

DETECCIÓN DE SOMNOLENCIA EN CONDUCTORES DE VEHÍCULOS POR MEDIO DE PROCESAMIENTO DE VIDEO

DROWSINESS DETECTION IN VEHICLE DRIVERS THROUGH VIDEO PROCESSING

Sergio Aarón Nieves Flores

Universidad Autónoma de Ciudad Juárez, México
AI158038@alumnos.uacj.mx

José David Díaz Román

Universidad Autónoma de Ciudad Juárez, México
david.roman@uacj.mx

Lidia Hortencia Rascón Madrigal

Universidad Autónoma de Ciudad Juárez, México
lrascon@uacj.mx

Juan de Dios Cota Ruiz

Universidad Autónoma de Ciudad Juárez, México
jcota@uacj.mx

Boris Jesús Mederos Madrazo

Universidad Autónoma de Ciudad Juárez, México
boris.mederos@uacj.mx

Recepción: 7/noviembre/2021

Aceptación: 7/marzo/2022

Resumen

Los accidentes automovilísticos son una de las principales causas de muerte y lesiones a nivel mundial. Muchos son causados por fatiga y somnolencia de los conductores. El presente estudio tiene como objetivo detectar somnolencia en conductores de vehículos. La metodología del trabajo consistió en las siguientes etapas: en primer lugar, se empleó un algoritmo para la detección del rostro del sujeto dentro de la cabina de un automóvil durante la simulación de conducción para identificar regiones que incluyan cada ojo. Posteriormente se construyó un clasificador para distinguir las regiones de cada ojo como: abierto o cerrado. Finalmente, se desarrolló un algoritmo para el seguimiento de las regiones de

interés para alimentar con imágenes al clasificador; para la detección se somnolencia se utiliza un criterio basado en una cantidad de fotogramas consecutivos presentando una identificación de ojos cerrados. El algoritmo presentó un 91.4% de exactitud en la detección de somnolencia.

Palabras Clave: Clasificación, inteligencia artificial, somnolencia.

Abstract

Car accidents are one of the leading causes of death and injury worldwide. Many are caused by fatigue and drowsiness in drivers. The present study is a work based on artificial intelligence applied to the field of video processing to determine a drowsy condition in vehicle drivers. The work methodology consisted of the following stages: first, using the first frame of the video, the subject's face is detected inside the car cabin during the driving simulation in order to identify the regions that include each eye. The next section consists of the construction of a classifier to identify the regions that each eye includes as: open eye or closed eye. Finally, an algorithm is developed to track the face and the regions of interest (regions that include the eyes) to feed the classifier with images; Detection of subject drowsiness is determined using criteria based on a series of consecutive frames having a closed eye identification. The algorithm presented a 91.4% accuracy in detecting drowsiness.

Keywords: Artificial intelligence, classification, drowsiness.

1. Introducción

Los accidentes automovilísticos son una de las principales causas de lesiones o muerte. Las estadísticas muestran que los accidentes automovilísticos son a nivel mundial la novena causa de muerte: 1,3 millones de personas mueren anualmente en accidentes automovilísticos, o 3287 por día. Estadísticamente más del 10% de los accidentes se deben a fatiga, la mayoría de los cuales ocurren en carreteras o después de conducir una gran cantidad de kilómetros [Vesselenyi, 2017].

Desde un punto de vista fisiológico la somnolencia puede ser definida como el estado en donde, para realizar las mismas tareas, es necesario un mayor volumen de tejido neural, lo cual provoca una pérdida progresiva de la eficiencia del

procesamiento cortical [Slater, 2008]. De este modo, conforme hay un aumento del reclutamiento cortical hay una mayor coherencia de la actividad eléctrica cortical. Esta actividad fisiológica es la base de la realización de trabajos para la detección de la somnolencia por medio del análisis de señales electroencefalográficas (EEG) [Chaabene, 2021] [Hong, 2021]. En el trabajo de [Dunbar, 2020] se diseñó un dispositivo de análisis de señales EEG para evaluar el estado del conductor en un entorno realista, poniendo a prueba la actividad cerebral de los pacientes, sin embargo, para emplear este método es necesario estar registrando constantemente la actividad cerebral y ocular mediante diversos sensores, aumentando así el costo del producto final y su implementación implicaría un elemento de incomodidad al conductor [Dunbar, 2020]. Un sujeto que se encuentra en un estado somnoliento es incapaz de mantener la actividad neuronal de los niveles más elevados de complejidad, debido a esto la realización de actividades diarias se ve afectada por la deficiente funcionalidad cognitiva que se presenta, y que se va perdiendo conforme existe una resistencia al sueño, comenzando dicha pérdida por las funciones cognitivas más complejas, incluyendo la relajación de otros sistemas fisiológicos autónomos. Diversas investigaciones hacen uso de esta característica de cambios en distintos sistemas fisiológicos con la finalidad de detectar somnolencia en personas en distintas situaciones; tal es el caso de la investigación de Siddiqui y colaboradores, quienes emplearon la medida de la frecuencia respiratoria como base para un sistema de detección de somnolencia en conductores [Siddiqui, 2021].

De forma similar, Satti et al. desarrollaron un sistema de detección de somnolencia, también aplicado durante la conducción, que se basó en el análisis de los componentes del espectro de frecuencias de señales de electromiografía obtenidas con un arreglo de electrodos conectado en un músculo del antebrazo del individuo que sostiene el volante del vehículo [Satti, 2021]. Por otra parte, Kundinger y su grupo propuso un sistema basado en un clasificador construido con algoritmos de *Machine Learning* que utiliza datos fisiológicos provenientes de un dispositivo de pulsera portátil que es vestido por el conductor del vehículo. Los datos analizados en este estudio fueron principalmente obtenidos del análisis de la señal de

fotopletiografía adquirida con el dispositivo colocado en la muñeca del individuo [Kunding, 2020].

Las personas que se encuentran en un estado de somnolencia y/o con fatiga exhiben ciertos cambios visuales en sus expresiones faciales que son fáciles de detectar, especialmente si se enfoca en los ojos, la cara y la cabeza, como en el trabajo de Flores y colaboradores, donde se presenta el desarrollo de un sistema autónomo para detectar la somnolencia y la distracción de los conductores, empleando la Visión por Computador e Inteligencia Artificial a partir del estudio del comportamiento del sujeto tras el volante [Flores, 2011]. Así pues, algunas de las características más comúnmente observadas en personas que tienen bajo nivel de alerta incluyen un movimiento de los párpados más lento, menor grado de apertura o cierre ocular, disminución de fuerza en los músculos de la nuca que provocan la caída de la cabeza (conocido regularmente como cabeceos), bostezos, el enfoque de la mirada hacia abajo reduciendo así el campo de visión; así las expresiones faciales presentan movimientos muy lentos o incluso sin movimiento y también se presenta una postura muy flácida [Ji, 2002]. Estas características pueden ser aprovechadas para la identificación de somnolencia en los conductores de vehículos empleando la visión por computadora (CV). En un trabajo reciente realizado por Maior et al., se llevó a cabo una comparación de tres modelos de *Machine Learning* (ML) para identificar los patrones de somnolencia, ya que, según los autores, el principal problema que enfrentaron fue que existe una variabilidad en cuanto a cómo se presenta la necesidad de dormir en cada individuo. Los modelos realizados fueron: perceptrón multicapa (MLP), bosque aleatorio (RF) y máquina de vectores de soporte (SVM) para entrenarlos con videos/imágenes y así inferir el estado de los ojos de las personas, trabajo en donde concluyeron que el empleo de la SVM fue la que presentó mejor precisión y tiempo de procesamiento [Maior, 2020]. Por otro lado, Nojiri et al., realizaron una comparación entre *Machine Learning* y *Deep Learning* para identificar la somnolencia en los conductores considerando el uso de maquillaje, en donde emplearon el histograma de gradientes orientados para extraer características y el SVM para clasificar el estado de los ojos, partiendo de igual forma del análisis de los cambios visuales que se presentaban en un conductor de

vehículos [Nojiri, 2019]. En este sentido, una metodología para llevar a cabo una adecuada detección de somnolencia basado en visión por computadora puede involucrar diferentes técnicas para detectar el rostro en primera instancia, identificar regiones en donde se incluyan los ojos, registrar un seguimiento a estas regiones, realizar un preprocesamiento a estas, y por medio de la creación de un modelo de clasificación, identificar si los ojos se encuentran abiertos o cerrados. Finalmente, por medio de un criterio definido basado en el estado de los ojos durante un período de tiempo, se puede generar una alerta oportuna de presencia de esta condición. El algoritmo de Kanade Lucas Tomasi (algoritmo KLT) es una técnica de registro de imágenes para identificar la posición de características, que posteriormente pueden ser empleadas para realizar el seguimiento de esos puntos a lo largo de los fotogramas de un video [Lucas, 1981]. De este modo, este algoritmo es empleado para la detección y seguimiento del rostro que, para lograr un seguimiento efectivo, es necesario seguir un proceso de tres pasos: detección del rostro, identificación de parámetros anatómicos de rastreo y finalmente el rastreo del rostro (con el algoritmo KLT) [MathWorks, 2020]. La localización de puntos de referencia faciales se ha convertido en una de las aplicaciones más comunes dentro de la CV, y para lograr un sistema como este se deben seguir cuatro pasos fundamentalmente [Liu, 2017] : a) detección y alineación de articulaciones, en donde se utiliza un detector de rostros para predecir el rectángulo de la cara y cinco puntos de referencia faciales semánticos; b) Regresión de forma en cascada basada en la pose, en la que se realiza una estimación de la forma en el dominio espacial; c) seguimiento de pose, donde se genera una forma inicial en relación con la alineación en el dominio del tiempo; y d) reinicialización, en la que se introduce un evaluador de alineación facial para medir la eficiencia de la alineación, cuando la puntuación del evaluador sea menor a la del umbral configurado, se realiza el reinicio de la forma compuesta por el detector y el rastreador facial [Liu, 2020].

2. Métodos

El algoritmo de procesamiento de los datos fue desarrollado en lenguaje de programación de Matlab® y se empleó un computador de escritorio del fabricante

Hewlett-Packard, modelo Compaq dc7900 con procesador Intel® Core 2 Duo y 6 GB de RAM.

Descripción de la base de datos empleada

Para este estudio fue necesaria la creación de dos bases de datos: una que incluyó videos de individuos simulando conducir en un estado de somnolencia y otra contentiva de imágenes de la región de interés de los ojos de personas; esta última extraída de los videos iniciales. Estas bases de datos fueron necesarias para entrenar y validar los modelos de clasificación del estado del ojo y la detección de somnolencia mediante un criterio definido.

La base de datos de videos cuenta con 31 videos cortos de entre 30 a 60 segundos en formato mp4, en donde se pidió a los individuos de prueba grabarse dentro de la cabina de un automóvil simulando que se encuentran en una situación real de conducción, realizando los movimientos que acostumbran normalmente, y en el transcurso de la grabación, simular que se encuentran en un estado de somnolencia. Para simular esta condición se les pidió mantener los ojos cerrados por al menos 2 segundos. Al permanecer con los ojos cerrados durante el tiempo requerido se considera que se presenta un micro sueño, indicativo importante de presencia de somnolencia y con peligrosa consecuencia durante la conducción. Los videos se grabaron a una frecuencia de 30 fotogramas por segundo (FPS) y se emplearon cámaras de celulares de distintas marcas y características, por lo que la resolución pudo ser distinta en todos los videos.

La segunda base de datos se utilizó para determinar el estado de los ojos, y las imágenes fueron extraídas de los videos grabados. Se almacenaron imágenes con dimensiones variantes dependiendo de la grabación de cada individuo, donde cada uno de los participantes presenta, respecto a la cámara de grabación, diferentes ángulos del rostro, posiciones de la cabeza y diferentes expresiones faciales. Se tienen imágenes donde se presentan los ojos abiertos o cerrados. Se realizó un procesamiento a dichas imágenes para obtener una región de interés que cubriera cada ojo (izquierdo y derecho) del individuo en la imagen. Las imágenes de estas regiones se redimensionaron a un tamaño de 101x101 pixeles. Finalmente, esta

base de datos contiene 9,065 imágenes únicamente de los ojos, agrupadas por pares. Es importante mencionar que las imágenes con cada ojo pueden contener parte de las cejas y/o nariz del sujeto, además de la porción respectiva del lente en caso de que la persona los estuviera usando; esto debido a que las imágenes se extrajeron de los videos y en ellos los individuos fueron grabados realizando una actividad común durante la conducción de un vehículo, por lo que sus rostros no estaban enfocados hacia el frente en todo momento. Esto permitió generar una base de datos robusta con imágenes para entrenar el modelo de clasificación.

Preprocesamiento de las imágenes de las regiones de interés

Para el desarrollo de un modelo de clasificación basado en aprendizaje automatizado, resulta de gran utilidad el realizar un preprocesamiento de los datos, en este caso de las imágenes que se utilizan para el entrenamiento. De esta manera se disminuye la carga de datos de entrenamiento debido a que el clasificador recibe únicamente las características de las regiones de interés y realiza menos tareas para poder arrojar una predicción de salida. De igual manera se reduce el tiempo de trabajo que le toma al modelo para generar una salida. En el presente caso, fue necesario identificar los ojos de los individuos en los fotogramas obtenidos de los videos y llevar a cabo un procesamiento únicamente de la región de interés (ROI) de los ojos.

El primer paso para lograr obtener la ROI de los ojos es la detección del rostro mediante el detector de objetos en cascada, el cual es un modelo que emplea el algoritmo de Viola-Jones [Viola, 2004]. Un ejemplo de la detección del rostro de un individuo se muestra en la figura 1. Las imágenes del rostro se redimensionaron a 300x300 píxeles. Para la creación de un método robusto que sea capaz de detectar los ojos, es necesario considerar diferentes parámetros como lo es: la iluminación ambiental, la cual en una situación real de conducción va cambiando constantemente, y los movimientos y gestos realizados por las personas al momento de estar detrás un volante, por lo que se ha planteado emplear una técnica que aproveche la información antropométrica del rostro para generar las regiones que incluyan diferente información de los ojos del sujeto y las regiones cercanas.

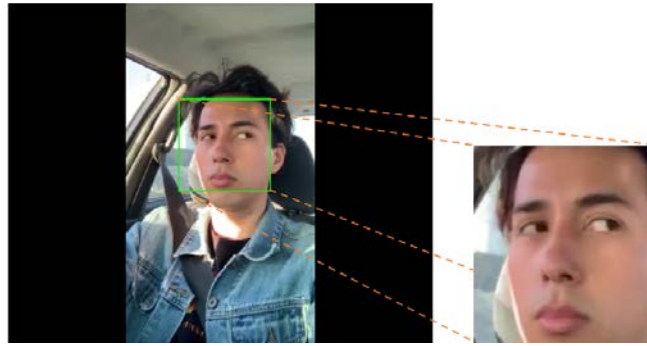


Figura 1 Detección del rostro empleando la cascada de objetos del algoritmo Viola-Jones.

En este estudio se empleó un algoritmo de detección de puntos de referencia faciales desarrollado por Asthana y colaboradores [Asthana, 2013], el cual detecta los bordes de ciertas regiones anatómicas como lo es el contorno de la cara, el puente de la nariz, los ojos, las cejas y la boca y genera 66 puntos de referencia, como se observa en la figura 2.



Figura 2 Implementación del algoritmo de Asthana.

A partir de la localización de estos puntos se pueden crear las regiones que contengan los ojos individualmente encerrados en recuadros. Para este caso se enfocó en la localización del primer y último punto sobre cada una de las cejas, posteriormente se realizó una medición de un punto al otro para cada ceja y con estas medidas se creó un rectángulo para contener únicamente la región de cada ojo, como se observa en la figura 3. Las regiones de cada ojo se redimensionaron a 101x101 píxeles, las cuales posteriormente se emplearían para la clasificación y en el seguimiento de los fotogramas del video. En la figura 4 se muestran los ojos individuales detectados mediante el uso de los puntos de referencia faciales.

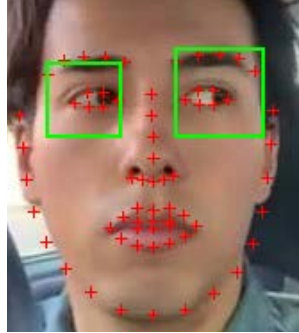
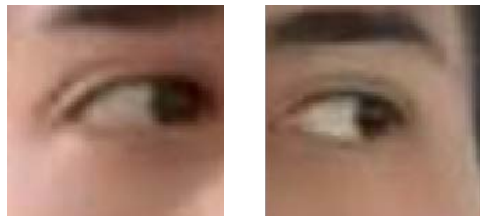


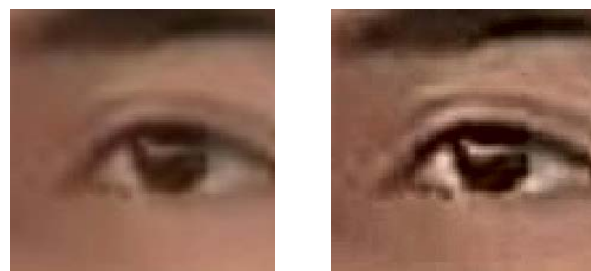
Figura 3 Identificación de las ROIs empleando los puntos de referencia faciales.



a) Ojo derecho b) Ojo izquierdo

Figura 4 Identificación y recorte de las regiones de interés de los ojos.

Posteriormente se empleó un filtrado para resaltar bordes y mejorar el contraste de las ROIs mediante el empleo del filtro *Sharpen* que permite controlar el enfoque de las imágenes mediante diferentes parámetros como el radio y la cantidad (*radius* y *amount*); el primero se ajustó a un valor de 5, para agudizar la nitidez de las regiones más anchas alrededor de los bordes; el segundo se refiere a la fuerza del efecto de nitidez y se ajusta típicamente en un rango de [0 2], en este caso se ajustó en 1.5. De esta manera se aumentó el contraste a lo largo de los bordes donde se encuentran diferentes píxeles con colores diferentes, como se puede observar en la figura 5.

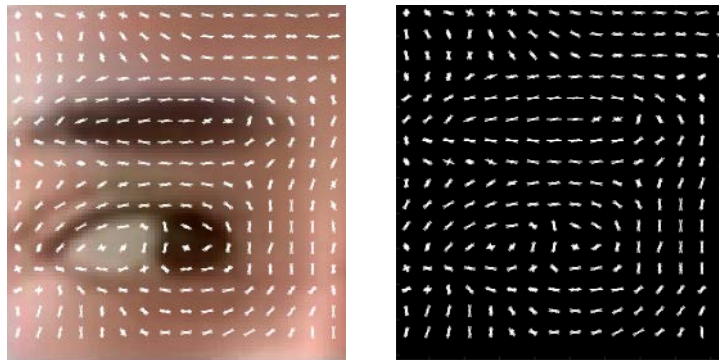


a) Antes del filtrado b) Después del filtrado

Figura 5 ROI del ojo.

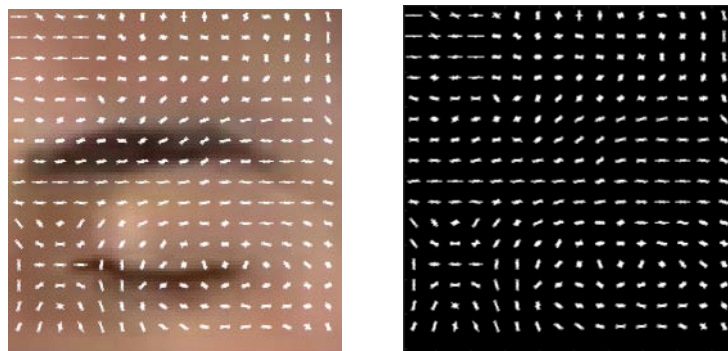
Extracción de características de la región de interés

La extracción de características consiste en discriminar toda la información irrelevante de los elementos de la base de datos y rescatar aquella que le permita al programa diferenciar una clase sobre otra. En el contexto de detectar la somnolencia en conductores, conviene emplear un algoritmo que sea capaz de detectar características antropométricas de la región ROI que no dependan de la luz ambiental debido a la diversidad de tonalidades que se adquieren al estar en una situación real de conducción. El Histograma de Orientaciones del Gradiente (HOG) se enfoca en determinar la forma de los objetos que se encuentran dentro de una imagen, en este caso de las cejas y los ojos como se observan en las figuras 6 y 7, y extraer esta información en un vector $1 \times N$, donde N representa el largo de las características HOG.



a) Características HOG sobre la ROI b) Información extraída de la ROI

Figura 6 Extracción de características HOG ojo abierto.



a) Características HOG sobre la ROI b) Información extraída de ROI

Figura 7 Extracción de características HOG ojo cerrado.

El tamaño de este vector depende tanto del tamaño de la imagen como de la celda que se emplee para obtener las características. Para agilizar el entrenamiento del clasificador y aligerar la carga de programación, se empleó un tamaño de celda de 8x8 con el cual el vector es del tamaño 4,356, siendo una medida adecuada para no perder información de la forma del ojo.

Modelo de clasificación del estado de los ojos

El modelo de clasificación fue programado para identificar nuevos datos de entrada con base en observaciones de ejemplos previamente etiquetados y entrenados. Con el algoritmo de aprendizaje máquina se ajusta un modelo que contiene el estado de los ojos que aparecen en los datos de entrenamiento, así que es posible realizar una discriminación entre los ojos de diferentes personas dentro de la base de datos y emplear el modelo para reconocer si un ojo está abierto o cerrado en imágenes de entrada.

El modelo de *machine learning* empleado consistió en la implementación de un clasificador basado en máquina de vectores de soporte para llevar a cabo la identificación de ambas clases (Abierto y Cerrado), en el que se emplearon 9,065 datos de entrenamiento y 1,240 datos de prueba, ambos debidamente etiquetados a su clase correspondiente. Se empleó un kernel de tipo Gaussiano para la construcción del clasificador y se utilizó la función *fitcsvm* con el parámetro *OptimizeHyperparameters* para el ajuste de los hiperparámetros (*BoxConstraint* y *KernelScale*) de la máquina de vectores de soporte, empleando como entrada las características HOG de las imágenes de entrenamiento.

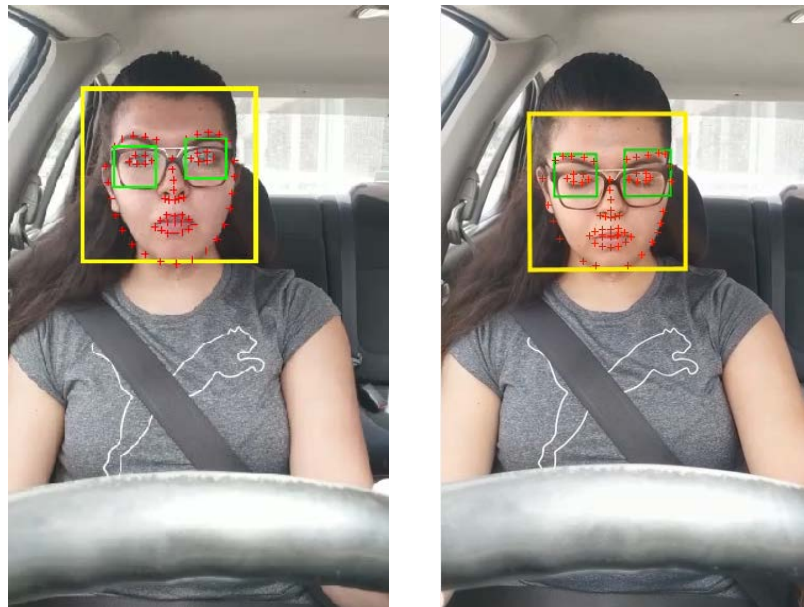
El modelo de clasificación arroja como argumentos de salida dos variables: una con la predicción de la imagen analizada, la cual debe ser únicamente “abierto” o “cerrado”; y otra con un valor numérico que establece la confianza de la predicción realizada.

Seguimiento de la región de interés en los videos

El algoritmo KLT, es un método de detección y seguimiento de objetos que es adecuado para la identificación de los ojos de los individuos en cada momento en

la conducción, ya que permite llevar a cabo un seguimiento del rostro que pudiera ser efectuado en tiempo real.

Para realizar un seguimiento de los ojos a lo largo de los cuadros del video, es necesario identificarlos inicialmente en el primer fotograma del video, en donde por medio del algoritmo de Viola-Jones [Viola, 2004] se identifica el rostro, adicional a esto, se emplea el modelo de detección de puntos de referencia faciales de Asthana [Asthana, 2013], el cual, como ya se mencionó antes, detecta los bordes de ciertas regiones anatómicas. Así pues, una vez aplicado este modelo, se procede a crear un recuadro (solo para fines visuales) en donde se identifica la ROI de los ojos para su seguimiento como se observa en la figura 8.



a) Identificación de ROI b) Seguimiento de los puntos y las ROIs

Figura 8 Identificación de ROI por medio de un sistema de detección de puntos de referencia faciales y seguimiento de los puntos y las ROIs.

El algoritmo KLT [Lucas, 1981] identifica un conjunto de puntos característicos en el rostro, los cuales le permitirán al rastreador encontrar, para cada punto del cuadro anterior, el punto correspondiente en el cuadro actual, y así estimar la traslación, rotación y la escala entre los puntos de un fotograma a otro dentro del recuadro que identifica al rostro y las ROIs de los ojos.

Sistema de detección de somnolencia

Una vez desarrollado el modelo de clasificación de ojos y establecida la técnica de seguimiento de estos, fue necesario implementar un algoritmo en donde se realicen los dos procedimientos en conjunto para conocer el estado de los ojos de los individuos en cada fotograma del video. Se realizó un análisis cuadro por cuadro de los videos de la base de datos para la elaboración e implementación de un criterio de detección de somnolencia. Una de las causas más letales en la conducción es la aparición de micro sueños debido a la falta de horas de sueño, con base en esto se implementó un criterio de detección de basado en la duración del ojo cerrado. Basado en que la duración típica de un pestañeo se encuentra como máximo en unos 0.4 segundos [Bacivarov, 2008], y que la frecuencia de grabación de los fotogramas de los videos es de 30 fotogramas por minuto (FPS), se consideró la presencia de somnolencia si se detectaba que el individuo permanecía con sus ojos cerrados continuamente por más de 3 veces la duración de un parpadeo normal. Esto claramente puede representar una pérdida de atención al conducir un vehículo. Tomando esta consideración se decidió establecer que la indicación de somnolencia se activaría si los ojos permanecieran cerrados continuamente por 40 fotogramas seguidos, lo cual representa una duración de 1.33 segundos con los ojos cerrados. En un estudio realizado por Maior y colaboradores, uno de los criterios definido para la detección de un “parpadeo largo”, considerado como indicación de somnolencia, fue la detección de ojos cerrados durante 20 fotogramas consecutivos, donde la frecuencia de grabación fue de 23 FPS, lo que representa un tiempo de 0.87 segundos (más de 2 veces la duración de un parpadeo normal) [Maior, 2020]. Una vez que un contador haya pasado el límite de fotogramas permitidos para tener los ojos cerrados (40 fotogramas seguidos), se despliega un mensaje de advertencia informando sobre la presencia de somnolencia, como se observa en la figura 9, y se mantendrá desplegado en pantalla hasta que el conductor vuelva a abrir los ojos. Para evitar que posibles errores de clasificación afecten el sistema de detección de somnolencia, el mensaje permanecerá en pantalla hasta que se detecten más de 5 fotogramas con los ojos abiertos luego de haber detectado esta condición previamente.

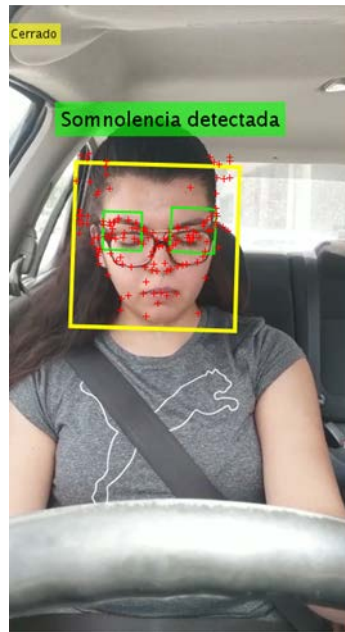


Figura 9 Mensaje de advertencia desplegado en pantalla como indicación de somnolencia.

3. Resultados

Modelo de clasificación del estado de los ojos

La validación del modelo de SVM para determinar el estado del ojo consistió en emplear 1,240 imágenes de prueba no usadas en el entrenamiento del clasificador, en donde se realizó el mismo procesamiento a las imágenes y se analizaron las predicciones arrojadas por el algoritmo, determinando la exactitud y otras métricas de evaluación del desempeño.

Se realizaron dos validaciones para este caso; el modelo de SVM analiza cada ojo individualmente y arroja una predicción para cada uno. Se probó el clasificador con un total de 2,480 imágenes de ojos para determinar su estado. Se consideró como “positivo” cuando el ojo se encuentra cerrado y “negativo” cuando se encuentra abierto. En la tabla 1 se muestra la matriz de confusión la cual desglosa los aciertos y errores obtenidos. Se observa que la exactitud del programa es de 97.74% en la predicción del estado del ojo individualmente.

En la tabla 2 se muestran los resultados de sensibilidad, especificidad y el valor predictivo de positivos (VPP) y el valor predictivo de negativos (VPN) en la evaluación del clasificador con referencia al mismo grupo de imágenes de los ojos.

Tabla 1 Matriz de confusión sobre la identificación de cada ojo.

| | | Predicción | | |
|-------|---------|------------|---------|-------|
| | | Cerrado | Abierto | Total |
| Clase | Cerrado | 1,206 | 32 | 1,238 |
| | Abierto | 24 | 1,218 | 1,242 |
| Total | | 1,230 | 1,250 | 2,480 |

Tabla 2 Desempeño del modelo de clasificación basado en imágenes de ojos individuales.

| Parámetro | Valor |
|------------------|---------------|
| Sensibilidad | 97.42% |
| Especificidad | 98.07% |
| VPP | 98.05% |
| VPN | 97.44% |
| Exactitud | 97.74% |

La segunda validación consistió en emplear 1,240 pares de imágenes de ojos para determinar el estado de ambos ojos al mismo tiempo, es decir, considerar que ambos ojos de una persona se encuentran cerrados o abiertos, a diferencia de la prueba anterior en donde se analizaba cada ojo por separado. En esta prueba se trabaja con los valores de predicción y de confianza que arroja el algoritmo para cada ojo; una vez realizada la predicción en ambos ojos individualmente se asigna la etiqueta de predicción a ambos ojos basada en el resultado de mayor confianza arrojado por el clasificador en cada ojo. En la tabla 3 se observa la matriz de confusión para esta prueba. Esta técnica arrojó una exactitud de 98.8%, lo que significa que es mayor a comparación de la prueba pasada, por lo que se decidió su empleo para la posterior detección de somnolencia.

Tabla 3 Matriz de confusión sobre la identificación de ambos ojos.

| | | Predicción | | |
|-------|---------|------------|---------|-------|
| | | Cerrado | Abierto | Total |
| Clase | Cerrado | 608 | 11 | 619 |
| | Abierto | 4 | 617 | 621 |
| Total | | 612 | 628 | 1,240 |

En la tabla 4 se muestran los valores de las métricas de desempeño para esta técnica, observando una ligera mejoría en comparación con los resultados obtenidos en la prueba pasada.

Tabla 4 Desempeño del modelo de clasificación para la identificación del estado de ambos ojos.

| Parámetro | Valor |
|------------------|---------------|
| Sensibilidad | 98.22% |
| Especificidad | 99.36% |
| VPP | 99.35% |
| VPN | 98.25% |
| Exactitud | 98.80% |

Sistema de detección de somnolencia

Para la evaluación del sistema de detección de somnolencia, se segmentaron los videos de la base de datos y se obtuvieron 140 videos cortos de 2 a 4 segundos de duración (de entre 60 y 120 fotogramas cada uno), los cuales presentan diferentes situaciones de la condición del sujeto sin la presencia de alguna interferencia que pudiera afectar el rastreo para la identificación de las ROIs. Las situaciones presentadas en los videos recortados son fragmentos en donde el conductor está: en alerta (ojos abiertos) durante todo el video, presenta somnolencia (ojos cerrados) en todo el video, fragmentos donde está en un estado de alerta y pasa a somnolencia, o se encuentra en somnolencia y pasa a estado de alerta.

En la tabla 5 se presenta la matriz de confusión especificando los aciertos y errores que el sistema presentó en la detección de somnolencia o estado de alerta.

Tabla 5 Matriz de confusión en la detección de somnolencia.

| | | Detección | | Total |
|-----------|-----------------|-------------|-----------------|-------|
| | | Somnolencia | Sin somnolencia | |
| Condición | Somnolencia | 79 | 7 | 86 |
| | Sin somnolencia | 9 | 91 | 100 |
| Total | | 88 | 98 | 186 |

Se consideró como “positivo” los instantes en donde el sujeto se encontraba en una simulación de somnolencia, y “negativo” presentaba sus ojos abiertos en señal de estar atento a la conducción. Dentro de los 140 videos examinados para evaluar el desempeño del sistema de detección de somnolencia se presentan 86 casos de simulación de somnolencia y 100 segmentos donde el conductor se encuentra en un estado de alerta que puede considerarse adecuado para la conducción.

A partir de los resultados obtenidos en la matriz de confusión, se determinaron las métricas de evaluación del desempeño del algoritmo. En la tabla 6 se muestran los resultados obtenidos.

Tabla 6 Contingencia sobre la exactitud del sistema de detección de somnolencia.

| Parámetro | Valor |
|------------------|---------------|
| Sensibilidad | 91.86% |
| Especificidad | 91.00% |
| VPP | 89.77% |
| VPN | 92.86% |
| Exactitud | 91.40% |

Evaluación de tiempos de ejecución sistema de detección de somnolencia

Para que este estudio pueda llegar a ser implementado para trabajar en tiempo real en un automóvil, es esencial que el algoritmo trabaje lo más rápido posible. En la tabla 7 se despliegan los tiempos promedios que tomaba el sistema en analizar un fotograma, especificando cada uno de los procedimientos que se realizan para determinar si hay somnolencia o no.

Tabla 7 Tiempos promedios de ejecución del sistema por fotograma.

| Proceso | Tiempo promedio (s) |
|--|---------------------|
| Detección de ojos, filtrado, redimensión | 0.0512 |
| Ext. Características HOG | 0.0142 |
| Predicción del estado de los ojos | 0.4002 |
| Total | 0.4656 |

La tabla 7 muestra que en promedio el sistema de detección de somnolencia tarda poco menos de medio segundo para determinar el estado de los ojos en cada fotograma, siendo un valor considerablemente lento para que la aplicación sea implementada en tiempo real. También muestra que el procedimiento que más

consume tiempo dentro del sistema es lo que tarda el modelo de clasificación en arrojar una predicción sobre si el ojo se encuentra abierto o cerrado, tomando en cuenta que este paso se realiza analizando ambos ojos. Esto indica que existe un área de oportunidad para mejorar el sistema de clasificación de manera que genere una respuesta en un tiempo más corto.

4. Discusión

La detección de la somnolencia parte de la identificación de las regiones de interés de los ojos, que como se ha mencionado, se emplearon algoritmos que han demostrado ser robustos como el de Viola-Jones y el de Asthana; la identificación de las zonas del rostro que contienen la región de los ojos en diferentes personas se ajusta adecuadamente aún con las variaciones que se presentan, como lo es el posicionamiento de la cabeza, ángulo e iluminación en la que se realizó la grabación, e incluso cuando el individuo portaba lentes de aumento.

El seguimiento adecuado de las ROIs de los ojos se logró al detectar los puntos característicos en la región del rostro, en el caso de este estudio se empleó la técnica propuesta en el algoritmo KLT (*good features to track*). Algo que se observó, y es muy común en esta etapa, fue que cierta cantidad de dichos puntos se pierden con el avance de los fotogramas del video dependiendo del movimiento que el individuo realizara, ya sea observar los retrovisores o mirar por las ventanas laterales. Sin embargo, el seguimiento de la región de interés se pudo llevar a cabo de forma adecuada, manteniendo en todo momento los ojos dentro de recuadros marcados que indicaban su detección.

Es importante mencionar el espacio de oportunidad que existe para optimizar el tiempo de ejecución del algoritmo, intentando minimizar el tiempo de respuesta del clasificador, el cual consume alrededor de 0.4 segundos en la identificación del estado de los ojos en un fotograma. En futuros trabajos se intentará implementar otro algoritmo que genere un vector de características de menor tamaño lo que permitiría al clasificador generar una respuesta más rápida, o bien, implementar algún algoritmo de reducción de dimensionalidad de los datos; también se plantea el uso de redes neuronales convolucionales para la tarea de clasificación del estado

de los ojos esperando una respuesta con menor tiempo. Cabe mencionar que se realizaron pruebas preliminares empleando la identificación de los puntos de interés del rostro y de los ojos en cada fotograma del video (empleando el algoritmo de Asthana), y utilizar la localización de esos puntos (y medidas de distancia entre ellos) como características para detectar el estado de los ojos, sin embargo se presentó el inconveniente de que el algoritmo fallaba al detectar todos los puntos cuando el sujeto giraba o bajaba ligeramente el rostro en relación a la ubicación frente a la cámara, lo que impedía tener todas las características en cada fotograma. Esta técnica ha sido empleada en investigaciones recientes, generando muy buenos resultados en la detección de somnolencia tanto en conductores [Celecia, 2020] como es otras situaciones fuera de la conducción [Maio, 2020].

Los autores de estas investigaciones no reportan haber tenido estos problemas, sin embargo, las condiciones en que se grabaron los videos en esos trabajos no fueron las mismas comparadas con las de la presente investigación, destacando mayor movilidad que presentaban los individuos en nuestras grabaciones.

5. Conclusiones

En la presente investigación se logró desarrollar un sistema de detección de somnolencia basado en la identificación del estado de los ojos. El preprocesamiento de las imágenes constó de diferentes procedimientos necesarios para facilitar el entrenamiento del modelo de clasificación, sin embargo, la extracción de características tuvo un papel fundamental tanto en la exactitud del clasificador, en la del detector de somnolencia y en el tiempo de ejecución del sistema; de ahí la necesidad de realizar una selección y aplicación de extracción de características adecuada. El uso de las características HOG se mostró adecuado en cuanto a las exactitudes de las dos fases del proyecto, sin embargo, la cantidad de datos a procesar por el clasificador hizo que este tomara más del tiempo deseado para generar una respuesta.

La detección del estado de los ojos de los conductores se llevó a cabo empleando la técnica de inteligencia artificial de SVM, con alta aplicación en proyectos de clasificación binaria, como lo es en este caso en la determinación de un ojo abierto

o cerrado. En combinación con la extracción de características HOG tuvo como resultado un modelo de clasificación con un porcentaje de exactitud elevado, obteniendo un 98.8% de aciertos.

El estado de alerta de los conductores fue evaluado bajo el criterio de determinar el número de fotogramas consecutivos en los que se identifica que el sujeto mantiene los ojos cerrados. Los resultados obtenidos demuestran que, con el empleo de este criterio de detección, se cumple con los objetivos planteados en el estudio, obteniendo un 91.4% de exactitud en la identificación de la somnolencia, activando la alarma de advertencia correctamente. No obstante, el margen de error observado se debe a factores externos al criterio establecido, como lo es el seguimiento de las regiones de interés y por lo tanto de la clasificación del estado del ojo.

Es importante resaltar que, aunque la técnica propuesta en el presente estudio no permite su aplicación en tiempo real, puede emplearse de manera práctica en situaciones donde se desee verificar (en tiempo diferido) la ocurrencia de episodios de somnolencia durante la conducción de vehículos durante largos períodos de tiempo. De esta forma, puede ser útil en determinar si la causa de un accidente es producto del estado de somnolencia del conductor, o bien hacer un análisis de la frecuencia de detección de somnolencia para recomendar períodos de descanso oportunos a los conductores, evitando así potenciales situaciones de peligro durante la conducción.

6. Bibliografía y Referencias

- [1] Asthana, Akshay and Zafeiriou, Stefanos and Cheng, Shiyang and Pantic, Maja, (2013). Robust discriminative response map fitting with constrained local models, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 3444–3451, doi: 10.1109/CVPR.2013.442.
- [2] Bacivarov, I., Ionita, M., and Corcoran, P., Statistical models of appearance for eye tracking and eye-blink detection and measurement. IEEE transactions on consumer electronics, 54(3), 1312-1320, 2008.
- [3] Celecia, A., Figueiredo, K., Vellasco, M., and González, R., A portable fuzzy driver drowsiness estimation system. Sensors, 20(15), 4093, 2020.

- [4] Chaabene, S., Bouaziz, B., Boudaya, A., Hökelmann, A., Ammar, A., and Chaari, L., Convolutional Neural Network for Drowsiness Detection Using EEG Signals. *Sensors*, 21(5), 1734, 2021.
- [5] Dunbar, J., Gilbert, J., and B. Lewis, B., (2020). Exploring differences between self-report and electrophysiological indices of drowsy driving: A usability examination of a personal brain-computer interface device, *J. Safety Res.*, Apr. 2020, doi: 10.1016/j.jsr.2020.04.006.
- [6] Flores, M., Armingol, J., and De La Escalera, A., (2011). Sistema Avanzado de Asistencia a la Conducción para la Detección de la Somnolencia, *Rev. Iberoam. Automática e Informática Ind.*, vol. 8, pp. 216–228, doi: 10.1016/j.riai.2011.06.009.
- [7] Hong, S., and Baek, H., Drowsiness Detection Based on Intelligent Systems with Nonlinear Features for Optimal Placement of Encephalogram Electrodes on the Cerebral Area. *Sensors*, 21(4), 1255, 2021.
- [8] Ji, Qiang and Yang, Xiaojie, (2002). Real-Time Eye, Gaze, and Face Pose Tracking for Monitoring Driver Vigilance, pp. 357–377, doi: 10.1006/rtim.2002.0279.
- [9] Kundinger, T., Sofra, N., and Riener, A., Assessment of the potential of wrist-worn wearable sensors for driver drowsiness detection. *Sensors*, 20(4), 1029, 2020.
- [10] Liu, Caifeng and Feng, Lin and Guo, Shuai and Wang, Huibing and Liu, Shenglan and Qiao, Hong, (2020). An incrementally cascaded broad learning framework to facial landmark tracking, *Neurocomputing*, vol. 410, pp. 125–137, Oct. 2020, doi: 10.1016/j.neucom.2020.05.025.
- [11] Liu, Qingshan and Yang, Jing and Deng, Jiankang and Zhang, Kaihua, (2017). Robust facial landmark tracking via cascade regression, *Pattern Recognit.*, vol. 66, pp. 53–62, Jun. 2017, doi:10.1016/j.patcog.2016.12.024.
- [12] MathWorks, (2020). Face Detection and Tracking Using the KLT Algorithm - MATLAB & Simulink - MathWorks América Latina. Disponible en: <https://la.mathworks.com/help/vision/examples/face-detection-and-tracking-using-the-klt-algorithm.html> [accessed Sep. 03, 2020].

- [13] Lucas, Bruce D and Kanade, Takeo and others, Iterative Image Registration Technique With an Application To Stereo Vision., vol. 2, no. April 1981, pp. 674–679, 1981.
- [14] Maior, C., das Chagas Moura, M., Santana, J., and Lins, I., Real-time classification for autonomous drowsiness detection using eye aspect ratio. *Expert Systems with Applications*, 158, 113505, 2020.
- [15] N. Nojiri, X. Kong, L. Meng and H. Shimakawa, (2019). Discussion on Machine Learning and Deep Learning based Makeup Considered Eye Status Recognition for Driver Drowsiness, *Procedia*, vol. 147, doi: 10.1016/j.procs.2019.01.252, pp. 264-270.
- [16] Satti, A., Kim, J., Yi, E., Cho, H., and Cho, S., Microneedle array electrode-based wearable EMG system for detection of driver drowsiness through steering wheel grip. *Sensors*, 21(15), 5091, 2021.
- [17] Siddiqui, H. U. R., Saleem, A. A., Brown, R., Bademci, B., Lee, E., Rustam, F., & Dudley, S., Non-invasive driver drowsiness detection system. *Sensors*, 21(14), 4833, 2021.
- [18] Slater, Jeremy D, (2008). A definition of drowsiness: One purpose for sleep?, *Med. Hypotheses*, vol. 71, no. 5, pp. 641–644, Nov. 2008, doi: 10.1016/j.mehy.2008.05.035.
- [19] Vesselenyi, T., Moca, S., Rus, A., Mitran, T., and Tătaru, B., (2017). Driver drowsiness detection using ANN image processing, *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 252, no. 1, 2017, doi: 10.1088/1757-899X/252/1/012097.
- [20] Viola, P., and Jones, M., (2004). Robust Real-Time Face Detection, *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004, doi: 10.1023/B:VISI.0000013087.49260.fb.