



Universidad Autónoma de Querétaro
Facultad de Ingeniería
Maestría en Instrumentación y Control Automático

Sistema embebido de bajo costo para la asistencia al conductor mediante procesamiento de expresiones faciales

TESIS

Que como parte de los requisitos para obtener el grado de
Maestría en Instrumentación y Control Automático

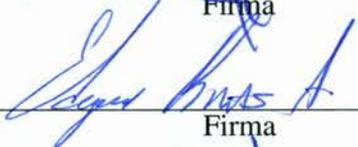
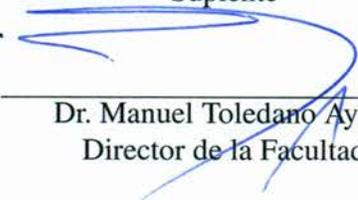
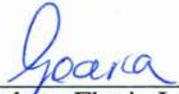
Presenta:

Ing. Cinthia Viviana Alcántara Montiel

Dirigido por:

Dr. Jesús Carlos Pedraza Ortega

SINODALES

Dr. Jesús Carlos Pedraza Ortega Presidente	 Firma
Dr. Juvenal Rodríguez Reséndiz Secretario	 Firma
Dr. Edgar Alejandro Rivas Araiza Vocal	 Firma
Dr. Juan Manuel Ramos Arreguín Suplente	 Firma
Dr. Efrén Gorrostieta Hurtado Suplente	 Firma
 Dr. Manuel Toledano Ayala Director de la Facultad	 Dra. Ma. Guadalupe Flavia Loarca Piña Director de Investigación y Posgrado

RESUMEN

El objetivo de la presente tesis es abordar el desarrollo de un algoritmo para la detección de somnolencia, mediante la aplicación de varias técnicas y conceptos de visión por computadora, el cual pueda ser empleado en sistemas para la asistencia a conductores. Para alcanzar dicho objetivo se desarrolló un detector facial basado en descriptores HOG y clasificadores en cascada. Los primeros se seleccionaron debido a su capacidad para generar descriptores robustos, los segundos fueron utilizados debido a su eficiencia en el entrenamiento de detectores de objetos.

Se aplicó un detector de puntos relevantes que permitiera el análisis de las expresiones faciales que pudieran indicar la presencia de somnolencia en el usuario y, con base en ello, decidir si ejecutar o no el sistema de alarma. Finalmente el algoritmo desarrollado se implementó en un centro de procesamiento de una tarjeta que, conectado a los periféricos necesarios para captura y reproducción de audio y vídeo, resulta en un sistema portable para la detección de somnolencia.

Este documento se divide en 5 capítulos más una sección de anexos, los cuales se describirán brevemente a continuación; en el capítulo 1 se presentan la descripción del problema y algunas estadísticas del problema social que podría atacarse con la apropiada implementación del algoritmo desarrollado, finalmente en el mismo capítulo se define la hipótesis y los objetivos de este proyecto.

En el capítulo 2 trata sobre los sistemas avanzados de asistencia al conductor (ADAS) y su relación con el procesamiento de imágenes, e incluye el estado del arte de los sistemas de detección de fatiga. Posteriormente, en el capítulo 3, se abordan los conceptos de detección de objetos que resultaran indispensables para el entendimiento del desarrollo del algoritmo del que es objeto este documento.

La metodología utilizada se describe en el capítulo 4. El capítulo 5 muestra los resultados de las pruebas realizadas una vez concluida la implementación del algoritmo de detección de objetos y se enlistan las conclusiones obtenidas tras el análisis de los resultados de la experimentación. Por último, en los «Anexos», se incluyen secciones de código utilizadas a lo largo del desarrollo del algoritmo.

Palabras clave: *Visión por computadora, procesamiento de imágenes, detección de objetos, ADAS, somnolencia.*

SUMMARY

The aim of this work is to describe the process to develop an algorithm for drowsiness detection, applied on driver assistance systems, by using computer vision techniques. In order to develop such algorithm, a face detector was created based on HOG descriptors and cascade classifiers; by combine such descriptors and classifiers it is possible get a robust face detector which can detect faces on different positions into sceneries which can vary on lighting.

Next, a landmarks detector was applied, its goal is to estimate the location of the facial landmarks and build data vectors that describes the facial components; eyes, chin, mouth, among others. In such a way it could be used to analyze facial expressions and define whether or not there are signs of drowsiness on the user.

This document is divided on five chapters plus one annex section; on first chapter, it is shown the information about the existent social problem that inspires this work and, also, the hypothesis and the goals for this thesis are set.

On chapter 2, ADAS are studied and also its relationship with computer vision and image processing is explained, to finally abord some works and concepts related to drowsiness detection. While in chapter 3, is all about object detection, concepts shown on this chapter are going to be very important all along the algorithm development.

Chapter 4 is about the proposed methodology and the way that every step of it helps fulfill the thesis goals. Then, on chapter 5, experimentation and results are discussed to finally expose the conclusions of this work. Finally, on annex section, fragments of code used along the algorithm development are shown.

Keywords: *Computer vision, image processing, object detection, ADAS, drowsiness.*

ÍNDICE GENERAL

RESUMEN	I
SUMMARY	II
ÍNDICE GENERAL	IV
ÍNDICE DE CUADROS	V
ÍNDICE DE FIGURAS	VIII
1. INTRODUCCIÓN	1
1.1. Motivación y descripción del problema	2
1.2. Hipótesis	6
1.3. Objetivos	6
1.3.1. Objetivos paticulares	6
2. ESTADO DEL ARTE	8
2.1. Estudio de la fatiga	8
2.1.1. Definición	8
2.1.2. Principales causas	9
2.1.3. Técnicas de medición de fatiga en conductores	10
2.2. Indicadores visuales de fatiga	15
2.3. Antecedentes del procesamiento de imágenes	17
2.4. ADAS	20
2.4.1. ADAS y su implementación actual	21
2.5. Sistema de detección de fatiga y distracción	30
2.6. Estudios relacionados	36
3. FUNDAMENTOS TEÓRICOS	39
3.1. Detección de objetos	39
3.1.1. Métodos basados en características	41
3.1.2. Métodos basados en plantillas	41
3.1.3. Métodos basados en clasificadores	42
3.1.4. Métodos basados en movimiento	42
3.2. Detección de rostro	43
3.2.1. Técnicas de detección	44
3.3. Características	51
3.3.1. Detectores de características	52
3.3.2. Descriptores de características	53
3.4. Clasificadores	55
3.4.1. Técnicas de aprendizaje	56
3.4.2. Técnicas de clasificación	57

4. METODOLOGÍA	59
4.1. Captura de Vídeo	60
4.2. Detección de rostro	64
4.2.1. Pruebas preliminares	64
4.2.2. Descriptores HOG en detección de rostro	65
4.2.3. Clasificadores en cascada	72
4.2.4. Entrenamiento del detector de rostro	73
4.2.5. Experimentación y discusión de resultados	78
4.3. Identificación de puntos relevantes en el rostro	83
4.4. Detección de indicadores visuales que impliquen fatiga o distracción	84
4.4.1. Apertura de ojos	84
4.4.2. Detección de bostezo	85
4.5. Procesamiento de los hallazgos.	86
5. EXPERIMENTACIÓN Y RESULTADOS	88
5.0.1. Publicación	94
BIBLIOGRAFÍA	95
ANEXOS	106
Anexo A	106
Anexo B	107
Anexo C	108
Anexo D	109
Anexo E	110
ACRÓNIMOS	117
NOMENCLATURA	119

ÍNDICE DE CUADROS

1.1. Relación establecida entre las características típicas de los accidentes que determinan la conducción con somnolencia como causa, según la NHTSA y las estadísticas aportadas por el INEGI.	4
2.2. Breve registro de algunos de los estudios relacionados con este trabajo de investigación, empleados como fuentes bibliográficas a lo largo de este documento. .	38
4.1. Características indicadas por el proveedor para la cámara de vídeo seleccionada; Logitech c525.	60
4.3. Comparativo de CPUs portables disponibles en el mercado.	63
4.4. Eficiencia del detector facial combinado, en relación con la posición facial del usuario.	81
4.5. Posiciones faciales adicionales en las que el detector combinado trabaja con una eficiencia por encima del 85 %.	82
5.1. Tabla comparativa del tiempo de procesamiento por cuadro de vídeo para tres diferentes sistemas de procesamiento.	89

ÍNDICE DE FIGURAS

1.1. (a) Población, (b) muertes por accidente de tránsito y (c) vehículos de motor matriculados por nivel de ingresos de los países.	3
2.1. Diagrama auxiliar para la definición de PERCLOS y AECS, figura original de Xiaojie (2001).	16
2.2. Ejemplos de ADAS, (Mobileye, 2015).	21
2.3. Imagen del sistema de navegación incluido en el BMW Serie 1 de tres puertas (BMW GROUP MÉXICO, 2016).	23
2.4. Imagen ilustrativa del uso del sistema para el estacionamiento asistido (Gotzig, 2016).	24
2.5. Activación de BLIS en una camioneta Odyssey 2011 de Toyota, imagen tomada del vídeo explicativo del sistema proporcionado por la marca en su sitio oficial (American Honda Motor, 2017).	25
2.6. Perfiles corporales bien definidos, ejemplo de siluetas que de forma óptima serían reconocidas por el sistema de detección de peatones del V40 2015 AT 1446 de Volvo (VOLVO, 2015).	26
2.7. Ejemplos de iluminación adaptativa ofrecida en el mercado; (a) patrones de iluminación utilizando faros LED dinámicos (Ford Motor Company, 2017b), (b) iluminación adaptativa de acuerdo al entorno de conducción (Audi AG, 2011a).	27
2.8. A la izquierda, (a) representación gráfica del NVA de Mercedes-Benz basado en cámaras infrarrojas (Daimler AG., 2016c); a la derecha, tomada de (Audi AG, 2016), (b) imagen ilustrativa sobre el funcionamiento del NVA implementado por AUDI utilizando cámaras termográficas.	28
2.9. Ejemplo de alerta visual desplegada por el sistema de detección de fatiga ofrecido por Nissan (Nissan, 2017).	30
3.1. Clasificación de los métodos para detección de objetos, según Shantaiya <i>et al.</i> (2013).	40
3.2. Técnicas para la detección de rostros, clasificadas según su enfoque.	44
3.3. Características Haar: (a) y (b) características propuestas por Papageorgiou <i>et al.</i> (1998), (c) conjunto de rectángulos creado por Viola y Jones (2003), las subfiguras (d), (e) y (f) muestran las características Haar presentadas por Lienhart y Maydt (2002), mientras que (g) y (h) presentan los rectángulos vecinos propuestos por Li <i>et al.</i> (2002) y finalmente (i) pertenece a las características Haar en diagonal de Viola y Jones (2003).	55
4.1. Metodología general propuesta.	59
4.2. Imagen ilustrativa de la cámara de vídeo utilizada.	61
4.3. Muestras de los resultados obtenidos de las pruebas preliminares, desarrolladas con la aplicación del detector de rostro frontal de OpenCV.	64
4.4. Resumen gráfico de la detección de rostros basada en un descriptor HOG.	65

4.5. Ejemplo de pre-procesamiento de la imagen; etapa inicial del proceso para la aplicación de los descriptores HOG. En (a) la imagen original o de entrada , en (b) la imagen normalizada en escala de grises.	66
4.6. Representación gráfica de las máscaras aplicables a la imagen para la obtención de sus derivadas espaciales a lo largo de sus ejes x (a) y y (b).	67
4.7. Imagen (I) tras la obtención de las derivadas espaciales a lo largo de sus ejes x (a) y y (b).	68
4.8. Magnitud del Gradiente ($ G $) de I HOG.	68
4.9. Ejemplo de división de 4.8 en celdas.	69
4.10. Ejemplo de los valores utilizados para la creación del histograma de una celda; (a) orientación de los pixeles de la celda, (b) magnitud del gradiente para cada pixel de la celda.	70
4.11. Ejemplo del histograma de una celda, basado en los valores mostrados en la Figura 4.10.	70
4.12. Histogramas para la imagen inicial (I), Figura 4.5(a), basados en la orientación de los pixeles agrupados en celdas.	71
4.13. Resultado final; (a) descriptores HOG utilizando la Figura 4.5(a) como imagen de entrada (), (b) descriptores HOG sobre la imagen original (I).	72
4.14. Representación gráfica del sistema de clasificación por cascada.	73
4.15. Grados de libertad considerados para la definición de posición del rostro.	74
4.16. Muestra del conjunto positivo de entrenamiento, estandarizado en tamaño y color.	75
4.17. Muestra del set de imágenes negativas de entrenamiento, estandarizado en color.	76
4.18. Ilustración del principio básico de funcionamiento de la función de MATLAB <code>trainCascadeObjectDetector()</code>	77
4.19. Nomenclatura utilizada para expresar la posición facial en términos de grados; signos positivos y negativos ilustrados, basada en la información proporcionada por <i>Gourier et al.</i> (2004).	79
4.20. Comparativo de la cantidad de detecciones hechas por cada detector, sin importar la precisión de la localización, solo se considera la apropiada clasificación de la imagen.	80
4.21. Gráfico de la eficiencia de detección para cada detector entrenado y el pre-entrenado de OpenCV; considerando el porcentaje de detecciones imprecisas, erróneas y acertadas (ver Figura 4.22).	80
4.22. Muestras de la clasificación que se da, en este proyecto, a la forma en que responden los detectores, de acuerdo con los requisitos establecidos previamente, ver subsección 4.2.1: (a) Ejemplo de detección incorrecta, (b) imagen ilustrativa de localización imprecisa, (c) muestra de detección correcta.	81
4.23. Gráfico de los resultados obtenidos mediante las pruebas realizadas para determinar el alcance del detector facial desarrollado.	83
4.24. Ejemplificación de los puntos faciales detectados con el predictor de Kazemi y Sullivan (2014).	84
4.25. Puntos de referencia generados sobre los ojos con el detector de LandMarks de Kazemi y Sullivan (2014).	84

4.26. Puntos descriptivos de la boca; (a) puntos proyectados sobre la boca por el detector de puntos relevantes de Kazemi y Sullivan (2014), (b) puntos seleccionados para el cálculo de apertura bucal.	86
4.27. Muestra de las alertas visuales emitidas por el sistema tras detectarse la presencia de indicadores visuales de somnolencia en el rostro del usuario; (a) activación de alerta por detección constante de ojos cerrados, (b) activación de alarma por bostezo.	87
5.1. Imágenes de las primeras pruebas realizadas al algoritmo.	88
5.2. Muestra de la indicación visual desplegada si el sistema no detecta el rostro del usuario.	89
5.3. Instalación de la cámara de vídeo en el interior del automóvil.	90
5.4. Fotografía de la superficie sobre la que se realizó la prueba del sistema implementada durante la conducción.	91
5.5. Muestra de detección realizada mientras el sujeto de prueba utiliza lentes transparentes.	91
5.6. Imágenes ilustrativas de la detección de rostro, y sus características principales cuando parte del rostro se halla cubierta.	92
5.7. Imágenes tomadas de las pruebas de detección a baja iluminación.	92
5.8. Sistema portátil de detección de somnolencia, instalado en el automóvil.	93
5.9. Distribución del <i>hardware</i> que compone el sistema portátil de detección de somnolencia; a la izquierda (a), la ubicación de la bocina inalámbrica utilizada para la emisión de la alarma sonora, a la derecha, (b) instalación de la cámara de vídeo de 8 MP, el banco de energía o batería portátil con una salida de +5V y 2.5 A, la Raspberry Pi 3B y, finalmente, la pantalla LCD de 3.5 pulgadas.	94

1. INTRODUCCIÓN

De la misma manera en la que la aparición del teléfono revolucionó la forma de comunicarnos a la distancia, la invención del automóvil cambió para siempre la concepción del transporte; de hecho, puede ser visto como una de las tecnologías que más empoderamiento dio a sus usuarios, pues les dio la capacidad de controlar su propio transporte de una forma nunca antes vista.

Viéndose en retrospectiva, los primeros automóviles no eran tan útiles como se esperaba puesto que se enfrentaban a problemas tales como la carencia de una interfaz intuitiva y estandarizada, sumado a que su velocidad, en muchos casos, no competía con la de los caballos (Lewis, 1999). Dado lo anterior es fácil notar como, con el transcurrir de los años, la industria del automóvil ha evolucionado enormemente y su tendencia a innovar parece no detenerse.

No obstante, a pesar de los grandes avances que la industria automotriz ha demostrado, y las ventajas que el crecimiento en la producción y consumo de automóviles han traído consigo, es importante reconocer que el uso de tal tecnología trajo también grandes riesgos, tanto para quienes conducen, como para aquellos que se encuentran vulnerables en las vías de tránsito. Ello se ha visto reflejado en una constante y notoria alza en el número de accidentes de tránsito, lo que hizo evidente la necesidad de implementar sistemas de seguridad cada vez más robustos que permitieran mejorar la experiencia de manejo haciendo de esta una actividad menos riesgosa.

En la década de los 70s se propusieron sistemas de asistencia al conductor, pero no fue sino hasta una década después que estos fueron lanzados al mercado; inicialmente, se diseñaron para cumplir tareas básicas que ayudaran al conductor a conservar el control del automóvil, y lo que empezó con sistemas básicos de asistencia como el de anti-bloqueo de frenos (ABS, por sus siglas en inglés) hoy continúa en lo que parece ser un camino directo a la conducción autónoma (Belbachir *et al.*, 2012; Auricht y Stark, 2014).

Actualmente, la demanda de automóviles más seguros ha generado un vasto campo de investigación y desarrollo que ha resultado en la obtención de sistemas de seguridad cada vez más sofisticados que son capaces de ayudar al conductor a prevenir errores o, como mínimo, reducir las consecuencias de estos, además de alertar y dar soporte durante la conducción; todo ello a partir de la observación de algunas condiciones de manejo tanto internas como externas. Tales sistemas se denominan sistemas avanzados de asistencia al conductor (ADAS), los cuales se describen con mayor detalle en la sección 2.4 ADAS, (Biondi *et al.*, 2017; Planning, 2014).

Sin embargo, a pesar de ser necesarios, los ADAS no se encuentran implementados a nivel global, debido principalmente al costo que ello representa; tanto si el automóvil se adquiere con estos sistemas integrados o si estos se obtienen de forma complementaria, se genera un gasto adicional debido a la necesidad de elementos de sensado específicos. Afortunadamente, las nuevas tecnologías y campos de estudio como el de visión por computadora, abren la posibilidad de desarrollar algunos de estos sistemas de asistencia a un menor costo.

El objetivo de esta tesis es documentar el desarrollo de un sistema embebido de bajo costo para la detección de fatiga aplicado a la asistencia al conductor mediante el procesamiento de expresiones faciales.

1.1. Motivación y descripción del problema

De acuerdo con OMS (2015), a pesar de que los países de ingresos bajos y medianos tienen aproximadamente la mitad de los vehículos del mundo, se producen en ellos más del 90 % de las muertes relacionadas con accidentes de tránsito (ver Figura 1.1).

En México, igual que en la mayor parte de países con bajos y medianos ingresos, las lesiones causadas por accidentes de tránsito han aumentado considerablemente, a tal grado de formar parte de las 10 primeras causas de muerte en el país. A pesar de lo cual se carece tanto de una apropiada legislación en términos de seguridad vial, como de una base de datos que permita determinar las causas principales de tales accidentes, lo cual entorpece la posibilidad de

desarrollo e implementación de estrategias para la prevención y atención de lesiones causadas por accidentes de tránsito. (STCONAPRA y de salud, 2015; OMS, 2015).

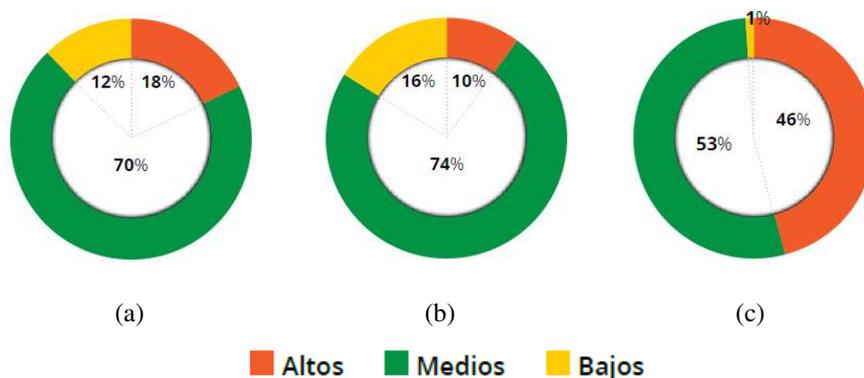


Figura 1.1: (a) Población, (b) muertes por accidente de tránsito y (c) vehículos de motor matriculados por nivel de ingresos de los países.

Una de las posibles causas de accidentes de tránsito que no se consideran en los informes presentados por la secretaría de salud del país respecto a seguridad vial en el país es la somnolencia en el conductor. Ello, a pesar de que según la National Highway Traffic Safety Administration (NHTSA) de los Estados Unidos de América, conducir con somnolencia incrementa la posibilidades de sufrir un accidente hasta en un 24 % (Klauer *et al.*, 2006).

Si bien es cierto que, la presencia o ausencia de somnolencia no puede determinarse mediante pruebas objetivas; como análisis sanguíneos. Imposibilitando, una vez ocurrido un accidente de tránsito, la posibilidad de determinar con precisión si la presencia de somnolencia pudo o no ser parte de las causas del mismo. Sin embargo, es posible analizar las características del accidente para saberlo. De acuerdo con la la NHTSA los accidentes viales relacionados con la somnolencia muestran ciertas características, mismas que incluyen (Strohl *et al.*, 1998);

- El accidente se desarrolla a altas horas de la noche, en la madrugada o a media tarde.
- El vehículo sale de la carretera sin causa aparente.
- El incidente se produce en carreteras de alta velocidad.
- Es evidente la inacción de conductor ante un inminente accidente.
- El conductor es la única persona dentro del vehículo.
- Es altamente probable que el accidente presente consecuencias graves.

Si se comparan dichas características con las estadísticas generadas al respecto en México (INEGI, 2016), los resultados demuestran que se debe prestar mayor atención al problema de la conducción con somnolencia. Esta comparativa se muestra en la Tabla 1.1.

Comparativo	
<i>Característica típica</i>	<i>Datos de accidentes de tránsito en México</i>
Mortalidad relacionada con hora del accidente	23.62 % de las muertes debidas a este tipo de accidentes se ocasiona en los incidentes ocurridos entre las 14 y las 16 horas, mientras que el 21.5 % debido a los accidentes de entre las 0 y las 6 horas.
El vehículo sale de la carretera sin causa aparente	El 9.2 % de los accidentes en zonas urbanas y suburbanas de México presentan esta característica.
No hay intención aparente de evitar el accidente	El 12.8 % de los choques en zonas urbanas y suburbanas del país son contra objetos estáticos.

Tabla 1.1: Relación establecida entre las características típicas de los accidentes que determinan la conducción con somnolencia como causa, según la NHTSA y las estadísticas aportadas por el INEGI.

Una implementación más amplia de ADAS enfocados a la detección y alerta de somnolencia supondría una contribución significativa a la reducción de accidentes viales. Mientras que si tales sistemas resultaran ser de bajo costo y fácil instalación se abriría la posibilidad de que aún en los países de bajos y medianos ingresos, como México, se incrementara su uso.

Actualmente existen cuantiosos estudios que abordan el problema de detección de somnolencia y otros factores de riesgo en conductores mediante la medición de bio-características. En algunos de tales estudios, se ha trabajado en establecer el nivel de somnolencia mediante EOG, mientras que el análisis de la carga cognitiva del sujeto de estudio, se mide a través de EEG, la obtención de ritmo cardíaco y presión sanguínea se realizan por medio de ECG y la temperatura del mismo se mide mediante termómetros (Liang *et al.*, 2009; Zhao *et al.*, 2012). Dicha clase de enfoques supone dos grandes inconvenientes; el primero se basa en que las bio-características serán variables de un sujeto a otro (considerando factores tales como edad y

genero, por ejemplo), el segundo de ellos radica en que para realizar la mediciones propuestas es necesario el uso de sensores especiales para ello que necesitan ser adheridos al cuerpo, lo que no resulta ni cómodo ni práctico cuando el sujeto de prueba se encuentra realizando una tarea tan compleja como la conducción, (Mihai *et al.*, 2015).

En estudios recientes es más común encontrar información sobre propuestas que se realizan para abordar no solo la detección de fatiga en el conductor sino algunas otras necesidades de los conductores, mediante la aplicación de técnicas de visión por computadora y procesamiento de imágenes. Algunos de los ADAS más abordados desde este enfoque son; el detector de cambio de carril y la detección de peatones, como ejemplo de ello se pueden encontrar los trabajos de Dalal (2006); Ruiz Sancho (2014); Gaikwad y Lokhande (2015).

Existen también, cuantiosos estudios sobre la detección de fatiga en el conductor mediante procesamiento de imágenes, sin embargo las soluciones dadas ofrecen siempre alguna variante que hace de cada sistema algo innovador. Mientras que inicialmente los estudios se dedicaban al análisis de una sola característica, generalmente de parpadeo(Ueno *et al.*, 1994; Suzuki *et al.*, 2006; Devi y Bajaj, 2008), estudios más recientes se encaminan a la elaboración de sistemas que combinan el análisis de más de una característica visual, tratando con ello de desarrollar sistemas robustos (Daza *et al.*, 2014).

En este proyecto se propone el desarrollo de un algoritmo que combine métodos de detección de objetos, detección de puntos sobresalientes en imágenes y asociaciones comparativas con indicadores visuales de fatiga sea capaz de detectar somnolencia y distracción en conductores, de tal forma que se pueda emitir una alarma en caso de encontrar indicadores faciales que impliquen la existencia de somnolencia en el usuario. Se espera que con ello se pueda participar en la reducción de accidentes viales debidos a la conducción con somnolencia, sin que sea necesario adquirir un sistema de gama alta y sin que ello implique altos costos de instalación, siendo este un sistema portable.

1.2. Hipótesis

Basada en la información mostrada en las secciones anteriores, se presenta a continuación la hipótesis de la posible solución del problema.

Hipótesis: *Mediante el procesamiento y reconocimiento de expresiones faciales específicas, es posible desarrollar un método de asistencia al conductor que, implementado en un sistema embebido, elimina la necesidad de emplear sensores especializados, disminuyendo así los costos de desarrollo, y reduciendo los gastos de instalación.*

1.3. Objetivos

El objetivo general del presente proyecto es el siguiente:

Objetivo general: *Desarrollar un sistema embebido capaz de detectar indicadores visuales de fatiga y/o distracción en conductores, mediante procesamiento de imágenes que incluya la conjugación de algoritmos de detección de rostro, métodos de localización de puntos relevantes en rostros y el análisis de los parámetros detectados en las expresiones faciales a fin de alertar sobre la existencia de fatiga y contribuir a la reducción de accidentes automovilísticos.*

1.3.1 Objetivos particulares

A continuación se definen los objetivos particulares, mediante los cuales se espera conseguir el cumplimiento satisfactorio del objetivo general, y por ende probar la hipótesis planteada:

- Establecer la arquitectura mínima con la que habrá de contar el sistema embebido, mediante el análisis detallado de los requerimientos básicos necesarios para que sea capaz de soportar las tareas de adquisición de datos, procesamiento de los mismos y despliegado de resultados.
- Desarrollar el algoritmo para la detección de rostro del usuario, empleando técnicas de extracción de características y clasificadores para determinar la existencia y, en su caso, la localización del rostro en la imagen de entrada.

- Identificar, mediante la localización de puntos de referencia, las principales características del rostro, tales como ojos y boca, para obtener de ellas información que ayude a establecer la existencia o carencia de indicadores visuales de fatiga.
- Establecer, a través de una investigación previa, los mecanismos de alarma del sistema para que estos sean activados, en caso de detectar fatiga en el usuario, sin que dichas alarmas tengan un impacto negativo en el conductor.
- Desarrollar métodos de prueba que, mediante la validación de cada etapa del proceso, garanticen el funcionamiento óptimo del algoritmo creado.

2. ESTADO DEL ARTE

2.1. Estudio de la fatiga

2.1.1 Definición

La fatiga, también conocida como somnolencia, puede definirse como la necesidad de dormir y es el resultado del ritmo biológico natural de los humanos conocido como ciclo de sueño-vigilia. Este ciclo se controla por factores tanto circádianos como homeostáticos; la homeóstasis se relaciona con la necesidad de dormir que acumula una persona a lo largo del día; mientras más larga haya sido la vigilia, mayor será la necesidad de dormir y más irresistible resultara hacerlo. El ciclo circadiano, por su parte, podría explicarse como un reloj interno humano que completa su ciclo aproximadamente cada 24 horas. De esta manera, los factores homeostáticos controlan a los circádianos para regular los tiempos de sueño-vigilia (Čolić *et al.*, 2014).

En el ciclo de sueño-vigilia, se generan patrones predecibles con picos de somnolencia que generalmente ocurren unas 12 horas después de la mitad del ciclo del sueño; si nos referimos a las personas que comúnmente duermen durante la noche, estos picos se presentarán en la tarde, y poco antes de la consolidación del siguiente período de sueño; generalmente durante la noche, antes de ir a dormir. Es importante mencionar que el ciclo de sueño-vigilia es influenciado también por los cambios de luz entre el día y la noche; los humanos, generalmente, alcanzan el umbral de vigilia durante el día, mientras que duermen a lo largo la noche (Strohl *et al.*, 1998).

Es importante resaltar que el ciclo de sueño-vigilia es intrínseco e inevitable, por lo tanto no es una condición que pueda ignorarse de forma voluntaria aún cuando actualmente se le reste importancia a la necesidad de dormir apropiadamente, dando preferencia a la realización a algunas otras actividades. Los individuos que, por razones variadas, duermen poco pueden experimentar somnolencia, rendimiento deficiente y una baja capacidad de alerta durante la

vigilia; respuestas neurobiológicas ante la deprivación del sueño (Čolić *et al.*, 2014). Se puede concluir que la fatiga es un proceso gradual y acumulativo, y que quien lo padece presenta una reducción en su desempeño mental, lo que incluye eficiencia y estado de alerta (Jimenez Moreno, 2011).

El ciclo del sueño, en términos médicos, se divide en dos fases; la fase de sueño lento, conocida como NREM por las siglas en inglés de Nonrapid Eye Movement (NREM), y la fase de sueño paradójico o REM (Rapid Eye Movement (REM)). Por su parte, la fase NREM se divide en 4 etapas, de las cuales la somnolencia es la primera (Yu, 2009).

2.1.2 Principales causas

Dejando de lado las sustancias tóxicas y medicamentos que pueden generar somnolencia, las principales causas de esta en personas sin trastornos del sueño, de acuerdo con Čolić *et al.* (2014) son:

- Restricción o pérdida del sueño: Dormir poco es un factor que contribuye ampliamente en la disminución de alerta durante la vigilia, dormir un aproximado de 8 horas por cada ciclo circadiano es lo más recomendable. Una restricción del sueño, aún cuando sea por solo un día, puede resultar en somnolencia extrema durante la vigilia. Las consecuencias de la deprivación del sueño son acumulativas; cuando se duerme poco se crea una especie de deuda de sueño, misma que solo podrá eliminarse al dormir.
- Fragmentación del sueño: Pasar 8 horas en cama, no garantiza que se duerma apropiadamente; las interrupciones del sueño pueden generar afectaciones negativas en el desempeño del individuo. La fragmentación del sueño puede deberse a causas internas tales como enfermedades o a causas externas como ruido ambiental o iluminación inadecuada.
- Factores circádianos: Tal como se mencionó en la sección 2.1.1, se considera normal, dentro del ciclo sueño-vigilia, que se presenten un par de picos de somnolencia durante la vigilia.

2.1.3 Técnicas de medición de fatiga en conductores

Las técnicas de detección y medición de fatiga en conductores suelen agruparse en cinco categorías, las cuales serán brevemente explicadas a lo largo de esta sección.

Mediciones subjetivas

Se sabe que, durante la vigilia, mientras mayor sea la fatiga presentada por el individuo mayor será la necesidad de dormir; de ello se puede deducir que la fatiga o somnolencia puede experimentarse en diferentes niveles. Las técnicas subjetivas empleadas para determinar el nivel de somnolencia se basan en cuestionarios mediante los cuales los individuos participantes son interrogados a cerca de su nivel de fatiga. Dado que la sensación de somnolencia de cada individuo no puede calificarse como un dato objetivo, se considera que la información obtenida tras los cuestionarios es subjetiva, lo que le da nombre a este tipo de técnicas (Sahayadhas *et al.*, 2012; Čolić *et al.*, 2014).

De acuerdo con Čolić *et al.* (2014) las pruebas de somnolencia más conocidas, pertenecientes a este grupo, son:

- Escala de somnolencia de Epworth (ESS): Esta prueba consta de ocho ítems mediante los cuales se que cuantifica la somnolencia del individuo en cuestión. Se basa en la tendencia, del individuo, de quedarse dormido en situaciones comunes dentro de su estilo de vida; tales como leer, ver la televisión o al detenerse mientras conduce. La respuesta a cada ítem o se valora en una escala de 0 a 3, refiriéndose el 0 a una nula posibilidad de dormir y el 3 a una alta probabilidad de que ello suceda; al finalizar la prueba, un puntaje acumulado menor a 10 indicará nula o poca somnolencia, entre 10 y 14 se considera somnolencia moderada, y un puntaje de 15 o más reflejará somnolencia severa (Strohl *et al.*, 1998; Čolić *et al.*, 2014).
- Test de latencias múltiples de sueño (MSLT): En esta prueba, los individuos son puestos en situaciones apropiadas para dormir; durante 20 minutos, cada dos horas a lo largo del día, mientras que se les ha indicado que en tales situaciones deben tratar de dormir. En este caso se analiza el tiempo que tarda cada individuo en quedar dormido; tardar

10 minutos o más es considerado normal, mientras que si toma 5 minutos o menos se considera somnolencia patológica (Strohl *et al.*, 1998).

- Test de mantenimiento de la vigilia (MWT): Para esta prueba, se indica a los individuos que deben tratar de mantenerse despiertos durante un periodo de 20 minutos; en caso de no lograrlo, se mide el tiempo que estos tardaron en quedarse dormidos. Aquellos sujetos que se duermen durante los primeros 15 minutos de la prueba se consideran incapaces de conducir debido a deprivación del sueño (Colic, 2014).
- Escala de somnolencia de Stanford (SSS): Consta de 7 enunciados a través de los cuales los individuos califican su nivel de alerta en una escala que va del 1 (muy despierto) hasta el 7 (muy cerca de dormir). Puede aplicarse de forma repetida a lo largo de un intervalo de 24 horas. Su mayor desventaja es que, al basarse en las respuestas de los individuos bajo estudio, es posible que obtener resultados inexactos si el sujeto decide mentir durante la prueba (Strohl *et al.*, 1998).
- Escala de somnolencia de Karolinska (KSS): Esta es la escala de somnolencia más comúnmente usada debido a que, al ser contrastada con pruebas de laboratorio y estudios de campo, se ha comprobado su eficiencia. Se basa en preguntas de respuestas abiertas que guían al individuo para que, mediante sus respuestas, provea un reporte relativo a su calidad de sueño (Čolić *et al.*, 2014).
- Escala visual analógica (VAS) para somnolencia: Se emplea una escala de 100 mm en la cual el extremo izquierdo representa «apunto de dormir» y el derecho «lo más despierto posible», en esta escala el individuo deberá crear una marca en el punto que indique que tan fatigados o somnolientos se sienten. El nivel de somnolencia se determina con la distancia existente entre la marca creada por el individuo y uno de los extremos de la escala. Se considera que el método utilizado en la VAS es bastante apropiado si se desea realizar pruebas repetidas pues es de aplicación rápida y simple (Čolić *et al.*, 2014).

En este tipo de técnicas para la medición de la fatiga, tanto la explicación de las instrucciones a seguir durante las pruebas como la interpretación de estas depende una persona. Esto da una amplia posibilidad de interpretaciones confusas de las pruebas aplicadas (Strohl *et al.*, 1998).

Mediciones basadas en el vehículo

En la mayoría de los casos, esta clase de pruebas se realizan en ambientes simulados, colocando sensores en algunas partes del automóvil, tales como el acelerador y el volante; la información obtenida de los sensores es analizada para determinar el nivel de somnolencia del individuo. Las características de manejo más comúnmente empleadas para este tipo de mediciones son el movimiento del volante y la desviación estándar de la posición del auto en el carril (Sahayadhas *et al.*, 2012).

El desarrollo de esta clase de pruebas se ha determinado de acuerdo con la evidencia arrojada por los reportes de accidentes automovilísticos presumiblemente debidos a un alto grado de fatiga del conductor. De estos reportes se obtiene que el comportamiento típico del conductor y su automóvil durante tales eventos, posee alguna o varias de las siguientes características (Čolić *et al.*, 2014):

- Velocidad de conducción alta en la que es evidente una muy baja o inexistente intención de frenado .
- El automóvil tiende a salirse del camino.
- Esta clase de accidentes suelen ocurrir en carreteras de alta velocidad.
- El accidente se produce en carreteras de alta velocidad.
- El conductor se mantiene inactivo durante el evento, no intenta evitarlo.
- El conductor viaja solo.

Mediciones por comportamiento

Este tipo de técnicas se basan en el hecho de que una persona cansada muestra movimientos faciales característicos, así como cambios poco controlados en la posición la cabeza. Para la medición de somnolencia mediante dichas características de comportamiento son ampliamente utilizados algunos sistemas computarizados no invasivos. Conforme a lo descrito por Sahayadhas *et al.* (2012), muchos de los estudios publicados acerca del uso de estas técnicas para determinar la existencia o no de fatiga en conductores se basan en el parpadeo y en el PER-

CLOS. En la sección 2.2 se detallan algunos de los indicadores visuales que sirven como base para la medición de fatiga por comportamiento.

Estas técnicas han demostrado ser eficaces en la predicción de somnolencia por lo que han sido aplicados en algunos sistemas de comerciales, tal como puede verse en la sección 2.4.1 de este documento. Es común encontrar que se realicen estudios y aplicaciones basados en el análisis, en conjunto, de varias de las características visuales que forman parte del comportamiento de un conductor fatigado, tales como cambios relativos a las cejas, bostezos, parpadeo, caída de mandíbula y posición de la cabeza (Sahayadhas *et al.*, 2012).

Pese a su eficacia y a la ventaja de ser métodos no invasivos, estos se encuentran aún desarrollo y en presencia de grandes retos asociados a campos como visión por computadora, procesamiento de imágenes y *machine learning*. Algunos de ellos son los referentes a la iluminación, pues en ambientes nocturnos o de escasa iluminación se torna más complicado el análisis de varios de los rasgos objeto de estudio de estas técnicas, otro de ellos es la necesidad de procesamiento en *tiempo real*; incluso aún en cuanto a montaje y selección de sensores aplicables hay mucho por estudiar y mejorar (Čolić *et al.*, 2014).

Medidas fisiológicas

A diferencia de las mediciones basadas en la observación del vehículo y las hechas en relación al comportamiento del conductor (descritas en párrafos anteriores), las mediciones fisiológicas pueden llevarse a cabo desde las etapas tempranas de la fatiga, teniendo plena certeza de que sus resultados serán aún más confiables. Para este tipo de mediciones son ampliamente usadas las siguientes señales electro-fisiológicas:

- **Electrocardiograma (ECG):** El ECG registra la actividad eléctrica del corazón y es la más usada de las técnicas para la medición de fatiga que se basan en condiciones fisiológicas. Mediante el empleo de esta técnica, se puede definir el estado de fatiga de una persona de forma bastante precisa; esto se logra a partir de la observación de los cambios generados en el comportamiento de su corazón, es decir, en las variaciones del ritmo cardíaco. Tales variaciones son comparadas mediante la medición de la Variabilidad del Ritmo Cardíaco

(HRV por sus siglas en inglés). La escala de HRV describe ambos límites; las frecuencias baja y alta a las que late el corazón a lo largo de dicha variabilidad: Si se encuentra una frecuencia de latido muy cercana a LF este será un indicio claro de somnolencia, mientras que un individuo con nula o escasa fatiga presentará un ritmo cardíaco con frecuencia muy parecida a la HF. La detección de fatiga, mediante este método, puede realizarse de forma progresiva y constante tan pronto como se presenten los cambios en el mesurando (Colic, 2014).

- Electroencefalograma (EEG): esta técnica registra la actividad eléctrica del cerebro. Al respecto, Sahayadhas *et al.* (2012) especifica lo siguiente: Si las señales registradas mediante el EEG poseen una frecuencia de entre 0.5 Hz a 4 Hz recibirán el nombre de ondas Delta, las cuales están asociadas a la actividad de dormir; las ondas Theta, por su parte, serán aquellas cuya frecuencia se encuentre entre los 4 Hz y los 8 Hz y se relacionan con la somnolencia; las ondas Alfa indican relajación y creatividad, y su frecuencia puede ir de los 8 Hz a los 13 Hz; por último, se tienen las ondas Beta, cuya frecuencia puede ir de los 13 Hz a los 25 Hz e indicarán un estado de alerta.
- Electrooculograma (EOG): Registra la diferencia de potencial eléctrico entre la córnea y la retina del ojo humano. Para llevarse a cabo, es necesario colocar un electrodo en la esquina exterior de cada ojo y uno más en el centro de la frente para tomarlo como referencia; a partir de dichos electrodos se monitorea el parpadeo, si se detecta un movimiento más lento de lo habitual se podrá concluir que el individuo está sintiéndose somnoliento. Aún cuando este método es muy preciso, tiene la desventaja de ser bastante invasivo, por lo que no es considerado como un método práctico de detección de fatiga (Čolić *et al.*, 2014).

Si bien, ya se mencionó que la medición de fatiga mediante características fisiológicas es más exacta que las anteriores, cabe mencionar que son las más invasivas. Se espera que con el avance de la tecnología, se puedan superar los inconvenientes de este tipo de técnicas.

Métodos híbridos

Una metodología híbrida para la medición de la fatiga es aquella que combina la aplicación de dos o más técnicas de medición (expuestas a lo largo de esta sección, la 2.1.3). La

creación de métodos híbridos es vista como una forma de combinar los beneficios de los métodos individuales, a la vez que se minimizan sus desventajas. Según Čolić *et al.* (2014), la combinación de métodos más prometedora es la que se conforma de técnicas basadas en mediciones de factores fisiológicos y la observación del comportamiento tanto del vehículo como del conductor.

2.2. Indicadores visuales de fatiga

En la sección 2.1.3, se realizó una breve definición de diferentes tipos de técnicas que pueden emplearse para la medición de fatiga, entre las que se encuentran aquellas denominadas 'mediciones por comportamiento'. Estas últimas encuentran su argumento en el hecho de que la fatiga puede ser inferida a partir de expresiones faciales y algunas otras variables observables. Atender a solo uno de los parámetros visibles que evidencian fatiga en un individuo no proporcionará información suficiente como para determinar el estado de alerta-fatiga del mismo; se recomienda, por lo tanto, combinar la información obtenida de varios de estos indicadores visuales. Esta sección, está dedicada a la definición de aquellos parámetros visuales que, de acuerdo con Zhang *et al.* (2014), son los más relevantes en la detección de fatiga, los cuales son:

- PERCLOS: Mide el porcentaje de cierre de los párpados sobre la pupila a lo largo del tiempo, diferenciando entre la 'caída' de los párpados debida a la fatiga (movimiento lento y prolongado) y el parpadeo normal durante la vigilia (movimientos de cierre rápidos). En 1994 se estableció el sistema de medición de somnolencia basado en PERCLOS, dicho sistema se basó en determinar que proporción de tiempo, de un total de un minuto, el cierre de los ojos era igual o mayor al 80 % (Dinges y Grace, 1998). Se sabe que los conductores en estado de alerta reportan un PERCLOS muy bajo comparado con aquellos que se sienten fatigados. Al determinar el PERCLOS se involucran dos de las características visuales medibles para el estudio de la fatiga que son relativas a los párpados; la frecuencia de parpadeo y el tiempo que los ojos permanecen cerrados (Xiaojie, 2001). Se considera que los ojos están cerrados cuando un alto porcentaje del área de las pupilas se encuentra oculta por los párpados, este porcentaje se ha definido en un cercano a 80 %.

Por lo tanto, haciendo referencia a las pupilas, cuándo su área visible se haya reducido hasta ser del 20 % de su tamaño se podrá inferir que los ojos se encuentran cerrados. Dado lo anterior y con el apoyo de la Figura 2.1, se puede decir que la duración del cierre de los ojos está definida por la diferencia entre los instantes consecutivos t_2 y t_3 , siendo el primero aquel en el que el área visible de la pupila es del 20 % o menos, mientras que el segundo de tales instantes es aquel en el que la pupila es visible al 100 % de su tamaño.

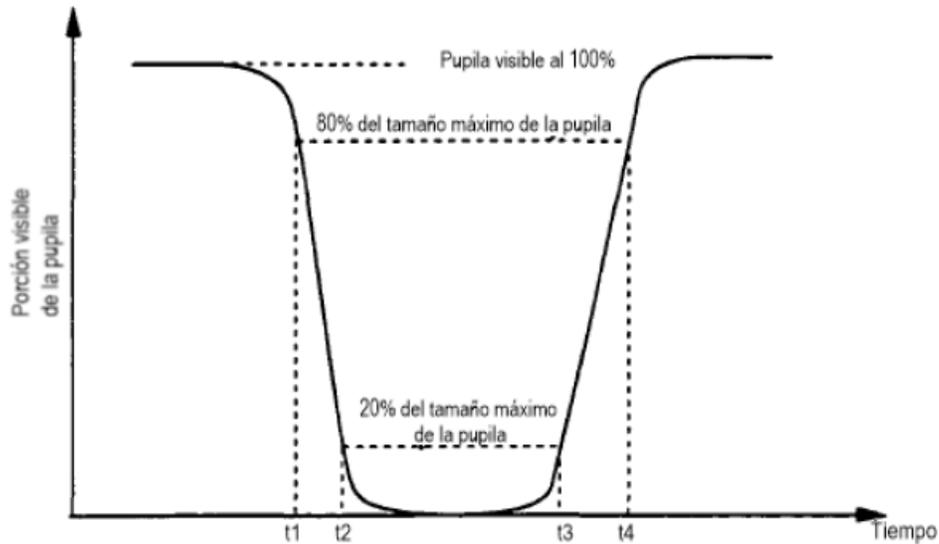


Figura 2.1: Diagrama auxiliar para la definición de PERCLOS y AECS, figura original de Xiaojie (2001).

- AECS: Se refiere a la velocidad promedio del cierre/apertura de los parpados. Se calcula como la media aritmética de la velocidad a la que se cerraron o abrieron los ojos durante un periodo de tiempo determinado (Xiaojie, 2001), Zhang *et al.* (2014) indica que la AECS en una persona despierta puede ser hasta 8 veces mayor que la medida en una persona somnolienta. Observando la Figura 2.1 es posible definir AECS como el periodo de tiempo transcurrido entre t_1 y t_2 o entre t_3 y t_4 ; lapsos de tiempo en los que el área visible de la pupila se encuentra entre el 20 % y el 80 % del total (Xiaojie, 2001).
- PERLVO: Alude al porcentaje de apertura vertical de la boca durante un tiempo determinado. Se emplea para detectar bostezos, y aún cuando puede ocurrir que una persona somnolienta no bostece, se considera que este parámetro debe incluirse el modelo de detección de fatiga si se espera obtener un sistema robusto (Zhang *et al.*, 2014; Gang *et al.*, 2017).

- PERTITL: Hace referencia a la proporción del tiempo que el individuo mantiene la cabeza inclinada, en relación con el tiempo total definido para la medición. Contabiliza el tiempo en que el sujeto mantiene su cara en dirección hacia abajo o hacia arriba. Se emplea como medida para determinar la existencia de cabeceo debido a la fatiga (Gang *et al.*, 2017).

A los cuales, según Jimenez Moreno (2011), se podrían sumar algunos otros como; el tono facial, dirección de la mirada, expresiones faciales y la detección de cabeceo repetitivo.

2.3. Antecedentes del procesamiento de imágenes

Tradicionalmente, se considera que los seres humanos estamos provistos de cinco mecanismos fisiológicos de percepción, comúnmente llamados sentidos; estos nos permiten observar y comprender el entorno que le rodea. De estos cinco, nuestro sistema de visión es, probablemente, el más preciso dando al cerebro una ilustración de lo que nos rodea en términos de percepción pues no solo es equiparable a un dispositivo de captura de imagen sino a un medio de procesamiento de la imagen capturada. El sistema de visión que poseemos incluye; detección, identificación, reconocimiento, análisis, síntesis y modelado (Das, 2015).

Dicho lo anterior, es fácil comprender que las imágenes jueguen un papel tan importante en nuestra vida, lo que a su vez explica el por qué el hombre se ha dado a la tarea de crear un sinnúmero de instrumentos con los cuales sea posible capturar imágenes; representaciones 2D de objetos tridimensionales.

Es importante mencionar que si bien, las imágenes tal cual son recibidas por el cerebro humano poseen una cantidad de variaciones de colores que bien podría definirse como infinita, la representación digital de las mismas no es más que una aproximación basada en los colores primarios, muestras espaciales e intensidades cuantificadas. Sin embargo, ya sea que se trate del ojo humano o de una cámara, existirán siempre limitaciones físicas en las que su capacidad de funcionamiento se vea mermada.

En los últimos años, los avances tecnológicos en cuanto al desarrollo de dispositivos de captura han sido enormes, siendo dichos avances, los mismos que han dado a las imágenes una nueva, y cada vez más amplia, gama de aplicaciones.

En cuanto a los nuevos dispositivos de captura, sus limitaciones, ya sean debidas al entorno o a la composición propia, pueden resultar en imágenes de baja calidad e inadecuadas para el fin por el que han sido creadas. Es quizá esta la razón por la que el estudio de técnicas mejoramiento de imágenes ha ganado relevancia con el tiempo.

Mediante el mejoramiento de imágenes se busca optimizar la apariencia de las mismas, conservando las características que así convengan y resaltando aquellas que sean de interés de acuerdo con la aplicación a la que estén destinadas.

Las técnicas de mejoramiento de imágenes pueden ser divididas en dos categorías, aquellas que trabajan para mejorar la información pictográfica ante la vista humana, y aquellas en las que el objetivo es el almacenamiento, transmisión y representación de datos para su procesamiento mediante máquinas o sistemas autónomos. También pueden clasificarse de acuerdo al dominio de la imagen sobre el que trabajan; las que se aplican en el dominio de la transformada operan modificando la frecuencia de la transformada de la imagen, mientras que las se enfocan en el dominio espacial, como el mejoramiento del contraste, operan directamente en el nivel de gris del pixel (Gonzalez y Woods, 2008).

Una de las primeras aplicaciones de imágenes digitales fue en la industria del periódico, cuando se comenzaron a enviar fotografías entre Londres y Nueva York a través de líneas de cable submarino mediante el método de transmisión denominado Bartlane. Este sistema, inventado a principios de la década de 1920, redujo el tiempo requerido para transportar una fotografía a través del Atlántico de más de una semana a no más de tres horas (Gonzalez y Woods, 2008). Dado que el sistema Bartlane solo podía codificar las imágenes en 5 niveles diferentes en escala de grises, la calidad de las imágenes se veía claramente disminuida, por lo que pronto se tuvo la necesidad de mejorar el método.

Desde entonces y hasta ahora, las mejoras han continuado incesablemente. Hoy en día, las ramas de aplicación del mejoramiento de imágenes son cuantiosas. No se ha llegado a un acuerdo general de donde termina el procesamiento de imágenes y comienzan otras áreas relacionadas como análisis de imágenes y visión por computadora. Sin embargo, no siempre es necesario separar unas de otras; existen aplicaciones en las que se requiere que dichas áreas se complementen entre sí. Una de tales aplicaciones se encuentra en el ámbito automotriz.

A lo largo de la historia, los avances en los medios de transporte han tenido un impacto tanto cultural como económico. La movilidad ha cambiado la forma de vida de las personas, y los automóviles continúan evolucionando y convirtiéndose en mecanismos cada vez más inteligentes mediante las tecnologías de vanguardia. Durante las últimas tres décadas hemos sido testigos del incremento de conocimientos en términos de visión por computadora, siendo éste el producto de investigaciones tanto industriales como educativas (Kisa y Gelautz, 2014).

No fue sino hasta principios del siglo XXI que la industria automotriz dio una aplicación sumamente interesante a los conocimientos obtenidos sobre visión por computadora, mediante los denominados ADAS. Estos sistemas están diseñados para aumentar la alerta y la seguridad del conductor proporcionándole información esencial acerca de situaciones que podrían poner en riesgo la integridad del auto, y por lo tanto la de sus ocupantes y, de ser necesario, tomando acciones que pudieran reducir la gravedad de un accidente. Lo anterior bajo la premisa de que los errores humanos existen, e incluso, aceptándolos como una característica inherente a nuestra naturaleza.

Sin embargo, para que los sistemas de visión por computadora puedan cumplir algunas de sus funciones dentro de los ADAS, es necesario obtener imágenes que describan el entorno de funcionamiento, lo que podría incluir imágenes tanto del interior como del exterior del vehículo. Si ponemos en contexto la situación, se puede hacer claro que durante la conducción, el vehículo se moverá a través de entornos variables; desde lugares con iluminación casi deslumbrante hasta sitios con escasa luminosidad. Es en este punto en el que entra en juego el procesamiento de imágenes, pues será a través de éste que las imágenes de entrada al sistema de visión por

computadora poseerán las características mínimas requeridas para el buen procesamiento de la información que éstas aporten.

2.4. ADAS

Los sistemas de asistencia al conductor tienen el potencial de cambiar el modo en que nos transportamos día a día; mediante el incremento de seguridad y eficiencia de los automóviles. En cierto grado, ya se han implementado sistemas de asistencia al conductor tales como los ABS, sin embargo, sistemas más complejos e innovadores están apenas en fase de difusión, denominados Sistemas Avanzados de Asistencia al Conductor (ADAS).

Los ADAS pueden considerarse como la colección de sistemas y subsistemas encaminados a la conducción autónoma. Los expertos concuerdan en que el veloz desarrollo de tales sistemas inevitablemente desembocará en automóviles “inteligentes”, detectando situaciones de riesgo y actuando en respuesta, de forma autónoma, para evitar el accidente o reducir el impacto del mismo (Planning, 2014).

Dichos sistemas están diseñados para, constantemente, monitorear ciertas condiciones del entorno de conducción, ya sea al interior o al exterior del automóvil, en caso de que los parámetros sobrepasen los límites configurados, el conductor será notificado.

Algunos ejemplos de ADAS son; indicador de cambio de carril, distancia entre vehículos, detección de fatiga y/o distracción, detección de peatones e indicadores de alerta por objeto oculto en el punto ciego de los espejos (Biondi *et al.*, 2017); en la Figura 2.2 se muestran algunos ejemplos de ADAS comerciales. Cada uno de estos sistemas puede realizar su función mediante métodos diversos, los cuales van desde la aplicación de cámaras para el monitoreo a distancia mediante comunicación satelital, sensores infrarrojos, cámaras termográficas hasta sensores para el monitoreo de los signos vitales del conductor. El nivel de sofisticación y exactitud de los ADAS a implementar dependerán de las posibilidades y necesidades tanto del proveedor como del usuario.



Figura 2.2: Ejemplos de ADAS, (Mobileye, 2015).

Es un hecho que los automóviles se han convertido en sistemas que deben ser capaces de colaborar con el conductor, lo que representa un reto para aquellos que involucran visión por computadora, pues operan en ambientes altamente cambiantes, (Fletcher *et al.*, 2005) por esta razón se ha de implementar un sistema de procesamiento de imágenes que provea señales 2D con la claridad necesaria para que los ADAS cumplan con su objetivo.

2.4.1 ADAS y su implementación actual

Actualmente, algunos sistemas de asistencia al conductor son implementados por las ensambladoras automotrices desde la fabricación de su producto, y aunque el alcance, las alertas que emiten, los sensores de los que se sirven, entre otras características dependerán siempre del fabricante y el sector del mercado al que se encuentren dirigidos los automóviles, su principio básico de funcionamiento es compartido. A continuación se hace una breve reseña de algunos de los ADAS presentes en el mercado y sus características básicas.

Es importante mencionar que cada empresa automotriz ofrece una diferente gama de sistemas de asistencia al conductor, de la cual se ha seleccionado una pequeña parte para desarrollar el tema de esta sección.

Sistema anti-bloqueo de frenos ABS

Como se mencionó en párrafos anteriores, este es un sistema ampliamente implementado, que ayuda a reducir la distancia de frenado y a aumentar la estabilidad del automóvil durante dicha acción. Una ventaja adicional que puede atribuirse a los ABS es una mayor eficiencia en las llantas, ya que se reduce la fricción de estas con la carretera durante el frenado.

El ABS permite que las llantas del automóvil mantengan contacto de tracción con la carretera, evitando que estas se bloqueen y patinen al frenar. Dicho sistema es avanzado y se usa los principios de umbral y ritmo de freno aplicados, en sistemas de frenado anteriores, por conductores experimentados; esto hace que el frenado pueda ser mucho más eficiente que al ser ejecutado por un conductor promedio sin asistencia (Khachane y Shrivastav, 2016).

De acuerdo con Khachane y Shrivastav (2016), la industria automotriz continúa trabajando en desarrollar mejoras que aumenten su efectividad; la distribución electrónica de la fuerza de frenado (EBD) es una de tales mejoras, esta incrementa el funcionamiento de los ABS en ambientes adversos, tales como la presencia de hilo o agua en la carretera mediante el uso de una unidad de control electrónica (ECU).

Sistema de navegación

Actualmente, el sistema de posicionamiento global (GPS) es usado de forma común como complemento de viaje para la localización de la dirección de destino, la descripción de la ruta más corta, o aquella en la que el tránsito muestra menores complicaciones.

Si bien es cierto, que inicialmente los sistemas GPS se comercializaron como un dispositivo independiente al automóvil, también lo es el hecho de que, con el avance de la tecnología, ha resultado posible la implementación de este sistema en el vehículo desde su fabricación; de tal manera que se pueda obtener y mostrar, en tiempo real, la información vial. De acuerdo con Tigadi *et al.* (2016) esta información puede ser provista por diversas fuentes, tales como la policía y sistemas de emergencia, así como por proveedores de datos de tránsito.

En Tigadi *et al.* (2016) encontramos que el principal beneficio de utilizar este sistema, radica en que el usuario podrá tomar mejores decisiones para su transporte, tales como la elección de ruta de viaje.

Al ser este uno de los sistemas de asistencia al conductor más conocidos es también uno de los que más presencia tiene en el mercado, siendo bastante amplia la lista de las marcas automotrices que lo han implementado en sus productos. Algunas de ellas son; FIAT, BMW, Ford y Chysler. En la Figura 2.3 se muestra el «Sistema de navegación profesional» que forma parte de los ADAS ofrecidos por BMW.



Figura 2.3: Imagen del sistema de navegación incluido en el BMW Serie 1 de tres puertas (BMW GROUP MÉXICO, 2016).

Sistema de control de velocidad de cruceo adaptativo ACC

El objetivo del ACC es controlar la velocidad del automóvil para mantener una distancia de seguridad, establecida previamente, respecto al automóvil que se encuentra delante. De esta manera, el conductor podrá mantener su atención en el volante en situaciones de tránsito pesado (Morán Cruz, 2013). Marcas como Volkswagen, Toyota, BMW y Ford indican en sus páginas oficiales que algunos de sus modelos cuentan con este sistema.

Aparcamiento asistido

Este sistema está diseñado para facilitar la labor del conductor al maniobrar en lugares con poco espacio, ayudando a que los movimientos necesarios para estacionar sean rápidos y seguros (Volvo, 2016).

Diversos proveedores de autos han incluido este sistema en algunos de sus modelos. En el caso de Volvo, por ejemplo, este sistema funciona a base de sensores que se activan a velocidades inferiores a los 10 km/h, estos sensores se encuentran ubicados en las partes delantera y trasera, así como en los laterales del auto; al detectar algún objeto a una distancia de 0.8 metros, una advertencia sonora se activa al interior del auto, al mismo tiempo que se muestra una imagen esquemática en la que se pueden monitorear la distancia entre el automóvil y el obstáculo detectado (Volvo, 2016).

Algunos fabricantes, como Volkswagen, han optado por implementar sistemas de aparcamiento asistido mucho más sofisticados en los cuales, una vez que se ha elegido el lugar de estacionamiento, el automóvil se encargará de los giros del volante de forma automática, mientras que el conductor únicamente deberá frenar o acelerar según se requiera. Para ello el automóvil que posee dicho sistema se encuentra provisto de sensores ultrasónicos e indicadores visuales sobre la posición del automóvil respecto a los objetos detectados (Volkswagen, 2016b).

Algunas otras automotrices que han implementado este sistema son: BMW, Ford, Porsche, Toyota, GM, Audi, entre otras. La calidad y el alcance de dichos sistemas dependerán de la marca y gama del automóvil al que se hayan integrado. Para información detallada es necesario consultar la página oficial de cada proveedor. También se sugiere la consulta del artículo de Gotzig (2016) dedicado a este sistema de asistencia, del que fue tomada la Figura 2.4, para un entendimiento más detallado del mismo.

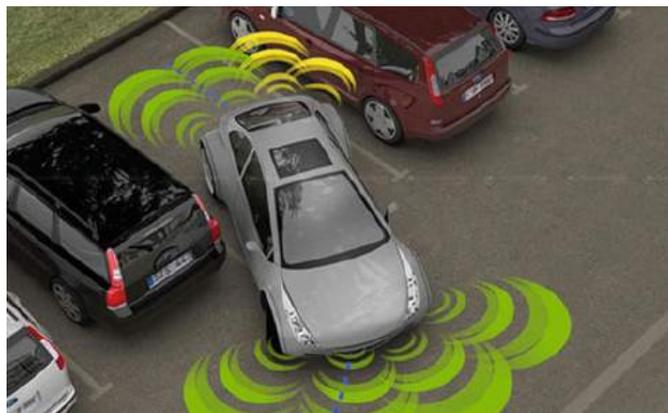


Figura 2.4: Imagen ilustrativa del uso del sistema para el estacionamiento asistido (Gotzig, 2016).

Sistema de información de puntos ciegos (BLIS)

El BLIS está basado en tecnología radar (generalmente), los sensores suelen instalarse en los espejos laterales del automóvil o, incluso, en el parachoques trasero. Cuando un vehículo se acerca tanto que se encuentre dentro del intervalo de detección de los sensores, se envía una señal al controlador del sistema para que este determine si existe o no alguna condición de alerta que puede ser visual, acústica o vibratoria; en caso afirmativo, se activará un indicador al interior del automóvil, de tal manera que el conductor pueda reaccionar de acuerdo a la situación (AAA Automobile Club, 2014).

La localización y tipo de los sensores, la ubicación de los indicadores y el tipo de alarma que emiten, así como el alcance del sistema dependen, como en el resto de los casos, de la gama, el modelo y la marca automotriz a la que pertenezca el vehículo, aunque el principio básico de funcionamiento es compartido. Varias marcas que ofrecen este servicio en sus productos, cuentan con espacios en línea en los que se muestra al usuario el funcionamiento básico y las características del sistema. Algunas de ellas son; Honda, Ford, Volvo, Volkswagen y Nissan. En la imagen 2.5 se muestra una imagen ilustrativa del sistema ofrecido por honda.



Figura 2.5: Activación de BLIS en una camioneta Odyssey 2011 de Toyota, imagen tomada del vídeo explicativo del sistema proporcionado por la marca en su sitio oficial (American Honda Motor, 2017).

Detección de peatones

Este sistema es un recurso auxiliar de apoyo al conductor que actuará frenando automáticamente el vehículo, en el mejor de los casos, o emitiendo una alerta que podría ser visual, sonora o una combinación de ambas si se trata de sistemas menos sofisticados, cuando se detecte la presencia de un peatón detenido frente al automóvil o caminando en la misma dirección.

Dependiendo de la desarrolladora automotriz responsable, podrá encontrarse como un sistema individual o combinado con algún otro de asistencia pre-colisión (VOLVO, 2015; Ford Motor Company, 2017a).

Se desea que con este tipo de sistemas se reduzca el número de accidentes de tránsito en los que se ven involucrados los peatones, diversos vídeos de las pruebas realizadas a estos sistemas pueden encontrarse en Internet.

Para este tipo de ADAS se suelen emplear instrumentos tales como cámaras y radares, las limitaciones de los sensores utilizados dependerá tanto de la marca como de la gama a la que pertenezca el vehículo. Por ejemplo, en el caso del V40 2015 AT 1446 de Volvo, se especifica que el peatón podría ser detectado únicamente si su estatura es mayor a los 80 cm, además de que si grandes partes de su cuerpo no están al alcance de la cámara la detección podría no realizarse, de tal manera que su fiabilidad es limitada en escenarios de baja luminosidad y nula durante la noche o dentro de túneles. En la Figura 2.6 se ilustran los perfiles corporales que permitirían una detección óptima con el sistema de detección de peatones del automóvil mencionado (VOLVO, 2015).

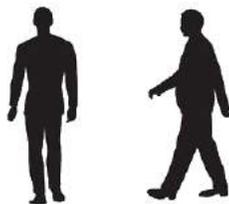


Figura 2.6: Perfiles corporales bien definidos, ejemplo de siluetas que de forma óptima serían reconocidas por el sistema de detección de peatones del V40 2015 AT 1446 de Volvo (VOLVO, 2015).

Sistemas de iluminación dinámica (DLS)

En el mercado se pueden encontrar sistemas de iluminación con cierto grado de automatización y adaptabilidad, dependiendo mayormente del proveedor automotriz y la gama del automóvil; encontrando en las gamas altas los sistemas de iluminación frontal más sofisticada.

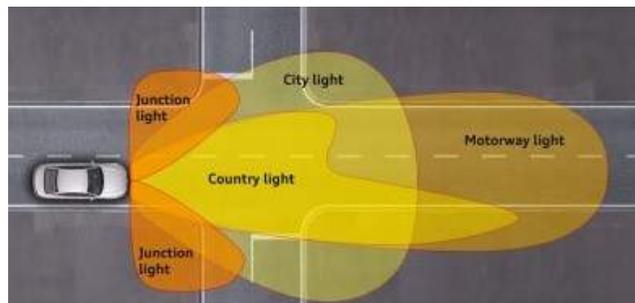
Los DLS más sencillos son aquellos que poseen un control automático para cambiar de luces largas a cortas y viceversa, según se requiera; durante la conducción nocturna, este

sistema puede mantener las luces largas encendidas siempre que no detecte tráfico delante del automóvil, para dar mayor visibilidad del camino, mientras que si detecta que un automóvil se acerca de frente o que hay tráfico delante, activa las luces de cruce o cortas. Ford y Toyota, son algunas de las marcas que ofrecen este tipo de sistemas (Ford Motor Company, 2017b; Toyota Motor Corporation, 2017).

Algunas otras marcas, como Volkswagen, ofrecen sistemas que se adaptan a los giros realizados al transitar por carreteras curvas. Mientras que marcas como Porsche, Audi y Mercedes-Benz han implementado sistemas que adaptan la iluminación dependiendo de condiciones de manejo tales como; detección de tráfico al frente, activación de direccionales, presencia de niebla, velocidad, zonas curvas, entre otras. Por su parte, Ford ha desarrollado un sistema adaptativo de luces delanteras que es capaz de crear patrones de iluminación bien definidos, aunque en la página que ha dedicado al tema, no deja muy claras las condiciones de activación para cada uno de ellos (Porsche Latin America, 2017b; Audi AG, 2011a; Daimler AG., 2016c; Ford Motor Company, 2017b). En la figura 2.7 se presentan algunos ejemplos de iluminación ofrecida por los DLS implementados actualmente.



(a)



(b)

Figura 2.7: Ejemplos de iluminación adaptativa ofrecida en el mercado; (a) patrones de iluminación utilizando faros LED dinámicos (Ford Motor Company, 2017b), (b) iluminación adaptativa de acuerdo al entorno de conducción (Audi AG, 2011a).

Asistente de visión nocturna (NVA)

Durante los trayectos realizados en la noche el asistente de visión nocturna ofrece un punto de referencia visual que va más allá de lo que los faros pueden aportar, de esta manera ayuda a que el conductor pueda reaccionar mejor ante la invasión de algún cuerpo en carretera.

Haciendo referencia a lo descrito por (Morán Cruz, 2013), estos sistemas son ofrecidos como un aditamento en automóviles de gama alta, y los hay de dos tipos; activos y pasivos, en los primeros se emplean sistemas de captación de imagen basados en luz infrarroja, mientras que los segundos trabajan con detección de calor mediante cámaras termográficas. Sin embargo, algunos sistemas ya combinan ambas tecnologías, tal es el caso de Porsche cuyo sistema de visión nocturna se compone de una cámara de infrarrojos que detecta a los peatones o animales grandes antes de que estos sean iluminados por los faros, se apoya también de una cámara termográfica para poder mostrar la imagen al conductor mediante la pantalla del cuadro de instrumentos; si el cuerpo detectado se encuentra a una distancia crítica se activa una advertencia acústica y el cuerpo se muestra en color rojo (Porsche Latin America, 2017a). En la Figura 2.8



Figura 2.8: A la izquierda, (a) representación gráfica del NVA de Mercedes-Benz basado en cámaras infrarrojas (Daimler AG., 2016c); a la derecha, tomada de (Audi AG, 2016), (b) imagen ilustrativa sobre el funcionamiento del NVA implementado por AUDI utilizando cámaras termográficas.

Aviso de salida de Carril

Al igual que para el resto de los ADAS expuestos en esta sección, este asistente puede ser tan complejo o simple como la gama a la que se haga referencia.

La función de este sistema es la de alertar un cambio de carril cuando este no parezca ser seguro. En algunos casos, se basa en una cámara instalada en el espejo retrovisor, con el lente hacia adelante, de tal manera que pueda adquirir información sobre las líneas entre carriles para determinar si son líneas continuas o no; si se detecta que el automóvil se encamina a cambiar de carril sin que se haya encendido la luz de giro (direccional), se activa una alarma vibratoria en el volante para alentar al conductor (Audi AG, 2011b). Algunos otros sistemas dan un plus activando una alarma sonora si la vibratoria ha sido ignorada.

Marcas como Mercedes-Benz ofrecen sistemas que se complementan con sensores de radar que ayudan a monitorear el entorno de conducción para alertar en caso de que se este a punto de cambiar de carril cuando en él se encuentre un auto y se origine la posibilidad de una colisión; de acuerdo con Mercedes-Benz, su sistema es capaz de analizar el comportamiento del conductor al interior del automóvil para decidir si se realiza el cambio de carril de forma intencional o no y establecer en qué condiciones activar las alertas (Daimler AG., 2016a).

Detección de fatiga

Los sistemas de detección de fatiga, dependiendo del desarrollador, analizan diversas variables que podrían indicar fatiga en el conductor. Algunos, como el de Ford, se basan en las imágenes obtenidas por una cámara frontal mediante las que se compara el comportamiento actual del conductor con el mostrado en un lapso de 6 minutos; en caso de encontrar signos de fatiga o falta de atención se activan alarmas visuales y auditivas cuya intensidad crece si el usuario no las desactiva manualmente, aunque no se define claramente los parámetros que considera (Ford Motor Company, N.d.).

Nissan por su parte, ofrece un sistema de detección de fatiga basado en el patrón de movimientos del volante (Nissan, 2017); Volkswagen añade a su algoritmo factores tales como el tiempo que ha durado el trayecto, así como la hora de día en que este ocurre (Volkswagen,

2016a); algunos, como Mercedes-Benz, monitorean aún más factores, como la aceleración longitudinal y lateral y las acciones de control que haya realizado el conductor en un intervalo de tiempo de entre 15 y 20 minutos (Daimler AG., 2016b).

Sin embargo, es el sistema de Jaguar Land Rover, desarrollado en combinación con Seeing Machines e Intel quien ha incluido sistemas de monitoreo visual que analiza los movimientos faciales del conductor de tal manera que le es posible definir si éste se encuentra cansado o distraído (Jaguar Land Rover, 2017). Sin embargo es bien sabido que los automóviles fabricados bajo la marca de Jaguar no son fácilmente adquiridos por los consumidores de clase media, debido a sus elevados costos. En la Figura 2.9 se muestra un ejemplo de alarma visual usada para alertar sobre la detección de factores indicativos de fatiga en el conductor.



Figura 2.9: Ejemplo de alerta visual desplegada por el sistema de detección de fatiga ofrecido por Nissan (Nissan, 2017).

2.5. Sistema de detección de fatiga y distracción

El sistema de detección de fatiga y distracción es uno de los ADAS en los que actualmente se trabaja, su función es dar aviso cuando detecte signos de cansancio en el conductor para que se tomen las medidas necesarias para evitar accidentes, lo que implica que se detenga la marcha y se tome un tiempo de descanso antes de continuar con el viaje. Las alertas usualmente generadas por los ADAS son visuales, auditivas y ocasionalmente, vibratorias (Biondi *et al.*, 2017); sin embargo, es necesario que dichas alarmas se encuentren diseñadas apropiadamente para evitar que sean los propios sistemas de asistencia los que distraigan al conductor.

La distracción ocurre cuando los conductores no se concentran en conducir (Regan y Strayer, 2014). Por ejemplo, además de controlar el vehículo, los conductores realizan tareas

secundarias no relacionadas con la conducción. En Strayer *et al.* (2011) se identifican tres fuentes de distracción: visual; cuando la vista no se enfoca en la carretera, manual; las manos no se encuentran en el volante, y cognitivas; cuando la atención no se centra en la conducción del automóvil.

Actualmente existen diversos métodos para la detección de fatiga en el conductor, entre los que podemos encontrar; gafas para monitoreo del comportamiento ocular del conductor, cámaras con visión nocturna incluida, e incluso sistemas altamente sofisticados que permiten el monitoreo de los signos vitales del conductor.

Hay, además, muchos estudios que abordan la idea de monitorear las características biológicas del conductor mediante las que se infiera su estado de salud. Los científicos han tratado de evaluar la carga cognitiva mediante electroencefalograma, mientras que el ritmo cardíaco y la presión arterial pueden ser medidos a través de un electrocardiograma, la temperatura corporal por termómetros y el nivel de somnolencia por una electrooculografía (Liang *et al.*, 2009; Zhao *et al.*, 2012). Sin embargo, al juntar cada uno de dichos procedimientos, resulta un conjunto de dispositivos cuya incorporación resulta incómoda para el usuario, y que, dada su naturaleza, no aportan datos tan útiles como los que podrían ser recopilados por sistemas de aplicación a distancia (Jung *et al.*, 2012).

Por su parte, mucha de la bio-información que pudiera emplearse para la detección de somnolencia presenta variaciones dependiendo de las características específicas de cada individuo, como su edad, genero, nivel de experiencia de manejo, así como los atributos relacionados al automóvil; como el espacio al interior, las capacidades del volante, la respuesta al cambio de velocidades y el desempeño del motor. Afortunadamente, se puede obtener información fidedigna acerca del conductor mediante métodos mucho más convenientes; uno de ellos es la visión por computadora.

La visión por computadora es un campo de estudio amplio, usado en numerosas aplicaciones; desde las más sofisticadas, como reconocimiento de patrones para diversos procedimientos tecnológicos, hasta las más simples, tales como el conteo de animales en grandes granjas.

Cuando se habla de visión por computadora en ADAS hallamos varias líneas de investigación, de estas podemos distinguir entre las que se centran en el entorno externo al automóvil, como; detección de cambio de carril, alerta de detección de peatón en carretera, y aquellos cuyo enfoque está al interior del vehículo, en tal clasificación se encuentran principalmente; el indicador (estimador) de fatiga y el sistema encargado de analizar el comportamiento del conductor. Es frecuente encontrar la aplicación de seguimiento ocular, en ADAS, con la intención de estudiar somnolencia y fatiga en conductores.

La somnolencia puede ser causada por muchos factores; largas jornadas de manejo, estrés, desordenes del sueño, incluso el uso de fármacos (Mihai *et al.*, 2015). Y de la misma manera, puede ser causante de varios eventos; básicamente, la sensación de sueño disminuye la concentración, la atención visual y el nivel de vigilancia del conductor, lo que incrementa la posibilidad de sufrir algún accidente de tránsito.

Los recientes avances en visión por computadora nos permiten tomar ventajas de las características visuales que se desprenden de la somnolencia, lo que es realmente aprovechable en los ADAS encargados de monitorear el cansancio en el conductor, pues permitiría estimar el nivel de somnolencia en el usuario para poder alertarle respecto a su estado actual de conducción, oportunamente. El nivel de somnolencia produce variadas pistas que pueden ser objeto de análisis, como la velocidad de parpadeo, cambio en expresiones faciales, movimientos visuales, bostezos (existencia y frecuencia), y la posición de la cabeza (Flores *et al.*, 2010).

De acuerdo con los análisis realizados, los científicos han clasificado los movimientos oculares en 5 tipos; sacádicos, de seguimiento, vestibular, vergencias y nystagmus fisiológico (Duchowski, 2007). Durante la conducción, los ojos realizan, de forma continua, movimientos de cualquiera de las ya mencionadas categorías con la finalidad de colocar la fovea (parte del ojo donde se enfocan los rayos luminosos) en la posición correcta para estabilizar las imágenes retínales.

Cada uno de nosotros realiza más de 100 000 movimientos del tipo sacádico por día. Los investigadores están de acuerdo en el hecho de que este tipo de movimientos pueden ser tanto voluntarios como involuntarios. Los movimientos de seguimiento son aquellos en que

el sujeto sigue, con la mirada, algún objetivo en movimiento; específicamente cuando el movimiento ocular se liga a la velocidad del objetivo. Los movimientos oculares de vergencia pueden ejemplificarse con el proceso de enfoque de imagen de una fotografía; aparece cuando el sujeto trata de ver objetos colocados a diferentes distancias pero en el mismo eje visual. Debido a la anatomía del ojo, los movimientos de vergencia son 20 veces más lentos que los sacádicos. Las fijaciones son los movimientos oculares que estabilizan la retina sobre un objeto estático de interés. Todos esos movimientos oculares deberían proporcionar información sobre la atención visual del conductor. Sin embargo, de acuerdo con Bărbuceanu *et al.* (2011), los más importantes para la obtención de datos sobre el estado de atención del sujeto y sus procesos de razonamiento son los sacádicos, las derivas y las fijaciones.

Dado que con el movimiento ocular viene también el parpadeo. La mayoría de los estudios realizados sobre somnolencia estiman una bio-característica específica: el porcentaje de cierre de los ojos (PERCLOS). De Dingus y Grace (1998) sabemos que el PERCLOS ha sido considerado un importante indicador de fatiga por al menos dos décadas. Para realizar el cálculo de PERCLOS mediante métodos basados en visión se deben seguir algunos pasos. La mayoría de los algoritmos se enfocarán primero en la localización o detección de la cara del sujeto bajo estudio; en esta parte del proceso se recomienda la mezcla de diversas técnicas para que se obtenga un algoritmo robusto que reduzca la posibilidad de errores, tales técnicas pueden incluir clasificadores Haar, Viola-Jones, modelo de emparejamiento de punto de referencia, segmentación de color, entre otras. Después de realizar la segmentación de imágenes, las partes restantes deberán ser clasificadas; para ello podría ser útil el desarrollo de alguna red neuronal, un algoritmo de optimización de enjambre de partículas o bien uno de emparejamiento con platillas (Coetzer y Hancke, 2011; Senaratne *et al.*, 2011; Bär *et al.*, 2012).

Posterior a la correcta detección del rostro, es necesario ubicar los ojos para seguirlos durante los fotogramas posteriores. El problema principal de este tipo de acercamiento es la alta sensibilidad a las diferencias de iluminación. Los investigadores han tratado de compensar dichas deficiencias con un mejor equipo de seguimiento, los cuales pueden ser divididos en remotos y montables o instalables en la cabeza (como cascos o lentes). Ambos tienen ventajas y desventajas.

Los sistemas remotos restringen el movimiento del sujeto a una distancia máxima de enfoque, pero son amigables con el usuario y son los más adecuados para la compilación de cuantiosos datos naturalistas para aplicaciones en las que el sujeto permanecerá relativamente quieto respecto al dispositivo de seguimiento. Por otro lado, mientras que los sistemas a instalar en la cabeza suelen ser incómodos son capaces de registrar os movimientos oculares con mucha más exactitud. Ejemplo de ello son las técnicas de seguimiento para las que se emplea el efecto de ojos rojos en cercanía la luz infrarroja para calcular la posición de la pupila. Los diodos emisores de luz infrarroja se colocan directamente en línea con el eje óptico de la cámara de vídeo, para lograr que la pupila de sujeto brille. Por lo tanto la pupila es más fácilmente detectable usando la diferencia entre las imágenes de pupilas brillantes y oscuras. Sin embargo, existen muy pocos automóviles con alguna fuente emisora de luz infrarroja, lo mismo ocurre con los actuales servicios móviles.

Mientras que, empleando únicamente técnicas clásicas de visión por computadora, la situación o el estado de los ojos puede determinarse mediante dos aproximaciones: analizando patrones desde el contorno del ojo con clasificadores de imagen o analizando el movimiento del parpado superior para estimar la dirección del movimiento. Otra forma de abordar este problema es mediante la segmentación de piel, basada en aproximación NN (Nearest Neighbor) de un histograma RGB del color de la piel (Lenskiy y Lee, 2012).

Hablando de la investigación sobre el comportamiento del conductor desde la perspectiva de visión por computadora, se sabe que esta gira alrededor de pocos pero importantes conceptos; orientación de la cabeza, dirección de la mirada y punto de intersección de la misma (GIP). Aun cuando esta área mantiene aspectos en común con la investigación acerca de la somnolencia en el conductor, también posee respecto a esta marcadas diferencias; estudiar el comportamiento del conductor se basa básicamente en encontrar tanto el punto de interés (POI) como la región de interés (ROI) configurados por el usuario. Es importante mencionar que en desarrollo de la presente tesis, este estudio sobre el comportamiento del conductor será aplicado únicamente para determinar el grado de atención del conductor. Por lo que los aspectos aplicables serán, la determinación de la orientación del rostro y la dirección de la vista.

El registro del movimiento ocular y su análisis aporta pistas de vital importancia para comprender las intenciones del usuario y así poder realizar acciones con las que se pueda prevenir algún accidente automovilístico, o bien, reducir sus consecuencias.

Los sistemas desarrollados actualmente, emplean de forma más común los siguientes parámetros; posición ocular, vector visual, GIP y datos de fijación. De los tres tipos de movimientos oculares más relevantes (derivadas, sacádicas y fijaciones) el más importante de todos parece ser el de las fijaciones. Desde hace aproximadamente dos décadas se ha concluido que la forma de definir las fijaciones, dada la aplicación, es la siguiente; una fijación es el acto de concentrarse en un punto cercano al foco de expansión que no es sino la línea horizontal en la que los objetos parecen anclados. Tales fijaciones, en este caso, se centran en detalles del camino o en señales de tránsito (Chapman y Underwood, 1998).

Para establecer si el conductor mantiene la vista en el camino, no es necesario poseer amplios detalles sobre la dirección de la vista, sin embargo es importante determinarla de tal forma que no se creen falsas alarmas. Para ello se evalúa la dirección de la vista en un intervalo que pueda ser considerado lo suficientemente amplio sin que genere fallas, y la forma más sencilla de hacerlo es determinar la posición facial, pues esta delimita ciertos parámetros para el ángulo en que el conductor puede estar dirigiendo la mirada.

Existen cuantiosos métodos para determinar la dirección facial, desde los más simples como la localización de los ojos, hasta algunos como el desarrollo de un modelo de apariencia activa (AAM) para encontrar características faciales en diversas circunstancias (Nuevo *et al.*, 2006). Dentro de los métodos basados en formas se puede encontrar la técnica del modelo facial cilíndrico (Kwon *et al.*, 2006), desafortunadamente, este método podría no ser adecuado para el modelado de la cabeza del conductor pues la estimación de parpadeo es inexacta cuando este gira la cabeza de forma considerable. Los métodos basados en características de textura funcionan encontrando la cara, o parte de ella, en vídeo y analizando la intensidad del patrón de la imagen de la cara para estimar su orientación. Actualmente, los métodos de seguimiento facial emplean una combinación de las técnicas mencionadas anteriormente, con la finalidad de ganar confiabilidad y robustez (Murphy-Chutorian y Trivedi, 2009).

Finalmente, es importante decir que, de acuerdo con Mihai *et al.* (2015) para ambas áreas de interés abordadas en este proyecto (fatiga y comportamiento del conductor), son muy pocos los estudios en los que se analiza la posibilidad de emplear únicamente una cámara como la única entrada de información para el sistema. La mayor parte de los trabajos se basan en el empleo de equipo especial de laboratorio desarrollado específicamente para este tipo de tareas.

2.6. Estudios relacionados

A continuación, en la tabla 2.2, se muestra un breve extracto de algunos de los trabajos relacionados con el proyecto que se desarrolla a lo largo de la presente tesis. Se eligieron para tal extracto aquellos trabajos de investigación cuyo aporte es fundamental para cumplimiento de los objetivos planteados en la sección 1.3.

De cada uno de los trabajos de investigación seleccionados se ha incluido el nombre con el que fueron publicados, el año de tal publicación, los autores de dichos trabajos y, finalmente, en la columna «Relación o aporte» se describe brevemente la razón por la que resultaron de tal relevancia para el desarrollo de esta tesis.

Breve registro de estudios relacionados con esta investigación			
<i>Nombre del estudio</i>	<i>Año de publicación</i>	<i>Autor (es)</i>	<i>Relación o aporte</i>
Changes in Physiological Parameters Induced by Indoor Simulated Driving: Effect of Lower Body Exercise at Mid-Term Break	2009	Liang, Wen Chieh; Yuan, John; Sun, Deh Chuan; Lin, Ming Han	Detección de somnolencia y otros factores de riesgo en conductores mediante la medición de bio-características

Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator	2012	Zhao, Chunlin; Zhao, Min; Liu, Jianpin; Zheng, Chongxun	Detección de somnolencia y otros factores de riesgo en conductores mediante la medición de bio-características
Finding People in Images and Videos	2006	Dalal, Navneet	Investigaciones que aplican el procesamiento de imágenes en ADAS diferentes a detección de somnolencia
Pedestrian Detection using a boosted cascade of Histogram of Oriented Gradients	2014	Ruiz Sancho, Cristina	
Lane Departure Identification for Advanced Driver Assistance	2015	Gaikwad, V; Lokhande, S	
Measurement of Driver's Consciousness by Image Processing -A Method for Presuming Driver's Drowsiness by Eye-Blinks coping with Individual Differences	2006	Suzuki, M; Yamamoto, N; Yamamoto, O; Nakano, T; Yamamoto, S	Detección de fatiga mediante análisis de un solo indicador visual
Driver fatigue detection based on eye tracking	2008	Devi, Mandalapu Sarada; Bajaj, Preeti R.	
Rapid object detection using a boosted cascade of simple features	2001	Viola, P.; Jones, M.	Aportes en detección de objetos; características y clasificadores
DHistograms of oriented gradients for human detection	2005	Dalal, N.; Triggs, B	

One Millisecond Face Alignment with an Ensemble of Regression Trees	2014	Kazemi, V.; Sullivan, J.	Presentación de detector de <i>landmarks</i> en rostros
---	------	--------------------------	---

Tabla 2.2: Breve registro de algunos de los estudios relacionados con este trabajo de investigación, empleados como fuentes bibliográficas a lo largo de este documento.

3. FUNDAMENTOS TEÓRICOS

3.1. Detección de objetos

La detección de objetos, atendiendo a la definición dada por Amit y Felzenszwalb (2014) y Ruiz Sancho (2014), puede ser definida como el procedimiento computacional cuyo objetivo es determinar si un de objeto, perteneciente a una clase específica, existe dentro de una imagen digital dada y, de ser así, localizarlo.

De acuerdo con Ruiz Sancho (2014) existen cuatro conceptos que deben diferenciarse claramente para que la definición de detección de objetos sea clara:

- Detección de la presencia del objeto, también llamada *soft detection* (Jalled y Voronkov, 2016): Establecer si una o más instancias de una clase de objetos específica está presente en una imagen dada, sin importar su localización.
- Detección\localización de objetos o *hard detection* (Jalled y Voronkov, 2016): Determina la existencia de instancias de una clase específica de objetos dentro de una imagen de entrada, además de indicar su localización específica.
- Reconocimiento de objetos: Detecta, localiza y, además, precisa la identidad de cada una de las instancias de una clase de objetos específica dentro de una imagen.
- Categorización de objetos: Determina la clase a la que pertenece cada objeto dentro de la imagen.

Es importante considerar que la apariencia de un objeto, perteneciente a una clase específica, puede variar drásticamente, ya sea por el proceso de obtención de la imagen (cambios en iluminación, digitalización, instrumentos utilizados, etc.) o por las diferencias intrínsecas en los objetos de una misma clase (tamaño, color, posición, etc.). Dado lo anterior, se puede intuir que el reto mayor del desarrollo de un algoritmo para la detección de objetos es el hecho de que este habrá de funcionar de forma invariable a pesar de las variaciones presentes en la apariencia del tipo de objeto a detectar\localizar (Amit y Felzenszwalb, 2014).

La detección de objetos se compone, generalmente, de dos sub-procesos; extracción de características y clasificación. El primero de ellos se encarga de codificar la apariencia visual de la imagen en características predefinidas, por su parte, durante la clasificación se utiliza la información obtenida en el primer sub-proceso para determinar si la imagen contiene o no el objeto de interés (Ruiz Sancho, 2014).

De acuerdo con Amit y Felzenszwalb (2014), los métodos de detección de objetos pueden clasificarse en generativos y discriminatorios. Los primeros se basan en modelos; suma los modelos de variabilidad de pose del objeto, de apariencia del objeto y el de fondo; en ellos, por ejemplo, se puede establecer un modelo para el objeto de interés, de tal manera que el sistema trabajará en encontrar coincidencias entre dicho modelo y diferentes partes de la imagen. Para los segundos, generalmente, se crea un clasificador que pueda distinguir entre imágenes positivas (contienen el objeto de interés) y las negativas (el objeto de interés no existe en ellas).

Sin embargo, en el trabajo de Mohan *et al.* (2001), se incluye una categoría más; aquellos que se basan en la coincidencia con un conjunto relacional de imágenes patrones. Sin embargo, una clasificación más detallada se puede encontrar en el trabajo de Shantaiya *et al.* (2013), en donde se puede ver un diagrama como el mostrado en la Figura 3.1. Cabe destacar que, aún cuando la terminología empleada para la clasificación de técnicas o enfoques para la detección de objetos en visión por computadora es muy variada, los conceptos básicos suelen coincidir de una fuente a otra.

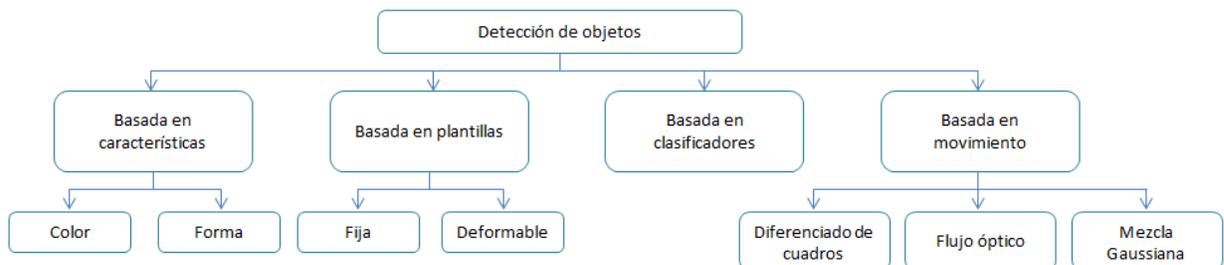


Figura 3.1: Clasificación de los métodos para detección de objetos, según Shantaiya *et al.* (2013).

3.1.1 Métodos basados en características

Este tipo de métodos deben su nombre a que los algoritmos utilizados reconocen a los objetos gracias a sus características específicas. Generalmente los objetos son descritos por varias características o atributos, para este tipo de métodos los atributos más comúnmente utilizados son; formas geométricas, color, tamaño y bordes. Cuando se utiliza este tipo de métodos, lo que se hace es, dada una imagen, buscar en ella el tipo de característica seleccionada, dicha característica se compara con la base de datos que contiene modelos del objeto de interés, para así determinar si el objeto se encuentra o no en la imagen de entrada (Achatz, 2016).

Las características y sus descriptores pueden dividirse en globales y locales, los primeros consideran, para la comparación, la imagen completa, mientras que los segundos utilizan pequeñas partes de la imagen, estos últimos son considerados más robustos que los primeros (Achatz, 2016).

3.1.2 Métodos basados en plantillas

Comúnmente usados debido a su simplicidad, si la plantilla del objeto de interés se encuentra disponible, la detección de objetos puede tratarse de comparar la plantilla (M) con la imagen de entrada (I). Los cálculos, para este tipo de detección, son sencillos; habrá de realizarse una comparación pixel por pixel, para ello deberá existir un umbral que permita definir si la imagen coincide con la plantilla o no (ver ecuación 3.1).

$$r(x, y) = \sum_{i \in M} \sum_{j \in M} (I(x + i, y + j) - M(i, j))^2 \quad (3.1)$$

Puede aplicarse un ajuste al resultado de la ecuación 3.1 para compensar pequeñas distorsiones, mediante la ecuación 3.2, em donde n se refiere al número de pixeles en M .

$$r = \frac{\sum IM - \sum I * \sum M}{\sqrt{(n \sum I^2 - (\sum I)^2) (n \sum M^2 - (\sum M)^2)}} \quad (3.2)$$

Esta clase de métodos, tal como se observa en la Figura 3.1, se divide en dos sub-tipos, los de plantilla fija y los de plantilla deformable; los primeros son útiles cuando la formas del objeto no cambia al cambiar el ángulo de visión de la cámara, mientras que los segundos son aplicables cuando el objeto podría sufrir deformaciones que afecten su apariencia; una plantilla deformable se obtiene aplicando una transformada de deformación parametrizada en el prototipo y una distribución probabilística de los parámetros de deformación para controlar la variación de la plantilla deformable.

3.1.3 Métodos basados en clasificadores

Según lo escrito por Shantaiya *et al.* (2013), este método fue propuesto para la extracción de un solo objeto en un cuadro de vídeo. Dicho método propone separar los objetos de la escena del fondo, para ello, la imagen o cuadro de vídeo se divide en bloques pequeños del mismo tamaño, los cuales se clasificaran en bloques de objeto o de fondo, dependiendo de si su pixel central pertenece al objeto buscado o a la escena de fondo. Posterior a una segmentación manual del primer cuadro de vídeo, los bloques obtenidos son utilizados como muestras para entrenar el clasificador que diferenciará entre objeto y fondo. Otra manera de generar un clasificador es entrenándolo mediante muestras negativas (imágenes que no contienen el objeto de interés) y positivas (imágenes en las que aparece el objeto de interés), para objetos más complejos es necesario integrar múltiples clasificadores, lo que se conoce como clasificadores en cascada o *boosting*.

3.1.4 Métodos basados en movimiento

Los métodos de detección basados en movimiento, se utilizan frecuentemente en la tarea de detectar y seguir un objeto de interés en una secuencia de vídeo. Uno de ellos es el método de diferenciado de cuadros, el cual se basa en calcular la diferencia entre dos cuadros de vídeo sucesivos. Otro de ellos el de flujo óptico, mediante este método se puede obtener toda la información del movimiento de un objeto y presenta un 85 % de exactitud, sin embargo la gran cantidad de cálculos necesarios para que pueda implementarse es un gran desventaja, a lo que se le suma su alta sensibilidad al ruido (Sukanya *et al.*, 2016). Por último, se tiene el método

de mezcla Gaussiana, en este método se modelan los valores de un pixel como una mezcla Gaussiana, con base en la persistencia y cambios de cada una de las Gaussianas de la mezcla se determina cual de ellas pertenece a los colores de fondo de la imagen, los valores de los pixeles que no se encuentren dentro de las distribuciones del fondo serán considerados como parte del primer plano, hasta que alguna mezcla Gaussiana los ubique con certeza suficiente en el fondo de la imagen (Stauffer y Grimson, 2000).

3.2. Detección de rostro

El término «detección de rostro» hace referencia a los mecanismos computacionales mediante los cuales es posible determinar si existe o no un rostro en una imagen digital dada, y en su caso determinar su posición en la misma, así como su tamaño (Modi y Macwan, 2014).

La detección de objetos es una tarea por demás estudiada y aplicada en el ramo de la visión por computadora, por ello no es de sorprender la cantidad tan elevada de trabajos relacionados con el tema. Para detectar un objeto específico, lo que se hace es buscar en la imagen aquellas características visuales que son inherentes a él, no obstante, cuando se trata de localizar o detectar un rostro el grado de dificultad se eleva dado que los rostros pueden encontrarse en un amplio rango de poses, con una infinidad de gestos y detalles visuales que modifican su apariencia; todo aquello que visualmente ayuda a nuestro cerebro a distinguir entre una persona y otra, resulta en un grado extra de complejidad en el caso de la localización facial mediante el procesamiento de imágenes y visión por computadora (Dalal y Triggs, 2005). En su trabajo, Ming-Hsuan Yang *et al.* (2002), en lista aquellos factores que se relacionan directamente con los retos de la detección de rostros, a continuación su listado:

- La variabilidad de componentes estructurales: La diferencias entre rostros incluye características intrínsecas del rostro, como la forma, color y tamaño, así como variaciones en la presencia o ausencia de características más específicas, como barba, bigote y lentes, por ejemplo.
- Pose: La posición en la que sea captado un rostro mediante la cámara influirá de forma directa en la apariencia de esta, la pose podría afectar la visibilidad de algunas partes

esenciales en la descripción de un rostro.

- **Expresión Facial:** Esta afecta de forma inmediata y directa la apariencia del rostro.
- **Oclusión:** Hace referencia a la posibilidad de encontrar, en la imagen de entrada, rostros parcialmente ocultos; ya sea por otra clase de objetos o por otros rostros.
- **Orientación de la imagen:** La posición de la cámara respecto al rostro que desea capturarse afectará la forma en que visualmente se aprecia el rostro.
- **Condiciones de la imagen:** Engloba todos los factores que pueden afectar el proceso de adquisición de imagen, tales como las características de la cámara, la iluminación, el movimiento, etc.

3.2.1 Técnicas de detección

Tanto en el trabajo de Hjelmas y Low (2001) como de Modi y Macwan (2014) se puede encontrar una clasificación de las técnicas empleadas para la detección de rostros, tal clasificación hace distinción del enfoque que se toma para llevar a cabo la tarea de detección asignada (ver Figura 3.2). De acuerdo a las citadas fuentes, existen dos tipos de detectores, según su enfoque, para realizar la detección de rostro; los basados en las características faciales y los que se basan en la imagen. En los primeros la detección se lleva a cabo mediante el análisis de ángulos, longitudes y áreas de las características visuales en búsqueda de determinados elementos que componen un rostro, por otra parte, los segundos se basan en comparar las imágenes de prueba con otras llamadas *de entrenamiento*.

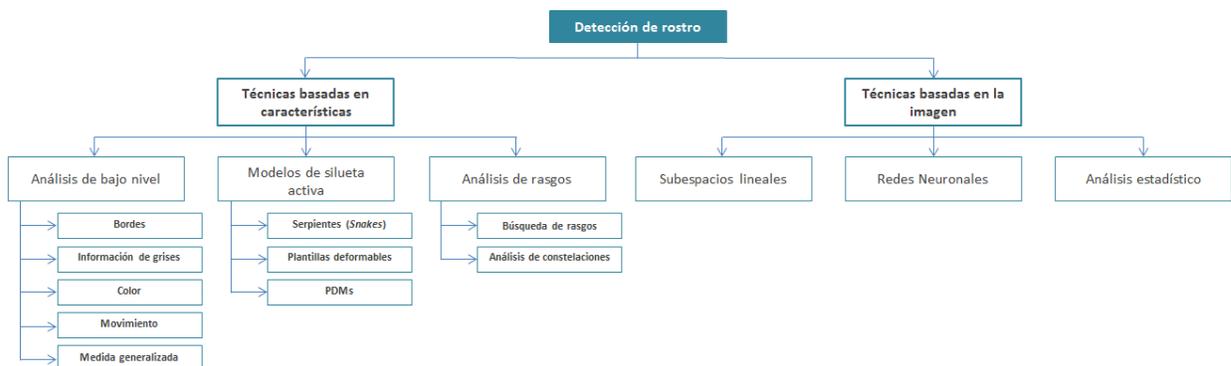


Figura 3.2: Técnicas para la detección de rostros, clasificadas según su enfoque.

Técnicas basadas en características

El desarrollo de este tipo de técnicas se divide en tres áreas, tal cual se observa en la Figura 3.2, las cuales se describen brevemente a continuación:

1. Análisis de bajo nivel:

- **Bordes:** Esta es una de las primeras en aplicarse para la detección de rostros. Para esta técnica, es necesario que los bordes se etiqueten de tal manera que puedan compararse con un modelo facial para verificar que las detecciones son correctas, de acuerdo con Hjelmas y Low (2001) Venu Govindaraju lo hizo etiquetando los bordes como lado izquierdo, línea capilar, o lado derecho de un rostro visto de frente, los comparó entonces contra un modelo facial usando el número áureo o número de oro para un rostro ideal (véase Ecuación 3.3).

$$\frac{\text{alto}}{\text{ancho}} = \frac{1 + \sqrt{5}}{2} \quad (3.3)$$

El algoritmo propuesto por Govindaraju, según Hjelmas y Low (2001), se basa en los siguientes pasos:

- Detección de bordes, estos se detectan mediante la utilización del operador de Marr-Hildreth.
- Adelgazamiento, este se realiza para que la línea representativa de cada borde sea de un pixel de ancho.
- Eliminación de picos, para ello cada componente conectado entre sí se reduce a su rama central.
- Filtrado, en este paso el algoritmo se encarga de seleccionar solamente los componentes con mayor probabilidad de formar parte de un rostro.
- Detección de esquinas, los componentes se separan basándose en dicha detección.
- Etiquetado, se realiza sobre los componentes que han pasado las etapas anteriores. Las etiquetas corresponderán a los lados izquierdo o derecho, o bien a la línea capilar.

Se considera que este algoritmo es completamente ineficiente si el rostro no se encuentra frontal en la imagen.

- Información de grises: El funcionamiento básico de esta técnica se basa en el hecho de que algunas características faciales (pupilas, cejas y labios) son más oscuras que las áreas faciales que las rodean . El algoritmo propuesto para esta técnica es el siguiente (Hjelmas y Low, 2001; Modi y Macwan, 2014; CATAM, N.d.):

- Aumentar el contraste, se realiza para acentuar en mayor proporción la diferencia entre las características faciales buscadas y el resto de la imagen.
- Umbral, este se establece de tal manera que, a partir de aquí, solo se trabaje con las áreas de la imagen que lo superen, mismas que presumiblemente formen parte de las cejas, pupilas y labios.
- Aplicación de plantillas ponderadas, estas se emplean para comparar el conjunto de tonos (o manchas) obtenidas en los pasos anteriores y así definir la existencia o ausencia de rostros en la imagen.

El color de piel es un factor de gran peso en este algoritmo, por lo que los resultados pueden variar dependiendo de dicho factor. Para corregir dicho problema se da paso a la técnica basada en el color.

- Color: Debido a las dimensiones adicionales del color, esta técnica es considerada más poderosa que la basada en información de grises. Se encontró que los diferentes tonos de piel humana dan lugar a un reducido grupo en espacios de color, lo que significa que la composición de color de la piel humana difiere de forma mínima entre individuos. El modelo RGB es uno de los más empleados para la representación de color, este se basa en la combinación de rojo, verde y azul (las siglas RGB pertenecen a los nombres en inglés de tales colores). Dado que la apariencia de de la piel se ve afectada por la luminosidad, entre otros factores, se recurre a la normalización de los colores RGB, para lo que se emplean las Ecuaciones 3.4, 3.5 y 3.6 (Hjelmas y Low, 2001).

$$r = \frac{R}{R + G + B} \quad (3.4)$$

$$g = \frac{G}{R + G + B} \quad (3.5)$$

$$b = \frac{B}{R + G + B} \quad (3.6)$$

Los pasos básicos de los algoritmos aplicables para esta técnica son (Modi y Macwan, 2014):

- Clasificar la región de piel dentro de los espacios de color.
 - Seleccionar un umbral para aplicarlo mediante una máscara en la región de piel.
 - Delimitar el rostro para poder extraerlo de la imagen.
- **Movimiento:** Si en donde se requiere detectar rostros es una secuencia de vídeo, esta técnica resulta conveniente para la detección de objetos en movimiento. Puede usarse para la detección de rostros mediante el uso de diferenciación de cuadros; por ejemplo, para localizar un par de ojos se mide la distancia horizontal y vertical entre dos áreas adyacentes que presenten posibilidades de formar parte de un rostro.
 - **Medida generalizada:** De acuerdo con lo descrito por Hjelmas y Low (2001), esta se basa en la propuesta de que un sistema de visión artificial debería comenzar con un cómputo previo de bajo nivel de las propiedades generalizadas de la imagen. Para esta técnica se utiliza el operador de simetría sugerido por Reisfeld y Yeshurun (1992), basado en una noción intuitiva de simetría más que en una noción matemática formal de la misma. Este método ha resultado altamente eficaz en la localización de ojos y boca, con un 95 % de detección exitosa (Hjelmas y Low, 2001).
2. **Modelos de silueta activa:** Emplean la información de los rasgos locales de la imagen, bordes y brillo, y se deforma gradualmente de tal manera que vaya tomando la forma del rasgo. Como se puede ver en la Figura 3.2, se considera que existen tres tipos de modelos de silueta activa (Hjelmas y Low, 2001).
- **Serpientes (*snakes*):** El nombre hace referencia al contorno genérico activo empleado para la detección de los bordes del rostro. La serpiente se inicializa en un punto cercano al rostro, y poco a poco se va ciñendo a su contorno hasta tomar la forma de este. La evolución de la serpiente se lleva a cabo minimizando su función de energía (E_s), misma que depende de la energía interna (E_i) y externa (E_e), tal como se muestra en la Ecuación 3.7:

$$E_s = E_e + E_i \quad (3.7)$$

Es importante, en este método, definir los términos de energía y la forma en que se realizará la minimización de energía; a menudo se emplea energía elástica como E_i , mientras que la E_e depende del tipo de características, de la imagen, consideradas. Finalmente, la minimización de energía se puede llevar a cabo mediante técnicas de optimización.

- Plantillas deformables: Método propuesto por Yuille *et al.* (1992), de acuerdo con su trabajo, las plantillas se definen por un conjunto de parámetros que permiten un conocimiento anticipado de la forma que se espera tengan las características que guiarán el proceso de detección. Tales plantillas son lo suficientemente flexibles como para cambiar algunos de los valores de sus parámetros, entre ellos el tamaño, para poder igualar a los datos de entrada. Los valores resultantes de los parámetros, al final del procedimiento, pueden emplearse para describir las características.

Yuille *et al.* (1992) aseguran que su método debería funcionar sin importar la inclinación, las variaciones de escala, cambios de luz ni la rotación de la cabeza.

- Modelos de distribución de punto (PDMs): Esta técnica se basa en la aseveración de que una vez que las formas se representan como vectores, es posible aplicar métodos estadísticos estándar, tal como podría hacerse con cualquier objeto multivariado (Modi y Macwan, 2014). De acuerdo con Hjelmas y Low (2001), Lanitis *et al.* (1994) fueron quienes desarrollaron esta técnica para detección facial. De hecho, Lanitis *et al.* (1994) explican que dado un conjunto de vectores de entrenamiento (X) se puede construir un modelo lineal flexible para representar las variaciones de las características, tales modelos se pueden crear aplicando análisis de componente principal. En el modelo se incluye la media de todas las características que componen el conjunto, así como las principales formas de variación para cada punto; ello se ve representado en 3.8

$$x = \bar{x} + Pu \tag{3.8}$$

Lanitis *et al.* (1994) desarrollaron un modelo flexible de la forma facial utilizando 152 puntos definidos, a mano, sobre 160 rostros encontrados en las imágenes de entrenamiento. Usando 16 pesos (u) iniciales se puede obtener una aproximación de hasta 95 % de las formas faciales en el conjunto de entrenamiento. Para ajustar a un

rostro el modelo de distribución obtenido, se coloca la forma promedio cerca de la cara, enseguida se aplica una estrategia de búsqueda local de escala de grises para mover cada punto a su posición correspondiente en el contorno del rostro (Lanitis *et al.*, 1994; Hjelmas y Low, 2001).

3. Análisis de rasgos: Este tipo de técnicas se consideran más certeras que aquellas llamadas de bajo nivel, dado que estas últimas pueden fácilmente generar resultados ambiguos, mientras que en el análisis de rasgos se aplica el conocimiento de la geometría facial de tal manera que las características o rasgos sean detectados y verificados. A continuación, una breve descripción de las técnicas que se incluyen en este grupo (Hjelmas y Low, 2001).

- Búsqueda de rasgos: Se inicia con la localización de características faciales destacadas. Ello permite, mediante mediciones antropométricas del rostro, generar una hipótesis sobre la existencia de características menos prominentes. La referencia más usada para comenzar la detección de rostro con base en estas técnicas es el par de ojos. Dentro de estos se pueden encontrar métodos como el *Viola-Jones*, propuesto por Viola y Jones (2001) y los filtros de Gabor (Sharif *et al.*, 2011), (Modi y Macwan, 2014; Hjelmas y Low, 2001).
- Análisis de constelaciones: Las constelaciones están formadas por un grupo de características faciales, ello convierte a este en un método de detección más robusto. Su utilización es ampliamente recomendada cuando se trata de detectar rostros bajo condiciones no controladas, lo que implica rostros en diversas posiciones e imágenes con fondos complejos (Hjelmas y Low, 2001).

Técnicas basadas en la imagen

En estas técnicas se basan en el reconocimiento de patrones faciales mediante entrenamiento, el cual clasifica las imágenes en dos clases prototipo; las que contienen rostros (positivas) y las que no (negativas). Comparar dichas clases con un arreglo 2D de intensidad, obtenido de una imagen de entrada, es lo que le permite determinar la existencia o ausencia de rostros en una imagen. En seguida una breve descripción de los métodos que se encuentran en este grupo (Hjelmas y Low, 2001).

1. Subespacios lineales: La técnica fue desarrollada a finales de la década de los 80s por Turk y Pentland (1991), con un enfoque a reconocimiento de rostro, de acuerdo con lo escrito por Hjelmas y Low (2001). Turk y Pentland (1991) explican que mediante esta técnica se extrae la información relevante contenida en la imagen de un rostro, para después codificarla tan eficientemente como sea posible y comparar esta codificación con los modelos codificados de la base de datos creada.

La creación de la base de datos de modelos codificados es considerada como la etapa de inicialización el sistema, para ello se adquiere un conjunto inicial de imágenes, al que se le denominará conjunto de entrenamiento. De dicho conjunto se obtienen los componentes principales de la distribución de rostros, lo que matemáticamente puede definirse como los eigenvectores de la matriz de covarianza del conjunto de entrenamiento, considerando las imágenes como puntos o vectores en un espacio dimensional muy alto. Los eigenvectores se ordenan; cada uno contabiliza una cantidad diferente de variaciones entre las imágenes de rostros.

Los eigenvectores pueden ser considerados como un conjunto de rasgos que caracterizan la variación entre imágenes. Cada punto en la imagen contribuye en diferente medida a cada eigenvector, de tal manera que al representar los eigenvectores se puede obtener la imagen de un rostro borroso, a dicha imagen le han denominado *eigenface*. El nombre de esta técnica proviene del hecho de que cada rostro puede ser representado de forma exacta por una combinación lineal de *eigenfaces* (Turk y Pentland, 1991).

2. Redes neuronales: Esta técnica ha sido aplicada de forma exitosa en la detección de objetos, y dado que la detección de rostros puede ser considerada como un problema de reconocimiento de patrones de dos clases, se establece como viable la aplicación de este método para la detección de rostros. La principal ventaja de las redes neuronales se basa en que su entrenamiento permite la obtención de un sistema capaz de capturar e interpretar patrones faciales complejos, mientras que su mayor desventaja radica en la el tiempo que ha de dedicarse a la creación y ajuste de una arquitectura que permita a la red tener un rendimiento óptimo (Ming-Hsuan Yang *et al.*, 2002).
3. Análisis estadístico: Cuando se habla de detección facial por análisis estadístico se habla de sistemas capaces de detectar la existencia y localización de rostros en una imagen

mediante la aplicación de información teórica; dentro de este conjunto se encuentran las máquinas de soporte vectorial (SVM) y el análisis de componentes principales (PCA), (Modi y Macwan, 2014; Hjelm y Low, 2001).

3.3. Características

Las características o *features*, en términos de procesamiento de imágenes y visión por computadora, son considerados como atributos individuales, relevantes, que describen a un objeto y que son de gran utilidad para su reconocimiento y diferenciación de este entre otros (Ruiz Sancho, 2014).

Es común que, para el proceso de detección de objetos, se elija trabajar con características y no con valores de píxeles, el principal motivo de ello es que las primeras pueden codificar información de alto nivel, mientras que los segundos son más lentos. Existen dos formas de clasificar las características; acuerdo con su complejidad se clasifican como características de bajo o alto nivel (Negri *et al.*, 2007; Ruiz Sancho, 2014);

- *Características de alto nivel*: Este tipo de características requieren de un método de procesamiento previo, generalmente se basan en las características de bajo nivel. La extracción de estos ocurre mediante la búsqueda de formas y objetos dentro de una imagen digital (Nixon y Aguado, 2012; Ruiz Sancho, 2014).
- *Características de bajo nivel*: Son aquellos que pueden extraerse, de la imagen, de forma automática, sin ninguna información de forma (Nixon y Aguado, 2012).

Mientras que, de acuerdo con la información que proporcionan, se dividen en características locales y globales (Awad y Hassaballah, 2016):

- *Características globales*: Estas describen a la imagen como un todo, pueden considerarse como un atributo particular de la imagen, mismo que involucra a todos sus píxeles (Awad y Hassaballah, 2016).
- *Características locales*: Estas se centran en puntos clave y en la descripción de las áreas que rodean tales puntos, dicho de otra forma, representan una pequeña área del total de la

imagen (Awad y Hassaballah, 2016; Ruiz Sancho, 2014).

3.3.1 Detectores de características

Estos detectores, pueden ser clasificados en tres categorías; detectores a una escala, detectores multiescala y detectores afines invariantes. Los primeros son detectores que son invariantes a transformaciones de la imagen, es decir, pueden detectar las características del objeto ante cambios tales como desplazamiento, rotación y cambios de iluminación, sin embargo no funcionan ante cambios de escala. los detectores multiescala tienen la capacidad de lidiar con los problemas de escalamiento, a diferencia de los primeros. Por último, los detectores afines invariantes pueden trabajar incluso con imágenes tomadas desde diferentes perspectivas visuales Awad y Hassaballah (2016).

Atributos de los detectores de características

Es importante conocer las propiedades básicas que un detector de características debe poseer para cumplir con su objetivo; proveer una adecuada representación de un objeto determinado, de tal manera que este pueda ser fácilmente identificable de entre otros en las imágenes dadas. Tales características, según Awad y Hassaballah (2016) son:

- **Exactitud:** La localización de características debe llevarse a cabo de forma precisa, principalmente en tareas de emparejamiento.
- **Repetibilidad:** Debe ser capaz de repetir las detecciones de características de los mismos objetos o escenas bajo diferentes condiciones de visualización.
- **Robustez:** Debe detectar las mismas características sin importar si estas se han desplazado, si presentan rotación o cambios de escalamiento.
- **Generalidad:** Debe tener la capacidad de detectar características que puedan utilizarse en diferentes aplicaciones.
- **Eficiencia:** Debe realizar las detecciones de forma rápida, de tal manera que pueda ser empleado en aplicaciones en tiempo real.
- **Alta densidad:** La cantidad o densidad de las características detectadas debe corresponder a la información contenida en la imagen. Es decir, que el detector deberá detectar la

mayoría de las características existentes en la imagen de entrada.

3.3.2 Descriptores de características

Una vez que el conjunto de características ha sido localizado en la imagen, su contenido debe ser codificado en un descriptor, por lo que los descriptores habrán de entenderse como representaciones compactas del contenido de una región de la imagen (Awad y Hassaballah, 2016; Prince, 2012). Algunos de ellos se listan a continuación:

- SIFT: Algoritmo presentado por Lowe (1999), es considerado como un descriptor basado en distribución, se encarga de la obtención de características en imágenes en escala de grises. Las características de este algoritmo son invariantes ante cambios de rotación, escala, corrimiento, y parcialmente invariantes a cambios de iluminación y perspectiva 3D. Las principales etapas de computación empleadas para generar el conjunto de características de la imagen mediante el algoritmo SIFT, son (Lowe, 1999; Flores y Braun, 2011; Lowe, 2004; Nixon y Aguado, 2012):
 1. Detección de extremos en el espacio-escala: Mediante la función de diferencia Gaussiana se identifican potenciales puntos clave que son invariantes en escala y orientación en la imagen.
 2. Localización de punto clave: La selección de estos puntos se realiza mediante la medición de estabilidad de los puntos potenciales hallados en la etapa anterior.
 3. Asignación de orientación: Una o más orientaciones se asignan a la ubicación de los puntos de interés, con base en las direcciones del gradiente local de la imagen.
 4. Descriptor de puntos clave: Se define un descriptor para cada punto clave, basado en su entorno, que le permite niveles significantes de distorsión y cambios en iluminación.
- Histograma de Gradientes Orientados HOG: Este descriptor, abordado por Dalal y Triggs (2005), forma parte de los descriptores basados en distribución, de acuerdo con Nixon y Aguado (2012). Este descriptor trabaja en la construcción de una caracterización más detallada de la estructura espacial con una pequeña ventana de imagen. Se basa en la afirmación de que la apariencia local de un objeto puede caracterizarse bastante bien mediante

La distribución de los gradientes locales de intensidad o las direcciones de los bordes en la imagen. Las fases de este algoritmo, son (Ruiz Sancho, 2014):

1. Normalización en gama y/o color: Pre-procesamiento de la imagen para asegurar los valores normalizados de color y gama.
 2. Calculo del gradiente. Calcular los gradientes vertical y horizontal.
 3. Codificación de los histogramas locales de orientación: La imagen se divide en bloques y estos en celdas, conformadas por cierto numero de pixeles. Se genera un histograma por cada celda, este se concatena para formar un descriptor normalizado para cada bloque.
 4. Normalización: Se aplica una normalización local de contraste para contrarrestar problemas de iluminación y variaciones en el fondo de la imagen.
 5. Concatenación. Se concatenen los histogramas de cada bloque en la ventana de detección, para obtener un vector que fungirá como descriptor.
- SURF: según lo escrito por Awad y Hassaballah (2016), este detector-descriptor fue desarrollado por Bay *et al.* (2006) y se considera como una alternativa al SIFT, siendo el SURF mucho más eficiente y robusto, aún cuando se basa en propiedades similares a las del SIFT. El primer paso de este algoritmo consiste en establecer una orientación reproducible, basada en la información de una región circular que rodee al punto de interés. Lo siguiente se trata de construir un área cuadrada, alineada a la orientación seleccionada, para finalmente, extraer el descriptor.
 - LBP: De acuerdo con Ruiz Sancho (2014) este es el principal descriptor, basado en textura, empleado en la detección de peatones. El operador describe cada pixel respecto a los niveles de grises relativos de los pixeles contiguos; si el nivel de gris del pixel vecino es menor que el del pixel de referencia, el valor se establece en 0, mientras que si es mayor o igual, el valor será de 1. El descriptor mostrará el resultado como un número binario, de acuerdo con 3.9, (Heikkilä *et al.*, 2009).

$$LBP_{Rd,N}(x, y) = \sum_{i=0}^{N-1} s(n_i - n_c) 2^i, s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & \text{otro} \end{cases} \quad (3.9)$$

- Haar: Los descriptores Haar fueron desarrollados por Viola y Jones sobre las bases planteadas por Papageorgiou *et al.* in 1998 (Soo, 2014).

Este tipo de descriptores utilizan regiones rectangulares compuestas por píxeles vecinos dentro de las ventanas de detección, suman las intensidades de los píxeles en cada una de las regiones y calcula la diferencia entre tales sumas. Dicha diferencia se utiliza para clasificar sub-secciones en la imagen (Soo, 2014).

Pavani *et al.* (2010) comparten las imágenes de las características Haar, incluyendo en ellos los valores que por defecto suelen darse a cada rectángulo que los compone, tales, dichas imágenes se muestran en la Figura 3.3.

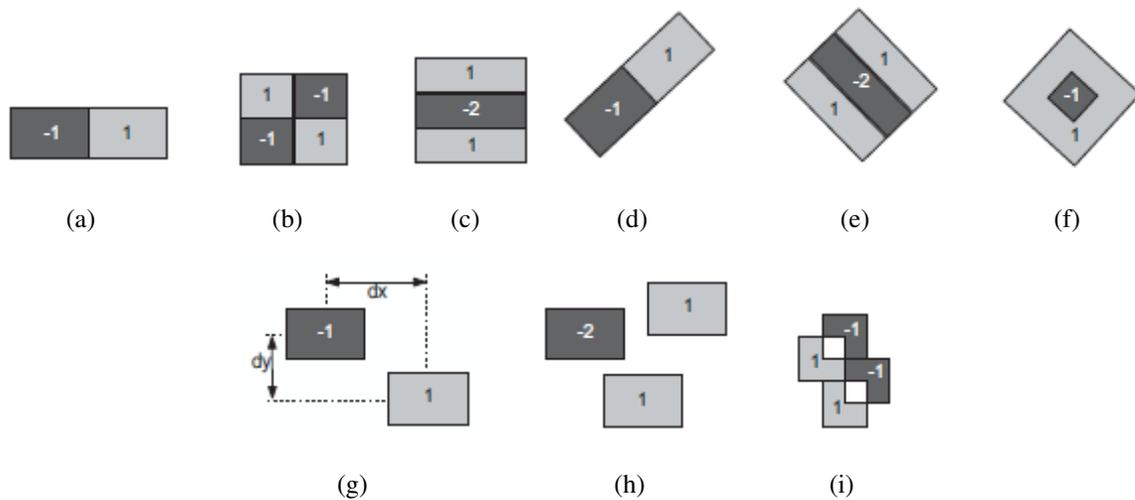


Figura 3.3: Características Haar: (a) y (b) características propuestas por Papageorgiou *et al.* (1998), (c) conjunto de rectángulos creado por Viola y Jones (2003), las sub-figuras (d), (e) y (f) muestran las características Haar presentadas por Lienhart y Maydt (2002), mientras que (g) y (h) presentan los rectángulos vecinos propuestos por Li *et al.* (2002) y finalmente (i) pertenece a las características Haar en diagonal de Viola y Jones (2003).

3.4. Clasificadores

De acuerdo con Ruiz Sancho (2014), el proceso de clasificación es aquel mediante el que elementos individuales son agrupados de acuerdo con la similitud entre sus características y la descripción de grupos preestablecidos, de tal manera que a los elementos de entrada se les asignará una etiqueta correspondiente a diferentes clases o categorías predefinidas con características comunes. Mientras que Tan *et al.* (2006) Dice de la clasificación que esta es la tarea de

aprender una función objetivo, o modelo de clasificación, la cual mapea cada uno de los atributos de las instancias de entrada en una de las clases predefinidas.

Un clasificador, también conocido como técnica de clasificación, es un enfoque sistemático para construcción de funciones objetivo a partir de un conjunto de datos de entrada. Cada uno de ellos necesita de una técnica o algoritmo de aprendizaje para identificar la función que mejor describa la relación entre el atributo buscado y la la clase del objeto de entrada (Tan *et al.*, 2006). De hecho, se considera que la clasificación es un proceso que consta de dos pasos (Ruiz Sancho, 2014):

- Etapa de aprendizaje o entrenamiento y construcción de modelo (o función objetivo).
- Fase de prueba o aplicación del modelo.

3.4.1 Técnicas de aprendizaje

Las técnicas de aprendizaje, en inteligencia artificial, pueden ser clasificadas como:

- Aprendizaje supervisado: Se caracteriza por necesitar, al menos, un vector de entrada y otro de salida; el primero de ellos contendrá la información que habrá de ser aprendida, mientras que el segundo, tal cual su nombre lo indica, contendrá la etiqueta de salida que habrá de considerarse correcta para dicho vector de entrada. En tipo de aprendizaje se realiza enseñándole al sistema la relación entre sus vectores de entrada y salida, estableciendo dicha relación como referencia del comportamiento que se espera que tenga. Como ejemplo de este tipo de aprendizaje pueden mencionarse las redes neuronales (Jimenez Moreno, 2011).
- Aprendizaje no supervisado: En este tipo de aprendizaje no se utiliza información previa, es conocido también como aprendizaje automático, la salida será la respuesta ante la similitud entre la entrada actual y la información que se le ha otorgado a la entrada en ocasiones previas. Está formado de un conjunto de reglas que ayudan a la red a aprender asociaciones entre los patrones que frecuentemente se les presentan, uno de los ejemplos de esta clase de aprendizaje es el *clustering* (Jimenez Moreno, 2011).
- Aprendizaje semi-supervisado: Este es una técnica que combina las cualidades de las téc-

nicas de aprendizaje supervisado y no supervisado, en ella el aprendizaje automático es reforzado mediante un entrenamiento previo (Jimenez Moreno, 2011).

3.4.2 Técnicas de clasificación

- Máquinas de vectores de soporte (SVM): Técnica que tiene su raíz en la teoría de aprendizaje estadístico, y fue propuesta por Vladimir Vapnik en 1995. Es parte del conjunto de técnicas de aprendizaje supervisado, por lo que emplea un conjunto de datos de entrenamiento para obtener el modelo de clasificación y un segundo conjunto de datos que sirve para probar la eficiencia del modelo. La función de un clasificador de este tipo es separar los datos de entrenamiento en el espacio dimensional más grande mediante la aplicación de una función kernel y así maximizar el margen entre los objetos que pertenecen a una clase (positivos) y los que no (negativos), (Tan *et al.*, 2006; Ruiz Sancho, 2014; Matuska *et al.*, 2014).
- Árboles de decisión: Está es una técnica bastante popular, se trata de construir una estructura jerárquica formada por nodos y ramas que clasifican las muestras de entrada de acuerdo con lo establecido en los nodos. Cada nodo representa una condición de prueba de la característica buscada, cada rama representa el valor que podría tomar dicha característica o atributo. Los árboles de decisión están formados por tres tipos de nodos (Ruiz Sancho, 2014):
 - Nodo raíz; es el nodo de comienzo por lo que no tiene ramas entrantes pero sí una o más ramas de salida.
 - Nodos interno; son aquellos que forman el cuerpo del árbol y cuentan con una rama de entrada y al menos dos de salida.
 - Nodo hoja o nodo terminal; Este representa la decisión final en la clasificación y está asignado a una clase. Estos tienen una rama de entrada, pero ninguna de salida.
- *Boosting*: Fue presentado por Schapire en 1990 como un método que podría mejorar el desempeño de un algoritmo débil de aprendizaje, se basa en el hecho de que hallar varias reglas básicas que ayuden a la clasificación puede ser mucho más sencillo que encontrar una sola regla de clasificación que resulte precisa, por lo que se trata de crear un clasificador robusto a partir de varios denominados *clasificadores débiles*. Esta técnica

de clasificación ha sido mejorada por Freund en 1990 y expandida en 1996 por el mismo Freund; recibiendo el nombre de *AdaBoost* que hace referencia a *Adaptive Boosting* (Mejoramiento adaptativo), (Bauer y Kohavi, 1999; Schapire, 2003).

- Cascada: Esta es un técnica de aprendizaje conjunto, se llama cascada debido a que se encuentra formada por etapas o niveles. Cada nivel esta formado, generalmente, de un clasificador más simple que el que se espera obtener con la cascada. El proceso de clasificación mediante cascada se inicia en un primer nivel que habrá de decidir si la muestra de entrada es positiva o negativa, si en este primer nivel la muestra se considera positiva, pasará a la segunda etapa, en la que se volverá a trabajar en clasificar la muestra; si nuevamente es positiva pasará a la etapa tres y así sucesivamente mientras la muestra sea considerada positiva, mientras que si en alguna de las etapas ésta es clasificada como negativa, entonces será rechazada inmediatamente, sin que se le dediquen mayores recursos de procesamiento, lo que reduce considerablemente el tiempo de clasificación (Viola y Jones, 2001; Ruiz Sancho, 2014).

4. METODOLOGÍA

Para cumplir con el objetivo de este proyecto, se propone que la metodología considere las siguientes etapas:

- Captura de vídeo.
- Localización de rostro.
- Identificación de características relevantes en el rostro.
- Detección de características visuales que impliquen fatiga o distracción.
- Procesamiento de los hallazgos.

La metodología propuesta se representa de forma gráfica en la Figura 4.1, es importante destacar que las etapas propuestas para la implementación del proyecto son consideradas como lineamientos generales del proceso a seguir, será en las secciones siguientes en las que se describirá con mayor profundidad el proceso de ejecución de cada una de las mencionadas etapas.

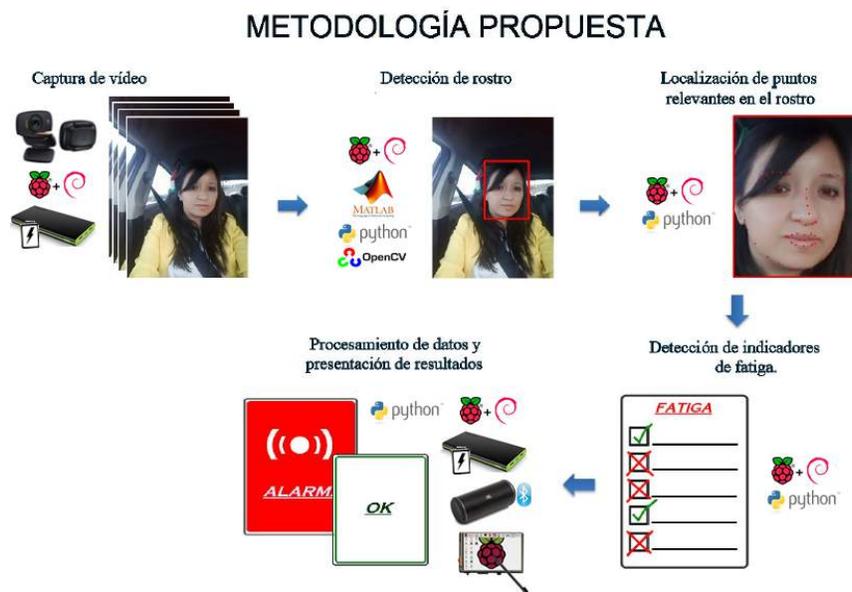


Figura 4.1: Metodología general propuesta.

4.1. Captura de Vídeo

Para llevar a cabo la primera fase de la metodología propuesta se realizaron pruebas con cámaras de vídeo de 3 MP, 5 MP y 8 MP de resolución, tras las pruebas se determinó que una cámara de 3 MP proporcionaba una imagen poco clara, la cual, dadas las necesidades del proyecto resultaba poco útil, por ello se decidió utilizar una cámara de al menos 5MP de resolución.

Para que la cámara resulte útil al proyecto, además de proporcionar una imagen clara, es indispensable que ésta se comunique con el procesador mediante un cable USB, es importante que pueda fijarse fácilmente, además, es recomendable que el ángulo de enfoque sea fácilmente modificable. Por todo lo anterior se seleccionó la cámara Logitech c525, mostrada en la Figura 4.2, la cual posee las características listadas en la Tabla 4.1.

Características de la Webcam	
Atributo	Valor / Descriptor
Resolución de vídeo	HD 720 pixeles
Calidad de fotografías	8 MP
Calidad de vídeo	Muy buena (en comparación con 2 modelos más, cuya calidad de vídeo se clasifica como buena y óptima)
Tecnología de objetivo	Estándar
Tipo de enfoque	Enfoque automático a 7cm o más
Micrófono incluido	Uno
Corrección de iluminación automática	De gama alta
Rotación completa de 360°	Sí
Cable de conexión	Conector USB 2.0, con longitud de 1.80 m

Tabla 4.1: Características indicadas por el proveedor para la cámara de vídeo seleccionada; Logitech c525.



Figura 4.2: Imagen ilustrativa de la cámara de vídeo utilizada.

Una vez que se ha seleccionado la cámara de vídeo a utilizar, es importante conocer los requisitos que, de acuerdo con el proveedor, se deben cumplir por la unidad de procesamiento para que su funcionamiento sea óptimo. Tales requerimientos se enlistan a continuación.

- Procesador de, al menos, 1 GHz
- 512 MB o más de memoria RAM
- 200 MB de espacio en disco
- puerto de conexión USB 1.1 (2.0 preferentemente)

Dada la lista anterior se realizó la búsqueda de un ordenador de placa reducida, de bajo costo, que cubriera dichos requerimientos. Para definir el ordenador a utilizar se realizó una comparación entre aquellos que mejor pudieran adaptarse a las necesidades del sistema, incluyendo portabilidad, la Tabla 4.3 muestra el comparativo.

Tabla comparativa de de CPUs, Parte I			
	Raspberry Pi 3 B	Raspberry Pi 2 B	Raspberry Pi B+
<i>Procesador</i>	1.2GHz 64-bit quad-core ARMv8	900MHz quad-core ARM Cortex-A7	ARM11 ARMv6 700 MHz
<i>GPU</i>	VideoCore IV	VideoCore IV	VideoCore IV
<i>Almacenamiento</i>	Tarjeta micro SD	Tarjeta micro SD	Tarjeta micro SD
<i>Memoria RAM</i>	1 GB	1 GB	520 MB

<i>Entrada de vídeo</i>	Conector MPI CSI-2 de 15 vías para cámara de vídeo HD Raspberry Pi	Conector MPI CSI-2 de 15 vías para cámara de vídeo HD Raspberry Pi	Conector MPI CSI-2 de 15 vías para cámara de vídeo HD Raspberry Pi
<i>Salida de vídeo</i>	Full HDMI / Salida de vídeo compuesto (PAL/NTSC) tipo jack de 3,5 mm 4 polos / Conector de interfaz serie de display de 15 vías	HDMI 1.4 a 1920x1200 píxeles / Salida de vídeo compuesto (PAL/NTSC)	HDMI 1.4 a 1920x1200 píxeles
<i>Salida de audio</i>	Audio digital por salida HDMI / Salida de audio estéreo compartida con la salida de vídeo compuesto	Audio digital por salida HDMI / Salida de audio estéreo compartida con la salida de vídeo compuesto	Audio digital por salida HDMI / Salida de audio estéreo compartida con la salida de vídeo compuesto
<i>Conectores USB</i>	4 x 2.0	4 x 2.0	4 x 2.0
<i>Alimentación</i>	+5V a 2.5A a través de conector hembra microUSB	+5v, 2A. A través de conector hembra micro USB	+5V a 2A a través de conector hembra microUSB

Tabla comparativa de de CPUs, Parte II

	Banana Pro	Banana Pi M2	ODROID-C2
<i>Procesador</i>	Allwinner A20 ARM Cortex-A7 dual core a 1GHz	Allwinner A31S ARM Cortex-A7 quad core a 1GHz	Amlogic quad core ARM Cortex-A53 1.5GHz
<i>GPU</i>	ARM Mali400 MP2	PowerVR SGX544MP2	Dual core Mali-450MP
<i>Almacenamiento</i>	Tarjeta micro SD	Tarjeta micro SD	Dual core Mali-450MP

<i>Memoria RAM</i>	1 GB	1 GB	2 GB
<i>Entrada de vídeo</i>	Tipo CSI para conectar módulos de cámara hardware	Tipo CSI para conectar módulos de cámara hardware	No incluye
<i>Salida de vídeo</i>	HDMI / Salida de vídeo compuesto (PAL/NTSC) tipo jack de 3,5 mm 4 polos / Conector LVDS/RGB	HDMI / Conector LVDS/RGB	Full HDMI
<i>Salida de audio</i>	Audio digital por salida HDMI / Salida de audio estéreo compartida con la salida de vídeo compuesto	Audio digital por salida HDMI / Tipo jack 3.5 mm	Audio digital por salida HDMI
<i>Conectores USB</i>	2 x 2.0	4 x 2.0	4 x 2.0
<i>Alimentación</i>	+5V a 2A a través de conector hembra microUSB	+5V a 2A a través de Power Jack	+5V a 2A

Tabla 4.3: Comparativo de CPUs portables disponibles en el mercado.

De acuerdo con la relación entre los requerimientos de procesamiento de la cámara, y los datos mostrados en la Tabla 4.3 se decidió que la mejor opción es utilizar la Raspberry Pi 3B, sin embargo se ha decidido que durante las pruebas posteriores se realizarán pruebas comparativas con diversos fines en 3 CPUs diferentes; Una computadora portátil con 12GB de memoria, procesador AMD A10 2.00GHz y una cámara de 5MP, la Raspberry Pi 3B y una Raspberry Pi 2B. Las pruebas realizadas y sus resultados se detallan en secciones posteriores.

4.2. Detección de rostro

4.2.1 Pruebas preliminares

Debido al amplio campo de aplicación de la detección de rostro, se ha convertido en un campo ampliamente estudiado, por lo que no es de sorprender la existencia de detectores pre-entrenados, basados en lenguajes tales como MATLAB y la librería de procesamiento de imágenes llamada OpenCV. Por esta razón, se tomó la decisión de realizar pruebas preliminares con uno de tales detectores; el detector de rostros frontales de Python. El detector de rostros, seleccionado para llevar a cabo las pruebas, se basa en el trabajo de Viola y Jones (2001) y fue entrenado para detectar rostros frontales (Mordvintsev y K, 2013). De acuerdo con las pruebas realizadas se pudo establecer que, tras configurar adecuadamente sus parámetros, dicho detector tiene un muy bajo porcentaje de errores de clasificación, sin embargo, el cuadro de detección que delimita la zona que forma parte del rostro encontrado no es consistente, mientras que, como parte de dichas inconsistencias, no siempre se extrae el rostro hallado desde la parte superior de la frente hasta justo debajo de la barbilla, es decir, se excluían segmentos del rostro.

Algunas muestras de los resultados de tales pruebas preliminares se muestran en la Figura 4.3. Cabe destacar que tales pruebas se llevaron a cabo sin un pre-procesamiento riguroso de las imágenes, dado que se consideraron únicamente como punto de partida para la primera etapa de la metodología propuesta, de tal manera que se tuvieran claros los factores a mejorar en cuanto a la detección de rostro.



Figura 4.3: Muestras de los resultados obtenidos de las pruebas preliminares, desarrolladas con la aplicación del detector de rostro frontal de OpenCV.

4.2.2 Descriptores HOG en detección de rostro

De acuerdo con Dalal y Triggs (2005) el primer paso para la localización de rostro es definir un conjunto robusto de características que permitan discernir si dentro de una imagen hay un rostro o no, aún si el fondo de la imagen es desordenado o si esta fue tomada en un entorno con mala iluminación, a tales conjuntos se les denomina descriptores.

En su trabajo, (Dalal y Triggs, 2005) concluye que usando descriptores locales normalizados basados en el método de (HOG) se podría realizar la detección con un excelente resultado en comparación con cualquier descriptor desarrollado, incluyendo Wavelet. En el mismo trabajo se indica que aún cuando los descriptores propuestos se procesan sobre una red densa de celdas uniformemente espaciadas y usan la superposición de la normalización del contraste local como un esfuerzo para mejorar su rendimiento, lo que los diferencia de algunos otros como los SIFT.

Un método de detección de rostro basado en los descriptores HOG, como el propuesto por Dalal y Triggs (2005) puede resumirse en el diagrama de la Figura 4.4.

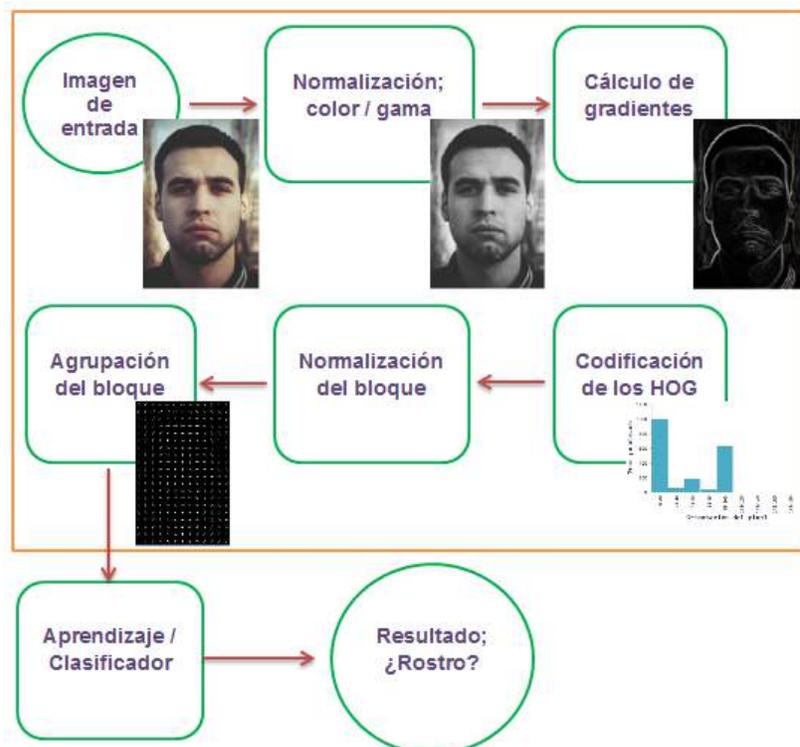


Figura 4.4: Resumen gráfico de la detección de rostros basada en un descriptor HOG.

Enseguida se describen brevemente las etapas que integran al algoritmo de aplicación del descriptor HOG, ilustradas en 4.4, con base en Dalal (2006); Del Abril De Mur (2017); Universitatea Tehnica din Cluj-Napoca (N.d.):

1. **Normalización color/gamma:** También llamada pre-procesamiento. En este punto, se ecualiza la gamma o el color de la imagen como medida para reducir la influencia de la iluminación (Dalal, 2006). En algunas fuentes se sugiere que este es un paso opcional (Del Abril De Mur, 2017) o bien, que la normalización más útil es la aplicada al tamaño de la imagen (Mallick, 2017), esto debido a que la normalización de gamma o color no representa una mejora significativa en los resultados. En la Figura 4.5 se muestra un ejemplo de normalización.

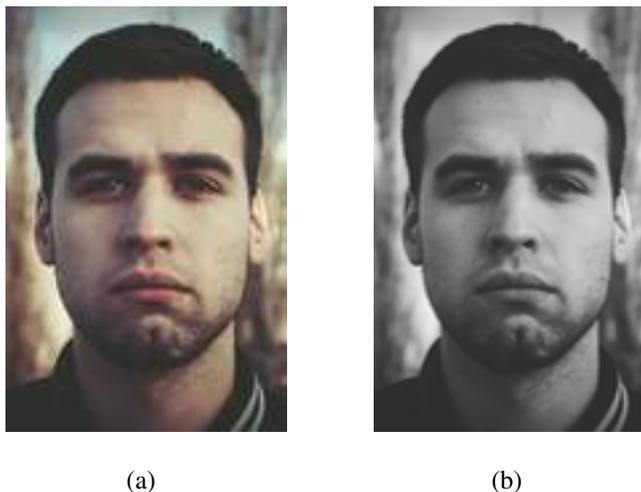


Figura 4.5: Ejemplo de pre-procesamiento de la imagen; etapa inicial del proceso para la aplicación de los descriptores HOG. En (a) la imagen original o de entrada, en (b) la imagen normalizada en escala de grises.

2. **Cálculo de gradientes:** El método más común para el cálculo de los gradientes es aplicar una máscara, derivada discreta 1D de punto centrado, tanto vertical como horizontalmente (Universitatea Tehnica din Cluj-Napoca, N.d.), filtrando así la imagen. Las máscaras a aplicar, en este caso, se ilustran en la Figura 4.6, que de acuerdo con Dalal (2006) son las que aportan un mejor desempeño al proceso. Con base en lo anterior, tenemos que; dada una imagen (I) se pueden obtener las derivadas espaciales en ambos ejes de la misma

mediante la aplicación de convolución, tal como se muestra en las ecuaciones (4.3).

En este punto se han de calcular tanto la magnitud (ecuación (4.4)) como la dirección (ecuación (4.5)) del gradiente para cada uno de los píxeles de la ROI, mediante los siguientes modelos matemáticos (Morán Cruz, 2013):

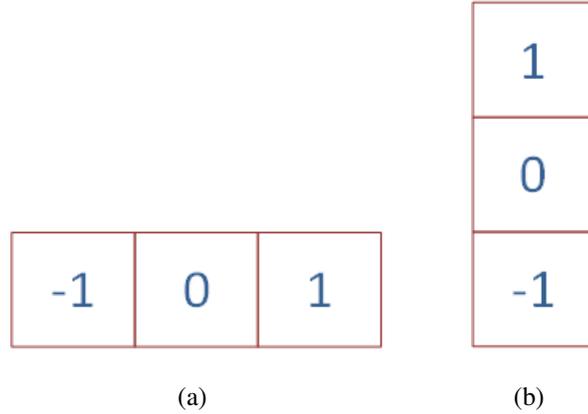


Figura 4.6: Representación gráfica de las máscaras aplicables a la imagen para la obtención de sus derivadas espaciales a lo largo de sus ejes x (a) y y (b).

$$D_x = \{-1 \ 0 \ 1\} \quad (4.1)$$

$$D_y = \begin{Bmatrix} 1 \\ 0 \\ -1 \end{Bmatrix} \quad (4.2)$$

$$I_x = I * D_x \quad I_y = I * D_y \quad (4.3)$$

$$|G| = \sqrt{I_x^2 + I_y^2} \quad (4.4)$$

$$\theta = \arctan \frac{I_x}{I_y} \quad (4.5)$$

Si consideramos que la imagen inicial I es la mostrada en la Figura 4.5(a), tras aplicar convolución para obtener I_x y I_y , se tiene como resultado las imágenes correspondientes

a la Figura 4.7(a) y 4.7(b) respectivamente. Mientras que la imagen, tras el cálculo de $|G|$, queda tal como se observa en la Figura 4.8.

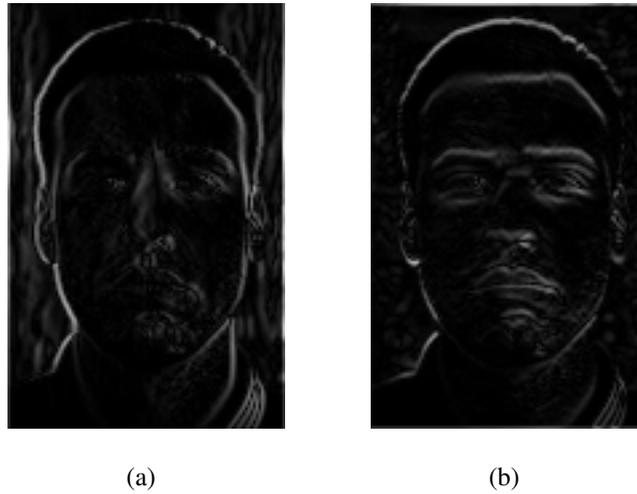


Figura 4.7: Imagen (I) tras la obtención de las derivadas espaciales a lo largo de sus ejes x (a) y y (b).



Figura 4.8: Magnitud del Gradiente ($|G|$) de I HOG.

3. **Codificación de los histogramas locales de orientación:** En este paso, la representación gráfica de $|G|$ (Figura 4.8) se divide en pequeñas regiones espaciales rectangulares denominadas celdas, las cuales se componen de una cantidad definida de píxeles adyacentes (ver Figura 4.9). Las celdas de 64 píxeles (la agrupación de $[8 \times 8]$ píxeles) son las que mejores resultados ofrecen (Dalal y Triggs, 2005).

A continuación se genera, para cada celda, un histograma basado en la orientación de los píxeles; dentro de la celda, cada píxel aporta un valor ponderado para un canal del histo-

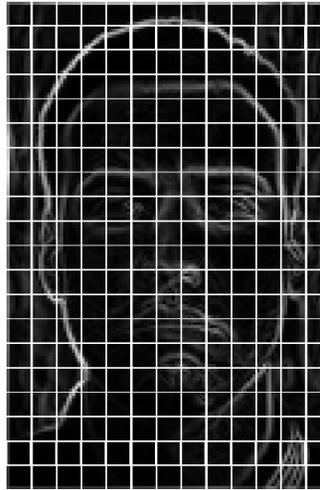


Figura 4.9: Ejemplo de división de 4.8 en celdas.

grama de acuerdo con los valores obtenidos durante el cálculo del gradiente. Los canales del histograma pueden definirse dentro de los intervalos $[0^\circ - 360^\circ]$ o $[0^\circ - 180^\circ]$ dependiendo de si se considera su signo o no. Es importante aclarar que el valor de θ , obtenido tras aplicar la Ec. 4.5 se expresa en radianes, por lo que habrá de usarse la Ec. 4.6 para convertirlo a grados. Estos histogramas son la representación básica del descriptor HOG. Es aquí donde se encuentra la no-linealidad del descriptor, pues aún cuando la codificación es sensible al contenido de la imagen, tolera leves cambios de pose o apariencia.

$$\alpha = \theta \cdot \left(\frac{180^\circ}{\pi} \right) \quad (4.6)$$

Para la obtención de los histogramas, Dalal (2006) sugiere el empleo de gradientes sin signo, es decir en el intervalo de $[0^\circ - 180^\circ]$, en conjunto con histogramas de 9 canales para un mejor desempeño del descriptor. El peso ponderado, asignado a cada pixel contenido en las celdas, puede ser igual a la magnitud del gradiente, la raíz cuadrada de este o bien su elevación al cuadrado.

En la Figura 4.10(a) se muestran, a modo de ejemplo, los valores de orientación de los pixeles que conforman una celda de $[8 \times 8]$ pixeles, mientras que en la Figura 4.10(b) se ejemplifican los valores de la magnitud de gradiente para cada uno de los pixeles de la misma celda; si se quisiera generar el histograma de nueve canales para la celda mostrada, considerando la orientación de los pixeles y tomando la magnitud del gradiente como

el peso ponderado aportado por cada pixel, se obtendría un histograma como el que se muestra en la Figura 4.11.

45	86	84	87	87	81	82	88
3	16	5	20	43	47	54	77
1	0	8	8	18	34	59	45
2	0	4	0	3	19	63	0
2	18	20	10	0	5	63	27
2	18	9	9	6	13	54	35
6	68	11	11	9	19	70	34
3	18	0	4	0	20	72	37

109	78	82	86	87	92	102	105
75	7.3	11	12	16	23	19	9.2
73	10	14	15	19	22	12	4.2
74	10	13	15	22	21	4.5	7
76	6.3	15	17	25	21	2.2	11
76	3.2	12	20	27	18	8.6	12
78	5.4	10	16	24	18	12	7.2
84	3.2	4	14	24	12	9.5	5

(a)
(b)

Figura 4.10: Ejemplo de los valores utilizados para la creación del histograma de una celda; (a) orientación de los pixeles de la celda, (b) magnitud del gradiente para cada pixel de la celda.

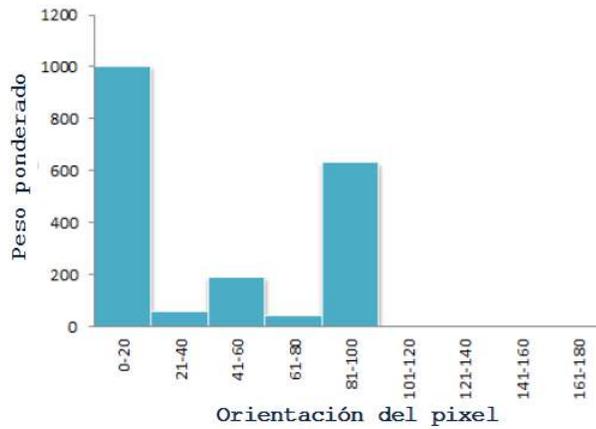


Figura 4.11: Ejemplo del histograma de una celda, basado en los valores mostrados en la Figura 4.10.

Tras finalizar el proceso de obtención de valores para cada celda mostrada en la Figura 4.9 se obtuvieron los histogramas presentados en la Figura 4.12.

4. **Normalización del bloque:** Para reducir los problemas debidos a las diferencias de iluminación y contraste, los gradientes habrán de normalizarse; para ello las celdas se agruparan en grandes bloques espacialmente conectados. Estos bloques, generalmente se traslapan de tal manera que cada celda contribuya en más de una vez en el descriptor final, lo que

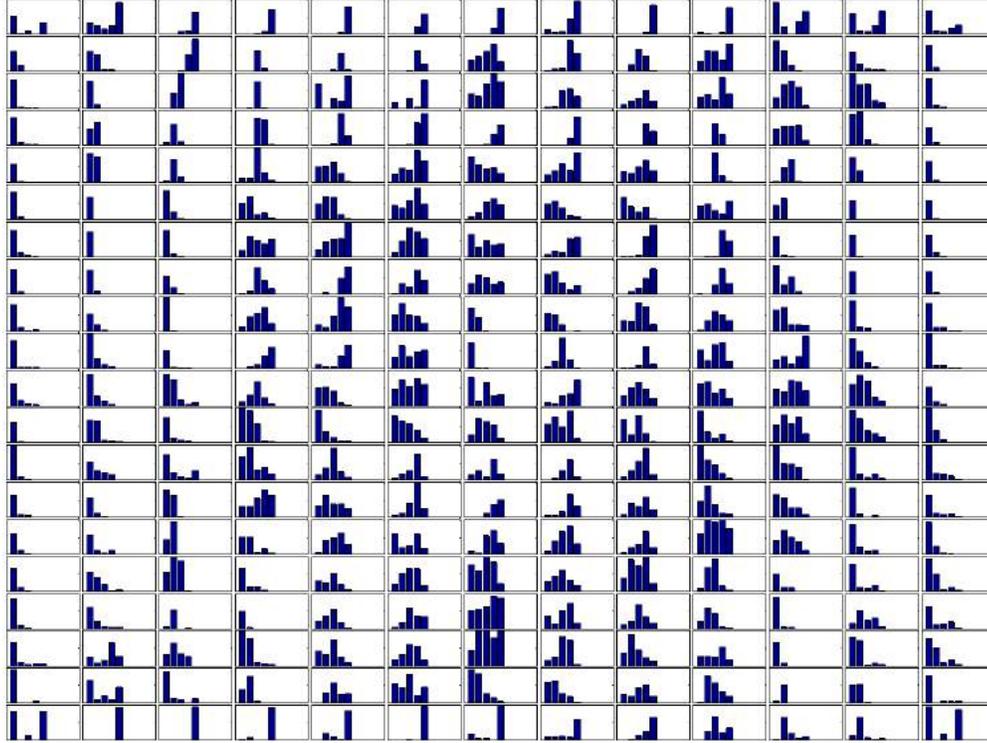


Figura 4.12: Histogramas para la imagen inicial (I), Figura 4.5(a), basados en la orientación de los pixeles agrupados en celdas.

mejorará el rendimiento del sistema. En el trabajo de Dalal (2006) se presentan tres diferentes métodos de normalización (Ecuaciones (4.7),(4.8) y (4.9)), de los cuales puede seleccionarse uno.

$$L2 - norm : f = \frac{v}{\sqrt{\|v\|_2^2 + e^2}} \quad (4.7)$$

$$L1 - norm : f = \frac{v}{\|v\|_1 + e} \quad (4.8)$$

$$L1 - sqrt : f = \sqrt{\frac{v}{\|v\|_1 + e}} \quad (4.9)$$

Dentro de las ecuaciones propuestas por Dalal (2006), Ecuaciones (4.7),(4.8) y (4.9), v describe al vector no normalizado que contiene a todos los histogramas que conforman un bloque, mientras que $\|v_k\|$ es su el vector k -normalizado de los histogramas por bloque, con $k = 1,2$. Y, finalmente, e es una constante cuyo valor es muy cercano a cero.

Para un mejor resultado, Dalal sugiere generar bloques de cuatro celdas, es decir un arreglo de [2x2] celdas, estando estas conformadas por 64 pixeles.

5. **Agrupación de los descriptores:** Finalmente se concatenan los descriptores de los bloques creados en el paso anterior. De esta manera, el descriptor final poseerá una descripción completa de la ventana de detección. La información obtenida hasta aquí con el descriptor, podrá entonces ser empleada en un clasificador. Siguiendo con el ejemplo donde I es la mostrada en la Figura 4.5(a), los descriptores HOG quedarán tal como se muestran en la Figura 4.13(a)

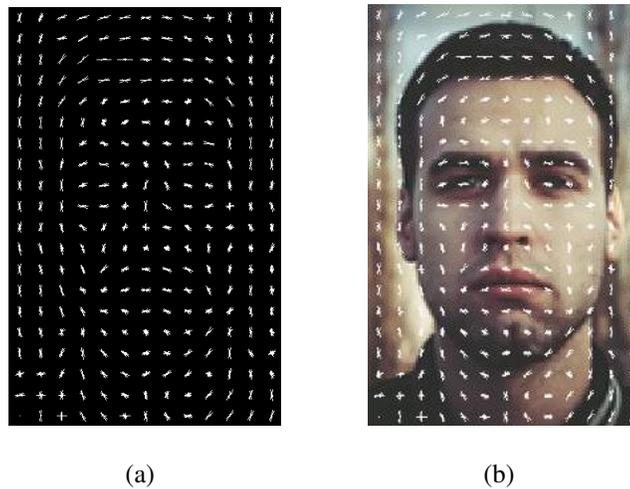


Figura 4.13: Resultado final; (a) descriptores HOG utilizando la Figura 4.5(a) como imagen de entrada (I), (b) descriptores HOG sobre la imagen original (I).

4.2.3 Clasificadores en cascada

Método propuesto por Viola y Jones (2001), el cuál, es descrito como una combinación sucesiva de clasificadores en una estructura en forma de cascada en la que la dificultad de los clasificadores incrementa etapa tras etapa.

De acuerdo con (Soo, 2014), el objeto buscado, en este caso el rostro, se localiza tras utilizar una sub-ventana que recorrerá la imagen de entrada (I), determinando si esta es positiva o negativa. Será considerada como positiva si, de acuerdo con los clasificadores, forma parte del objeto de interés, en ese caso la sub-ventana será evaluada por clasificadores cada vez más

complejos, si llega hasta la última etapa entonces se entenderá que objeto ha sido localizado dentro de la imagen; mientras que será declarada negativa si en cualquiera de las etapas se determina que dicha sub-ventana no forma parte del objeto buscado. En la Figura 4.14 se muestra un diagrama del funcionamiento de dichos clasificadores.

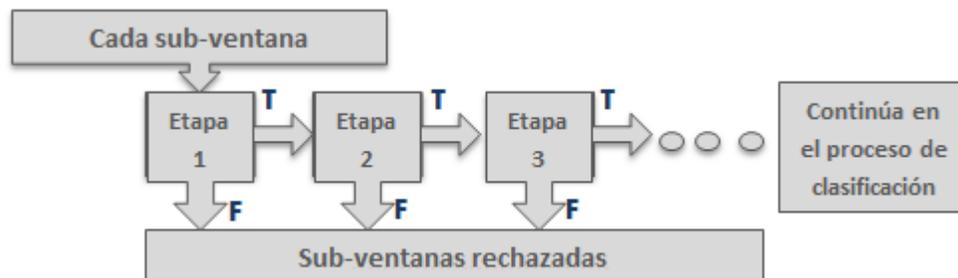


Figura 4.14: Representación gráfica del sistema de clasificación por cascada.

El uso de los clasificadores en cascada incrementa considerablemente la velocidad de respuesta del clasificador pues centra su atención en las regiones de la imagen que resultan prometedoras para la localización del rostro, lo que desecha rápidamente las regiones que conforman el fondo de la imagen de entrada (Viola y Jones, 2001).

4.2.4 Entrenamiento del detector de rostro

Tras los resultados obtenidos durante las pruebas preliminares, se decidió entrenar un nuevo detector, con lo que se esperaba mejorar la capacidad de detección en rostros con ángulos mayores de inclinación y giro (ver Figura 4.15), es decir, reducir las restricciones de posición en el rostro para que este pueda ser detectado es un factor importante en el desarrollo del proyecto, debido a la aplicación en la que se enfoca, la asistencia en la conducción. También se espera que el recuadro de detección del nuevo detector sea consistente; seleccionando los rostros desde la parte superior de la frente hasta el límite inferior de la barbilla.

Selección y pre-procesamiento de imágenes de entrenamiento

- **Imágenes positivas:**

Se le da el nombre de imágenes positivas a aquellas que conforman el grupo de imágenes de entrenamiento que contienen el objeto que se desea detectar, en este caso el rostro.

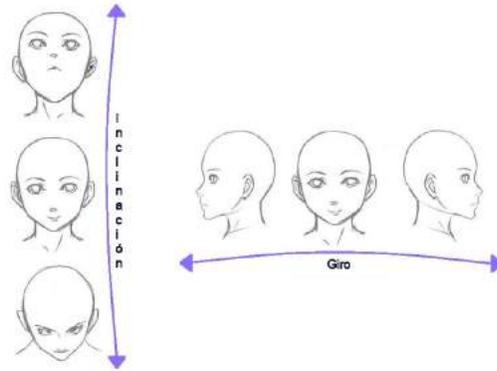


Figura 4.15: Grados de libertad considerados para la definición de posición del rostro.

Existen un gran número de bases de datos con imágenes faciales (Grgic y Delac, 2017), el seleccionar unas u otras depende del uso que se les quiera dar. Para aumentar la robustez del detector propuesto para este proyecto se decidió usar un alto número de imágenes positivas, por lo que fue necesario emplear más de una base de datos, para la selección de las mismas se consideraron las siguientes condiciones;

- Los rostros en las imágenes deberían ser de tamaño y resolución muy similar.
- Las imágenes deberían presentar diferencias de iluminación y fondo.
- Las imágenes debían mostrar variaciones de posición en el rostro contenido. Considerando el propósito para el que es creado tal detector, se establece, además, que las posiciones de los rostros en las imágenes deben permitir la localización de ambos ojos, nariz, boca y frente.
- Se emplearon únicamente bases de datos con imágenes de rostros reales, descartando aquellas que se conforman por imágenes de rostros creados con animación, tanto 2D como 3D.

Tras atender los puntos mencionados, se eligió un conjunto de poco más de 19 000 imágenes para fungir como muestras positivas para el entrenamiento. Dichas imágenes fueron tomadas de: *The ORL Database of Faces* (AT&T Laboratories Cambridge, 2002), *JAFFE Database* (Lyons *et al.*, 1998), *FEI Face Database* Thomaz y Giraldi (2010), Samaria y Harter (1994) y Bainbridge *et al.* (2013).

Tras seleccionar las imágenes, estas fueron pre-procesadas;

- Las imágenes fueron recortadas de tal manera que se retirara el fondo innecesario,

dejando únicamente el área de interés (la que contiene al rostro). En este paso es importante que el área de interés recortada sea consistente con los requisitos de detección establecidos, de tal manera que los rostros en las imágenes seleccionadas luzcan exactamente como se espera que sean detectados los rostros una vez finalizado el detector.

- Se estandarizó el tamaño y color de las imágenes, todas ellas quedaron de (576 x 768) pixeles y convertidas a escala de grises. Una muestra de tales imágenes se presenta en la Figura 4.16



Figura 4.16: Muestra del conjunto positivo de entrenamiento, estandarizado en tamaño y color.

- Las imágenes resultantes se guardan en un directorio específico, y sus nombres se listan en un archivo *.txt, que será necesario en el proceso de entrenamiento, en este archivo se anotan también las coordenadas de localización del rostro en la imagen, así como el tamaño del rostro
- **Imágenes negativas:**
Se conocen como imágenes negativas de entrenamiento aquellas que carecen del objeto buscado, que en este caso son rostros. Para la selección de imágenes se buscó que mostraran escenas en las que es común encontrar rostros, tales como oficinas e interiores de autos, por mencionar algunos. Además se incluyeron imágenes que mostraran componentes de rostros que por sí mismos no pudieran ser considerados como un rostro en sí, como ojos, nariz y labios; con ello se buscó reducir los falsos positivos en los que se localizaran solo pequeñas partes del rostro, pues dada la aplicación a la que se enfoca este proyecto es importante que se realice una apropiada detección de rostros de tal manera que en ellos

sean visibles ambos ojos, la frente y la boca. Es importante que la cantidad de muestras negativas sea considerablemente superior a la de imágenes de entrenamiento positivas, por tal motivo se seleccionaron alrededor de 66 500 imágenes para este conjunto, estas fueron tomadas de páginas de Internet de acceso público en su mayoría, aunque se incluyen también algunas imágenes tomadas de los repositorios *online* de la Universidad de Oxford (Department of Engineering Science, 2004), dichas imágenes fueron utilizadas únicamente para la fase de entrenamiento. Las imágenes seleccionadas se estandarizaron en color y, aún cuando no se estandarizaron en tamaño, se buscó que este fuera cerca de 4 veces mayor al de las muestras positivas. En la Figura 4.17 se puede ver una muestra de las imágenes negativas colectadas y pre-procesadas para el entrenamiento del detector.



Figura 4.17: Muestra del set de imágenes negativas de entrenamiento, estandarizado en color.

Este conjunto de imágenes se almacenó en un directorio dedicado, en el que se guardó también un registro de los nombres de cada una de las imágenes que conforman este grupo, lo cual será necesario durante el proceso de entrenamiento.

Entrenamiento del clasificador

Una vez que las imágenes de entrenamiento, tanto positivas como negativas, se encontraron listas era factible el entrenamiento del clasificador.

El clasificador fue entrenado mediante la función *trainCascadeObjectDetector()* de MATLAB (Ver Código 5, en los anexos), la cual soporta la identificación de objetos con ayuda de tres diferentes descriptores; Haar, LBP y HOG (The MathWorks Inc., 2017).

De entre dichos descriptores se decidió utilizar los HOG, esto considerando la información proporcionada por Dalal y Triggs (2005) respecto al desempeño efectivo y eficiente de dicho descriptor. Al configurar la función para utilizar tales descriptores esta trabajará con los parámetros con los que Dalal (2006) obtuvo los mejores resultados, es decir, bloques de $[2 \times 2]$ celdas, las cuales se conforman de $[8 \times 8]$ píxeles.

Por otra parte, la clasificación se realiza mediante la aplicación de clasificadores cascada; una ventana recorre la imagen y mediante la cascada de clasificadores se define si dicha ventana contiene o no el objeto de interés, en este caso rostros (The MathWorks Inc., 2017). En la Figura 4.18 se ilustra el funcionamiento de la función utilizada.

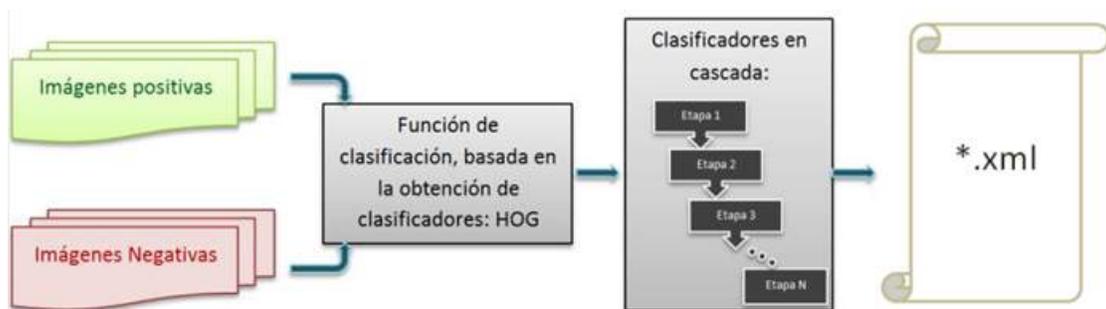


Figura 4.18: Ilustración del principio básico de funcionamiento de la función de MATLAB *trainCascadeObjectDetector()*.

La función utilizada (*trainCascadeObjectDetector()*) necesita, para su funcionamiento, definir algunos parámetros, los cuales incluyen; información de las imágenes positivas y negativas, tasa de falsa alarma y verdaderos positivos. De acuerdo con la página dedicada a dicha función en el sitio web de MathWorks (The MathWorks Inc., 2017), al definir tales parámetros se debe considerar lo siguiente:

- Al utilizar un conjunto grande de imágenes de entrenamiento, como en este caso, se deberá incrementar la cantidad de etapas de entrenamiento y establecer una tasa de falso positivo más alta para cada etapa.
- Si se emplea un conjunto pequeño de imágenes de entrenamiento, se debería reducir el número de etapas de entrenamiento de la cascada, y establecer una tasa baja de falso positivo para cada etapa.

- Incrementar la tasa de falso positivo permitirá reducir la posibilidad de pasar por alto el objeto de interés en la imagen de entrada, sin embargo, es importante considerar que elevar desmedidamente dicha tasa podría ocasionar un elevado número de detecciones falsas.
- Para reducir el número de detecciones falsas se propone elevar el número de etapas de detección o reducir el índice de falsa alarma por etapa.

Tras la obtención del primer detector entrenado se realizaron algunas pruebas de detección utilizando imágenes tomadas con una cámara de 8 MP. Si bien estas pruebas no se realizaron con metodologías estrictas se trataba de probar la funcionalidad básica del detector y eran realizadas con la finalidad de detectar las fallas más notorias del mismo, a partir de ellas se ajustaron los parámetros de entrenamiento, entre ellos; la cantidad de etapas de la cascada, la cantidad de imágenes de entrenamiento, el tamaño de las mismas, entre otros.

Al finalizar, se contaba con 11 detectores obtenidos mediante entrenamientos ejecutados bajo parámetros diferentes. Se esperaba que el último de ellos fuera el mejor de todos, para garantizarlo fue necesario realizar pruebas de parámetros controlados, tales pruebas se detallan a continuación.

4.2.5 Experimentación y discusión de resultados

Para las pruebas formales de detección de rostro se utilizó la base de datos citada por Gourier *et al.* (2004), la cual contiene 2 511 imágenes de rostros en 93 posiciones diferentes, la posición en la que se halla cada rostro en su respectiva imagen se encontraba en términos de los grados de libertad definidos en la Figura 4.15.

Dadas las necesidades del proyecto para la detección de fatiga, se eligieron, de entre las 93 posiciones faciales, solamente 14, las cuales permiten que ambos ojos, nariz y boca sean fácilmente visibles. Dichas posiciones se enlistan a continuación, utilizando la notación de posición en grados basada en los parámetros establecidos por Gourier *et al.* (2004), tal como se muestra en la Figura 4.19:

- $-15^\circ / -15^\circ$ ■ $-15^\circ / 0^\circ$ ■ $-15^\circ / +15^\circ$ ■ $0^\circ / -15^\circ$

- $0^\circ / 0^\circ$ ■ $+15^\circ / -15^\circ$ ■ $+15^\circ / +30^\circ$ ■ $+30^\circ / +15^\circ$
- $0^\circ / +15^\circ$ ■ $+15^\circ / 0^\circ$ ■ $+30^\circ / -15^\circ$
- $0^\circ / +30^\circ$ ■ $+15^\circ / +15^\circ$ ■ $+30^\circ / 0^\circ$

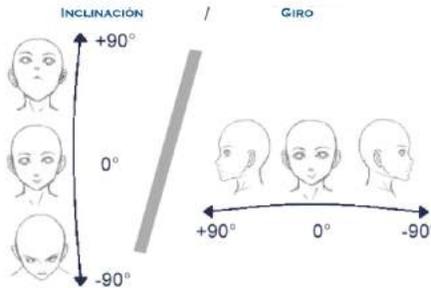


Figura 4.19: Nomenclatura utilizada para expresar la posición facial en términos de grados; signos positivos y negativos ilustrados, basada en la información proporcionada por Gourier *et al.* (2004).

En la Figura 4.19 se puede ver que si la inclinación del rostro se encuentra en términos negativos indica que el sujeto de prueba se encuentra con el rostro inclinado hacia enfrente, siendo -90° el indicador de que la mirada del sujeto es perpendicular al piso; mientras que si el signo es positivo, la inclinación es hacia atrás, $+90^\circ$ de inclinación indicarán que el sujeto mira directamente al techo inclinando lo más posible su cabeza hacia atrás. Por otra parte, un giro negativo hace referencia a un rostro girado a la izquierda, siendo -90° el indicador de que el sujeto ha girado su rostro lo más posible hacia la izquierda; mientras que un giro con signo positivo indicaría que el sujeto de prueba ha girado el rostro hacia su derecha, de tal forma que $+90^\circ$ indicaran que el sujeto ha girado la cabeza a la derecha hasta casi alinear la posición de su nariz con la posición de su hombro derecho.

Tras descartar las imágenes que se encontraban en posiciones faciales diferentes a las 14 listadas previamente, únicamente quedaron 378 imágenes, con las cuales se realizaron las pruebas. En ellas, los detectores entrenados fueron comparados entre sí y contra el detector pre-entrenado de OpenCV. En la comparación se consideraron aspectos como el porcentaje de detección de verdaderos positivos, falsos positivos y la adecuada localización de los rostros.

Si se tratase de detección más que de localizar el rostro y determinar su tamaño apropiadamente, el detector pre-entrenado de OpenCV resultaría ser la mejor opción, tal como se

muestra en la gráfica de la Figura 4.20. Sin embargo, como se mencionó en la subsección 4.2.1 Preliminares, el detector de OpenCV no es consistente en la localización de los rostros, pues no es muy preciso al momento de determinar las coordenadas de su ubicación ni el tamaño del mismo o por lo menos no siempre considerando el rostro desde la frente hasta la parte inferior de la barbilla, lo que lo hace ineficiente para este proyecto.

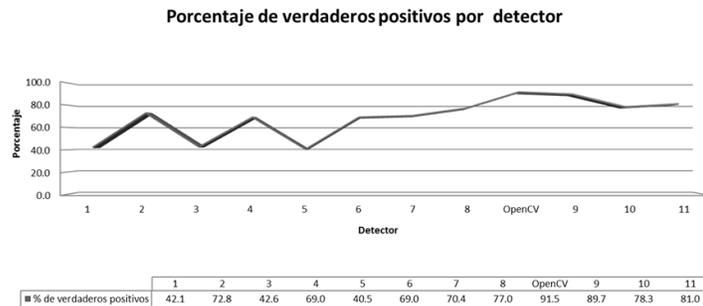


Figura 4.20: Comparativo de la cantidad de detecciones hechas por cada detector, sin importar la precisión de la localización, solo se considera la apropiada clasificación de la imagen.

Como resultado de las pruebas se determinó que podía mantener constantes sus resultados únicamente en cinco de las posiciones seleccionadas para realizar las pruebas, ello se refleja en un porcentaje de 78 % de detección imprecisa en las imágenes seleccionadas. En la Figura 4.21 se puede observar el porcentaje de eficiencia relativa para cada uno de los detectores entrenados y el detector de OpenCV, para ello se consideran únicamente las detecciones realizadas, clasificándolas en detecciones imprecisas, erróneas y acertadas (ver Figura 4.22).

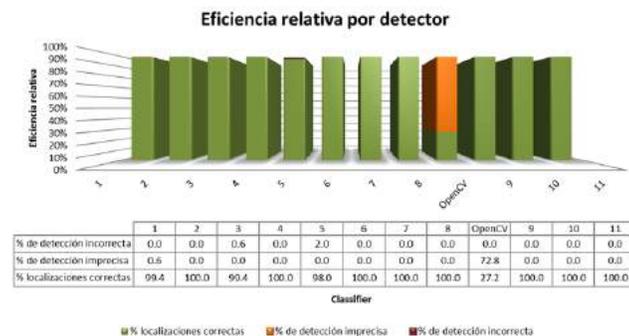


Figura 4.21: Gráfico de la eficiencia de detección para cada detector entrenado y el pre-entrenado de OpenCV; considerando el porcentaje de detecciones imprecisas, erróneas y acertadas (ver Figura 4.22).

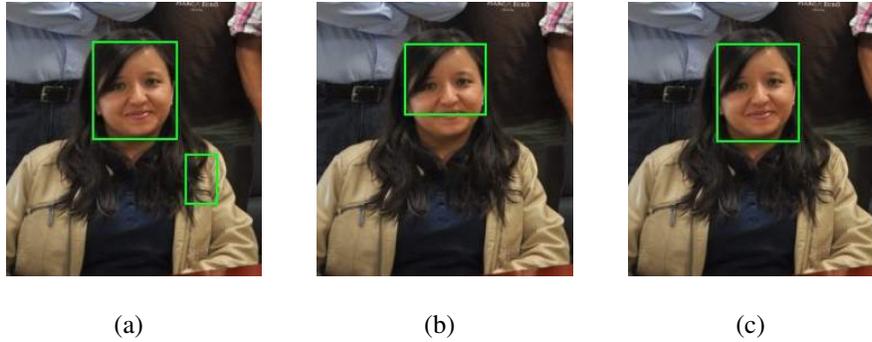


Figura 4.22: Muestras de la clasificación que se da, en este proyecto, a la forma en que responden los detectores, de acuerdo con los requisitos establecidos previamente, ver subsección 4.2.1: (a) Ejemplo de detección incorrecta, (b) imagen ilustrativa de localización imprecisa, (c) muestra de detección correcta.

Después de comparar la eficiencia relativa de cada detector, se decidió combinar los tres con mejor desempeño de tal manera que se dotara de robustez a la etapa de detección de rostro pues es una parte esencial para el cumplimiento del objetivo general de este proyecto. Una vez que se implementó el detector combinado, fue sometido a las mismas pruebas a las que se habían sometido los detectores de forma individual. En la Tabla 4.4 se muestran los porcentajes de detección que se obtuvieron con el detector combinado al finalizar las pruebas.

Eficiencia del detector combinado			
Relación posición facial / eficiencia		Relación posición facial / eficiencia	
<i>Posición facial</i>	<i>Porcentaje de detección</i>	<i>Posición facial</i>	<i>Porcentaje de detección</i>
-15° / -15°	92.6 %	+15° / -15°	96.3 %
-15° / 0°	100 %	+15° / 0°	96.3 %
-15° / +15°	92.6 %	+15° / +15°	96.3 %
0° / -15°	96.3 %	+15° / +30°	96.3 %
0° / 0°	100 %	+30° / -15°	96.3 %
0° / +15°	100 %	+30° / 0°	96.3 %
0° / +30°	96.3 %	+30° / +15°	85.2 %

Tabla 4.4: Eficiencia del detector facial combinado, en relación con la posición facial del usuario.

Con base en la información proporcionada en la Tabla 4.4 se puede establecer que el detector de rostros creado satisface los requerimientos de detección con un alto porcentaje de eficiencia, salvo en la posición (+30° / +15°) en la que el porcentaje de eficiencia puede ser considerado bueno pero no óptimo.

Adicionalmente, se probó el detector con las 2 511 imágenes de la base de datos citada por Gourier *et al.* (2004), de tal manera que se pudiera determinar el alcance del detector de rostros respecto a la posición de los mismos (ver Figura 4.19) en la imagen. Tras dicha prueba se determinó que, además de las 14 posiciones mostradas en la Tabla 4.4, otras 9 posiciones faciales permiten la detección de rostro en imágenes, mediante el algoritmo de detección desarrollado, con un porcentaje de eficiencia por encima del 85 %. En la Tabla 4.5 se muestra el porcentaje de eficiencia de detección obtenida respecto a tales posiciones.

Eficiencia del detector combinado			
Relación posición facial / eficiencia		Relación posición facial / eficiencia	
<i>Posición facial</i>	<i>Porcentaje de detección</i>	<i>Posición facial</i>	<i>Porcentaje de detección</i>
-30° / -15°	85.2 %	-30° / 0°	85.2 %
-30° / 15°	88.9 %	-15° / +30°	92.6 %
0° / -30°	88.9 %	+15° / -30°	96.3 %
+30° / -30°	92.6 %	+60° / -15°	92.6 %
60° / 0°	92.6 %		

Tabla 4.5: Posiciones faciales adicionales en las que el detector combinado trabaja con una eficiencia por encima del 85 %.

En la Figura 4.23 se muestran los resultados obtenidos para cada una de las 93 posiciones faciales con las que fue probado el algoritmo del detector facial. En esta figura se muestran en color verde las posiciones faciales en las que el detector realizaba su cometido con un porcentaje de eficiencia por encima del 85.0 %, lo que se consideró como una apropiada eficiencia de detección facial. En color amarillo se encuentran las posiciones faciales en las que el porcentaje de detección se encontró por encima del 75.0 %, a lo que se denominó eficiencia aceptable de de-

tección facial. Finalmente, en color rojo, las posiciones faciales en las que el nivel de detección se consideró ineficiente para la aplicación a la que está destinado.

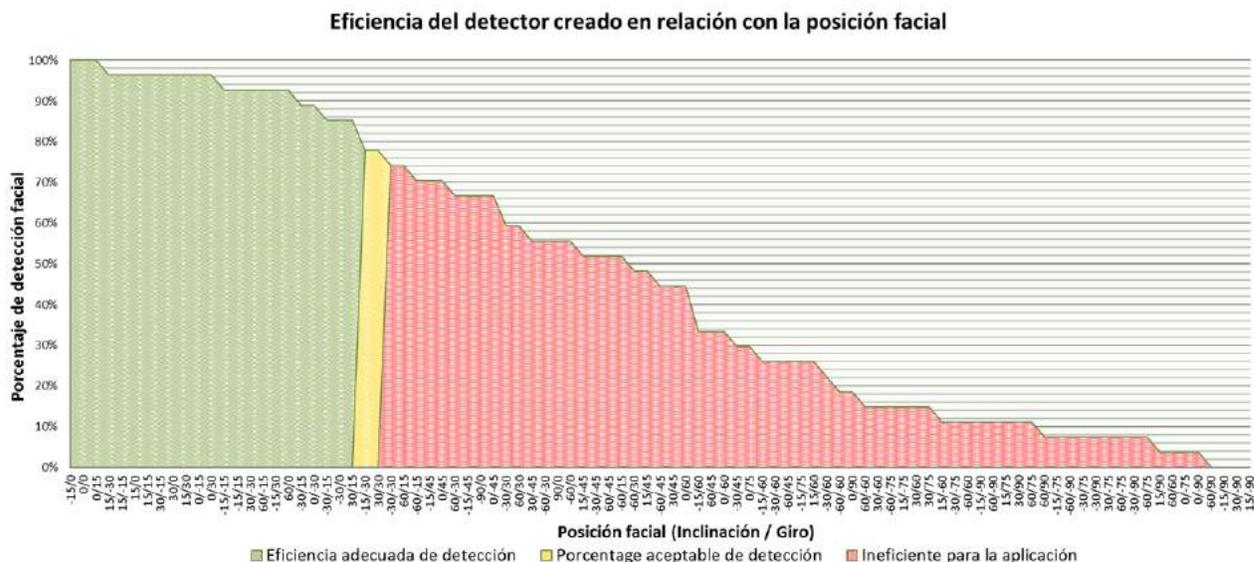


Figura 4.23: Gráfico de los resultados obtenidos mediante las pruebas realizadas para determinar el alcance del detector facial desarrollado.

De acuerdo con los datos mostrados en las tablas 4.4 y 4.5 y en la Figura 4.23, se concluye que la localización de rostro podrá llevarse a cabo de forma eficiente mediante el detector entrenado. Esto garantiza que la primera fase del proceso, establecido en la metodología, podrá realizarse de tal forma que las fases subsecuentes puedan ejecutarse sin dificultades heredadas.

4.3. Identificación de puntos relevantes en el rostro

Para la identificación de puntos relevantes, también llamados *landmarks*, del rostro se decidió utilizar el detector implementado por Kazemi y Sullivan (2014), el cual es capaz de determinar la posición de 68 puntos mediante los cuales se puede definir la localización de las características más representativas del rostro, tal como se muestra en la Figura 4.24.

Una vez que se han localizado los *Landmarks*, se obtiene un vector con las coordenadas de tales puntos, cuya posición en tal vector corresponde a los números que se muestran en la Figura 4.24; sabiendo lo anterior, es fácil extraer los datos correspondientes a las características

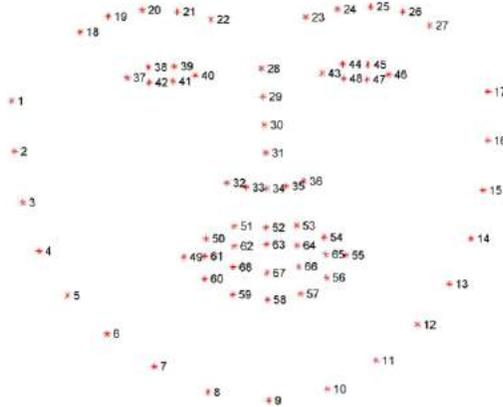


Figura 4.24: Ejemplificación de los puntos faciales detectados con el predictor de Kazemi y Sullivan (2014).

que pueden utilizarse para la determinación de somnolencia, tal como se detalla en la siguiente sección.

4.4. Detección de indicadores visuales que impliquen fatiga o distracción

4.4.1 Apertura de ojos

Mediante la utilización del detector de *Landmarks* de Kazemi y Sullivan (2014) es posible obtener un vector con la ubicación de 6 puntos descriptivos para cada ojo, tal como se muestra en la Figura 4.25. A partir de tales puntos se puede determinar la relación de aspecto ocular EAR entre la altura y el ancho del ojo, mediante la aplicación de la Ecuación (4.10) en la cual se sigue la nomenclatura de identificación de puntos mostrada en la Figura 4.25, (Soukupová y Cech, 2016). Ver el Código 5 en el capítulo de Anexos.

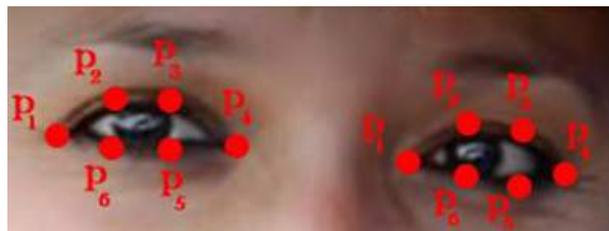


Figura 4.25: Puntos de referencia generados sobre los ojos con el detector de LandMarks de Kazemi y Sullivan (2014).

$$EAR = \frac{||p_2 - p_6|| + ||p_3 - p_5||}{2||p_1 - p_4||} \quad (4.10)$$

El valor de la EAR cuando el ojo está abierto se mantiene constante, o con variaciones muy pequeñas, por lo que es sencillo establecer un punto de comparación con el que se determine el estado de apertura del ojo. Dentro del código desarrollado en Python para la determinación de apertura del ojo, se decidió calcular el promedio de la EAR de ambos ojos para garantizar que el resultado del cálculo reflejara el estatus de apertura de ambos ojos, reduciendo así errores debidos a guiños o cualquier gesto que implique aperturas impar entre los ojos.

Una vez que se ha obtenido el cálculo combinado de la EAR, es importante establecer un tiempo máximo en el que el los ojos podrían permanecer cerrados sin ser considerado como signo de somnolencia, sino como parte de parpadeo o algún otro gesto que implique cierre momentáneo de los ojos. En caso de que el tiempo establecido sea rebasado se considerará como un signo inequívoco de somnolencia. Tras realizar diversas pruebas de funcionamiento en una computadora portátil con 12 GB de memoria RAM, procesador AMD A10 2.00GHz y una cámara de 5MP se ha definido el tiempo límite de cierre de ojos en 0.6 segundos.

4.4.2 Detección de bostezo

Para la detección de bostezo se aplica un método muy similar al aplicado para el calculo de la EAR; el primer paso es obtener los puntos que describen la boca dentro del rostro detectado en la imagen de entrada, para ello es necesario utilizar nuevamente los datos obtenidos tras utilizar el detector de *lanmarks* seleccionado (Kazemi y Sullivan, 2014).

Tras la aplicación del detector de *lanmarks*, se obtienen 12 puntos de referencia para la boca (ver Figura 4.26(a)), sin embargo, tal como se vio en la subsección 4.4.1 «Apertura de ojos», bastan 6 puntos para determinar la relación de apertura; por ello se desechó la mitad de los puntos obtenidos y seleccionando unicamente los puntos mostrados en la Figura 4.26(b) para el cálculo de apertura de la boca.

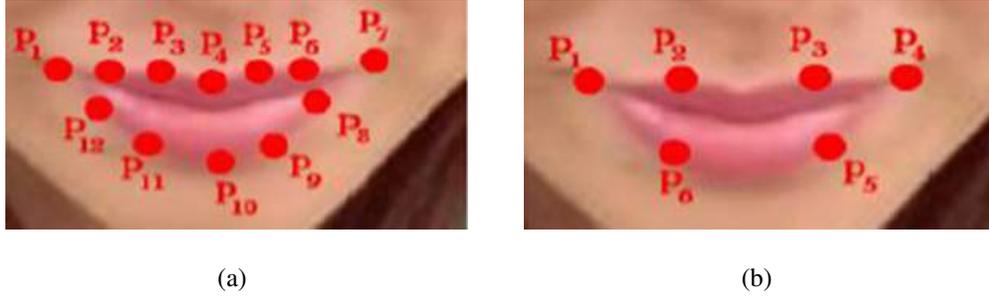


Figura 4.26: Puntos descriptivos de la boca; (a) puntos proyectados sobre la boca por el detector de puntos relevantes de Kazemi y Sullivan (2014), (b) puntos seleccionados para el cálculo de apertura bucal.

Tras definir los puntos a utilizar para el cálculo, se aplica (4.11), siguiendo la nomenclatura de identificación de puntos de la Figura 4.26(b), ver también Código 5.

$$MAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|} \quad (4.11)$$

A continuación se establece un valor de apertura máxima de la boca, en la que pudiera considerarse que tal apertura forma parte de gestos o expresiones faciales ajenas a la somnolencia; tras superar dicho punto de referencia, por un tiempo determinado, se tomará dicha apertura como indicador de la existencia de bostezo. El tiempo establecido para dicho fin es de 0.6 segundos, mientras que el punto de referencia de apertura máxima se establece en 0.62.

4.5. Procesamiento de los hallazgos.

Tras analizar los datos aportados por los *landmarks* encontrados, se puede determinar si existen o no indicadores visuales de fatiga, y en caso de que el resultado sea afirmativo se activa una alarma que alerte al conductor con el propósito de que éste tome acciones a fin de evitar un accidente automovilístico debido a la somnolencia.

Se decidió que se emitirían dos alertas diferentes, una de advertencia y duración corta para el bostezo y una más aguda y prolongada en caso de detección de ojos cerrados cuyo objetivo es alertar al conductor en caso de que este pudiera presentar micro-sueños. En ambos

casos las alarmas emitidas son tanto sonoras como visuales, en la Figura 4.27(a) se ilustra la alerta visual mostrada al detectar que el usuario ha mantenido los ojos cerrados por al menos 0.6 segundos, mientras que en la Figura 4.27(b) se puede ver la alerta visual activada tras la detección de bostezo.



(a)



(b)

Figura 4.27: Muestra de las alertas visuales emitidas por el sistema tras detectarse la presencia de indicadores visuales de somnolencia en el rostro del usuario; (a) activación de alerta por detección constante de ojos cerrados, (b) activación de alarma por bostezo.

5. EXPERIMENTACIÓN Y RESULTADOS

Inicialmente las pruebas se realizaron utilizando una computadora portátil con 12GB de memoria, procesador AMD A10 2.00GHz y una cámara de 5MP. La finalidad de tales pruebas fue la de corroborar el funcionamiento del algoritmo desarrollado, en estas pruebas participaron sujetos de diferentes complejiones y características faciales. En la Figura 5.1 se muestran imágenes de algunas de tales pruebas.

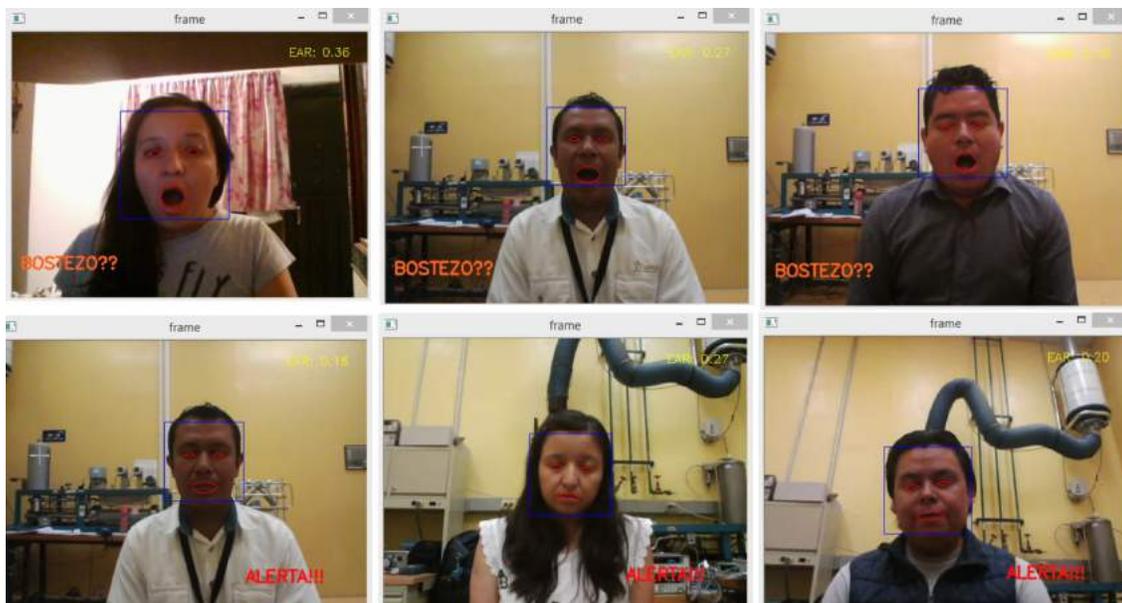


Figura 5.1: Imágenes de las primeras pruebas realizadas al algoritmo.

Además de las alertas por bostezo y cierre, por tiempo prolongado, de los ojos se muestra una indicación visual en caso de que el rostro del sujeto se encuentre en una posición en la que el algoritmo no pueda detectarlo en al menos 3 cuadros de vídeo consecutivos, tal como se muestra en la Figura 5.2.



Figura 5.2: Muestra de la indicación visual desplegada si el sistema no detecta el rostro del usuario.

Una vez que se comprobó el apropiado funcionamiento del algoritmo desarrollado se realizaron pruebas en un par de CPUs más, con procesamiento del vídeo en 5 resoluciones diferentes. Estas pruebas fueron diseñadas para conocer el tiempo de respuesta del algoritmo implementado bajo diferentes condiciones. Los resultados de tales pruebas se muestran en la Tabla 5.1

Tiempo de procesamiento por cuadro de vídeo en relación con el CPU utilizado			
	Computadora portátil	Raspberry Pi 2 B	Raspberry Pi 3 B
<i>Resolución de prueba (píxeles)</i>	<i>Tiempo promedio de procesamiento por cuadro de vídeo (segundos)</i>		
<i>1024 x 768</i>	0.531	7.47	3.51
<i>720 x 576</i>	0.252	3.17	1.32
<i>640 x 480</i>	0.212	2.51	1.05
<i>320 x 240</i>	0.075	0.503	0.179
<i>160 x 120</i>	0.036	0.167	0.058

Tabla 5.1: Tabla comparativa del tiempo de procesamiento por cuadro de vídeo para tres diferentes sistemas de procesamiento.

De la tabla 5.1 se concluye que a menor resolución, mayor velocidad de procesamiento de vídeo, sin embargo, es importante hacer notar que en las pruebas realizadas a una resolución

de prueba de 160 x 120 píxeles resulta complicado el análisis de los parámetros proporcionados por el detector de *landmarks* dado que la distribución de los puntos se realiza en un plano mucho más reducido, por lo que una variación mínima en las expresiones faciales podría mal interpretarse; además, por la misma razón, es notoria la falta de estabilidad del sistema para la detección de rostro a ángulos de posición facial ligeramente alejados de la posición 0° / 0°.

Por lo anterior, se ha decidido que la unidad de procesamiento que formará parte del sistema de detección de fatiga será la Raspberry Pi 3 B, la cual procesará el vídeo de entrada a una resolución de 320 x 240 píxeles.

Tras determinar la velocidad de procesamiento, la siguiente etapa en la fase de pruebas se enfocó en estudiar la reacción del sistema una vez instalado en un automóvil. Para dichas pruebas se instaló la cámara en el tablero del auto, justo frente al conductor, y se realizaron algunas pruebas de conducción con el sistema de detección de fatiga encendido. De entre todas las pruebas realizadas, incluidas estas últimas, se concluyó lo siguiente:

- La instalación de la cámara de vídeo en el interior del automóvil es un paso que puede realizarse de forma sencilla (ver Figura 5.3), sin embargo es importante llevar a cabo pruebas de enfoque pues es necesario que la cámara sea capaz de captar el rostro del conductor en todo momento.

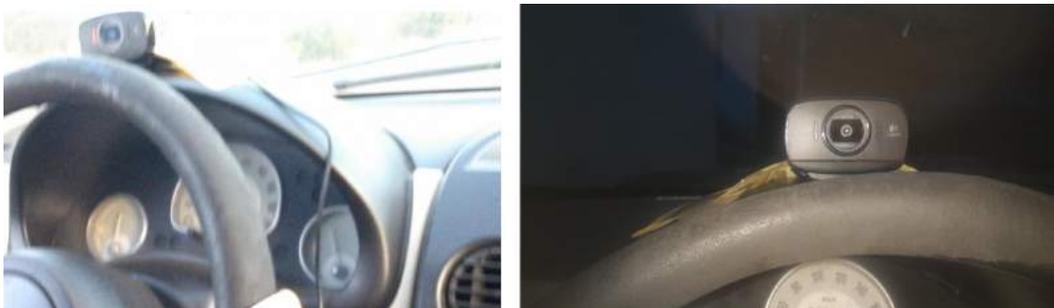


Figura 5.3: Instalación de la cámara de vídeo en el interior del automóvil.

- La vibración generada por el automóvil, mientras este se encuentra en movimiento, no afecta la capacidad de detección del sistema, ello incluye detección de rostro y *landmarks*. Es importante mencionar que las pruebas de conducción se realizaron sobre pavimento asfáltico cubierto en pequeñas secciones por terrecería 5.4.



Figura 5.4: Fotografía de la superficie sobre la que se realizó la prueba del sistema implementada durante la conducción.

- Es posible detectar el rostro del conductor aun mientras este realiza una inspección rápida a los espejos laterales del automóvil, pues generalmente no es necesario sobregirar el rostro para poder ver a través de tales espejos, lo que hace que los ángulos de detección de rostro alcanzados por el sistema (ver Tablas 4.4 y 4.5) sean suficientes para cubrir esta tarea inherente a la conducción.
- El sistema mantiene su capacidad de respuesta ante el uso de anteojos transparentes (Figura 5.5), mientras que su uso con lentes oscuras desecha la garantía de buenos resultados.

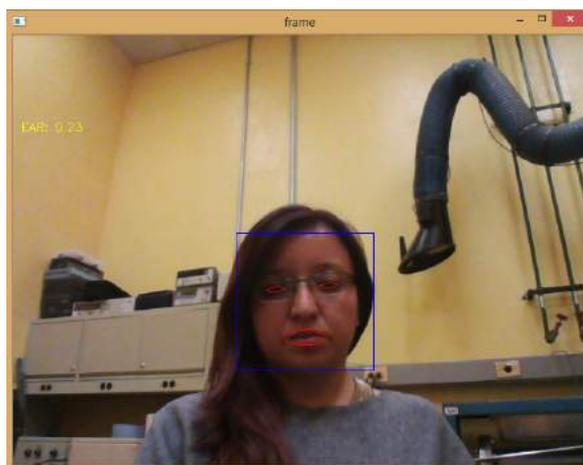


Figura 5.5: Muestra de detección realizada mientras el sujeto de prueba utiliza lentes transparentes.

- El rostro del usuario puede ser detectado aún cuando este tenga cubierta una pequeña porción del rostro de forma momentánea (Figura 5.6), si se cubre al menos una cuarta parte del rostro por tiempo prolongado (más de 5 cuadros de vídeo) o un ojo (o ambos) no

se garantiza una buena detección de rostro, lo que deriva en una alta probabilidad de fallo en la detección de somnolencia.

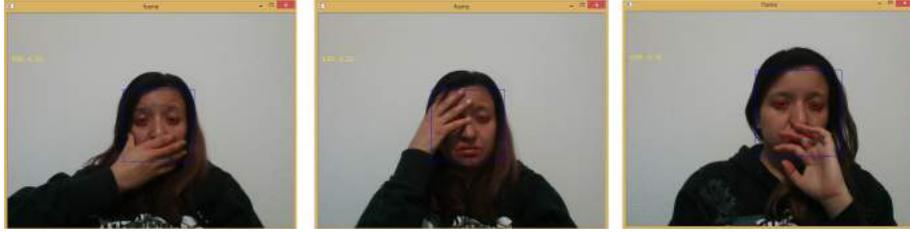


Figura 5.6: Imágenes ilustrativas de la detección de rostro, y sus características principales cuando parte del rostro se halla cubierta.

- La iluminación juega un papel importante en los sistemas de visión por computadora, por lo que el algoritmo diseñado garantiza resultados siempre y cuando la iluminación del entorno de conducción sea lo suficientemente buena como para distinguir claramente, mediante el procesamiento de imágenes, el rostro y sus principales características. De las pruebas de detección a baja iluminación (ver Figura 5.7) se concluye que aun cuando el sistema sea capaz de detectar el rostro en entornos oscuros, no es capaz de detectar con precisión los ojos ni la boca por lo que resulta inútil para la detección de somnolencia; por ello, se concluye que el algoritmo no presentará un buen desempeño en entornos de baja iluminación, es decir, no podrá ser utilizado eficientemente durante la noche.

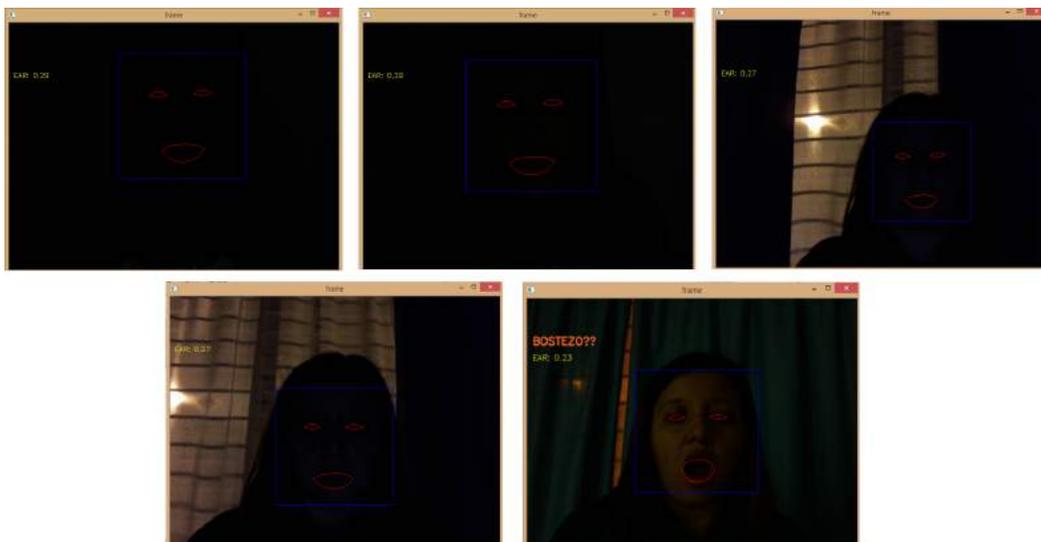


Figura 5.7: Imágenes tomadas de las pruebas de detección a baja iluminación.

- Finalmente, se concluye que un sistema portátil compuesto por:
 - Una cámara de vídeo de entre 5 MP y 8 MP
 - Una Raspberry Pi 3B
 - Una Batería portátil con una salida de +5V y 2.5 A
 - Una Bocina con conexión mediante Bluetooth
 - Una Pantalla táctil LCD, compatible con RaspBerry Pi

en el que se implemente el algoritmo desarrollado para la detección de somnolencia, procesando el vídeo a una resolución de (320 x 240) pixeles puede llevar a cabo, de forma eficiente, la tarea de detectar somnolencia en conductores. En la Figura 5.8 se muestra como se ha instalado el sistema con dichas características, de tal manera que sea capaz de cumplir su función, dentro del automóvil utilizado para las pruebas.



Figura 5.8: Sistema portátil de detección de somnolencia, instalado en el automóvil.

Por último, en la Figura 5.9 se indica la ubicación de cada uno de los componentes del sistema.

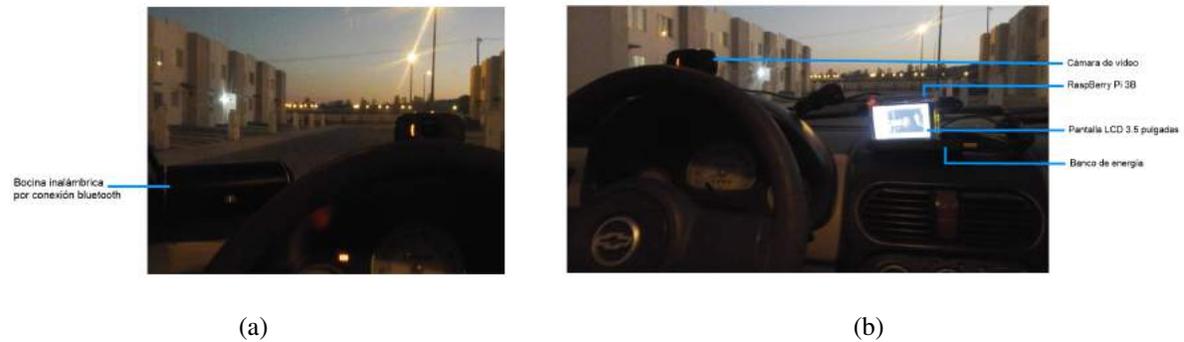


Figura 5.9: Distribución del *hardware* que compone el sistema portátil de detección de somnolencia; a la izquierda (a), la ubicación de la bocina inalámbrica utilizada para la emisión de la alarma sonora, a la derecha, (b) instalación de la cámara de vídeo de 8 MP, el banco de energía o batería portátil con una salida de +5V y 2.5 A, la Raspberry Pi 3B y, finalmente, la pantalla LCD de 3.5 pulgadas.

5.0.1 Publicación

Como parte de los resultados del presente proyecto, se realizó la publicación del artículo titulado «**Sistema de detección de somnolencia mediante procesamiento de expresiones faciales**» (Ver Anexo 5), la presentación de dicho artículo se llevó a cabo en el **XIV Congreso Internacional sobre Innovación y Desarrollo Tecnológico CIINDET 2018** mismo que contó con el respaldo del IEEE, sección Morelos. Las memorias del CIINDET serán publicadas bajo el nombre de «**Herramientas y tecnologías específicas para beneficio de las instituciones y comunidades**», y cuyo ISBN es 978-607-95255-8-3. Para mayor información, recurrir a la página <http://ciindet.org/index.html>.

BIBLIOGRAFÍA

- AAA Automobile Club. 2014. Evaluation of Blind Spot Monitoring and Blind Spot. Technical report AAA.
- Achatz, Simon. 2016. State of the Art of Object Recognition Techniques. Technical report Neuroscientific System Theory Munich.
<https://www.nst.ei.tum.de/fileadmin/w00bqs/www/publications/as/SS2016{ }AS{ }ObjectRecogn.pdf>
- American Honda Motor. 2017. Blind Spot Information System.
- Amit, Yali y Pedro Felzenszwalb. 2014. Object Detection. Computer Vision: A Reference Guide :537–542.
- AT&T Laboratories Cambridge. 2002. The ORL Database of Faces.
- Audi AG. 2011a. Adaptive light.
- Audi AG. 2011b. Audi active lane assist.
- Audi AG. 2016. Night vision assistant.
- Auricht, Maik y Rainer Stark. 2014. How to consider emotional reactions of the driver within the development of Advanced Driver Assistance Systems (ADAS)? Procedia CIRP 21(figure 1):70–74.
- Awad, Ali Ismail y Mahmoud Hassaballah. 2016. Image Feature Detectors and Descriptors. 630 of Studies in Computational Intelligence Springer International Publishing Cham.
- Bainbridge, Wilma A., Phillip Isola y Aude Oliva. 2013. The Instrict Memorability of Face Photographs. Journal of Experimental Psychology 142(4):1323–1334.
- Bär, Tobias, Jan Felix Reuter y J. Marius Zöllner. 2012. Driver head pose and gaze estimation based on multi-template ICP 3-D point cloud alignment. 2012 15th International IEEE Conference on Intelligent Transportation Systems. :1797–1802.

- Bauer, Eric y Ron Kohavi. 1999. An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants. *Machine Learning* 36(1):105–139.
- Bay, Herbert, Tinne Tuytelaars y Luc Van Gool. 2006. SURF: Speeded Up Robust Features. Berlin, Heidelberg Springer Berlin Heidelberg :404–417.
https://doi.org/10.1007/11744023{}_32
http://link.springer.com/10.1007/11744023{}_32
- Belbachir, Assia, Jean-Christophe Smal, Jean-Marc Blosseville y Dominique Gruyer. 2012. Simulation-Driven Validation of Advanced Driving-Assistance Systems. *Procedia - Social and Behavioral Sciences* 48:1205–1214.
- Biondi, Francesco, David L. Strayer, Riccardo Rossi, Massimiliano Gastaldi y Claudio Mulatti. 2017. Advanced driver assistance systems: Using multimodal redundant warnings to enhance road safety. *Applied Ergonomics* 58:238–244.
- BMW GROUP MÉXICO. 2016. Sistemas de asistencia al conductor BMW ConnectedDrive en el BMW Serie 1 tres puertas.
- Bărbuceanu, Florin, Mihai Duguleană, Stoianovici Vlad y Adrian Nedelcu. 2011. Evaluation of the Average Selection Speed Ratio between an Eye Tracking and a Head Tracking Interaction Interface. Berlin, Heidelberg Springer Berlin Heidelberg :181–186.
http://dx.doi.org/10.1007/978-3-642-19170-1{}_20
- CATAM. N.d. DETECCIÓN DE CARAS Y ANÁLISIS DE EXPRESIONES FACIALES.
- Chapman, Peter R y Geoffrey Underwood. 1998. Visual search of driving situations : Danger and experience. *Perception* 27(8):951–964.
- Coetzer, R C y G P Hancke. 2011. Eye detection for a real-time vehicle driver fatigue monitoring system. 2011 IEEE Intelligent Vehicles Symposium (IV). :66–71.
- Colic, Aleksandar. 2014. Design and implementation of driver drowsiness detection system PhD thesis Florida Atlantic University.

- Čolić, Aleksandar, Oge Marques y Borko Furht. 2014. Driver Drowsiness Detection. Springer-Briefs in Computer Science Springer International Publishing Cham.
- Daimler AG. 2016a. Active Lane Keeping Assist.
- Daimler AG. 2016b. ATTENTION ASSIST system.
- Daimler AG. 2016c. Night View Assist.
- Dalal, N y B Triggs. 2005. Histograms of oriented gradients for human detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). 1 :886–893 vol. 1.
<https://hal.inria.fr/file/index/docid/548512/filename/hog{ }cvpr2005.pdf>
- Dalal, Navneet. 2006. Finding People in Images and Videos Theses Institut National Polytechnique de Grenoble - INPG.
- Das, Apurba. 2015. Guide to signals and patterns in image processing: Foundations, methods and applications. Springer International Publishing Switzerland 2015.
- Daza, Iván G, Luis M Bergasa, Sebastián Bronte, J Javier Yebes, Javier Almazán y Roberto Arroyo. 2014. Fusion of optimized indicators from Advanced Driver Assistance Systems (ADAS) for driver drowsiness detection. Sensors 14(1):1106–1131.
- Del Abril De Mur, Antonio. 2017. Detección de personas en secuencias de imágenes en entornos interiores PhD thesis UNIVERSIDAD DE ALCALÁ.
- Department of Engineering Science. 2004. No Title.
- Devi, Mandalapu Sarada y Preeti R Bajaj. 2008. Driver fatigue detection based on eye tracking. Emerging Trends in Engineering and Technology, 2008. ICETET'08. First International Conference on. IEEE :649–652.
- Dinges, David F. y Richard Grace. 1998. PERCLOS : A Valid Psychophysiological Measure of Alertness As Assessed by Psychomotor Vigilance.
<https://ntl.bts.gov/lib/10000/10100/10114/tb98-006.pdf>

- Duchowski, Andrew T. 2007. *Eye Tracking Methodology: Theory and Practice*. Springer-Verlag New York, Inc. Secaucus, NJ, USA.
- Fletcher, Luke, Gareth Loy, Nick Barnes y Alexander Zelinsky. 2005. Correlating driver gaze with the road scene for driver assistance systems. *Robotics and Autonomous Systems* 52:71–84.
- Flores, Marco Javier, José María Armingol y Arturo de la Escalera. 2010. Real-Time Warning System for Driver Drowsiness Detection Using Visual Information. *Journal of Intelligent {&} Robotic Systems* 59(2):103–125.
- Flores, Pablo y Juan Braun. 2011. Algoritmo SIFT: fundamento teórico.
- Ford Motor Company. 2017a. Asistente de pre-colisión con detección de peatones.
- Ford Motor Company. 2017b. Tecnología de iluminación Ford.
- Ford Motor Company. N.d. Detector de Fatiga.
- Gaikwad, V y S Lokhande. 2015. Lane Departure Identification for Advanced Driver Assistance. *IEEE Transactions on Intelligent Transportation Systems* 16(2):910–918.
- Gang, Longhui, Xiaolin Song, Mingheng Zhang, Baozhen Yao y Liping Zhou. 2017. Analysis of driver fatigue causations based on the Bayesian network model. *SIMULATION* 93(7):553–565.
- Gonzalez, Rafael C y Richard E. Woods. 2008. *Digital Image Processing Third Edition*.
- Gotzig, Heinrich. 2016. *Parking Assistance*. Cham Springer International Publishing :1077–1092.
https://doi.org/10.1007/978-3-319-12352-3_{_}45
- Gourier, N., D. Hall y J.L. Crowley. 2004. Estimating face orientation from robust detection of salient facial structures. *FG Net Workshop on Visual Observation of Deictic Gestures (POINTING)* :17–25.
- Grgic, Mislav y Kresimir Delac. 2017. *FACE RECOGNITION HOMEPAGE: DATABASES*.

- Heikkilä, Marko, Matti Pietikäinen y Cordelia Schmid. 2009. Description of interest regions with local binary patterns. *Pattern Recognition* 42(3):425–436.
- Hjelmas, Erik y Boon Kee Low. 2001. Face Detection : A Survey. *Computer vision and image understanding* 83(3):236–274.
- INEGI. 2016. ESTADÍSTICAS A PROPÓSITO DEL DÍA MUNDIAL EN RECUERDO DE LAS VÍCTIMAS DE LOS ACCIDENTES DE TRÁFICO; TERCER DOMINGO DEL MES DE NOVIEMBRE. :1–9.
- Jaguar Land Rover. 2017. JAGUAR LAND ROVER UNVEIL NEW DRIVER MONITOR SYSTEM THAT IS SET TO REVOLUTIONISE ROAD SAFETY.
- Jalled, Fares y Ilia Voronkov. 2016. Object Detection using Image Processing. :1–6.
- Jimenez Moreno, Robinson. 2011. SISTEMA DE DETECCION CONDUCTORES MEDIANTE ´ POR TECNICAS DE VISI ON PhD thesis Universidad Nacional de Colombia.
- Jung, Sang-Joong, Heung-Sub Shin y Wan-Young Chung. 2012. Highly sensitive driver health condition monitoring system using nonintrusive active electrodes. *Sensors and Actuators B: Chemical* 171:691–698.
- Kazemi, Vahid y Josephine Sullivan. 2014. One Millisecond Face Alignment with an Ensemble of Regression Trees. *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. CVPR '14 Washington, DC, USA IEEE Computer Society* :1867–1874.
<http://dx.doi.org/10.1109/CVPR.2014.241>
- Khachane, Divyata y Prof Anjali Shrivastav. 2016. Antilock Braking System and Its Advancement. *International Research Journal of Engineering and Technology (IRJET)* 03(05):2221–2225.
- Kisa, Branislav y Margrit Gelautz. 2014. *Advances in Embedded Computer Vision*. Springer International Publishing Switzerland 2014.

- Klauer, S.G., T. A. Dingus, V. L. Neale, J.D. Sudweeks y D.J. Ramsey. 2006. The Impact of Driver Inattention On Near-Crash / Crash Risk : An Analysis Using the 100-Car Naturalistic Driving Study Data. p. 226.
- Kwon, O, J Chun y P Park. 2006. Cylindrical Model-Based Head Tracking and 3D Pose Recovery from Sequential Face Images. 2006 International Conference on Hybrid Information Technology. 1 :135–139.
- Lanitis, A, C J Taylor y T F Cootes. 1994. Automatic tracking, coding and reconstruction of human faces, using flexible appearance models. *Electronics Letters* 30(19):1587–1588.
- Lenskiy, Artem A y Jong-Soo Lee. 2012. Driver's eye blinking detection using novel color and texture segmentation algorithms. *International Journal of Control, Automation and Systems* 10(2):317–327.
- Lewis, Bob. 1999. A look at the history of the automobile offers a glimpse of the PC's future. *InfoWorld* p. 76.
- Li, Stan Z, Long Zhu, ZhenQiu Zhang, Andrew Blake, HongJiang Zhang y Harry Shum. 2002. Statistical learning of multi-view face detection. *European Conference on Computer Vision*. Springer :67–81.
- Liang, Wen Chieh, John Yuan, Deh Chuan Sun y Ming Han Lin. 2009. Changes in Physiological Parameters Induced by Indoor Simulated Driving: Effect of Lower Body Exercise at Mid-Term Break. *Sensors* 9(9):6913–6933.
- Lienhart, R. y J. Maydt. 2002. An extended set of Haar-like features for rapid object detection. *Proceedings. International Conference on Image Processing* 1:I–900–I–903.
- Lowe, David G. 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60(2):91–110.
- Lowe, D.G. 1999. Object recognition from local scale-invariant features. *Proceedings of the Seventh IEEE International Conference on Computer Vision*. 2 IEEE :1150–1157 vol.2.
<http://ieeexplore.ieee.org/document/790410/>

- Lyons, M, S Akamatsu, M Kamachi y J Gyoba. 1998. Coding facial expressions with Gabor wavelets. Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition. :200–205.
- Mallick, Sataya. 2017. Histogram of Oriented Gradients | Learn OpenCV.
- Matuska, Slavomir, Robert Hudec, Patrik Kamencay, Miroslav Benco y Martina Zachariasova. 2014. Classification of wild animals based on SVM and local descriptors. AASRI Procedia 9:25–30.
- Mihai, Dugulean\ua, G\^\irbacia Florin y Mogan Gheorghe. 2015. Using Dual Camera Smartphones As Advanced Driver Assistance Systems: NAVIEYES System Architecture. Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments. PETRA '15 New York, NY, USA ACM :23:1—23:8.
<http://doi.acm.org/10.1145/2769493.2769513>
- Ming-Hsuan Yang, D.J. Kriegman y Narendra Ahuja. 2002. Detecting faces in images: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(1):34–58.
- Mobileye. 2015. Sistemas Mobileye de Asistencia Avanzada al Conductor.
- Modi, Mitul y Fedrik Macwan. 2014. Face Detection Approaches : A Survey. International Journal of Innovative Research in science, engineering and tchnology 3(4):11107–11116.
- Mohan, Anuj, Constantine Papageorgiou y Tomaso Poggio. 2001. Example-based object detection in images by components. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(4):349–361.
- Morán Cruz, Natalia. 2013. Desarrollo de un sistema avanzado de asistencia a la conducción en tiempo real para la detección de peatones en entornos urbanos complejos PhD thesis UNIVERSIDAD CARLOS III DE MADRID.
- Mordvintsev, Alexander y Abid K. 2013. Face Detection using Haar Cascades.
- Murphy-Chutorian, E y M M Trivedi. 2009. Head Pose Estimation in Computer Vision: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(4):607–626.

Negri, Pablo, Xavier Clady y Lionel Prevost. 2007. Benchmarking haar and histograms of oriented gradients features applied to vehicle detection. ICINCO-RA (1). :359–364.

<https://www.researchgate.net/profile/Lionel{ }Prevost/publication/221644952{ }Benchmarking{ }haar{ }and{ }histograms{ }of{ }or links/00b495162c8fd06f80000000.pdf>

Nissan. 2017. Nissan's «Driver Attention Alert» helps detect erratic driving caused by drowsiness and inattention.

Nixon, Mark S y Alberto S Aguado. 2012. Feature extraction & image processing for computer vision. 3rd ed. Academic Press.

Nuevo, J, L M Bergasa, M A Sotelo y M Ocana. 2006. Real-time robust face tracking for driver monitoring. 2006 IEEE Intelligent Transportation Systems Conference. :1346–1351.

OMS. 2015. Informe Sobre La Situación Mundial De La Seguridad Vial 2015. Ginebra, Suiza :1–12.

Papageorgiou, Constantine P, Michael Oren y Tomaso Poggio. 1998. A general framework for object detection. Computer vision, 1998. sixth international conference on. IEEE :555–562.

Pavani, Sri-kaushik, David Delgado y Alejandro F Frangi. 2010. Haar-like features with optimally weighted rectangles for rapid object detection. Pattern Recognition 43(1):160–172.

Planning, Patrick. 2014. Innovation Acceptance; The case of Advanced Driver-Assistance Systems. Springer Gabler Stuttgart , Germany.

Porsche Latin America. 2017a. Asistente de visión nocturna.

Porsche Latin America. 2017b. Faros de matriz LED con Porsche Dynamic Light System Plus (PDLS Plus).

Prince, Simon J D. 2012. Computer vision: models, learning, and inference. Cambridge University Press.

Regan, Michael A y David L Strayer. 2014. Towards an Understanding of Driver Inattention : Taxonomy and Theory. Annals of Advances in Automotive Medicine 58:5–14.

- Reisfeld, Daniel y Yehezkel Yeshurun. 1992. Robust Detection of Facial Features by Generalized Symmetry The Generalized Symmetry Opera-. :117–120.
- Ruiz Sancho, Cristina. 2014. Pedestrian Detection using a boosted cascade of Histogram of Oriented Gradients.
- Sahayadhas, Arun, Kenneth Sundaraj y Murugappan Murugappan. 2012. Detecting Driver Drowsiness Based on Sensors: A Review. *Sensors* 12(12):16937–16953.
- Samaria, F S y A C Harter. 1994. Parameterisation of a stochastic model for human face identification. *Proceedings of 1994 IEEE Workshop on Applications of Computer Vision*. :138–142.
- Schapire, Robert E. 2003. The boosting approach to machine learning: An overview. *Nonlinear estimation and classification*. Springer :149–171.
- Senaratne, Rajinda, Budi Jap, Sara Lal, Arthur Hsu, Saman Halgamuge y Peter Fischer. 2011. Comparing two video-based techniques for driver fatigue detection: classification versus optical flow approach. *Machine Vision and Applications* 22(4):597–618.
- Shantaiya, Sanjivani, Keshri Verma y Kamal Mehta. 2013. A Survey on Approaches of Object Detection. *International Journal of Computer Applications* 6518(18):975–8887.
- Sharif, Muhammad, Adeel Khalid, Mudassar Raza y Sajjad Mohsin. 2011. Face Recognition using Gabor Filters. *Journal of Applied Computer Science & Mathematics* (11).
- Soo, Sander. 2014. Object detection using Haar-cascade Classifier. *Institute of Computer Science, University of Tartu* .
- Soukupová, Tereza y Jan Cech. 2016. Real-Time Eye Blink Detection using Facial Landmarks. *21st Computer Vision Winter Workshop* .
- Stauffer, Chris y W Grimson. 2000. Learning Patterns of Activity Using Real-Time Tracking. *Ieee Transactions on Pattern Analysis and Machine Intelligence* 22(8):747–757.
- STCONAPRA y Secretaría de salud. 2015. Informe s obre la S ituación de la S eguridad V ial, México 2015. 1ra ed. Secretaría de Salud Ciudad de México, México.

- Strayer, David L., Jason M. Watson y Frank A. Drews. 2011. Cognitive Distraction While Multitasking in the Automobile. *The Psychology of Learning and Motivation*. Burlington chapter 2, :29–58.
- Strohl, Kingman P., Jesse Blatt, Forrest Council, Kate Georges, James Kiley, Roger Kurrus, Anne T. McCartt, David Willis, Pat Waller, Jane Stutts, Thomas Roth, Susan Rogus, Allan I. Pack y Sharon L. Merritt. 1998. *Drowsy Driving and Automobile Crashes*. National Highway Traffic Safety Administration.
- Sukanya, C M, Roopa Gokul y Vince Paul. 2016. A Survey on Object Recognition Methods. *International Journal of Science, Engineering and Computer Technology* 6(1):48–52.
- Suzuki, M, N Yamamoto, O Yamamoto, T Nakano y S Yamamoto. 2006. Measurement of Driver's Consciousness by Image Processing -A Method for Presuming Driver's Drowsiness by Eye-Blinks coping with Individual Differences -. 2006 IEEE International Conference on Systems, Man and Cybernetics. 4 :2891–2896.
- Tan, Pang-ning, Michael Steinbach y Vipin Kumar. 2006. Introduction to data mining.
- The MathWorks Inc. 2017. Train a Cascade Object Detector.
- Thomaz, Carlos Eduardo y Gilson Antonio Giralddi. 2010. A new ranking method for principal components analysis and its application to face image analysis. *Image and Vision Computing* 28(6):902–913.
- Tigadi, Arun, Rudrappa Gujanatti y Anil Gonchi. 2016. Advanced Driver Assistance Systems. *International Journal of Engineering Research and General Science* 4(3):151–158.
- Toyota Motor Corporation. 2017. Automatic High Beam (AHB).
- Turk, Matthew y Alex Pentland. 1991. Eigenfaces for recognition. *Journal of cognitive neuroscience* 3(1):71–86.
- Ueno, H, M Kaneda y M Tsukino. 1994. Development of drowsiness detection system. *Vehicle Navigation and Information Systems Conference, 1994. Proceedings., 1994.* :15–20.

- Universitatea Tehnica din Cluj-Napoca. N.d. Pattern recognition systems – Lab 5 Histograms of Oriented Gradients.
- Viola, P. y M. Jones. 2001. Rapid object detection using a boosted cascade of simple features. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001 1:I-511–I-518.
- Viola, Paul y Michael Jones. 2003. Fast multi-view face detection. Mitsubishi Electric Research Lab TR-20003-96 3(14):2.
- Volkswagen. 2016a. Detector de fatiga.
- Volkswagen. 2016b. Park Assist.
- VOLVO. 2015. Aviso de colisión* - detección de peatones.
- Volvo. 2016. Aparcamiento asistido.
- Xiaojie, Yang. 2001. Real Time Visual Cues Extraction for Monitoring Driver Vigilance PhD thesis University of Nevada.
- Yu, Xun. 2009. Real-time Nonintrusive Detection of Driver Drowsiness. Technical Report 2 Center for Transportation Studies University of Minnesota Washington, DC USA.
<http://www.cts.umn.edu/Publications/ResearchReports/reportdetail.html?id=1768>
<http://www.cts.umn.edu/Research/ProjectDetail.html?id=2008017>
- Yuille, Alan L, Peter W Hallinan y David S Cohen. 1992. Feature extraction from faces using deformable templates. International journal of computer vision 8(2):99–111.
- Zhang, Mingheng, Gang Longhui, Zhe Wang, Xiaoming Xu, Baozhen Yao y Liping Zhou. 2014. Hybrid Model for Early Onset Prediction of Driver Fatigue with Observable Cues. Mathematical Problems in Engineering 2014(385716):1–9.
- Zhao, Chunlin, Min Zhao, Jianpin Liu y Chongxun Zheng. 2012. Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator. Accident Analysis & Prevention 45:83–90.

ANEXOS

Anexo A

Código en MATLAB; Entrenamiento del detector

```
1 % Inicializacion de variables
2 clc
3 clear all
4 nPositivas=19398; %numero de imagenes que integran el set de
   entrenamiento positivo
5 % crear lista de positivas; se edita una lista en formato *.txt
   con la informacion del set de entrenamiento positivo
6 positiveData= struct('imageFilename', 'positiva', '
   objectBoundingBoxes', 'coordenadas');
7 for i=1:nPositivas
8 positiveData(i).imageFilename = strcat(['Positivas\positiva'
   num2str(i) '.bmp']);
9 positiveData(i).objectBoundingBoxes = [1, 1, 240, 320];
10 end
11
12 % asignando folder de negativas
13 negativeFolder= fullfile('C:\Users\VIA\Documents\
   Train_FaceDetector\Training_060417\Negativas');
14 disp('listo el folder negativo... empezando entrenamiento');
15 % entrenamiento (Utiliza HOG por defecto, por eso no se incluye
   en parametros)
16 trainCascadeObjectDetector('faceTrain_190517.xml', positiveData
   , negativeFolder, 'FalseAlarmRate',0.3, 'NumCascadeStages',15)
   ;
```

Anexo B

Código en Python; Detección de rostro

```
1 # Iniciar captura de video
2 print("[INFO] starting video stream thread...")
3 print("[INFO] Cuadro a 640")
4 vs = cv2.VideoCapture(1)
5 # loop sob cuadros de video
6 noCamCounter=0
7 while True:
8 # Capturar cuadro por cuadro
9 ret, frame = vs.read()
10 camTime=0
11 tInicioCiclo= time.time()
12 frame = imutils.resize(frame, width=640)
13 # Las operaciones sobre los cuadros de video comienzan aqui
14 #se convierte a grayscale
15 gray = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)
16 # Detectar rostros en los cuadros de video en escala de grises
17 faces = detector_3.detectMultiScale(gray, 1.3, 5)
18 print "on getFace"
19 print ("%s" % len(faces))
20 if len(faces) < 1: # face found using first detector
21 faces = detector.detectMultiScale(gray, 1.07, 8, 0, (120, 120))
22 if len(faces) < 1: # face found using first detector
23 faces = detector_2.detectMultiScale(gray, 1.07, 6, 0, (120, 120))
24 if len(faces) < 1: # face found using third detector
25 time.sleep(0.001)
26 noCamCounter=noCamCounter+1
27 if noCamCounter >=3:
28 cv2.putText(frame, "NO FACE DETECTED!", (10, 30), cv2.
    FONT_HERSHEY_SIMPLEX, 0.7, (15, 255, 15), 2)
```

Anexo C

Código en Python; Cálculo de EAR

```
1 def eye_aspect_ratio(eye):
2     # Calcular las distancias euclidianas entre dos conjuntos de
      cordenadas (x,y) verticales de los landmarks localizados en los
      ojos
3     A = dist.euclidean(eye[1], eye[5])
4     B = dist.euclidean(eye[2], eye[4])
5
6     # Calcular la distancia euclidiana entre las cordenadas (x,y)
      verticales de los lanmarks oculares.
7     C = dist.euclidean(eye[0], eye[3])
8
9     # Calcular EAR
10    ear = (A + B) / (2.0 * C)
11
12    # Entregar EAR
13    return ear
```

Anexo D

Código en Python; Cálculo de apertura bucal

```
1 def mouth_aspect_ratio(mouth):
2     # Calcular las distancias euclidianas entre dos conjuntos de
      cordenadas (x,y) verticales de los landmarks localizados en la
      boca
3     AM = dist.euclidean(mouth[2], mouth[10])
4     BM = dist.euclidean(mouth[4], mouth[8])
5
6     # Calcular la distancia euclidiana entre las cordenadas (x,y)
      verticales de los lanmarks bucales.
7     # eye landmark (x, y)-coordinates
8     CM = dist.euclidean(mouth[0], mouth[6])
9
10    # Calcular la relacion de aspecto de apertura bucal
11    mar = (AM + BM) / (2.0 * CM)
12
13    # Entregar resultado
14    return mar
```

Anexo E



Constancia de presentación de artículo en el XIV Congreso Internacional sobre Innovación y Desarrollo Tecnológico CIINDET 2018

Sistema de detección de somnolencia mediante procesamiento de expresiones faciales

Cintia V. Alcántara-Montiel^a, J.Carlos Pedraza-Ortega^b

Universidad Autónoma de Querétaro, Facultad de Ingeniería

^a*cinthia.vamontiel@hotmail.com*, ^b*caryoko@yahoo.com*

Resumen—Los accidentes de tránsito en los que se involucra la conducción con somnolencia forman parte de un porcentaje considerable del total de los accidentes anuales en nuestro país, mientras que resulta alarmante la relación de mortandad por accidentes de tránsito y el nivel de ingresos de un país. Este trabajo busca desarrollar un algoritmo mediante el cual detectar somnolencia en tiempo real para alertar al conductor de manera oportuna, considerando para ello varias técnicas y conceptos de visión por computadora. Para lograrlo, se inicia con la detección facial del usuario, posteriormente se establecen puntos de referencia para describir ojos y boca, de ellos se analizan los valores de apertura; de los primeros para definir el parpadeo, de la segunda para establecer la existencia de bostezo, para finalmente, emitir alarma de ser necesario. Se realizaron pruebas con resultados satisfactorios pero con muestras claras de las oportunidades de mejora en el algoritmo.

Palabras clave. ADAS, visión por computadora, procesamiento de imágenes, Sistemas embebidos.

1. Introducción

El automóvil, desde que fuera lanzado al mercado por primera vez, empoderó al usuario respecto a la forma de transportarse. Si bien los primeros automóviles no eran realmente sencillos de manipular ni mantener, y mostraban claras deficiencias de desempeño, la industria automotriz nunca ha detenido su avance en los campos de investigación y desarrollo [1]. Al pasar de los años, la presencia automotriz en la vida cotidiana ha crecido considerablemente y con ello, ha aumentado también el número de accidentes de tránsito. En los años 70 se respondió a dicho problema con la propuesta de implementar sistemas de asistencia al conductor (DAS, por sus siglas en inglés) que hicieran de la conducción una tarea más segura de lo que era hasta el momento. Sin embargo, fue hasta los años 80 que tales sistemas se pusieron al alcance del consumidor. Estos sistemas también han evolucionado; comenzaron como apoyo a tareas básicas de conducción, tales como el frenado, y actualmente pueden verse como una posibilidad real de llegar a la conducción autónoma del automóvil [2,3].

Tras su evolución, los DAS han pasado a ser ADAS (sistemas avanzados de asistencia al conductor), cuya función es ayudar al conductor a prevenir errores que puedan resultar en un accidente de tráfico o, en el peor de los casos, reducir el impacto de tales errores. Los ADAS deben dar soporte al conductor a lo largo del viaje y alertarlo en caso de detectar factores de riesgo, para ello deben considerar condiciones de manejo tanto internas como externas. Los sistemas de anticollisión, aviso de cambio de carril, monitoreo de punto ciego y detección de somnolencia son algunos ejemplos de ADAS [4,5].

La presencia de sistemas básicos de asistencia al conductor, tales como el sistema antibloqueo de freno, es bastante alta, sin embargo, los sistemas más complejos no corren con la misma suerte. Una de las principales causas de ello estriba en el costo de tales sistemas. De acuerdo con la OMS, a pesar de que alrededor del 50% de los automóviles en el mundo se concentran en países de bajos y medianos ingresos es en tales países en donde ocurre el 90% de las muertes debidas a accidentes de tránsito [6]. En México, las lesiones ocasionadas por accidentes de tránsito se han elevado tanto que forman parte de las 10 principales causas de muerte [7], lo anterior sin que se implemente en el país una apropiada legislación en términos de seguridad vial y sin que se genere una base de datos que permita determinar las causas principales de tales accidentes. Un ejemplo de las posibles causas de accidentes de tránsito que es completamente ignorada por la secretaría de salud del país en sus informes, es la somnolencia, esto a pesar de que un conductor somnoliento tiene un 24% más de probabilidad de ocasionar un accidente [8].

Si bien la somnolencia no posee características medibles mediante pruebas objetivas, como las basadas en análisis de sangre, es posible analizar las características de un accidente de tránsito para determinar si la somnolencia pudo o no ser parte de las causas del mismo. Tales características incluyen; el vehículo salió de la carretera sin causa aparente, el accidente ocurre durante la las 14 y las 16 horas o entre las 0 horas y el amanecer, no hay intención aparente de evitar el accidente, el conductor viaja solo, el incidente ocurre en carreteras de alta velocidad y generalmente sus consecuencias son graves [9]. Al comparar tales características con las estadísticas en México se tiene la información mostrada en la Tabla I [10].

TABLA I. POSIBLE RELACIÓN ENTRE LA SOMNOLENTIA Y LOS ACCIDENTES DE TRÁNSITO EN ZONAS URBANAS Y SUBURBANAS DE MÉXICO

Característica típica	Datos de accidentes de tránsito en México
Mortalidad relacionada con hora del accidente	23.62% de las muertes debidas a este tipo de accidentes se ocasiona en los incidentes ocurridos entre las 14 y las 16 horas, mientras que el 21.5 % debido a los accidentes de entre las 0 y las 6 horas.
El vehículo sale de la carretera sin causa aparente	El 9.2 % presentan esta característica.
No hay intención aparente de evitar el accidente	El 12.8% de los choques son contra objetos estáticos.

El objetivo de este proyecto, es por lo tanto, desarrollar un algoritmo, el cual pueda ser implementado en un sistema embebido y sea capaz de detectar y procesar los indicadores visuales relacionados con la somnolencia, con la finalidad de detectarlos durante el manejo de tal manera que pueda emplearse como un ADAS de fácil implementación.

2. Parte técnica del artículo

Para cumplir el objetivo de este proyecto, se consideraron como referencia los pasos sugeridos en [11] y [12], así como las definiciones de los descriptores encontradas en [13], por lo que el procedimiento general a seguir es:

- Localización de rostro.
- Identificación de las principales características del rostro (ojos, boca, nariz, etc.).
- Detección de características visuales que impliquen fatiga o distracción.
- Procesamiento de los hallazgos.

A. Localización de rostro

Para la localización del rostro se tomó la decisión de entrenar un detector capaz de realizar dicha tarea. El primer paso para crear un detector es establecer el descriptor mediante el que se determinará la existencia o no de un rostro dentro de una imagen dada, entre tales descriptores se encuentran los descriptores locales normalizados basados en el método de Histograma de Gradientes Orientados (HOG) mediante los cuales se puede realizar la detección con un excelente resultado en comparación con cualquier descriptor desarrollado, incluyendo Wavelet [14], razón por la cual el entrenamiento del detector se basó en HOG, cuyo proceso se resume en la Fig. 1.

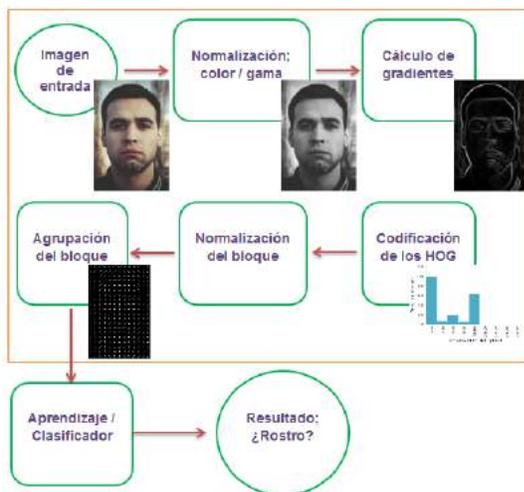


Fig 1. Descripción gráfica del procedimiento para el desarrollo del detector basado en HOG

A continuación se describen brevemente las etapas que integran al algoritmo para la obtención de descriptores HOG, ilustradas en la Fig. 1.

- 1) Normalización color/gamma (pre-procesamiento): Etapa considerada opcional [15], se puede ecualizar la gamma o el color de la imagen como medida para reducir la influencia de la iluminación [16] o bien, normalizar el tamaño de la misma. Esto último se considera de mayor utilidad debido a que la normalización de gamma o color no representa una mejora significativa en los resultados [17].
- 2) Cálculo de gradientes: El método más común para el cálculo de los gradientes es filtrar la imagen (I) aplicando una máscara derivativa discreta 1D de punto centrado, tanto vertical como horizontalmente. En este caso, las máscaras a aplicar son $D_x = \{-1, 0, 1\}$ para el eje x , y $D_y = \begin{Bmatrix} 1 \\ 0 \\ -1 \end{Bmatrix}$ pues aportan un mejor desempeño al proceso [16]. Al aplicar (1) y (2) se obtienen las derivadas espaciales en ambos ejes de I .

$$I_x = I * D_x \quad (1)$$

$$I_y = I * D_y \quad (2)$$

Para calcular la magnitud y la dirección del gradiente para cada uno de los píxeles de la región de interés (ROI), se aplican (3) y (4) respectivamente [18].

$$|G| = \sqrt{I_x^2 + I_y^2} \quad (3)$$

$$\theta = \arctan \frac{I_x}{I_y} \quad (4)$$

- 3) Codificación de los histogramas locales de orientación: La representación gráfica de $|G|$ se divide en celdas que se componen de cierta cantidad de píxeles adyacentes. Las celdas de 64 píxeles (la agrupación de $[8 \times 8]$ píxeles) son las que mejores resultados ofrecen [14]. A continuación se genera, para cada celda, un histograma basado en la orientación de los píxeles; dentro de la celda, cada píxel aporta un valor ponderado para un canal del histograma de acuerdo con los valores obtenidos durante el cálculo del gradiente. Los canales del histograma pueden definirse dentro de los intervalos $[0^\circ - 360^\circ]$ o $[0^\circ - 180^\circ]$ dependiendo de si se considera su signo o no. Es importante aclarar que el resultado obtenido tras aplicar (4) se expresa en radianes, por lo que habrá de usarse (5) para convertirlo a grados.

$$\alpha = \theta \cdot \left(\frac{180}{\pi}\right) \quad (5)$$

Para mejores resultados se sugiere generar el histograma con 9 canales en el intervalo de $[0^\circ - 180^\circ]$, mientras que el peso ponderado que cada píxel aportará al canal correspondiente a su orientación puede ser la magnitud de su gradiente calculado [14,19].

- 4) Normalización del bloque: Esta normalización se realiza para reducir la posibilidad de afectaciones debidas a la

iluminación y contraste. Las celdas se agrupan en grandes bloques espacialmente conectados que comparten algunas celdas entre sí de tal manera que cada celda contribuya más de una vez en el descriptor final, proporcionando un mejor desempeño. Para ello se puede aplicar alguna de las siguientes expresiones matemáticas [16]:

$$L2 - norm: f = \frac{v}{\sqrt{\|v\|_2^2 + e^2}} \quad (6)$$

$$L1 - norm: f = \frac{v}{\|v\|_1 + e} \quad (7)$$

$$L1 - sqrt: f = \frac{v}{\sqrt{\|v\|_1 + e}} \quad (8)$$

Para mejores resultados se sugiere usar bloques de 4 celdas, es decir un arreglo de [2 x 2] celdas, en combinación de celdas de 64 pixeles [16].

- 5) Agrupación de los descriptores: Finalmente, los bloques creados deben concatenarse para que el descriptor final agrupe la información completa sobre la ventana de detección. La información resultante puede emplearse en el clasificador, ver Fig.2 para ejemplo.

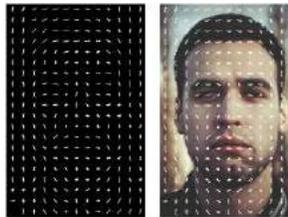


Fig 2. Ejemplo de Descriptores HOG; a la derecha los descriptores HOG sobre la imagen inicial, a la izquierda únicamente los descriptores.

Una vez que los descriptores se han definido, se puede aplicar el concepto de ‘*machine learning*’ enfocado a detección visual de objetos, descrito por Viola y Jones en el 2001, conocido como clasificadores en cascada [20]. Para ello es necesario establecer un conjunto de imágenes de entrenamiento, este deberá estar dividido en dos subgrupos; imágenes positivas e imágenes negativas, las primeras poseen el objeto de interés (caras, en este caso), mientras que las segundas no. Los clasificadores en cascada reducen considerablemente el tiempo de procesamiento y se basa en combinar clasificadores cuya complejidad incrementa en cada etapa, por ejemplo; si una sub-ventana de la imagen de entrada es clasificada como positiva (T) tras el primer clasificador, que es también el más simple, se inicia una reacción en cadena para que tal sub-ventana sea procesada por el resto de los clasificadores, hasta el más complejo. Si, en cualquier punto de la ‘cascada’, la sub-ventana es clasificada como negativa (F) esta es rechazada inmediatamente, evitando que siga procesándose por los clasificadores subsecuentes. De esta forma, los clasificadores iniciales sirven para descartar el fondo de la imagen, dejando así,

reducida la ROI. Este método se describe gráficamente en la Fig. 3.

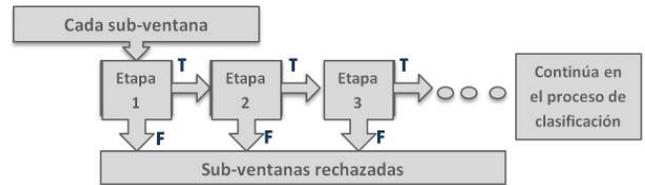


Fig 3. Representación gráfica del sistema de detección por cascada propuesto por Viola y Jones, 2001 [20].

Combinando los conceptos de los descriptores HOG y de clasificadores en cascada, el detector de rostro podría describirse esquemáticamente mediante la Fig. 4.

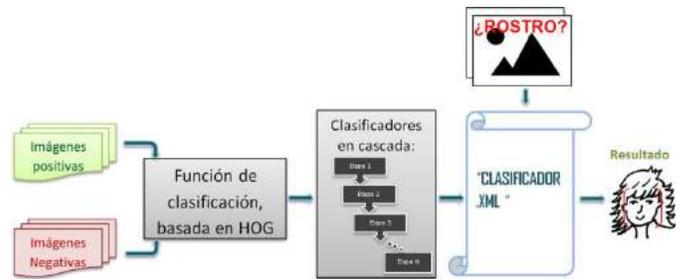


Fig 4. Descripción gráfica del detector facial sugerido

Para este proyecto se eligió un grupo de poco más de 19 000 imágenes positivas, tomando en consideración lo siguiente: los rostros en las imágenes deben poseer una resolución similar, las imágenes deben presentar variaciones de iluminación, la imágenes que conforman este grupo incluyen solo fotografías de rostros reales y no representaciones gráficas en 3D de rostros, finalmente, debía variar la posición de los rostros considerando dos grados de libertad, mostrados en Fig. 5, cuidando siempre que a pesar de la posición del rostro se mantengan visibles ambos ojos, nariz y boca. Estas imágenes fueron seleccionadas de al menos 5 bases de datos especializadas; The ORL Database of Faces, JAFFE Database, FEI Face Database, the caltech faces y 10k US Adult Faces Database, además de algunas imágenes de libre acceso [21,22, 23, 24, 25].

Para las imágenes negativas se seleccionaron imágenes de acceso libre, de mayor tamaño que las positivas, se almacenaron en un directorio dedicado a ellas, se les dio un nombre consecutivo mismo que fue enlistado en un archivo *.txt necesario para el entrenamiento del detector de rostro. Este grupo se conformó por poco más de 66 500 imágenes.

Para el entrenamiento del clasificador se empleó la función `trainCascadeObjectDetector()`, disponible en MATLAB, algunos de los parámetros que se consideran en esta función

son; la información de imágenes positivas y negativas, la cantidad de etapas para la cascada, la tolerancia a falsos positivos, el tipo de descriptor a usar, el nombre del archivo de salida, entre otros.

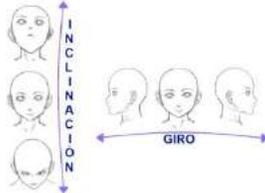


Fig 5. Grados de libertad considerados en la posición del rostro.

Tras entrenar un clasificador se realizaron pruebas para definir sus debilidades de tal forma que pudieran mejorarse en el entrenamiento de uno nuevo; tales pruebas se realizaron inicialmente con imágenes tomadas con una cámara de 8 MP, al final, mediante la aplicación de la técnica de clasificación en cascada se obtuvieron 11 clasificadores, descritos en archivos “*.xml” (ver Fig. 4), basados en descriptores HOG; utilizando para ellos bloques de 4 celdas de 68 píxeles cada una. Dichos clasificadores se compararon entre sí y contra el clasificador proporcionado por OpenCV, el cual se basa en el método de Viola-Jones [20]. Para tal comparación se utilizaron de imágenes pertenecientes a una base de datos especializada [26], en la que las imágenes incluyen en su descripción los ángulos de giro e inclinación en que se encuentran los rostros; esta base de datos se conforma por 2 511 imágenes en 93 diferentes posiciones, sin embargo se seleccionaron solo las imágenes en las que se muestra un rostro en cualquiera de las 14 posiciones que permitieran observar ambos ojos, nariz y boca, dichas posiciones se listan a continuación (inclinación/desviación):

- -15 / -15
- -15 / 0
- -15 / 15
- 0 / -15
- 0 / 0
- 0 / 15
- 0 / 30
- 15 / -15
- 15 / 0
- 15 / 15
- 15 / 30
- 30 / -15
- 30 / 0
- 30 / 15

En la expresión de la posición: El signo negativo para la inclinación indica que la cara del sujeto se encuentra inclinada hacia arriba, mientras que un ángulo positivo indicaría inclinación hacia abajo; en cuanto al giro, un signo negativo indica que el rostro se ha girado hacia la izquierda, mientras que el signo contrario se ha girado también en dirección opuesta. La posición neutral es considerada como (0 / 0), y es cuando el sujeto mira de frente a la cámara sin girar.

Tras comparar la eficiencia relativa de cada clasificador, se tomó la decisión de combinar los 3 de mejor desempeño, de tal forma que se mejorara el desempeño del sistema. El

detector combinado mostró una gran eficiencia en la detección de rostros cuya posición se hallara dentro de las listadas anteriormente, los resultados se muestran en la Tabla II. Se encontró, además, que se tiene una respuesta bastante favorable en otras 6 posiciones (ver Tabla III), mientras que se obtuvo una eficiencia relativa en la detección de rostros de entre el 85.2% al 77.8 % para otras 7 posiciones.

TABLA II. PORCENTAJE DE EFICIENCIA DEL DETECTOR COMBINADO, DE ACUERDO CON LA POSICIÓN DE LOS ROSTROS

Posición de rostro	Porcentaje de detección	Posición de rostro	Porcentaje de detección
-15 / -15	92.6%	15 / -15	96.3%
-15 / 0	100.0%	15 / 0	96.3%
-15 / 15	92.6%	15 / 15	96.3%
0 / -15	96.3%	15 / 30	96.3%
0 / 0	100.0%	30 / -15	96.3%
0 / 15	100.0%	30 / 0	96.3%
0 / 30	96.3%	30 / 15	85.2%

TABLA III. POSICIONES ADICIONALES EN LAS QUE EL DETECTOR COMBINADO TRABAJA CON UNA ALTA EFICIENCIA

Posición de rostro	Porcentaje de detección	Posición de rostro	Porcentaje de detección
-15 / 30	92.6%	30 / -30	92.6%
0 / 30	96.3%	60 / -15	92.6%
15 / -30	96.3%	60 / 0	92.6%

B. Detección de apertura de los ojos

Para esta parte del proyecto se eligió utilizar un detector de puntos de referencia (*landmarks*) que, aplicado al rostro pudiera ser de utilidad para detectar ojos, nariz, boca y cejas. Este detector es parte de la implementación hecha en [27]. Con tal detector se obtiene un vector de 6 coordenadas que indican puntos importantes de cada ojo, tal como se puede ver en la Fig. 6.



Fig 6. Puntos de referencia generados sobre los ojos con el detector

A partir de los puntos expuestos en la Fig. 6 se puede calcular la relación de aspecto ocular (EAR) entre la altura y el ancho del ojo utilizando (9), de acuerdo con [28].

$$EAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|} \quad (9)$$

La EAR es muy constante cuando el ojo se encuentra abierto, por lo que es fácil determinar el punto de comparación con el que se determine el estado del ojo, dentro del código creado en Python se ha decidido, además, calcular el promedio de las EAR de ambos ojos para garantizar un mejor resultado en el cálculo.

Aun cuando existen trabajos en los que se determina el estatus de la alarma dependiendo del número de cuadros de video en que los ojos son detectados como cerrados, en esta ocasión la activación de la alarma dependerá del tiempo en que los ojos sean detectados como cerrados. En la Fig. 7 se muestra un ejemplo del funcionamiento del sistema, en el ejercicio mostrado, el sujeto ha cerrado los ojos por al menos 0.6 segundos, además del letrero de alerta, se activa por conveniencia una alarma sonora.



Fig 7. Ejemplificación de activación de alerta por detección constante de cerrado de ojos.

C. Bostezo

El proceso para la detección del bostezo es muy similar al aplicado para calcular la EAR, por lo que el primer paso en obtener los puntos que describen a la boca del sujeto, mediante el detector seleccionado, en este caso el desarrollado en [27]. Con esto se obtienen 12 puntos de referencia (ver Fig. 8 lado izquierdo), sin embargo, tal como se vio en la subsección B, bastan 6 puntos para determinar la relación de apertura; por ello se desechó la mitad de los puntos obtenidos y se utilizan para el cálculo de apertura de la boca aquellos que se muestran a la derecha de la Fig. 8.



Fig 8. A la Izquierda los puntos proyectados sobre la boca por el detector de puntos relevantes, a la derecha los puntos seleccionados para el cálculo de apertura bucal.

Considerando los 6 puntos de la segunda imagen de la Fig. 8 se puede aplicar (6), solo que ahora la condición de activación de la alarma será mediante una apertura máxima de la boca, puesto que es una característica del bostezo.

El bostezo puede indicar presencia de sueño o fatiga, pero indica también que el individuo se encuentra aún despierto, por ello se decidió implementar alarmas visual y sonora diferentes a las que se activan al detectar ojos cerrados por un tiempo prolongado más allá de un parpadeo.

En la Fig. 9 se muestra un ejemplo de la activación de alarma por bostezo, en la imagen se ha activado la alarma luego de que el sujeto mantuviera la boca abierta por poco más de 0.6 segundos.



Fig 9. Ejemplo de activación de alarma por bostezo

D. Consideraciones experimentales

Las pruebas realizadas con el algoritmo implementado, de las cuales forman parte las imágenes mostradas en las Fig. 7 y Fig. 9, se llevaron a cabo utilizando una computadora portátil con 12 GB de memoria RAM, procesador AMD A10 2.00GHz y una cámara de 5MP; en ellas participaron sujetos de diferentes complejiones y características faciales, con lo que se probó la importancia de hacer del algoritmo un sistema más adaptable, puesto que factores como la altura del sujeto de interés, combinado con el ángulo de posicionamiento de la cámara afectan el cálculo de la EAR al probar el algoritmo con rostros en posiciones alejadas del (0 / 0). Se desarrollaron también pruebas para conocer la velocidad de procesamiento por cada cuadro de video, a 5 resoluciones diferentes, tanto en la computadora portátil antes mencionada como en dos CPU's más, los resultados se resumen en la Tabla IV.

TABLA IV. TIEMPO DE PROCESAMIENTO POR CUADRO DE VIDEO PARA 3 DIFERENTES UNIDADES DE PROCESAMIENTO

Resolución de video (píxeles)	Tiempo promedio de procesamiento por cuadro de video (segundos)		
	Computadora portátil	Raspberry Pi 2 B	Raspberry Pi 3 B
1024 x 768	0.531	7.47	3.51
720 x 576	0.252	3.17	1.32
640 x 480	0.212	2.51	1.05
320 x 240	0.075	0.503	0.179
160 x 120	0.036	0.167	0.058

E. Acciones futuras

Se trabaja en la modificación del algoritmo de tal manera que se reduzca el impacto de algunos factores tales como la complejión del usuario, el ángulo de instalación de la cámara y la iluminación; en la Fig. 10 se muestran algunos de los resultados obtenidos en entornos de baja iluminación. Adicionalmente, se considera la posibilidad de integrar el análisis de un mayor número de factores visuales para la detección de somnolencia.



Fig 10. Ejemplo de detección a baja iluminación.

3. Conclusiones

En este trabajo se muestra que mediante la aplicación de diferentes conceptos y técnicas de visión por computadora es posible desarrollar un sistema que pueda ayudar a la prevención de accidentes de tránsito debidos a la somnolencia sin ser necesario adquirir automóviles de gama alta.

Se demuestra que la combinación de técnicas enriquece el resultado del ejercicio, dando como resultado un mejor rendimiento del algoritmo. Y finalmente, se comprueba que, al tratarse de un sistema que basa su funcionamiento en la detección apropiada del rostro, usar los descriptores HOG resulta en un porcentaje elevado de precisión.

4. Referencias

- [1] B. Lewis, "A look at the history of the automobile offers a glimpse of the PC's future", *InfoWorld*, p. 76, Dec. 1999.
- [2] A. Belbachir, J. C. Smal, J. M. Blossville and D. Gruyer, "Simulation-Driven Validation of Advanced Driving-Assistance Systems", *Procedia - Social and Behavioral Sciences*, vol. 48, pp. 1205-1214, 2012.
- [3] M. Auricht and R. Stark, "How to Consider Emotional Reactions of the Driver within the Development of Advanced Driver Assistance Systems (ADAS)?", *Procedia CIRP* 21, pp. 70-74, 2014.
- [4] F. Biondi, D. L. Strayer, R. Rossi, M. Gastaldi and C. Mulatti, "Advanced driver assistance systems: Using multimodal redundant warnings to enhance road safety", *Applied Ergonomics*, vol. 58, pp. 238-244, 2017.
- [5] P. Planing, *Innovation Acceptance; The case of Advanced Driver-Assistance Systems*, Springer, 2014.
- [6] Organización Mundial de la Salud, *Informe Sobre La Situación Mundial De La Seguridad Vial 2015*, 2015.
- [7] STCONAPRA, *Informe sobre la situación de la seguridad vial, México 2015*, 1ra. Ed., Secretaría de salud, 2016.
- [8] S. G. Klauer, T. A. Dingus, V. L. Neale, J. D. Sudweeks and D. J. Ramsey, *The Impact of Driver Inattention On Near-Crash / Crash Risk : An Analysis Using the 100-Car Naturalistic Driving Study Data*, National Technical Information Service, Apr. 2006.
- [9] K. P. Strohl, J. Blatt, F. Council, K. Georges, J. Kiley, R. Kurrus, A. T. McCartt, D. Willis, P. Waller, J. Stutts, T. Roth, S. Rogus, A. I. Pack and S. L. Merritt, *Drowsy Driving and Automobile Crashes; Report and Recommendations*, 1998.
- [10] Instituto Nacional de Estadística y Geografía, *Estadísticas a propósito del día mundial en recuerdo de las víctimas de los accidentes de tráfico (tercer domingo del mes de noviembre); datos nacionales*, Nov. 2016.
- [11] G.L. Masala, E. Grosso, "Real time detection of driver attention: Emerging solutions based on robust iconic classifiers and dictionary of poses", *Transportation Research Part C: Emerging Technologies*, vol. 49, pp. 32-42, Nov. 2014.
- [12] D. Mihai, G. Florin and M. Gheorghe, "Using Dual Camera Smartphones As Advanced Driver Assistance Systems: NAVIEYES System Architecture", *Proceedings of the 8th ACM International Conference on Pervasive Technologies Related to Assistive Environments*, pp. 23:1--23:8, 2015.
- [13] R. Senaratne, D. Hardy, B. Vanderaa and S. Halgamuge, "Driver Fatigue Detection by Fusing Multiple Cues", *Advances in Neural Networks -- ISNN 2007: 4th International Symposium on Neural Networks*, ISSN 2007, Nanjing, China, June 3-7, 2007, *Proceedings*, Part II, pp. 801-809, 2007.
- [14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 886-893, Jun. 2005.
- [15] A. Del Abril De Mur, *Detección de personas en secuencias de imágenes en entornos interiores (Tesis de Ingeniería)*, Universidad de Alcalá; Escuela politécnica superior, 2017.
- [16] N. Dalal, *Finding People in Images and Videos (Tesis doctoral)*, Institut National Polytechnique de Grenoble, 2006.
- [17] S. Mallick, "Histogram of Oriented Gradients | Learn OpenCV", 2017, <http://www.learnopencv.com/histogram-of-oriented-gradients/>, consultado en Julio 2017.
- [18] N. Morán Cruz, "Desarrollo de un sistema avanzado de asistencia a la conducción en tiempo real para la detección de peatones en entornos urbanos complejos" (Tesis Ingeniería), Universidad Carlos III de Madrid. 2013.
- [19] Universitatea Tehnica din Cluj-Napoca, "Pattern recognition systems – Lab 5 Histograms of Oriented Gradients" (Laboratory practice).
- [20] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, pp. I-511-I-518, 2001.
- [21] AT&T Laboratories Cambridge, *The ORL Database of Faces*, 2002.
- [22] M. Lyons, S. Akamatsu, M. Kamachi and J. Gyoba, "Coding facial expressions with Gabor wavelets", *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 200-205, 1998.
- [23] C. Thomaz and G. Giraldi, "A new ranking method for principal components analysis and its application to face image analysis", *Image and Vision Computing*, vol. 29, pp. 902-913, 2010.
- [24] M. Weber, *Frontal face dataset*, California Institute of Technology.
- [25] W. A. Bainbridge, P. Isola and A. Oliva, "The Instruct Memorability of Face Photographs", *Journal of Experimental Psychology*, vol. 142, pp. 1323-1334, 2013.
- [26] Gourier, N., Hall, D., & Crowley, J. L. (2004). Estimating face orientation from robust detection of salient facial structures. *FG Net Workshop on Visual Observation of Deictic Gestures (POINTING)*, 17–25. <https://doi.org/citeulike-article-id:1595218>
- [27] V. Kazemi, J. Sullivan, "One Millisecond Face Alignment with an Ensemble of Regression Trees", *CVPR '14 Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1867-1874, Jun. 2014.
- [28] T. Soukupova and J. Cech, "Real-Time Eye Blink Detection using Facial Landmarks", *21st Computer Vision Winter Workshop*, 2016.

5. Biografías

A. Cinthia Viviana Alcántara Montiel

Se tituló como Técnico Superior Universitario en Electrónica y Automatización en la Universidad Tecnológica de Tula-Tepeji, en la misma institución llevo a cabo sus estudios de Licenciatura en Ingeniería Mecatrónica. Actualmente se encuentra matriculada en el programa ofrecido por la Universidad Autónoma de Querétaro para obtener el grado de M. en C. con especialidad en Instrumentación y control. Ejerce, además, su profesión como metrologo en el Centro Nacional de Metrología.

B. Jesús Carlos Pedraza Ortega

El Doctor Jesús Carlos Pedraza Ortega realizó sus estudios de Licenciatura, con especialidad en Electrónica en el Instituto Tecnológico de Celaya, sus estudios de Maestría en la Facultad de Ingeniería Mecánica, Eléctrica y Electrónica (FIMEE – Universidad de Guanajuato) con especialidad en Instrumentación y Sistemas Digitales donde trabajó con sistemas embebidos y optoelectrónica. El Doctorado en Ingeniería Mecánica con especialidad en Robótica - Sistemas de Reconstrucción 3D lo realizó en la University of Tsukuba en Japón, donde trabajó con técnicas de visión pasiva. Como docente, ha impartido diferentes cursos en 3 niveles de estudios (Licenciatura, Maestría y Doctorado) desde 1997 en varias instituciones como; la FIMEE de la Universidad de Guanajuato, entre otras, y actualmente en la Universidad Autónoma de Querétaro.

ACRÓNIMOS

AAM	Active Appearance Model
ABS	Antilock Brake System
ACC	Adaptive Cruise Control
ADAS	Advanced Driver Assistance Systems
AECS	Average Eye Closure Speed
BLIS	Blind Spot Information System
DLS	Dynamic Light System
EAR	Relación de Aspecto Ocular
EBD	Electronic Brake-force Distribution
ECG	Electrocardiograma
ECU	Electronic Control Unit
EEG	Electroencefalograma
EOG	Electrooculograma
ESS	Epworth Sleepiness Scale
GIP	Gaze Intersection Point
GPS	Global Positioning System
HF	High Frequency, (se refiere al límite superior de la HRV)
HOG	Histogram of Oriented Gradient
HRV	Heart Rate Variability
KSS	The Karolinska Sleepiness Scale
LBP	Local Binary Pattern
LF	Low Frequency, (referida al límite inferior de la HRV)
NHTSA	National Highway Traffic Safety Administration
NN	Nearest Neighbor
NREM	Nonrapid Eye Movement
MAR	Relación de Aspecto Bucal
MB	Megabyte

MP Megapíxeles
MSLT Multiple Sleep Latency Test
MWT Maintenance of Wakefulness Test
PCA Principle Component Analysis
PDMs Point distribution models
PERCLOS Percentage of eyelid Closure
PERLVO Percentage of Mouth Large Vertical Open
PERTITL Percentage of Head Tilt
POI Point Of Interest
SIFT Scale-invariant Feature Transform
SSS Stanford Sleepiness Scale
SURF Speeded-Up Robust Features Descriptor
SVM Support Vector Machine
REM Rapid Eye Movement
RGB Red,Green,Blue
ROI Region Of Interest
VAS Visual Analogue Scale

NOMENCLATURA

α	Dirección del gradiente, expresada en grados
B	Valor original del componente azul en una imagen
b	Valor normalizado del azul en una imagen
D_x	Kernel o máscara para filtrado de I en el eje x
D_y	Kernel o máscara para filtrado de I en el eje y
e	Constante de valor mínimo (muy cercano a cero)
E_e	Energía externa, provoca que la serpiente tome la forma de las características cercanas
E_i	Energía interna, define la evolución natural de la serpiente
E_s	Energía de la serpiente, función que define su evolución
$ G $	Magnitud del gradiente
G	Valor original del componente verde en una imagen
g	Valor normalizado del verde en una imagen
I	Imagen inicial o de entrada
I_x	Derivada espacial de I en el eje x
I_y	Derivada espacial de I en el eje y
M	Plantilla descriptora de un objeto específico
N	Cantidad de pixeles vecinos, espaciados equitativamente en un circulo de radio Rd
n	Número de pixeles en la plantilla de un objeto
n_c	Nivel de gris del pixel central o de referencia
n_i	Nivel de gris de cada pixel vecino del pixel central o de referencia
P	Matriz de los eigenvectores de la covariancia de desviaciones, en PDMs
R	Valor original del componente rojo en una imagen
r	Valor normalizado del rojo en una imagen
Rd	Radio del circulo
θ	Dirección del gradiente, expresada en radianes
u	Vector de pesos, en PDMs

v	Vector no normalizado de los histogramas por bloque
$\ v_k\ $	Vector k -normalizado de los histogramas por bloque, con $k = 1,2$
x	Variable que representa un punto en los PDMs
X	Conjunto de vectores de entrenamiento en el método de PDMs
\bar{x}	Media de la característica en el conjunto de entrenamiento para el punto x , en PDMs