	✓	✗	✓	✓	✓
Kircher et al. [145]	✗	✓	✗	✓	✓	✓
Victor et al. [146]	✗	✓	✗	✓	✓	✗
Zhang et al. [178]	✗	✗	✓	✗	✓	✗
Liang et al. [179]	✗	✗	✓	✗	✓	✗
Liang et al. [180]	✗	✗	✓	✗	✓	✗
Liang et al. [181]	✗	✗	✓	✗	✓	✗
Liang et al. [27]	✗	✗	✓	✗	✓	✗
Liang et al. [182]	✗	✗	✓	✗	✓	✗
Liang et al. [183]	✗	✗	✓	✗	✓	✗
Miyaji et al. [184]	✗	✗	✓	✗	✓	✗
Miyaji et al. [156]	✗	✗	✓	✗	✓	✗
Yang et al. [187]	✗	✗	✓	✗	✓	✗
Li et al. [194]	✗	✓	✓	✗	✓	✗
Craye et al. [195]	✓	✓	✗	✗	✓	✗
Liu et al. [196]	✗	✓	✓	✗	✓	✗
Ragab et al. [197]	✓	✓	✗	✗	✓	✗

There are many factors that can modulate distraction. For example, as discussed in Section 7, emotional information can modulate attention and decision-making abilities. Additionally, numerous studies link highly aroused stress states with impaired decision-making capabilities [278], decreased situational awareness [279], and degraded performance, which could impair driving ability [280]. Another driver state, often responsible for traffic violations and even road accidents that can lead to distraction,

is confusion or irritation, as it is related to loss of self-control and, therefore, loss of vehicle control, which can be provoked by non-intuitive user interfaces or defective navigation systems as well as by complex traffic conditions, mistakable signs and complicated routing. Moreover, the amount of information that needs to be processed simultaneously during driving is a source of confusion especially for older people [281], who have slower perception and reaction times. Just like stress, confusion or irritation leads to impairment of driving capabilities including driver's perception, attention, decision making, and strategic planning. Nervousness corresponds to a level of arousal above the "normal" one, which best suits to the driving task [211]. It is an affective state with negative impact both on decision-making process and strategic planning. Nervousness can be induced by a variety of reasons either directly related to the driving task like novice drivers or by other factors like personal/physical conditions [211].

The system should be validated, firstly, in a driver simulator and afterwards, in real conditions, where various factors including variations in lighting and noise can also affect both the driver's attention and the performance of the developed algorithms. Therefore, public data sets covering simulated and real driving environments should be released. The driver's physiological responses could be different in a driver simulator from those in real conditions [11,167,257,263]. Hence, while developing an inattention detection system, the simulated environment must be a perfect replica of the real environment. However, they are normally used in research and simulated scenarios, but not in real ones, due to the problems of vision systems working in outdoor environments (lighting changes, sudden movements, etc.). Moreover, they do not work properly with users wearing glasses and may need high computational requirements.

Data-driven applications will require large amount of labeled images for both training and testing the system. Both manual data reduction and labeling of data are time-consuming and they are also subject to interpretation of the reductionist. Therefore, to deal with this problem, two approaches are emerging from the literature: (1) unsupervised or semi-supervised learning and (2) automatic data reduction. For example, in connection with the first approach, Liu et al. [189] commented the benefits of SSL methods. Specifically, they explained the benefits of using SSL increased with the size of unlabeled data set showing that by exploring the data structure without actually labeling them, extra information to improve models performance can be obtained. On the other hand, there has been a hype in data reduction using vehicle dynamics and looking outside on large scale naturalistic driving data [282–284], and looking in at the driver [285].

In many distraction detection systems, the use of commercial sensors is usually performed [77,239,241,245–247]. We understand that the reason from this is twofold: these systems are well-established solutions offering both head and gaze tracking in the car environment and the efforts of the investigation can be focused to detect and predict distraction from the outputs from these commercial sensors instead of developing a new sensor from the very beginning. These commercial sensors can operate using one camera [239,245–247], two cameras [241] or even up to 8 cameras [77] placed all over the vehicle cabin. What we find missing is some research works trying to compare these commercial sensors in order to highlight the pros and cons of each one. Also, missing from the literature is the comparison between a new sensor and a commercial one trying to offer a competitive solution from the sake of the research community.

Author Contributions: Alberto Fernández, Rubén Usamentiaga, Juan Luis Carús have made the review of the already existing methods and have written the paper as well. Alberto Fernández, Rubén Usamentiaga and Rubén Casado have designed paper concept and structure and have revised its content.

Conflicts of Interest: The authors declare no conflict of interest.

References

- World Health Organization. Global Status Report on Road Safety 2015. Available online: http://apps.who.int/iris/bitstream/10665/189242/1/9789241565066_eng.pdf?ua=1 (accessed on 2 July 2016).
- Llerena, L.E.; Aronow, K.V.; Macleod, J.; Bard, M.; Salzman, S.; Greene, W.; Haider, A.; Schupper, A. An evidence-based review: Distracted driver. *J. Trauma Acute Care Surg.* **2015**, *78*, 147–152.
- Carsten, O.; Merat, N. Protective or not? (visual distraction). In Proceedings of the 2015 4th International Conference on Driver Distraction and Inattention, Sydney, Australia, 9–11 November 2015.
- Dingus, T.A.; Klauer, S.; Neale, V.; Petersen, A.; Lee, S.; Sudweeks, J.; Perez, M.; Hankey, J.; Ramsey, D.; Gupta, S.; et al. *The 100-Car Naturalistic Driving Study, Phase II—Results of the 100-Car Field Experiment*; National Highway Traffic Safety Administration: Washington, DC, USA, 2006.
- Ranney, T.A.; Mazzae, E.; Garrott, R.; Goodman, M.J. NHTSA driver distraction research: Past, present, and future. In *Driver Distraction Internet Forum*; National Highway Traffic Safety Administration: Washington, DC, USA, 2000; Volume 2000.
- Klauer, S.G.; Neale, V.L.; Dingus, T.A.; Ramsey, D.; Sudweeks, J. Driver inattention: A contributing factor to crashes and near-crashes. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Orlando, FL, USA, 26–30 September 2005; SAGE Publications: Thousand Oaks, CA, USA; Volume 49, pp. 1922–1926.
- Klauer, S.G.; Dingus, T.A.; Neale, V.L.; Sudweeks, J.D.; Ramsey, D.J. *The Impact of Driver Inattention on Near-Crash/crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data*; National Highway Traffic Safety Administration: Washington, DC, USA, 2006.
- Talbot, R.; Fagerlind, H. Exploring inattention and distraction in the SafetyNet accident causation database. *Accid. Anal. Prev.* **2009**, *60*, 445–455.
- Bennakhi, A.; Safar, M. Ambient Technology in Vehicles: The Benefits and Risks. *Procedia Comput. Sci.* **2016**, *83*, 1056–1063.
- Lee, J.D. Driving safety. In *Review of Human Factors*; Nickerson, R.S., Ed.; Human Factors and Ergonomics Society: San Francisco, CA, USA, 2006; pp. 172–218.
- Engström, J.; Johansson, E.; Östlund, J. Effects of visual and cognitive load in real and simulated motorway driving. *Transp. Res. Part F Traffic Psychol. Behav.* **2005**, *8*, 97–120.
- Liu, T.; Yang, Y.; Huang, G.B.; Lin, Z. Detection of Drivers' Distraction Using Semi-Supervised Extreme Learning Machine. In *Proceedings of ELM-2014*; Springer: Berlin/Heidelberg, Germany, 2015; Volume 2, pp. 379–387.
- Simons-Morton, B.G.; Guo, F.; Klauer, S.G.; Ehsani, J.P.; Pradhan, A.K. Keep your eyes on the road: Young driver crash risk increases according to duration of distraction. *J. Adolesc. Health* **2014**, *54*, S61–S67.
- Dingus, T.A.; Neale, V.L.; Klauer, S.G.; Petersen, A.D.; Carroll, R.J. The development of a naturalistic data collection system to perform critical incident analysis: An investigation of safety and fatigue issues in long-haul trucking. *Accid. Anal. Prev.* **2006**, *38*, 1127–1136.
- Harbluk, J.L.; Noy, Y.I.; Trbovich, P.L.; Eizenman, M. An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance. *Accid. Anal. Prev.* **2007**, *39*, 372–379.
- Recarte, M.A.; Nunes, L.M. Mental workload while driving: Effects on visual search, discrimination, and decision making. *J. Exp. Psychol. Appl.* **2003**, *9*, 119–137.
- Klauer, S.G.; Guo, F.; Simons-Morton, B.G.; Ouimet, M.C.; Lee, S.E.; Dingus, T.A. Distracted driving and risk of road crashes among novice and experienced drivers. *N. Engl. J. Med.* **2014**, *370*, 54–59.
- Bergmark, R.W.; Gliklich, E.; Guo, R.; Gliklich, R.E. Texting while driving: The development and validation of the distracted driving survey and risk score among young adults. *Inj. Epidemiol.* **2016**, *3*, doi:10.1186/s40621-016-0073-8.
- Administration, N.H.T.S. Traffic Safety Facts: Distracted Driving 2014. Available online: <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812260> (accessed on 2 July 2016).
- Carmona, J.; García, F.; Martín, D.; Escalera, A.D.L.; Armingol, J.M. Data fusion for driver behaviour analysis. *Sensors* **2015**, *15*, 25968–25991.
- Sahayadhas, A.; Sundaraj, K.; Murugappan, M. Detecting driver drowsiness based on sensors: A review. *Sensors* **2012**, *12*, 16937–16953.

22. Daza, I.G.; Bergasa, L.M.; Bronte, S.; Yebes, J.J.; Almazán, J.; Arroyo, R. Fusion of optimized indicators from Advanced Driver Assistance Systems (ADAS) for driver drowsiness detection. *Sensors* **2014**, *14*, 1106–1131.
23. Hoff, J.; Grell, J.; Lohrman, N.; Stehly, C.; Stoltzfus, J.; Wainwright, G.; Hoff, W.S. Distracted driving and implications for injury prevention in adults. *J. Trauma Nurs.* **2013**, *20*, 31–34.
24. Stothart, C.; Mitchum, A.; Yehnert, C. The attentional cost of receiving a cell phone notification. *J. Exp. Psychol. Hum. Percept. Perform.* **2015**, *41*, 893–897.
25. Kushlev, K.; Proulx, J.; Dunn, E.W. Silence Your Phones: Smartphone Notifications Increase Inattention and Hyperactivity Symptoms. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 1011–1020.
26. Angell, L.; Perez, M.; Soccolich, S. *Identification of Cognitive Load in Naturalistic Driving*; Virginia Tech Transportation Institute: Blacksburg, VA, USA, 2015.
27. Liang, Y. Detecting Driver Distraction. Ph.D. Thesis, University of Iowa, Iowa City, IA, USA, May 2009.
28. Caird, J.K.; Johnston, K.A.; Willness, C.R.; Asbridge, M.; Steel, P. A meta-analysis of the effects of texting on driving. *Accid. Anal. Prev.* **2014**, *71*, 311–318.
29. Ohn-Bar, E.; Trivedi, M.M. Looking at Humans in the Age of Self-Driving and Highly Automated Vehicles. *IEEE Trans. Intell. Veh.* **2016**, *1*, 90–104.
30. Toole, L.M. Crash Risk and Mobile Device Use Based on Fatigue and Drowsiness Factors in Truck Drivers. Ph.D. Thesis, Virginia Tech, Blacksburg, VA, USA, October 2013.
31. Lee, J.D.; Young, K.L.; Regan, M.A. *Defining Driver Distraction*; CRC Press: Boca Raton, FL, USA, 2008.
32. Foley, J.P.; Young, R.; Angell, L.; Domeyer, J.E. Towards operationalizing driver distraction. In Proceedings of the 7th International Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, Bolton Landing, NY, USA, 17–20 June 2013; pp. 57–63.
33. Young, R.A. A Tabulation of Driver Distraction Definitions. Available online: <https://www.yumpu.com/en/document/view/22871401/a-tabulation-of-driver-distraction-definitions-toyota/5> (accessed on 27 October 2016).
34. Young, R. Cognitive distraction while driving: A critical review of definitions and prevalence in crashes. *SAE Int. J. Passeng. Cars Electron. Electr. Syst.* **2012**, *5*, 326–342.
35. Regan, M.A.; Hallett, C.; Gordon, C.P. Driver distraction and driver inattention: Definition, relationship and taxonomy. *Accid. Anal. Prev.* **2011**, *43*, 1771–1781.
36. Engström, J.; Monk, C.; Hanowski, R.; Horrey, W.; Lee, J.; McGehee, D.; Regan, M.; Stevens, A.; Traube, E.; Tuukkanen, M.; et al. *A Conceptual Framework and Taxonomy for Understanding and Categorizing Driver Inattention*; European Commission: Brussels, Belgium, 2013.
37. Lee, J.D. Dynamics of driver distraction: The process of engaging and disengaging. *Ann. Adv. Autom. Med.* **2014**, *58*, 24–32.
38. Viola, P.; Jones, M. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001), Kauai, HI, USA, 8–14 December 2001; Volume 1.
39. Zafeiriou, S.; Zhang, C.; Zhang, Z. A Survey on Face Detection in the wild: Past, present and future. *Comput. Vis. Image Underst.* **2015**, *138*, 1–24.
40. Viola, P.; Jones, M.J. Robust real-time face detection. *Int. J. Comput. Vis.* **2004**, *57*, 137–154.
41. Jensen, O.H. Implementing the Viola-Jones Face Detection Algorithm. Ph.D. Thesis, Technical University of Denmark, Lyngby, Denmark, September 2008.
42. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), Miami, FL, USA, 20–25 June 2005; Volume 1, pp. 886–893.
43. Dlib. Make Your Own Object Detector. Available online: <http://blog.dlib.net/2014/02/dlib-186-released-make-your-own-object.html> (accessed on 25 November 2015).
44. Ranjan, R.; Patel, V.M.; Chellappa, R. A Deep Pyramid Deformable Part Model for Face Detection. 2015, arXiv preprint arXiv:1508.04389.
45. Wang, Q.; Wu, J.; Long, C.; Li, B. P-FAD: Real-time face detection scheme on embedded smart cameras. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2013**, *3*, 210–222.

46. Markuš, N.; Friljak, M.; Pandžić, I.S.; Ahlberg, J.; Forchheimer, R. A Method for Object Detection Based on Pixel Intensity Comparisons Organized in Decision Trees. 2013, arXiv preprint arXiv:1305.4537
47. Reese, K.; Zheng, Y.; Elmaghraby, A. A comparison of face detection algorithms in visible and thermal spectrums. In Proceedings of the International Conference on Advances in Computer Science and Application, New Delhi, India, 25–27 May 2012.
48. Abtahi, S.; Omidyeganeh, M.; Shirmohammadi, S.; Hariri, B. YawDD: A yawning detection dataset. In Proceedings of the 5th ACM Multimedia Systems Conference, Singapore, 19–27 March 2014; pp. 24–28.
49. Vicente, F.; Huang, Z.; Xiong, X.; de la Torre, F.; Zhang, W.; Levi, D. Driver Gaze Tracking and Eyes Off the Road Detection System. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 2014–2027.
50. Murphy-Chutorian, E.; Trivedi, M.M. Head pose estimation in computer vision: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 607–626.
51. Tawari, A.; Martin, S.; Trivedi, M.M. Continuous Head Movement Estimator for Driver Assistance: Issues, Algorithms, and On-Road Evaluations. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 818–830.
52. Zhu, Z.; Ji, Q. Real time and non-intrusive driver fatigue monitoring. In Proceedings of the 2004 7th International IEEE Conference on Intelligent Transportation Systems, Washington, DC, USA, 3–6 October 2004; pp. 657–662.
53. Garcia-Mateos, G.; Ruiz, A.; Lopez-de Teruel, P.E.; Rodriguez, A.L.; Fernandez, L. Estimating 3D facial pose in video with just three points. In Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'08), Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.
54. Kaminski, J.Y.; Knaan, D.; Shavit, A. Single image face orientation and gaze detection. *Mach. Vis. Appl.* **2009**, *21*, 85–98.
55. Martin, S.; Tawari, A.; Murphy-Chutorian, E.; Cheng, S.Y.; Trivedi, M. On the design and evaluation of robust head pose for visual user interfaces: Algorithms, databases, and comparisons. In Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Portsmouth, NH, USA, 17–19 October 2012; pp. 149–154.
56. Oyini Mbouna, R.; Kong, S.G.; Chun, M.G. Visual analysis of eye state and head pose for driver alertness monitoring. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 1462–1469.
57. Xiong, X.; de la Torre, F. Supervised descent method and its applications to face alignment. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, OR, USA, 25–27 June 2013; pp. 532–539.
58. Asthana, A.; Marks, T.K.; Jones, M.J.; Tieu, K.H.; Rohith, M. Fully automatic pose-invariant face recognition via 3D pose normalization. In Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, 6–13 November 2011; pp. 937–944.
59. Murphy-Chutorian, E.; Trivedi, M.M. Head pose estimation and augmented reality tracking: An integrated system and evaluation for monitoring driver awareness. *IEEE Trans. Intell. Transp. Syst.* **2010**, *11*, 300–311.
60. Peláez, G.; de la Escalera, A.; Armingol, J. Head Pose Estimation Based on 2D and 3D Information. *Phys. Procedia* **2014**, *22*, 420–427.
61. Tawari, A.; Trivedi, M.M. Robust and continuous estimation of driver gaze zone by dynamic analysis of multiple face videos. In Proceedings of the 2014 IEEE Intelligent Vehicles Symposium Proceedings, Ypsilanti, MI, USA, 8–11 June 2014; pp. 344–349.
62. Ohn-Bar, E.; Tawari, A.; Martin, S.; Trivedi, M.M. Predicting driver maneuvers by learning holistic features. In Proceedings of the 2014 IEEE Intelligent Vehicles Symposium Proceedings, Ypsilanti, MI, USA, 8–11 June 2014; pp. 719–724.
63. Ohn-Bar, E.; Tawari, A.; Martin, S.; Trivedi, M.M. On surveillance for safety critical events: In-vehicle video networks for predictive driver assistance systems. *Comput. Vis. Image Underst.* **2015**, *134*, 130–140.
64. La Cascia, M.; Sclaroff, S.; Athitsos, V. Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3D models. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 322–336.
65. Zhu, X.; Ramanan, D. Face detection, pose estimation, and landmark localization in the wild. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 16–21 June 2012; pp. 2879–2886.

66. Uricár, M.; Franc, V.; Hlavác, V. Facial Landmark Tracking by Tree-based Deformable Part Model Based Detector. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Santiago, Chile, 11–18 December 2015; pp. 10–17.
67. Uricár, M.; Franc, V.; Thomas, D.; Sugimoto, A.; Hlavác, V. Real-time multi-view facial landmark detector learned by the structured output SVM. In Proceedings of the 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia, 4–8 May 2015; Volume 2, pp. 1–8.
68. Cao, X.; Wei, Y.; Wen, F.; Sun, J. Face alignment by explicit shape regression. *Int. J. Comput. Vis.* **2014**, *107*, 177–190.
69. Ren, S.; Cao, X.; Wei, Y.; Sun, J. Face alignment at 3000 fps via regressing local binary features. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1685–1692.
70. Wang, N.; Gao, X.; Tao, D.; Li, X. Facial feature point detection: A comprehensive survey. 2014, arXiv preprint arXiv:1410.1037
71. Rezaei, M.; Klette, R. Adaptive Haar-like classifier for eye status detection under non-ideal lighting conditions. In Proceedings of the 27th Conference on Image and Vision Computing New Zealand, Dunedin, New Zealand, 26–28 November 2012; pp. 521–526.
72. Song, F.; Tan, X.; Chen, S.; Zhou, Z.H. A literature survey on robust and efficient eye localization in real-life scenarios. *Pattern Recognit.* **2013**, *46*, 3157–3173.
73. Kircher, K.; Ahlström, C. *Evaluation of Methods for the Assessment of Minimum Required Attention*; Swedish National Road and Transport Research Institute (VTI): Linköping, Sweden, 2015.
74. Victor, T. Keeping Eye and Mind on the Road. Ph.D. Thesis, Uppsala University, Uppsala, Sweden, 2005.
75. Flannagan, C.A.; Bao, S.; Klinich, K.D. *Driver Distraction From Cell Phone Use and Potential for Self-Limiting Behavior*; University of Michigan: Ann Arbor, MI, USA, 2012.
76. Ahlstrom, C.; Dukic, T. Comparison of eye tracking systems with one and three cameras. In Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research, Eindhoven, The Netherlands, 24–27 August 2010.
77. SmartEye. Smart Eye Pro. Available online: <http://smarte.se/products/smart-eye-pro/> (accessed on 3 January 2016).
78. Sigari, M.H.; Pourshahabi, M.R.; Soryani, M.; Fathy, M. A Review on Driver Face Monitoring Systems for Fatigue and Distraction Detection. *Int. J. Adv. Sci. Technol.* **2014**, *64*, 73–100.
79. Zhao, S.; Grigat, R.R. Robust eye detection under active infrared illumination. In Proceedings of the 2006 IEEE 18th International Conference on Pattern Recognition (ICPR 2006), Washington, DC, USA, 20–24 August 2006; Volume 4, pp. 481–484.
80. Zhao, S.; Grigat, R.R. An automatic face recognition system in the near infrared spectrum. In *Machine Learning and Data Mining in Pattern Recognition*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 437–444.
81. Smith, P.; Shah, M.; Lobo, N.D.V. Determining driver visual attention with one camera. *IEEE Trans. Intell. Transp. Syst.* **2003**, *4*, 205–218.
82. Sigari, M.H. Driver hypo-vigilance detection based on eyelid behavior. In Proceedings of the IEEE 7th International Conference on Advances in Pattern Recognition (ICAPR'09), Kolkata, India, 4–6 February 2009; pp. 426–429.
83. Batista, J. A drowsiness and point of attention monitoring system for driver vigilance. In Proceedings of the 2007 IEEE Intelligent Transportation Systems Conference (ITSC 2007), 30 September–3 October 2007; pp. 702–708.
84. Flores, M.J.; Armingol, J.M.; de la Escalera, A. Driver drowsiness warning system using visual information for both diurnal and nocturnal illumination conditions. *EURASIP J. Adv. Signal Process.* **2010**, *2010*, 438205.
85. Rezaei, M.; Klette, R. Simultaneous analysis of driver behaviour and road condition for driver distraction detection. *Int. J. Image Data Fusion* **2011**, *2*, 217–236.
86. Craye, C.; Karray, F. Multi-distributions particle filter for eye tracking inside a vehicle. In *Image Analysis and Recognition*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 407–416.
87. Rezaei, M.; Klette, R. Look at the driver, look at the road: No distraction! No accident! In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 129–136.

88. Zhao, C.; Zhang, B.; He, J.; Lian, J. Recognition of driving postures by contourlet transform and random forests. *Intell. Transp. Syst. IET* **2012**, *6*, 161–168.
89. Zhao, C.; Gao, Y.; He, J.; Lian, J. Recognition of driving postures by multiwavelet transform and multilayer perceptron classifier. *Eng. Appl. Artif. Intell.* **2012**, *25*, 1677–1686.
90. Zhao, C.H.; Zhang, B.L.; Zhang, X.Z.; Zhao, S.Q.; Li, H.X. Recognition of driving postures by combined features and random subspace ensemble of multilayer perceptron classifiers. *Neural Comput. Appl.* **2013**, *22*, 175–184.
91. Zhao, C.; Zhang, B.; He, J. Vision-based classification of driving postures by efficient feature extraction and bayesian approach. *J. Intell. Robot. Syst.* **2013**, *72*, 483–495.
92. Yan, C.; Coenen, F.; Zhang, B. Driving posture recognition by joint application of motion history image and pyramid histogram of oriented gradients. *Int. J. Veh. Technol.* **2014**, *2014*, 719413.
93. Yan, C.; Coenen, F.; Zhang, B. Driving posture recognition by convolutional neural networks. *IET Comput. Vis.* **2016**, *10*, 103–114.
94. Bosch, A.; Zisserman, A.; Munoz, X. Representing shape with a spatial pyramid kernel. In Proceedings of the 6th ACM International Conference on Image and Video Retrieval, Hyderabad, India, 6–12 January 2007; pp. 401–408.
95. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110.
96. Zhang, X.; Zheng, N.; Wang, F.; He, Y. Visual recognition of driver hand-held cell phone use based on hidden CRF. In Proceedings of the 2011 IEEE International Conference on Vehicular Electronics and Safety (ICVES), Beijing, China, 10–12 July 2011; pp. 248–251.
97. Artan, Y.; Bulan, O.; Loce, R.P.; Paul, P. Driver cell phone usage detection from HOV/HOT NIR images. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Columbus, OH, USA, 23–28 June 2014; pp. 225–230.
98. Berri, R.A.; Silva, A.G.; Parpinelli, R.S.; Girardi, E.; Arthur, R. A Pattern Recognition System for Detecting Use of Mobile Phones While Driving. 2014, arXiv preprint arXiv:1408.0680.
99. Xu, B.; Loce, R.P. A machine learning approach for detecting cell phone usage. In Proceedings of the IS&T/SPIE Electronic Imaging, International Society for Optics and Photonics, San Francisco, CA, USA, 4 March 2015.
100. Seshadri, K.; Juefei-Xu, F.; Pal, D.K.; Savvides, M.; Thor, C.P. Driver Cell Phone Usage Detection on Strategic Highway Research Program (SHRP2) Face View Videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Boston, MA, USA, 7–12 June 2015.
101. Li, Y.; Zhou, G.; Li, Y.; Shen, D. Determining driver phone use leveraging smartphone sensors. *Multimed. Tools Appl.* **2015**, *1*, doi:10.1007/s11042-015-2969-7.
102. Ohn-Bar, E.; Trivedi, M. In-vehicle hand activity recognition using integration of regions. In Proceedings of the 2013 IEEE Intelligent Vehicles Symposium (IV), Gold Coast City, Australia, 23–26 June 2013; pp. 1034–1039.
103. Ohn-Bar, E.; Martin, S.; Trivedi, M.M. Driver hand activity analysis in naturalistic driving studies: Challenges, algorithms, and experimental studies. *J. Electron. Imaging* **2013**, *22*, 041119.
104. Martin, S.; Ohn-Bar, E.; Tawari, A.; Trivedi, M.M. Understanding head and hand activities and coordination in naturalistic driving videos. In Proceedings of the 2014 IEEE Intelligent Vehicles Symposium Proceedings, Ypsilanti, MI, USA, 8–11 June 2014; pp. 884–889.
105. Ohn-Bar, E.; Martin, S.; Tawari, A.; Trivedi, M. Head, eye, and hand patterns for driver activity recognition. In Proceedings of the IEEE 2014 22nd International Conference on Pattern Recognition (ICPR), Stockholm, Sweden, 24–28 August 2014; pp. 660–665.
106. Ohn-Bar, E.; Trivedi, M.M. The power is in your hands: 3D analysis of hand gestures in naturalistic video. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Portland, OR, USA, 23–28 June 2013; pp. 912–917.
107. Lee, S.R.; Bambach, S.; Crandall, D.J.; Franchak, J.M.; Yu, C. This hand is my hand: A probabilistic approach to hand disambiguation in egocentric video. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Columbus, OH, USA, 23–28 June 2014; pp. 557–564.
108. Drews, F.A.; Yazdani, H.; Godfrey, C.N.; Cooper, J.M.; Strayer, D.L. Text messaging during simulated driving. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2009**, *51*, 762–770.

109. Tsimhoni, O.; Arbor, A. *Time-Sharing of a Visual in-Vehicle Task While Driving: Findings from the Task Occlusion Method*; University of Michigan, Transportation Research Institute: Ann Arbor, MI, USA, 2003.
110. Victor, T.W.; Harbluk, J.L.; Engström, J.A. Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transp. Res. Part F Traffic Psychol. Behav.* **2005**, *8*, 167–190.
111. Donmez, B.; Boyle, L.N.; Lee, J.D. Safety implications of providing real-time feedback to distracted drivers. *Accid. Anal. Prev.* **2007**, *39*, 581–590.
112. Klauer, S.G.; Guo, F.; Sudweeks, J.; Dingus, T.A. *An Analysis of Driver Inattention Using a Case-Crossover Approach on 100-Car Data: Final Report*; National Highway Traffic Safety Administration: Washington, DC, USA, 2010.
113. Ahlstrom, C.; Kircher, K.; Kircher, A. A gaze-based driver distraction warning system and its effect on visual behavior. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 965–973.
114. Lee, S.J.; Jo, J.; Jung, H.G.; Park, K.R.; Kim, J. Real-time gaze estimator based on driver's head orientation for forward collision warning system. *IEEE Trans. Intell. Transp. Syst.* **2011**, *12*, 254–267.
115. Wollmer, M.; Blaschke, C.; Schindl, T.; Schuller, B.; Farber, B.; Mayer, S.; Trefflich, B. Online driver distraction detection using long short-term memory. *IEEE Trans. Intell. Transp. Syst.* **2011**, *12*, 574–582.
116. Neale, V.L.; Dingus, T.A.; Klauer, S.G.; Sudweeks, J.; Goodman, M. *An Overview of the 100-Car Naturalistic Study and Findings*; National Highway Traffic Safety Administration: Washington, DC, USA, 2005.
117. Boyraz, P.; Yang, X.; Hansen, J.H. Computer vision systems for context-aware active vehicle safety and driver assistance. In *Digital Signal Processing for in-Vehicle Systems and Safety*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 217–227.
118. Hammoud, R.I.; Wilhelm, A.; Malawey, P.; Witt, G.J. Efficient real-time algorithms for eye state and head pose tracking in advanced driver support systems. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), San Diego, CA, USA, 20–25 June 2005; Volume 2.
119. Tivesten, E.; Dozza, M. Driving context and visual-manual phone tasks influence glance behavior in naturalistic driving. *Transp. Res. Part F Traffic Psychol. Behav.* **2014**, *26*, 258–272.
120. Morimoto, C.H.; Koons, D.; Amir, A.; Flickner, M. Pupil detection and tracking using multiple light sources. *Image Vis. Comput.* **2000**, *18*, 331–335.
121. Ji, Q.; Yang, X. Real time visual cues extraction for monitoring driver vigilance. In *Computer Vision Systems*; Springer: Berlin/Heidelberg, Germany, 2001; pp. 107–124.
122. Ji, Q.; Yang, X. Real-time eye, gaze, and face pose tracking for monitoring driver vigilance. *Real Time Imaging* **2002**, *8*, 357–377.
123. Ji, Q. 3D Face pose estimation and tracking from a monocular camera. *Image Vis. Comput.* **2002**, *20*, 499–511.
124. Gu, H.; Ji, Q.; Zhu, Z. Active facial tracking for fatigue detection. In Proceedings of the 6th IEEE Workshop on Applications of Computer Vision (WACV 2002), Orlando, FL, USA, 3–4 December 2002; pp. 137–142.
125. Batista, J.P. A real-time driver visual attention monitoring system. In *Pattern Recognition and Image Analysis*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 200–208.
126. Bergasa, L.M.; Nuevo, J.; Sotelo, M.; Barea, R.; Lopez, M.E. Real-time system for monitoring driver vigilance. *IEEE Trans. Intell. Transp. Syst.* **2006**, *7*, 63–77.
127. Hansen, D.W.; Ji, Q. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 478–500.
128. D'Orazio, T.; Leo, M.; Guaragnella, C.; Distanto, A. A visual approach for driver inattention detection. *Pattern Recognit.* **2007**, *40*, 2341–2355.
129. Nuevo, J.; Bergasa, L.M.; Sotelo, M.; Ocaña, M. Real-time robust face tracking for driver monitoring. In Proceedings of the IEEE Intelligent Transportation Systems Conference (ITSC'06), Toronto, ON, Canada, 17–20 September 2006; pp. 1346–1351.
130. Bergasa, L.M.; Buenaposada, J.M.; Nuevo, J.; Jimenez, P.; Baumela, L. Analysing driver's attention level using computer vision. In Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems (ITSC 2008), Beijing, China, 12–15 October 2008; pp. 1149–1154.
131. Flores, M.J.; Armingol, J.M.; De la Escalera, A. Driver drowsiness detection system under infrared illumination for an intelligent vehicle. *IET Intell. Transp. Syst.* **2011**, *5*, 241–251.
132. Jimenez-Pinto, J.; Torres-Torriti, M. Face salient points and eyes tracking for robust drowsiness detection. *Robotica* **2012**, *30*, 731–741.

133. Jiménez-Pinto, J.; Torres-Torriti, M. Optical flow and driver's kinematics analysis for state of alert sensing. *Sensors* **2013**, *13*, 4225–4257.
134. Cyganek, B.; Gruszczyński, S. Hybrid computer vision system for drivers' eye recognition and fatigue monitoring. *Neurocomputing* **2014**, *126*, 78–94.
135. Murphy-Chutorian, E.; Doshi, A.; Trivedi, M.M. Head pose estimation for driver assistance systems: A robust algorithm and experimental evaluation. In Proceedings of the 2007 IEEE Intelligent Transportation Systems Conference (ITSC 2007), Bellevue, WA, USA, 30 September–3 October 2007; pp. 709–714.
136. Hattori, A.; Tokoro, S.; Miyashita, M.; Tanaka, I.; Ohue, K.; Uozumi, S. Development of Forward Collision Warning System Using the Driver Behavioral Information. Available online: <http://papers.sae.org/2006-01-1462/> (accessed on 26 October 2016).
137. You, C.W.; Lane, N.D.; Chen, F.; Wang, R.; Chen, Z.; Bao, T.J.; Montes-de Oca, M.; Cheng, Y.; Lin, M.; Torresani, L.; et al. Carsafe app: Alerting drowsy and distracted drivers using dual cameras on smartphones. In Proceedings of the 11th Annual International Conference on Mobile Systems, Applications, and Services, Taipei, Taiwan, 25–28 June 2013; pp. 13–26.
138. Flores, M.J.; Armingol, J.M.; de la Escalera, A. Real-time warning system for driver drowsiness detection using visual information. *J. Intell. Robot. Syst.* **2010**, *59*, 103–125.
139. Yuying, J.; Yazhen, W.; Haitao, X. A surveillance method for driver's fatigue and distraction based on machine vision. In Proceedings of the 2011 International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE), Changchun, China, 16–18 December 2011; pp. 727–730.
140. Pohl, J.; Birk, W.; Westervall, L. A driver-distraction-based lane-keeping assistance system. *Proc. Inst. Mech. Eng. Part I J. Syst. Control Eng.* **2007**, *221*, 541–552.
141. Lee, J.; Moeckli, J.; Brown, T.; Roberts, S.; Victor, T.; Marshall, D.; Schwarz, C.; Nadler, E. Detection of driver distraction using vision-based algorithms. In Proceedings of the 23rd International Conference on Enhanced Safety of Vehicles, Seoul, Korea, 27–30 May 2013.
142. Donmez, B.; Boyle, L.N.; Lee, J.D. Mitigating driver distraction with retrospective and concurrent feedback. *Accid. Anal. Prev.* **2008**, *40*, 776–786.
143. Kircher, K.; Kircher, A.; Ahlström, C. *Results of a Field Study on a Driver Distraction Warning System*; Swedish National Road and Transport Research Institute (VTI): Linköping, Sweden, 2009.
144. Kircher, K.; Ahlstrom, C.; Kircher, A. Comparison of two eye-gaze based real-time driver distraction detection algorithms in a small-scale field operational test. In Proceedings of the 5th International Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Big Sky, MT, USA, 22–25 June 2009; pp. 16–23.
145. Kircher, K.; Kircher, A.; Claezon, F. *Distraction and Drowsiness—A Field Study*; Swedish National Road and Transport Research Institute (VTI): Linköping, Sweden, 2009.
146. Victor, T. The Victor and Larsson (2010) distraction detection algorithm and warning strategy. *Volvo Technol.* **2010**, *1*, 0–6.
147. Angell, L.S.; Auflick, J.; Austria, P.; Kochhar, D.S.; Tijerina, L.; Biever, W.; Diptiman, T.; Hogsett, J.; Kiger, S. *Driver Workload Metrics Task 2 Final Report*; National Highway Traffic Safety Administration: Washington, DC, USA, 2006.
148. Recarte, M.A.; Nunes, L.M. Effects of verbal and spatial-imagery tasks on eye fixations while driving. *J. Exp. Psychol. Appl.* **2000**, *6*, 31.
149. Harbluk, J.L.; Noy, Y.I.; Eizenman, M. *The Impact of Cognitive Distraction on Driver Visual Behaviour and Vehicle Control*; Transport Canada: Ottawa, ON, Canada, 2002.
150. May, J.G.; Kennedy, R.S.; Williams, M.C.; Dunlap, W.P.; Brannan, J.R. Eye movement indices of mental workload. *Acta Psychol.* **1990**, *75*, 75–89.
151. Miyaji, M.; Kawanaka, H.; Oguri, K. Driver's cognitive distraction detection using physiological features by the adaboost. In Proceedings of the IEEE 12th International IEEE Conference on Intelligent Transportation Systems (ITSC'09), St. Louis, MO, USA, 4–7 October 2009; pp. 1–6.
152. Yang, Y. The Effects of Increased Workload on Driving Performance and Visual Behaviour. Ph.D. Thesis, University of Southampton, Southampton, UK, December 2011.
153. He, J.; Becic, E.; Lee, Y.C.; McCarley, J.S. Mind wandering behind the wheel performance and oculomotor correlates. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2011**, *53*, 13–21.

154. He, J. Identify Mind-Wandering Behind the Wheel. Ph.D. Thesis, University of Illinois at Urbana-Champaign, Champaign, IL, USA, December 2010.
155. Kahneman, D.; Tursky, B.; Shapiro, D.; Crider, A. Pupillary, heart rate, and skin resistance changes during a mental task. *J. Exp. Psychol.* **1969**, *79*, 164–167.
156. Miyaji, M.; Kawanaka, H.; Oguri, K. Effect of pattern recognition features on detection for driver's cognitive distraction. In Proceedings of the 2010 13th International IEEE Conference on Intelligent Transportation Systems (ITSC), Funchal, Portugal, 19–22 September 2010; pp. 605–610.
157. Heeman, P.A.; Meshorer, T.; Kun, A.L.; Palinko, O.; Medenica, Z. Estimating cognitive load using pupil diameter during a spoken dialogue task. In Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Eindhoven, The Netherlands, 27–30 October 2013; pp. 242–245.
158. Pfleging, B.; Fekety, D.K.; Schmidt, A.; Kun, A.L. A Model Relating Pupil Diameter to Mental Workload and Lighting Conditions. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 5776–5788.
159. Mehler, B.; Reimer, B.; Coughlin, J.F. Physiological reactivity to graded levels of cognitive workload across three age groups: An on-road evaluation. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, San Francisco, CA, USA, 27 September–1 October 2010; SAGE Publications: Thousand Oaks, CA, USA; Volume 54, pp. 2062–2066.
160. Partin, D.L.; Sultan, M.F.; Thrush, C.M.; Prieto, R.; Wagner, S.J. Monitoring Driver Physiological Parameters for Improved Safety. Available online: <http://papers.sae.org/2006-01-1322/> (accessed on 26 October 2016).
161. Liu, S.H.; Lin, C.T.; Chao, W.H. The short-time fractal scaling of heart rate variability to estimate the mental stress of driver. In Proceedings of the 2004 IEEE International Conference on Networking, Sensing and Control, Taipei, Taiwan, 21–23 March 2004; Volume 2, pp. 829–833.
162. Apparies, R.J.; Riniolo, T.C.; Porges, S.W. A psychophysiological investigation of the effects of driving longer-combination vehicles. *Ergonomics* **1998**, *41*, 581–592.
163. Wickens, C.D.; Lee, J.D.; Liu, Y.; Gordon-Becker, S. *Introduction to Human Factors Engineering*; Pearson Education: Upper Saddle River, NJ, USA, 1998.
164. Mulder, L. Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biol. Psychol.* **1992**, *34*, 205–236.
165. Miller, E.E. Effects of Roadway on Driver Stress: An On-Road Study using Physiological Measures. Ph.D. Thesis, University of Washington, Seattle, WA, USA, July 2013.
166. Zhao, C.; Zhao, M.; Liu, J.; Zheng, C. Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator. *Accid. Anal. Prev.* **2012**, *45*, 83–90.
167. Östlund, J.; Nilsson, L.; Törnros, J.; Forsman, Å. Effects of cognitive and visual load in real and simulated driving. *Transp. Res. Part F Traffic Psychol. Behav.* **2006**, *8*, 97–120.
168. Poh, M.Z.; McDuff, D.J.; Picard, R.W. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt. Express* **2010**, *18*, 10762–10774.
169. Poh, M.Z.; McDuff, D.J.; Picard, R.W. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 7–11.
170. Fernandez, A.; Carus, J.L.; Usamentiaga, R.; Alvarez, E.; Casado, R. Unobtrusive health monitoring system using video-based physiological information and activity measurements. In Proceedings of the IEEE 2015 International Conference on Computer, Information and Telecommunication Systems (CITS), Paris, France, 11–13 June 2015; pp. 1–5.
171. Wu, H.Y.; Rubinstein, M.; Shih, E.; Guttag, J.V.; Durand, F.; Freeman, W.T. Eulerian video magnification for revealing subtle changes in the world. *ACM Trans. Graph.* **2012**, *31*, 65.
172. Chambino, P. Android-Based Implementation of Eulerian Video Magnification for Vital Signs Monitoring. Ph.D. Thesis, Faculdade de Engenharia da Universidade do Porto, Porto, Portugal, July 2013.
173. Balakrishnan, G.; Durand, F.; Guttag, J. Detecting pulse from head motions in video. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Washington, DC, USA, 23–28 June 2013; pp. 3430–3437.
174. Li, G.; Chung, W.Y. Detection of driver drowsiness using wavelet analysis of heart rate variability and a support vector machine classifier. *Sensors* **2013**, *13*, 16494–16511.

175. Solovey, E.T.; Zec, M.; Garcia Perez, E.A.; Reimer, B.; Mehler, B. Classifying driver workload using physiological and driving performance data: Two field studies. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Toronto, ON, Canada, 26 April–1 May 2014; pp. 4057–4066.
176. Lee, D.; Kim, J.; Kwon, S.; Park, K. Heart rate estimation from facial photoplethysmography during dynamic illuminance changes. In Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milano, Italy, 25–29 August 2015; pp. 2758–2761.
177. Jeanne, V.; Asselman, M.; den Brinker, B.; Bulut, M. Camera-based heart rate monitoring in highly dynamic light conditions. In Proceedings of the 2013 IEEE International Conference on Connected Vehicles and Expo (ICCVE), Las Vegas, NV, USA, 2–6 December 2013; pp. 798–799.
178. Zhang, Y.; Owechko, Y.; Zhang, J. Driver cognitive workload estimation: A data-driven perspective. In Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems, Washington, DC, USA, 3–6 October 2004; pp. 642–647.
179. Liang, Y.; Reyes, M.L.; Lee, J.D. Real-time detection of driver cognitive distraction using support vector machines. *IEEE Trans. Intell. Transp. Syst.* **2007**, *8*, 340–350.
180. Liang, Y.; Lee, J.D.; Reyes, M.L. Nonintrusive detection of driver cognitive distraction in real time using Bayesian networks. *Transp. Res. Rec. J. Transp. Res. Board* **2007**, *2018*, 1–8.
181. Liang, Y.; Lee, J.D. Comparing Support Vector Machines (SVMs) and Bayesian Networks (BNs) in detecting driver cognitive distraction using eye movements. In *Passive Eye Monitoring: Algorithms, Applications and Experiments*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 285–300.
182. Liang, Y.; Lee, J.D. Using a Layered Algorithm to Detect Driver Cognitive Distraction. In Proceedings of the seventh International Driving Symposium on Human Factors in Driver assessment, Training, and Vehicle Design, New York, NY, USA, 17–20 June 2013; pp. 327–333.
183. Liang, Y.; Lee, J.D. A hybrid Bayesian Network approach to detect driver cognitive distraction. *Transp. Res. Part C Emerg. Technol.* **2014**, *38*, 146–155.
184. Miyaji, M.; Danno, M.; Kawanaka, H.; Oguri, K. Driver's cognitive distraction detection using AdaBoost on pattern recognition basis. In Proceedings of the IEEE International Conference on Vehicular Electronics and Safety (ICVES 2008), Columbus Ohio, OH, USA, 22–24 September 2008; pp. 51–56.
185. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501.
186. Huang, G.B.; Zhou, H.; Ding, X.; Zhang, R. Extreme learning machine for regression and multiclass classification. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2012**, *42*, 513–529.
187. Yang, Y.; Sun, H.; Liu, T.; Huang, G.B.; Sourina, O. Driver Workload Detection in On-Road Driving Environment Using Machine Learning. In *Proceedings of ELM-2014*; Springer: Berlin/Heidelberg, Germany, 2015; Volume 2, pp. 389–398.
188. Li, N.; Busso, C. Predicting Perceived Visual and Cognitive Distractions of Drivers With Multimodal Features. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 51–65.
189. Liu, T.; Yang, Y.; Huang, G.B.; Yeo, Y.; Lin, Z. Driver Distraction Detection Using Semi-Supervised Machine Learning. *IEEE Trans. Intell. Transp. Syst.* **2015**, *17*, 1–13.
190. Belkin, M.; Niyogi, P.; Sindhvani, V. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *J. Mach. Learn. Res.* **2006**, *7*, 2399–2434.
191. Joachims, T. Transductive inference for text classification using support vector machines. In Proceedings of the Sixteenth International Conference on Machine Learning (ICML), Bled, Slovenia, 27–30 June 1999; Volume 99, pp. 200–209.
192. Lemerrier, C.; Pêcher, C.; Berthié, G.; Valéry, B.; Vidal, V.; Paubel, P.V.; Cour, M.; Fort, A.; Galéra, C.; Gabaude, C.; et al. Inattention behind the wheel: How factual internal thoughts impact attentional control while driving. *Saf. Sci.* **2014**, *62*, 279–285.
193. Ho, C.; Gray, R.; Spence, C. To what extent do the findings of laboratory-based spatial attention research apply to the real-world setting of driving? *IEEE Trans. Hum. Mach. Syst.* **2014**, *44*, 524–530.
194. Li, N.; Busso, C. Analysis of facial features of drivers under cognitive and visual distractions. In Proceedings of the 2013 IEEE International Conference on Multimedia and Expo (ICME), San Jose, CA, USA, 15–19 July 2013; pp. 1–6.
195. Craye, C.; Karray, F. Driver distraction detection and recognition using RGB-D sensor. 2015, arXiv preprint arXiv:1502.00250.

196. Liu, T.; Yang, Y.; Huang, G.B.; Lin, Z.; Klanner, F.; Denk, C.; Rasshofer, R.H. Cluster Regularized Extreme Learning Machine for Detecting Mixed-Type Distraction in Driving. In Proceedings of the 2015 IEEE 18th International Conference on Intelligent Transportation Systems (ITSC), Las Palmas de Gran Canaria, Spain, 15–18 September 2015; pp. 1323–1326.
197. Ragab, A.; Craye, C.; Kamel, M.S.; Karray, F. A Visual-Based Driver Distraction Recognition and Detection Using Random Forest. In *Image Analysis and Recognition*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 256–265.
198. Tian, Y.L.; Kanade, T.; Cohn, J.F. Handbook of face recognition. In *Ch Facial Expression Analysis*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 487–519.
199. Ekman, P.; Friesen, W. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*; Consulting Psychologists: San Francisco, CA, USA, 1978.
200. Tian, Y.; Kanade, T.; Cohn, J.F. Facial expression recognition. In *Handbook of Face Recognition*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 487–519.
201. Shan, C.; Gong, S.; McOwan, P.W. Facial expression recognition based on local binary patterns: A comprehensive study. *Image Vis. Comput.* **2009**, *27*, 803–816.
202. Ekman, P.; Friesen, W.V.; Press, C.P. *Pictures of Facial Affect*; Consulting Psychologists Press: Palo Alto, CA, USA, 1975.
203. Kang, H.B. Various approaches for driver and driving behavior monitoring: A review. In Proceedings of the 2013 IEEE International Conference on Computer Vision Workshops (ICCVW), Sydney, Australia, 1–8 December 2013; pp. 616–623.
204. Chan, M.; Singhal, A. The emotional side of cognitive distraction: Implications for road safety. *Accid. Anal. Prev.* **2013**, *50*, 147–154.
205. Chan, M.; Singhal, A. Emotion matters: Implications for distracted driving. *Saf. Sci.* **2015**, *72*, 302–309.
206. Deffenbacher, J.L.; Lynch, R.S.; Filetti, L.B.; Dahlen, E.R.; Oetting, E.R. Anger, aggression, risky behavior, and crash-related outcomes in three groups of drivers. *Behav. Res. Ther.* **2003**, *41*, 333–349.
207. Jeon, M.; Walker, B.N.; Gable, T.M. Anger effects on driver situation awareness and driving performance. *Presence Teleoper. Virtual Environ.* **2014**, *23*, 71–89.
208. Eyben, F.; Wöllmer, M.; Poitschke, T.; Schuller, B.; Blaschke, C.; Färber, B.; Nguyen-Thien, N. Emotion on the road—Necessity, acceptance, and feasibility of affective computing in the car. *Adv. Hum. Comput. Interact.* **2010**, *2010*, 263593.
209. James, L. *Road Rage and Aggressive Driving: Steering Clear of Highway Warfare*; Prometheus Books: Amherst, NY, USA, 2000.
210. Grimm, M.; Kroschel, K.; Harris, H.; Nass, C.; Schuller, B.; Rigoll, G.; Moosmayr, T. On the necessity and feasibility of detecting a driver's emotional state while driving. In *Affective Computing and Intelligent Interaction*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 126–138.
211. Katsis, C.D.; Rigas, G.; Goletsis, Y.; Fotiadis, D.I. Emotion Recognition in Car Industry. In *Emotion Recognition: A Pattern Analysis Approach*; Wiley: New York, NY, USA, 2015; pp. 515–544.
212. Stein, F. The challenge of putting vision algorithms into a car. In Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Providence, RI, USA, 16–21 June 2012; pp. 89–94.
213. Nieto Doncel, M.; Arróspide Laborda, J.; Salgado Álvarez de Sotomayor, L.; García Santos, N. Video-Based Driver Assistance Systems. Ph.D. Thesis, Robert Bosch SRL, Cluj Napoca, Romania, October 2008.
214. Kim, K.; Choi, K. SoC Architecture for Automobile Vision System. In *Algorithm & SoC Design for Automotive Vision Systems*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 163–195.
215. Dedeoğlu, G.; Kisaçanin, B.; Moore, D.; Sharma, V.; Miller, A. An optimized vision library approach for embedded systems. In Proceedings of the 2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Colorado Springs, CO, USA, 20–25 June 2011; pp. 8–13.
216. Satzoda, R.K.; Lee, S.; Lu, F.; Trivedi, M.M. Snap-DAS: A Vision-based Driver Assistance System on a Snapdragon TM Embedded Platform. In Proceedings of the 2015 IEEE Intelligent Vehicles Symposium (IV), Seoul, Korea, 28 June–1 July 2015.
217. Nieto, M.; Otaegui, O.; Vélez, G.; Ortega, J.D.; Cortés, A. On creating vision-based advanced driver assistance systems. *IET Intell. Transp. Syst.* **2014**, *9*, 59–66.

218. Pelaez, C.; Garcia, F.; de la Escalera, A.; Armingol, J. Driver Monitoring Based on Low-Cost 3-D Sensors. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 1855–1860.
219. Velez, G.; Cortés, A.; Nieto, M.; Vélez, I.; Otaegui, O. A reconfigurable embedded vision system for advanced driver assistance. *J. Real Time Image Process.* **2014**, *10*, 1–15.
220. Forster, F. Heterogeneous Processors for Advanced Driver Assistance Systems. *ATZelektronik Worldw.* **2014**, *9*, 14–18.
221. Anders, J.; Mefenza, M.; Bobda, C.; Yonga, F.; Aklah, Z.; Gunn, K. A hardware/software prototyping system for driving assistance investigations. *J. Real Time Image Process.* **2013**, *11*, 1–11.
222. Young, R.; Zhang, J. Safe Interaction for Drivers: A Review of Driver Distraction Guidelines and Design Implications. Available online: <http://papers.sae.org/2015-01-1384/> (accessed on 26 October 2016).
223. Dasgupta, A.; George, A.; Happy, S.; Routray, A. A vision-based system for monitoring the loss of attention in automotive drivers. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 1825–1838.
224. Krishnasree, V.; Balaji, N.; Rao, P.S. A Real Time Improved Driver Fatigue Monitoring System. *WSEAS Trans. Signal Process.* **2014**, *10*, 146.
225. Veeraraghavan, H.; Papanikolopoulos, N.P. *Detecting Driver Fatigue Through the Use of Advanced Face Monitoring Techniques*; University of Minnesota: Minneapolis, MN, USA, 2001.
226. Karuppusamy, S.; Jerome, J.; Shankar, N. Embedded implementation of facial landmarks detection using extended active shape model approach. In Proceedings of the IEEE 2014 International Conference on Embedded Systems (ICES), Coimbatore, India, 3–5 July 2014; pp. 265–270.
227. De Marsico, M.; Nappi, M.; Riccio, D.; Wechsler, H. Robust face recognition for uncontrolled pose and illumination changes. *IEEE Trans. Syst. Man Cybern. Syst.* **2013**, *43*, 149–163.
228. Malinowski, A.; Yu, H. Comparison of embedded system design for industrial applications. *IEEE Trans. Ind. Inform.* **2011**, *7*, 244–254.
229. Moreno, F.; Aparicio, F.; Hernández, W.; Paez, J. A low-cost real-time FPGA solution for driver drowsiness detection. In Proceedings of the 29th Annual Conference of the IEEE Industrial Electronics Society, Roanoke, VA, USA, 2–6 June 2003; Volume 2, pp. 1396–1401.
230. Wang, F.; Qin, H. A FPGA based driver drowsiness detecting system. In Proceedings of the IEEE International Conference on Vehicular Electronics and Safety, Xian, China, 14–16 October 2005; pp. 358–363.
231. Sanz, R.; Salvador, R.; Alarcon, J.; Moreno, F.; López, I. Embedded Intelligence on Chip: Some FPGA-Based Design Experiences. Available online: <http://www.intechopen.com/books/pattern-recognition-recent-advances/embedded-intelligence-on-chip-some-fpgabased-design-experiences> (accessed on 26 October 2016).
232. Samarawickrama, M.; Pasqual, A.; Rodrigo, R. FPGA-based compact and flexible architecture for real-time embedded vision systems. In Proceedings of the 2009 International Conference on Industrial and Information Systems (ICIIS), Peradeniya, Sri Lanka, 28–31 December 2009; pp. 337–342.
233. Mielke, M.; Schafer, A.; Bruck, R. Asic implementation of a gaussian pyramid for use in autonomous mobile robotics. In Proceedings of the 2011 IEEE 54th International Midwest Symposium on Circuits and Systems (MWSCAS), Seoul, Korea, 7–10 August 2011; pp. 1–4.
234. Stein, G.P.; Rushinek, E.; Hayun, G.; Shashua, A. A computer vision system on a chip: A case study from the automotive domain. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, Miami, FL, USA, 20–25 June 2005; p. 130.
235. Saussard, R.; Bouzid, B.; Vasiliu, M.; Reynaud, R. Towards an Automatic Prediction of Image Processing Algorithms Performances on Embedded Heterogeneous Architectures. In Proceedings of the 2015 44th International Conference on Parallel Processing Workshops (ICPPW), Beijing, China, 1–4 September 2015; pp. 27–36.
236. Ebert, C.; Jones, C. Embedded software: Facts, figures, and future. *Computer* **2009**, *42*, 42–52.
237. Macher, G.; Stolz, M.; Armengaud, E.; Kreiner, C. Filling the gap between automotive systems, safety, and software engineering. *E I Elektrotech. Informationstech.* **2015**, *132*, 1–7.
238. Jin, L.; Niu, Q.; Jiang, Y.; Xian, H.; Qin, Y.; Xu, M. Driver sleepiness detection system based on eye movements variables. *Adv. Mech. Eng.* **2013**, *5*, 648431.
239. Lumeway. EyeAlert Distracted Driving and Fatigue Warning Systems. Available online: <http://www.lumeway.com/EA.htm> (accessed on 3 January 2016).

240. Craye, C.; Rashwan, A.; Kamel, M.S.; Karray, F. A Multi-Modal Driver Fatigue and Distraction Assessment System. *Int. J. Intell. Transp. Syst. Res.* **2015**, *14*, 1–22.
241. SeeingMachines. Advanced Driver Fatigue and Distraction Detection. Available online: <http://www.seeingmachines.com/> (accessed on 6 January 2016).
242. Fletcher, L.; Apostoloff, N.; Petersson, L.; Zelinsky, A. Vision in and out of vehicles. *IEEE Intell. Syst.* **2003**, *18*, 12–17.
243. Fletcher, L.; Loy, G.; Barnes, N.; Zelinsky, A. Correlating driver gaze with the road scene for driver assistance systems. *Robot. Auton. Syst.* **2005**, *52*, 71–84.
244. Friedrichs, F.; Yang, B. Camera-based drowsiness reference for driver state classification under real driving conditions. In Proceedings of the 2010 IEEE Intelligent Vehicles Symposium (IV), San Diego, CA, USA, 21–24 June 2010; pp. 101–106.
245. VisageTechnologies. Face Tracking and Analysis. Available online: <http://www.visagetechnologies.com/> (accessed on 6 January 2016).
246. Edenborough, N.; Hammoud, R.; Harbach, A.; Ingold, A.; Kisaćanin, B.; Malawey, P.; Newman, T.; Scharenbroch, G.; Skiver, S.; Smith, M.; et al. Driver state monitor from delphi. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2005; Volume 2, pp. 1206–1207.
247. SMI. Eye and gaze tracking systems. Available online: www.smivision.com (accessed on 6 January 2016).
248. Juhola, M.; Aalto, H.; Joutsijoki, H.; Hirvonen, T.P. The classification of valid and invalid beats of three-dimensional nystagmus eye movement signals using machine learning methods. *Adv. Artif. Neural Syst.* **2013**, *2013*, doi:10.1155/2013/972412.
249. Nissan. Nissan Introduces Driver Attention Alert for Passenger Safety. Available online: <http://automobiletechnology.automotive-business-review.com/news/nissan-introduces-driver-attention-alert-for-passenger-safety-020415-4546074> (accessed on 6 March 2016).
250. Ford. Ford Technology: Driver Alert, Lane Departure Warning. Available online: <http://technology.fordmedia.eu/documents/newsletter/FordTechnologyNewsletter082010.pdf> (accessed on 6 March 2016).
251. Nabo, A. Driver attention-dealing with drowsiness and distraction. Available online: <http://smarteye.se/wp-content/uploads/2015/01/Nabo-Arne-IVSS-Report.pdf> (accessed on 26 October 2016).
252. Ishiguro, H.; Hayashi, T.; Naito, T.; Kasugai, J.; Ogawa, K.; Ohue, K.; Uozumi, S. Development of facial-direction detection sensor. In Proceedings of the 13th Its World Congress, London, UK, 8–12 October 2006.
253. Volvo. Volvo Cars Introduces New Systems for Alerting Tired And Unconcentrated Drivers. Available online: <http://www.mobileye.com/wp-content/uploads/2011/09/Volvo.Driver.Alert.pdf> (accessed on 6 January 2016).
254. MobileEye. MobileEye Advanced Vehicle Technologies Power Volvo Car's Driver Alert Control (DAC) System. Available online: <http://www.mobileye.com/wp-content/uploads/2011/09/MobileEyeAdvanceVehicleTechnologiesPowerVolvo.pdf> (accessed on 6 January 2016).
255. SAE. Hyundai HCD-14 Genesis Concept Previews Eye-Tracking, Gesture-Recognition Technologies. Available online: <http://articles.sae.org/11727/> (accessed on 17 January 2016).
256. Velez, G.; Otaegui, O. Embedded Platforms for Computer Vision-based Advanced Driver Assistance Systems: A Survey. 2015, arXiv preprint arXiv:1504.07442.
257. Rosey, F.; Auberlet, J.M.; Moisan, O.; Dupré, G. Impact of Narrower Lane Width. *Transp. Res. Rec. J. Transp. Res. Board* **2009**, *2138*, 112–119.
258. Konstantopoulos, P.; Chapman, P.; Crundall, D. Driver's visual attention as a function of driving experience and visibility. Using a driving simulator to explore drivers' eye movements in day, night and rain driving. *Accid. Anal. Prev.* **2010**, *42*, 827–834.
259. Auberlet, J.M.; Rosey, F.; Anceaux, F.; Aubin, S.; Briand, P.; Pacaux, M.P.; Plainchault, P. The impact of perceptual treatments on driver's behavior: From driving simulator studies to field tests-First results. *Accid. Anal. Prev.* **2012**, *45*, 91–98.
260. Johnson, M.J.; Chahal, T.; Stinchcombe, A.; Mullen, N.; Weaver, B.; Bedard, M. Physiological responses to simulated and on-road driving. *Int. J. Psychophysiol.* **2011**, *81*, 203–208.

261. Mayhew, D.R.; Simpson, H.M.; Wood, K.M.; Lonero, L.; Clinton, K.M.; Johnson, A.G. On-road and simulated driving: Concurrent and discriminant validation. *J. Saf. Res.* **2011**, *42*, 267–275.
262. De Winter, J.; Happee, P. Advantages and Disadvantages of Driving Simulators: A Discussion. In Proceedings of the Measuring Behavior, Utrecht, The Netherlands, 28–31 August 2012; pp. 47–50.
263. Bach, K.M.; Jæger, M.G.; Skov, M.B.; Thomassen, N.G. Evaluating driver attention and driving behaviour: Comparing controlled driving and simulated driving. In Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction, Swindon, UK, 1–5 September 2008; pp. 193–201.
264. Masala, G.; Grosso, E. Real time detection of driver attention: Emerging solutions based on robust iconic classifiers and dictionary of poses. *Transp. Res. Part C Emerg. Technol.* **2014**, *49*, 32–42.
265. Martin, S.; Tawari, A.; Trivedi, M.M. Toward privacy-protecting safety systems for naturalistic driving videos. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 1811–1822.
266. Newton, E.M.; Sweeney, L.; Malin, B. Preserving privacy by de-identifying face images. *IEEE Trans. Knowl. Data Eng.* **2005**, *17*, 232–243.
267. Martin, S.; Tawari, A.; Trivedi, M.M. Balancing Privacy and Safety: Protecting Driver Identity in Naturalistic Driving Video Data. In Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, New York, NY, USA, 17–19 September 2014; pp. 1–7.
268. la Torre, F.D. Vision-Based Systems for Driver Monitoring and De-Identification. Available online: <http://fot-net.eu/wp-content/uploads/sites/7/2015/09/Fernando-de-la-Torre.pdf> (accessed on 30 April 2016).
269. Young, R. Revised Odds Ratio Estimates of Secondary Tasks: A Re-Analysis of the 100-Car Naturalistic Driving Study Data. Available online: <http://papers.sae.org/2015-01-1387/> (accessed on 26 October 2016).
270. Strayer, D.L.; Drews, F.A.; Johnston, W.A. Cell phone-induced failures of visual attention during simulated driving. *J. Exp. Psychol. Appl.* **2003**, *9*, 23–32.
271. Apostoloff, N.; Zelinsky, A. Vision in and out of vehicles: Integrated driver and road scene monitoring. *Int. J. Robot. Res.* **2004**, *23*, 513–538.
272. Shin, B.S.; Xu, Z.; Klette, R. Visual lane analysis and higher-order tasks: A concise review. *Mach. Vis. Appl.* **2014**, *25*, 1519–1547.
273. Hirayama, T.; Mase, K.; Takeda, K. Analysis of temporal relationships between eye gaze and peripheral vehicle behavior for detecting driver distraction. *Int. J. Veh. Technol.* **2013**, *2013*, doi:10.1155/2013/285927.
274. Ohn-Bar, E.; Tawari, A.; Martin, S.; Trivedi, M.M. Vision on wheels: Looking at driver, vehicle, and surround for on-road maneuver analysis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 24–27 June 2014; pp. 185–190.
275. Tawari, A.; Sivaraman, S.; Trivedi, M.M.; Shannon, T.; Toppelhofer, M. Looking-in and looking-out vision for urban intelligent assistance: Estimation of driver attentive state and dynamic surround for safe merging and braking. In Proceedings of the 2014 IEEE Intelligent Vehicles Symposium Proceedings, Dearborn, MI, USA, 8–11 June 2014; pp. 115–120.
276. Martin, S.; Rangesh, A.; Ohn-Bar, E.; Trivedi, M.M. The rhythms of head, eyes and hands at intersections. In Proceedings of the 2016 IEEE Intelligent Vehicles Symposium (IV), Gothenburg, Sweden, 19–22 June 2016; pp. 1410–1415.
277. Jain, A.; Koppula, H.S.; Raghavan, B.; Soh, S.; Saxena, A. Car that knows before you do: Anticipating maneuvers via learning temporal driving models. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 13–16 December 2015; pp. 3182–3190.
278. Baddeley, A.D. Selective attention and performance in dangerous environments. *Br. J. Psychol.* **1972**, *63*, 537–546.
279. Vidulich, M.A.; Stratton, M.; Crabtree, M.; Wilson, G. Performance-based and physiological measures of situational awareness. *Aviat. Space Environ. Med.* **1994**, *65*, 7–12.
280. Helmreich, R.L.; Chidester, T.R.; Foushee, H.C.; Gregorich, S.; Wilhelm, J.A. How effective is cockpit resource management training. *Flight Saf. Dig.* **1990**, *9*, 1–17.
281. Ball, K.; Rebok, G. Evaluating the driving ability of older adults. *J. Appl. Gerontol.* **1994**, *13*, 20–38.
282. Satzoda, R.K.; Trivedi, M.M. Drive analysis using vehicle dynamics and vision-based lane semantics. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 9–18.

283. Satzoda, R.K.; Gunaratne, P.; Trivedi, M.M. Drive quality analysis of lane change maneuvers for naturalistic driving studies. In Proceedings of the 2015 IEEE Intelligent Vehicles Symposium (IV), Seoul, Korea, 18 June–1 July 2015; pp. 654–659.
284. Kusano, K.D.; Montgomery, J.; Gabler, H.C. Methodology for identifying car following events from naturalistic data. In Proceedings of the 2014 IEEE Intelligent Vehicles Symposium Proceedings, Dearborn, MI, USA, 8–11 June 2014; pp. 281–285.
285. Martin, S.; Ohn-Bar, E.; Trivedi, M.M. Automatic Critical Event Extraction and Semantic Interpretation by Looking-Inside. In Proceedings of the 2015 IEEE 18th International Conference on Intelligent Transportation Systems, Las Palmas, Spain, 15–18 September 2015; pp. 2274–2279.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).

4.1.8. Sistema Automático Para la Detección de Distracción y Somnolencia en Conductores por Medio de Características Visuales Robustas

- Fernández, A., Usamentiaga, R. & Casado, R. Sistema Automático Para la Detección de Distracción y Somnolencia en Conductores por Medio de Características Visuales Robustas. Revista Iberoamericana de Automática e Informática Industrial. 2017.
- Para el año 2015 la Revista Iberoamericana de Automática e Informática Industrial (ISSN: 1697-7912) tuvo un índice de impacto de 0.475 (Q4) y un índice de impacto a 5 años de 0.362.

D. José Luis Díez Ruano, miembro de la Secretaría de la Revista Iberoamericana de Automática e Informática Industrial (RIAI), a petición de los autores,

HAGO CONSTAR QUE: el artículo con título

***“Sistema Automático Para la Detección de Distracción y Somnolencia en
Conductores por Medio de Características Visuales Robustas”***

y autores:

Alberto Fernández Villán, Rubén Usamentiaga Fernández, y Rubén Casado Tejedor

ha sido aceptado para su publicación en RIAI.

En Valencia, a 3 de mayo de 2017



Fdo. José Luis Díez Ruano

Sistema Automático Para la Detección de Distracción y Somnolencia en Conductores por Medio de Características Visuales Robustas

Alberto Fernández Villán^{a,*}, Rubén Usamentiaga Fernández^b, Rubén Casado Tejedor^b

^aGrupo TSK, Parque Científico y Tecnológico de Gijón, 33203 Gijón, Asturias, España

^bUniversidad de Oviedo, Campus de Viesques, 33204 Gijón, Asturias, España

Resumen

De acuerdo con un reciente estudio publicado por la Organización Mundial de la Salud (OMS), se estima que 1.25 millones de personas mueren como resultado de accidentes de tráfico. De todos ellos, muchos son provocados por lo que se conoce como inatención, cuyos principales factores contribuyentes son tanto la distracción como la somnolencia. En líneas generales, se calcula que la inatención ocasiona entre el 25 % y el 75 % de los accidentes y casi-accidentes. A causa de estas cifras y sus consecuencias se ha convertido en un campo ampliamente estudiado por la comunidad investigadora, donde diferentes estudios y soluciones han sido propuestos, pudiendo destacar los métodos basados en visión por computador como uno de los más prometedores para la detección robusta de estos eventos de inatención. El objetivo del presente artículo es el de proponer, construir y validar una arquitectura especialmente diseñada para operar en entornos vehiculares basada en el análisis de características visuales mediante el empleo de técnicas de visión por computador y aprendizaje automático para la detección tanto de la distracción como de la somnolencia en los conductores. El sistema se ha validado, en primer lugar, con bases de datos de referencia testeando los diferentes módulos que la componen. En concreto, se detecta la presencia o ausencia del conductor con una precisión del 100 %, 90.56 %, 88.96 % por medio de un marcador ubicado en el reposacabezas del conductor, por medio del operador LBP, o por medio del operador CS-LBP, respectivamente. En lo que respecta a la validación mediante la base de datos CEW para la detección del estado de los ojos, se obtiene una precisión de 93.39 % y de 91.84 % utilizando una nueva aproximación basada en LBP (LBP.RO) y otra basada en el operador CS-LBP (CS-LBP.RO). Tras la realización de varios experimentos para ubicar la cámara en el lugar más adecuado, se posicionó la misma en el salpicadero, pudiendo aumentar la precisión en la detección de la región facial de un 86.88 % a un 96.46 %. Las pruebas en entornos reales se realizaron durante varios días recogiendo condiciones lumínicas muy diferentes durante las horas diurnas involucrando a 16 conductores, los cuales realizaron diversas actividades para reproducir síntomas de distracción y somnolencia. Dependiendo del tipo de actividad y su duración, se obtuvieron diferentes resultados. De manera general y considerando de forma conjunta todas las actividades se obtiene una tasa media de detección del 93.11 %.

Palabras Clave:

Detección distracción y somnolencia, Visión por computador, Percepción y reconocimiento, Aprendizaje automático, Monitorización y supervisión

1. Introducción

La conducción es una actividad que requiere un alto grado de concentración por parte de la persona que la realiza, ya que un pequeño descuido es suficiente para sufrir un accidente con las consiguientes pérdidas materiales y/o humanas. De acuerdo al más reciente estudio publicado por la Organización Mundial de la Salud (OMS) en 2013, se estimó que 1.25 millones de personas mueren como resultado de accidentes de tráfico y entre

20 y 50 millones más sufren accidentes sin perder la vida pero pudiendo derivar en dolencias crónicas (Organization (2016)). Todas estas muertes y accidentes no sólo afectan de manera directa a los familiares de las víctimas, sino que, además, tienen un alto coste sobre los presupuestos de los gobiernos, que se estima entre un 3 y un 5 % del producto interior bruto (Peden et al. (2016)).

De todos estos accidentes, muchos son provocados por lo que se conoce como inatención. Este término engloba diferentes estados del conductor, como pueden ser la distracción y la somnolencia, siendo precisamente éstos los que más fatalidades ocasionan. Existen muchas publicaciones e investigaciones que intentan poner cifras que indiquen la cantidad producida por la

* Autor en correspondencia.

Correos electrónicos: alberto.fernandez@grupotsk.com (Alberto Fernández Villán), rusamentiaga@uniovi.es (Rubén Usamentiaga Fernández), rcasado@lsi.uniovi.es (Rubén Casado Tejedor)

inatención (y sus subtipos), pero no existe una figura exacta sobre los accidentes causados por la inatención puesto que todos estos estudios están realizados en diferentes lugares, diferentes marcos temporales, y por tanto, en diferentes condiciones. En líneas generales, se calcula que la inatención ocasiona entre el 25 % y el 75 % de los accidentes y casi-accidentes (Talbot et al. (2013)).

Uno de los trabajos que mejor trata de definir estos conceptos y su relación es el propuesto por Regan et al. (2011), el cual define una taxonomía para la inatención, con sus diferentes subtipos y que la define como *‘insuficiente o no atención a las actividades críticas para una conducción segura’*. En esta taxonomía, a) la somnolencia la engloban en el tipo *‘Conductor con Atención Restringida’*, definida como *‘insuficiente o no atención a las actividades críticas para una conducción segura debido a factores biológicos en los que el conductor no es capaz de procesar la información crítica, como por ejemplo, en eventos de micro-sueños, parpadeos, o procesos de somnolencia’*, y b) la distracción la engloban en el tipo *‘Conductor con Atención Desviada’*, que la definen como *‘desviación de la atención de las actividades críticas para una conducción segura’*.

En un estudio realizado en 10 países europeos acerca de la somnolencia y la conducción, la fatiga incrementa el tiempo de reacción en un 86 % y es la cuarta causa de muerte en las carreteras españolas (RACE (2016)). Además, cabe destacar que el 75 % de los conductores españoles han sufrido episodios de somnolencia mientras conducían, muy superior a la media del 47 % que han admitido este hecho. Además, otro factor importante a tener en cuenta es que aunque los accidentes producidos por la somnolencia suelen ser muy graves (vistas las estadísticas anteriores de mortalidad), muchos conductores infravaloran esta situación y conducen aunque noten la presencia de sus síntomas. Bostezos frecuentes, cabeceos, visión borrosa, caída de párpados y esfuerzos por mantener tanto la atención como los ojos abiertos son signos habituales de somnolencia (RACE (2016)). Respecto a la distracción, éste es uno de los factores que más fatalidades ocasiona en España. Por ejemplo, de acuerdo con la Dirección General de Tráfico (DGT), la distracción es la primera infracción detectada en los accidentes con víctimas, con un 13,15 % de los casos (StopChatear (2016)).

Por todo esto, este campo ha sido vastamente explorado por la comunidad investigadora, donde los diferentes estudios y soluciones para luchar contra la inatención se pueden agrupar en tres grandes grupos.

El primero de ellos se corresponde con los métodos basados en el comportamiento vehicular. Estos métodos detectan el estado del conductor analizando constantemente ciertas métricas como pueden ser la posición del coche, los movimientos del volante, la presión del acelerador o del freno, el cambio de marchas (entre otros), y si en alguno se sobrepasa un determinado umbral, es probable que el conductor esté somnoliento o distraído (Liu et al. (2009); Forsman et al. (2013); Sahayadhas et al. (2012)). En líneas generales, el principal inconveniente de estos métodos es que su eficacia depende principalmente de las características individuales del vehículo, conductor y carretera (Sahayadhas et al. (2012); Selvakumar et al. (2015); Jo et al. (2014)). Dentro de los métodos basados en el comportamiento

vehicular, empiezan a desarrollarse alternativas que requieren la comunicación entre vehículos para operar correctamente (Sla- wiński et al. (2015)).

El segundo de los grupos se basa en el análisis de variables fisiológicas, principalmente para la detección de la somnolencia. Son métodos muy robustos pues permiten la detección de la somnolencia en sus fases tempranas con una baja tasa de falsos positivos (Sahayadhas et al. (2012)). En este grupo destacan los métodos basados en: a) electroencefalograma (EEG), b) electromiograma (EMG), c) electrocardiograma (ECG), d) electrooculograma (EOG). De entre todos estos métodos, el más común para la detección de la somnolencia es EEG, dónde se analizan diferentes bandas de frecuencia (Sahayadhas et al. (2012)). Sin embargo, estos métodos requieren contacto con el conductor para la realización de las medidas, lo que ocasiona que su implementación en entornos reales no sea ni lo más adecuado ni lo más práctico (Dasgupta et al. (2013); Sahayadhas et al. (2012)).

Finalmente, el tercero de los grupos se basa en el análisis de características visuales que presenta un conductor distraído o bajo un estado somnoliento. Un conductor distraído se caracteriza por no mantener la atención puesta en la carretera, por lo que son continuos los movimientos de cabeza hacia ambos lados, sin mantener fija la mirada en la carretera. En cuanto a la somnolencia, las características visuales que la describen son muy variadas, incluyendo movimientos faciales, parpadeos rápidos y constantes, cabeceos y bostezos frecuentes. Hacer constar que, estas características visuales de la somnolencia aparecen en espacios temporales diferentes y normalmente bien definidos (Jo et al. (2014)). De manera específica, los bostezos ocurren generalmente antes de que el conductor entre en somnolencia mientras que, normalmente, los cabeceos ocurren cuando el conductor se duerme. Es por ello que los métodos basados tanto en los bostezos como en los cabeceos no son capaces de detectar con exactitud cuando un conductor está empezando a estar somnoliento. Sin embargo, los métodos basados en obtener información de los ojos pueden detectar con precisión este punto, es decir, son los métodos visuales más adecuados para la detección de la somnolencia (Vural et al. (2007)). Sin embargo, cabe decir que, puesto que existen esas diferencias temporales entre los distintos signos visuales, un punto importante puede ser la combinación de varias de estas características para aumentar la robustez final de la solución (Jo et al. (2014); Sahayadhas et al. (2012)).

Basado en las premisas anteriores, el objetivo del presente artículo es el de proponer, construir y validar una arquitectura basada en el análisis de características visuales mediante el empleo de técnicas de visión por computador y aprendizaje automático para la detección tanto de la distracción como de la somnolencia en los conductores. En concreto, se propone una arquitectura de procesamiento especialmente diseñado para operar en entornos vehiculares, con una carga computacional muy baja y fácilmente integrable en dispositivos con reducidas capacidades de cómputo y capaz de lidiar con distintas condiciones de imágenes muy presentes en este tipo de entornos, como pueden ser las condiciones lumínicas, la resolución de la imagen y la apariencia y pose del rostro del conductor en la

imagen.

En resumen, la principal contribución es el hecho de presentar una solución completamente autónoma para detección de distracción y somnolencia pues son dos de los tipos en la inatención que más accidentes ocasionan. Además, dicha solución presenta las siguientes características: 1) detección automática de la presencia del conductor en el entorno vehicular para dirigir el flujo del algoritmo, 2) detección facial adaptada al entorno vehicular, 3) normalización facial para enfrentarse a características de imagen difíciles, 4) detección rápida y robusta de distracción y 5) detección rápida y robusta de somnolencia.

La organización del resto del trabajo es como se expone a continuación. En la Sección 2 se comentan los principales trabajos relacionados, los cuales sirven como base a la propuesta aquí presentada. A continuación, en la Sección 3 se comenta la metodología empleada, donde se detallan las principales fases del sistema acorde con las contribuciones antes comentadas. En la Sección 4 se comentan los principales puntos en cuanto a la implementación que puedan servir para replicar el trabajo aquí presentado. En la Sección 5 se comentan los resultados obtenidos, tanto con bases de datos de referencia como su validación en entornos reales. Por último, en la Sección 6 se comentan las conclusiones y el trabajo futuro.

2. Estado del arte

Puesto que el presente sistema trata de abordar tanto la distracción como la somnolencia utilizando técnicas de visión por computador, expondremos los principales trabajos en este ámbito, distribuidos en tres grupos acorde a su objetivo (distracción, somnolencia y ambos tipos). Además, extraeremos unas conclusiones de trabajos previos, que servirán de base para proponer nuestra arquitectura.

Para la detección de la distracción existen básicamente dos aproximaciones ¹(Fernández et al. (2016)). Por un lado, están las aproximaciones que detectan la distracción por medio de cámaras de alta resolución colocadas por toda la cabina del conductor con el objetivo de poder observar los ojos del conductor independientemente de la posición de la cabeza del conductor. Existen varios inconvenientes en esta aproximación (Hansen and Ji (2010)) como son: a) necesidad de calibración periódica entre las cámaras, b) imposibilidad de captar los ojos con robustez debido a desviaciones en la cabeza del conductor que causan oclusión, o c) requerimiento computacional elevado. Sin embargo, muchos sistemas de ayuda a la conducción no necesitan disponer de la posición exacta de hacia dónde está mirando el conductor con precisión, sino que es suficiente con una aproximación de la misma (Lee et al. (2011)) para, por ejemplo, saber si el conductor está mirando al frente o a alguno de los lados. Estas aproximaciones se basan en que la posición de la cabeza puede ser suficiente para extraer la información de hacia dónde está mirando el conductor (Boyraz et al. (2012)).

Esta aproximación funciona bien porque la pose de la cabeza es un indicador de hacia dónde tiene puesto el foco de atención el conductor y, por tanto, hacia donde está mirando (Murphy-Chutorian and Trivedi (2010)). Siguiendo esta aproximación, en Hattori et al. (2006) se propone un sistema que determina la distracción del conductor en caso de que no se detecte una cara frontal. En esta línea, existen otras publicaciones que se comentan a continuación. You et al. (2013) entrenan varios modelos para la detección facial, captando varias categorías: a) cara no encontrada, b) cara frontal, c) cara de perfil hacia la derecha, y d) cara de perfil hacia la izquierda. Otro sistema similar es el implementado en Flores et al. (2010) donde, con el objetivo de detectar la distracción, se genera una alerta si se detecta que el conductor no está mirando de frente.

La mayoría de los trabajos para la detección de la somnolencia se basan en la obtención, de la manera más robusta posible, de lo que se conoce como PERCLOS (Dinges and Grace (1998)) - PERcentage of eye CLOSure - por considerarse la métrica más aceptada y extendida para la detección de la somnolencia (Dong et al. (2011)). Dicha métrica ha sido validada usando tanto EEG como evaluación subjetiva, por ejemplo, encuestas (Dong et al. (2011)). Con el objetivo de aumentar esta robustez, Sigari (2009) propone la extracción de signos de fatiga en la parte mitad superior de la región facial del conductor. Es decir, captan signos de somnolencia de la región facial sin realizar un preprocesamiento de dicha región. En nuestra solución se utiliza dicha aproximación para establecer la zona de la región facial que usaremos para extraer a posteriori los signos de somnolencia. Otra opción, es combinar información de otros sensores. Recientemente, López Romero (2016) propone un sistema para la detección de la somnolencia. Utiliza técnicas de visión por computador para extraer del rostro información de somnolencia y se complementa con un sensor de oximetría para la extracción del pulso, que se coloca alrededor del dedo de la mano izquierda, lo que hace que el sistema sea intrusivo.

Algunos trabajos intentan extraer tanto la distracción como la somnolencia. Por ejemplo, en Flores et al. (2011) se detectan ambos factores de inatención tanto de día como de noche. Para ello, realizan el seguimiento de los ojos y el rostro para posteriormente extraer los índices de somnolencia y distracción.

Conclusiones de trabajos previos

El hecho de incorporar varias características permite aumentar la robustez de los sistemas para la detección tanto de distracción como de somnolencia. Para ello, se combinan diferentes técnicas y algoritmos (Jo et al. (2014); Sahayadhas et al. (2012); Noori and Mikaeili (2016)). A modo de ejemplo, se puede destacar la reciente publicación de Noori and Mikaeili (2016), donde abordan la detección de la somnolencia fusionando la información del encefalograma, electrooculograma y señales de la conducción. Sin embargo, los métodos basados en visión por computador son una buena alternativa para monitorizar al conductor de forma no intrusiva sin que interfiera con la conducción. Por ello, el objetivo del presente trabajo es el de proponer una arquitectura basada en diferentes características visuales extraídas mediante procesamiento de imagen y apren-

¹En general cuando se habla de distracción sin especificar el tipo, los autores se refieren a distracción visual, aunque también existen otros tipos de distracción como son la cognitiva y la manual. Así que siempre que se hable de distracción nos referiremos a distracción visual

dizaje automático para aumentar la robustez del sistema de detección.

Para la detección de la somnolencia el factor clave es la detección del estado de los ojos de manera robusta. Algunas aproximaciones para la detección del estado de los ojos en los conductores se basan en la detección de ciertas características que permitan discriminar el estado del ojo. Estos algoritmos se pueden agrupar en lo que se conoce como ‘*métodos basados en características*’. Ejemplos de este tipo pueden ser la aproximación del iris por medio de elipses o la obtención del estado de los ojos contabilizando la distribución de ausencia/presencia de iris y del blanco ocular en función de si está abierto o cerrado. Para ello, se suele recurrir a acumular las intensidades de los píxeles tanto en el eje vertical como horizontal de la imagen, que reciben el nombre de proyecciones. Caracterizando las curvas resultantes, se puede obtener el estado del ojo (Zhang and Zhang (2006); Devi and Bajaj (2008); Lu et al. (2011)). Otra aproximación, conocida como ‘*métodos basados en apariencia*’, se presenta como una alternativa más robusta para la detección de ciertas características y atributos faciales, entre los que está el estado de los ojos (Song et al. (2014)), y se basa en obtener la apariencia mediante operadores robustos como LBP, wavelets de Gabor o similares (Song et al. (2014)). En lo que a la detección del estado del ojo se refiere y utilizando los métodos basados en apariencia, antes comentados, es común el uso de aproximaciones de aprendizaje automático en la etapa de clasificación: Redes Neuronales (NN), Máquinas de Soporte Vectorial (SVM), Árboles de Decisión y Adaboost. De todas éstas, las SVM son las que se comportan de manera más robusta (Song et al. (2014)). Es por ello, que para la aproximación actual, se usa SVM como herramienta para el aprendizaje automático en la etapa final de clasificación del estado de los ojos.

Para la detección de la distracción, el factor clave es la detección facial. El algoritmo de Viola & Jones es un algoritmo que permite la detección facial con alta robustez y que puede ser ejecutado en tiempo real. Como muestra de su aplicabilidad, muchos de los sistemas aquí comentados para la detección de la distracción y somnolencia lo usan (Hattori et al. (2006); Sigari (2009); You et al. (2013); Flores et al. (2010, 2011); Hong and Qin (2007); López Romero (2016)). Sin embargo, este método es computacionalmente exigente y no es el más adecuado para su integración en dispositivos con reducidas capacidades de cómputo en entorno vehicular. Es por ello que en la presentación actual se usará, validará e integrará en el sistema, una alternativa más adecuada que el algoritmo de Viola & Jones.

3. Metodología

En esta sección se comentan los puntos principales de la arquitectura desarrollada, cuyo esquema de funcionamiento puede verse en la Figura 1. El funcionamiento del sistema se basa en tres módulos principales: ‘*detección de pose*’, ‘*detección de ausencia conductor*’ y ‘*detección de estado de los ojos*’, que permiten la detección tanto de la distracción como de la somnolencia en entornos vehiculares. En primer lugar, se hará una descripción del flujo de las imágenes y cómo éstas se recogen en los buffers CABEZA y OJOS, que recogen los estados de la

cabeza y de los ojos, para disparar, en último lugar, las reglas que permiten detectar tanto la distracción como la somnolencia respectivamente.

En concreto, el sistema propuesto va capturando imágenes de la cámara y las va procesando secuencialmente. En primer lugar, se ejecuta el módulo de ‘*detección de pose*’, que a partir de tres detectores faciales (que abreviadamente identificaremos como frontal, izquierdo y derecho) computacionalmente muy ligeros, se consigue estimar la orientación de la cabeza del conductor.

El hecho de que ninguno de los detectores faciales consiga detectar la presencia de la cara del conductor es debido a: 1) el conductor no está presente, o 2) la cabeza del conductor presenta una inclinación o posición excesiva y por tanto refleja una pose que ocasiona que no sea ‘detectable’ por ninguno de los detectores faciales². En este caso, se ejecuta el módulo ‘*detección de ausencia del conductor*’, que permite discernir si se está en el caso 1) o caso 2) antes comentados.

En caso de que el módulo de detección de pose detecte la presencia del conductor, se comprueba si la pose (o detección) es frontal para determinar el estado de los ojos del conductor mediante el módulo ‘*detección de estado de ojos*’. Como se ha comentado, sólo se detecta el estado de los ojos si la pose es frontal, pues por pruebas realizadas, no se puede detectar correctamente (se obtiene un elevado número tanto de falsos positivos como de falsos negativos) el estado de los ojos si la pose no es frontal, en cuyo caso la detección facial ha sido ‘proporcionada’ por el detector de perfil izquierdo o derecho. Sin embargo, este factor no es importante, pues en ese caso, el conductor sería clasificado como distraído por no estar ‘de frente’ y, por tanto, dicha situación sería tenida en cuenta.

Por tanto, para la detección de la somnolencia, se tiene en cuenta el estado de los ojos del conductor, que se recoge en el buffer **OJOS**. El estado de los ojos puede presentar tres estados (*abiertos, cerrados, indefinido*), donde el valor indefinido es establecido cuando la pose no es frontal, y por tanto, no se detecta el estado de los ojos. Sin embargo, hemos de notar que dicho caso es tenido en cuenta porque sería catalogado como un caso de distracción al no presentar una pose frontal.

Por otro lado, para la detección de la distracción y en función del flujo de ejecución, se tienen en cuenta cinco posibles estados en el conductor, que se registran en el buffer **CABEZA**. Estos cinco posibles estados (*frontal, izquierda, derecha, ausencia e inatención*) se obtienen como se comenta a continuación. Tres de ellos son obtenidos por el módulo ‘*detección de pose*’ y se corresponden con las detecciones positivas en los detectores faciales, identificados en la Figura 1 como **CABEZA.resultado**, donde **CABEZA.resultado** = frontal, izquierda, derecha. Existen dos estados adicionales que son proporcionados por el módulo ‘*detección de ausencia conductor*’, que se corresponden con **CABEZA.ausencia** - obtenido en caso de

²Por las pruebas realizadas cuando los detectores faciales no son capaces de detectar la cara del conductor (y éste está presente) es debido a que presenta una pose excesiva por estar distraído (por ejemplo buscando algo dentro del vehículo), o por estar demasiado somnoliento y la cabeza presenta una gran desviación - hecho conocido como ‘cabezadas’

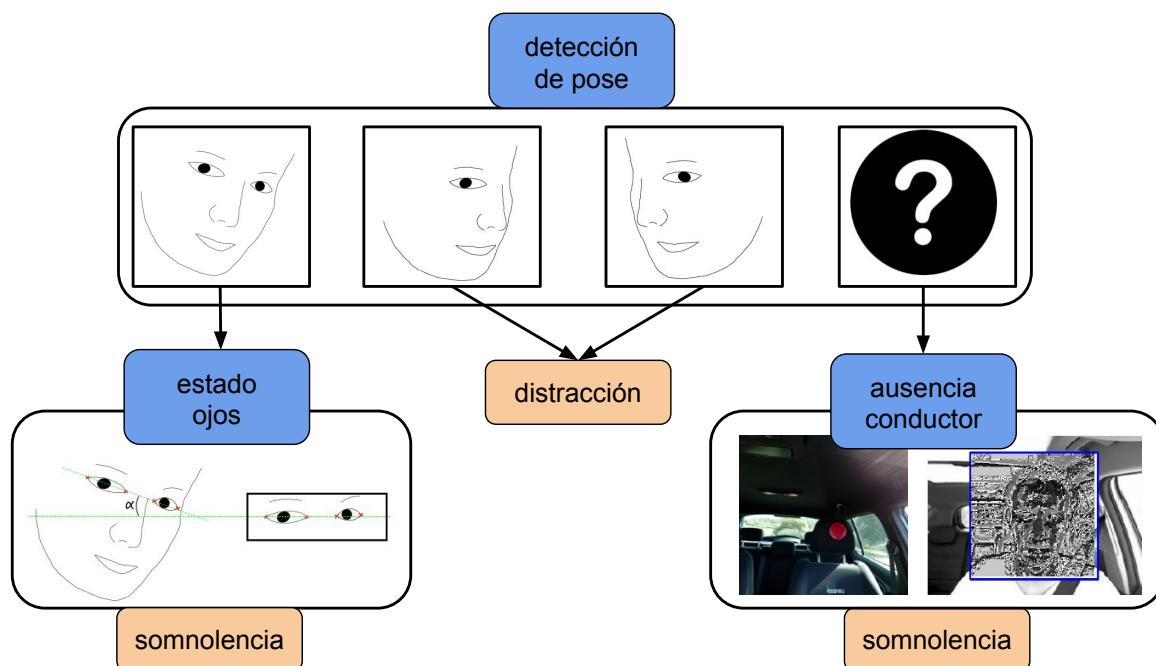


Figura 1: Principales puntos del sistema propuesto

que el conductor no se encuentre en el vehículo - y con *CABEZA.inatencion* - que se corresponde con el caso de que la cabeza del conductor presenta una pose demasiado excesiva y se puede deber tanto a un estado de distracción como de somnolencia (por ello se cataloga como inatención, pues es un factor que engloba a ambos tipos).

A continuación, se proponen y especifican los tres módulos antes comentados. Los detalles de implementación se especifican en la Sección 4, y en la Sección 5, cuando se validan en entorno real, se especifican los tamaños de los buffers (*CABEZA* y *OJOS*), así como las reglas que permiten caracterizar el estado del conductor.

3.1. Detección de estado de los ojos

Los principales signos de somnolencia aparecen en los ojos y, por tanto, se comentarán los puntos del algoritmo implementado para el cálculo de PERCLOS por ser el indicador visual más adecuado.

Una vez que la cara ha sido detectada, el siguiente punto radica en aplicar un detector robusto de características faciales basado en 'Modelos de Partes Deformables' (DPM) propuesto por Uříčár et al. (2012). La salida del detector se corresponde con estimaciones de localizaciones para un conjunto de puntos característicos en la imagen: esquinas de los ojos, esquinas de la boca y nariz. A continuación, se aplica un algoritmo para que las caras sean rotadas y alineadas de manera que los ojos siempre se encuentren en las mismas coordenadas en la imagen final. Para ello, con el objetivo de calcular el ángulo de desviación de

la cara, se calcula una recta de regresión que utiliza los cuatro puntos de los ojos. Por último, se calcula la región facial por encima y por debajo de los ojos, para que únicamente información relevante se procese en las etapas siguientes del algoritmo. Para más detalles de este algoritmo, se puede consultar el Apéndice B. Pre-procesamiento facial para obtener el estado de los ojos, donde se describe dicha etapa de pre-procesamiento con mayor profundidad.

Una vez la región facial a procesar ha sido normalizada, se aplica un *framework* especialmente diseñado e implementado para la normalización de una imagen ante condiciones lumínicas adversas (Tan and Triggs (2010)), situaciones muy comunes en entornos vehiculares. Se trata de un *framework* computacionalmente muy ligero, pero que permite mejorar considerablemente los algoritmos posteriores que se apliquen sobre la región facial pues dicha región se encuentra normalizada en cuanto a la iluminación. Son tres las operaciones que se realizan en este *framework*: 1) corrección gamma, 2) diferencias gaussianas, y c) ecualización de contraste.

A continuación, se aplican tanto el operador LBP (Ojala et al. (1996)), como el operador Center-Symmetric Local Binary Pattern (CS-LBP) (Heikkilä et al. (2009)), que es una modificación del operador LBP computacionalmente más ligera y produce prácticamente los mismos resultados. A continuación se hace una introducción a ambos operadores y para más información, se recomienda la lectura del Apéndice C. Operadores LBP y CS-LBP.

El operador Local Binary Pattern (LBP) (Ojala et al. (1996))

es uno de los descriptores de texturas más populares. Esto es debido principalmente a varios factores: a) fácil implementación, b) invariante a cambios monotónicos de iluminación, y c) complejidad computacional baja. Dicho operador fue introducido en 1996 como un método para sintetizar la estructura del nivel de grises en imágenes. A pesar de que originalmente fue propuesto para el análisis de texturas, el método LBP se ha propuesto para muy diversas tareas en lo que a la visión por computador y el aprendizaje automático se refiere. Relacionado con el procesamiento facial, ha sido empleado para muy diversas tareas de reconocimiento (Losada et al. (2013)), como por ejemplo, reconocimiento facial (Ahonen et al. (2006); Villan et al. (2016)), extracción de variables fisiológicas (Fernández et al. (2015a); Fernandez et al. (2017)) clasificación de género (Shan (2012)), clasificación de expresiones faciales (Shan et al. (2009)), clasificación de la edad (Hadid and Pietikäinen (2013)) e, incluso, detección de gafas (Fernández et al. (2015b)). Dicho operador tiene en cuenta un vecindario local de píxeles alrededor de un píxel central (Ojala et al. (1996)). Seguidamente, umbraliza los píxeles del vecindario con el valor del píxel central y usa el resultado como un número en binario como descriptor para ese vecindario y así sucesivamente para toda la imagen. Fue originalmente propuesto para un vecindario de 8, con 8 bits para codificar los valores binarios. El operador fue posteriormente extendido para incorporar vecindarios de píxeles de diferentes tamaños, haciendo por tanto posible lidiar con las texturas a diferentes escalas (Ojala et al. (2002)).

Para una representación facial eficiente, las características extraídas por el operador LBP deben considerar información espacial. Para ello, la imagen se divide en m regiones $\{R_0, \dots, R_{m-1}\}$ y para cada una de esas regiones, se construye el histograma a partir de la imagen LBP generada tras aplicar el operador. De esta manera, el histograma básico se puede extender recibiendo el nombre de 'histograma espacial extendido' (Ahonen et al. (2006)), codificando tanto la apariencia como las relaciones espaciales de las regiones de la cara. El histograma tendrá una longitud de $B = 256 \times m$, siendo 256, el número de patrones diferentes que se pueden producir para el operador LBP con 8 vecinos y m , el número de regiones. Existe una variante del operador LBP, que se conoce como uniforme (Ojala et al. (2002)), y que reduce el número de patrones de 256 a 59, con lo que el histograma generado se reduce en dimensionalidad.

Si bien el operador es computacionalmente ligero y fácil de implementar, puede producir longitudes de histograma bastante grandes si el número de divisiones es considerable, lo cual, como hemos comprobado en una publicación anterior, tiene influencia directa en el grado de acierto (Fernández et al. (2015c)). Esto puede tener implicaciones de rendimiento en dispositivos con reducidas capacidades de cómputo al tener que tratar con estos histogramas. Se propone usar en vez del operador LBP, una modificación eficiente de este operador para obtener histogramas mucho más reducidos, pero sin perder poder discriminatorio, es decir, manteniendo el rendimiento. Para ello, se hace uso del operador CS-LBP (Heikkilä et al. (2009)), una modificación del operador LBP, que es: 1) computacionalmente más ligero que el operador LBP, 2) más fácil de implementar que el operador LBP, y 3) produce vectores de carac-

terísticas más compactos. Para computar el operador CS-LBP se comparan pares de píxeles opuestos en vez de comparar todos los píxeles con el píxel central. El operador CS-LBP también se emplea para calcular el histograma espacial extendido. En la Figura 2 se puede ver la diferencia en la forma de computar ambos operadores. Como se puede ver en esta Figura, el operador CS-LBP produce 16 patrones diferentes. Si se compara con los 256 producidos por el operador LBP (ó los 59 producidos por la extensión uniforme), se trata de una reducción considerable. Además, añadiendo el umbral T se aumenta la robustez en regiones con poco contraste (Heikkilä et al. (2009)).

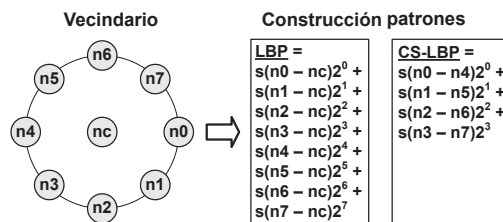


Figura 2: Comparación en la construcción del patrón LBP y CS-LBP para un vecindario de 8

Una vez construido el histograma (mediante alguno de los operadores antes comentados), el último paso consiste en aplicar la etapa de clasificación. Para ello, se hizo uso de las Máquinas de Soporte Vectorial (Support Vector Machines, SVMs). Por tanto, y a modo de resumen, se calcula el histograma espacial extendido (Ahonen et al. (2006)) de la región de los ojos (Fernández et al. (2015c)) normalizada respecto a la iluminación (Tan and Triggs (2010)) aplicando tanto los operadores LBP (Ojala et al. (1996)) como CS-LBP (Heikkilä et al. (2009)). Para clasificar dichos histogramas, se hace uso de la librería LIBSVM (A Library for Support Vector Machines) (Chang and Lin (2011)).

Se realizaron pruebas preliminares para establecer unos valores iniciales de configuración en los operadores LBP y CS-LBP. Tras analizar dichas pruebas, se llega a la conclusión de que, en el caso de LBP, los patrones uniformes $LBP_{P,R}^{u2}$ arrojan mejores resultados que otros tipos de patrones LBP. En cuanto al vecindario, $P = 8$ y un radio de $R = 2$ es la configuración que mejores resultados arroja tanto para LBP como para CS-LBP.

3.2. Módulo detección de pose

Existen muchas aproximaciones para la detección facial en general y otras, para entorno vehicular en particular. El algoritmo más común y extendido es el algoritmo de Viola & Jones (Viola and Jones (2004)). Esto es debido, entre otras cosas, a que es un algoritmo bastante robusto y también porque está listo para usarse en librerías de visión por computador, como OpenCV. Si bien es un algoritmo que se puede ejecutar en tiempo real en un PC, su rendimiento en dispositivos con reducidas capacidades de cómputo se decremente considerablemente. En este trabajo se propone usar una variante del algoritmo de Viola & Jones, especialmente pensada para ejecutarse en este tipo de dispositivos y además, con una tasa de falsos positivos

menor. Dicho algoritmo recibe el nombre de PICO (Pixel Intensity Comparison-based Object detection Markuš et al. (2014)) y proporciona resultados comparables a algoritmos de vanguardia (Li et al. (2015)), pero con un coste computacional muy bajo. Para poner en perspectiva a este algoritmo, se adjunta la tabla 4, donde se puede observar el coste computacional comparado con las implementaciones que ofrece la librería de OpenCV para el algoritmo de Viola & Jones. Para una descripción más exhaustiva de PICO y su comparación con el algoritmo de Viola & Jones, se recomienda la lectura del Apéndice D. Comparación entre los algoritmos de Viola & Jones y PICO.

Siguiendo la aproximación propuesta por Asthana et al. (2011), en el sistema aquí propuesto se hace uso de tres clasificadores. El detector frontal detecta caras con un ángulo de *yaw* entre -40 grados y 40 grados, otro detector de perfil izquierdo que detecta caras con un ángulo de *yaw* entre 30 y 60 y un detector de perfil derecho que detecta caras con un ángulo de *yaw* entre -30 y -60 grados. Además, para estos detectores se estima que se detectan caras entre -30 y 30 grados en el *pitch*. De manera resumida, mediante estos tres detectores se abarca prácticamente todo el ángulo *yaw*, cuyo procesamiento nos daría la distracción. Sin embargo, estos detectores tienen una limitación, que para la detección de la inatención (somnia y distracción) puede ser una virtud que conviene explotar. Estos detectores no detectan caras si tienen un *pitch* excesivo (es lo que se identifica como CABEZA.inatención). Una cara presenta un ángulo *pitch* excesivo sobre todo producto de las típicas cabezadas cuando el conductor está demasiado somnoliento o completamente distraído. Estos ángulos se pueden ver de manera esquemática en la Figura 3.

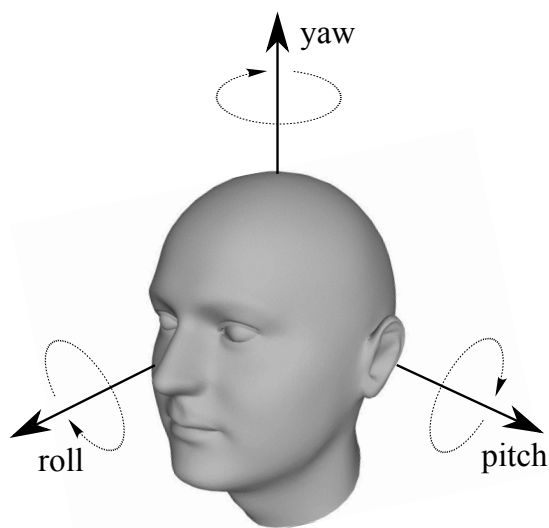


Figura 3: La pose se puede descomponer en los ángulos conocidos como roll, pitch y yaw. El ángulo yaw es especialmente útil para la detección de la distracción y el ángulo pitch para la somnolencia

3.3. Detección de ausencia del conductor

Este módulo detecta si el conductor se encuentra presente en el habitáculo. Como punto de partida, comentar que este módulo se ejecuta tras el módulo de obtención de pose en entorno vehicular. Mediante el presente módulo y el anterior, se permite caracterizar la cabeza del usuario en los cinco estados comentados en la introducción de esta sección: *frontal, izquierda, derecha, ausencia e inatención*). Para la detección de ausencia del conductor, se han implementado dos aproximaciones que se comentan a continuación: a) detección de ausencia de conductor mediante marcador visual en el asiento del conductor, y b) detección de ausencia mediante operador CS-LBP y aprendizaje supervisado.

3.3.1. Detección de ausencia de conductor mediante marcador visual en el asiento del conductor

El primer paso consiste en transformar la imagen del espacio de color RGB al HSV, donde la crominancia y la iluminación están separados, con lo que establecer umbrales para el filtrado posterior es más fácil y más robusto. En segundo lugar, establecemos dichos umbrales para la detección de zonas rojas en la imagen. Una vez filtrada la imagen, se calculan los contornos que envuelven a dichas zonas. Por último, se filtran por tamaño y forma los contornos detectados para obtener el marcador circular. De manera más específica en la etapa de filtrado, se calcula como primer paso el cuadrado mínimo que envuelve al potencial círculo y se considera que el potencial contorno se trata del marcador, si cumple las cuatro condiciones siguientes: a) el contorno tiene más de 6 vértices, b) el cociente entre el ancho y el alto del cuadrado envolvente es aproximadamente de 1, c) el área del contorno es aproximadamente πr^2 (siendo el radio - r - la mitad del ancho de la envolvente), y d) el área del contorno tienen un determinado tamaño en píxeles. Las pruebas del algoritmo se pueden ver en la Sección Pruebas. El hecho de utilizar un marcador, que se ubicaría en el reposacabezas del asiento, permitiría no sólo saber si el conductor está presente o no en su lugar de conducción, sino también realizar aspectos de calibración y establecer heurísticos de manera fácil. A modo de ejemplo, nos permitiría establecer una zona donde buscar la cara en la imagen, para limitar la búsqueda de la región facial.

3.3.2. Detección de ausencia mediante operadores LPB/CS-LBP y aprendizaje supervisado

Comentar que esta aproximación surge a raíz de la implementación de los operadores LBP y CS-LBP para la obtención del estado de los ojos. En concreto, el operador CS-LBP fue propuesto originalmente para describir regiones de interés en imágenes, donde demostró ser tolerante a cambios de iluminación, ruido en la imagen y pequeños cambios de perspectiva, factores bastante frecuentes en el interior de los vehículos. Es por ello que se optó por detectar la ausencia (o presencia) del conductor mediante la aproximación que se comenta a continuación. En este caso, se entrena un clasificador para discernir si el conductor está presente en el vehículo o por el contrario el conductor no está en el mismo. Para ello, se recolectaron imágenes para entrenar el sistema. El procedimiento es muy similar al

Tabla 1: Tiempo medio requerido para procesar una imagen con resolución de 640x480 píxeles

Dispositivo	CPU	Tiempo [ms]		
		PICO	Viola & Jones (OpenCV)	LBP (OpenCV)
PC1	3.4GHz Core i7-2600	2.4	16.9	4.9
PC2	2.53GHz Core 2 Duo P8700	2.8	25.4	6.3
iPhone 5	1.3GHz Apple A6	6.3	175.3	47.3
iPad 2	1GHz ARM Cortex-A9	12.1	347.6	103.5
iPhone 4S	800MHz ARM Cortex-A9	14.7	430.3	129.2

usado para el cálculo de los ojos, lo que cambia es: 1) las imágenes utilizadas para entrenar y validar el sistema, y 2) la región de la imagen utilizada. Por lo tanto, se calcula el histograma espacial extendido de la región de interés normalizada respecto a la iluminación aplicando tanto los operadores LBP como CS-LBP. Respecto a las imágenes, se recopilieron 1000 imágenes negativas donde el conductor no está presente y se recopilieron 1000 imágenes positivas donde el conductor está presente. Con el objetivo de reducir el tamaño de la imagen a procesar y de que el clasificador aprenda mejor las características que le permitan discernir las dos clases, se restringe el espacio de búsqueda (ROI), pero teniendo en cuenta un cierto margen para facilitar el ‘set up’ inicial de la cámara. Como herramienta de clasificación se utilizó SVM. Los resultados del algoritmo y las imágenes utilizadas para entrenar y validar el sistema se pueden ver en la Sección 5.

4. Implementación

El sistema aquí propuesto ha sido desarrollado en C++ y haciendo uso de la librería de OpenCV - Open Source Computer Vision Library, que es una biblioteca de visión por computador multiplataforma, publicada bajo la licencia BSD, que permite ser usada tanto para uso académico como comercial. Incluye más de 500 algoritmos. La última versión estable es la 2.4.13, que es la que ha sido usada en la implementación del sistema. Como herramienta para entrenar los modelos basados en SVM, se hizo uso de la librería LibSVM Chang and Lin (2011). El sistema ha sido construido de manera modular. Cada uno de los módulos que componen el sistema son los descritos anteriormente. Para la detección facial, se hizo uso del framework PICO. Para la detección de los principales puntos faciales, se hizo uso de la librería Flandmarks. Para la normalización de la iluminación de la región facial, se hizo uso de la implementación del algoritmo Tan and Triggs (2010), que requiere únicamente tres llamadas a tres funciones de la librería OpenCV, que se corresponden con los tres pasos propuestos por el framework para la normalización de la iluminación. En relación al operador LBP como del operador, se hizo uso de un algoritmo implementado en C++ y OpenCV. Para el alineamiento de la región facial, y posteriormente quedarnos con la región del entorno de los ojos, se hizo uso también de un algoritmo implementado en C++ y OpenCV.

En un primer momento, el sistema se implementó bajo el sistema operativo Windows, haciendo uso de la librería OpenCV antes comentada y enteramente en C++. De esta manera, se rea-

lizaron los primeros tests y pruebas con bases de datos de referencia y con imágenes capturadas desde entornos vehiculares. Dichos tests se corresponden con la subsección ‘Validación de los módulos que componen el sistema con diferentes bases de datos’.

Dada la modularidad del sistema, el siguiente paso fue portar los algoritmos de visión por computador a la plataforma Android usando el framework JNI. Es un framework que permite que partes de la aplicación en Android se comuniquen con los algoritmos de visión artificial cuya implementación seguiría estando en C++. De esta manera, el sistema no perdería excesivamente en rendimiento. El objetivo de portar los algoritmos a la plataforma Android fue el de construir una aplicación para los dispositivos móviles para validarla y ejecutarla en un entorno real. Dichos tests se corresponden con la subsección ‘Validación en entorno real controlado’.

5. Resultados

Para evaluar el sistema actual con rigurosidad, pero sin poner en riesgo la integridad de cualquier persona involucrada en los experimentos o terceras personas ajenas a los mismos, se llevaron a cabo dos procedimientos. El primero de ellos fue testear con bases de datos los diferentes módulos que componen el sistema y con el segundo, se realizaron diferentes pruebas del sistema completo en un entorno real controlado. Cabe destacar que las pruebas realizadas fueron en condiciones diurnas.

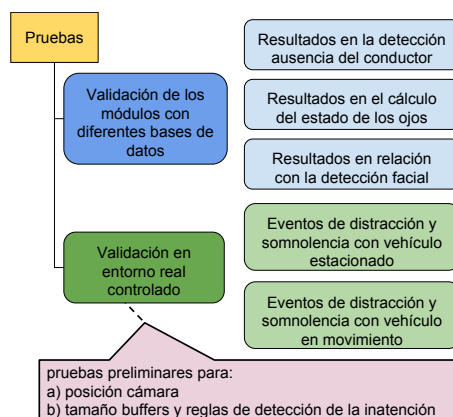


Figura 4: Organización y estructura de las pruebas realizadas



Figura 5: Resultados en la detección ausencia del conductor

Todas estas pruebas se resumen en la Figura 4. A continuación, se comentan los dos procedimientos antes introducidos.

5.1. Validación de los módulos que componen el sistema con diferentes bases de datos

Para la validación tanto de los resultados de ausencia del conductor por medio de aprendizaje automático y la detección del estado de los ojos se hace uso de los operadores LBP, CS-LBP en la etapa de extracción de características y SVM en la etapa de clasificación. En los Apéndices C. Operador LBP y E. Entrenamiento con SVM se pueden ver de manera más detallada aspectos más técnicos y específicos con el objetivo de reproducir los resultados aquí obtenidos.

Resultados en la detección ausencia del conductor. En la Figura 5 se puede ver de manera gráfica cómo funciona el algoritmo para la detección de ausencia del conductor por medio de un marcador. Pruebas preliminares nos demostraron que el factor más influyente en este algoritmo eran las diferentes condiciones lumínicas en el interior vehicular, más que el propio interior propiamente dicho. Las diferentes condiciones lumínicas producen ruido en la imagen, que en la etapa de clasificación por forma circular es capaz de rechazar, con lo que se puede discriminar el contorno circular correspondiente al marcador. Es por ello que sólo se realizaron pruebas en dos vehículos, pero con entornos lumínicos muy diferentes. Se realizaron pruebas con $N = 200$ imágenes y en todas ellas se detectó el marcador correctamente. Podemos, por tanto, decir que este simple

algoritmo es capaz de detectar con robustez la presencia o ausencia de un conductor por medio de un marcador, alcanzando una tasa de precisión del 100 %. El marcador mostrado en la Figura 5 presenta un radio $r = 4\text{cm}$, el cual es suficiente para su detección en condiciones diurnas. Un tamaño más pequeño (por ejemplo $r = 2\text{cm}$) podría presentar problemas, sobre todo cuando existe mucho ruido debido a la iluminación.

Por otro lado, a continuación se comentan las pruebas realizadas para la detección del conductor por medio de procesamiento de imagen y de aprendizaje supervisado. Para entrenar el algoritmo se recolectaron imágenes positivas (el conductor está presente) e imágenes negativas (el conductor no está presente y, por tanto, se ve únicamente el interior del vehículo). Para la recolección de imágenes positivas, se tomaron 1000 imágenes de las siguientes bases de datos: a) YawDD (Abtahi et al. (2014)), b) RS-DMV (Nuevo et al. (2010)), c) Set 11 (Daniluk et al. (2014)), y d) imágenes recogidas para la elaboración de la presente publicación y las diversas pruebas realizadas. Como imágenes negativas, se capturaron un total de 1000 imágenes desde el interior de cinco vehículos diferentes (200 imágenes de cada uno) para obtener cierta variabilidad, y que el clasificador no clasificara características irrelevantes. El *dataset*, por tanto, consta de 2000 imágenes en total.

Como herramienta de clasificación se hizo uso de SVM. Se realizaron pruebas variando parámetros en los diferentes operadores (LBP y CS-LBP), obteniendo resultados similares en ambos casos. De manera más específica, la tasa de acierto media del clasificador basado en el operador LBP fue de 90.56 %,

Tabla 2: Tasa detección para la detección de la ausencia del conductor utilizando tanto un marcador, como los operadores LBP y CS-LBP

Marcador	LBP	CS-LBP
100 %	90.56 %	88.96 %

y del 88.96 % utilizando el operador CS-LBP. En la Figura 6 se puede ver la imagen preprocesada que sirve de entrada al clasificador SVM, donde se puede ver el ROI establecido y el resultado de aplicar tanto el operador LBP como el CS-LBP. El ROI seleccionado nos permite, por un lado, que la variabilidad en la imagen sea ofrecida por la presencia/ausencia del conductor y no por otros elementos como pueden ser cambios en el exterior del vehículo debido al movimiento o la presencia de pasajeros. Sin embargo, con objeto de que no requiera de calibración, no hemos establecido una ROI bastante precisa, para que se pudiera adaptar a diferentes ‘set up’ de cámara.

Los resultados para la detección del conductor se pueden ver de manera resumida en la Tabla 2. No existe comparación posible, pues no se han encontrado publicaciones que lleven a cabo dicha detección.

Resultados en el cálculo del estado de los ojos. Para el cálculo del estado de los ojos, se hizo uso de la base de datos CEW (Closed Eyes in The Wild) Song et al. (2014), recientemente propuesta para este propósito. Esta base de datos constituye una buena base de test, pues contiene imágenes en condiciones no controladas, con iluminaciones muy variadas, baja resolución, diferentes poses o gafas, entre otras. A continuación, se comentan los resultados obtenidos puestos en contexto con otras aproximaciones, los cuales pueden verse de manera resumida en la Tabla 3. En la Figura 7 se puede ver un ejemplo

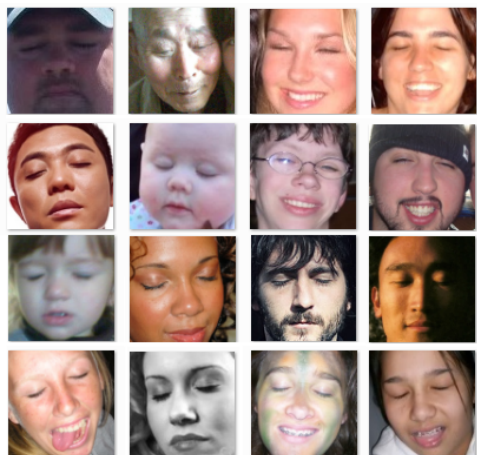


Figura 7: Ilustración de las imágenes de la base de datos CEW

visual de la variabilidad incluida en dicha base de datos.

Los métodos basados en proyecciones, obtienen una tasa de reconocimiento del 70.1 %, muy lejos de la aproximación más robusta que propone el uso de una modificación del al-

Tabla 3: Comparativa de rendimiento para la detección del estado de los ojos en la base de datos CEW y algoritmos *state-of-the-art*

Aproximación	Precisión (%)	Tiempo de ejecución [ms]
Proyección	70.1	-
Escala de grises + SVM	82.85	0.32
LBP + SVM	91.12	1.96
LTP + SVM	92.94	6.67
Gabor + SVM	91.16	17.65
HOG + SVM	93.10	12.61
MultiHOG + SVM	93.31	19.81
HPOG + SVM	93.13	18.55
MultiHPOG + SVM	93.51	38.47
LBP_RO + SVM	93.39	2
CS-LBP_RO + SVM	91.84	1

goritmo basado en la computación de Histogramas Orientados de Gradientes (HOG), originalmente propuesto para la detección de personas Dalal and Triggs (2005), pero que ha demostrado ser robusto para la detección de objetos en general. Dicha modificación del algoritmo HOG, que los autores llaman ‘MultiHPHOG’, obtiene una tasa de acierto del 93.51 % Song et al. (2014). Nuestra aproximación, basada tanto en LBP como en CS-LBP, obtiene unas tasas de acierto del 93.39 % y del 91.84 %, respectivamente. Además, la arquitectura aquí propuesta tiene dos ventajas principales frente a otras aproximaciones: 1) la tasa de reconocimiento es de las más altas sin llevar a cabo una localización exhaustiva del ojo, es decir, es suficiente con detectar de forma más o menos precisa la región de los ojos, lo que aumenta su robustez frente a otros métodos que requieren localizar el ojo, y 2) es una aproximación computacionalmente muy ligera. Por poner en contexto la aproximación más robusta, basada en *MultiHPHOG*, con tasa de reconocimiento del 93.51 % (Song et al. (2014)), tiene un tiempo de ejecución aproximado de unos 40ms. Nuestra aproximación, basada tanto en LBP (93.39 %) como en CS-LBP (91.84 %), tiene un tiempo de ejecución aproximado de unos 2ms y 1ms aproximadamente. Aquí es donde entendemos que es la principal diferencia, pues los algoritmos son unas 20 veces más rápidos, con un ligero decremento de la efectividad, pero aún así, mucho más robustos que la mayoría de algoritmos. En la Figura 8 se puede ver el preprocesamiento y el procesamiento de la imagen para la aproximación que se usa para la detección de los ojos, donde se muestran todos los pasos para llevar a cabo dicha detección.

Resultados en relación con la detección facial. Pruebas realizadas arrojaron que la posición de la cámara juega un papel fundamental en la detección facial y, por tanto, en la detección posterior tanto de la distracción como la somnolencia. Es por ello que se realizaron pruebas para determinar la posición óptima de la cámara sin ocasionar oclusión o distracción al conductor por su posible ubicación. Las pruebas realizadas en este punto se incluyen en la subsección siguiente, pues han sido llevadas a cabo en entorno real.

5.2. Validación en entorno real controlado

La validación en entorno real controlado incluye dos tipos de pruebas: 1) eventos de distracción y somnolencia con el

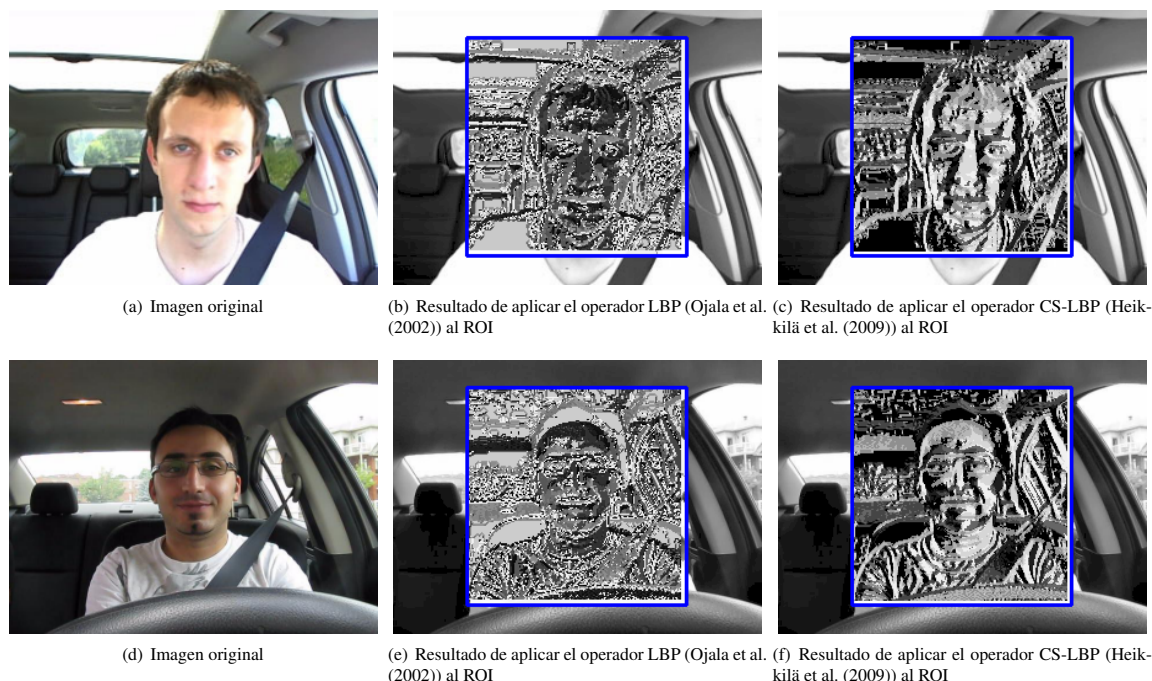


Figura 6: Descripción gráfica de la detección ausencia del conductor mediante procesamiento de la imagen y aprendizaje supervisado

vehículo estacionado y 2) eventos de distracción y somnolencia con vehículo en movimiento en un recorrido preestablecido. En este sentido, tanto: a) posición de la cámara y b) tamaño de los buffers CABEZA y OJOS (introducidos en la Sección 3) juegan un papel fundamental y serán detallados antes de comentar los resultados de estas pruebas.

Para las pruebas se instaló la aplicación en un dispositivo móvil. En concreto, las pruebas fueron realizadas con un móvil de gama media con procesador Qualcomm APQ8064T Snapdragon 600 Quad-Core a 1.9 GHz y 2 GB de RAM. El dispositivo tiene instalada la versión Android 4.2.2. y la aplicación en este dispositivo funciona a unos 10 frames por segundo (fps).

Los vehículos usados fueron: Citroen C4, Skoda Fabia RS y Kia cee'd. Cada vehículo fue conducido por las mismas 16 personas, con una edad media (μ) de 34.5 años desviación estándar (σ) de 6.65. De esos 16 conductores, 8 de ellos ($\mu = 34.25$, $\sigma = 5.91$) eran hombres y los 8 restantes ($\mu = 34.75$, $\sigma = 7.81$) mujeres. A su vez, de los 8 hombres, 4 llevaban gafas en el momento de las pruebas ($\mu = 34$, $\sigma = 8.22$) y 4 no las llevaban ($\mu = 34.5$, $\sigma = 1.5$). En lo que respecta a las 8 mujeres, 4 llevaban gafas ($\mu = 39.5$, $\sigma = 8.29$) y 4 no las llevaban ($\mu = 30$, $\sigma = 2.83$). Ninguno de los conductores llevaba gafas de sol.

Posición de la cámara. Para ambos tipos de pruebas, se tuvo especial cuidado en la ubicación de la cámara. La ubicación de la misma, y en este caso por tanto, del dispositivo móvil, tiene un papel fundamental, pues dependiendo de la imagen adquirida, la cabeza del conductor será encontrada de forma más o menos robusta. Se realizaron pruebas con dos ubicaciones,

que se pueden ver en la Figura 9. En la primera, el dispositivo móvil se sitúa en el espejo retrovisor, y en la segunda, sobre el salpicadero, tratando de capturar las imágenes del conductor lo más frontalmente posible. Respecto a este punto, en la Figura 10 se incluye la posición final escogida, por proporcionar una detección más robusta de la cabeza del conductor, como se comenta de forma más detallada a continuación. Dicha posición, por tanto, fue tenida en cuenta para la validación en entorno real controlado y los dos tipos de pruebas antes introducidos.

El hecho de contar con la cámara frente a la cabeza del conductor presenta varias ventajas. La primera de ellas es que el detector frontal presenta una tasa de reconocimiento mayor. En la Tabla 4, se presenta el rendimiento de ambas posiciones en lo que al grado de detección facial se refiere. Para ello, se grabaron varias secuencias cortas de vídeo contemplando distintas condiciones y el conductor manteniendo una conducción normal, comprobando los retrovisores en caso necesario o cambiando de marchas, pero sin mostrar signos de distracción (como, por ejemplo, mirar fuera del vehículo mediante un giro acusado de la cabeza, lo cual no sería ya una detección frontal). Otra de las ventajas es el hecho de detectar diferentes eventos de distracción, pues se corresponden directamente con la salida de los detectores, es decir, en cuanto la cabeza no se capture frontalmente, es que el conductor está distraído.

Tamaño de los buffers y reglas de detección de la inatención (somnolencia y distracción). La información capturada por el sistema se resume en los buffers CABEZA y OJOS de la Figura 1. El objetivo es la detección de eventos de somnolen-

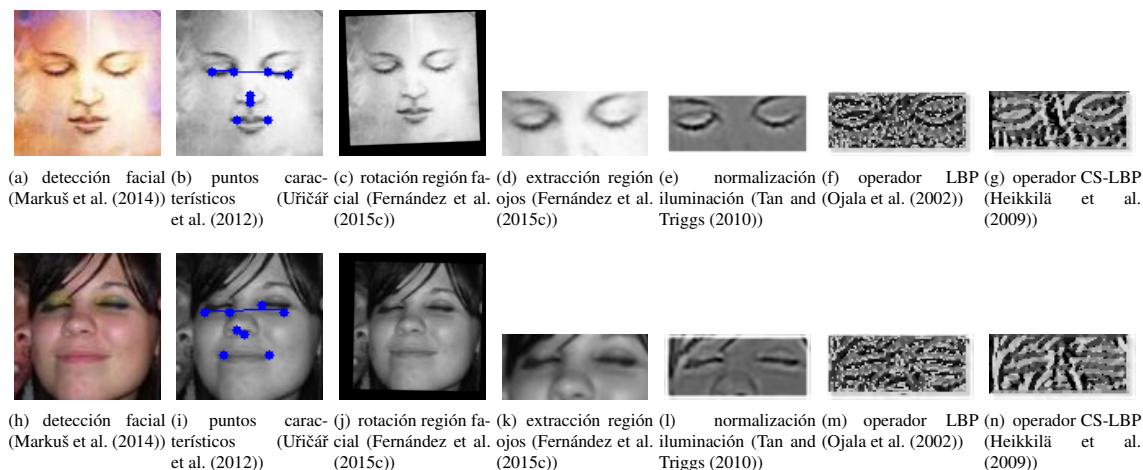


Figura 8: Representación visual de los principales pasos para la detección del estado de los ojos

Tabla 4: Tasa detección facial para las dos ubicaciones de la cámara propuestas: 1) debajo del espejo retrovisor central y 2) en el salpicadero centrado al conductor

		Posición 1 (debajo espejo)	Posición 2 (salpicadero)
Detector facial	frontal(Markuš et al. (2014))	86.88 %	96.46 %

cia y distracción de una manera robusta, pero también teniendo en cuenta una baja tasa de falsos positivos (por ejemplo, que el usuario compruebe los espejos retrovisores y el sistema detecte esta situación como de distracción). Para establecer el tamaño de los buffers, se realizaron: a) diversas pruebas preliminares, b) encuestas a los usuarios en función de resultados intermedios y c) análisis del estado del arte. En este sentido, las tres opciones arrojaron resultados muy similares, que se comentan a continuación.

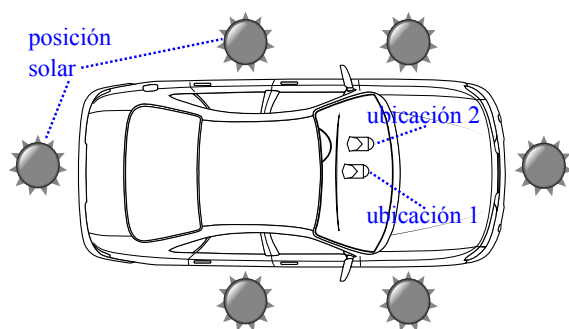


Figura 9: Ubicaciones seleccionadas para la instalación del dispositivo móvil. Además, las pruebas realizadas incluyeron diferentes posiciones solares para conseguir diferentes condiciones lumínicas. Las pruebas realizadas tuvieron lugar tanto en verano como en invierno

Debe ser tenido en cuenta, que del estudio ‘100-Car Study’ cerca del 80 % de los accidentes y del 65 % de los casi-accidentes involucraban algún tipo de inatención por parte del conductor en los últimos 3 segundos antes del evento (Martin (2006)). Además, según otro estudio, se deben detectar eventos de inatención en ventanas de 4 segundos (of Transportation (2016)). Partiendo de estos datos, algunos autores han seguido estas aproximaciones para ajustar las ventanas temporales de sus algoritmos a 3 (You et al. (2013), Berri et al. (2014)) y 4 (Mbouna et al. (2013)) segundos respectivamente.

Tras realizar diversas pruebas y ajustes, se estableció el tamaño temporal de 4 segundos para la detección de eventos de distracción y somnolencia. Esto supone el análisis de unos 40 frames (la aplicación se ejecuta a unos 10fps) y, por tanto, el análisis de unos 40 elementos almacenados tanto para el buffer CABEZA como para el de OJOS. Cada elemento almacenado en los buffers tiene su *timestamp* para disparar los eventos de detección de distracción y somnolencia teniendo en cuenta los 4 segundos antes comentados. Es necesario almacenar el *timestamp* porque el tiempo de ejecución del sistema completo varía en función del flujo de ejecución (ver Figura 1). Cada vez que se almacena un frame (elemento X) en los buffers, se contabiliza el tiempo respecto a la última inserción en el buffer (elemento X-1) y se anota dicho tiempo al elemento X-1. Estos tiempos nos permiten computar después los estados del conductor para disparar las reglas. Las reglas que disparan las detecciones son bastante sencillas. Se estableció una para la somnolencia (ver regla 1) y otra para la distracción (ver regla 2).

La regla para la detección de la somnolencia se basa en procesar los frames contenidos en el buffer OJOS. En caso de que los frames almacenados en dicho buffer, etiquetados como cerrado (CER), superen los 3 segundos, el conductor se considera dormido. En caso contrario, se considera que está despierto. Para la distracción, se considera que el conductor está distraído si los frames etiquetados como izquierda (IZQ), derecha (DER) e inatención (INA) para el buffer CABEZA, superan los 3 segun-

Regla 1 Deteccion Somnolencia**OJOS**

- 1: **if** *tiempos*(OJOS, [CER]) $\geq 3s$ **then**
- 2: Conductor dormido
- 3: **else**
- 4: Conductor despierto
- 5: **end if**

Tabla 5: Actividades a realizar por los conductores tanto para las pruebas con el vehículo estacionado como para el vehículo en movimiento sobre recorrido preestablecido

Eventos de distracción y somnolencia a realizar por los conductores

- ✓A01_X: Cerrar los ojos, con $X=\{4,5,6\}$ seg
- ✓A02_X: Tocar con la barbilla en el pecho, con $X=\{4,5,6\}$ seg
- ✓A03_X: Tocar con la nuca en la espalda, con $X=\{4,5,6\}$ seg
- ✓A04_X: Girar la cabeza hacia la derecha, con $X=\{4,5,6\}$ seg
- ✓A05_X: Girar la cabeza hacia la izquierda, con $X=\{4,5,6\}$ seg

dos. La función *tiempos*(X, [Y]) devuelve el tiempo en el buffer X para los estados contenidos en el array Y.

Regla 2 Deteccion Distraccion**CABEZA**

- 1: **if** *tiempos*(CABEZA, [IZQ, DER, INA]) $\geq 3s$ **then**
- 2: Conductor distraido
- 3: **else**
- 4: Conductor atento
- 5: **end if**

En caso de que la aplicación se instale en un dispositivo con mejores capacidades de cómputo, la métrica *fps* tendrá un valor más elevado, y por tanto, se almacenarán más elementos en los buffers para llegar a cubrir los 4 segundos antes comentados. En caso de ser un dispositivo con menos recursos, se almacenarán menos elementos en los buffers. Por pruebas realizadas, hasta los 5 *fps* se obtienen unos resultados similares, pero para *fps* menores, la aplicación no funciona correctamente.

5.2.1. Eventos de distracción y somnolencia con vehículo estacionado

En parte, estas pruebas sirvieron para ‘ajustar’ ciertos aspectos de funcionamiento de la aplicación antes comentados: a) posición de la cámara, b) tamaño de los buffers, y c) ajustes más finos y precisos en la implementación de los algoritmos. Las pruebas fueron realizadas con el vehículo estacionado. La lista de actividades a realizar por los usuarios se puede ver en la Tabla 5. Se trata de 5 eventos con 3 duraciones diferentes. A modo de ejemplo, la actividad A01.4 consiste en cerrar los ojos durante 4 segundos. Los resultados de estas prueba se recogen en la subsección 5.2.3.

5.2.2. Eventos de distracción y somnolencia con vehículo en movimiento

En la Figura 11 se puede ver el recorrido para las pruebas en entorno controlado en movimiento. Para facilitar que los conductores realizaran el mismo recorrido entre las distintas ite-

raciones, se dispusieron conos para delimitar el recorrido. Las pruebas en movimiento se realizaron a una velocidad de 4km/h durante todas las iteraciones y para todos los conductores. Para ello, se dispuso del Skoda Fabia RS con cambio automático, en el cual, introduciendo la marcha ‘D’, se inicia el movimiento del vehículo manteniendo dicha velocidad constante siempre y cuando no se altere mediante el acelerador o el freno. El hecho de establecer dicha velocidad no es otro que el de permitir, para el recorrido seleccionado, reproducir por parte del conductor varios episodios de distracción y somnolencia con suficiente tiempo, sin poner en riesgo la integridad del conductor debido a la baja velocidad alcanzada. Además, el recorrido se realizó en ambos sentidos. El hecho de capturar las imágenes a una mayor velocidad no supone ningún cambio en las características de las mismas. Los tramos azules se corresponden con zonas en las que está ‘permitido’ realizar las actividades de distracción y somnolencia encomendadas (ver Figura 11). Las actividades son las mismas que para el vehículo estacionado, y se corresponden con las de la Tabla 5. Los resultados de estas prueba se recogen en la subsección 5.2.3.

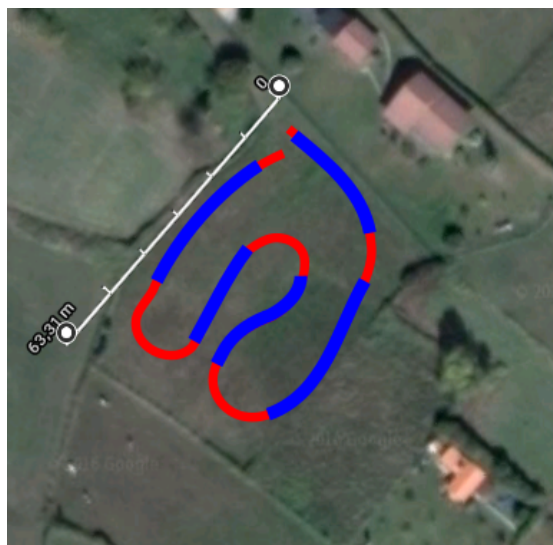


Figura 11: Ruta para las maniobras controladas realizadas en el experimento

5.2.3. Resultados para entorno real controlado

En este apartado se resumen los resultados de las pruebas comentadas anteriormente en las subsecciones 5.2.1 y 5.2.2. Como ya se ha comentado, las pruebas se realizaron con 16 personas, 3 vehículos diferentes y en 2 condiciones diferentes (vehículo parado - 5.2.1 - y en movimiento - 5.2.2). Las pruebas fueron realizadas a lo largo de un total 22 días, con lo que diferentes condiciones lumínicas fueron tenidas en cuenta al realizar las pruebas. Además las pruebas fueron realizadas en diferentes estaciones, tanto en verano (a lo largo de los meses de Agosto y Septiembre) y en invierno (a lo largo de los meses de Enero y Febrero).



Figura 10: Posición (posición 2 - salpicadero) finalmente seleccionada para la ubicación del dispositivo móvil. De esta manera se consigue una detección facial más robusta

Cada prueba (se denominará sesión para una mayor claridad), consistió en realizar las 5 pruebas descritas en la Tabla 5 con las 3 duraciones diferentes (4,5 y 6 segundos). Es decir, cada sesión involucró 15 actividades diferentes. Se realizaron un total de 96 sesiones (16 personas, 3 vehículos y 2 tipos - vehículo estacionado y en movimiento), con una duración aproximada para realizar todas las actividades de una sesión de 25 minutos. Los resultados del sistema para clasificar cada una de estas 96 sesiones se pueden ver en las Tablas 6, 7, 8, las cuales se comentan en detalle a continuación. De la Tabla 6 se puede extraer la media de las 96 sesiones ($\mu = 93.11$, $\sigma = 1.6$).

Como se puede apreciar, se obtienen resultados similares para los eventos de somnolencia A02 y A03, puesto que la mecánica de funcionamiento es la misma, es decir, en caso de que no se puede obtener la pose del conductor (detección negativa en los algoritmos de detección facial), se comprueba si el conductor está presente o no en el habitáculo. También se obtienen resultados similares para la detección de los eventos de distracción (A04 y A05), puesto que la cámara está posicionada justo en frente del conductor y la pose respecto a la cámara es la misma al realizar los eventos de distracción. Respecto al cálculo del estado de los ojos, se obtienen resultados muy robustos, que pueden llegar a una tasa de acierto del 95.43 %. Sin embargo, es en esta actividad donde se encuentran unas diferencias significativas para los conductores que llevaban gafas en el momento de realizar las pruebas. En la Tabla 7 se pueden ver los resultados correspondientes a los 8 conductores (4 hombres y 4 mujeres) con gafas y su tasa de detección en la actividad A01. En la Tabla 8 se pueden ver los resultados para los otros 8 conductores (4 hombres y 4 mujeres) sin gafas.

Para los conductores que no llevan gafas se produce una tasa de detección media de 95.54 %, que puede llegar al 96.88 % si la actividad tiene una duración de 6 segundos. Si la actividad tiene una duración de 4 segundos la tasa de detección es de 94.42 %. Por el contrario, para los conductores que llevan gafas, la tasa de detección media se decrecienta, y se obtiene un valor de 93.51 %, que puede llegar al 94.30 % si la duración de la actividad es de 6 segundos. Si la duración de la actividad es de 4 segundos la tasa de detección es de 92.44 %.

No se encontraron diferencias significativas en caso de que el conductor fuera un hombre o una mujer. En este sentido cabe comentar que las mujeres involucradas en el experimento no presentaban maquillaje y el pelo estaba recogido, de manera que no presentaba ninguna posible oclusión en este aspecto.

Tabla 6: Resumen de los resultados (tasa de acierto) para las pruebas en entorno controlado. Se muestra la precisión al clasificar las 96 sesiones como eventos de somnolencia o no (para A01/A02/A03) o de distracción o no (para A04/A05)

	4 seg	5 seg	6 seg
A01	93.50	94.66	95.43
A02	91.25	92.80	93.61
A03	92.42	92.94	93.27
A04	92.06	92.66	93.37
A05	92.14	93.08	93.39

Tabla 7: Resumen de los resultados (tasa de acierto) para la Actividad del tipo A01.X al clasificar el estado de los ojos para 8 conductores (4 hombres y 4 mujeres) con gafas

	4 seg	5 seg	6 seg
A01	94.42	95.33	96.88

to. Tampoco se encontraron diferencias significativas tanto si el vehículo estaba estacionado como en funcionamiento.

6. Discusión y conclusiones

La comparación mediante datasets y bases de datos disponibles para la comparación de los diferentes algoritmos, como puede ser la base de datos CEW en este caso, permite una comparación 'justa' entre las diferentes aproximaciones. En este sentido, de los resultados desprendidos tras la realización de la validación de los módulos que componen el sistema con diferentes bases de datos (ver Tablas 2, 3), se puede decir que los diferentes módulos se comportan de forma robusta. La detección de ausencia del conductor por medio de un marcador es un método simple y efectivo para detectar si el conductor está presente en el puesto de conducción. En lo que respecta al cálculo del estado de los ojos, la base de datos utilizada para entrenar y validar los algoritmos faciales es un aspecto muy

Tabla 8: Resumen de los resultados (tasa de acierto) para la Actividad del tipo A01.X al clasificar el estado de los ojos para 8 conductores (4 hombres y 4 mujeres) sin gafas

	4 seg	5 seg	6 seg
A01	92.44	93.79	94.30

importante, pues aspectos como la pose, iluminación y resolución de las imágenes juegan un papel fundamental. De manera específica, y como se puede ver en Song et al. (2014), varias aproximaciones basadas en LBP, LTP, Gabor, HOG que en la base de datos ZJU (Pan et al. (2007)) obtienen una precisión de 89.19, 91.39, 85.04, 90.90 pero en cambio evaluados con la base de datos CEW obtienen 81.00, 83.59, 85.53, 90.35 respectivamente, lo que supone un decremento considerable. Es por ello que en la presente publicación se ha optado por emplear la base de datos CEW (ver Figura 3 para visualizar ejemplos de imágenes de dicha base de datos) para el cálculo del estado de los ojos, obteniendo el segundo algoritmo más robusto, sólo por detrás de una aproximación que presenta una carga computacional prácticamente 20 veces más. Además, mediante el empleo de LBP uniforme y CS-LBP se consiguen vectores con dimensionalidad relativamente baja, lo que conlleva una computación más fácil (Jung et al. (2016)).

De los resultados obtenidos tras la ejecución de las pruebas en entorno real controlado, se puede deducir que se detecta con bastante precisión los eventos de distracción y somnolencia, obteniéndose los resultados más elevados para la detección del estado de los ojos (A01). El hecho de que el conductor lleve gafas en el momento de las pruebas ocasionó que la tasa de reconocimiento se decrementara aproximadamente en un 2 %. Por otro lado, se trata de una solución completamente autónoma, apoyada por el módulo de detección de presencia, el cual se puede llevar a cabo, bien por medio de técnicas de aprendizaje automático (sin realizar ninguna modificación en el vehículo), o por medio de un marcador que se situaría en el reposacabezas del asiento; siendo esta alternativa más robusta. Por último, comentar que tampoco hay excesivas diferencias en relación a la duración del evento a detectar. Si bien es cierto que para los eventos con una duración de 4 segundos la tasa es menor, se achaca este decremento al instante temporal inicial y final marcado para dichos eventos, pues los conductores tardaban más en empezar a obedecer la orden dada (por ejemplo, cerrar los ojos) que en finalizar la orden dada (por ejemplo, abrir los ojos).

Por otro lado, la posición de la cámara juega un papel fundamental, pues la detección facial es el punto principal sobre el que pivota el resto del sistema. Esto es así porque los algoritmos de detección facial están entrenados con caras con una determinada pose para poder ser detectadas de frente o de perfil. Así pues, las caras capturadas en la imagen desde el interior vehicular deben tener una apariencia similar a aquellas con las que fueron entrenadas. Si la cámara se sitúa en el salpicadero en frente del conductor, se puede incrementar la detección facial en, aproximadamente, 8 puntos porcentuales respecto a si la cámara está en el espejo retrovisor (de un 86.88 % se puede pasar a un 96.46 %), lo cual está en consonancia con otros trabajos, que establecen que el salpicadero es el mejor lugar para la ubicación de la cámara (Vicente et al. (2015); Abtahi et al. (2014); López Romero (2016)).

El grado de detección facial logrado con la ubicación final de la cámara y utilizando el algoritmo de detección facial antes comentado, es superior al de todos los trabajos analizados (Hattori et al. (2006); Sigari (2009); You et al. (2013); Flores et al. (2010, 2011); Hong and Qin (2007); López Romero (2016)).

Puesto que el resto de la arquitectura de estos sistemas trabaja sobre la detección de la cara del conductor, este es un factor determinante.

Además, el Algoritmo de Viola & Jones es frecuentemente usado por su facilidad de uso y por sus resultados más o menos robustos. En los trabajos analizados (Hattori et al. (2006); Sigari (2009); You et al. (2013); Flores et al. (2010, 2011); Hong and Qin (2007); López Romero (2016)) es usado como detector facial, lo que puede conllevar alguna restricción en el procesamiento. Por ejemplo, en You et al. (2013), con el objetivo de satisfacer los requisitos de 'tiempo real', se vieron obligados a bajar la resolución a 320×240 píxeles, lo que no ocurre en el sistema aquí propuesto.

Respecto a la detección del estado de los ojos, en You et al. (2013) se obtiene una precisión del 92 % utilizando para su evaluación 1780 imágenes. En Sigari (2009) se obtiene una precisión de 91.5 % evaluado con 817 imágenes. En la presente publicación se obtiene una precisión superior teniendo en cuenta 3000 imágenes (E.16). Flores et al. (2010) obtienen una tasa de clasificación del 94 % sobre una base de datos de 500 ojos cerrados y 1000 ojos abiertos. Si bien es una tasa bastante elevada, dicha base de datos ya se encuentra procesada, es decir, está compuesta por *patches* con imágenes de únicamente los ojos, con lo que la parte de pre-procesamiento no se tiene en cuenta.

En este sentido, en la alternativa aquí propuesta para detectar el estado de los ojos (LBP_RO) y (CS-LBP_RO) no se lleva a cabo una localización exhaustiva del ojo, es decir, es suficiente con detectar de forma más o menos precisa la región de los ojos, lo que aumenta su robustez frente a otros métodos que requieren localizar de forma específica el ojo.

Como se ha comentado anteriormente, el hecho de que las imágenes faciales presenten gafas, puede dificultar las tareas de clasificación. En la Figura 12 se pueden observar un conjunto de imágenes que han sido clasificadas de forma errónea en lo que corresponde al estado de los ojos debido a que las mismas presentan diferentes condiciones lumínicas y la presencia de gafas que dificulta las tareas de clasificación. Dichas imágenes han sido extraídas de la base de datos FERET (Phillips et al. (2000)), también usada para comparación de algoritmos de procesamiento facial.

También comentar que se han tenido en cuenta diversos estudios (Martin (2006); of Transportation (2016); You et al. (2013); Berri et al. (2014); Mbouna et al. (2013)) para establecer las ventanas temporales para la detección de los eventos de somnolencia y distracción, llegando a sus mismas conclusiones.

6.1. Limitaciones del sistema propuesto

Existen básicamente dos limitaciones en el sistema propuesto. El primero de ellos es que está orientado como sistema para la detección tanto de la distracción como de la somnolencia en condiciones diurnas. El sistema se debería completar con los algoritmos adecuados para contemplar condiciones nocturnas para ofrecer un funcionamiento ininterrumpido.

Por otro lado, y relacionado con la detección de la distracción visual, el sistema utiliza lo que se conoce como una apro-



Figura 12: Imágenes clasificadas erróneamente ya que en todas ellas los individuos presentan los ojos abiertos y han sido clasificadas con los ojos cerrados

ximación *coarse*, lo que podría suponer alguna limitación en alguna situación muy concreta. Por ejemplo, cuando el conductor esté realizando una tarea secundaria que involucre intervención visual, en la que presente una orientación en la cabeza detectada por el algoritmo de tracking como de distracción, pero pudiendo alternar de manera continua y constante la mirada entre la carretera y el foco de atención de dicha tarea secundaria.

Trabajo futuro

La arquitectura aquí propuesta podría completarse con las limitaciones antes comentadas (funcionamiento en condiciones nocturnas y detección precisa de la ubicación de los ojos para la obtención del foco de atención y poder establecer hacia donde está mirando el conductor) y con tres aspectos que contribuirían a mejorar la robustez. En primer lugar, se podría añadir al flujo de ejecución de la arquitectura un algoritmo para la detección de bostezos. Los bostezos aparecen como uno de los primeros síntomas de la somnolencia, con lo que su detección podría desencadenar recomendaciones y advertencias al conductor. Tanto para entrenar como para testear el grado de acierto del algoritmo a implementar, se podría hacer uso de la base de datos propuesta por Abtahi et al. (2014), que contiene dos bases de datos de vídeo para la comparación de modelos y algoritmos en la detección de bostezos.

En segundo lugar, se podría proponer el uso de un algoritmo de detección de gafas de sol en los conductores, pues en el caso de que las lleve, es imposible la detección del estado de los ojos. Por tanto, dicho algoritmo debería ser aplicado previamente a la detección del estado de los ojos y también antes de aplicar el algoritmo de detección del foco de atención del conductor (en caso de llevar a cabo su implementación).

Y en tercer lugar, el sistema se integrará con el GPS para la obtención de la posición real con el objetivo de posicionar el vehículo en un mapa. De esta manera, se podría obtener si el conductor está en un cruce. En tal situación, se podría diferenciar si el conductor está distraído o si, por el contrario, se está asegurando antes de cruzarlo.

English Summary

Automatic System to Detect Both Distraction and Drowsiness in Drivers Using Robust Visual Features.

Abstract

According to the most recent studies published by the World Health Organization (WHO) in 2013, it is estimated that 1.25

million people die as a result of traffic crashes. Many of them are caused by what it is known as inattention, whose main contributing factors are both distraction and drowsiness. Overall, it is estimated that inattention causes between 25 % and 75 % of the crashes and near-crashes. That is why this is a thoroughly studied field by the research community, where solutions to combat distraction and drowsiness, in particular, and inattention, in general, can be classified into three main categories, and, where computer vision has clearly become a non-obtrusive effective tool for the detection of both distraction and drowsiness. The aim of this paper is to propose, build and validate an architecture based on the analysis of visual characteristics by using computer vision techniques and machine learning to detect both distraction and drowsiness in drivers. Firstly, the modules have been tested with all its components independently using several datasets. More specifically, the presence/absence of the driver is detected with an accuracy of 100 %, 90.56 %, 88.96 % by using a marker positioned onto the headrest, the LBP operator and the CS-LBP operator, respectively. Regarding the eye closeness validation with CEW dataset, an accuracy of 93.39 % and 91.84 % is obtained using a new method using both LBP (LBP_RO) and CS-LBP (CS-LBP_RO). After performing several tests, the camera is positioned on the dashboard, increasing the accuracy of face detection from 86.88 % to 96.46 %. In connection with the tests performed in real-world settings, 16 drivers were involved performing several activities imitating different signs of sleepiness and distraction. Overall, an accuracy of 93.11 % is obtained considering all activities and all drivers.

Keywords:

Distraction and drowsiness detection Computer vision Perception and recognition Machine learning Monitoring and supervision

Agradecimientos

El origen de las actividades del presente trabajo ha sido realizado parcialmente gracias al apoyo tanto de la Fundación para el fomento en Asturias de la investigación científica aplicada y la tecnología (FICYT) y de la empresa SINERCO SL, por medio de la ejecución del proyecto '*Creación de algoritmos de visión artificial*', con referencia IE09-511. El presente trabajo se engloba en la tesis doctoral de Alberto Fernández Villán.

Apéndice A. Detección distracción visual en condiciones diurnas

En un trabajo previo recientemente publicado (Fernández et al. (2016)) analizamos un total de 1500 publicaciones con el objetivo de categorizar los diferentes métodos propuestos para la detección de la distracción en general, y de la distracción visual en particular, utilizando únicamente algoritmos de visión por computador. De acuerdo a dicha publicación y en lo que respecta a la distracción visual, existen principalmente dos aproximaciones.

La primera de ellas es la que se conoce como *coarse*, y se basa principalmente en asumir que el foco de atención del conductor (hacia dónde dirige la mirada) se puede aproximar utilizando la orientación de la cabeza. En la segunda aproximación, que se conoce como *fine*, los investigadores consideran tanto la orientación de la cabeza como la orientación de los ojos para obtener un foco de atención más preciso.

Cada una de las aproximaciones anteriores presenta ventajas e inconvenientes. En lo que respecta a la primera de las aproximaciones, se ha demostrado que la orientación de la cabeza es un indicador robusto del foco de atención del conductor (Murphy-Chutorian and Trivedi (2010)), y ambas medidas están estrechamente ligadas (Hammoud et al. (2005)).

Sin embargo, el conductor, en algunas ocasiones y sobre todo cuando lleva a cabo alguna tarea secundaria que requiere involucración manual o visual (por ejemplo buscar algo en la guantera del vehículo) es común que posicione su cabeza en una posición intermedia entre las dos situaciones que requieren su atención (en este caso, la guantera y la carretera) y mediante continuos y constantes movimientos oculares lleva a cabo la inspección visual de ambas tareas. En esta situación, un algoritmo de *tracking* facial reconocería esta situación como de distracción visual al basarse únicamente en la posición de la cabeza. Por lo tanto, en una situación ideal, se deberían considerar tanto la posición de la cabeza como la posición de los ojos.

Sin embargo, el hecho de obtener la posición de los ojos en todo momento no es una tarea fácil (Song et al. (2013)), debido principalmente a aspectos como: a) las expresiones faciales, b) oclusiones (por ejemplo si el conductor lleva gafas o gafas de sol, o parpadeos), c) pose (por ejemplo puede ser posible que no sea posible recoger la posición de los dos ojos en una posición de perfil), d) diferentes condiciones de la imagen y su calidad (iluminación o baja calidad de la imagen, entre otros). Es por ello que, para lidiar con estos aspectos antes comentados, se instalen diversas cámaras distribuidas por el habitáculo.

A medida que aumenta el número de cámaras, es posible obtener mayor información. Sin embargo, esta información habría que fusionarla (se obtienen datos de cada una de las cámaras) para obtener unos resultados finales. Además, estos sistemas son difíciles de configurar, requieren de calibración periódica para comprobar el estado de las cámaras, pues pequeños desajustes pueden ocasionar problemas (Ahlstrom and Dukic (2010)), y pueden presentarse como intrusivos al conductor. Además, existen muchas situaciones en las que es imposible obtener la posición exacta de los ojos debido a las condiciones propias del entorno vehicular: a) reflejos en las gafas, b) parpa-

deos, c) pose muy acusada del conductor (por ejemplo si presenta su cabeza mirando hacia el techo o el suelo del vehículo es imposible recoger la información de sus ojos - ambas posiciones posibles con un conductor dormido), d) diferentes condiciones de iluminación (Fernández et al. (2016)).

Sin embargo, y con el objeto de profundizar en algoritmos para la detección de los ojos en condiciones no controladas, se hicieron pruebas con dos algoritmos recientemente propuestos para la detección de los ojos. En este sentido, uno de los algoritmos que mejor se ha comportado en vista de los resultados obtenidos de las diversas publicaciones analizadas (Song et al. (2013)) es el algoritmo ASEF (Average of synthetic exact filters), el cual destaca por tener una carga computacional bastante baja, aspecto fundamental para su integración en un entorno embebido, propuesto en Bolme et al. (2009).

Es por ello que se realizaron pruebas preliminares a partir de su implementación en C++ que permite una integración directa con OpenCV. También se experimentó con el algoritmo propuesto en Timm and Barth (2011) para la detección de los ojos, pues por los resultados mostrados en el mismo, son más que prometedores. Además, el algoritmo destaca, como en el caso anterior, por su baja carga computacional. Es por ello que también se realizaron pruebas preliminares a partir de su implementación en C++ que permite una integración directa con OpenCV. En ambos casos, los resultados preliminares (utilizando imágenes con los ojos abiertos de la base de datos CEW) no fueron buenos, no considerándose adecuado para lidiar con las condiciones de iluminación y pose en los entornos vehiculares. En este sentido, los resultados fueron peores cuando:

- Las condiciones lumínicas afectaban la imagen
- La cara presentaba una pose acusada
- Las caras presentaban gafas

Además, comentar que ambos algoritmos (Timm and Barth (2011); Bolme et al. (2009)) fueron usados como etapa de pre-procesamiento para el alineamiento facial, con el objetivo de obtener el ángulo de inclinación obtenido a partir de la posición de los ojos. Tras realizar pruebas con diferentes bases de datos faciales, se comprobó experimentalmente que es más robusto utilizar un detector de *landmarks* faciales (Uřičář et al. (2012)), el cual usamos finalmente en la etapa de pre-procesamiento.

Para más información en este punto, se recomienda leer el Apéndice Pre-procesamiento facial para obtener el estado de los ojos. Es por ello que, finalmente, en la presente publicación, se ha optado por utilizar la primera aproximación (*coarse*), al requerir una solución más compacta, más fácil de instalar y configurar, sin perder de vista los algoritmos para la detección de los ojos, con vistas a una futura integración en la arquitectura.

Finalmente y por completar la información presentada, la detección de distracción visual en condiciones nocturnas se realiza, generalmente, de forma más precisa, pues los sistemas de iluminación infrarroja hacen que la pupila adquiera un brillo característico que posibilita una fácil localización de los ojos. De día, dichos sistemas se ven perjudicados por la luz infrarroja solar (Fernández et al. (2016)).

Apéndice B. Pre-procesamiento facial para obtener el estado de los ojos

Es necesario realizar un pre-procesamiento facial antes de extraer las características de la región facial (en este caso, aplicar el operador LBP/CS-LBP) para obtener el estado de los ojos. Con el objetivo de lidiar con el escalado, rotación, pose e inexactitudes del algoritmo de detección facial, se propuso en Fernández et al. (2015c) un algoritmo para normalizar la región facial (Algoritmo 3). Los principales pasos se comentan a continuación. Una vez que la cara ha sido detectada, el siguiente punto radica en aplicar un detector robusto de características faciales basado en ‘Modelos de Partes Deformables’ (DPM) propuesto por Uříčář et al. (2012). La salida del detector se corresponde con estimaciones de localizaciones para un conjunto de puntos característicos en la imagen: esquinas de los ojos, esquinas de la boca y nariz. A continuación, se aplica un algoritmo para que las caras sean rotadas y alineadas de manera que los ojos siempre se encuentren en las mismas coordenadas en la imagen final. Para ello, con el objetivo de calcular el ángulo de desviación de la cara, se calcula una recta de regresión que utiliza los cuatro puntos de los ojos. Por último, se calcula la región facial por encima y por debajo de los ojos, para que únicamente información relevante se procese en las etapas siguientes del algoritmo. Para más detalles de este algoritmo, se recomienda la lectura de una publicación previa, donde se comprueba que este algoritmo de normalización incrementa la tasa de reconocimiento significativamente Fernández et al. (2015c). De esta manera, se consigue extraer la región facial normalizada alrededor de los ojos de manera totalmente automatizada para posteriormente extraer el vector de características mediante el operador correspondiente.

Algoritmo 3 Normalización facial del ROI

eyes_distance_r, eye_line_r, size

- 1: Se localizan 8 puntos faciales. Se calcula recta de regresión basada en los 4 puntos de los ojos. El ángulo α de desviación es calculado y la imagen se rota teniendo en cuenta dicho ángulo para alinear las imágenes
- 2: La distancia Euclídea d_0 se calcula entre los ojos en la imagen rotada
- 3: La distancia entre los ojos en la imagen redimensionada se calcula según la fórmula $d_t = \text{size.w} * \text{eyes_distance_r}$
- 4: El ratio r se calcula como $r = d_0/d_t$
- 5: La anchura w_0 and la altura h_0 de la región alrededor de los ojos se calcula como $w_0 = r * \text{size.w}$ y $h_0 = r * \text{size.h}$
- 6: Las coordenadas de las esquinas de la región facial en la imagen rotada se calculan como $x_l = x_e - w_0/2$, $y_l = y_e - h_0/\text{eye_line_r}$, $x_r = x_l + w_0$ y $y_b = y_l + h_0$, donde x_l es la coordenada x del borde izquierdo, x_e es la coordenada x del punto medio entre los ojos, y_l es la coordenada y del borde superior, y_e es la coordenada y de los ojos, x_r es la coordenada x del borde derecho e y_b es la coordenada y del borde inferior. Este ROI es extraído de la imagen
- 7: **return ROI**

Para establecer dicha región, los parámetros de entrada al

algoritmo que se han usado son los siguientes: eyes_distance_r = 0.7, eye_line_r = 2, size = 80x36 píxeles, que de manera gráfica se pueden ver en la Figura B.13.

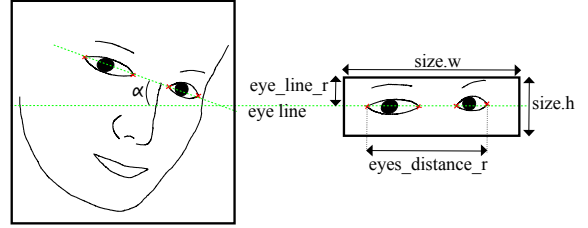


Figura B.13: Representación visual del algoritmo de pre-procesamiento de la región facial mostrando los parámetros de entrada del mismo

Apéndice C. Operadores LBP y CS-LBP

Formalmente, el operador LBP toma la siguiente forma:

$$LBP(x_c, y_c) = \sum_{p=0}^7 2^p s(g_p - g_c) \quad (C.1)$$

donde p recorre los 8 vecinos alrededor del pixel central c , g_c y g_p son los valores del nivel de gris en c y p respectivamente y

$$s(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (C.2)$$

Este operador fue extendido para usar vecindarios de diferentes tamaños Ojala et al. (2002), posibilitando lidiar con texturas a diferentes escalas. Este hecho se denota mediante (P, R) , donde P representa el número de vecinos y R representa el radio para el vecindario. Cuando los puntos no se corresponden con posiciones enteras, el valor de intensidad para un determinado punto es bilinealmente interpolado. A esta implementación se le conoce como LBP ($LBP_{P,R}$):

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} 2^p s(g_p - g_c) \quad (C.3)$$

Otra extensión al operador original es lo que se conoce como patrones uniformes Ojala et al. (2002). Un patrón LBP es uniforme si contiene, como máximo, dos transiciones de 0 a 1 (o viceversa) visto como un *buffer* circular de bits. Por ejemplo, 00000000, 00011110 and 10000011 son los tres patrones uniformes. La uniformidad de los patrones es un hecho importante, pues codifica información como esquinas y bordes. A pesar de que sólo 58 de los 256 patrones (para un vecindario de 8) son uniformes, alrededor del 90 % de los patrones observados son uniformes Ojala et al. (2002). Para identificar a los patrones uniformes, se utiliza la siguiente notación: $LBP_{P,R}^{u2}$. Es por ello que los patrones uniformes se pueden considerar un método efectivo para reducir la dimensionalidad de los datos. Otras extensiones al operador original, también propuestas en la mencionada publicación, se corresponden con: 1) los patrones invariantes a

la rotación, los cuales adquieren la siguiente notación: $LBP_{P,R}^{ri}$ y representan un total de 36 del total de 256 patrones para un vecindario de 8 píxeles; 2) la conjunción de patrones uniformes e invariantes a la rotación: $LBP_{P,R}^{riu2}$ y representan un total de 10 del total de los 256.

Se realizaron pruebas preliminares teniendo en cuenta estos tipos ($LBP_{P,R}^{u2}, LBP_{P,R}^{ri}, LBP_{P,R}^{riu2}$) para diferentes valores de P y R . Según las publicaciones analizadas y la experiencia previa en otras publicaciones (Fernández et al. (2015c); Villan et al. (2016); Losada et al. (2013)) $P = \{8, 16\}$ y $R = \{1, 2, 3\}$.

Tras analizar dichas pruebas, se llega a la conclusión de que los valores con mayor tasa de reconocimiento son los patrones uniformes $LBP_{P,R}^{u2}$ con un número de vecinos de $P = 8$ y un radio de $R = 2$.

Respecto al operador CS-LBP, formalmente toma la siguiente formulación:

$$CS - LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{(P/2)-1} 2^p s(g_p - g_{p+(P/2)}) \quad (C.4)$$

Con este operador también se realizaron pruebas preliminares, obteniendo los mejores resultados, como en el caso del operador LBP con un número de vecinos de $P = 8$ y un radio de $R = 2$.

Una vez se han definidas las características de ambos operadores, el siguiente paso consiste en dividir la imagen en m regiones $\{R_0, \dots, R_{m-1}\}$ y para cada una de esas regiones, se construye el histograma a partir de la imagen LBP generada tras aplicar el operador seleccionado. El número de divisiones juega, por tanto, un papel importante en la longitud final del histograma generado. Escoger este número de divisiones no es tampoco tarea fácil y se requieren diversas pruebas para establecer dicho número.

En lo que se refiere a la tarea de detección de ausencia del conductor mediante LBP y CS-LBP el número de divisiones que mejores resultados arrojó fueron de 5×5 . Para la tarea de detección del estado de los ojos, el número de divisiones óptimo fue de 6×3 .

Apéndice D. Comparación entre los algoritmos de Viola & Jones y PICO

En relación a las necesidades computacionales, el algoritmo de PICO presenta unas prestaciones mucho superiores. En relación a la precisión en dicha detección, en la propia publicación de PICO, comparan su algoritmo con el algoritmo de Viola & Jones. Para ello, hacen uso de la base de datos FDDB (Jain and Learned-Miller (2010)), que contiene 5171 imágenes de caras adquiridas en diferentes condiciones y su uso es común para la comparación de algoritmos de detección facial.

La comparación puede verse en las Figuras D.14, D.15, donde se representan, las curvas ROC discretas y continuas, respectivamente. Como se puede observar, el método de PICO presenta mejores resultados.

Si bien en la presente publicación no se realizaron pruebas cuantitativas para obtener de manera precisa la diferencia

entre ambos algoritmos, de manera cualitativa sí que se puede concluir que el algoritmo de PICO presenta mejores resultados. Este hecho unido a que la carga computacional del algoritmo PICO es mucho inferior, lo hace mucho más interesante para su inclusión en sistemas con reducidas capacidades de cómputo.

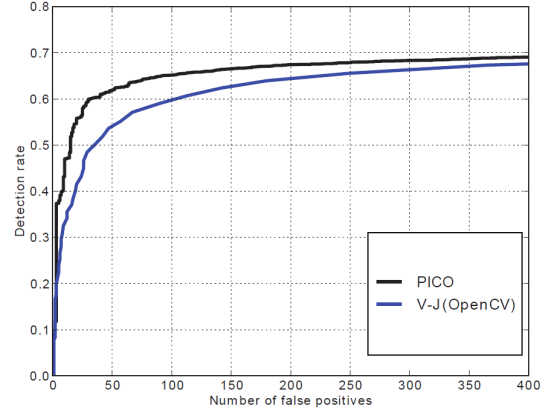


Figura D.14: Comparación de las curvas ROC discretas tanto para los métodos de PICO y de Viola & Jones empleando la base de datos FDDB

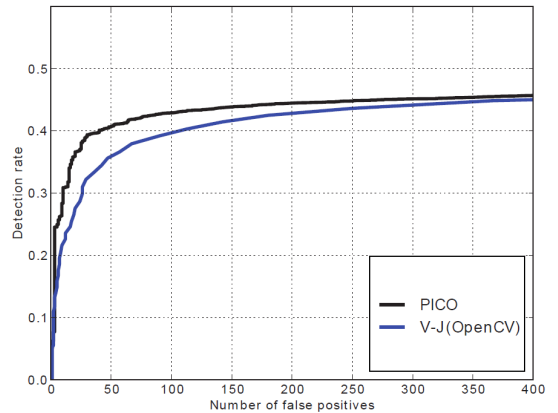


Figura D.15: Comparación de las curvas ROC continuas tanto para los métodos de PICO y de Viola & Jones empleando la base de datos FDDB

Apéndice E. Entrenamiento con SVM

Las Máquinas de Soporte Vectorial (SVMs, por sus siglas en inglés) representan un método de aprendizaje automático muy popular para tareas de clasificación, regresión y otras de aprendizaje Chang and Lin (2011), es también una poderosa herramienta para tareas relacionadas con la extracción de información de la región facial.

La SVM construye un hiperplano o conjunto de hiperplanos en un espacio de dimensionalidad muy alta (o incluso infinita) que, de forma óptima (hiperplano que tenga la máxima

distancia - margen - con los puntos que estén más cerca de él mismo), separe a los puntos de una clase de la de otra Vapnik (1998). Dado una equivalencia no lineal Φ que transforma los datos a uno con mayor dimensionalidad, los *kernels* presentan la siguiente formulación $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$, siendo $\{(x_i, y_i), i = 1, \dots, l\}$ un conjunto de datos de entrenamiento etiquetados, donde $x_i \in R^n, y_i \in \{1, -1\}$.

A pesar de que se están proponiendo nuevos *kernels* Shan et al. (2009), los más populares son: a) lineal, b) polinomial, y c) función de base radial (RBF, por sus siglas en inglés) (Hsu et al. (2003)).

Para trabajar con las SVMs, se hizo uso de la librería LIBSVM (Chang and Lin (2011)). A continuación, y como está sugerido en Hsu et al. (2003), se llevaron a cabo los siguientes pasos:

1. Transformar los datos de entrada al formato de la librería.
2. Realizar un escalado de los datos.
3. Experimentar con diferentes *kernels* y seleccionar el que mejores resultados arroje.
4. Realizar validación cruzada y lo que se conoce como *grid-search* para establecer los parámetros de la SVM.

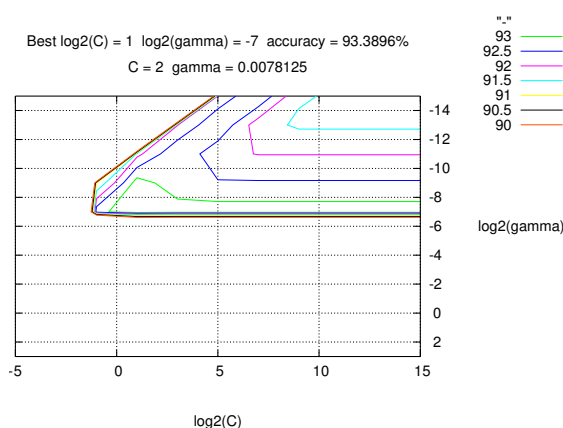


Figura E.16: Grid-search sobre los parámetros C y γ mediante el empleo de la técnica de *cross-validation* para el operador LBP sobre la región de los ojos utilizando el kernel de tipo RBF correspondiente a la SVM para la detección del estado de los ojos

Para el escalado de los datos de entrada se hicieron pruebas con dos rangos de datos recomendados (Hsu et al. (2003)): $[-1, +1]$ y $[0, +1]$, obteniendo mejores resultados con el intervalo $[-1, +1]$, con lo que se estableció dicho intervalo de escalado. Tras realizar diversas pruebas con los tres *kernels* antes comentados (lineal, polinomial y RBF), se seleccionó éste último por presentar mejores resultados. Una vez seleccionado, se realizó lo que se conoce como *cross-validation* y *grid-search* para establecer los mejores parámetros en dicho *kernel*, los cuales se gobiernan principalmente mediante dos parámetros: C y γ .

El hecho de identificar estos dos parámetros (C, γ) es con el objetivo de que el clasificador pueda predecir con robustez

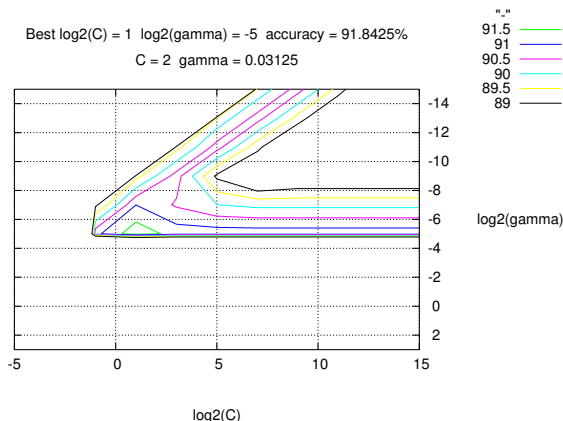


Figura E.17: Grid-search sobre los parámetros C y γ mediante el empleo de la técnica de *cross-validation* para el operador CS-LBP sobre la región de los ojos utilizando el kernel de tipo RBF correspondiente a la SVM para la detección del estado de los ojos

los datos desconocidos (por ejemplo, los datos de entrenamiento). De manera más específica, se llevó a cabo el proceso de *grid-search* de los parámetros C y γ usando *cross-validation*. Es decir, se cogen pares de valores de (C, γ) y para cada uno de ellos se realiza *cross-validation*. Al final del proceso, el par de valores que mejores resultados obtiene es el que se selecciona.

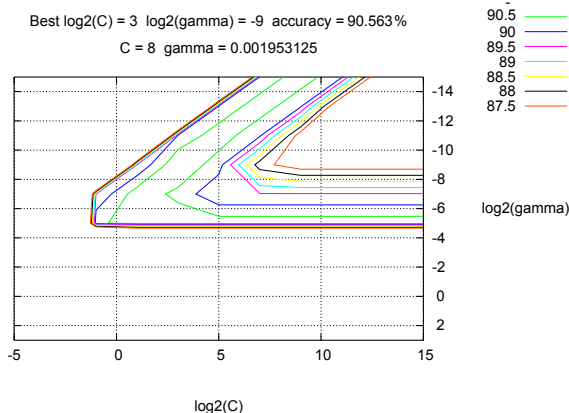


Figura E.18: Grid-search sobre los parámetros C y γ mediante el empleo de la técnica de *cross-validation* para el operador LBP sobre la ROI utilizando el kernel de tipo RBF correspondiente a la SVM para la detección de ausencia del conductor

Para evidenciar de manera visual este procedimiento, mostramos, en las Figuras E.16 y E.17 los resultados tras las diversas ejecuciones en lo que al cálculo del estado de los ojos se refiere. En la Figura E.16 se puede ver que el mejor resultado obtenido es de 93.39% y se corresponde a aplicar el operador LBP sobre la región de los ojos.

En la Figura E.17 se puede ver que el mejor resultado es de 91.84% y se corresponde a aplicar el operador CS-LBP sobre

la región de los ojos. Además, en relación a la detección de la ausencia del conductor utilizando tanto LBP como CS-LBP y SVM, en las Figuras E.18 y E.19 se muestran los resultados utilizando el mismo procedimiento antes comentado.

En la Figura E.18 se puede ver que el mejor resultado obtenido es de 90.56 % y se corresponde a aplicar el operador LBP sobre el ROI establecido para la detección automática de la presencia del conductor. En la Figura E.19 se puede ver que el mejor resultado obtenido es de 88.96 % y se corresponde a aplicar el operador CS-LBP sobre el ROI establecido.

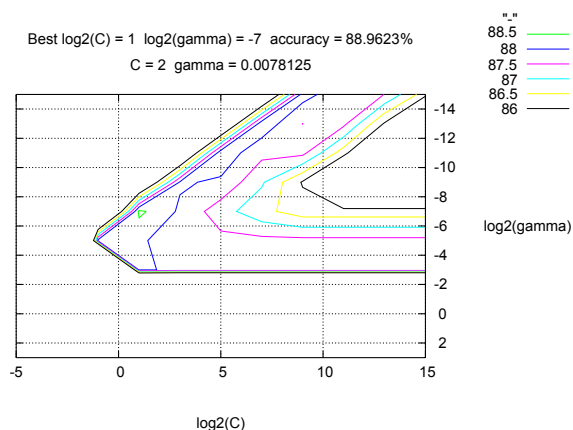


Figura E.19: Grid-search sobre los parámetros C y γ mediante el empleo de la técnica de cross-validation para el operador CS-LBP sobre la ROI utilizando el kernel de tipo RBF correspondiente a la SVM para la detección de ausencia del conductor

Referencias

- Abtahi, S., Omidyeganeh, M., Shirmohammadi, S., Hari, B., 2014. Yawdd: a yawning detection dataset. In: Proceedings of the 5th ACM Multimedia Systems Conference. ACM, pp. 24–28.
- Ahlstrom, C., Dukic, T., 2010. Comparison of eye tracking systems with one and three cameras. In: Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research. ACM, p. 3.
- Ahonen, T., Hadid, A., Pietikainen, M., 2006. Face description with local binary patterns: Application to face recognition. IEEE transactions on pattern analysis and machine intelligence 28 (12), 2037–2041.
- Asthana, A., Marks, T. K., Jones, M. J., Tieu, K. H., Rohith, M., 2011. Fully automatic pose-invariant face recognition via 3d pose normalization. In: 2011 International Conference on Computer Vision. IEEE, pp. 937–944.
- Berri, R. A., Silva, A. G., Parpinelli, R. S., Girardi, E., Arthur, R., 2014. A pattern recognition system for detecting use of mobile phones while driving. In: Computer Vision Theory and Applications (VISAPP), 2014 International Conference on. Vol. 2. IEEE, pp. 411–418.
- Bolme, D. S., Draper, B. A., Beveridge, J. R., 2009. Average of synthetic exact filters. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, pp. 2105–2112.
- Boyraz, P., Yang, X., Hansen, J. H., 2012. Computer vision systems for context-aware active vehicle safety and driver assistance. In: Digital Signal Processing for In-Vehicle Systems and Safety. Springer, pp. 217–227.
- Chang, C.-C., Lin, C.-J., 2011. Libsvm: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST) 2 (3), 27.
- Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). Vol. 1. IEEE, pp. 886–893.
- Daniluk, M., Rezaei, M., Nicolescu, R., Klette, R., 2014. Eye status based on eyelid detection: A driver assistance system. In: International Conference on Computer Vision and Graphics. Springer, pp. 171–178.
- Dasgupta, A., George, A., Happy, S., Routray, A., Shanker, T., 2013. An on-board vision based system for drowsiness detection in automotive drivers. International Journal of Advances in Engineering Sciences and Applied Mathematics 5 (2-3), 94–103.
- Devi, M. S., Bajaj, P. R., 2008. Driver fatigue detection based on eye tracking. In: 2008 First International Conference on Emerging Trends in Engineering and Technology. IEEE, pp. 649–652.
- Dinges, D. F., Grace, R., 1998. Perclos: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance. US Department of Transportation, Federal Highway Administration, Publication Number FHWA-MCRT-98-006.
- Dong, Y., Hu, Z., Uchimura, K., Murayama, N., 2011. Driver inattention monitoring system for intelligent vehicles: A review. IEEE transactions on intelligent transportation systems 12 (2), 596–614.
- Fernández, A., Carús, J., Usamentiaga, R., Alvarez, E., Casado, R., 2017. Wearable and ambient sensors to health monitoring using computer vision and signal processing techniques. Journal of Networks In press.
- Fernández, A., Carús, J. L., Usamentiaga, R., Alvarez, E., Casado, R., 2015a. Unobtrusive health monitoring system using video-based physiological information and activity measurements. In: Computer, Information and Telecommunication Systems (CITS), 2015 International Conference on. IEEE, pp. 1–5.
- Fernández, A., Casado, R., Usamentiaga, R., 2015b. A real-time big data architecture for glasses detection using computer vision techniques. In: Future Internet of Things and Cloud (FiCloud), 2015 3rd International Conference on. IEEE, pp. 591–596.
- Fernández, A., García, R., Usamentiaga, R., Casado, R., 2015c. Glasses detection on real images based on robust alignment. Machine Vision and Applications 26 (4), 519–531.
- Fernández, A., Usamentiaga, R., Carús, J. L., Casado, R., 2016. Driver distraction using visual-based sensors and algorithms. Sensors 16 (11), 1805.
- Flores, M. J., Armingol, J. M., de la Escalera, A., 2010. Real-time warning system for driver drowsiness detection using visual information. Journal of Intelligent & Robotic Systems 59 (2), 103–125.
- Flores, M. J., de la Escalera, A., et al., 2011. Sistema avanzado de asistencia a la conducción para la detección de la somnolencia. Revista Iberoamericana de Automática e Informática Industrial RIAI 8 (3), 216–228.
- Forsman, P. M., Vila, B. J., Short, R. A., Mott, C. G., Van Dongen, H. P., 2013. Efficient driver drowsiness detection at moderate levels of drowsiness. Accident Analysis & Prevention 50, 341–350.
- Hadid, A., Pietikainen, M., 2013. Demographic classification from face videos using manifold learning. Neurocomputing 100, 197–205.
- Hammoud, R. I., Wilhelm, A., Malawey, P., Witt, G. J., 2005. Efficient real-time algorithms for eye state and head pose tracking in advanced driver support systems. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 2. IEEE, pp. 1181–vol.
- Hansen, D. W., Ji, Q., 2010. In the eye of the beholder: A survey of models for eyes and gaze. IEEE Transactions on pattern analysis and machine intelligence 32 (3), 478–500.
- Hattori, A., Tokoro, S., Miyashita, M., Tanaka, I., Ohue, K., Uozumi, S., 2006. Development of forward collision warning system using the driver behavioral information. Tech. rep., SAE Technical Paper.
- Heikkilä, M., Pietikainen, M., Schmid, C., 2009. Description of interest regions with local binary patterns. Pattern recognition 42 (3), 425–436.
- Hong, T., Qin, H., 2007. Drivers drowsiness detection in embedded system. In: Vehicular Electronics and Safety, 2007. ICVES. IEEE International Conference on. IEEE, pp. 1–5.
- Hsu, C.-W., Chang, C.-C., Lin, C.-J., et al., 2003. A practical guide to support vector classification.
- Jain, V., Learned-Miller, E. G., 2010. Fddb: A benchmark for face detection in unconstrained settings. UMass Amherst Technical Report.
- Jo, J., Lee, S. J., Park, K. R., Kim, I.-J., Kim, J., 2014. Detecting driver drowsiness using feature-level fusion and user-specific classification. Expert Systems with Applications 41 (4), 1139–1152.
- Jung, J.-Y., Kim, S.-W., Yoo, C.-H., Park, W.-J., Ko, S.-J., 2016. Lbp-ferns-based feature extraction for robust facial recognition. IEEE Transactions on Consumer Electronics 62 (4), 446–453.
- Lee, S. J., Jo, J., Jung, H. G., Park, K. R., Kim, J., 2011. Real-time gaze estima-

- tor based on driver's head orientation for forward collision warning system. *IEEE Transactions on Intelligent Transportation Systems* 12 (1), 254–267.
- Li, H., Lin, Z., Shen, X., Brandt, J., Hua, G., 2015. A convolutional neural network cascade for face detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 5325–5334.
- Liu, C. C., Hosking, S. G., Lenné, M. G., 2009. Predicting driver drowsiness using vehicle measures: Recent insights and future challenges. *Journal of safety research* 40 (4), 239–245.
- López Romero, W. L., 2016. Sistema de control del estado de somnolencia en conductores de vehículos.
- Losada, D. G., López, G. A. R., Acevedo, R. G., Villán, A. F., 2013. Avieut-artificial vision to improve the user experience. In: *New Concepts in Smart Cities: Fostering Public and Private Alliances (SmartMILE)*, 2013 International Conference on. IEEE, pp. 1–6.
- Lu, L., Ning, X., Qian, M., Zhao, Y., 2011. Close eye detected based on synthesized gray projection. In: *Advances in Multimedia, Software Engineering and Computing Vol. 2*. Springer, pp. 345–351.
- Markuš, N., Frljak, M., Pandžić, I. S., Ahlberg, J., Forchheimer, R., 2014. Object detection with pixel intensity comparisons organized in decision trees. *arXiv preprint arXiv:1305.4537*.
- Martin, E., 2006. Breakthrough research on real-world driver behavior released. National Highway Traffic Safety Administration.
- Mbouna, R. O., Kong, S. G., Chun, M.-G., 2013. Visual analysis of eye state and head pose for driver alertness monitoring. *IEEE transactions on intelligent transportation systems* 14 (3), 1462–1469.
- Murphy-Chutorian, E., Trivedi, M. M., 2010. Head pose estimation and augmented reality tracking: An integrated system and evaluation for monitoring driver awareness. *IEEE Transactions on intelligent transportation systems* 11 (2), 300–311.
- Noori, S. M. R., Mikaeili, M., 2016. Driving drowsiness detection using fusion of electroencephalography, electrooculography, and driving quality signals. *Journal of medical signals and sensors* 6 (1), 39.
- Nuevo, J., Bergasa, L. M., Jiménez, P., 2010. Rsmat: Robust simultaneous modeling and tracking. *Pattern Recognition Letters* 31 (16), 2455–2463.
- of Transportation, D., 2016. Pennsylvania driver's manual. <https://goo.gl/XCER8C>, accessed: 2016-09-018.
- Ojala, T., Pietikäinen, M., Harwood, D., 1996. A comparative study of texture measures with classification based on featured distributions. *Pattern recognition* 29 (1), 51–59.
- Ojala, T., Pietikäinen, M., Maenpää, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence* 24 (7), 971–987.
- Organization, W. H., 2016. Global status report on road safety 2015. <http://goo.gl/jMoJ41>, accessed: 2016-07-01.
- Pan, G., Sun, L., Wu, Z., Lao, S., 2007. Eyeblick-based anti-spoofing in face recognition from a generic webcam. In: *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, pp. 1–8.
- Peden, M., Toroyan, T., Krug, E., Iaych, K., et al., 2016. The status of global road safety: The agenda for sustainable development encourages urgent action. *Journal of the Australasian College of Road Safety* 27 (2), 37.
- Phillips, P. J., Moon, H., Rizvi, S. A., Rauss, P. J., 2000. The feret evaluation methodology for face-recognition algorithms. *IEEE Transactions on pattern analysis and machine intelligence* 22 (10), 1090–1104.
- RACE, A. y. I. D., 2016. Los conductores españoles reconocen sufrir más somnolencia al volante que los usuarios europeos. <http://goo.gl/mui9S3>, accessed: 2016-07-01.
- Regan, M. A., Hallett, C., Gordon, C. P., 2011. Driver distraction and driver inattention: Definition, relationship and taxonomy. *Accident Analysis & Prevention* 43 (5), 1771–1781.
- Sahayadhas, A., Sundaraj, K., Murugappan, M., 2012. Detecting driver drowsiness based on sensors: a review. *Sensors* 12 (12), 16937–16953.
- Selvakumar, K., Jerome, J., Rajamani, K., Shankar, N., 2015. Real-time vision based driver drowsiness detection using partial least squares analysis. *Journal of Signal Processing Systems*, 1–12.
- Shan, C., 2012. Learning local binary patterns for gender classification on real-world face images. *Pattern Recognition Letters* 33 (4), 431–437.
- Shan, C., Gong, S., McOwan, P. W., 2009. Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing* 27 (6), 803–816.
- Sigari, M. H., 2009. Driver hypo-vigilance detection based on eyelid behavior. In: *Advances in Pattern Recognition, 2009. ICAPR '09. Seventh International Conference on*. IEEE, pp. 426–429.
- Slawiński, E., Mut, V., Penizzotto, F., 2015. Sistema de alerta al conductor basado en realimentación vibro-táctil. *Revista Iberoamericana de Automática e Informática Industrial RIAI* 12 (1), 36–48.
- Song, F., Tan, X., Chen, S., Zhou, Z.-H., 2013. A literature survey on robust and efficient eye localization in real-life scenarios. *Pattern Recognition* 46 (12), 3157–3173.
- Song, F., Tan, X., Liu, X., Chen, S., 2014. Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients. *Pattern Recognition* 47 (9), 2825–2838.
- StopChatear, 2016. Uso de los smartphones en la conducción. <http://goo.gl/67dvtv>, accessed: 2016-07-01.
- Talbot, R., Fagerlind, H., Morris, A., 2013. Exploring inattention and distraction in the safety net accident causation database. *Accident Analysis & Prevention* 60, 445–455.
- Tan, X., Triggs, B., 2010. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE transactions on image processing* 19 (6), 1635–1650.
- Timm, F., Barth, E., 2011. Accurate eye centre localisation by means of gradients. *VISAPP* 11, 125–130.
- Ufíčá, M., Franc, V., Hlaváč, V., 2012. Detector of facial landmarks learned by the structured output svm. *VISAPP* 12, 547–556.
- Vapnik, V., 1998. *Statistical learning theory* wiley new york google scholar.
- Vicente, F., Huang, Z., Xiong, X., De la Torre, F., Zhang, W., Levi, D., 2015. Driver gaze tracking and eyes off the road detection system. *IEEE Transactions on Intelligent Transportation Systems* 16 (4), 2014–2027.
- Villan, A. F., Candas, J. L. C., Fernandez, R. U., Tejedor, R. C., 2016. Face recognition and spoofing detection system adapted to visually-impaired people. *IEEE Latin America Transactions* 14 (2), 913–921.
- Viola, P., Jones, M. J., 2004. Robust real-time face detection. *International journal of computer vision* 57 (2), 137–154.
- Vural, E., Cetin, M., Ercil, A., Littlewort, G., Bartlett, M., Movellan, J., 2007. Drowsy driver detection through facial movement analysis. In: *International Workshop on Human-Computer Interaction*. Springer, pp. 6–18.
- You, C.-W., Lane, N. D., Chen, F., Wang, R., Chen, Z., Bao, T. J., Montes-de Oca, M., Cheng, Y., Lin, M., Torresani, L., et al., 2013. Carsafe app: alerting drowsy and distracted drivers using dual cameras on smartphones. In: *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM, pp. 13–26.
- Zhang, Z., Zhang, J.-s., 2006. Driver fatigue detection based intelligent vehicle control. In: *18th International Conference on Pattern Recognition (ICPR'06)*. Vol. 2. IEEE, pp. 1262–1265.

Bibliografía

- [1] World Health Organization. Global status report on road safety 2015. <http://goo.gl/jMoJ41>, 2016. Accessed: 2016-07-01.
- [2] Margaret Peden, Tomitza Toroyan, Etienne Krug, Kacem Iaych, et al. The status of global road safety: The agenda for sustainable development encourages urgent action. *Journal of the Australasian College of Road Safety*, 27(2):37, 2016.
- [3] Jane C Stutts, Donald W Reinfurt, Loren Staplin, and Eric A Rodgman. The role of driver distraction in traffic crashes, 2001.
- [4] Thomas A Dingus, Sheila G Klauer, Vicki L Neale, A Petersen, SE Lee, JD Sudweeks, MA Perez, J Hankey, DJ Ramsey, S Gupta, et al. The 100-car naturalistic driving study, phase ii-results of the 100-car field experiment. Technical report, 2006.
- [5] Thomas A Ranney, Elizabeth Mazzae, Riley Garrott, and Michael J Goodman. Nhtsa driver distraction research: Past, present, and future. In *Driver distraction internet forum*, volume 2000, 2000.
- [6] E Donald Sussman, H Bishop, B Madnick, and R Walter. Driver inattention and highway safety. *Transportation Research Record*, 1047:40–48, 1985.
- [7] Jing-Shiarn Wang, Ronald R Knipling, and Michael J Goodman. The role of driver inattention in crashes: New statistics from the 1995 crashworthiness data system. In *40th annual proceedings of the Association for the Advancement of Automotive Medicine*, volume 377, page 392, 1996.
- [8] David W Eby and Lidia P Kostyniuk. Driver distraction and crashes: An assessment of crash databases and review of the literature. *Ann Arbor*, 1001(481):09, 2003.
- [9] Sheila G Klauer, Vicki L Neale, Thomas A Dingus, David Ramsey, and Jeremy Sudweeks. Driver inattention: A contributing factor to crashes and near-crashes. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 49, pages 1922–1926. SAGE Publications, 2005.

- [10] Vicki L Neale, Thomas A Dingus, Sheila G Klauer, Jeremy Sudweeks, and Michael Goodman. An overview of the 100-car naturalistic study and findings. *National Highway Traffic Safety Administration, Paper*, (05-0400), 2005.
- [11] Sheila G Klauer, Thomas A Dingus, Vicki L Neale, Jeremy D Sudweeks, and David J Ramsey. The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data. Technical report, 2006.
- [12] Rebecca L Olson, Richard J Hanowski, Jeffrey S Hickman, and Joseph L Bocanegra. Driver distraction in commercial vehicle operations. Technical report, 2009.
- [13] Ralph H Craft and Brian Preslopsky. Driver distraction and inattention in the usa large truck and national motor vehicle crash causation studies. In *1st International Conference on Driver Distraction and Inattention (DDI 2009)*, 2009.
- [14] Rachel Talbot, Helen Fagerlind, and Andrew Morris. Exploring inattention and distraction in the safetynet accident causation database. *Accident Analysis & Prevention*, 60:445–455, 2013.
- [15] Marc Starnes. Large-truck crash causation study: An initial overview. Technical report, 2006.
- [16] Queensland Gonvernment. 2004 road traffic crashes in queensland. department of transport and main roads, may 2009. 2009.
- [17] Alberto Fernández, Rubén Usamentiaga, Juan Luis Carús, and Rubén Casado. Driver distraction using visual-based sensors and algorithms. *Sensors*, 16(11):1805, 2016.
- [18] Laura Marie Toole. *Crash risk and mobile device use based on fatigue and drowsiness factors in truck drivers*. PhD thesis, Virginia Tech, 2013.
- [19] John D Lee, Kristie L Young, and Michael A Regan. Defining driver distraction. *Driver distraction: Theory, effects, and mitigation*, 13(4):31–40, 2008.
- [20] Michael A Regan, Charlene Hallett, and Craig P Gordon. Driver distraction and driver inattention: Definition, relationship and taxonomy. *Accident Analysis & Prevention*, 43(5):1771–1781, 2011.

-
- [21] ANFABRA y la DGT RACE. Los conductores españoles reconocen sufrir más somnolencia al volante que los usuarios europeos. <http://goo.gl/mui9S3>, 2016. Accessed: 2016-07-01.
- [22] StopChatear. Uso de los smartphones en la conducción. <http://goo.gl/67dvtn>, 2016. Accessed: 2016-07-01.
- [23] DGT. 1.160 fallecidos en accidente de tráfico. <https://goo.gl/HygC2j>, 2017. Accessed: 2017-01-01.
- [24] prensapress. El año 2016 acaba con 1.160 muertos en las carreteras. <https://goo.gl/a821EW>, 2017. Accessed: 2017-01-01.
- [25] Charles C Liu, Simon G Hosking, and Michael G Lenné. Predicting driver drowsiness using vehicle measures: Recent insights and future challenges. *Journal of safety research*, 40(4):239–245, 2009.
- [26] Pia M Forsman, Bryan J Vila, Robert A Short, Christopher G Mott, and Hans PA Van Dongen. Efficient driver drowsiness detection at moderate levels of drowsiness. *Accident Analysis & Prevention*, 50:341–350, 2013.
- [27] Arun Sahayadhas, Kenneth Sundaraj, and Murugappan Murugappan. Detecting driver drowsiness based on sensors: a review. *Sensors*, 12(12):16937–16953, 2012.
- [28] K Selvakumar, Jovitha Jerome, Kumar Rajamani, and Nishanth Shankar. Real-time vision based driver drowsiness detection using partial least squares analysis. *Journal of Signal Processing Systems*, pages 1–12, 2015.
- [29] Jaeik Jo, Sung Joo Lee, Kang Ryoung Park, Ig-Jae Kim, and Jaihie Kim. Detecting driver drowsiness using feature-level fusion and user-specific classification. *Expert Systems with Applications*, 41(4):1139–1152, 2014.
- [30] Emanuel Slawiński, Vicente Mut, and Franco Penizzotto. Sistema de alerta al conductor basado en realimentación vibro-táctil. *Revista Iberoamericana de Automática e Informática Industrial RIAI*, 12(1):36–48, 2015.
- [31] Anirban Dasgupta, Anjith George, SL Happy, Aurobinda Routray, and Tara Shanker. An on-board vision based system for drowsiness detection in automotive drivers. *International Journal of Advances in Engineering Sciences and Applied Mathematics*, 5(2-3):94–103, 2013.
- [32] Esra Vural, Mujdat Cetin, Aytul Ercil, Gwen Littlewort, Marian Bartlett, and Javier Movellan. Drowsy driver detection through facial movement analysis. In *International Workshop on Human-Computer Interaction*, pages 6–18. Springer, 2007.

- [33] Gorka Velez and Oihana Otaegui. Embedded platforms for computer vision-based advanced driver assistance systems: a survey. *arXiv preprint arXiv:1504.07442*, 2015.
- [34] Yanchao Dong, Zhencheng Hu, Keiichi Uchimura, and Nobuki Murayama. Driver inattention monitoring system for intelligent vehicles: A review. *IEEE transactions on intelligent transportation systems*, 12(2):596–614, 2011.
- [35] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7):971–987, 2002.
- [36] Erno Mäkinen and Roope Raisamo. An experimental comparison of gender classification methods. *Pattern Recognition Letters*, 29(10):1544–1556, 2008.
- [37] Hui-Cheng Lian and Bao-Liang Lu. Multi-view gender classification using local binary patterns and support vector machines. In *International Symposium on Neural Networks*, pages 202–209. Springer, 2006.
- [38] Caifeng Shan. Learning local binary patterns for gender classification on real-world face images. *Pattern Recognition Letters*, 33(4):431–437, 2012.
- [39] Zhiguang Yang and Haizhou Ai. Demographic classification with local binary patterns. In *Advances in Biometrics*, pages 464–473. Springer, 2007.
- [40] Jamie R Lyle, Philip E Miller, Shrinivas J Pundlik, and Damon L Woodard. Soft biometric classification using periocular region features. In *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, pages 1–7. IEEE, 2010.
- [41] Caifeng Shan, Shaogang Gong, and Peter W McOwan. Robust facial expression recognition using local binary patterns. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 2, pages II–370. IEEE, 2005.
- [42] Caifeng Shan, Shaogang Gong, and Peter W McOwan. Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*, 27(6):803–816, 2009.
- [43] Damon L Woodard, Shrinivas J Pundlik, Jamie R Lyle, and Philip E Miller. Periocular region appearance cues for biometric identification. In

- Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2010 *IEEE Computer Society Conference on*, pages 162–169. IEEE, 2010.
- [44] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen. Face description with local binary patterns: Application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12):2037–2041, 2006.
- [45] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.
- [46] Daniel Gonzalez Losada, Gustavo Adolfo Rosai Lopez, Rodrigo Garcia Acevedo, and Alberto Fernandez Villan. Aviu-e-artificial vision to improve the user experience. In *New Concepts in Smart Cities: Fostering Public and Private Alliances (SmartMILE)*, 2013 *International Conference on*, pages 1–6. IEEE, 2013.
- [47] Alberto Fernández, Rodrigo García, Rubén Usamentiaga, and Rubén Casado. Glasses detection on real images based on robust alignment. *Machine Vision and Applications*, 26(4):519–531, 2015.
- [48] Xiaoyi Jiang, Michael Binkert, Bernard Achermann, and H Bunke. Towards detection of glasses in facial images. *Pattern Analysis & Applications*, 3(1):9–18, 2000.
- [49] Antitza Dantcheva, Carmelo Velardo, Angela D’angelo, and Jean-Luc Dugelay. Bag of soft biometrics for person identification. *Multimedia Tools and Applications*, 51(2):739–777, 2011.
- [50] Zhong Jing and Robert Mariani. Glasses detection and extraction by deformable contour. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 2, pages 933–936. IEEE, 2000.
- [51] Daniel A Vaquero, Rogério Schmidt Feris, Duan Tran, Lisa Brown, Arun Hampapur, and Matthew Turk. Attribute-based people search in surveillance environments. In *Applications of Computer Vision (WACV)*, 2009 *Workshop on*, pages 1–8. IEEE, 2009.
- [52] Bo Wu, Haizhou Ai, and Ran Liu. Glasses detection by boosting simple wavelet features. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 1, pages 292–295. IEEE, 2004.

- [53] Haiyuan Wu, Genki Yoshikawa, Tadayoshi Shioyama, T Lao, and T Kawade. Glasses frame detection with 3d hough transform. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 2, pages 346–349. IEEE, 2002.
- [54] Shiguang Shan, Wen Gao, Yan Lu, Bo Cao, Xilin Chen, Debin Zhao, and Wenbin Zeng. Unified framework for classifying facial images based on facial attribute-specific subspaces and minimum reconstruction error. In *Proc. of ACCV*, volume 2. Citeseer, 2002.
- [55] Chenyu Wu, Ce Liu, Heung-Yueng Shum, Ying-Qing Xy, and Zhengyou Zhang. Automatic eyeglasses removal from face images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(3):322–336, 2004.
- [56] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages 1–511. IEEE, 2001.
- [57] Jingu Heo, Seong G Kong, Bisma R Abidi, and Mongi A Abidi. Fusion of visual and thermal signatures with eyeglass removal for robust face recognition. In *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on*, pages 122–122. IEEE, 2004.
- [58] Yang Zhao, Wei Jia, Rong-Xiang Hu, and Hai Min. Completed robust local binary pattern for texture classification. *Neurocomputing*, 106:68–76, 2013.
- [59] Patrick Lucey, Jeffrey F Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 94–101. IEEE, 2010.
- [60] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.
- [61] Neeraj Kumar, Alexander C Berg, Peter N Belhumeur, and Shree K Nayar. Attribute and simile classifiers for face verification. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 365–372. IEEE, 2009.
- [62] Wilman W Zou, Pong C Yuen, and Rama Chellappa. Low-resolution face tracker robust to illumination variations. *Image Processing, IEEE Transactions on*, 22(5):1726–1739, 2013.

- [63] P Jonathon Phillips, Hyeonjoon Moon, Syed A Rizvi, and Patrick J Rauss. The feret evaluation methodology for face-recognition algorithms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(10):1090–1104, 2000.
- [64] Bruce Mehler, Bryan Reimer, and Joseph F Coughlin. Physiological reactivity to graded levels of cognitive workload across three age groups: An on-road evaluation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 54, pages 2062–2066. SAGE Publications, 2010.
- [65] Shing-Hong Liu, Chin-Teng Lin, and Wen-Hung Chao. The short-time fractal scaling of heart rate variability to estimate the mental stress of driver. In *Networking, Sensing and Control, 2004 IEEE International Conference on*, volume 2, pages 829–833. IEEE, 2004.
- [66] LJM Mulder. Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological psychology*, 34(2):205–236, 1992.
- [67] Erika E Miller. *Effects of Roadway on Driver Stress: An On-Road Study using Physiological Measures*. PhD thesis, University of Washington, 2013.
- [68] Dale L Partin, Michel F Sultan, Christopher M Thrush, Ray Prieto, and Steve J Wagner. Monitoring driver physiological parameters for improved safety. Technical report, SAE Technical Paper, 2006.
- [69] Chunlin Zhao, Min Zhao, Jianpin Liu, and Chongxun Zheng. Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator. *Accident Analysis & Prevention*, 45:83–90, 2012.
- [70] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics express*, 18(10):10762–10774, 2010.
- [71] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE transactions on biomedical engineering*, 58(1):7–11, 2011.
- [72] Alberto Fernández, Juan Luis Carús, Rubén Usamentiaga, Eduardo Alvarez, and Rubén Casado. Unobtrusive health monitoring system using video-based physiological information and activity measurements. In *Computer, Information and Telecommunication Systems (CITS), 2015 International Conference on*, pages 1–5. IEEE, 2015.

- [73] Magdalena Lewandowska, Jacek Rumiński, Tomasz Kocejko, and Jędrzej Nowak. Measuring pulse rate with a webcam-a non-contact method for evaluating cardiac activity. In *Computer Science and Information Systems (FedCSIS), 2011 Federated Conference on*, pages 405–410. IEEE, 2011.
- [74] Alexander Trumpp, Hagen Malberg, and Sebastian Zaunseder. Signal extraction in camera-based photoplethysmography using a modified wiener filter. 2014.
- [75] Philipp V Rouast, Marc TP Adam, Verena Dorner, and Ewa Lux. Remote photoplethysmography: Evaluation of contactless heart rate measurement in an information systems setting. 2016.
- [76] Haiying Xia, Zhouxiao Bao, and Haomiao Jiang. Mobile cardiac pulse measurements. *Electrical Engineering*, 2012.
- [77] Gerard de Haan and Vincent Jeanne. Robust pulse rate from chrominance-based rppg. *IEEE Transactions on Biomedical Engineering*, 60(10):2878–2886, 2013.
- [78] Guha Balakrishnan, Fredo Durand, and John Guttag. Detecting pulse from head motions in video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3430–3437, 2013.
- [79] Xiaobai Li, Jie Chen, Guoying Zhao, and Matti Pietikainen. Remote heart rate measurement from face videos under realistic situations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4264–4271, 2014.
- [80] A. Fernandez, J.L. Carus, R. Usamentiaga, E. Alvarez, and R. Casado. Wearable and ambient sensors to health monitoring using computer vision and signal processing techniques. *Journal of Networks*, In press., 2017.
- [81] Dongseok Lee, Jeehoon Kim, Sungjun Kwon, and Kwangsuk Park. Heart rate estimation from facial photoplethysmography during dynamic illuminance changes. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 2758–2761. IEEE, 2015.
- [82] Vincent Jeanne, Michel Asselman, Bert den Brinker, and Murtaza Bulut. Camera-based heart rate monitoring in highly dynamic light conditions. In *Connected Vehicles and Expo (ICCVE), 2013 International Conference on*, pages 798–799. IEEE, 2013.

- [83] Alberto Fernández, Rubén Casado, and Rubén Usamentiaga. A real-time big data architecture for glasses detection using computer vision techniques. In *Future Internet of Things and Cloud (FiCloud), 2015 3rd International Conference on*, pages 591–596. IEEE, 2015.
- [84] Bert Taube. Leveraging big data and real-time analytics to achieve situational awareness for smart grids.
- [85] E Dumbil. What is big data? an introduction to the big data landscape, 2012.
- [86] Rubén Casado and Muhammad Younas. Emerging trends and technologies in big data processing. *Concurrency and Computation: Practice and Experience*, 27(8):2078–2091, 2015.
- [87] George Bebis, Richard Boyle, Bahram Parvin, Darko Koracin, Ryan McMahan, Jason Jerald, Hui Zhang, Steven Drucker, Kambhamettu Chandra, Zhigang Deng, et al. *Advances in Visual Computing: 10th International Symposium, ISVC 2014, Las Vegas, NV, USA, December 8-10, 2014, Proceedings*, volume 8887. Springer, 2014.
- [88] Amir Gandomi and Murtaza Haider. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2):137–144, 2015.
- [89] Tolga Soyata, Rajani Muraleedharan, Colin Funai, Minseok Kwon, and Wendi Heinzelman. Cloud-vision: Real-time face recognition using a mobile-cloudlet-cloud acceleration architecture. In *Computers and Communications (ISCC), 2012 IEEE Symposium on*, pages 000059–000066. IEEE, 2012.
- [90] A Vinay, Vinay S Shekhar, J Rituparna, Tushar Aggrawal, KN Balasubramanya Murthy, and S Natarajan. Cloud based big data analytics framework for face recognition in social networks using machine learning. *Procedia Computer Science*, 50:623–630, 2015.
- [91] Ankit Jain and Anand Nalya. *Learning storm*. Packt Publishing, 2014.
- [92] Muhammad Hussain Iqbal and Tariq Rahim Soomro. Big data analysis: Apache storm perspective. *International Journal of Computer Trends and Technology*, pages 9–14, 2015.
- [93] Alberto Fernandez Villan, Juan Luis Carus Candas, Ruben Usamentiaga Fernandez, and Ruben Casado Tejedor. Face recognition and spoofing detection system adapted to visually-impaired people. *IEEE Latin America Transactions*, 14(2):913–921, 2016.

- [94] Shachar Maidenbaum, Shlomi Hanassy, Sami Abboud, Galit Buchs, Daniel-Robert Chebat, Shelly Levy-Tzedek, and Amir Amedi. The eyecane, a new electronic travel aid for the blind: technology, behavior & swift learning. *Restorative neurology and neuroscience*, 32(6):813–824, 2014.
- [95] KM Kramer, DS Hedin, and DJ Rolkosky. Smartphone based face recognition tool for the blind. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 4538–4541. IEEE, 2010.
- [96] Luigi Balduzzi, Giovanni Fusco, Francesca Odone, S Dini, Marco Mesiti, Augusto Destrero, and Alberto Lovato. Low-cost face biometry for visually impaired users. In *Biometric Measurements and Systems for Security and Medical Applications (BIOMS), 2010 IEEE Workshop on*, pages 45–52. IEEE, 2010.
- [97] Xiaoyang Tan and Bill Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE transactions on image processing*, 19(6):1635–1650, 2010.
- [98] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.
- [99] Jukka Määttä, Abdenour Hadid, and Matti Pietikäinen. Face spoofing detection from single images using micro-texture analysis. In *Biometrics (IJCB), 2011 international joint conference on*, pages 1–7. IEEE, 2011.
- [100] Ivana Chingovska, André Anjos, and Sébastien Marcel. On the effectiveness of local binary patterns in face anti-spoofing. In *Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG-Proceedings of the International Conference of the*, pages 1–7. IEEE, 2012.
- [101] Azeddine Benlamoudi, Djamel Samai, Abdelkrim Ouafi, Abdelmalik Taleb-Ahmed, Salah Eddine Bekhouche, and Abdenour Hadid. Face spoofing detection from single images using active shape models with stasm and lbp. In *Proceeding of the Troisième CONFERENCE INTERNATIONALE SUR LA VISION ARTIFICIELLE CVA 2015*, 2015.
- [102] Azeddine Benlamoudi, Djamel Samai, Abdelkrim Ouafi, Salah Eddine Bekhouche, Abdelmalik Taleb-Ahmed, and Abdenour Hadid. Face spoofing detection using local binary patterns and fisher score. In *Control, Engineering & Information Technology (CEIT), 2015 3rd International Conference on*, pages 1–5. IEEE, 2015.

- [103] Tiago de Freitas Pereira, André Anjos, José Mario De Martino, and Sébastien Marcel. Can face anti-spoofing countermeasures work in a real world scenario? In *Biometrics (ICB), 2013 International Conference on*, pages 1–8. IEEE, 2013.
- [104] Sujitha Martin, Ashish Tawari, and Mohan Manubhai Trivedi. Toward privacy-protecting safety systems for naturalistic driving videos. *IEEE Transactions on Intelligent Transportation Systems*, 15(4):1811–1822, 2014.
- [105] Elaine M Newton, Latanya Sweeney, and Bradley Malin. Preserving privacy by de-identifying face images. *IEEE transactions on Knowledge and Data Engineering*, 17(2):232–243, 2005.
- [106] CH Zhao, BL Zhang, Jinwei He, and Jie Lian. Recognition of driving postures by contourlet transform and random forests. *Intelligent Transport Systems, IET*, 6(2):161–168, 2012.
- [107] Chihang Zhao, Yongsheng Gao, Jie He, and Jie Lian. Recognition of driving postures by multiwavelet transform and multilayer perceptron classifier. *Engineering Applications of Artificial Intelligence*, 25(8):1677–1686, 2012.
- [108] Chihang H Zhao, Bailing L Zhang, Xiaozheng Z Zhang, Sanqiang Q Zhao, and Hanxi X Li. Recognition of driving postures by combined features and random subspace ensemble of multilayer perceptron classifiers. *Neural Computing and Applications*, 22(1):175–184, 2013.
- [109] Chihang Zhao, Bailing Zhang, and Jie He. Vision-based classification of driving postures by efficient feature extraction and bayesian approach. *Journal of Intelligent & Robotic Systems*, 72(3-4):483–495, 2013.
- [110] Anna Bosch, Andrew Zisserman, and Xavier Munoz. Representing shape with a spatial pyramid kernel. In *Proceedings of the 6th ACM international conference on Image and video retrieval*, pages 401–408. ACM, 2007.
- [111] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.
- [112] Chao Yan, Frans Coenen, and Bailing Zhang. Driving posture recognition by joint application of motion history image and pyramid histogram of oriented gradients. *International journal of vehicular technology*, 2014, 2014.

- [113] Chao Yan, Frans Coenen, and Bailing Zhang. Driving posture recognition by convolutional neural networks. *IET Computer Vision*, 10(2):103–114, 2016.
- [114] Xuetao Zhang, Nanning Zheng, Fei Wang, and Yongjian He. Visual recognition of driver hand-held cell phone use based on hidden crf. In *Vehicular Electronics and Safety (ICVES), 2011 IEEE International Conference on*, pages 248–251. IEEE, 2011.
- [115] Yusuf Artan, Orhan Bulan, Robert P Loce, and Peter Paul. Driver cell phone usage detection from hov/hot nir images. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on*, pages 225–230. IEEE, 2014.
- [116] Rafael A Berri, Alexandre G Silva, Rafael S Parpinelli, Elaine Girardi, and Rangel Arthur. A pattern recognition system for detecting use of mobile phones while driving. *arXiv preprint arXiv:1408.0680*, 2014.
- [117] Beilei Xu and Robert P Loce. A machine learning approach for detecting cell phone usage. In *IS&T/SPIE Electronic Imaging*, pages 94070A–94070A. International Society for Optics and Photonics, 2015.
- [118] Keshav Seshadri, Felix Juefei-Xu, Dipan K Pal, Marios Savvides, and Craig P Thor. Driver cell phone usage detection on strategic highway research program (shrp2) face view videos. 2015.
- [119] Eshed Ohn-Bar and Mohan M Trivedi. The power is in your hands: 3d analysis of hand gestures in naturalistic video. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*, pages 912–917. IEEE, 2013.
- [120] Sebastien Martin, Eshed Ohn-Bar, Ashish Tawari, and Mohan Manubhai Trivedi. Understanding head and hand activities and coordination in naturalistic driving videos. In *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, pages 884–889. IEEE, 2014.
- [121] Eshed Ohn-Bar, Sujitha Martin, Ashish Tawari, and M Trivedi. Head, eye, and hand patterns for driver activity recognition. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pages 660–665. IEEE, 2014.
- [122] Luis M Bergasa, Jesús Nuevo, Miguel Sotelo, Rafael Barea, María Elena Lopez, et al. Real-time system for monitoring driver vigilance. *Intelligent Transportation Systems, IEEE Transactions on*, 7(1):63–77, 2006.

- [123] Sung Joo Lee, Jaeik Jo, Ho Gi Jung, Kang Ryoung Park, and Jaihie Kim. Real-time gaze estimator based on driver's head orientation for forward collision warning system. *Intelligent Transportation Systems, IEEE Transactions on*, 12(1):254–267, 2011.
- [124] F. Vicente, Zehua Huang, Xuehan Xiong, F. de la Torre, Wende Zhang, and D. Levi. Driver gaze tracking and eyes off the road detection system. *Intelligent Transportation Systems, IEEE Transactions on*, 16(4):2014–2027, 2015.
- [125] Bogusław Cyganek and Sławomir Gruszczyński. Hybrid computer vision system for drivers' eye recognition and fatigue monitoring. *Neurocomputing*, 126:78–94, 2014.
- [126] Birsen Donmez, Linda Ng Boyle, and John D Lee. Mitigating driver distraction with retrospective and concurrent feedback. *Accident Analysis & Prevention*, 40(2):776–786, 2008.
- [127] Katja Kircher, Albert Kircher, and Christer Ahlström. Results of a field study on a driver distraction warning system. 2009.
- [128] Katja Kircher, Christer Ahlstrom, and Albert Kircher. Comparison of two eye-gaze based real-time driver distraction detection algorithms in a small-scale field operational test. In *Proc. 5th Int. Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, pages 16–23, 2009.
- [129] Katja Kircher, Albert Kircher, and Fredrich Claezon. Distraction and drowsiness—a field study. *Linköping, Sweden: VTI, Swedish National Road and Transport Research Institute*, 2009.
- [130] T Victor. The victor and larsson (2010) distraction detection algorithm and warning strategy. *Volvo Technology*, 1(0.8):0–6, 2010.
- [131] Yilu Zhang, Yuri Owechko, and Jing Zhang. Driver cognitive workload estimation: A data-driven perspective. In *Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference on*, pages 642–647. IEEE, 2004.
- [132] Yulan Liang, Michelle L Reyes, and John D Lee. Real-time detection of driver cognitive distraction using support vector machines. *Intelligent Transportation Systems, IEEE Transactions on*, 8(2):340–350, 2007.
- [133] Yulan Liang, John D Lee, and Michelle L Reyes. Nonintrusive detection of driver cognitive distraction in real time using bayesian networks. *Transportation Research Record: Journal of the Transportation Research Board*, 2018(1):1–8, 2007.

- [134] Masahiro Miyaji, Haruki Kawanaka, and Koji Oguri. Effect of pattern recognition features on detection for driver's cognitive distraction. In *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*, pages 605–610. IEEE, 2010.
- [135] Yan Yang, Haoqi Sun, Tianchi Liu, Guang-Bin Huang, and Olga Sourina. Driver workload detection in on-road driving environment using machine learning. In *Proceedings of ELM-2014 Volume 2*, pages 389–398. Springer, 2015.
- [136] T. Liu, Y. Yang, G.-B. Huang, Y.K. Yeo, and Z. Lin. Driver distraction detection using semi-supervised machine learning. *Intelligent Transportation Systems, IEEE Transactions on*, PP(99):1–13, 2015.
- [137] Nanxiang Li and Carlos Busso. Analysis of facial features of drivers under cognitive and visual distractions. In *Multimedia and Expo (ICME), 2013 IEEE International Conference on*, pages 1–6. IEEE, 2013.
- [138] Céline Craye and Fakhri Karray. Driver distraction detection and recognition using rgb-d sensor. *arXiv preprint arXiv:1502.00250*, 2015.
- [139] Tianchi Liu, Yan Yang, Guang-Bin Huang, Zhiping Lin, Felix Klanner, Cornelia Denk, and Ralph H Raschhofer. Cluster regularized extreme learning machine for detecting mixed-type distraction in driving. In *Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference on*, pages 1323–1326. IEEE, 2015.
- [140] Amira Ragab, Celine Craye, Mohamed S Kamel, and Fakhri Karray. A visual-based driver distraction recognition and detection using random forest. In *Image Analysis and Recognition*, pages 256–265. Springer, 2014.
- [141] A. Fernandez, R. Usamentiaga, and R. Casado. Automatic system to detect both distraction and drowsiness in drivers using robust visual features. *RIAI: Revista Iberoamericana de Automatica e Informatica industrial*, In press., 2017.
- [142] Fengyi Song, Xiaoyang Tan, Xue Liu, and Songcan Chen. Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients. *Pattern Recognition*, 47(9):2825–2838, 2014.
- [143] Francisco Vicente, Zehua Huang, Xuehan Xiong, Fernando De la Torre, Wende Zhang, and Dan Levi. Driver gaze tracking and eyes off the road detection system. *IEEE Transactions on Intelligent Transportation Systems*, 16(4):2014–2027, 2015.

-
- [144] Shabnam Abtahi, Mona Omidyeganeh, Shervin Shirmohammadi, and Behnoosh Hariri. Yawdd: a yawning detection dataset. In *Proceedings of the 5th ACM Multimedia Systems Conference*, pages 24–28. ACM, 2014.
- [145] Wilfrido Leonel López Romero. Sistema de control del estado de somnolencia en conductores de vehículos. 2016.
- [146] Akira Hattori, Setsuo Tokoro, Michihiro Miyashita, Isahiko Tanaka, Kenichi Ohue, and Shigeyasu Uozumi. Development of forward collision warning system using the driver behavioral information. Technical report, SAE Technical Paper, 2006.
- [147] Mohamad Hoseyn Sigari. Driver hypo-vigilance detection based on eyelid behavior. In *Advances in Pattern Recognition, 2009. ICAPR'09. Seventh International Conference on*, pages 426–429. IEEE, 2009.
- [148] Chuang-Wen You, Nicholas D Lane, Fanglin Chen, Rui Wang, Zhenyu Chen, Thomas J Bao, Martha Montes-de Oca, Yuting Cheng, Mu Lin, Lorenzo Torresani, et al. Carsafe app: alerting drowsy and distracted drivers using dual cameras on smartphones. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*, pages 13–26. ACM, 2013.
- [149] Marco Javier Flores, José María Armingol, and Arturo de la Escalera. Real-time warning system for driver drowsiness detection using visual information. *Journal of Intelligent & Robotic Systems*, 59(2):103–125, 2010.
- [150] Marco J Flores, Arturo de la Escalera, et al. Sistema avanzado de asistencia a la conducción para la detección de la somnolencia. *Revista Iberoamericana de Automática e Informática Industrial RIAI*, 8(3):216–228, 2011.
- [151] Tianyi Hong and Huabiao Qin. Drivers drowsiness detection in embedded system. In *Vehicular Electronics and Safety, 2007. ICVES. IEEE International Conference on*, pages 1–5. IEEE, 2007.
- [152] Juan Carmona, Fernando García, David Martín, Arturo de la Escalera, and José María Armingol. Data fusion for driver behaviour analysis. *Sensors*, 15(10):25968–25991, 2015.
- [153] Iván G Daza, Luis M Bergasa, Sebastián Bronte, J Javier Yebes, Javier Almazán, and Roberto Arroyo. Fusion of optimized indicators from advanced driver assistance systems (adas) for driver drowsiness detection. *Sensors*, 14(1):1106–1131, 2014.

- [154] Nicholas Apostoloff and Alexander Zelinsky. Vision in and out of vehicles: Integrated driver and road scene monitoring. *The International Journal of Robotics Research*, 23(4-5):513–538, 2004.
- [155] Takatsugu Hirayama, Kenji Mase, and Kazuya Takeda. Analysis of temporal relationships between eye gaze and peripheral vehicle behavior for detecting driver distraction. *International journal of vehicular technology*, 2013, 2013.
- [156] Eshed Ohn-Bar, Ashish Tawari, Sujitha Martin, and Mohan M Trivedi. Vision on wheels: Looking at driver, vehicle, and surround for on-road maneuver analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 185–190, 2014.
- [157] Ashish Tawari, Sayanan Sivaraman, Mohan Manubhai Trivedi, Trevor Shannon, and Mario Toppelhofer. Looking-in and looking-out vision for urban intelligent assistance: Estimation of driver attentive state and dynamic surround for safe merging and braking. In *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, pages 115–120. IEEE, 2014.
- [158] Sujitha Martin, Akshay Rangesh, Eshed Ohn-Bar, and Mohan M Trivedi. The rhythms of head, eyes and hands at intersections. In *Intelligent Vehicles Symposium (IV), 2016 IEEE*, pages 1410–1415. IEEE, 2016.
- [159] Ashesh Jain, Hema S Koppula, Bharad Raghavan, Shane Soh, and Ashutosh Saxena. Car that knows before you do: Anticipating maneuvers via learning temporal driving models. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3182–3190, 2015.
- [160] CTIS. Índice aceptación congreso ctis 2015. <https://goo.gl/V7aHgw>, 2015. Accessed: 2017-01-01.