



Análisis de algoritmos de procesamiento de imágenes para el reconocimiento facial de fatiga y somnolencia en conductores

Por: Michael Rendón Villa y Johan Sebastian Vargas Castro

Facultad de Ingenierías Ingeniería de Sistemas y Computación

Noviembre del 2020

TABLA DE CONTENIDO

DEDICATORIAS Y AGRADECIMIENTOS.....	4
RESUMEN	5
ABSTRACT.....	6
PARTE I INTRODUCCIÓN A LA INVESTIGACIÓN.....	8
CAPITULO 1 INTRODUCCIÓN	9
<i>1.1. PLANTEAMIENTO Y JUSTIFICACIÓN DEL PROBLEMA</i>	<i>10</i>
1.1.1. FORMULACIÓN DEL PROBLEMA.....	10
1.1.2. JUSTIFICACIÓN	11
<i>1.2. HIPOTESIS Y OBJETIVOS.....</i>	<i>12</i>
1.2.1. OBJETIVO GENERAL.....	13
1.2.2. OBJETIVOS ESPECÍFICOS	13
<i>1.3. METODOLOGIA PROPUESTA PARA EL DESARROLLO DE LA MONOGRAFIA</i>	<i>13</i>
<i>1.4. ESTRUCTURA DE LA MONOGRAFIA.....</i>	<i>14</i>
PARTE II ESTADO DEL ARTE.....	16
<i>CAPITULO 2 FATIGA Y SOMNOLENCIA</i>	<i>17</i>
2.1. MARCO HISTORICO	18
2.2. MARCO REFERENCIAL.....	20
PARTE III DESARROLLO DE LA MONOGRAFÍA.....	24
<i>CAPITULO 3 FACTORES Y CONSECUENCIAS DE LA FATIGA.....</i>	<i>25</i>
<i>CAPITULO 4 SISTEMAS DE RECONOCIMIENTO FACIAL</i>	<i>27</i>
4.1. SISTEMA DE DETECCIÓN DE MICROSUEÑO MEDIANTE EL USO DE LA TÉCNICA EAR (EYE ASPECT RATIO ALGORITHM)	28
4.2. SISTEMA DE DETECCIÓN DE SOMNOLENCIA PARA LA PERFECCIÓN DE LA INTERFAZ CEREBRO-COMPUTADORA MEDIANTE EL ALGORITMO DE VIOLA-JONES.....	37
<i>CAPITULO 5 ALGORITMOS PARA EL RECONOCIMIENTO FACIAL.....</i>	<i>46</i>
5.1. VIOLA-JONES.....	48
5.2. ALGORITMO EAR (EYE ASPECT RATIO ALGORITHM) CON DETECCIÓN DE PARPADEO EN TIEMPO REAL MEDIANTE MARCAS FACIALES.	56
PARTE IV CONCLUSIONES Y BIBLIOGRAFIA.....	71
<i>CAPITULO 6: CONCLUSIONES.....</i>	<i>72</i>
<i>BIBLIOGRAFIA.....</i>	<i>75</i>

TABLA DE ILUSTRACIONES

<i>Ilustración 1: Sistema de detección de microsueño</i>	29
<i>Ilustración 2: Secuencia de detección ocular para detección de microsueños 2</i>	30
<i>Ilustración 3: Puntos de referencia faciales (landmarks)</i>	31
<i>Ilustración 4: Hito de seis puntos para la detección de ojos</i>	32
<i>Ilustración 5: (a) Apertura de ojos y (b) cierre basado en el umbral EAR <3.0</i>	33
<i>Ilustración 6: Efecto de la oscuridad circundante en la detección del sistema (a) menos oscuridad (b) completamente oscuro</i>	34
<i>Ilustración 7: Efecto de gafas de sol</i>	34
<i>Ilustración 8: Efecto de anteojos en el sistema de detección</i>	35
<i>Ilustración 9: Resultados EAR</i>	36
<i>Ilustración 10: Representación típica de la señal SSVEP en el dominio de frecuencia (Hz)</i>	38
<i>Ilustración 11: Diagrama de bloques generalizado para representación HCI o BCI</i>	39
<i>Ilustración 12: Representación de conceptos para la detección del nivel de somnolencia</i>	39
<i>Ilustración 13: Monitorización del sujeto para detección ocular</i>	40
<i>Ilustración 14: Representación del trabajo global propuesto</i>	41
<i>Ilustración 15: Diagrama de flujo del trabajo propuesto</i>	42
<i>Ilustración 16: Ojo detectado usando el algoritmo de Viola-jones a) estado de ojo cerrado b) estado de ojo abierto</i>	43
<i>Ilustración 17: Nivel de somnolencia debido a varios estados oculares</i>	43
<i>Ilustración 18: Nivel de relajación en la banda de señal del EEG (Alfa y Beta) debido al nivel de somnolencia de los intérpretes BCI</i>	44
<i>Ilustración 19: Diagrama bloques</i>	48
<i>Ilustración 20: Imagen integral</i>	49
<i>Ilustración 21: Ejemplo de una sumatoria de imagen integral</i>	50
<i>Ilustración 22: Características rectangulares</i>	50
<i>Ilustración 23: Representación haar (filtro)</i>	52
<i>Ilustración 24: (a) XH, (b) XIVS, (c) XFSDD, (d) XENM, (e) XSEG</i>	53
<i>Ilustración 25: Resultados experimentales de Viola-Jones original</i>	54
<i>Ilustración 26: Experimentos de uso del método Viola-Jones basado en características compuestas en un conjunto de datos FDDB</i>	55
<i>Ilustración 27: Comparación completa viola-jones</i>	55
<i>Ilustración 28: Ejemplo de parpadeos detectados. Las gráficas de la relación de aspecto del ojo (EAR) en Ecuación 3, resultados del umbral de EAR (umbral establecido en 0,2), los parpadeos detectados por EAR SVM y las etiquetas sobre la secuencia de vídeo (verdad fundamental o Ground-truth). Imagen de entrada con puntos de referencia detectados (el frame está representado con una línea roja).</i>	59

<i>Ilustración 29: Imágenes de ejemplo del conjunto de datos 300-VW con puntos de referencia obtenidos por Chehra (A. Asthana, S. Zafeoriou, S. Cheng, & M. Pantic, 2014) e Intrafase (X. Xiong & F. De la Torre, 2013). Imágenes originales (izquierda) con distancia inter-ocular (IOD) igual a 63 (arriba) y 53 (abajo) píxeles. Imágenes sub-muestreadas (derecha) a IOD igual a 6,3 (arriba) y 17 (abajo)</i>	61
<i>Ilustración 30: Histograma acumulativo del error de localización promedio de los 49 puntos de referencia (arriba) y los 12 puntos de referencia de los ojos (abajo). Los histogramas son calculados para imágenes de resolución original (líneas continuas) y un subconjunto de imágenes pequeñas ($IOD \leq 50px$)</i>	63
<i>Ilustración 31: Precisión de localización de puntos de referencia en función de la resolución de la imagen de la cara calculada para todos los puntos de referencia y puntos de referencia del ojo únicamente</i>	64
<i>Ilustración 32: Precisión del radio de apertura del ojo en función de la resolución de la imagen de la cara. Arriba: para imágenes con un radio verdadero bajo (en su mayoría ojos cerrados o medio cerrados) e inferior: imágenes con una proporción más alta (ojos abiertos)</i>	65
<i>Ilustración 33: Ejemplo de parpadeos detectados donde falla el umbral de EAR mientras que EAR SVM tiene éxito. Las gráficas de la relación de aspecto del ojo EAR en Ecuación 3, resultados del umbral de EAR (umbral establecido en 0,2), los parpadeos detectados por EAR SVM y las etiquetas de verdad sobre el terreno sobre la secuencia de vídeo. Imagen de entrada con puntos de referencia detectados. La representación del frame está marcada con una línea roja)</i>	67
<i>Ilustración 34: Curvas de recuperación de precisión de los clasificadores de umbral EAR y EAR SVM medidos en (a) la ZJU y (b) las bases de datos Eyeblink8. Resultados publicados de los métodos A - Drutarovsky y Fogelton (T. Drutarovsky & A. Fogelton, 2014), B - Lee y col. (W. H. Lee, E. C. Lee, & K. E. Park, 2010), C - Danisman y col. (T. Danisman, I. Bilasco, C. Djeraba, & N. Ihaddadene, 2010)</i>	68
<i>Ilustración 35: Precisión del detector de parpadeo del ojo (medida por AUC) en función de la resolución de la imagen (IOD promedio) al sub-muestrear el conjunto de datos ZJU</i>	69

DEDICATORIAS Y AGRADECIMIENTOS

1. A Dios gracias por ayudarnos en el día a día para el cumplimiento de este arduo trabajo.
2. A nuestro director Saulo de Jesús Rengifo Torres por la disponibilidad y acompañamiento en el desarrollo de esta monografía.
3. Y a nuestras familias y allegados que aportaron su granito de arena para motivarnos y ver cumplidas nuestras metas.

RESUMEN

Una vez identificada la cantidad de accidentes provocados por la presencia de síntomas como la somnolencia en conductores (Se calcula que el 2,5% de los accidentes fatales y el 2% de los accidentes no fatales se relacionan con la presencia de somnolencia (Sueño, 2016, pág. 21)), quienes requieren ser mitigados para disminuir la tasa de mortandad en estos siniestros. En la parte I, capítulo 1, se formula el problema y se justifica el diseño de una “Metodología para el desarrollo de la monografía”, donde se plantean interrogantes, como:

1. ¿Cómo afecta la fatiga y la somnolencia a los conductores mientras conducen?
2. ¿Existen sistemas informáticos que permitan el reconocimiento facial de fatiga y somnolencia en conductores?
3. ¿Qué algoritmos se implementan en los sistemas informáticos que reconocen gestos faciales de fatiga y somnolencia?
4. ¿Qué tan eficaces son los algoritmos al momento del procesamiento de imágenes para el reconocimiento de fatiga y somnolencia?

Donde las respuestas a dichas preguntas, con la comprobación de las hipótesis relacionadas se resuelven a lo largo de 6 capítulos en esta monografía.

ABSTRACT

Once the number of accidents caused by the presence of symptoms such as drowsiness in drivers has been identified (It is estimated that 2.5% of fatal accidents and 2% of non-fatal accidents are related to the presence of drowsiness (Sleep, 2016, page 21)), those who need to be mitigated to reduce the death rate in these claims. In part I, chapter 1, the problem is formulated and the design of a "Methodology for the development of the monograph" is justified, where questions are raised, such as:

1. How does fatigue and drowsiness affect drivers while driving?
2. Are there computer systems that allow facial recognition of fatigue and drowsiness in drivers?
3. Which algorithms are implemented in computer systems that recognize facial gestures of fatigue and drowsiness?
4. How effective are the algorithms at the time of image processing for the recognition of fatigue and drowsiness?

Where the answers to these questions, with the verification of the related hypotheses are resolved throughout 6 chapters in this monograph.

PARTE I INTRODUCCIÓN A LA INVESTIGACIÓN

CAPITULO 1 INTRODUCCIÓN

En este capítulo se plantea el problema y se justifica la investigación acerca de algoritmos de procesamiento de imágenes para el reconocimiento de fatiga y somnolencia en conductores, teniendo como base el gran porcentaje de accidentes ocasionados por estos dos factores y la falta de concientización por parte de los conductores que no respetan los límites humanos.

1.1. PLANTEAMIENTO Y JUSTIFICACIÓN DEL PROBLEMA

1.1.1. FORMULACIÓN DEL PROBLEMA

Según la Sociedad Española de Sueño, 1 de cada 5 accidentes de tráfico se genera por causa del sueño y la fatiga, posicionándose junto con el consumo de alcohol y alucinógenos como las principales causas de siniestros en las carreteras con víctimas.

El sueño es una de las actividades más imprescindibles del ser humano, ya que este dedica prácticamente un tercio de su vida en ella para lograr el correcto funcionamiento tanto físico como psicológico. A pesar de la recomendación por parte de las organizaciones de la salud de dedicar 8 horas diarias a descansar, la encuesta ‘¿Cómo duermen los españoles?’ realizada por IPSOS para Philips en colaboración con la Sociedad Española de Sueño (SES) con motivo del Día Mundial del Sueño, pone en manifiesto que

“Más de la mitad de los españoles no duerme bien y utiliza el fin de semana para recuperar las horas de sueño perdidas. El porcentaje de españoles que duerme mal (58%) superó al de los que duerme bien (42%), destacando que un 13% de los encuestados no descansa “nada bien”. Con este estudio se pretende demostrar que la calidad del sueño de los españoles no es óptima, ya que los españoles duermen una media de 6,8 horas entre semana, mientras que se usa el fin de semana para recuperar el sueño perdido. El insomnio es otro de los problemas fundamentales que saca a la luz el informe. El 75% de los españoles se despierta al menos una vez por la noche, y 3 de cada 10 afirman directamente que padecen insomnio.” (europapress.tv, 2019).

Además una cantidad alta de conductores que mezclan la privación del sueño con demasiadas horas frente al volante, esto produce un conflicto entre las necesidades fisiológicas y las actividades tanto sociales como profesionales de un conductor de automóvil.

Por consiguiente, esta investigación sobre algoritmos de procesamiento de imágenes para el reconocimiento de fatiga y somnolencia en conductores, con el fin de aproximarse a una prevención tecnológica oportuna de los accidentes, busca las respuestas a los siguientes interrogantes:

1. ¿Cómo afecta la fatiga y la somnolencia a los conductores mientras conducen?

2. ¿Existen sistemas informáticos que permitan el reconocimiento facial de fatiga y somnolencia en conductores?
3. ¿Qué algoritmos se implementan en los sistemas informáticos que reconocen gestos faciales de fatiga y somnolencia?
4. ¿Qué tan eficaces son los algoritmos al momento del procesamiento de imágenes para el reconocimiento de fatiga y somnolencia?

1.1.2. JUSTIFICACIÓN

Según el Observatorio Nacional de Seguridad Vial entre enero y octubre de 2019 murieron en nuestro país 52 personas en 119 accidentes, estos están relacionados con la pérdida de la conciencia y la realidad por un espacio de tiempo muy breve. (De La Hoz, 2019)

El sueño y la fatiga está catalogado entre las principales causas de accidentes viales y siniestros en carretera, por esta razón consideramos que la realización de esta investigación es importante para reconocer los factores y las consecuencias de la fatiga y la somnolencia en conductores, además de encaminar una posible prevención tecnológica para la reducción de accidentes de tránsito, como lo podrían ser sistemas que reconozcan a partir de algoritmos de procesamiento de imágenes el estado de fatiga y la somnolencia en un conductor.

1.2. HIPOTESIS Y OBJETIVOS

Con el fin de analizar algunos aspectos de salud sobre la fatiga y la somnolencia, y algoritmos que puedan reconocerlos, se estudiará ya analizará cada una de las siguientes hipótesis que se presenta a continuación:

1. La fatiga y la somnolencia se presentan en el 3% de los accidentes fatales y el 2.5% de los accidentes no fatales.
2. Existen sistemas informáticos que reconocen la fatiga y la somnolencia en conductores a partir de visión computacional.
3. Los algoritmos que procesan imágenes para el reconocimiento de fatiga y somnolencia, identifican correctamente patrones de fatiga y somnolencia en imágenes en escenarios con datos favorables (Luminosidad, Contraste, balance de blancos etc...), mientras que en escenarios reales tienen una tasa más baja de reconocimiento ya que influyen variables como el clima, el movimiento del conductor etc...

1.2.1. OBJETIVO GENERAL

Investigar y analizar algoritmos de reconocimiento de imágenes y video para reconocer gestos faciales de fatiga y somnolencia en conductores.

1.2.2. OBJETIVOS ESPECÍFICOS

1. Investigar factores faciales que influyan en la fatiga y somnolencia.
2. Determinar algoritmos que reconozcan la fatiga y somnolencia.
3. Analizar y comparar la eficiencia de los algoritmos en la detección de gestos faciales como fatiga y somnolencia.

1.3. METODOLOGIA PROPUESTA PARA EL DESARROLLO DE LA MONOGRAFIA

A continuación, se describen las fases requeridas para el buen proceso, convirtiéndose en la guía de desarrollo de la monografía.

- Búsqueda de factores faciales que influyan en la fatiga y somnolencia:
- Búsqueda de sistemas que reconozcan gestos de fatiga y somnolencia.
- Analizar de los algoritmos que usan los sistemas previamente buscados.
- Comparar cuales algoritmos o técnicas pueden cumplir mejor las expectativas de reconocimiento.

Fase I. Búsqueda de factores faciales que influyan en la fatiga y somnolencia.

Esta etapa es vital para el desarrollo de esta monografía, pues aquí se analizan los factores que influyen en la fatiga y somnolencia de las personas a la hora de conducir, y responder a las preguntas de ¿cómo ocurre?, ¿Cuánto tiempo dura? ¿Qué causa?, ¿Cómo se previene?, etc...

Fase II. Búsqueda de sistemas que reconozcan gestos de fatiga y somnolencia.

En esta etapa se buscarán sistemas informáticos ya implementados que reconozcan gestos faciales, específicamente aquellos que se enfaticen en la detección de factores de fatiga o somnolencia en personas.

Fase III. Analizar de los algoritmos que usan los sistemas previamente buscados.

En esta etapa se van a implementar los algoritmos encontrados previamente en la etapa anterior, para así saber su funcionalidad y características que lo destacan.

Fase IV. Analizar y comparar cuales algoritmos o técnicas pueden cumplir mejor las expectativas de reconocimiento.

Luego de ser implementados los algoritmos, se analizarán y se sacará una conclusión de cada uno mostrando los pros y contras, para luego hacer una comparación entre estos y llegar a una conclusión final de que algoritmo o algoritmos obtienen los mejores resultados.

1.4. ESTRUCTURA DE LA MONOGRAFIA.

Esta monografía está estructurada en 4 partes que incluyen 6 capítulos.

Parte I de introducción que comprende:

- a) Donde se realiza la formulación del problema, la justificación, se hace el planteamiento de 3 interrogantes, dos hipótesis y se determinan los objetivos específicos requeridos para alcanzar el objetivo general de desarrollar una monografía sobre “Algoritmos de reconocimiento de imágenes y video para reconocer gestos faciales de fatiga y somnolencia en conductores.”.

Parte II Estado del Arte: que comprende:

- a) Donde se realiza una búsqueda de información sobre factores de salud y consecuencias de la somnolencia, a través de 4 etapas.

Parte III Desarrollo de la monografía:

- a) Donde se trabaja por buscar las respuestas y comprobación a los interrogantes e hipótesis planteados.

Parte IV Parte que comprende:

- a) El capítulo 6 Conclusiones y bibliografía. Donde se resumen los resultados a los interrogantes e hipótesis planteados inicialmente, así como también los alcances logrados con esta investigación.
- b) Bibliografía en formato APA de todas las referencias utilizadas en este trabajo.

PARTE II ESTADO DEL ARTE

CAPITULO 2 FATIGA Y SOMNOLENCIA

La fatiga o la somnolencia, ambos son un campo muy extenso y complejo en muchos ámbitos, así que hacer una investigación exhaustiva estaría fuera del límite planteado para la monografía, por esta razón solo se abordaran desde el ámbito de la salud, conociendo sus consecuencias tanto psicológicas como físicas y también como afecta la vida social y laboral de quienes las padecen.

2.1. MARCO HISTORICO

La palabra “fatiga” es etimológicamente proveniente del latín y su significado es “cansarse” (lat. fatigare) o “estar saciado o extenuado” (lat. ad fatim, affatum) (University, 2020). La Real Academia de la Lengua Española también define este término como una “Molestia ocasionada por un esfuerzo más o menos prolongado o por otras causas, y que en ocasiones produce alteraciones físicas.” (Real Academia De la Lengua Española, 2019).

En el ámbito de la biomedicina, la fatiga es para algunos investigadores definida como algo que ocurre a lo durante o después de diversas actividades realizadas por el ser humano. El estrés prolongado suele ser el efecto principal que genera la fatiga. Dependiendo del individuo y sus circunstancias, la manifestación e intensidad de la fatiga como se citó en (Bruno Pedraz-Petrozzi, 2018).

La fatiga es un síntoma que se manifiesta como antelación o acompañamiento a otros eventos patológicos según otros investigadores. Por ejemplo, Coirault definió la fatiga como un estado de sufrimiento nervioso (en conjunto con el insomnio) que constituye el usualmente el inicio de muchas enfermedades mentales como se citó en (Bruno Pedraz-Petrozzi, 2018).

En la historia de la fatiga se pueden considerar 4 etapas importantes, llenas de aportes para la investigación de la fatiga como se planteó en (Bruno Pedraz-Petrozzi, 2018).

PRIMERA ETAPA

En la antigua Roma, fue Galeno (130 - 201) quien describió por primera vez pacientes que presentaban variedad de molestias en el hemiabdomen superior y que a la vez estas se encontraban relacionadas con alteraciones emocionales que a la vez se relacionaban con fatiga. A ello se le denominó “Morbus hypocraticus”, la cual describía un subtipo de melancolía que se creía estar relacionado con la inundación del bazo en el hipocondrio con bilis negra.

SEGUNDA ETAPA

Cerca de la segunda mitad del siglo XIX, George Miller Beard empezó el uso del término “neurastenia” (“A practical treatise on nervous exhaustion -Neurasthenia-: its symptoms, nature, sequences and treatment”) como forma de agrupación de todas las investigaciones previas con el fin

de describir a aquellas personas que padecían desórdenes psicológicos cuya queja principal era fatiga, ansiedad, tristeza y miedo.

TERCERA ETAPA

Llegando a la primera mitad del siglo XX, se generó una insostenibilidad de la terminología expuesta por Beard se procedió a eliminar de los diccionarios médicos el término “Neurastenia” desapareciendo por completo en la década de los 60. Ello se debió al marco de la psicoanalítica. Según a la concepción freudiana de salud-enfermedad, los pacientes “neurasténicos” son neuróticos y no existe ninguna explicación biológica posible para poder sostener la sintomatología de dicho diagnóstico, en esto se incluyó la fatiga.

CUARTA ETAPA

Hacia finales del siglo XX, la aparecieron nuevas descripciones sobre enfermedades inflamatorias e infecciosas, esto involucró también la descripción de fatiga crónica y cansancio similar como fue descrito por Beard con “neurastenia”, esto forzó a algunos investigadores y clínicos a “emplear” nuevamente este concepto. Después de un tiempo se desarrolló un nuevo concepto denominado “síndrome de fatiga crónica”.

La fatiga es un síntoma usual al momento de un accidente, Este estado son muy común en el diario vivir, ya sea por actividades laborarles o simplemente sociales. Sin embargo, muchos conductores desconocen o hacen caso omiso a los riesgos asociados a este síntoma. Los accidentes que puede ocasionar un conductor en estado de fatiga o sueño acumulado van desde colisiones con otros vehículos o personas hasta salidas de la vía transitada.

“Hay una clara evidencia de que la somnolencia, es un importante factor de riesgo para los accidentes de tráfico. La National Highway Traffic Safety Administration calcula que el 2,5% de los accidentes fatales y el 2% de los accidentes no fatales se relacionan con la presencia de somnolencia.” (Sueño, 2016, pág. 21).

2.2. MARCO REFERENCIAL

Muchas personas han indagado por estos senderos de los algoritmos para el reconocimiento de fatiga y somnolencia en conductores, por eso es importante que sean leídos, ya que aportan mucho al momento de realizar este estudio. En este espacio se enuncian y describen algunos proyectos que se han realizado con anterioridad.

2.2.1. Estudio comparativo de clasificadores para el reconocimiento de expresiones faciales:

realizado por : Víctor M. Álvarez - Claudia N. Sánchez - Sebastián Gutiérrez - Julieta Domínguez-Soberanes - Ramiro Velázquez

En este artículo los autores presentan un algoritmo para clasificar expresiones faciales en alguna de las ocho categorías básicas que propone Paul Ekman, como lo son enojo, desprecio, disgusto, miedo, alegría, tristeza, sorpresa y neutral. Para esto utilizaron la base de datos de Cohn-Kanade. Los clasificadores usados fueron: Redes Bayesianas Simples, máquina de Soporte Vectorial, árbol de Decisiones, bosque Aleatorio, adaBoost, perceptrón Multicapa. Es importante destacar que el algoritmo que mejor les funcionó fue el perceptrón, mostrando una precisión de 89%. (Domínguez Soberanes, Manuel Álvarez Pato, Sánchez Gómez, Gutiérrez Calderón, & Velázquez Guerrero, 2019)

2.2.2. Algoritmos en tiempo real para el reconocimiento de emociones faciales: una comparación de diferentes enfoques:

Realizado por: Aneta Kartali - Miloš Roglić - Marko Barjaktarović - Milica Đurić-Jovičić - Milica M. Janković.

En el artículo, los autores compararon tres enfoques de aprendizaje profundo basados en redes neuronales convolucionales y dos enfoques convencionales para la clasificación de las características del histograma de gradientes orientados. Lo anterior para el reconocimiento facial de emociones tales como: Enojo, Miedo, Felicidad y Tristeza. (Kartali, Roglić, Barjaktarović, Đurić-Jovičić, & Janković, 2018)

2.2.3. Algoritmos de detección de objetos para la detección y seguimiento de ojos:

realizado por: Miguel Ángel Antúnez Galindo

El proyecto que el autor plantea, consiste en diseñar un software capaz de identificar por medio del el análisis de imágenes indicios de fatiga en un conductor. Implementó los algoritmos de Viola-Jones y transformada de Hough para la detección y rastreo de ojos.

El proyecto aporta una valiosa información, ya que podemos analizar cómo funciona solo detectar los ojos y el tiempo de parpadeo. (Antúnez Galindo, 2019)

2.2.4. Aplicación para teléfono móvil con sistema operativo android que permita detectar somnolencia y emitir una alarma sonora a conductores de automóvil mediante procesamiento de imágenes:

Realizado por: Brayan Alirio Baquero Cardozo - Brayan Alexandert Torres Malaver

Los autores desarrollaron un aplicativo para teléfono móvil que mediante el procesamiento digital de imágenes permitirá dar un estimado del estado de somnolencia del conductor mediante el análisis de los ojos, y en dependencia de éste dar una alerta sonora al usuario. Implementaron tres algoritmos para probar la eficacia de cada uno realizaron pruebas teniendo en cuenta parámetros como tiempos de procesamiento, estabilidad en ambientes de luz no controlados y cantidad de aciertos y errores en la detección de somnolencia. (Torres Malaver & Baquero Cardozo, 2019)

2.2.5. Detección de estado peligroso del controlador basada en bibliotecas OpenCV y

Dlib mediante procesamiento de video móvil:

Realizado por: Igor Lashkov - Alexey Kashevnik - Nikolay Shilov - Vladimir Parfenov - Anton Shabaev.

En otra investigación, los autores diseñaron una aplicación que usa la cámara frontal de los celulares inteligentes para rastrear continuamente las características faciales y reconocer temprano los estados peligrosos de somnolencia y distracción del conductor. Las imágenes del conductor capturadas por la cámara pueden describir sus características faciales, como los movimientos de la cabeza, el estado de los ojos, el

estado de la boca y, luego, identificar un nivel actual de estado de fatiga. El reconocimiento de los estados peligrosos se hace en modos online y offline. En el modo online, dichos estados son determinados en tiempo real con la ayuda de librerías de visión por computador de OpenCV y Dlib mientras se conduce. Para el modo offline, el reconocimiento se basa en resultados de procesos analíticos proporcionados por un servicio en la nube, que utiliza no solo las estadísticas acumuladas en tiempo real, sino también las herramientas de aprendizaje automático recopiladas, almacenadas y producidas anteriormente. (Lashkov, Kashevnik, Shilov, Parfenov, & Shabaev, 2019)

2.2.6. Deep Learning con red neuronal convolucional para detectar estados de microsueño:

Realizado por: Venkat Krishnamoorthy - Reza Shoorangiz - Stephen J. Weddell - Lutz Beckert - Richard D. Jones

En un estudio, se exploró la capacidad de una red neuronal convolucional (CNN) para detectar estados de micro sueño a partir de datos de EEG (electroencefalograma) de 16 canales de 8 sujetos, realizando un visomotor 1D. Los datos estaban altamente desequilibrados. Cuando se promedió en 8 sujetos, hubo 17 estados de respuesta para cada estado de micro sueño. Se utilizaron dos enfoques para manejar la capacitación de CNN con desequilibrio de datos: sobre muestreo de la clase minoritaria y el aprendizaje basado en los costos. El EEG se analizó usando una época de 4 s con un tamaño de paso de 0.25 s. La validación cruzada de dejar un tema por fuera se utilizó para evaluar el rendimiento. (Krishnamoorthy, Shoorangiz, Weddell, Beckert, & Jones, 2019)

2.2.7. Sistema detector de somnolencia en secuencias de vídeo de conductores manejando usando visión computacional:

Realizado por: Juan Carlos Crespín Luis - Julián García Raúl Alexander

El autor desarrolla un sistema que detecta somnolencia en conductores de automóviles, este sistema hace parte de los (S.A.A.C). Para la investigación necesitaron consultar conceptos tales como: Ingeniería de Software, somnolencia en

conductores, y visión computacional. El sistema detecta rostros, ojos y muestra el estado del conductor. Utilizaron python para la programación del sistema con la ayuda de las librerías OpenCV y Dlib. (Crespin Luis & Raúl Alexander , 2019)

2.2.8. Sistema basado en la detección y notificación de somnolencia para conductores de autos:

Realizado por: Eduardo Luis Buenavides Muñoz - Muricio Miguel Medina Mendez. El autor de proyecto desarrolló un sistema que permite el reconocimiento y notificación de gestos como bostezo, cabeceo y ojos cerrados en personas con somnolencia. Para esto utilizaron como hardware un dispositivo Kinect y una cámara web modificada para capturar imágenes en ausencia de luz, y como software usaron Kinect SDK 1.8 y las librerías de visión artificial OpenCV y LuxandFace SDK, también utilizaron el algoritmo de Viola & Jones que se encuentra implementado en OpenCV. El sistema desarrollado permitió la detección de bostezo y cabeceo con facetracking de Kinect y LuxandFace en el día. De noche se logró detectar ojos cerrados, cabeceo y bostezo con LuxandFace, ojos cerrados y bostezo con OpenCvSharp un wrapper de OpenCV. (Buenavides Muñoz & Medina Mendez, 2018)

2.2.9. Reconocimiento de somnolencia en conductores bajo condiciones simuladas:

Realizado por: Luis Angel Rondon Condori - Frederick Jacinto Paucara Nuñez. El objetivo principal de los autores del proyecto es la utilización de algoritmos de detección de objetos para reconocer el estado de somnolencia en conductores. Para reconocer el estado de somnolencia se capturan fotogramas del conductor usando una cámara, evalúan cada fotograma buscando primero detectar un rostro, si un rostro es detectado entonces se evalúa el estado de los ojos ("abiertos" o "cerrados"), con la información del estado de los ojos de los 10 últimos fotogramas se calcula el porcentaje de ojos cerrados o PERCLOS, para un PERCLOS mayor a 40% consideraron que el estado de somnolencia del conductor es peligroso y muestran una señal de alarma. (Rondon Condori & Paucara Nuñez, 2013)

**PARTE III DESARROLLO DE LA
MONOGRAFÍA.**

CAPITULO 3 FACTORES Y CONSECUENCIAS DE LA FATIGA

El origen de la fatiga puede deberse a gran variedad de factores, en esta ocasión nombraremos algunos de ellos: Cansancio acumulado, estar demasiado tiempo frente al volante, consumo de alcohol o alucinógenos, pereza o también el consumo de medicamentos que producen el sueño.

Las consecuencias de la fatiga que afectan a la conducción son las siguientes:

Tiempo de reacción: La somnolencia incrementa el tiempo que una persona tarda en reaccionar ante un suceso ocurrido el ambiente del tráfico. Por ejemplo si una persona se encuentra en estado de somnolencia mientras conduce y frente a el otro conductor frena inesperadamente, el primer conductor no podría reacción rápidamente y se ocasionaría un choque.

Poca concentración y muchas distracciones: Estar en estado de somnolencia hace muy difícil mantener la concentración. Por tal razón se pueden presentar con facilidad muchas distracciones en poco tiempo. Esto se presenta usualmente en ambientes monótonos o en condiciones de escaso tráfico.

Toma de decisiones erróneas y de manera lenta: La somnolencia retarda el procesamiento de la información que provee el ambiente y por consiguiente es mayor el tiempo de reacción frente a un suceso inesperado.

Perturbaciones motoras: Los músculos del cuerpo humano se relajan, cuando se está bajo los efectos de la somnolencia, produciendo movimientos más lentos y con poca precisión. Es muy posible que se presenten tenues temblores en las manos o en otras partes del cuerpo.

Microsueños: “El microsueño se puede describir como un período breve e involuntario de pérdida de conciencia, los ojos se cierran al intentar inesperadamente. Es una de las consecuencias de un descanso nocturno deficiente. Puede resultar peligroso e incluso fatal, según la actividad que estemos realizando cuando ocurre el episodio.” (ideal, 2018). Los microsueños son los que provocan comúnmente accidentes de extrema gravedad.

Perturbaciones sensoriales: En un estado de somnolencia se necesitan que los estímulos sean más fuertes de lo normal para poder que sean captados con precisión. Aunque la somnolencia afecta todos los sentidos, el sentido de la vista es uno de los más importantes, este se deteriora significativamente. Resulta más complicado enfocar la vista, la visión de vuelve borrosa y se produce fatiga ocular.

Cambios comportamentales: Mientras se conduce en estado de somnolencia es posible que se produzcan sensaciones de tensión, nerviosismo y usualmente agresividad. Además de eso también se pueden generar comportamientos riesgosos tales como apresurarse al lugar de destino cuanto se está más cerca de él para poder llegar y descansar.

CAPITULO 4 SISTEMAS DE RECONOCIMIENTO FACIAL

En este capítulo se hablan de sistemas informáticos que implementan algoritmos basados en modelos matemáticos.

4.1. SISTEMA DE DETECCIÓN DE MICROSUEÑO MEDIANTE EL USO DE LA TÉCNICA EAR (EYE ASPECT RATIO ALGORITHM)

Los ingenieros Mohd Arif Ngasri, Iza Sazanita Isa, Siti Noraini Sulaiman, Zainal Hisham Che Soh de la Universidad Tecnológica UiTM Cawangan Pulau Pinang de malasia, desarrollaron un sistema de detección basado en video de apertura de ojos específicamente para eventos de microsueños entre conductores. Este sistema se utiliza para ayudar a evitar accidentes automovilísticos causados por microsueños entre los conductores mientras conducen. Principalmente, el objetivo de este sistema propuesto por los ingenieros, es detectar el evento de microsueño mientras se conduce y alertar a los conductores ya que, nadie puede alertar al conductor con microsueños, especialmente cuando conducen solos. Ellos utilizaron programación en Python con técnica de procesamiento de imágenes y módulo Raspberry Pi para que sea aplicable como dispositivo independiente. El algoritmo EAR es el de mayor importancia, ya que la detección del microsueño se realiza mediante el cierre de ojos. Probaron varias condiciones de apertura y cierre de ojos, luz y oscuridad, y con o sin anteojos / gafas de sol. A partir del resultado, el sistema mostró que la capacidad de la detección automática de microsueños propuesta basada en imágenes de vídeo ha mostrado un rendimiento prometedor. Sin embargo, la efectividad de la detección ocular debe tolerar las condiciones circundantes que también se deben al tipo de cámara utilizada. Por lo tanto, ellos sugieren que la cámara pueda funcionar con funciones de visión nocturna como el modo infrarrojo (IR) para que se pueda aplicar durante el día o la noche. (H. Nguyen & P. Examiner, 2007).

El sistema de detección de microsueños está diseñado para detectar la somnolencia entre el conductor y también para alertar al conductor. Utilizaron el método de técnicas de procesamiento de imágenes para detectar el microsueño basado en el cierre de los ojos del conductor. La cámara se utiliza para la transmisión de video para capturar la cara del conductor y monitorear el cierre de los ojos del conductor. A partir de ahí, se determina el microsueño del conductor. Este sistema fue diseñado para ser aplicable como dispositivo móvil que es pequeño y de fácil manejo, por lo que es aplicable para ser instalado en cualquier vehículo. Para la fuente de alimentación de este dispositivo, se puede utilizar la toma de corriente auxiliar de 12V o también conocida como encendedor de cigarrillos en el

vehículo. Para el diseño de hardware, el sistema de detección de microsueños consiste en el componente principal, que es el módulo de placa Raspberry Pi 3 B (1GB RAM) con WLAN incorporada como microcontrolador. Desde el microcontrolador, la cámara Raspberry Pi 5mp se usa como entrada que también actúa como sensor para capturar la cara del conductor en la transmisión de video. El puerto de interfaz en serie de la cámara (CSI) se utiliza para conectar la cámara (resolución de 5Mp) y uno de los pines GPIO utilizados para conectar con el zumbador. Para fuente de alimentación, esta placa construida con micro USB 5V1, 2.5A. Como imagen de entrada del flujo de video, el sistema fue programado para detectar la cercanía de los ojos del conductor. Según la configuración del umbral en la programación, el sistema de detección se activará para el comportamiento de microsueño que depende de la apertura y el cierre de los ojos del conductor. El timbre se conectará a este microcontrolador como una de las salidas del sistema. El timbre se usa para alertar al conductor cuando se detecta el evento de microsueño. En la Ilustración 1 se muestra la idea general del sistema

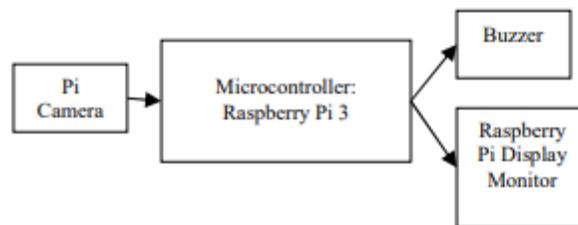


Ilustración 1: Sistema de detección de microsueño

Mientras tanto, para el diseño del software, usaron Python versión 3 para programar este sistema como se muestra en la Ilustración 2. En primer lugar, el sistema está programado para detectar la cara en la secuencia de video usando el algoritmo Haar Cascade de Open CV. Este algoritmo se utiliza para detectar el objeto en la imagen. El algoritmo se puede entrenar para clasificar el rostro a partir de muchas imágenes, incluida la imagen positiva (con rostro) y la imagen negativa (sin rostro) (P. & M., 2014). Open CV proporciona muchos clasificadores previamente entrenados para objetos en archivos XML.

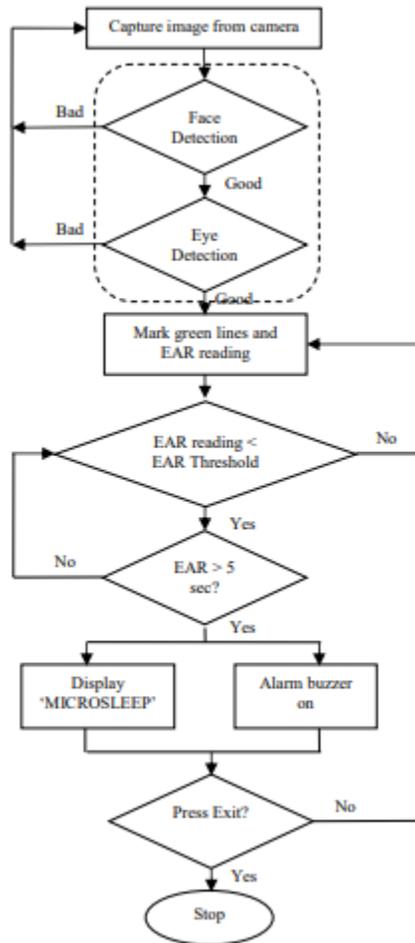


Ilustración 2: Secuencia de detección ocular para detección de microsueños 2

A partir de la detección de rostros, los ojos se ubican aplicando puntos de referencia (Landmarks) faciales (Dlib's). El predictor de puntos faciales de Dlib's se utiliza para localizar 68 puntos de la imagen de la cara como se muestra en la Ilustración 3. La localización de la marca de 68 puntos para la cara está codificada dentro del diccionario FACIAL_LANDMARKS_IDXS en la biblioteca de "imutils" (Rosebrock, 2017). Lo utilizan para localizar las características de la cara como ojos, cejas, nariz, boca y línea de la mandíbula. Por lo tanto, la localización de los ojos se aplica cuando importan la biblioteca imutils al programa de detección de microsueños.



Ilustración 3: Puntos de referencia faciales (landmarks)

Para la detección de ojos ellos utilizan el algoritmo de relación de aspecto del ojo (EAR) o “Eye aspect ratio”, definen la condición normal de los ojos (ojos abiertos) para un valor constante. Este algoritmo se utiliza para estimar el estado de apertura de los ojos que se deriva de la relación entre la distancia del punto de referencia vertical del ojo y la distancia del punto de referencia horizontal (Cech, 2016). Basado en el algoritmo EAR, la imagen del ojo representa seis puntos de coordenadas como se muestra en la Ilustración 4, el límite amarillo se usa para localizar el par de ojos. Por lo tanto, basado en la coordenada de localización, la ecuación de EAR es Ecuación 1: Detección de nivel EAR a partir de los 6 puntos en los ojos:

$$\text{Microsueño}_{ojos} = \frac{(P2 - P6)(P3 - P5)}{2(P1 - P4)}$$

Ecuación 1: Detección de nivel EAR a partir de los 6 puntos en los ojos

El umbral de EAR lo establecieron en 0,3 siguiendo el resultado promedio, tomado en los datos de 15 sujetos diferentes. Cuando el nivel de EAR es más bajo que el umbral, significa que el nivel de cierre de ojos es bajo. Sin embargo, este estado aún no está declarado como evento de microsueño. El microsueño es alertado solo cuando este estado cuente durante 5 segundos. Después de 5 segundos, como el valor de EAR detectado es menor que el umbral, el evento del microsueño comienza a alertar. Por último, la condición “Microsueño_{ojos}” Se define por el valor medido de EAR en comparación con el umbral EAR como se ve en la Ecuación 2 : medición a partir de umbral.

$$\mathbf{Microsueño}_{ojos} = \begin{cases} \text{abiertos si } EAR \geq \text{umbral} \\ \text{cerrados si } EAR < \text{umbral} \end{cases}$$

Ecuación 2 : medición a partir de umbral

El estado de alerta se detiene después de que el sistema detecta que los ojos permanecen en su estado normal. De esta manera el sistema no se activará si el ojo del conductor parpadea mientras conduce.

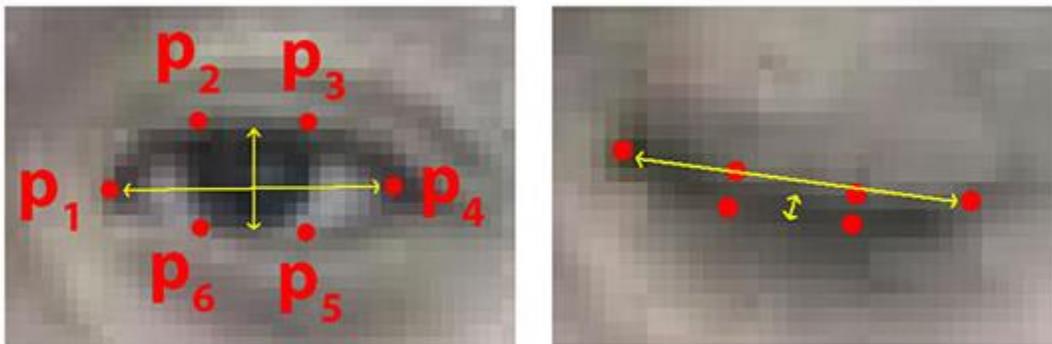


Ilustración 4: Hito de seis puntos para la detección de ojos

4.1.1. RESULTADOS EAR

4.1.1.1. DETECCIÓN DEL SISTEMA EN DIFERENTES CONDICIONES

Este estudio para la detección de microsueños basado en la apertura y el cierre de los ojos se probaron en varias condiciones de apertura y cierre de los ojos, luz y oscuridad y con o sin gafas / anteojos de sol que pueden representar los efectos de una aplicación en real. La observación se realizó durante el día y la noche. La **¡Error! No se encuentra el origen de**

la referencia. muestra la pantalla de flujo de video que puede detectar el par de ojos y el resultado cuando el sistema de detección ha activado el estado de microsueño. Los límites verdes con forma convexa se utilizaron para marcar el par de ojos. Se incluye en la pantalla un valor medido de la relación de aspecto de los ojos (EAR) de la persona / conductor. Si el valor EAR medido es superior a un umbral que es 3.0, la condición del conductor se declara como estado normal y no se da ninguna alerta de lo contrario se activará la alerta por el sistema de detección.



Ilustración 5: (a) Apertura de ojos y (b) cierre basado en el umbral EAR < 3.0

A continuación los ingenieros aplicaron diferentes factores para ver la eficiencia en la detección, la Ilustración 6 muestra el resultado de un entorno con menos luz. Esta situación simula el rendimiento del sistema durante la conducción nocturna con iluminación. El sistema pudo detectar la apertura de los ojos y también pudo cambiar un estado de alerta de microsueño. Sin embargo, en la condición de totalmente oscuro (sin luz) muestra que el sistema es insuficiente para detectar los ojos en la imagen. Por lo tanto, el sistema no puede leer el estado del EAR del conductor en un lugar completamente

OSCURO.



Ilustración 6: Efecto de la oscuridad circundante en la detección del sistema (a) menos oscuridad (b) completamente oscuro

A partir del resultado que obtuvieron, el sistema de detección de microsueños funciona con éxito durante el día y la noche con un brillo mínimo. Sin embargo, este sistema no puede detectar pares de ojos en condiciones de oscuridad total. Esto se debe a que la cámara Raspberry Pi de 5 MP utilizada en este sistema no es compatible con la visión nocturna. Esto abre una brecha de análisis que hablaremos en el siguiente capítulo de esta monografía ya que encontramos hipotéticamente que las limitaciones son con respecto a hardware y no a software.



Ilustración 7: Efecto de gafas de sol



Ilustración 8: Efecto de anteojos en el sistema de detección

En la Ilustración 7 se muestra que el sistema de detección de microsleeps no detecta un par de ojos en una persona que usa gafas de sol y en la Ilustración 8 se muestra que el sistema de detección de microsleeps, en ocasiones, no detecta correctamente un par de ojos en una persona que usa gafas medicadas de sol usa gafas tampoco pudo dar el buen resultado. El valor EAR también proporciona una lectura diferente en comparación con una persona sin gafas. Por tanto, este sistema es eficazmente adecuado cuando el conductor no lleva gafas de sol y puede no ser eficaz para el conductor con gafas.

4.1.1.2. MEDICIÓN Y ANÁLISIS DE EAR

El análisis EAR tabulado en la Ilustración 9 muestra el resultado que tomaron para observar el valor EAR de 15 sujetos diferentes en 3 estados variables que son ojos abiertos, ojos estrechos y ojos cerrados. A partir del resultado, se calcularon los datos para los valores promedio de EAR de todos los estados. A partir del promedio de lectura EAR durante ojos abiertos y datos de ojos estrechos, el valor umbral más óptimo para declarar el microsleep es 0.3. En este estudio hecho en la Universidad Cawangan Pulau, los patrones de ojos estrechos muestran que el estado ha caído en la condición de microsleep, por lo tanto, EAR 0.3 es el umbral más preciso. Mientras tanto, encuentran que la EAR para el estado de ojos completamente cerrados es, en promedio, menos de 0,2.

Sujeto	Relación de aspecto de ojos (EAR)		
	<i>Ojos abiertos</i>	<i>Ojos estrechos</i>	<i>Ojos cerrados</i>
1	0.358	0,245	0,195
2	0.359	0,277	0,170
3	0.304	0,244	0,193
4	0.313	0,296	0,234
5	0.320	0,265	0,184
6	0.344	0,295	0,196
7	0.320	0,268	0,179
8	0.309	0,279	0,180
9	0.316	0,254	0,192
10	0.321	0,295	0,185
11	0.311	0,285	0,175
12	0.305	0,245	0,188
13	0.334	0,271	0,168
14	0.328	0,255	0,191
15	0.318	0,234	0,193
Promedio	0.324	0,267	0,188

Ilustración 9: Resultados EAR

4.1.2. ANÁLISIS

El sistema de detección de microsueños tuvo éxito en la detección de apertura y cierre de ojos mediante el uso de una técnica de procesamiento de video. El sistema es aplicable en varias condiciones de día y de noche con algunas limitaciones según lo vimos en los resultados. El sistema de detección de microsueños es capaz de detectar el estado de microsueños mediante la apertura y el cierre de los ojos. Cuando usan el módulo Raspberry Pi, el sistema se puede aplicar como un sistema independiente para que el conductor sea alertado mientras conduce. Sin embargo, el diseño del sistema tiene algunas limitaciones, especialmente para la aplicación de conducción nocturna. Por último, la cámara sugerida para usar es la cámara RPi NoIR ya que está diseñada sin filtro de infrarrojos que le permite capturar la imagen en condiciones de poca luz. Esto ayudaría bastante en el tema de detección nocturna porque en términos generales el sistema cumple funcionalmente con el propósito final y lo que marca ese punto final es el algoritmo EAR.

4.2. SISTEMA DE DETECCIÓN DE SOMNOLENCIA PARA LA PERFECCIÓN DE LA INTERFAZ CEREBRO-COMPUTADORA MEDIANTE EL ALGORITMO DE VIOLA-JONES

INTRODUCCIÓN

Los ingenieros Md. Kamrul Hasan, S. M. Hasnat Ullah, Shantanu Sen Gupta y Mohiuddin Ahmad del departamento de Ingeniería Eléctrica y Electrónica (EEE) de la Universidad de Ingeniería y Tecnología de Khulna (KUET) desarrollaron un sistema de detección de somnolencia para la perfección de la interfaz cerebro-computadora (BCI), utilizando el algoritmo de Viola-Jones y la electroencefalografía (EEG), que es una metodología fácil y de bajo costo (en comparación con otros), para el monitoreo de la función cerebral.

En el BCI basado en EEG, el potencial evocado visualmente en estado estacionario (SSVEP) actúa como una señal cruda vital que se procesa y extrae para las características deseables. La perfección de cualquier dispositivo BCI depende de la perfección de la señal EEG. En la Fig. 1 se muestra una señal EEG saludable típica basada en SSVEP en el dominio de frecuencia (Hz) (**M. K. Hasan, R. Z. Rusho, & M. Ahmad, 2010**). La somnolencia es el estado de la persona que la siente peculiarmente somnolienta durante el día o la noche en el horario laboral. La presencia de somnolencia es la principal contaminación en la señal del EEG y también dificulta la perfección de BCI. La somnolencia en la señal del EEG es el estado momentáneo entre la conciencia y la inconsciencia. La estimación o predicción del nivel de somnolencia en la señal de EEG para un funcionamiento perfecto de BCI utilizando señales de EEG es el campo de investigación clave en la ingeniería biomédica (BME).

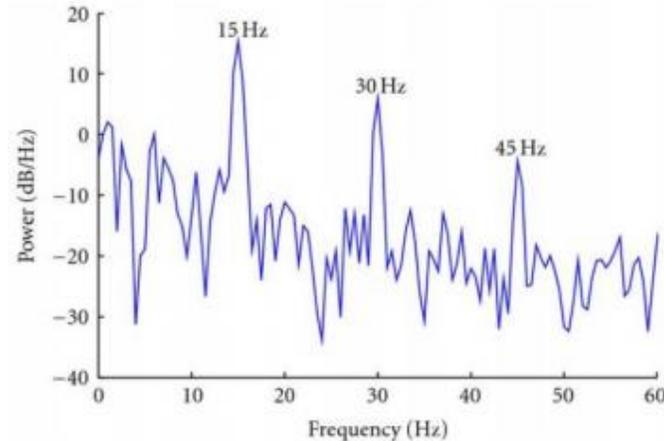


Ilustración 10: Representación típica de la señal SSVEP en el dominio de frecuencia (Hz)

La persona que contiene somnolencia puede quedarse dormida en situaciones desafortunadas en cualquier momento inadecuado. Existen varios motivos de somnolencia de las personas (**Ryan James Kimmel, y otros, 2019**) como:

- I. Dolor continuo.
- II. Diabetes.
- III. Trabajando mucho tiempo sin descansar.
- IV. Variaciones en la densidad de sodio en sangre
- V. Medicamentos del sueño y antihistamínicos
- VI. El tiempo para dormir no es suficiente y el trastorno del sueño
- VII. Demasiado calcio en la sangre y tiroides hipo-activa.

Cualquiera que sea la causa de la somnolencia, el objetivo principal de los ingenieros es determinar el nivel de somnolencia a partir de la señal de EEG del paciente. Una vez que detecte el nivel de somnolencia, el funcionamiento de BCI será como el comportamiento de la señal EEG de los pacientes de interés.

El diagrama de bloques básico de la Interacción humano-computadora (HCI) o la Interfaz cerebro-computadora (BCI) se muestra en la **Ilustración 11**. El EEG sin procesar se procesa para la extracción de características adecuadas, la señal sin procesar se debe filtrar para la eliminación de porciones no deseadas. Pero la somnolencia es la contaminación prominente en la señal del EEG que degrada el rendimiento de BCI o

HCI. En esta investigación la detección de somnolencia se hace con el algoritmo de Viola-jones, como se muestra en la **Ilustración 12** Ilustración 11: Diagrama de bloques generalizado para representación HCI o BCI. La somnolencia debe eliminarse de la señal EEG de los pacientes ya que la perfección de BCI depende de la perfección de la señal SSVEP.

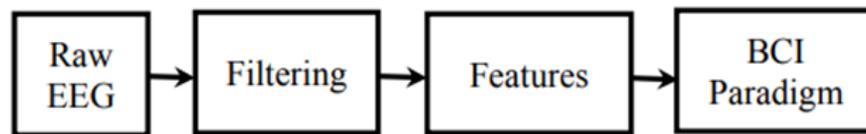


Ilustración 11: Diagrama de bloques generalizado para representación HCI o BCI

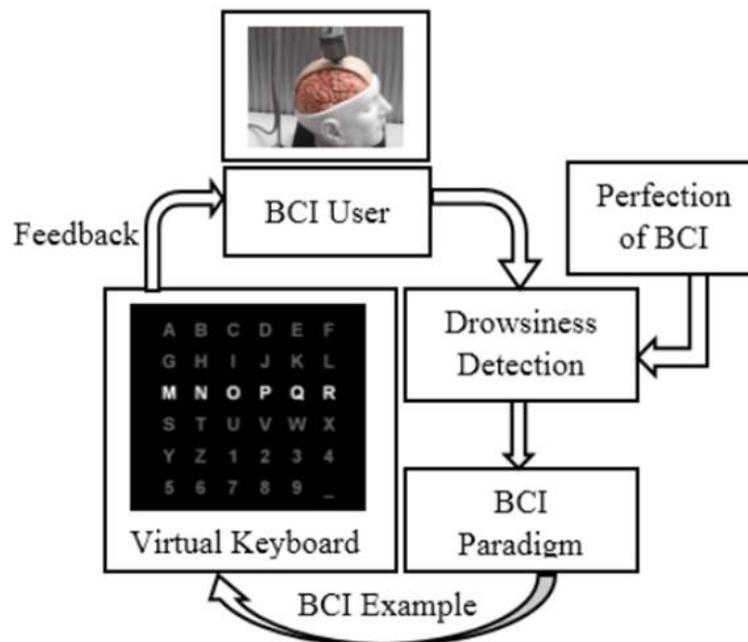


Ilustración 12: Representación de conceptos para la detección del nivel de somnolencia

El marco de investigación general que los ingenieros propusieron se describe en varias secciones que se indican a continuación:

A. Configuración del sujeto

La configuración experimental propuesta se muestra en la **Ilustración 13**. Se coloca una cámara de monitoreo de tal manera que pueda captar la cara completa de los usuarios de BCI cuyo nivel de somnolencia se va a detectar. A partir de la imagen

capturada, el ojo se puede detectar y separar mediante el algoritmo de Viola-Jones. Después de analizar este ojo detectado, podemos llegar a la conclusión de que el ojo del usuario está cerrado, semi-abierto o completamente abierto, etc. La condición del ojo (cerrado, semi-abierto o completamente abierto) es la principal preocupación de la detección del nivel de somnolencia.

B. Marco general

En la **Ilustración 14** se muestra un marco general. La imagen de los usuarios de BCI se puede capturar desde la cámara. Utilizando el algoritmo Viola-Jones, en primer lugar, la detección de rostros y luego la detección de ojos se han realizado con una precisión de grados. Luego, el ojo se ha recortado y se ha convertido en una imagen binaria. Comparando esta imagen con la plantilla ideal, se puede estimar el nivel de somnolencia de los usuarios de BCI. Entonces se puede decidir si el nivel de somnolencia es aceptable o no.

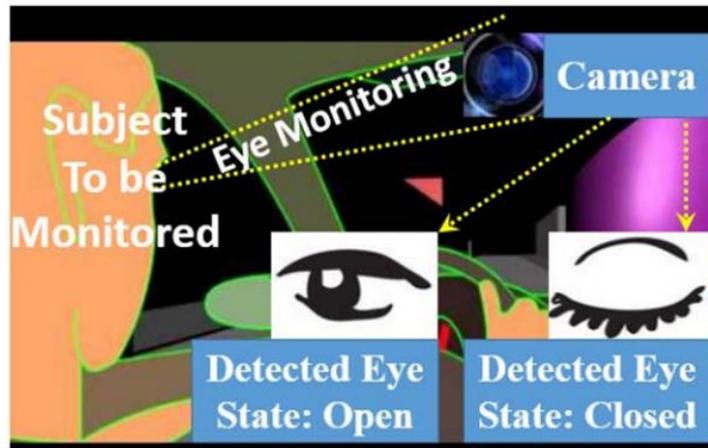


Ilustración 13: Monitorización del sujeto para detección ocular

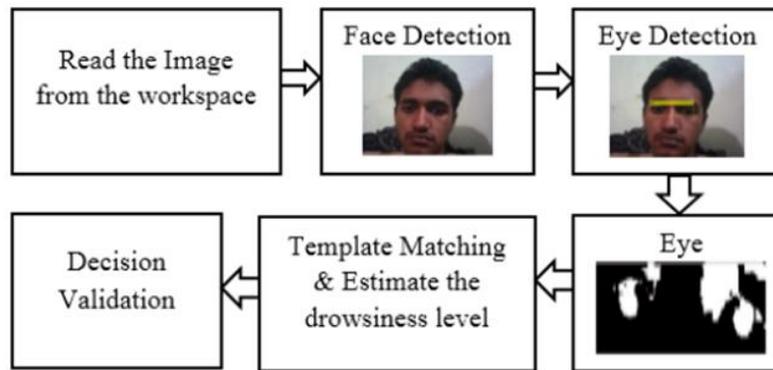


Ilustración 14: Representación del trabajo global propuesto

C. Diagrama de flujo del marco

El diagrama de flujo de este marco se muestra en la **Ilustración 15**. Se ha obtenido una imagen de cámara preestablecida de los usuarios de BCI. Después de tomar la imagen, la cara de los usuarios ha sido detectada y anotada usando el algoritmo de detección de objetos de Viola-jones. Por otra parte, utilizando el algoritmo de Viola-jones, se ha detectado y separado el ojo. Comparando el ojo con una plantilla ideal que predefinida, llega la decisión de establecer si el ojo está cerrado, semi-abierto o completamente abierto. Si está completamente abierto, es aceptable y el proceso puede continuar. Si el ojo está semi-abierto, la condición es tolerable pero debe proporcionarse una alarma. Si el ojo está cerrado, se decidirá ya que la condición es anormal y el proceso debe detenerse allí.

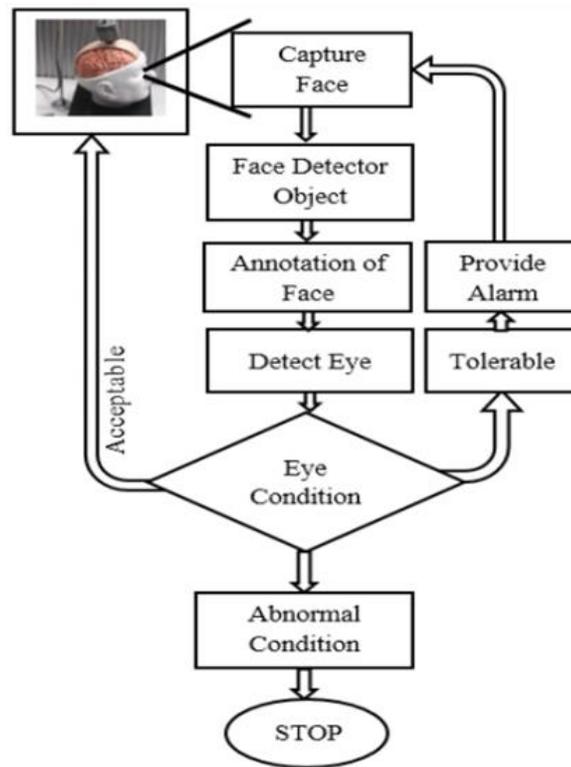


Ilustración 15: Diagrama de flujo del trabajo propuesto

D. Algoritmo de Viola-Jones

El algoritmo Viola-jones es el algoritmo para la detección del objeto que es el primer marco de detección de objetos en la detección en tiempo real. En 2001, Paul Viola y Michael Jones propusieron este sofisticado algoritmo que se utiliza principalmente para la detección de rostros (**Wikipedia, wikipedia, 2020**) - (**P. Viola & M. Jones, 2004**). Hay algunas ventajas vitales de este algoritmo que se mencionan a continuación en algunos puntos:

1. Proporciona alta velocidad de cálculo.
2. La selección de funciones más eficaz.
3. escalado de las funciones.
4. Tasa de detección muy alta.

I. RESULTADOS DE LA SIMULACIÓN Y DISCUSIÓN

La cara (Salida de la cámara) que se toma del espacio de trabajo se carga en MATLAB. Después de tomar esta cara, se detecta el ojo mediante el algoritmo de Viola-Jones. En

la **Ilustración 16** se muestra un ojo típico detectado. La ilustración (a) es para el ojo cerrado de los usuarios BCI y la ilustración (b) para el ojo abierto de los usuarios BCI.

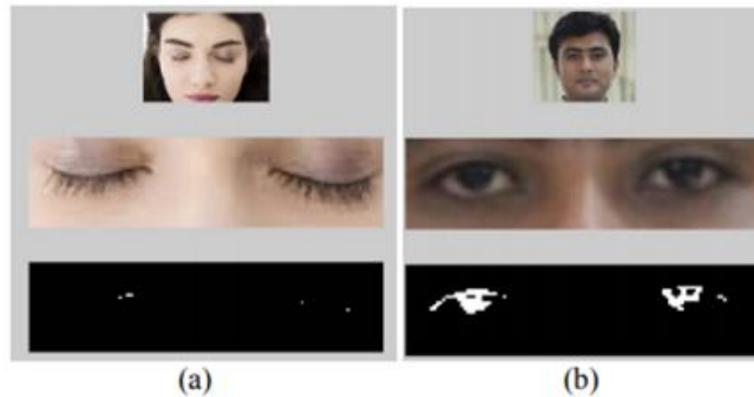


Ilustración 16: Ojo detectado usando el algoritmo de Viola-jones a) estado de ojo cerrado b) estado de ojo abierto

El nivel de somnolencia debido a varios estados del ojo (Abierto, Cerrado, Parcialmente abierto o cerrado, medio abierto) se muestra en la **Ilustración 17**. En la ilustración, se nota que durante el estado completamente cerrado del ojo hay una gran cantidad de somnolencia en la señal de EEG comparando el estado completamente abierto del ojo. Hay un cambio significativo en la somnolencia desde el estado parcialmente cerrado de los ojos hasta el estado completamente cerrado.

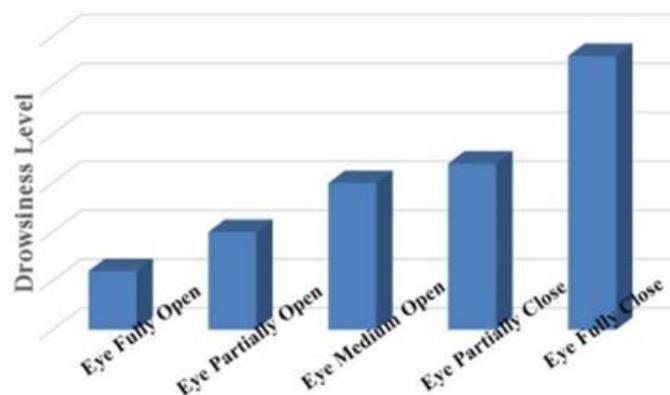


Ilustración 17: Nivel de somnolencia debido a varios estados oculares

El nivel de relajación en las bandas alfa y beta (después de filtrar la señal de EEG sin procesar) del usuario BCI se muestra en la **Ilustración 18** con respecto al nivel de somnolencia. La mayor potencia de la banda alfa y la menor potencia de la banda beta indica que los sujetos están en un estado más relajado (**M. K. Hasan, M. S.**

Hossain, T. K. Ghosh, & M. Ahmad, A SSVEP Based EEG Signal Analysis to Discriminate the Effects of Music Levels, 2015). De la Ilustración 18, también está claro que al comienzo de la ejecución de BCI, la relajación está en el mismo punto, pero después de un tiempo cuando la somnolencia aumenta debido al cansancio, la pendiente del nivel de relajación de la banda alfa es más pronunciada que la pendiente del nivel de relajación de la banda beta. Indica que la potencia de la banda alfa aumenta más que la potencia de la banda beta. Como consecuencia, los artistas de BCI se sienten más relajados y, en última instancia, el desempeño de los paradigmas de BCI cae.

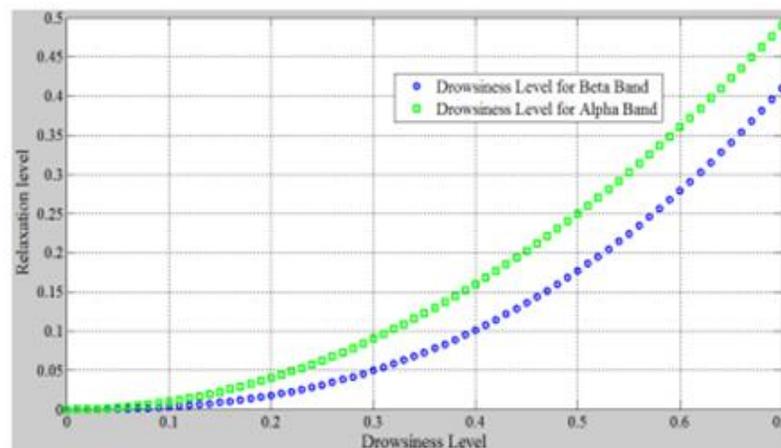


Ilustración 18: Nivel de relajación en la banda de señal del EEG (Alfa y Beta) debido al nivel de somnolencia de los intérpretes BCI

II. CONCLUSIÓN Y APLICACIONES

La somnolencia es una de las contaminaciones más dominadas que degrada la calidad de la señal de EEG y también degrada la perfección de los paradigmas de BCI. En esta investigación los ingenieros proponen una técnica sencilla para la detección de la somnolencia que se basa en la técnica de detección ocular basada en el algoritmo Viola-Jones. A partir de la salida de la simulación de este sistema propuesto, se enfoca la detección eficiente de rostro así como el nivel de somnolencia. El nivel de potencia es más para la banda alfa que para la banda beta de la señal EEG para el valor creciente de la somnolencia. A más poder de la banda alfa de EEG indica mayor relajación de los usuarios BCI. Cuanto mayor es la relajación de los usuarios BCI, hay una degradación

de la calidad de la señal EEG basada en SSVEP. Dado que SSVEP depende de los efectos de visualización en el ojo. El sistema que proponen los ingenieros puede aplicarse fácilmente al Sistema de Transporte Inteligente (ITS) porque al usar este sistema se puede monitorear la condición de fatiga, lo que puede reducir los accidentes de tránsito.

CAPITULO 5 ALGORITMOS PARA EL RECONOCIMIENTO FACIAL

Alrededor de los años 70 se dio origen a los primeros algoritmos. Cabe aclarar que muchos años antes en el 1840-43 fue la matemática informática británica Ada Lovelace quien gracias a sus trabajos dedicados a la calculadora de uso general dio origen a lo que podría considerarse como el primer algoritmo cuyo destino sería ser procesado por una máquina. Antes de continuar explicaremos de manera breve lo que es un algoritmo. Se trata de todas las instrucciones o reglas definidas y finitas que dan paso a la solución de un problema gracias a su organización, en cual tiene sus bases en la capacidad de invención en lo que respecta y análisis de medidas del cuerpo humano. Teniendo esto claro podemos proseguir con la línea temporal en la cual aparecen los primeros algoritmos que podemos concebir como algoritmos modernos, si consideramos los estudios de Lovelace. Así pues, durante los años 70 el avance tecnológico era muy reducido como se cabría de esperar, fue por esta razón que dicha invención no pudo alcanzar su potencial en este periodo de tiempo, ya que no parecía poseer ninguna utilidad para las necesidades de la época. Sin embargo, ya en los 90, cuando dicha investigación se reanudó contando con nueva tecnología a su disposición. Convirtiéndose esto en una carrera similar a la espacial, la corporación japonesa Fujifilm se postuló como la primera en utilizar dicha tecnología direccionada al reconocimiento facial. A pesar de esto, en 2005 fue la empresa Nikon, la cual también es japonesa, la primera en lanzar un producto que hiciera uso de esta tecnología. Teniendo en cuenta que Nikon es la empresa líder del sector fotográfico y óptico, se sobre entiende que la utilización de algoritmos para reconocimiento facial es más frecuente y mayor en el campo de la fotografía digital

Habiendo ya explicado brevemente el contexto histórico alrededor de los algoritmos utilizados para reconocimiento facial, pasaremos pues a hablar de los diferentes tipos existentes abriendo paso así al algoritmo de Viola & Jones. Fue en el 2001 cuando este método de detección de objetos fue propuesto por Paul Viola y Michael Jones, de ahí su nombre. A pesar de que puede utilizarse para la detección de objetos en tiempo real, este es preferido para la detección de rostros por las siguientes razones: la probabilidad de

verdaderos positivos alcanza un 99,9%, por otro lado la de falsos positivos solo ronda el 3,33%, dicho de otra manera, la probabilidad de éxito es 30 veces mayor a la de fracaso. Esto lo convierte en un método sorprendentemente robusto y rápido en el procesamiento de imágenes. Además, gracias a que usa una representación de imagen llamada cascada, no utiliza la imagen en cuestión directamente. Por último, ya que el algoritmo es capaz de dividir la imagen integral en regiones de diferentes tamaños puede determinar qué áreas de la imagen corresponder a un rostro discriminando aquellas que no.

5.1. VIOLA-JONES

El algoritmo viola-jones fue propuesto por P. Viola, and M. Jhones con la descripción de “Robust Real-Time Face Detection” o detección de rostros robusta en tiempo real, donde muestra una metodología que lleva consigo 3 etapas como se muestra en la **Ilustración 19**. **(Rodríguez Baque, 2017)** , podemos ver que se inicia con una Imagen de entrada que se enlaza a una imagen integral en la primer etapa, en este punto la imagen se transforma y queda integrada, cuando hablamos de una imagen integral nos referimos a una imagen del mismo tamaño a la original pero donde el valor de cada píxel será la suma de todos los píxeles situados a la izquierda y arriba en la imagen que ingresamos anteriormente **(López Peña, Valveny, & Vanrell, 2015)**.

Llegando a la segunda etapa se realiza la extracción de las características o propiedades usando filtros de segmentación o base Haar para finalmente llegar a la etapa 3 en donde se usa el algoritmo de aprendizaje (boosting) para la construcción de clasificadores en cascada utilizando toda la información recopilada de la salida de un clasificador como información para el siguiente clasificador **(Wikipedia, 2020)**.

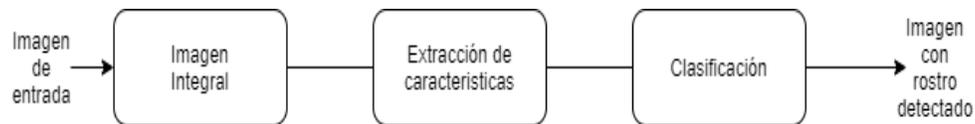


Ilustración 19: Diagrama bloques

5.1.1. IMAGEN INTEGRAL

Como anteriormente se mencionaba la imagen integral toma los píxeles situados en la parte superior izquierda de la imagen original Ilustración 20: Imagen integral esta nueva imagen permite extraer rápidamente características a escalas diferentes ya que se trabaja con una imagen acumulativa, esto quiere decir que se construye a partir de operaciones básicas.

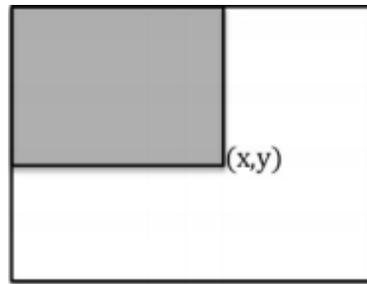


Ilustración 20: Imagen integral

Como podemos observar la **Ilustración 20** en la parte superior izquierda marcada en un área gris (x,y) tiene la suma de píxeles de la imagen original que mencionamos anteriormente, la sumatoria contiene la suma de todos los rectángulos de la imagen y ello se hace con el siguiente cálculo

$$I(x,y) = \sum_{x' \leq x, y' \leq y} Im(x',y')$$

Donde $I(x, y)$ es la imagen integral e $Im(x, y)$ es la imagen original.

5.1.2. CALCULO DE IMAGEN INTEGRAL

Cada píxel es la sumatoria de los píxeles por encima y a la izquierda de la imagen integral. Para ilustrar en la **Ilustración 21**, La imagen integral se rellena con ceros (0) a la izquierda y a la parte superior para permitir el cálculo. Miremos los píxeles como una matriz donde cada campo es una fila y una columna identificadas. El valor de píxel en (1,1) en la imagen original se convierte en el valor de píxel (2,2), como mencionamos anteriormente dicho valor en la imagen integral se obtiene de la suma por encima y a la izquierda de este píxel, es este caso en el píxel (1,1) obtendremos el mismo valor al no haber números en la parte superior y en la parte izquierda. El valor de píxel en (1,2) en la imagen original se convierte en el valor de píxel (2,3) en la imagen integral luego de hacer la sumatoria del píxel por encima de él que sería (4+2) y a la izquierda (6+0). El i en (2,2) en la imagen original con el valor 8 se convierte en el píxel en (3,3) en la imagen integral con el valor 24 luego de

agregar el valor de píxel por encima de él con la suma (8+10) y agregar el píxel a la izquierda de él con la otra suma ((18+6).

2	10
4	8

Imagen de entrada

0	0	0
0	2	12
0	6	24

Imagen integral

Ilustración 21: Ejemplo de una sumatoria de imagen integral

5.1.3. CARACTERÍSTICAS HAAR

Para representar la imagen en el método Viola-Jones se utilizan características Haar (**P. Viola & M.J. Jones, Rapid Object Detection using a Boosted Cascade of Simple Features**”, proceedings IEEE Conf. on Computer Vision and Pattern Recognition, 2001). Para simplificar, se utilizan para clasificar los segmentos de la imagen en características rectangulares **Ilustración 22**.

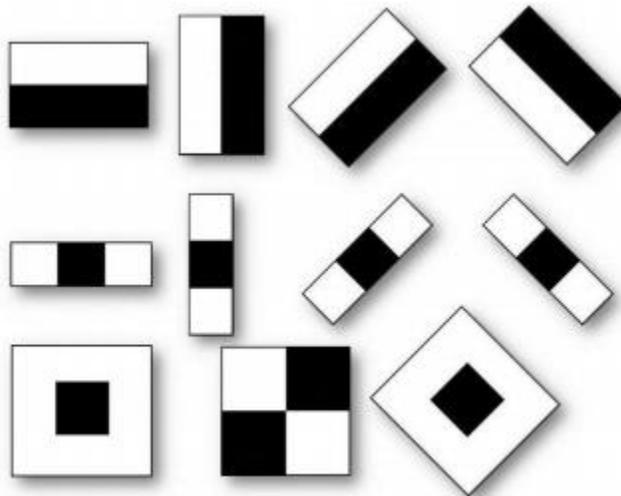


Ilustración 22: Características rectangulares

Para comprender el objetivo de utilizar estas funciones, debemos imaginar el proceso de superposición de cualquier función en nuestra imagen. El siguiente paso es calcular la suma del brillo de los píxeles debajo del área blanca y negra de la característica y, usando la ecuación (1), calculamos su diferencia, que es la característica general de la anisotropía de algún segmento de la imagen. El valor calculado de dicha característica será:

$$F = X - Y \quad (1)$$

Dónde X, Y son valores totales del brillo de los píxeles, que se superponen con el área de característica en blanco y negro. De esta manera, una agrupación de características se pueden usar para para codificar contrastes en rostros humanos.

Cuando hablamos de características se consideran 3 tipos: (**Ilustración 23**)

- Dos rectángulos: se calcula teniendo la diferencia entre la suma de los píxeles en las regiones rectangulares.
- Tres rectángulos: se calcula con la suma de los píxeles que están dentro de dos regiones rectangulares contrarias a las internas (exteriores).
- Cuatro rectángulos: se calcula obteniendo la diferencia de las diagonales de los pares de rectángulos.

Para calcular estas características en la imagen dada, usamos la representación integral de la imagen. El principal problema aquí es una gran cantidad de funciones derivadas. Incluso para imágenes pequeñas, la cantidad de características superpuestas es muy alta. Para manejar este problema, usamos el algoritmo AdaBoost. Este algoritmo permite elegir solo aquellas características, que representan el segmento dado de la imagen. Básicamente, se trata de un algoritmo de aprendizaje supervisado por máquina, este es adaptativo en el sentido de que los clasificadores posteriores construidos se modifican a favor de aquellas instancias mal clasificadas por clasificadores anteriores.

200	200	100	100	200	200	100	100
250	250	50	50	250	250	50	50
255	255	255	255	100	100	100	100
255	255	255	255	100	100	100	100
200	200	100	100	200	200	100	100
250	250	50	50	250	250	50	50
255	255	255	255	100	100	200	200
255	255	255	255	100	100	250	250

Ilustración 23: Representación haar (filtro)

Debido a que la dimensión de la característica de Haar generada es enorme, para calcular la característica rápidamente, el mapa integral se usa para dividir la celda de la imagen, y se establece la imagen en escala de grises de las coordenadas bidimensionales, y los valores de cada punto de píxel son rápidamente ubicado, por lo tanto, se asegura que la suma de píxeles de la región de la imagen se calcule en el estado eficiente. En el cálculo del gráfico integral, la esquina superior izquierda de la celda de la imagen se toma como punto de partida. Cualquier punto de la imagen hasta el punto de inicio puede formar un rectángulo y se puede calcular la suma de todos los píxeles del rectángulo. Luego, los píxeles se guardan como píxeles de gráficos integrales de esta región. Por lo tanto, cuando necesitamos extraer los píxeles de una determinada región, podemos obtenerlos directamente.

5.1.4. VIOLA-JONES CON CARACTERÍSTICAS COMPUESTAS

El algoritmo de Viola-Jones tiene el problema de que los objetos rígidos reducen la tasa de reconocimiento. Para solucionar este problema, LU Wen-yao y YANG Ming de la Universidad integral de Chongqing proponen un método de detección de rostros basado en el algoritmo de Viola-Jones utilizando características compuestas (W. LU & M. YANG, 2019). El proceso principal de este método es el siguiente:

Se ingresa una imagen y determina un marco rectangular del rostro humano mediante el algoritmo Viola-Jones;

Después de calibrar las caras dentro del marco rectangular, se procesan en cuatro tipos de sub-imágenes. Como se muestra en la Ilustración 24, X^H es la cara calibrada dentro del

rectángulo, X^{ENM} , X^{SEG} es la sub-imagen obtenida cortando parte del rostro, X^{ENM} es la sub-imagen compuesta por las principales características de los ojos, nariz y boca, X^{SEG} son las sub-imágenes divididas en cuatro partes y X^{FSDD} , X^{IVS} es una sub-imagen obtenida por seleccionar píxeles útiles en función del análisis discriminante.

Se utiliza NLDA (análisis discriminante lineal de espacio cero) para extraer características de la imagen de cara completa obtenida y cuatro sub-imágenes.

Se Evalúa la validez de todas las características extraídas (características globales y locales) por distancia discriminante.

Selecciona las regiones de características con una gran distancia de los valores discriminantes para formar nuevos vectores de características compuestos y luego se introducen en el clasificador para el reconocimiento facial.

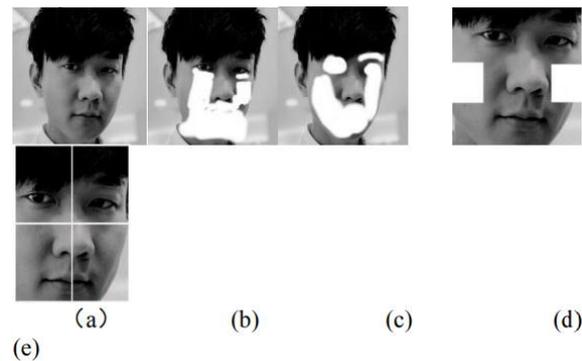


Ilustración 24: (a) XH, (b) XIVS, (c) XFSDD, (d) XENM, (e) XSEG

5.1.5. CONSTRUYENDO NUEVAS CARACTERÍSTICAS COMPUESTAS

El vector de proyección de la cara completa (X_h) y la sub-imagen

$$X^L, \in \{ IVS, ENM, FSDD, SEG \}$$

Puede ser obtenido utilizando NLDA (Zero Space Linear Discriminant Analysis).

Ellos son $W_{Opt}^H \in R^{m \times C-1}$, $W_{Opt}^L \in R^{m \times C-1}$ respectivamente.

Luego de ser obtenido el vector de proyección se evalúa la validez de las características que se obtuvieron como se mencionó anteriormente

El clasificador compuesto probado por LU Wen-yao y YANG Ming da datos más precisos a la hora del reconocimiento facial, usando Fddb obtienen un conjunto de datos en esta prueba. Este conjunto de datos contiene 2.845 imágenes con un total de 5.171 caras. Las imágenes faciales en este conjunto de datos tienen cambios de ángulo, iluminación y escena. En la prueba, seleccionaron al azar 100 imágenes del conjunto de datos Fddb (Face Detection Data Set and Benchmark) para la detección de rostros. Y contaron el número de rostros realmente incluidos en las 100 imágenes seleccionadas, el número de detección .falsa y detección perdida. Sobre esta base, se realizaron un total de 10 pruebas y se utilizaron 1000 imágenes como datos experimentales.

La Ilustración 25, Ilustración 26 e Ilustración 27, como se muestra a continuación, son los resultados experimentales específicos en el conjunto de datos Fddb del algoritmo Viola-Jones original y la aplicación del método de características compuestas realizado por LU Wen-yao y YANG Ming.

Numero de Prueba	Numero de Rostro real	Numero de rostros no encontrados	Numero de rostros falsos	porcentaje de rostros no encontrados	porcentaje de error de deteccion
1	126	16	3	0,127	0,024
2	150	34	3	0,227	0,02
3	128	23	1	0,18	0,008
4	134	19	0	0,142	0
5	152	22	1	0,145	0,007
6	153	16	4	0,105	0,026
7	131	18	4	0,137	0,031
8	126	14	1	0,111	0,008
9	147	18	2	0,122	0,014
10	125	9	2	0,072	0,016

Ilustración 25: Resultados experimentales de Viola-Jones original

Numero de Prueba	Numero de Rostro real	Numero de rostros no encontrados	Numero de rostros falsos	procentaje de rostros no encontrados	porcentaje de error de deteccion
1	126	5	1	0,04	0,008
2	150	8	0	0,053	0
3	128	8	0	0,063	0
4	134	6	0	0,044	0
5	152	10	0	0,065	0
6	153	10	0	0,065	0
7	131	7	0	0,053	0
8	126	8	0	0,063	0
9	147	9	0	0,061	0
10	125	4	0	0,032	0

Ilustración 26: Experimentos de uso del método Viola-Jones basado en características compuestas en un conjunto de datos Fddb

Algoritmo	Total Numero de Rostro real	Total Numero de rostros no encontrados	Total Numero de rostros falsos	Total procentaje de rostros no encontrados	Total porcentaje de error de deteccion
Viola-Jones	1372	189	21	0,138	0,015
Viola-Jones Compuesto		75	1	0,055	0,0007

Ilustración 27: Comparación completa viola-jones

5.1.6. ANÁLISIS FINAL

Al analizar los resultados experimentales en la sección anterior, podemos saber que el algoritmo original de Viola-Jones tiene muchas detecciones de caras falsas y pérdidas en el conjunto de datos de Fddb. Sin embargo, el método utilizado por LU Wen-yao y YANG Ming es obviamente más bajo que el algoritmo de Viola-Jones en términos de tasa de detección de pérdida y tasa de detección falsa. A partir de esto, podemos ver que la combinación del algoritmo de Viola-Jones y las características compuestas construidas evita la detección falsa y pérdida de rostro original de manera efectiva, lo que constituye una buena base para garantizar la precisión del reconocimiento facial.

5.2. ALGORITMO EAR (EYE ASPECT RATIO ALGORITHM) CON DETECCIÓN DE PARPADEO EN TIEMPO REAL MEDIANTE MARCAS FACIALES.

Tereza Soukupova y Jan Čech, Ingenieros de la Universidad Técnica Checa en Praga propusieron un algoritmo en tiempo real para detectar parpadeos en una secuencia de video de una cámara estándar. Los puntos de referencia y los conjuntos de datos exhiben una excelente robustez frente a la orientación de la cabeza con respecto a una cámara, iluminación variable y expresiones faciales. Los ingenieros muestran que los puntos de referencia se detectan con la suficiente precisión para estimar de manera confiable el nivel de apertura de los ojos. Por lo tanto, el algoritmo que propusieron estima las posiciones de los puntos de referencia, extrae una sola cantidad escalar y la relación de aspecto del ojo (EAR), que caracteriza la apertura del ojo en cada cuadro. Finalmente, un clasificador SVM detecta parpadeos como un patrón de valores EAR en una ventana temporal

Otros métodos de detección por lo general, imponen implícitamente requisitos demasiado estrictos en la configuración, en el sentido de una pose relativa de cara a cámara (orientación de la cabeza), resolución de imagen, iluminación, dinámica de movimiento, etc. Como es el caso de Viola Jones que hablamos en capítulos anteriores de esta monografía.

Sin embargo, hoy en día, los detectores de puntos faciales son más robustos y capturan la mayoría de los puntos característicos de una imagen de rostro humano, incluidas las esquinas de los ojos y los párpados. La mayoría de los detectores de puntos (Landmarks) de última generación formulan un problema de regresión, donde hay mapeo de una imagen en posiciones de puntos de referencia (**X. Xiong & F. De la Torre, 2013**) o parametrización de puntos de referencia (**A. Asthana, S. Zafeoriou, S. Cheng, & M. Pantic, 2014**). Estos detectores de puntos de referencia modernos están entrenados en "conjuntos de datos naturales (luminosidad, oscuridad, etc...)" y, por lo tanto, son resistentes a la iluminación variable, las diversas expresiones faciales y las rotaciones moderadas de la cabeza no frontales. Una media de error de la localización de los puntos de referencia de un detector de última generación suele ser inferior al

cinco por ciento de la distancia inter-ocular. Los métodos recientes se ejecutan incluso significativamente en tiempo súper real (**S. Ren, X. Cao, Y. Wei, & J. Sun, 2014**).

Partiendo de esto crean el algoritmo EAR, un algoritmo simple pero eficiente para detectar parpadeos mediante el uso de un detector de marcas faciales. Con una única cantidad escalar que refleja un nivel de apertura del ojo y se deriva de los puntos de referencia. Finalmente se tiene un fotograma (per-frame) de las estimaciones de apertura de ojos, los parpadeos se encuentran mediante un clasificador SVM que está entrenado con ejemplos de patrones de parpadeo y no parpadeo.

Existe otro modelo de segmentación facial presentado en (**F. M. Sukno, S.-K. Pavani, C. Butakoff, & A. F. Frangi, 2009**) que es similar al método propuesto. Sin embargo, su sistema se basa en modelos de formas activas con un tiempo de procesamiento de aproximadamente 5 segundos por fotograma para la segmentación, y la señal de apertura del ojo se normaliza mediante estadísticas estimadas al observar una secuencia más larga. Por lo tanto, el sistema sólo se puede utilizar para el procesamiento normal. El algoritmo EAR saca ventaja de esto ya que se ejecuta en tiempo real y los tiempos de procesamiento son mínimos.

5.2.1. FUNCIONAMIENTO DEL ALGORITMO EAR

El parpadeo es un cierre y una reapertura rápida del ojo humano. Cada individuo tiene un patrón de parpadeo un poco diferente. El patrón se diferencia en la velocidad de cierre y apertura, un grado de apriete del ojo y la duración del parpadeo. El parpadeo dura aproximadamente 100-400 ms. El algoritmo EAR utiliza detectores de puntos de referencia faciales de última generación para localizar los ojos y los contornos de los párpados. A partir de los puntos de referencia detectados en la imagen, obtienen la relación de aspecto del ojo (EAR) que se utiliza como una estimación del estado de apertura del ojo. Dado que es posible que el per-frame EAR no reconozca necesariamente que el ojo parpadea correctamente, entrenan un clasificador que tiene en cuenta una ventana temporal más grande de un frame.

5.2.1.1. DESCRIPCIÓN DE CARACTERÍSTICAS

Para cada cuadro de video, se detectan los puntos de referencia del ojo. Se calcula la relación de aspecto del ojo (EAR) entre la altura y el ancho del ojo.

$$\text{EAR} = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|},$$

Ecuación 3

Donde P1 hasta P6 son los puntos de referencia del ojo.

EAR es mayormente constante cuando un ojo está abierto y se acerca a cero al cerrar un ojo. Es parcialmente insensible a la persona y la cabeza. La relación de aspecto del ojo abierto tiene una pequeña variación entre los individuos y es completamente invariante a una escala uniforme de la imagen y la rotación en el plano de la cara. Dado que el parpadeo lo realizan ambos ojos de forma sincrónica, se promedia el EAR de ambos ojos. La ilustración 29 e Ilustración 33 muestran un ejemplo de una señal EAR sobre la secuencia de video.

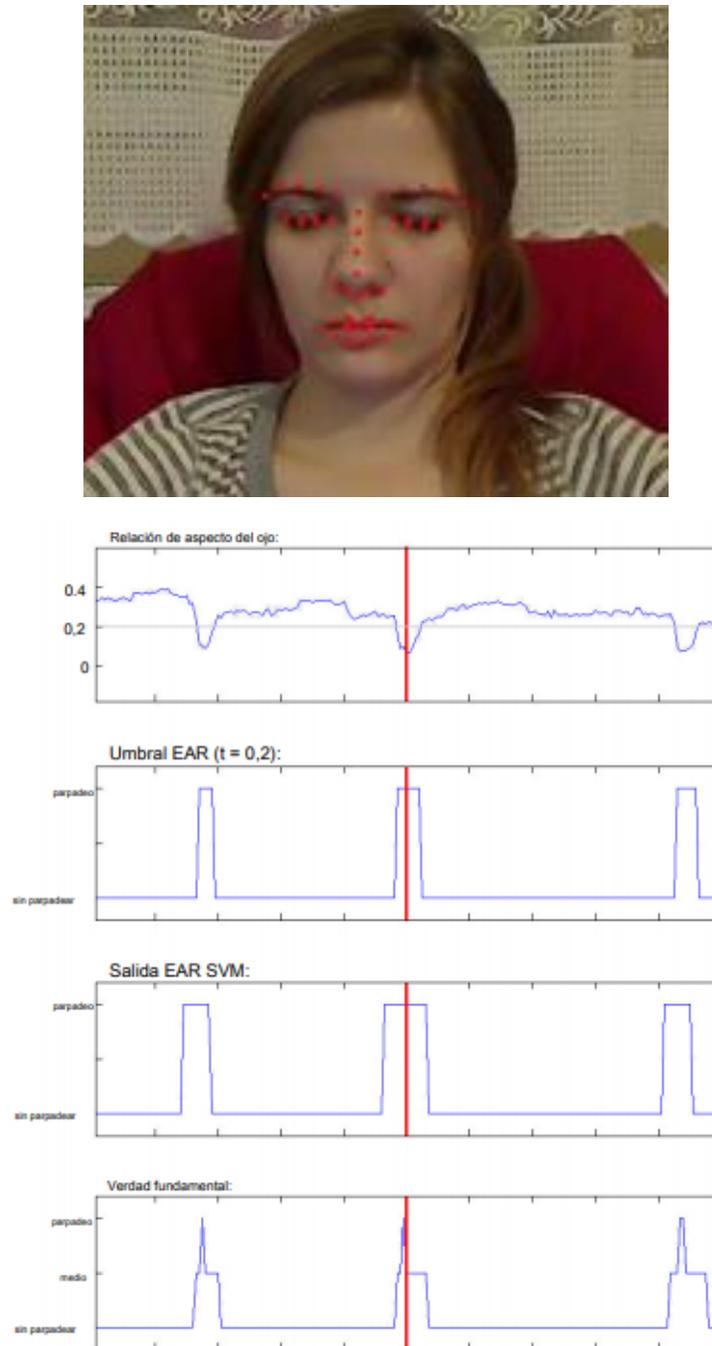


Ilustración 28: Ejemplo de parpadeos detectados. Las gráficas de la relación de aspecto del ojo (EAR) en Ecuación 3, resultados del umbral de EAR (umbral establecido en 0,2), los parpadeos detectados por EAR SVM y las etiquetas sobre la secuencia de vídeo (verdad fundamental o Ground-truth). Imagen de entrada con puntos de referencia detectados (el frame está representado con una línea roja)

5.2.1.2. CLASIFICACIÓN

El algoritmo EAR tiene un clasificador que toma como entrada una ventana temporal más grande de un frame. Para los videos de 30 fps, descubrieron experimentalmente que ± 6 fotogramas pueden tener un impacto significativo en la detección de parpadeo para un fotograma en el que un ojo está más cerrado al parpadear.

Por lo tanto, para cada fotograma, se recopila una característica de 13 dimensiones concatenando los EAR de sus ± 6 fotogramas vecinos.

Esto se implementa mediante un clasificador SVM lineal (llamado EAR SVM) entrenado a partir de secuencias anotadas manualmente. Los ejemplos positivos se recopilan como parpadeos de verdad, mientras que los negativos son aquellos que se muestrean de partes de los videos donde no se produce ningún parpadeo, con un espaciado de 5 fotogramas y un margen de 7 fotogramas desde los parpadeos de verdad. Durante la prueba, ejecutan un clasificador en forma de ventana de exploración. EAR SVM calcula y clasifica una característica de 13 dimensiones para cada fotograma, excepto el principio y el final de una secuencia de vídeo.

5.2.1.3. PRUEBAS

Tereza Soukupova y Jan Čech llevaron a cabo dos tipos de prueba: La prueba que mide la precisión de los detectores de hitos y las pruebas que evalúan el rendimiento del algoritmo de detección de parpadeo de ojos completos.

5.2.1.4. PRECISIÓN DE LOS DETECTORES DE PUNTOS DE REFERENCIA

Para evaluar la precisión de los detectores de puntos probados, utilizaron el conjunto de datos 300-VW (S. Zafeiriou, G. Tzimiropoulos, & M. Pantic, 2015). Es un conjunto de datos que contiene 50 videos donde cada cuadro tiene asociada una anotación precisa de puntos de referencia faciales. Los videos son "en la naturaleza", en su mayoría grabados desde un televisor. El propósito de las pruebas era demostrar que los detectores de puntos de referencia actuales son particularmente robustos y precisos para detectar ojos, es decir, las esquinas de los ojos y el contorno de los párpados. Por lo tanto, prepararon un conjunto de datos, un subconjunto del 300-VW, que contiene imágenes de muestra con los ojos

abiertos y cerrados. Más precisamente, teniendo la anotación del punto de referencia de la verdad fundamental, ordenaron los fotogramas para cada sujeto por la relación de aspecto del ojo (EAR en la **Ecuación 3**) y tomaron 10 fotogramas de la relación más alta (ojos bien abiertos), 10 fotogramas de la proporción más baja (en su mayoría ojos bien cerrados) y 10 fotogramas muestreados al azar. Así recopilaron 1500 imágenes. Además, todas las imágenes fueron sub-muestreadas posteriormente (sucesivamente 10 veces por factor 0,75) para evaluar la precisión de los detectores probados en imágenes de caras pequeñas. Probaron dos detectores de puntos de última generación: Chehra (A. Asthana, S. Zafeoriou, S. Cheng, & M. Pantic, 2014) e Intraface (X. Xiong & F. De la Torre, 2013). Ambos se ejecutan en tiempo real. Las muestras del conjunto de datos se muestran en la Fig. 3. Hay que tener en cuenta que los rostros no siempre son frontales a la cámara, la expresión no siempre es neutral, las personas a menudo hablan emocionalmente o sonríen, etc. A veces las personas usan anteojos, el cabello ocasionalmente puede ocluir parcialmente uno de los ojos. Ambos detectores funcionan generalmente bien, pero el Intraface es más robusto para imágenes de caras muy pequeñas, como se muestra en la Ilustración 29.

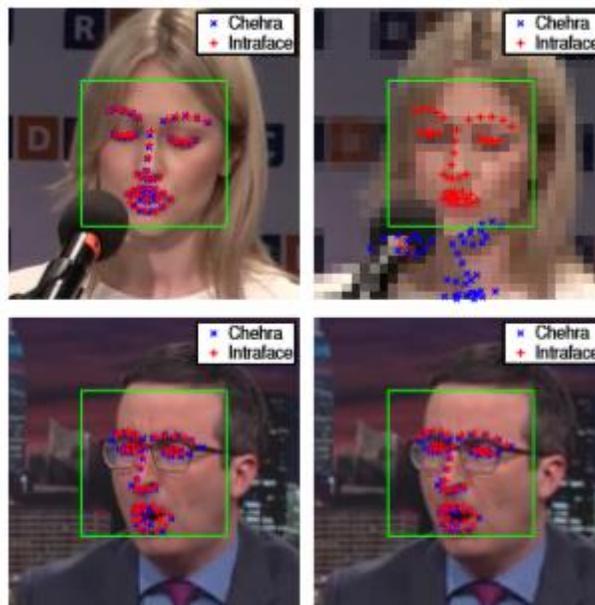


Ilustración 29: Imágenes de ejemplo del conjunto de datos 300-VW con puntos de referencia obtenidos por Chehra (A. Asthana, S. Zafeoriou, S. Cheng, & M. Pantic, 2014) e Intraface (X. Xiong & F. De la Torre, 2013). Imágenes originales (izquierda) con distancia inter-ocular (IOD) igual a 63 (arriba) y 53 (abajo) píxeles. Imágenes sub-muestreadas (derecha) a IOD igual a 6,3 (arriba) y 17 (abajo)

Cuantitativamente, la precisión de la detección de puntos de referencia para una imagen facial se mide por el error de localización de puntos de referencia relativa promedio, definido normalmente así:

$$\epsilon = \frac{100}{\kappa N} \sum_{i=1}^N \|x_i - \hat{x}_i\|_2,$$

Ecuación 4

Donde X_i es la ubicación real del punto de referencia (verdad fundamental) i en la imagen, \hat{X} es una ubicación de puntos de referencia estimados por un detector, N es el número de puntos de referencia y factor de normalización, es decir, la distancia interocular (IOD), Distancia euclidiana entre los centros del ojo en la imagen.

En primer lugar, se calcula un histograma acumulativo estándar del error medio de localización de puntos de referencia relativos, Ilustración 30, para un conjunto completo de 49 puntos de referencia y un subconjunto de 12 puntos de referencia de los ojos solamente, ya que estos puntos de referencia se utilizan en el detector de parpadeo. Los resultados se calculan para todas las imágenes originales que tienen un IOD promedio de alrededor de 80 px, y también para todas las imágenes de caras "pequeñas" (incluidas las submuestreadas) que tienen un $IOD \leq 50$ px. Para todos los puntos de referencia,

Chehra tiene más casos de errores muy pequeños (hasta el 5 por ciento del IOD), pero el Intraface tiene más casos de errores por debajo del 10% del IOD. Solo para puntos de referencia del ojo,

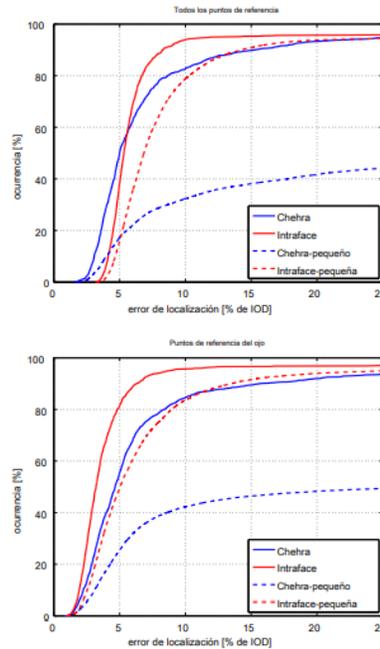


Ilustración 30: Histograma acumulativo del error de localización promedio de los 49 puntos de referencia (arriba) y los 12 puntos de referencia de los ojos (abajo). Los histogramas son calculados para imágenes de resolución original (líneas continuas) y un subconjunto de imágenes pequeñas (IOD ≤ 50 px)

el Intrafase es siempre más preciso que Chehra. Como ya se mencionó, el Intrafase es mucho más robusto para imágenes pequeñas que Chehra. Este comportamiento se observa además en la siguiente prueba que hicieron los ingenieros de la universidad de Praga.

Toman un conjunto de las 1500 imágenes, miden un error de localización medio μ en función de la resolución de la imagen facial determinada por el IOD. Precisamente: $\mu = \frac{1}{|S|} \sum_{j \in S} \epsilon_j$, el error medio sobre el conjunto de imágenes faciales S que tienen el IOD en un rango dado. Los resultados se muestran en la Ilustración 31. Los gráficos tienen barras de error de desviación estándar. Se ve que Chehra falla rápidamente para imágenes con IOD < 20 px. Para caras más grandes, el error medio es comparable, aunque ligeramente mejor en Intrafase para los puntos de referencia del ojo.

La última prueba está directamente relacionada con el detector de parpadeo. Midieron la precisión de EAR en función del IOD. El error EAR medio se define como una diferencia media absoluta entre el EAR verdadero y el estimado. Las gráficas se calculan para dos subconjuntos en proporción verdadera promedio 0.05 ± 0.05 y ojos abiertos (relación real media 0.4 ± 0.1). El error es mayor para los ojos cerrados. La razón es que es más probable que ambos detectores muestren ojos abiertos en caso de falla. Se ve que el error de relación para IOD < 20 px causa una gran confusión

entre los estados de ojos abiertos / cerrados para Chehra, sin embargo, para caras más grandes, la proporción se estima con la suficiente precisión para garantizar una detección confiable del parpadeo.

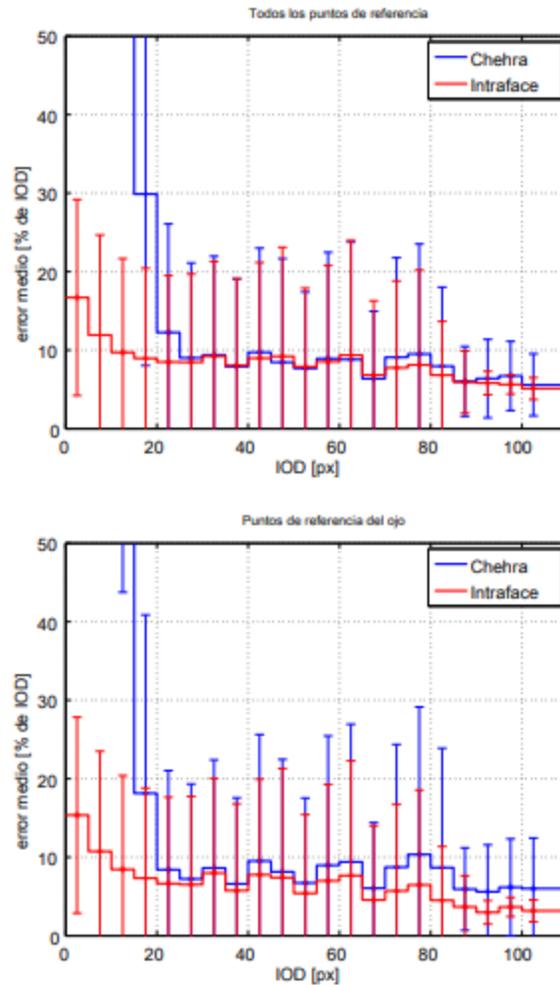


Ilustración 31: Precisión de localización de puntos de referencia en función de la resolución de la imagen de la cara calculada para todos los puntos de referencia y puntos de referencia del ojo únicamente

5.2.1.5. EVALUACIÓN DEL DETECTOR DE PARPADO

El algoritmo EAR es evaluado en dos bases de datos estándar con anotaciones de parpadeos de verdad fundamental (ground-truth). El primero es ZJU (G. Pan, L. Sun, Z. Wu, & S. Lao, 2007) que consta de 80 videos cortos de 20 sujetos. Cada sujeto tiene 4 videos: 2 con y 2 sin lentes, 3 videos son frontales y 1 es una vista hacia arriba. Los videos de 30fps son

de tamaño 320×240 px. La duración promedio de un video es de 136 cuadros y contiene aproximadamente 3.6 parpadeos en promedio. Un IOD promedio es de 57,4 píxeles. En esta base de datos, los sujetos no realizan ninguna expresión facial notable. Miran directamente a la cámara a corta distancia, casi no se mueven, no sonríen ni hablan. Un parpadeo ground-truth (verdad fundamental) se define por su fotograma inicial, fotograma máximo y fotograma final.

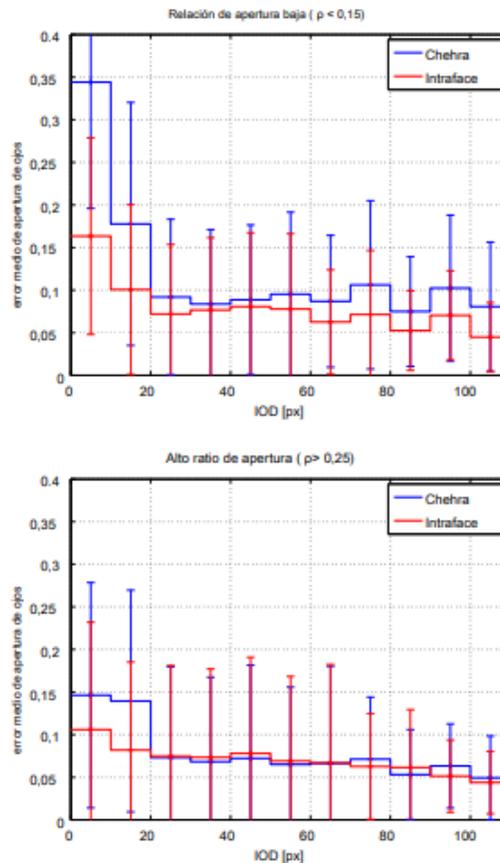


Ilustración 32: Precisión del radio de apertura del ojo en función de la resolución de la imagen de la cara. Arriba: para imágenes con un radio verdadero bajo (en su mayoría ojos cerrados o medio cerrados) e inferior: imágenes con una proporción más alta (ojos abiertos)

La segunda base de datos Eyeblick8 (T. Drutarovsky & A. Fogelton, 2014) es más desafiante. Consiste en 8 videos largos de 4 sujetos que están sonriendo, girando la cabeza de forma natural, cubriendo la cara con las manos, bostezando, bebiendo y mirando hacia abajo en un teclado. Estos videos tienen una duración de 5k a 11k fotogramas, también 30 fps, con una resolución de 640×480 píxeles y un IOD promedio de 62,9 píxeles.

Contienen alrededor de 50 parpadeos en promedio por video. Cada cuadro que pertenece a un parpadeo está anotado por el estado de los ojos medio abierto o cerrado. Se consideran medio parpadeos a los que no alcanzan el estado de cierre, como parpadeos completos para ser coherentes con el ZJU. Allí también probaron los métodos EAR SVM, que están entrenados para detectar el patrón de parpadeo específico, los compararon con un método de línea de base simple, que solo establece el umbral de la EAR en la Ecuación 3. Los clasificadores EAR SVM los probaron con detectores de puntos de referencia Chehra (A. Asthana, S. Zafeoriou, S. Cheng, & M. Pantic, 2014) e Intrafase (X. Xiong & F. De la Torre, 2013).

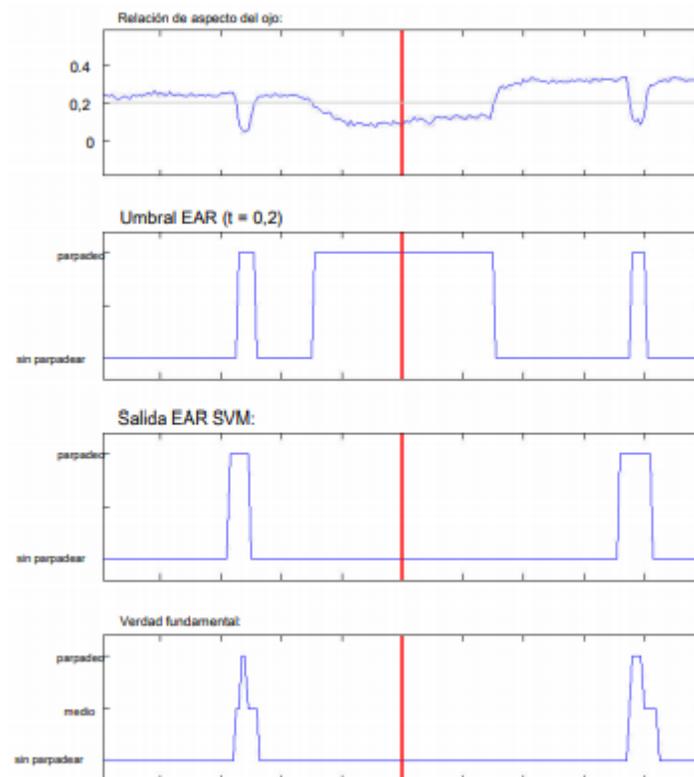


Ilustración 33: Ejemplo de parpadeos detectados donde falla el umbral de EAR mientras que EAR SVM tiene éxito. Las gráficas de la relación de aspecto del ojo EAR en Ecuación 3, resultados del umbral de EAR (umbral establecido en 0,2), los parpadeos detectados por EAR SVM y las etiquetas de verdad sobre el terreno sobre la secuencia de vídeo. Imagen de entrada con puntos de referencia detectados. La representación del frame está marcada con una línea roja)

El experimento con EAR SVM se realiza en forma de conjuntos de datos cruzados. Significa que el clasificador SVM está entrenado en el Eyeblink8 y probado en el ZJU y viceversa.

Para evaluar la exactitud de la detección, los parpadeos predichos se comparan con los parpadeos reales. El número de verdaderos positivos se determina como un número de parpadeos de verdad (ground truth) que tienen una sección no vacía con parpadeos detectados.

El número de falsos negativos se cuenta como un número de parpadeos de “ground truth” que no se cruzan con los parpadeos detectados. El número de falsos positivos es igual al número de parpadeos detectados menos el número de verdaderos positivos más un costo por detectar parpadeos demasiado largos. El costo se cuenta solo para detectar parpadeos dos veces más largos que un parpadeo promedio de longitud A . Cada parpadeo largo de longitud L se cuenta $\frac{L}{A}$ veces como un falso positivo. El número de todos los parpadeos posiblemente detectables se calcula como el número de fotogramas de una secuencia de vídeo dividido por la duración media del parpadeo del sujeto (**T. Drutarovsky & A. Fogelton, 2014**). La base de datos de ZJU en su mayoría sostiene que cada cierre de ojos es un parpadeo. En consecuencia, las curvas de recuperación de precisión que se muestran en la Ilustración 34 (a) del umbral de EAR y ambos clasificadores de EAR SVM son casi idénticas. Estas curvas se calcularon abarcando un umbral de la puntuación de salida de EAR y SVM respectivamente.

Las curvas de recuperación de precisión en la Ilustración 34 (b) muestran la evaluación en la base de datos Eyeblink8. Se observa que en esta base de datos, el umbral de EAR va por detrás de ambos clasificadores EAR SVM. El umbral falla cuando un sujeto sonrío (tiene los ojos entrecerrados; vea un ejemplo en la Ilustración 33, tiene una vista lateral o cuando el sujeto cierra los ojos por un tiempo mayor que un parpadeo. Ambos detectores de SVM funcionan mucho mejor, el SVM basado en el detector intrafase es incluso un poco mejor que el Chehra SVM. Ambos detectores

EAR SVM superan al método de Drutarovsky y Fogelton (T. Drutarovsky & A. Fogelton, 2014) por un margen significativo. Finalmente, la precisión total del detector de parpadeo en el IOD promedio sobre el conjunto de datos fue medida por los ingenieros. Cada fotograma de la base de datos ZJU fue sub-muestreado al 90%, 80%,..., 10% de su resolución original. Para la evaluación se utilizaron tanto Chehra-SVM como Intrafase-SVM. Para cada resolución, fue calculada el área bajo la curva de recuperación de precisión (AUC). El resultado se muestra en la Ilustración 35. Podemos ver que con los puntos de referencia de Chehra la precisión permanece muy alta hasta que la IOD promedio es de aproximadamente 30px. El detector falla en imágenes con el IOD < 20px. Los puntos de referencia intrafase son mucho mejores en bajas resoluciones.

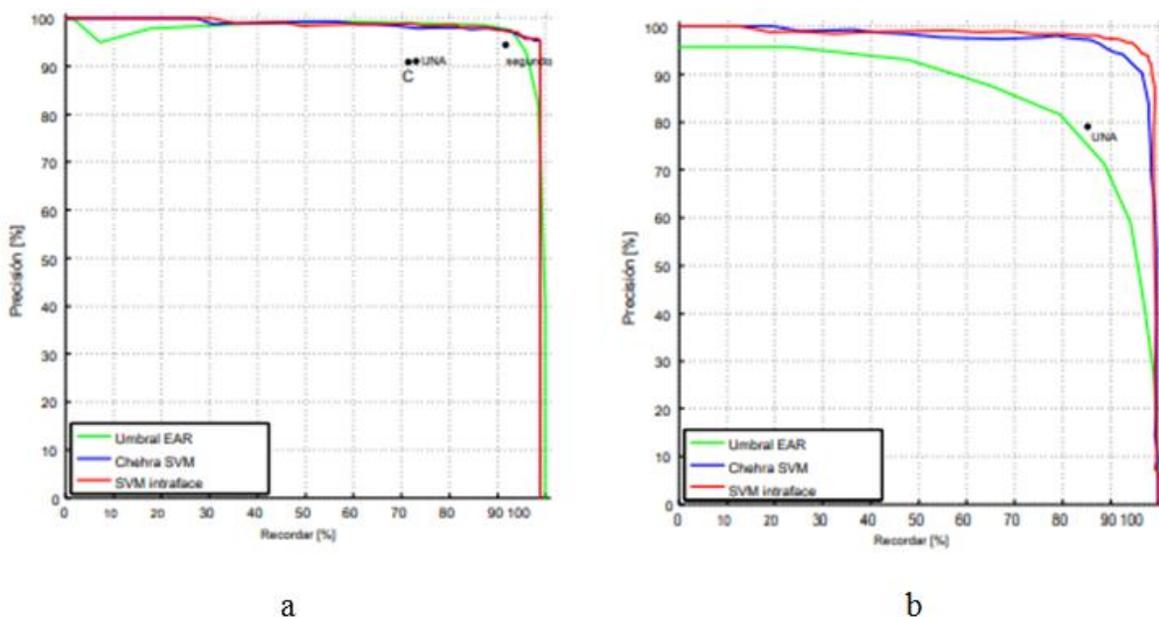


Ilustración 34: Curvas de recuperación de precisión de los clasificadores de umbral EAR y EAR SVM medidos en (a) la ZJU y (b) las bases de datos Eyeblink8. Resultados publicados de los métodos A - Drutarovsky y Fogelton (T. Drutarovsky & A. Fogelton, 2014), B - Lee y col. (W. H. Lee, E. C. Lee, & K. E. Park, 2010), C - Danisman y col. (T. Danisman, I. Bilasco, C. Djeraba, & N. Ihaddadene, 2010)

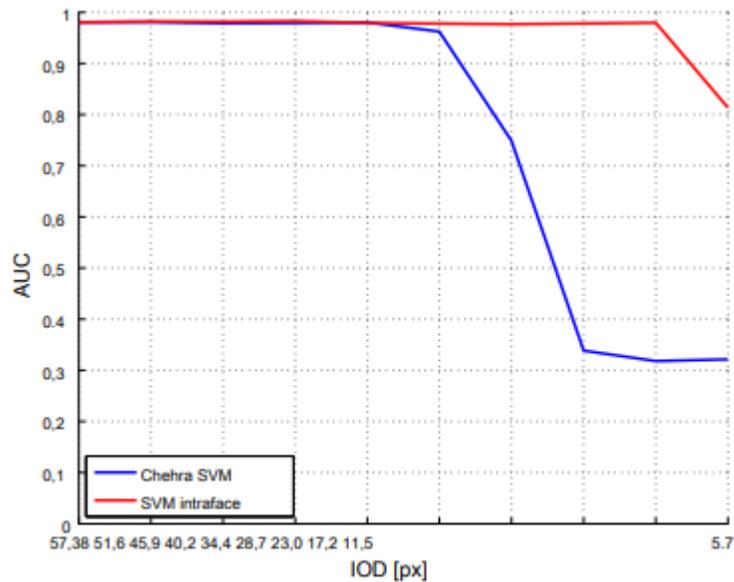


Ilustración 35: Precisión del detector de parpadeo del ojo (medida por AUC) en función de la resolución de la imagen (IOD promedio) al sub-muestrear el conjunto de datos ZJU

5.2.2. ANÁLISIS FINAL

El algoritmo EAR implementa detectores de puntos de referencia faciales (landmarks) basados en regresión y son lo suficientemente precisos como para estimar de manera confiable un nivel de apertura ocular. Si bien son robustos a baja calidad de imagen (baja resolución de imagen) y en la naturaleza, fenómenos como no frontalidad, mala iluminación, expresiones faciales, etc. Logra la técnica en dos conjuntos de datos estándar utilizando el detector de puntos de referencia robusto seguido de una detección simple de parpadeo basada en el SVM. El algoritmo se ejecuta en tiempo real, ya que los costos computacionales adicionales para la detección del parpadeo son insignificantes además de los detectores de puntos de referencia en tiempo real. El método SVM propuesto que utiliza una ventana temporal de la relación de aspecto del ojo (radio de apertura) (EAR), supera el umbral de EAR. Por otro lado, el umbral se puede utilizar como un clasificador de imagen única para detectar el estado del ojo, en caso de que no se disponga de una secuencia más larga. Vemos una limitación de que se asumió una duración de parpadeo fija para todos los sujetos, aunque el parpadeo de todos dura de manera diferente. Los resultados podrían mejorarse con un enfoque adaptativo. Otra limitación está en la estimación de apertura de ojos. Si bien la EAR se estima a partir de una imagen 2D, es bastante insensible a la orientación de la cabeza, pero puede perder capacidad de discriminación para rotaciones fuera del plano. Una solución podría ser definir

el EAR en 3D. Hay detectores de puntos de referencia que estiman una pose 3D (posición y orientación) de un modelo 3D de puntos de referencia, por ejemplo [(**A. Asthana, S. Zafeoriou, S. Cheng, & M. Pantic, 2014**), (**J. Cech, V. Franc, & J. Matas, 2014**)].

PARTE IV CONCLUSIONES Y BIBLIOGRAFIA.

CAPITULO 6: CONCLUSIONES

Este capítulo se creó con el fin de responder evaluar y verificar los interrogantes, hipótesis y objetivos que se propusieron:

1. Primer interrogante: ¿Cómo afecta la fatiga y la somnolencia a los conductores mientras conducen?

Como se planteó en el capítulo 3 la fatiga y somnolencia puede deberse a gran variedad de factores, cansancio acumulado, estar demasiado tiempo frente al volante, consumo de alcohol o alucinógenos, pereza o también el consumo de medicamentos que producen el sueño etc... Y sus consecuencias afectan:

- a. Tiempo de reacción
- b. concentración y genera muchas distracciones
- c. Toma de decisiones erróneas y de manera lenta
- d. Perturbaciones motoras
- e. Microsueños
- f. Perturbaciones sensoriales
- g. Cambios comportamentales

Primera hipótesis: La fatiga y la somnolencia se presentan en el 3% de los accidentes fatales y el 2.5% de los accidentes no fatales.

Podemos ver en el capítulo 2 que se citan unas estadísticas de la La National Highway Traffic Safety Administration, esta calcula que el 2,5% de los accidentes fatales y el 2% de los accidentes no fatales se relacionan con la presencia de somnolencia, así que con se puede concluir que la hipótesis estuvo cerca de ser la correcta, ya que tiene cifras cercanas a las de las estadísticas planteadas.

2. Segundo interrogante: ¿Existen sistemas informáticos que permitan el reconocimiento facial de fatiga y somnolencia en conductores?

Como se muestra en el capítulo 4, evidenciamos que existen sistemas para el reconocimiento facial enfocados en el estado de los ojos para detectar la somnolencia en conductores.

Segunda hipótesis: Existen sistemas informáticos que reconocen la fatiga y la somnolencia en conductores a partir de visión computacional.

Como lo vimos en el capítulo 4, la visión computacional hace parte del sistema para la detección de somnolencia. Cabe destacar que la cámara no detecta la somnolencia, es el sistema programado que a partir de los datos transmitidos por la cámara evalúa si se cumple o no con el estado de somnolencia.

3. Tercer interrogante: ¿Qué algoritmos se implementan en los sistemas informáticos que reconocen gestos faciales de fatiga y somnolencia?

En el capítulo 5, se hizo una búsqueda y se encontró que los algoritmos comúnmente usados son el algoritmo Viola-Jones y el algoritmo EAR.

4. Cuarto interrogante ¿Qué tan eficaces son los algoritmos al momento del procesamiento de imágenes para el reconocimiento de fatiga y somnolencia?

Como lo vimos en los capítulos 4 y 5, se encontró que el algoritmo de Viola-Jones compuesto es bueno en escenarios con condiciones favorables (Luminosidad, Contraste, balance de blancos... etc.), ya que en estas condiciones el umbral de error es bajo como se muestra en la Ilustración 27: Comparación completa viola-jones, pero en condiciones no favorables, como poca luminosidad o accesorios como gafas formuladas, el algoritmo EAR destaca resultados más eficientes.

Tercera hipótesis: Los algoritmos que procesan imágenes para el reconocimiento de fatiga y somnolencia, identifican correctamente patrones de fatiga y somnolencia en imágenes en escenarios con datos favorables (Luminosidad, Contraste, balance de blancos etc...), mientras que en escenarios reales tienen una tasa más baja de reconocimiento ya que influyen variables como el clima, el movimiento del conductor etc...

Como lo vimos en los capítulos 4 y 5, se encontró que existen algoritmos que son eficientes no solo en condiciones favorables sino también en condiciones no favorables como lo es el algoritmo EAR.

BIBLIOGRAFIA

- A. Asthana, S. Zafeoriou, S. Cheng, & M. Pantic. (2014). Incremental face alignment in the wild. *In Conference on Computer Vision and Pattern Recognition*, (págs. 1, 2, 3, 4, 5, 7).
- Antúnez Galindo, M. (29 de 1 de 2019). *UPCommons*. Recuperado el 16 de 10 de 2020, de <https://upcommons.upc.edu/handle/2117/129222>
- Bruno Pedraz-Petrozzi. (17 de 09 de 2018). Recuperado el 28 de 10 de 2020, de <http://www.scielo.org.pe/pdf/rnp/v81n3/a05v81n3.pdf>
- Buenavides Muñoz, E. L., & Medina Mendez, M. M. (27 de 1 de 2018). *Repositorio Universidad de Córdoba* . Recuperado el 16 de 10 de 2020, de <https://repositorio.unicordoba.edu.co/handle/ucordoba/526>
- Cech, J. (2016). En J. Cech, *Real-Time Eye Blink Detection using Facial Landmarks*.
- Crespin Luis, J. C., & Raúl Alexander , J. G. (2019). *Universidad Nacional de Trujillo*. Recuperado el 16 de 10 de 2020, de <http://dspace.unitru.edu.pe/handle/UNITRU/15096>
- De La Hoz, A. (10 de 11 de 2019). *El heraldo*. Recuperado el 21 de 10 de 2020, de <https://www.elheraldo.co/barranquilla/microsueno-el-enemigo-oculto-de-los-conductores-679276>
- Domínguez Soberanes, J., Manuel Álvarez Pato, V., Sánchez Gómez, C. N., Gutiérrez Calderón, J. S., & Velázquez Guerrero, R. (21 de 10 de 2019). *Tecnologico Nacial de Mexico*. Recuperado el 16 de 10 de 2020, de <http://www.itc.mx/ojs/index.php/pistas/article/download/2091/1624>
- europapress.tv. (14 de 03 de 2019). *europapress.tv*. Recuperado el 10 de 21 de 2020, de <https://www.europapress.tv/vn/429185/1/encuesta-duermen-espanoles>
- F. M. Sukno, S.-K. Pavani, C. Butakoff, & A. F. Frangi. (2009). Automatic assessment of eye blinking patterns through statistical shape models. *In ICVS*, (págs. 1,2).
- G. Pan, L. Sun, Z. Wu, & S. Lao. (2007). Eyeblink-based anti-spoofing in face recognition from a generic webcam. *In ICCV*, (págs. 1, 2, 5).
- H. Nguyen, & P. Examiner. (2007). (12) *United States Patent vol. 2*. Recuperado el 28 de 10 de 2020

- ideal, S. (1 de 2 de 2018). *¿Qué son los microsueños y cómo prevenirlos?* Recuperado el 16 de 10 de 2020, de <https://salud.ideal.es/lomonaco/que-son-los-micro-suenos>
- J. Cech, V. Franc, & J. Matas. (2014). A 3D approach to facial landmarks: Detection, refinement, and tracking. *In Proc. International Conference on Pattern Recognition*, (pág. 7).
- Kartali, A., Roglić, M., Barjaktarović, M., Đurić-Jovičić, M., & Janković, M. M. (11 de 2018). *IEEE Xplore*. Recuperado el 16 de 10 de 2020, de <https://ieeexplore.ieee.org/document/8587011>
- Krishnamoorthy, V., Shoorangiz, R., Weddell, S. J., Beckert, L., & Jones, R. D. (7 de 2019). *IEEE Xplore*. Recuperado el 16 de 10 de 2020, de <https://ieeexplore.ieee.org/abstract/document/8857588/>
- Lashkov, I., Kashevnik, A., Shilov, N., Parfenov, V., & Shabaev, A. (8 de 2019). *IEEE Xplore*. Recuperado el 16 de 10 de 2020, de <https://ieeexplore.ieee.org/abstract/document/8919515>
- López Peña, A., Valveny, E., & Vanrell, M. (1 de 12 de 2015). *Coursera*. Recuperado el 28 de 9 de 2020, de <https://es.coursera.org/lecture/deteccion-objetos/15-3-imagen-integral-CROQ4>
- M. K. Hasan, M. S. Hossain, T. K. Ghosh, & M. Ahmad. (2015). A SSVEP Based EEG Signal Analysis to Discriminate the Effects of Music Levels. En *A SSVEP* (págs. 27-33).
- M. K. Hasan, R. Z. Rusho, & M. Ahmad. (2010). *.hindawi*. Recuperado el 02 de 11 de 2020, de <https://www.hindawi.com/journals/cin/2010/702357/fig2/>
- P. V., & M. J. (2014). En *Rapid Object Detection Using a Boosted Cascade of Simple Features*.
- P. Viola, & M. Jones. (2004). En P. V. Jones, *Robust Real-time Object Detection*”, *International vol 57* (págs. 137-154).
- P. Viola, & M.J. Jones. (2001). En *Rapid Object Detection using a Boosted Cascade of Simple Features*”, *proceedings IEEE Conf. on Computer Vision and Pattern Recognition*.
- Real Academia De la Lengua Española. (2019). *RAE*. Recuperado el 28 de 10 de 2020, de <https://dle.rae.es/fatiga>
- Rodríguez Baque, K. E. (9 de 2017). *Repositorio Digital Universidad Técnica del Norte*. Recuperado el 28 de 9 de 2020, de

<http://repositorio.utn.edu.ec/bitstream/123456789/7315/1/04%20MEC%20204%20TRABAJO%20DE%20GRADO.pdf>

Rondon Condori, L. A., & Paucara Nuñez, F. J. (2013). *Repositorio Institucional Digital de la Universidad Nacional de San Antonio Abad del Cusco*. Recuperado el 13 de 10 de 2020, de <http://repositorio.unsaac.edu.pe/bitstream/handle/UNSAAC/950/253T20130066.pdf?sequence=1&isAllowed=y>

Rosebrock, A. (10 de 4 de 2017). *pyimagesearch*. Recuperado el 28 de 10 de 2020, de Detect eyes, nose, lips, and jaw with dlib, OpenCV, and Python: <https://www.pyimagesearch.com/2017/04/10/detect-eyes-nose-lips-jaw-dlib-opencv-python/>

Ryan James Kimmel, MD, Medical Director of Hospital Psychiatry at the Uni, Seattle, WA. Also reviewed by David Zieve, MHA, Medical Director, y otros. (7 de 7 de 2019). *medlineplus*. Recuperado el 3 de 11 de 2020, de <https://medlineplus.gov/ency/article/003208.htm>

S. Ren, X. Cao, Y. Wei, & J. Sun. (2014). Face alignment at 3000 fps via regressing local binary features. *In Proc*, (pág. 2).

S. Zafeiriou, G. Tzimiropoulos, & M. Pantic. (2015). *The 300 videos in the wild (300-VW) facial landmark tracking in-the-wild challenge*. In *ICCV Workshop*. Recuperado el 25 de 10 de 2020, de <http://ibug.doc.ic.ac.uk/resources/300-VW/>

Sueño, S. E. (10 de 2016). *Revista de Neurología*. Recuperado el 16 de 10 de 2020, de <https://ses.org.es/docs/rev-neurologia2016.pdf>

T. Danisman, I. Bilasco, C. Djeraba, & N. Ihaddadene. (2010). Drowsy driver detection system using eye blink patterns. In *Machine and Web Intelligence. ICMWI*, (págs. 1, 6, 7).

T. Drutarovsky, & A. Fogelton. (2014). Eye blink detection using variance of motion vectors. In *Computer Vision. In ECCV Workshops*. 1, 2, 5, 6, 7.

Torres Malaver, B. A., & Baquero Cardozo, B. A. (18 de 9 de 2019). *Repositorio Institucional Universidad Distrital - RIUD*. Recuperado el 16 de 10 de 2020, de <http://repository.udistrital.edu.co/handle/11349/22359>

- University, O. (2020). *Oxford English and Spanish Dictionary, Thesaurus, and Spanish to English Translator*. Recuperado el 28 de 10 de 2020, de <https://en.oxforddictionaries.com/definition/fatigue>
- W. H. Lee, E. C. Lee, & K. E. Park. (2010). Blink detection robust to various facial poses. *Journal of Neuroscience Methods.*, (págs. 1, 3, 6, 7).
- W. LU, & M. YANG. (2019). En *Face Detection Based on Viola-Jones Algorithm Applying Composite Features* (págs. 80-85).
- Wikipedia. (25 de 10 de 2020). *wikipedia*. Obtenido de en.wikipedia.org/wiki/Viola%E2%80%93Jones_object_detection_frame
- Wikipedia. (23 de 7 de 2020). *Wikipedia*. Recuperado el 28 de 9 de 2020, de https://es.qaz.wiki/wiki/Cascading_classifiers
- X. Xiong, & F. De la Torre. (2013). En *Supervised descent methods and its applications to face alignment*. In *Proc. CVPR* (págs. 2, 3, 4, 5).