Universidad de Lima
Facultad de Ingeniería y Arquitectura
Carrera de Ingeniería de Sistemas

ADAPTACIÓN Y COMPARACIÓN DE DOS METODOLOGÍAS DE RECONOCIMIENTO FACIAL APLICADOS A LA DETECCIÓN DE SOMNOLENCIA EN CONDUCTORES

Tesis para optar el Título Profesional de Ingeniero de Sistemas

Johanna Paola Gorvenia Ynfanzón

Código 20131858

Fernando Antonio Tello Sánchez

Código 20132291

Asesor

Juan Gutiérrez Cárdenas

Lima – Perú

Noviembre, 2018

UNIVERSIDAD DE LIMA
MCMLXII
SCIENTIA ET PRAXIS

ADAPTACIÓN Y COMPARACIÓN DE DOS METODOLOGÍAS DE RECONOCIMIENTO FACIAL APLICADOS A LA DETECCIÓN DE SOMNOLENCIA EN CONDUCTORES

TABLA DE CONTENIDO

RESUMEN **1**

ABSTRACT **2**

INTRODUCCIÓN **3**

DESCRIPTORES TEMÁTICOS **4**

CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA **5**

1.1 Formulación del problema 5

1.2 Objetivo de la investigación 7

1.1.1 Objetivo general 7

1.1.2 Objetivos específicos 7

1.3 Justificación 7

CAPÍTULO II: ESTADO DEL ARTE **9**

2.1 Reconocimiento facial 9

2.2 Reconocimiento de la somnolencia. 12

2.3 Método para la detección de somnolencia: Viola Jones y *Regression Based Facial Landmark Detection.* 18

CAPÍTULO III: MARCO TEÓRICO **23**

3.1 Detección de rostros 23

3.2 Somnolencia 25

3.3 Accidentes de tráfico 28

3.4 Consideraciones 30

3.4.1 Base de datos NTHU *Drowsy Driver Detection* (NTHU-DDD): 30

3.4.2 Participantes 30

3.4.3 Población 30

3.4.4 Muestra 31

3.5 Definición de metodologías 31

3.5.1 Viola-Jones Object Detection Framework 31

3.5.2*Support Vector Machine* (SVM).....36
CAPÍTULO IV: DESARROLLO DE LA SOLUCIÓN PROPUESTA 50
4.1 Diagrama de bloque.....50
4.2 Desarrollo de la experimentación.....51
4.2.1Captura de la imagen.....51
4.2.2Metodología Viola-Jones.....51
4.2.3Metodología *Regression Based Facial Landmark Detection.*.....56
4.3 Implementación de la propuesta.....60
4.3.1Implementación de la metodología Viola-Jones.....60
4.3.2Implementación de la metodología *Regression Based Facial Landmark Detection.*.....67
CAPÍTULO V: PRUEBAS Y RESULTADOS.....71
5.1 Metodología de la experimentación.....71
5.2 Características de los videos de la base de datos NTHU-DDD. ...72
5.3 Características de los videos de nuestra propia base de datos.....74
5.4 Experimentación usando videos de la base de datos NTHU-DDD. 76
5.4.1Situación 1: Sin oclusión y luminosidad alta.....76
5.4.2Situación 2: Sin oclusión y luminosidad baja.....77
5.4.3Situación 3: Con oclusión y luminosidad alta. 80
5.4.4Situación 4: Con oclusión y luminosidad baja.....82
5.5 Experimentación usando videos de nuestra propia base de datos. 84
5.5.1Situación 1: Sin oclusión y luminosidad alta.....84
5.5.2Situación 2: Sin oclusión y luminosidad baja.....86
5.5.3Situación 3: Con oclusión y luminosidad alta. 89
5.5.4Situación 4: Con oclusión y luminosidad baja.....91
5.6 Resultados.....93
5.6.1Resultados con la base de datos NTHU-DDD.....93

5.6.2Resultados con nuestra propia base de datos. 94
CONCLUSIONES.....97
TRABAJOS FUTUROS 100
GLOSARIO DE TÉRMINOS.....101
REFERENCIAS 104

ÍNDICE DE TABLAS

Tabla 3.1 Métodos de detección de rostros: Holísticos, Basados en características e Híbridos. 25
Tabla 3.2 Número total de accidentes de tránsito por departamento en el año 2014. 29
Tabla 5.1 Resultados - Situación 1: Sin oclusión y luminosidad alta - Método: *Regression Based Facial Landmark Detection.* 76
Tabla 5.2 Resultados - Situación 1: Sin oclusión y luminosidad alta - Método: Viola-Jones. 77
Tabla 5.3 Resultados - Situación 2: Sin oclusión y luminosidad baja - Método: *Regression Based Facial Landmark Detection.* 78
Tabla 5.4 Resultados - Situación 2: Sin oclusión y luminosidad baja - Método: Viola-Jones. 79
Tabla 5.5 Resultados - Situación 3: Con oclusión y luminosidad alta - Método: *Regression Based Facial Landmark Detection.* 80
Tabla 5.6 Resultados - Situación 3: Con oclusión y luminosidad alta - Método: Viola-Jones. 81
Tabla 5.7 Resultados - Situación 4: Con oclusión y luminosidad baja - Método: *Regression Based Facial Landmark Detection.* 82
Tabla 5.8 Resultados - Situación 4: Con oclusión y luminosidad baja - Método: Viola-Jones.....83
Tabla 5.9 Resultados - Situación 1: Sin oclusión y luminosidad alta - Método: *Regression Based Facial Landmark Detection.* 84
Tabla 5.10 Resultados - Situación 1: Sin oclusión y luminosidad alta - Método: Viola-Jones. 85
Tabla 5.11 Resultados - Situación 2: Sin oclusión y luminosidad baja - Método: *Regression Based Facial Landmark Detection.* 86
Tabla 5.12 Resultados - Situación 2: Sin oclusión y luminosidad baja - Método: Viola-Jones. 88
Tabla 5.13 Resultados - Situación 3: Con oclusión y luminosidad alta - Método: *Regression Based Facial Landmark Detection.* 89
Tabla 5.14 Resultados - Situación 3: Con oclusión y luminosidad alta - Método: Viola-Jones. 90
Tabla 5.15 Resultados - Situación 4: Con oclusión y luminosidad baja - Método: *Regression Based Facial Landmark Detection.* 91
Tabla 5.16 Resultados - Situación 4: Con oclusión y luminosidad baja - Método: Viola-Jones.....92
Tabla 5.17 Resultados de AAC finales de cada situación. 93
Tabla 5.18 Resultados de DR finales de cada situación. 94
Tabla 5.19 Resultados de AAC finales de cada situación con Base de datos Propia 95
Tabla 5.20 Resultados de DR finales de cada situación con Base de datos Propia 95

ÍNDICE DE FIGURAS

Figura 3.1 Detección de las características de un rostro. 24
Figura 3.2 Detección de Somnolencia de un usuario al volante: Variables del comportamiento ocular (frecuencia de parpadeo, duración del cierre del parpadeo, porcentaje de ojos cerrados). 26
Figura 3.3 Curva de porcentaje de Población urbana de 1985-2015.Población Urbana (% del Total). . 28
Figura 3.4 Las 10 principales causas de defunción en países de ingresos bajos según la OMS. 29
Figura 3.5 Metodología Viola-Jones aplicado a un video del NTHU *dataset* en escenario de luminosidad alta y otro escenario con luminosidad baja. 32
Figura 3.6 Se aplican los cinco tipos de *Haar Features* a la sub-imagen obtenida en la metodología Viola-Jones para determinar que parte de la sub-imagen puede ser una característica relevante ya sea los ojos, boca o nariz. Por ejemplo, el tipo 3 es usado para la detección de la nariz. 33
Figura 3.7 A partir de los cinco tipos *Haar Features* se crean variaciones que luego son aplicadas de a las sub-imágenes para una mejor aplicabilidad. 33
Figura 3.8 Representación gráfica de la Imagen integral. 34
Figura 3.9 El algoritmo Adaboost determina que características son relevantes para la detección de una facción de la cara en particular y cual es irrelevante. 35
Figura 3.10 A partir de una Base de Datos de imágenes de caras e imágenes de no-caras se entrena al clasificador para que pueda determinar, en una imagen de entrada que parte de esta podría ser una cara. 36
Figura 3.11 El SVM hace una clasificación y determinación del estado de los ojos del usuario, es decir si se encuentran abiertos o cerrados. 37
Figura 3.12 Etapa pre procesamiento: Relación de 1:2 de los pixeles para poder crear un parche de una imagen. 39
Figura 3.13 Calculo del descriptor HOG Núcleos de una filtración de imagen. 39
Figura 3.14 Izquierda: valor absoluto del gradiente x. Centro: valor absoluto del gradiente y. Derecha: magnitud del gradiente. 40
Figura 3.15 Cálculo del histograma de gradientes de 8 x 8 cedulas, El parche RGB y gradientes se representan mediante flechas. 41
Figura 3.16 Representación de gradientes en cédulas de 8 x 8. 42
Figura 3.17 Creación del Histograma en cédulas de 8x8. 43
Figura 3.18 Creación del Histograma en cédulas de 8 x 8. 44
Figura 3.19 Creación del histograma 9-bin. El eje y indica las contribuciones de todos los pixeles de las células 8 x 8. 45
Figura 3.20 Bloque de normalización 16 x 16. 46
Figura 3.21 Visualización del histograma de gradientes orientados. 47
Figura 3.22 *Facial Landmarks Detection*. Puntos fáciles en el rostro. Aproximado de unos 80 puntos. 48
Figura 3.23 Coordenadas bidimensionales del ojo (formados por 6 puntos). 48
Figura 4.1 Método de Reconocimiento facial- Diagrama de Bloques. 50
Figura 4.2 Diagrama de detección de somnolencia con metodología Viola-Jones. 53
Figura 4.3 *Haar Features* aplicados a una imagen. 54
Figura 4.4 Representación esquemática del método cascada, todas las sub-ventanas entran al método de cascada para determinar si esa sub-ventana en específico podría ser una cara o no, en caso que, si podría serlo, pasa al siguiente estado, caso contrario se rechaza. 55
Figura 4.5 Diagrama de detección de somnolencia con metodología *Regression Based Facial Landmark Detection.* 57
Figura 4.6 *Facial Landmark* delimitados en la zona de los ojos (izquierdo y derecho). 59
Figura 4.7 Coordenadas en un plano bidimensional de los ojos. 59
Figura 4.8 Primeras pruebas con el reconocimiento de rostro Viola-Jones, El algoritmo reconoce el rostro y lo encasilla en un recuadro de color rojo. 60
Figura 4.9 A partir del reconocimiento del rostro, empieza el reconocimiento de ojos y boca del usuario para identificar si estos se encuentran abiertos o cerrados. 61
Figura 4.10 La implementación de la propuesta hecha a partir de la metodología Viola-Jones, detecta el rostro y arroja como resultado el estado "despierto". Reconocimiento de somnolencia webcam. Estado: despierto. 62

Figura 4.11 Se detecta en este caso la presencia de "fatiga" por parte del participante al Bostezar, el programa lo reconoce y lo alerta. Reconocimiento de somnolencia webcam. Bostezo, Estado: somnoliento. 62
Figura 4.12 En este caso el participante cierra los ojos por un tiempo determinado por eso el programa reconoce su "fatiga" y lanza la alarma. Reconocimiento de somnolencia webcam. Estado: somnoliento. 63
Figura 4.13 Segundo Experimento, el participante se presenta en un escenario con alta luminosidad. Reconocimiento de somnolencia webcam. Estado: despierto, luz solar. 63
Figura 4.14 Segundo experimento, el programa reconoce el estado de somnolencia del usuario ante la presencia de los ojos cerrados. Reconocimiento de somnolencia webcam. Estado: fatigado, luz solar. 64
Figura 4.15 Se presenta la limitación de Viola-Jones al momento de la presencia de rotación del rostro, el participante al girar su cara más de 45 grados, la metodología dicha no detecta el rostro de una manera eficiente. 65
Figura 4.16 Experimento video, se presenta el usuario en un escenario de alta luminosidad. Reconocimiento de somnolencia video. Estado: despierto. 66
Figura 4.17 Experimento Video, con una luminosidad alta, el programa reconoce el estado de fatiga del participante. Reconocimiento de somnolencia video. Estado: somnoliento. 67
Figura 4.18 Detección del rostro gracias al algoritmo *Histogram of oriented gradients* (HOG), luego se detecta los ojos gracias a que se aplica los *Facial Landmark*. 68
Figura 4.19 Aplicación de la metodología *Regression Based Facial Landmark Detection y Eye aspect ratio* (EAR). Primera prueba: rostro frontal a la cámara con los ojos abiertos y luminosidad alta. Resultado: Despierto. 68
Figura 4.20 Aplicación de la metodología *Regression Based Facial Landmark Detection* y *eye aspect ratio*. El EAR determina que ambos ojos se encuentran abiertos y el sistema seguirá pasando los *frames* hasta que vea un comportamiento contrario. 69
Figura 4.21 Aplicación de la metodología *Regression Based Facial Landmark Detection and eye aspect ratio*. El EAR determina que ambos ojos se encuentran cerrados. 69
Figura 4.22 Aplicación de la metodología *Regression Based Facial Landmark Detection and eye aspect ratio*. Detección de somnolencia, si los ojos se mantienen cerrados por un determinado tiempo, el sistema reconocerá que existe somnolencia y votará una alerta. 70
Figura 5.1 Capturas de imágenes de los tipos de escenarios existentes en la base de datos puestos a prueba (Imágenes capturadas de los videos proporcionados por en la base de datos NTHU DDD). 73
Figura 5.2 Capturas de imágenes de los tipos de escenarios existentes en nuestra base de datos puestos a prueba (elaboración propia con nuestra base de datos). 75

RESUMEN

Actualmente, la congestión vehicular se agudiza por el número de vehículos que aumenta cada año sobre todo en las tardes a partir de las 17:00 horas, esta situación produce en el conductor: estrés, aburrimiento, cansancio y sueño. Los efectos por la falta de sueño ocasionan: la disminución del tiempo de reacción, cansancio ocular, la visión se torna borrosa, menor concentración, etc.; los cuales influyen de manera directa en el incremento de accidentes de tráfico.

Esta tesis busca realizar una comparación entre dos metodologías de reconocimiento de expresiones faciales: Viola-Jones y *Regression Based Facial Landmark Detection*, los cuales han sido adaptados para la detección de somnolencia, para conocer cuál de ellas es la óptima y se adecua mejor a las condiciones variables de: oclusión, rotación de rostro e iluminación.

El análisis cuantitativo, utilizando la matriz de confusión, puso a prueba las metodologías propuestas en 4 escenarios (sin oclusión y con luminosidad alta, sin oclusión y con luminosidad baja, con oclusión y luminosidad alta, con oclusión y luminosidad baja), se realizó una comparación con los resultados obtenidos. Las pruebas realizadas, para el caso de la metodología Viola-Jones, en todos los escenarios, la inclusión de la detección del bostezo empeoraba la precisión de la detección de la somnolencia; el resultado cambia en la metodología *Regression Based Facial Landmark Detection*, ya que, al solo concentrarse en la apertura de los ojos, obtuvo un mejor resultado que Viola-Jones. Finalmente, de acuerdo a los resultados generalizados en la mayoría de los escenarios, la metodología *Regression Based Facial Landmark Detection* detecta mejor la somnolencia que Viola-Jones.

ABSTRACT

Currently, vehicle congestion is exacerbated by the number of vehicles that increases each year in the afternoons from 17:00 hours, this situation produces in the driver: stress, boredom, fatigue and sleep. The effects of lack of sleep cause: decreased reaction time, eye fatigue, vision becomes blurred, less concentration, etc .; The traffic flows.

This thesis seeks to make a comparison between two facial expression recognition methodologies: Detection of facial milestones based on Jones regression and regression, which have been adapted for the detection of drowsiness, for the knowledge of the answers is the best and better is better suited to the variable conditions of: occlusion, face rotation and lighting.

The quantitative analysis, using the matrix of confusion, tested the proposed methodologies in 4 stages (without occlusion and with high luminosity, without occlusion and with low luminosity, with occlusion and high luminosity, with occlusion and low luminosity), a Comparison with the corresponding results. The tests performed, in the case of the Viola-Jones methodology, in all scenarios, the inclusion of yawn detection worsened the detection accuracy of drowsiness; The result changes in the methodology Detection of facial milestones based on the regression, since, when only concentrating on the opening of the eyes, it obtains a better result than Viola-Jones. Finally, according to the generalized results in most of the scenarios, the methodology of Detection of facial milestones based on regression better detects drowsiness than Viola-Jones.

INTRODUCCIÓN

En la actualidad, el índice de accidentes automovilísticos tiene como principal causa la somnolencia en los conductores, ya que un conductor cansado o que se duerme durante su actividad no puede realizar maniobras evasivas para evitar colisiones o despistes. Según la Dirección General de Circulación Terrestre del Ministerio de Transportes y Comunicaciones (MTC) del Perú, el año pasado se registró 54,970 sujetos heridos y 2,826 murieron por accidentes de tránsito (MTC - Secretaría Técnica del Consejo Nacional de Seguridad Vial, 2018). Por estas cifras alarmantes de accidentes, no solo en el Perú, sino en el mundo entero, es necesario la implementación de un sistema de detección somnolencia efectivo en los vehículos para minimizar las probabilidades de que estos accidentes ocurran.

En el caso de Lima, la congestión vehicular se agudiza por el número de vehículos que sigue aumentando año tras año, la infraestructura está llegando a su máxima capacidad sobre todo en horas donde hay mayor número de conductores al volante, provocando en ellos: estrés, aburrimiento, cansancio y sueño. Ocurren diferentes efectos por la falta de sueño como la disminución del tiempo de reacción, cansancio ocular, la visión se torna borrosa, menor concentración, etc.; dichos factores influyen de manera directa en el incremento de accidentes de tráfico.

Fruto de la investigación que realizamos, se elaboró el artículo del mismo nombre "Adaptación Y Comparación De Dos Metodologías De Reconocimiento Facial Aplicados A La Detección De Somnolencia En Conductores" donde se desarrollan los temas más importantes de esta investigación como la explicación de los algoritmos que usamos, las pruebas realizadas y las conclusiones que llegamos. Este artículo se expuso ante el I CIIS (Congreso Internacional de Ingeniería de Sistemas) y actualmente, se encuentra publicado en las Actas del mismo.

DESCRIPTORES TEMÁTICOS

Detección de somnolencia.

Reconocimiento facial.

Accidentes de tráfico.

Viola-Jones.

Regression Based Facial Landmark Detection.

Eye Aspect Ratio (EAR).

CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA

En este capítulo se da la definición del problema de nuestro tema elegido y se describen los aspectos que han motivado la realización de esta investigación. Se justifica el interés tecnológico y social que actualmente tiene que resolver este problema. A continuación, se describen la formulación del problema en el punto 1.1, luego se definirán el objetivo general en el punto 1.1.1 y los objetivos específicos en el punto 1.1.2. Finalmente, en el punto 1.3 se describirá la justificación de la investigación.

1.1 Formulación del problema

El presente trabajo trata el tema de accidentes vehiculares, en el caso específico de somnolencia en los conductores, accidentes que podrían evitarse con ciertas pautas fáciles de seguir, como dormir adecuadamente. Principalmente en Lima, ciudad en creciente desarrollo, al no haber una infraestructura preparada para una gran cantidad de autos que circulen que, a su vez, producen un congestionamiento en horas punta, el número de accidentes automovilísticos aumentan. Según la Dirección General de Circulación Terrestre del Ministerio de Transportes y Comunicaciones (MTC) del Perú, el año 2017 se registró 54,970 sujetos heridos y 2,826 murieron por accidentes de tránsito (MTC - Secretaría Técnica del Consejo Nacional de Seguridad Vial, 2018).

Tratar de disminuir el número de heridos y muertos en los accidentes de tráfico causados por la somnolencia nos parece una tarea solidaria, ayudar a que ya no ocurran este tipo de situaciones mediante la detección temprana del sueño en conductores para que se tomen medidas preventivas.

Existen varias técnicas para el reconocimiento de expresiones faciales; sin embargo, se encuentran dificultades al momento de extraer las características faciales, tales como la oclusión, luminosidad, calidad de la imagen, la cercanía o lejanía del rostro, etc. Estos problemas mencionados entorpecen los resultados del reconocimiento, por lo tanto, disminuye la tasa de efectividad de estos.

Los problemas más importantes en la detección de expresiones faciales:

- Oclusión facial: Se presenta cuando objetos o accesorios dificultan la detección del rostro como gafas, gorras, cabello en la cara. Los sistemas de

reconocimientos faciales pueden alcanzar sólo el 10 por ciento de precisión bajo estas condiciones (Akhtar y Rattani, 2017).

- Luminosidad: La variación de la iluminación es una preocupación para el reconocimiento, en diferentes condiciones de iluminación se encuentran sustanciales diferencias en cuantos a los valores de los píxeles en una misma imagen. Por ejemplo, se tiene dos imágenes con rostros iguales y diferente iluminación, estas presentan mayor variación que dos imágenes con rostros diferentes con la misma iluminación (Akhtar y Rattani, 2017).
- Calidad de Imagen: Los cambios inducidos por la calidad de imagen podrían ser mayores que las diferencias entre individuos para los sistemas basados en la comparación de imágenes entonces se clasifica erróneamente la imagen de entrada.
- Ángulo de inclinación del rostro: El reconocimiento *multiview* es una manera de abordar las variaciones de pose, pero los algoritmos requieren imágenes de galería en cada pose, lo cual es poco práctico (Akhtar y Rattani, 2017).
- Fondo de imagen: La detección del rostro solo, en un fondo no controlado sigue siendo problemático, porque aumenta las complejidades en la detección, el ambiente podría tener una iluminación deficiente, por ejemplo, o un desorden de fondo (Akhtar y Rattani, 2017).

El problema radica en que no se tienen una metodología que pueda superar todos estos problemas o por lo menos los más importantes y conseguir un porcentaje de reconocimiento bastante alto para llegar a ser preciso.

1.2 Objetivo de la investigación

1.1.1 Objetivo general

Comparar dos metodologías de reconocimiento de expresiones faciales: Viola-Jones y *Regression Based Facial Landmark Detection*, adaptados para la detección de somnolencia, dando a conocer cuál de ellas es superior en términos de eficacia y se adecua mejor a las condiciones variables, considerando las restricciones de: oclusión, rotación de rostro, iluminación.

1.1.2 Objetivos específicos

- Modificar la metodología de reconocimiento facial de Viola-Jones para el reconocimiento de somnolencia.
    - Modificar la metodología de reconocimiento facial de *mark Detection* para el reconocimiento de somnolencia.
- Generar una base de datos de personas con rasgos latinos para la posterior comparación de los algoritmos.
- Comparar las metodologías Viola-Jones y *Regression Based Facial Landmark Detection*, identificar la más eficaz y fiable que sea capaz de procesar la información de los frames de video, reduciendo el número de falsos negativos, es decir, que no se detecta la somnolencia.

1.3 Justificación

La presente investigación se enfocará en estudiar dos metodologías de reconocimiento facial y demostrar cuál de ellas obtiene una mayor tasa de efectividad, ya que al momento de extraer las características faciales surgen numerosas dificultades como la luminosidad, la calidad de imagen, la oclusión, la distancia del rostro frente a la cámara, que entorpecen los resultados del reconocimiento disminuyendo así la tasa de efectividad.

Respecto al ámbito social, el sistema contribuye en la disminución de los accidentes alertando a los conductores. Un accidente automovilístico puede traer graves consecuencias, desde dejar a una persona afectada físicamente o en el peor de lo casos

causarle la muerte. Por ello el sistema de detección de somnolencia ayudará a los conductores a evitar estos accidentes.

Finalmente, en el campo tecnológico, se pretende brindar una solución a una de las principales causas de accidentes automovilísticos haciendo uso de la visión computacional y llegar a la determinar cuál de estas metodologías de reconocimiento facial (Viola-Jones o *Regression Based Facial Landmark Detection*) funciona mejor para la detección la somnolencia en conductores y se adapta mejor a las condiciones variables tales como rotación de rostro, luminosidad y oclusión.

CAPÍTULO II: ESTADO DEL ARTE

Diversos trabajos de investigación relacionados con el tema del análisis del comportamiento del conductor han sido publicados en esta última década dentro del sector automotriz, dando a conocer que los principales factores que puntuales durante el proceso de conducción y que en muchos casos terminan en accidentes son: la somnolencia y la distracción. Por ello, en este capítulo se presenta el estado del arte y los antecedentes sub dividiéndolo en 3 secciones: la primera sección se tratará sobre las investigaciones que hacen referencia a los temas de reconocimiento de expresiones faciales en general, la segunda sección tratará sobre los temas relacionados directamente con la detección de somnolencia; y finalmente, la tercera sección será los temas sobre las metodologías de Viola Jones y *Regression Based Facial Landmark Detection* que son los métodos a comparar en este trabajo de investigación.

2.1 Reconocimiento facial

En esta primera sección se presentan algunos artículos relacionados netamente con el reconocimiento facial ya sea en tiempo real para reconocer una expresión facial en específica o alguna técnica para reconocer el rostro de una persona dentro de un conjunto de datos.

En el artículo "A Face in any Form: New Challenges and Opportunities for Face *Recognition Technology*." (Akhtar y Rattani, 2017) nos comenta una serie de métodos de reconocimiento facial, los algoritmos más representativos, los problemas que se presentan con más frecuencia, como las complejidades de la pose de la cara (el ángulo de rotación de la misma), orientación, ubicación en la imagen/vídeo y por último la oclusión (parte del rostro está oculto u obstruido). El articulo nos menciona algunos métodos como: *Local binary pattern* (LBP), *Active appearance model* (AAM), *Simulated annealing for 3D face recognition*, *DeepFace deep neural network*, entre otras; que se usan para el reconocimiento facial, los problemas que llega a ocurrir en la aplicación de estos y finalmente algunas posibles soluciones que se podrían dar para robustecerlos. Los sistemas de reconocimiento facial (SRF), además de almacenar las imágenes de los rostros también se guarda una plantilla de conjunto de características que son una especie de metadatos extraídos de los rasgos que posee el individuo, tales

como la edad, el género, la raza, entre otros. Se plantean también los escenarios de seguridad y privacidad y cómo estas imágenes deben estar totalmente protegidas, ya que existen ataques, como el que un impostor se disfraza del usuario auténtico a través de una foto/video. Para esto se plantean 4 propiedades:

1. Se debe prohibir la interconexión entre bases de datos.
2. Se debería ser capaz de revocar una plantilla comprometida.
3. Tener una seguridad alta al momento del acceso a la información.
4. No debe degradar el rendimiento del FRS.

Actualmente, el problema recae en que todos los métodos convencionales de reconocimiento facial tienen varias deficiencias y estas causan la disminución del porcentaje de precisión de la detección. Todos los métodos al momento de probarlo lo hacen en un ambiente controlado y este ambiente se dan unos resultados buenos, pero el problema viene cuando se trata de probarlo en un ambiente no controlado, el porcentaje de precisión disminuye por razones de iluminación, pose de la cara, uso de lentes incluso el dejarse crecer la barba o bigote disminuye la ratio. Este trabajo nos muestra una serie de posibles soluciones a los problemas que disminuyen drásticamente la tasa efectividad de reconocimiento y algunas tendencias en este ámbito, aplicando nuevas técnicas como la de las imágenes *multiview* la cual consiste en almacenar la misma imagen, pero con diferentes ángulos de rotación para entrenar mejor a los métodos y puedan reconocer una imagen que no está 100% centrada mirando hacia la cámara.

Después de abordar un poco en los diferentes problemas que presentan la mayoría de métodos de reconocimientos facial, el artículo "*Real Time Facial Expression Recognition App Development on Mobile Phones*" (Alshamsi y Meng, 2016) propone un sistema automático para reconocer las emociones de las imágenes faciales en un teléfono inteligente en tiempo real. Se explica que las etapas principales para el reconocimiento de la expresión facial son: la extracción de características que contiene la detección y el seguimiento de los movimientos del rostro haciendo uso de cascadas HAAR aplicando el método *Binary Robust Independent Elementary Features* (BRIEF). Este es un descriptor binario que se basa en la comparación entre 2 intensidades de pixel suavizadas asignando el binario 0 al pixel que tiene menor intensidad y al que tiene mayor intensidad el binario 1. La otra etapa es la clasificación de características donde se hace uso del algoritmo *K-Nearest Neighbor* (K-NN) que estima el vecino más cercano depende del espacio entre los ejemplos de la investigación y las pruebas de muestra de entrenamiento. Este algoritmo define lo más cercano a la

luz de la distancia euclidiana. Luego de extraer las características, estas se almacenan en una base de datos a modo de prueba; cuando se ejecuta el código, esté requiere de las imágenes de entrenamiento y luego las imágenes de prueba. Para llevar a cabo con su objetivo, hizo uso de un método híbrido bajo la consideración de 3 algoritmos bases: BRIEF, CAPP y SPTS. Con ello, obtuvo una tasa de efectividad del 89.5% en el rango de detección que es mayor a la tasa de efectividad (70.34%) cuando se hace uso de los métodos por sí solos; esto se debe a que se examinó un mayor número de imágenes de diferentes bases de datos cada una con diferentes rasgos faciales, diferentes expresiones. Este artículo muestra que se utilizó un método híbrido en base a 3 algoritmos y este le proporciona una mayor tasa de rendimiento. Sin embargo, este algoritmo funciona a nivel de pixeles y por lo tanto es sensible al ruido de las imágenes, así que se tendría que contar con una cámara con una resolución media-alta para que se dé la efectividad deseada.

En el artículo "*Fuzzy Linear Regression Discriminant Projection for Face Recognition*" (Huang, Gao, Qian, Yang, y Yang, 2017) se propone un nuevo algoritmo de extracción denominado *Fuzzy Linear Regression Discriminant Projection* (FLRDP) este busca generar un sub espacio de características eficiente para el método *Linear Regression Classification* (LRC) y podrá manejar eficazmente las variaciones entre las imágenes faciales, ya que capturar características distintivas de las imágenes faciales es difícil, ya que existen variaciones de iluminación, poses, expresiones, etc. Además, la extracción (reducción de dimensionalidad y reducir la información irrelevante); técnica conocida: *Principal Component Analysis* (PCA) y la clasificación de características (clasificador para reconocer las características extraídas; método conocido: LRC) son dos problemas fundamentales; ya que la mayoría de técnicas de extracción se diseñaron independientes de los métodos clasificadores.

Para llevar a cabo con su objetivo, primero, se calcula los grados de pertenencia gradual de cada muestra a las clases correspondientes haciendo uso del algoritmo *fuzzy k-nearest neighbor* (FKNN) para evaluar las contribuciones de las muestras de entrenamiento. Luego, dicha información se incorpora a la construcción de los errores de reconstrucción entre clases y dentro de clases. Finalmente, la función criterio se obtiene maximizando la relación entre el error de reconstrucción entre clases difuso y el error de reconstrucción dentro de la clase difuso. Las ventajas de este algoritmo es que sigue la regla de decisión del método LRC y al evaluar las contribuciones de las muestras de entrenamiento, se adoptan los grados de membresía gradual de cada

muestra a las clases correspondientes haciendo que esta se vuelva más robusta e insensible a las variaciones (iluminación, poses, expresiones). Todas las técnicas mencionadas buscan sobrepasar los problemas mencionados anteriormente en el artículo "*A Face in any Form: New Challenges and Opportunities for Face Recognition Technology*" (Akhtar y Rattani, 2017) que se tiene a la hora de obtener las imágenes o a la hora de procesarlas para el reconocimiento facial. En este *paper* en específico nos da a conocer otra técnica de extracción de característica: *fuzzy linear regression discriminant projection* (FLRDP), la cual se adapta al método LRC para la clasificación, haciendo más robustas las características e insensibles a las variaciones de luminosidad, poses y expresiones, ya que se sabe que las imágenes son muy sensibles y esto puede afectar a la efectividad del reconocimiento facial (Huang, Gao, Qian, Yang, y Yang, 2017).

2.2 Reconocimiento de la somnolencia.

En esta segunda sección se resumirán algunos de los trabajos investigados sobre el reconocimiento de somnolencia y se hablará sobre qué investigaciones se han hecho en este ámbito, como las aplicaciones y usos que se le da para la vida cotidiana para ello el primero que se comentará será sobre los métodos para la detección de la somnolencia.

En el artículo "*Driver drowsiness detection based on non-intrusive metrics considering individual specifics*" (Wang y Xu, 2015), mantiene que la detección de la somnolencia puede separarse en dos categorías de métodos: los intrusivos y no intrusivos. Los métodos intrusivos, como electroencefalogramas (EEGs) o electrocardiogramas (EKG), muestran una buena precisión de detección, sin embargo, se limitan al laboratorio de investigación. Por el contrario, los métodos basados en medidas no intrusivas detectan la somnolencia al medir el comportamiento de conducción y a veces los rasgos oculares, por lo que son útiles para situaciones de conducción del mundo real. Hasta el día de hoy, los métodos no intrusivos han sido menos fiables que los métodos intrusivos. En la investigación, las diferencias individuales de los métodos no intrusivos se mencionan con frecuencia, tanto para el comportamiento de conducción y características del ojo. Para el comportamiento de conducción, se informó que una medida es el *Standard deviation of lateral position* (SDLP), y esta investigación encontró que, para el mismo nivel de somnolencia, diferentes conductores tienen SDLP diferentes. En cuanto a los rasgos oculares,

también se observaron diferencias individuales en la duración del cierre del párpado (PERCLOS) en muchos estudios (Wang y Xu, 2015).

En el trabajo de Zhang, Cheng y Lin exponen en su artículo "*Driver Drowsiness Recognition Based on Computer Vision Technology*" (Cheng, Zhang, y Lin, 2012) por su parte, su objetivo de presentar un método de detección de la somnolencia no intrusivo utilizando el rastreo ocular mediante una cámara y el procesamiento de imágenes, se introduce un algoritmo de detección de ojos robusto *Active Shape Model* (ASM), que se utiliza para localizar con precisión los ojos humanos a partir de las imágenes de vista frontal, para sobrepasar los problemas causados por la postura del conductor y los cambios en la iluminación. Para determinar el nivel de somnolencia de la persona se proponen seis medidas: porcentaje de cierre del párpado, duración máxima del cierre, frecuencia de parpadeo, nivel medio de apertura de los ojos, velocidad de apertura de los ojos y velocidad de cierre de los ojos. Para mejorar la velocidad y precisión del algoritmo ASM, en cuanto al problema de la luminosidad, se utilizarán las imágenes *Self-quotient*, estas son imágenes que se les ha eliminado el efecto de iluminación, ya que el contorno facial se resalta en estas imágenes, en vez de las imágenes originales captadas por la cámara. Los resultados con seis participantes en experimentos simuladores de conducción demuestran la viabilidad de este método de reconocimiento de somnolencia basado en video que proporcionó una precisión del 86% (Cheng, Zhang, y Lin, 2012). El artículo explica una metodología para la detección de somnolencia no intrusiva, el algoritmo ASM, y en esta investigación al algoritmo se le introduce una variante para aumentar la precisión, se utiliza como entrada las imágenes *Self-quotient*, en vez de las imágenes originales para resolver el problema de la luminosidad cambiante que ocurre cuando uno se transporta en auto.

Otra técnica para detectar la somnolencia en conductores se explica en el artículo "*Real-time driver drowsiness feedback improves driver alertness and self-reported driving performance*" (Aidman, Chadunow, Johnson, y Reece, 2015) donde los autores explican que se realizan pruebas a quince miembros del personal de la Reserva del Ejército utilizando el *Optalert Alertness Monitoring System* (OAMS) para monitorear los movimientos del párpado. El sistema utiliza un emisor de IR y un sensor montado en un marco de gafas para medir continuamente la velocidad de parpadeo de los ojos, de los que se derivan los niveles de somnolencia. El OAMS ha sido utilizado para la detección y monitoreo de la somnolencia del conductor en la minería y en las

industrias de transporte por carretera. Su estudio tiene como objetivo examinar la eficacia de OAMS en la mejora del estado funcional del conductor y se centra en la respuesta individual y la comprensión de la utilidad de la retroalimentación en tiempo real a los conductores individuales por ejemplo se generan sonidos de pitido cuando los puntajes alcanzaron niveles de somnolencia de riesgo medio y alto. La retroalimentación y advertencias dio como resultado un 28% menos de advertencias de riesgo medio y un 41% menos de advertencias de alto riesgo comparadas con la condición base sin retroalimentación (Aidman, Chadunow, Johnson, y Reece, 2015). Alertar al conductor en tiempo real emitiendo un sonido tuvo buenos resultados y según los autores del presente trabajo se consiguió un resultado mejor que de los que no les alertaba.

En la investigación *"Drowsy Driver Detection using Representation Learning*" (Dwivedi, Biswaranjan, y Sethi, 2014) se utilizan las redes convolucionales para la detección de somnolencia en conductores y ver cómo estas ayudan en aumentar la eficacia en predecir si el sujeto se encuentra somnoliento o no. En este artículo propone un esquema basado en la extracción de rasgos visuales de los datos sin intervención humana. Estas características visuales han sido aprendidas usando un modelo de aprendizaje profundo conocido como redes neuronales convolucionales para capturar explícitamente varias características faciales latentes y las complejas interacciones de características no lineales. Se creó un conjunto de datos para capacitar al clasificador y evaluar el desempeño del esquema. Se creó un conjunto de datos diverso que involucra a 30 sujetos. Se adoptaron diferentes condiciones de iluminación para hacer el conjunto de datos aún más universal, teniendo en cuenta las diferentes condiciones de brillo en escenarios de la vida real. Por lo tanto, el clasificador sería más robusto y eficiente en todas las circunstancias. Los sujetos también usan anteojos en pocas secuencias de vídeo para agregar aún más a la diversa naturaleza y dificultad del conjunto de datos. El método propuesto tiene como objetivo clasificar los marcos en videos basados en rasgos faciales especiales aprendidos a través de la red neuronal convolucional. En primer lugar, los cuadros se extraen del vídeo. Estos marcos alimentan a un Viola y Jones como características basadas detector de cara. Sobre la base de estas características, el clasificador de capa *softmax* fue entrenado. Una vez que el clasificador ha sido entrenado, el veinte por ciento de las imágenes extraídas anteriormente se prueban en el clasificador entrenado (con el 80% de las imágenes). Para que una señal de alerta sea

entregada al conductor. Se llevó a cabo otro experimento en el que se entrena el clasificador en un conjunto de sujetos y la prueba se realiza en una variedad absolutamente diferente de personas que tienen diferentes características físicas y faciales. En este caso se encontró un promedio satisfactorio de un 78% de precisión entre los sujetos. Por lo tanto, el clasificador basado en el aprendizaje profundo propuesto detecta la somnolencia del conductor basada únicamente en rasgos faciales visuales de manera eficiente en un conjunto de datos diverso (Dwivedi, Biswaranjan, y Sethi, 2014). Después de los experimentos realizados, podemos concluir que la detección de somnolencia en conductores es capaz de dar resultados razonables con éxito en un conjunto de datos diverso, pero todavía hay cierto margen para aumentar su desempeño. Las redes neuronales convolucionales basadas en la representación característica enfoque de aprendizaje proporciona un conjunto automatizado y eficiente de características que nos ayudan a clasificar el conductor como somnoliento o no somnolencia con un porcentaje de precisión muy alto.

Siguiendo con los métodos no intrusivos, en el artículo "*Real-time physiological and vision monitoring of vehicle driver for non-intrusive drowsiness detection*" (Lee, Jung, y Chung, 2011) adiciona un segundo elemento para mejorar la precisión. Este estudio presenta un nuevo enfoque para detectar la somnolencia del conductor mediante la aplicación de dos métodos distintos: La visión por computadora y el procesamiento de imágenes. El objetivo de este estudio fue combinar ambos métodos en un solo perfil para mejorar la precisión de detección de somnolencia del conductor. Por lo tanto, se desarrolla un sistema de monitoreo de somnolencia no intrusivo para alertar al conductor si el conductor cae en estado somnoliento. En la parte fisiológica, la fotopletismografía (*photoplethysmography* PPG) se analiza por sus cambios en la forma de onda de las señales desde el estado despierto al estado somnoliento. Mientras tanto, el patrón de los ojos o el movimiento en el procesamiento de la imagen se dirige a detectar la fatiga del conductor. El sistema desarrollado consta de dos módulos, que son el módulo de bio-señales para analizar los cambios de forma de onda PPG y el módulo de expresión de características faciales para analizar las características de movimiento. El sistema adquirió dos fuentes de entrada, señales de PPG a través de la piel del pulgar debajo de la oximetría de pulso y capturó el movimiento facial a través de los marcos de secuencia de vídeo con una cámara web USB PC, un filtro de paso de banda atenúa todas las frecuencias de señales PPG que están fuera del paso de banda. Se utiliza un

algoritmo genético (*Genetic algorithm* GA) para buscar el mejor ajuste en la clasificación de índice de sueño para ambos movimientos de los ojos y las variantes de PPG. Los valores de aptitud derivados en GA pueden predecir el estado actual del estado de alerta del conductor. Si el valor de la aptitud máxima cae por debajo del umbral especificado, se sospecha que el conductor está en el nivel de somnolencia crítico y se dispara una alarma para advertir al conductor de las precauciones de seguridad. Finalmente, el GA puede ser reutilizado para indicar los parámetros de visión y bio-señal, y finalmente para la predicción de somnolencia. Se realizó el experimento con un total de seis personas con cuatro hombres y dos mujeres. Cada conductor realizó experimentos durante una hora continua en tres situaciones diferentes. El sistema indicó que después de la hora del almuerzo, las personas A y B empiezan a caer en estado somnoliento debido a las insuficientes horas de reposo y la condición mental incómoda. En conclusión, el sistema puede indicar con éxito el estado de alerta de cada conductor con precisión alcanzada hasta el 97% (Lee, Jung, y Chung, 2011).

Luego, en el artículo "*Heart Rate Variability Can Be Used to Estimate Sleepiness-related Decrements in Psychomotor Vigilance during Total Sleep Deprivation*" (Chua, y otros, 2012) se incluye la medición del ritmo cardiaco para la detección de somnolencia. Dado que la somnolencia aumenta el riesgo de una persona de fallas de atención, errores y accidentes, se han desarrollado muchas tecnologías para evaluar la vigilancia y los niveles de rendimiento en tiempo real. Estas tecnologías apuntan a advertir al usuario cuando es probable que ocurra una caída en el rendimiento y, si se combina con una intervención apropiada, podría evitar que ocurra un accidente. Las tecnologías que buscan monitorear en línea el estado fisiológico de un individuo se han centrado principalmente en las características de los ojos, la cara y la cabeza. Se han examinado varias medidas oculares como indicadores de fatiga, incluidos parpadeos en los ojos, respuestas de la pupila, movimientos de los ojos y porcentaje de cierre del párpado sobre la pupila a lo largo del tiempo (PERCLOS). De estas medidas, se ha demostrado que PERCLOS se correlaciona con los lapsos en la atención visual y las medidas del rendimiento de conducción simulado durante la pérdida de sueño. Además, las nuevas tecnologías permiten el registro sin contacto de las mediciones de latidos cardíacos. Recientemente, se demostró que algunas medidas de la variabilidad del ritmo cardíaco (*heart rate variability* - HRV) se pueden usar para predecir los niveles de rendimiento diurnos.

En el presente estudio, estuvieron presentes 24 hombres participantes con una edad media de 26 años. Los sujetos se mantuvieron despiertos continuamente durante 40 horas en condiciones ambientales constantes. Cada 2 horas, los sujetos completaron una Tarea de Vigilancia Psicomotora (*psychomotor vigilance task* - PVT) de 10 minutos para evaluar su capacidad para mantener la atención visual. Durante cada PVT, examinamos el electrocardiograma (ECG), el electroencefalograma (EEG) y el porcentaje de tiempo que los ojos estuvieron cerrados (PERCLOS). Durante las horas de sueño habituales, el tiempo de reacción y el número de fallas de la PVT por sesión aumentaron considerablemente, alcanzando sus niveles más altos aproximadamente 24 horas después de la vigilia. A partir de entonces, el rendimiento de PVT mejoró, pero no se recuperó completamente a los niveles de referencia cuando los sujetos estaban descansados (es decir, el rendimiento durante las primeras 16 horas de vigilia). Esta mejora parcial en el rendimiento de la PVT después de 24 horas de vigilia se debió probablemente a un aumento del impulso circadiano para el estado de alerta, que sigue de cerca el ritmo circadiano de la temperatura corporal. El electrocardiograma (ECG), transporta información sobre el estado de vigilancia de una persona. Por lo tanto, las medidas de HRV podrían potencialmente usarse para predecir cuándo un individuo tiene un mayor riesgo de falla atencional (Chua, y otros, 2012).

Finalmente, en la investigación "*Drowsiness detection using heart rate variability*" (Vicente, Laguna, Bartra, & Bailón, 2016) complementa información al anticulo mencionado previamente sobre la detección de somnolencia usando el ritmo cardiaco, Los trastornos del sueño tienen una mayor tasa de accidentes en comparación con individuos que se encuentran despiertos. Se estima que el 10-30% de las muertes en carretera están relacionados con la conducción somnolienta. Por ello, se están realizando estudios para poder detectar la somnolencia en conductores basada en señales biológicas. La actividad del sistema nervioso autónomo (SNA) presenta alteraciones durante los episodios de estrés, fatiga extrema y somnolencia. Se puede medir de forma no invasiva a partir de la señal de variabilidad de la frecuencia cardíaca (HRV) obtenida del electrocardiograma de superficie. Se definen 2 estados: estado de vigila que se caracterizan por un aumento de la actividad simpática y / o disminución de la actividad parasimpática y estado de relajación extrema que se caracteriza por un aumento de la actividad parasimpática y / o una disminución de la actividad simpática. Se planteó la hipótesis de que se podría detectar la somnolencia del conductor a través

de la variabilidad de frecuencia cardiaca (HRV). Se analizó 3 bases de datos de conducción proporcionado por Fico Mirrors S.A y se desarrollaron 2 detectores de somnolencia diferentes basados en HRV. El primero, es un detector de episodios de somnolencia, el cual evalúa periódicamente el estado del conductor como despierto (ON) o somnoliento (OFF) y su salida podría ser utilizada para activar una alarma de alerta. El segundo, es un detector de privación del sueño que evalúa el estado global del conductor, diferenciando si un conductor era adecuado para conducir o no; en función a su estado de privación de sueño, a partir de los 3 primeros minutos de conducción. Una de las limitaciones del estudio el número de episodios despiertos frente a somnolencia fue desequilibrado entre las bases de datos RDB y SDB. En el primero, los sujetos no fueron privados de sueño, y la mayoría de sus minutos fueron etiquetados como despiertos. Por otro lado, en la segunda base, los sujetos estaban privados de sueño, y tenían la mayoría de sus minutos etiquetados como somnolientos. Para superar este desequilibrio se realizó el análisis de los datos por tipo, agrupando todos los minutos de todas las asignaturas en ambas bases de datos. Se debería considerar usar muestras de mayor cantidad de personas y / o usar un diseño de estudio cruzado para poder dar cuenta la variabilidad de las características del sujeto. De la prueba, 59% de los episodios de somnolencia son detectados, 96% de estos son detectado adecuadamente. Esto sugiere que el sistema podría ayudar a reducir los accidentes relacionados con la somnolencia, especialmente si se combinan con los sistemas existentes de prevención de seguridad del automóvil. La incorporación de la evaluación de la somnolencia basada en la señal HRV puede agregar mejoras significativas a los sistemas de seguridad existentes en los automóviles (Vicente, Laguna, Bartra, y Bailón, 2016).

2.3 Método para la detección de somnolencia: Viola Jones y *Regression Based Facial Landmark Detection.*

Para esta última sección se comentarán las distintas investigaciones que se han hecho sobre los dos métodos para la detección de somnolencia sobre los que estamos comparando sus resultados en este trabajo de investigación.

El primer artículo "*Study of Viola-Jones Real Time Face Detector*" (Cen, 2016) nos presenta un método de detección de rostros, Viola-Jones que es capaz de procesar imágenes muy rápidamente y lograr altas tasas de detección, existen 3 componentes clave para este algoritmo de detección: el primero, la "Imagen Integral", una

representación de imagen, permite que las características usadas por el detector sean computadas muy rápidamente; el segundo es un algoritmo de aprendizaje, basado en *AdaBoost*, este selecciona una cantidad mínima de características visuales que son críticas, de un conjunto más grande y produce clasificadores extremadamente eficientes; el último componente central es un método para combinar clasificadores cada vez más complejos en una "cascada", permite que el fondo de la imagen sea descartado rápidamente mientras se gasta más cálculo en regiones prometedoras de tipo cara, luego de introducir los 3 componentes se realiza una implementación y se analiza las limitaciones que tuvieron de rendimiento e introduce una solución experimental utilizando las características de *Histogram of Oriented Gradients* (HOG) con el clasificador *Support Vector Machines* (SVM). Este artículo explica e implementa la metodología Viola-Jones para la detección de rostros, la detección del rostro significa la primera parte de nuestro tema de investigación para luego adicionarle un componente para que detecte la somnolencia. Sin embargo, también se presenta la dificultad de las imágenes con rostros no frontales, pero utiliza HOG + SVM y aumenta la tasa de efectividad en un 20% aproximadamente, pero también aumenta la tasa de falsos positivos de 1.7% a 5.9%.

La técnica de *Support Vector Machine*, se utiliza en múltiples campos, pero en esta investigación "*Driver drowsiness detection with eyelid related parameters by Support Vector Machine*" se describe un sistema para la detección de somnolencia al reconocer si los ojos del conductor están abiertos o cerrados y, si están abiertos, calcular el grado de apertura. El propósito de este, es estudiar el enfoque de usar múltiples funciones de movimiento del párpado para detectar la somnolencia de los conductores con una técnica de aprendizaje de máquina, *Support Vector Machine* (SVM), para detectar y realizar la predicción de la somnolencia de los conductores con múltiples funciones de movimiento del párpado. Con más de una característica, se espera que tanto la precisión como la robustez de la predicción puedan mejorarse significativamente. A partir de los datos experimentales del *Electrooculogram* (EOG), se extraen las funciones de movimiento de párpados necesarias para el entrenamiento y la validación de SVM. El modelo SVM (clasificador) que se construyó en este trabajo será entrenado con las 11 características relacionadas con el párpado extraídas del EOG con tres niveles diferentes de somnolencia y luego usarlas para detectar la condición del conductor, con un conjunto de datos de validación. Los resultados muestran que la

precisión de detección de somnolencia mediante esta técnica es bastante alta, especialmente cuando los sujetos están muy somnolientos, lo que faltó fue la implementación de un tipo de alerta hacia el sujeto que se va quedando dormido en pleno manejo del automóvil, opción que sí se planteó en otras investigaciones y aumentó el *ratio* de precisión.

Otro de las metodologías que se usará es el *Regression based Facial Landmark Detection*, que estaremos comparando con Viola-Jones más adelante en el trabajo de investigación. En el artículo "*Real-Time Eye Blink Detection using Facial Landmarks*" (Soukupova & Cech, 2016) se habla sobre la desafiante tarea de la detección de seres humanos en imágenes debido a su apariencia variable y la amplia gama de poses que pueden adoptar. El autor nos menciona que lo primero que se necesita es un conjunto de características robusto que permita que la forma humana sea discriminada limpiamente, incluso en fondos desordenados bajo una iluminación difícil. Bajo los estudios sobre el tema de los conjuntos de características para la detección humana, que muestran que los descriptores de Histograma de gradiente orientado (HOG) normalizados localmente proporcionan un rendimiento excelente en relación con otros conjuntos de características existentes. Este método se basa en la evaluación de histogramas locales bien normalizados de orientaciones de gradiente de imagen en una grilla densa. La idea básica es que la apariencia y forma del objeto local se pueda caracterizar por la distribución de los gradientes de intensidad locales o las direcciones de los bordes, incluso sin un conocimiento preciso de la gradiente correspondiente o las posiciones de los bordes. Esto se implementa al dividir la ventana de imagen en pequeñas regiones espaciales (celdas), para cada celda que acumula un histograma 1-D local de direcciones de gradiente u orientaciones de borde sobre los píxeles de la celda. Para una mejor invariancia a la iluminación, sombreado, etc., también es útil normalizar el contraste de las respuestas locales antes de usarlas. Esto se puede hacer acumulando una medida de "energía" del histograma local sobre regiones espaciales algo más grandes ("bloques") y utilizando los resultados para normalizar todas las celdas del bloque. La representación HOG / SIFT tiene varias ventajas: captura la estructura de borde o degradado que es muy característica de forma local, y lo hace en una representación local con un grado de invariancia fácilmente controlable a las transformaciones geométricas y fotométricas locales. Para la detección humana, el muestreo espacial más bien grueso, el muestreo de orientación fina y la fuerte normalización fotométrica local

resultan ser la mejor estrategia, presumiblemente porque permite que las extremidades y los segmentos corporales cambien de apariencia y se muevan de un lado a otro siempre que mantengan una orientación vertical.

Un indicador principal en la detección de somnolencia dentro del algoritmo de *Regression based Landmark detector* es el parpadeo de los ojos. La detección de parpadeos es importante en sistemas que monitorean la vigilancia de un operador humano, por ejemplo, en la somnolencia de conductores, en sistemas que advierten a un usuario de computadora mirando a la pantalla sin parpadear durante mucho tiempo para evitar el ojo seco y los síndromes de visión por computadora. Según el artículo "*Real-Time Eye Blink Detection using Facial Landmarks*" (Soukupova & Cech, 2016), el autor propone un algoritmo en tiempo real para detectar parpadeos en una secuencia de video desde una cámara estándar. Esta estima las posiciones históricas, extrae una cantidad escalar única que refleja un nivel de apertura del ojo - relación de aspecto del ojo (EAR) - que caracteriza la apertura de los ojos en cada marco. Finalmente, un clasificador SVM detecta los parpadeos como un patrón de valores EAR en una ventana temporal corta.

El parpadeo del ojo es un cierre rápido y la reapertura de un ojo humano. Cada individuo tiene un patrón de parpadeos un poco diferente. Este patrón difiere en la velocidad de cierre y apertura, un grado de apretar el ojo y en un abrir y cerrar de ojos. El parpadeo del ojo tiene una duración aproximada de 100-400 ms. Se propone explotar detectores de referencia faciales de vanguardia para localizar los ojos y los contornos de los párpados. A partir de los puntos de referencia detectados en la imagen, obtenemos la relación de aspecto del ojo (EAR) que se utiliza como una estimación del estado de apertura de los ojos. Si bien son robustos a la baja calidad de imagen (baja resolución de imagen en gran medida) y en los fenómenos naturales como la no frontalidad, mala iluminación, expresiones faciales, etc. Este indicador también tiene unas limitaciones, por ejemplo, se asumió una duración de parpadeo fija para todos los sujetos, aunque el parpadeo de todos tenga una duración diferente. Los resultados podrían ser mejorados por un enfoque adaptativo Otra limitación está en la estimación de apertura de ojos. Mientras que EAR se estima a partir de una imagen en 2D, es bastante insensible a la orientación de la cabeza, pero puede perder la capacidad de discriminación para las rotaciones fuera del plano. Una solución podría ser definir el EAR en 3D. Hay

detectores de puntos de referencia que estiman una pose en 3D (posición y orientación) de un modelo tridimensional de puntos de referencia.

CAPÍTULO III: MARCO TEÓRICO

En este capítulo se explica a mayor detalle sobre los temas relacionados a la presente investigación tales como: la detección de rostros, somnolencia, accidentes de tráfico, definición de las metodologías utilizadas: Viola-Jones y *Regression based facial landmark detection* y finalmente algunas consideraciones que tuvimos para la realización de nuestra validación que servirán para sustentar nuestra propuesta de solución al problema planteado.

3.1 Detección de rostros

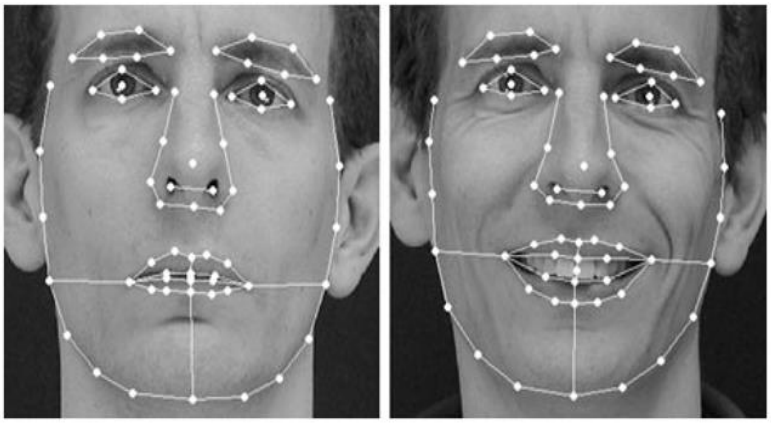
Las primeras investigaciones sobre la detección de rostros se realizaron en la década de los 50's en el área de psicología (Bruner y Tagiuri, 1954), y en la década de los 60's en la literatura ingenieril (Bledsoe, 1964). En la actualidad, este concepto ha sido un área de investigación activa en el campo de la visión por computador debido al innumerable número de aplicaciones que requieren detección de rostros (Triantafyllidou y Tefas, 2016).

La detección del rostro y sus características físicas han levantado curiosidad en los investigadores desde hace un tiempo atrás, debido a que este brinda una gran cantidad de información visual que puede dar sobre una persona tales como: su género, edad, estado de ánimo, la actitud en la que se encuentra. Gracias a los avances en la tecnología, este tema está involucrado en diferentes campos de aplicación, por ejemplo: los sistemas de video de vigilancia, sistemas de teleconferencias, realidad virtual, juegos de video, sistemas de identificación de rostros, biometría, entre otras (García y Crespin, 2014).

La detección del rostro se define a la acción de reconocer a una persona mediante su rostro, teniendo en consideración todas las características visibles y medibles. El rostro, no es un objeto que se pueda definir y detectar fácilmente, ya que este es único por persona; es por esto que, se hace uso de diferentes metodologías, algoritmos que deben ser lo suficientemente robusto para poder detectar caras de diferentes edades, géneros, tamaños, formas, distintos cabellos, lentes, bigotes; sin dejar de tener en cuenta los distintos tipos de iluminación y fondo (Riveros, 2008).

Figura 3.1

Detección de las características de un rostro.



Fuente: Reconocimiento Facial Mediante Imágenes Estereoscópicas Para Control de Ingreso, Pereyra Pamela. (2013).

En la **Figura 3.1** se puede observar que la detección del rostro brinda una gran cantidad de información visual que se puede proporcionar sobre una persona como: género, edad, estado de ánimo, actitud. En esta imagen se observa que es un hombre, aproximadamente de 20-30 años, en la imagen de la izquierda se encuentra serio y en la imagen de la derecha, contento. Los puntos dibujados en su rostro son puntos de referencia faciales que se utilizan para localizar y representar regiones de la cara como: ojos, cejas, nariz, boca, mandíbula (Rosebrock, 2017).

Existen varios métodos que se han propuesto para la detección facial; por ello, es complicado realizar una clasificación de estos. Se puede identificar tres clases, como se muestra en la **Tabla 3.1**: los holísticos que hacen uso de la región facial y han demostrados ser efectivos en experimentaciones con grandes volúmenes de datos como las bases de datos: "*eigenface*" o "*fisher facer*"; los basados en características que busca obtener y hacer uso de características locales como son los ojos, la nariz, la boca y utiliza un clasificador estructural; y los híbridos que utiliza un sistema de percepción similar al de un humano, ya que combina métodos holísticos con métodos basados en características faciales (Pereyra, 2012).

Tabla 3.1

Métodos de detección de rostros: Holísticos, Basados en características e Híbridos.

METODOS	TRABAJOS
Holísticos	*Principal Component Analisis (PCA)*
	Eigenface
	Probabilistic eigenface
	Fisherfaces/subspace LDA
	SVM
	Evolution pursuit
	Features Lines
	ICA
	Otros
	LDA/FLD
	PDBNN
Basados en características	*Pure geometric methods*
	Dynamic link architecture
	Hidden Markov model
	Convolutional Neural Network
	Modular eigenface
Híbridos	*Hybrid LFA*
	Shape-normalized
	Component-bases

Fuente: Reconocimiento Facial Mediante Imágenes Estereoscópicas Para Control de Ingreso, Pereyra P. (2013)

3.2 Somnolencia

Según la Real Academia Española (RAE), la somnolencia es "la sensación de pesadez y torpeza de los sentidos motivadas por el sueño" (Real Academia Española, 2017). Este produce la disminución de la capacidad de vigila afectando negativamente a los sentidos, disminuyendo la capacidad de percepción, reconocimiento y control del auto; aumentado la probabilidad de sufrir algún accidente o un sobresalto en la carretera (Ji y Yang, 2002).

La somnolencia es causada por múltiples factores, como, por ejemplo: el aburrimiento, la conducción por un tiempo considerable, el consumo de alcohol y/o drogas, los desórdenes de sueño, estrés laboral, consumo de medicamentos que producen sueños (García y Crespín, 2014).

Está comprobado que la somnolencia provoca en el conductor: aumento en el tiempo de reacción, menor concentración y mayor distracción, errores en la toma de decisiones, alteraciones motoras: pequeños temblores, movimientos involuntarios y

relajación muscular, se dan los llamados *microsleep*, estos son momentos de escasos segundos en los que el cerebro permanece ajeno a todo lo que suceda a su alrededor. Ese pequeño tiempo es más que suficiente para perder el control del vehículo y sufrir un accidente automovilístico. Las alteraciones sensoriales, de percepción y los cambios contantes de comportamiento, da lugar a mucha tensión en el individuo, lo que favorece la agresividad y los cambios de humor (Marks, 2013).

Figura 3.2

Detección de Somnolencia de un usuario al volante: Variables del comportamiento ocular (frecuencia de parpadeo, duración del cierre del parpadeo, porcentaje de ojos cerrados).

Fuente: Drowsiness detector wakes sleepy miners if they‘re dozing off, Marks (2013).

Una de las tantas formas en las que se manifiesta la somnolencia es a través del comportamiento ocular, como se muestra en la Figura 3.2, el cual ha sido cuantificado con diferentes variables: frecuencia de parpadeos, duración del cierre de párpados y el porcentaje de tiempo en los que los ojos permanecen cerrados (Hamada, Ito, Adachi, Nakano y Yamamoto, 2003).

La detección de la somnolencia, no es un tema dirigido solamente a los conductores vehiculares, sino que también han sido utilizado en diferentes ámbitos; por ejemplo, muchos estudios han analizado la atención del conductor mediante la aplicación de tecnologías de detección de imágenes. Si este mecanismo se pudiera aplicar en las aulas de clase, ayudaría a los maestros a mantener a los estudiantes atentos y así reducir la carga del maestro durante la clase. Este estudio aplica principalmente el análisis de la lógica difusa de las imágenes faciales de los estudiantes

al participar en clase. Los trabajos de simulación se realizan para evaluar el efecto del sistema propuesto bajo diversas condiciones. Los resultados de la simulación indicaron que el sistema propuesto es efectivo para detectar la atención del alumno en clase. El grado de atención del estudiante afecta los resultados del aprendizaje. En las clases con muchos alumnos, los maestros generalmente tienen dificultad para determinar si todos los alumnos están atentos. La detección de imágenes se puede aplicar a las imágenes faciales del estudiante y determinar sus comportamientos basados en imágenes faciales, como los ojos cerrados, el movimiento de la cabeza y la somnolencia. Para la prueba del sistema de investigación, la simulación había sido procesada. La imagen facial había sido captada prácticamente, y aplicado el programa del sistema al análisis. Los comportamientos de simulación se han centrado en el objeto de investigación que incluyen no salir en la fotografía, girar la cabeza, somnolencia y sin movimiento. Estos comportamientos habían sido simulados por el alumno, e introducido estas imágenes en el sistema para ser análisis (Hwang y Yang, 2009). El sistema de detección de somnolencia se puede extender a diversos campos, tal es el caso del artículo en cuestión, se planeaba determinar qué alumnos no prestaban la debida atención en clase para tomar alguna acción correctiva.

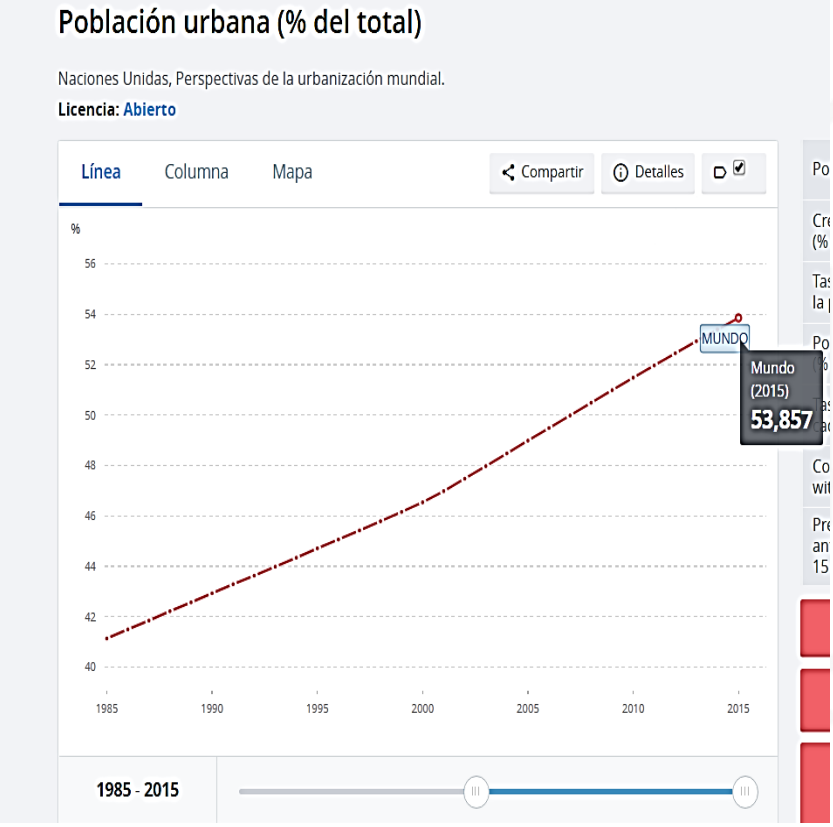
Otro ámbito en el que se ha utilizado la detección de la somnolencia es en la industria de la minería. Un conductor somnoliento al volante de un camión de minería de más de 400 toneladas de carga es un verdadero peligro, hay una alta probabilidad que suceda algún accidente. Por ello, la empresa estadounidense Caterpillar ha incorporado un detector de somnolencia en todos sus camiones mineros. El Sistema de administración de riesgos por fatiga (*Fatigue Risk Management System*, FRMS), definido por Australia, está diseñado para prevenir los accidentes causados por cansancio, haciendo uso de una cámara infrarroja que puede ver a través de los lentes de sol y una computadora de procesamiento de imágenes, evalúa la frecuencia, duración y velocidad del parpadeo del conductor para detectar la probabilidad de desatención y *microsleep*. Si el FMS detecta que el conductor está distraído, se dispara una alarma advirtiendo a los conductores de camiones de minería de forma inalámbrica. Según Marks, en sus primeras pruebas, el FMS disminuyó los eventos de fatiga en un 72% (Marks, 2013).

3.3 Accidentes de tráfico

Según el Banco Mundial, al año 2015, la población mundial cuenta con un 53.875% de población urbana con un crecimiento de 2.058% anual (Grupo Banco Mundial, 2017), la necesidad de movilizarse, en especial en las grandes ciudades, ha generado que se incremente la demanda de vehículos, la contaminación y la congestión vehicular. Estos factores, entre otros, generan estrés y cansancio en la población y particularmente en los conductores; por consiguiente, estos influyen directamente en el incremento de los accidentes de tráfico como se muestra en la Figura 3.3

Figura 3.3

Curva de porcentaje de Población urbana de 1985-2015.Población Urbana (% del Total).



Fuente: Grupo Banco Mundial (2017).

Según la Organización Mundial de la Salud (OMS), Los traumatismos causaron 5 millones de muertes en 2015, más de una cuarta parte (el 27%) como consecuencia de accidentes de tránsito. La mortalidad por estos accidentes, que a nivel mundial es de 18,3 defunciones por cada 100,000 habitantes, es más acusada en los países de ingresos bajos, donde alcanza las 28,5 defunciones por cada 100,000 habitantes. Los accidentes de tránsito se encuentran entre las 10 causas principales de muerte en los países de

ingresos bajos, como se muestra en la Figura 3.4, tanto en la franja inferior como en la franja superior (Organización Mundial de la Salud, 2017).

Figura 3.4

Las 10 principales causas de defunción en países de ingresos bajos según la OMS.

Fuente: Las 10 principales causas de defunción en países de ingresos bajos según la OMS (2017). Recuperado de http://www.who.int/mediacentre/factsheets/fs310/es/index1.html).

En el año 2014, solo en Lima ocurrieron más de sesenta mil accidentes de tráfico, como se muestra en la siguiente Tabla 3.2.

Tabla 3.2

Número total de accidentes de tránsito por departamento en el año 2014.

Total de Accidentes de tránsito Año 2014=123,786		
N	Departamento	Cantidad de Accidentes de transito
1	Lima	60,060.00
2	Arequipa	8,299.00
3	La Libertad	6,325.00
4	Cusco	3,237.00
5	Piura	4,912.00
6	Junín	4,222.00
7	Prov. Const. Del Callao	4,069.00
8	Ancash	3,957.00

Fuente: Consejo Nacional de Seguridad Ciudadana. (2014)

3.4 Consideraciones

3.4.1 Base de datos NTHU *Drowsy Driver Detection* (NTHU-DDD):

La Base de datos NTHU-DDD, brindada por la universidad *National Tsing Hua*, es un conglomerado de vídeos de personas, tanto hombres como mujeres de diferentes etnias, que simulan estar en un auto; se desarrollaron 5 tipos de escenarios de los cuales usamos 4 de ellos para la realización de nuestras pruebas. Los escenarios contienen

1. Rostro sin gafas.
2. Rostro con gafas.
3. Rostro con gafas de sol.
4. Rostro sin gafas en la noche.
5. Rostro con gafas de noche.

Cada *frame* de los videos estarán etiquetados con el estado de somnoliento o no somnoliento.

Los videos están en 640x480 píxeles, se encuentra en formato AVI sin audio, los que cuentan con el escenario de luminosidad baja presentan 15 fotogramas/segundo y los demás escenarios, luminosidad alta, presentan 30 fotogramas/segundo.

3.4.2 Participantes

Para el presente trabajo, los participantes están conformados por los videos proporcionados por la Base de datos NTHU-DDD. Además, se reclutará a jóvenes entre 20-24 años de edad con rostros latinos, en específico peruanos, para que simulen el estado de somnolencia al volante y así poder obtener los videos para la Base de datos propia.

3.4.3 Población

- Los *frames* obtenidos de los videos de la base de datos *NTHU Drowsy Driver Detection (*NTHU-DDD).
- Los *frames* obtenidos de los videos de la Base de datos propia.

3.4.4 Muestra

- Conformada por las secuencias de video de personas simulando conducir un auto, para luego ser procesados por el sistema para detectar si el participante muestra signos de somnolencia.

3.4.4.1 Tamaño de la muestra:

- Se hará uso de la siguiente fórmula para poder calcular el tamaño de nuestra muestra (García y Crespin, 2014).

$$n = \frac{Z^2 * P * Q}{i^2} \quad \ldots\ldots (1)$$

Donde:

- n= tamaño muestral.
- z=valor correspondiente a la distribución de Gauss, en este caso el valor a tomarse es 1.96 ya que consideramos un nivel de confianza del 95%.
- p= prevalencia esperada del parámetro a evaluar, como se desconoce, el valor a tomarse es 0.5.
- q= 1-p.
- i=precisión, en este caso será de un 12%.

Resolviendo esta ecuación, se obtiene que la muestra estará conformada por 66 secuencias de videos aproximadamente.

3.5 Definición de metodologías

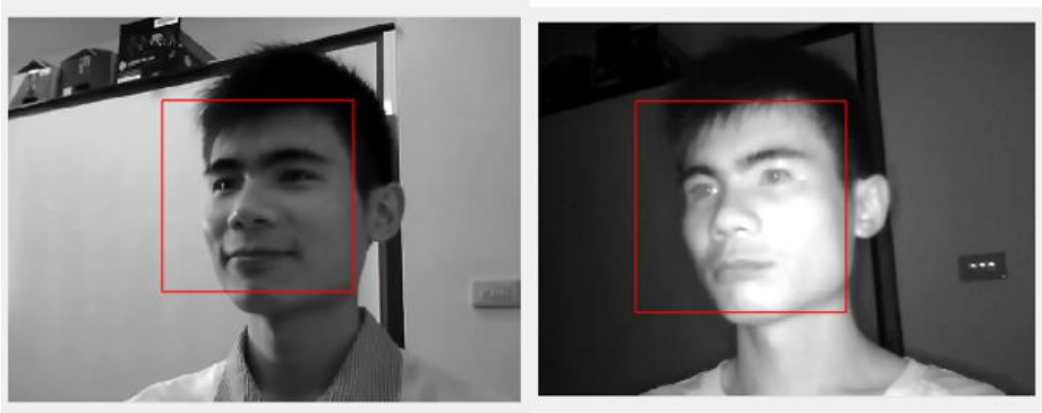
3.5.1 Viola-Jones Object Detection Framework

Es un *framework* desarrollado por: Paul Viola y Michael Jones (Viola y Jones, 2004) utiliza una serie de algoritmos e ideas para una robusta y rápida detección visual. Este marco está compuesto de tres principales contribuciones: un entrenador de clasificadores, basado en *AdaBoost*, un algoritmo de aprendizaje, para la detección de objetos utiliza la clasificación en "cascada" y la "imagen integral" que es una representación de la imagen. Inicialmente, en la detección del objeto se hace uso de la imagen integral, para la evaluación de las características, para acelerar el proceso; esta imagen se obtiene a través de unas operaciones por pixel y que al finalizar permiten la

búsqueda de características en subregiones de 24 x 24 pixeles, todo lo detallado se transforma en una tarea de tiempo constante, sin importar su posición o escala en la subregión. Una vez obtenida la imagen, el algoritmo la parte en subregiones de diferentes tamaños y utiliza los clasificadores, cada uno con una serie de características visuales, finalmente para discriminar si en la imagen se encuentra el objeto o no.

Figura 3.5

Metodología Viola-Jones aplicado a un video del NTHU *dataset* en escenario de luminosidad alta y otro escenario con luminosidad baja.



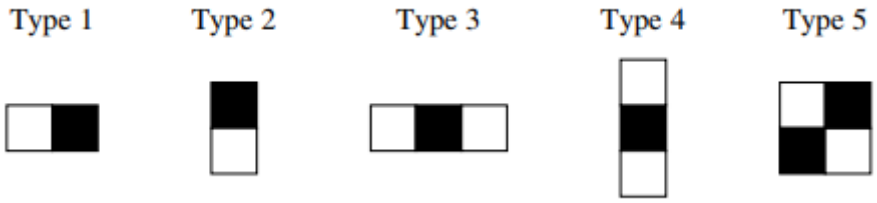
Fuente: Elaboración propia

3.5.1.1 *Haar Features*

Haar Features son usados para detectar la presencia de alguna característica en una imagen, existen varios tipos para detectar ciertas características, que se usan generalmente en Viola-Jones que se aplican a la imagen. A continuación, a la región negra se le aplica el valor "+1" y a la región blanca "-1", cuando se le aplique esta técnica a la sub-ventana en la que se está trabajando se suman todos los valores que se encuentran en la región del mismo color y finalmente se restan los valores de los pixeles que están en diferentes regiones para obtener un valor único siendo este el resultado Figura 3.6.

Figura 3.6

Se aplican los cinco tipos de *Haar Features* a la sub-imagen obtenida en la metodología Viola-Jones para determinar que parte de la sub-imagen puede ser una característica relevante ya sea los ojos, boca o nariz. Por ejemplo, el tipo 3 es usado para la detección de la nariz.



Fuente: Viola y Jones (2004)

Ocurren diferentes variaciones de tamaños y posiciones al aplicar los diferentes tipos de *Haar Features* a las imágenes, recorren la misma aumentando el tamaño de pixeles y moviéndose de posición, se puede calcular que se pueden dar más de 160 000 *features* en una sub-ventana de 24 x 24 pixeles, ver Figura 3.7.

Figura 3.7

A partir de los cinco tipos *Haar Features* se crean variaciones que luego son aplicadas de a las sub-imágenes para una mejor aplicabilidad.

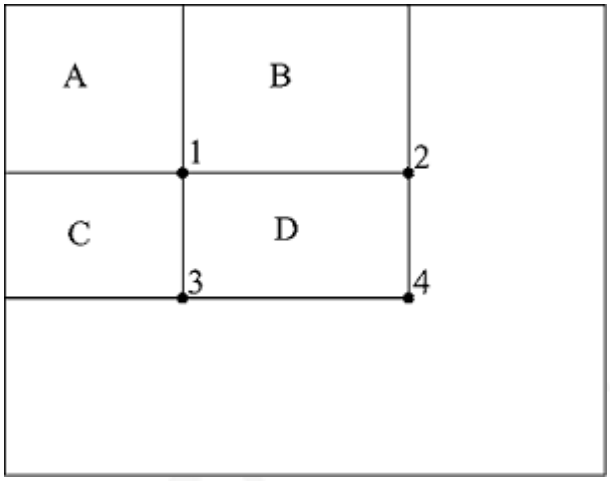
Fuente: Face Reco. Recuperado de http://siret.ms.mff.cuni.cz/facereco/method.

3.5.1.2 *Integral Image*

Imagen integral es un método para la suma de pixeles dentro de una región en concreto para mejorar la *performance*. Utilizando los *corner values* se suma los valores de las esquinas para determinar la cantidad de pixeles, ver Figura 3.8.

Figura 3.8

Representación gráfica de la Imagen integral.



Fuente: Face Reco. Recuperado de http://siret.ms.mff.cuni.cz/facereco/method.

En este caso, para hallar la representación gráfica de la imagen integral, es la suma de los pixeles de la región "A" es igual al punto "1", la región "B" es igual al punto "2" y así sucesivamente.

Suma de todos los pixeles en:

$D = 1 + 4 - (2 + 3)$..... (2)

$D = A + (A + B + C + D) - (A + C + A + B)$..... (3)

El punto "1" es la suma de todos los pixeles de la región "A" el punto "2" es la suma de los pixeles de la región "A" y "B" y así sucesivamente. Después de sumar los valores de la misma diagonal y se resta con la diagonal opuesta, finalmente da como resultado la suma de los pixeles de esa área.

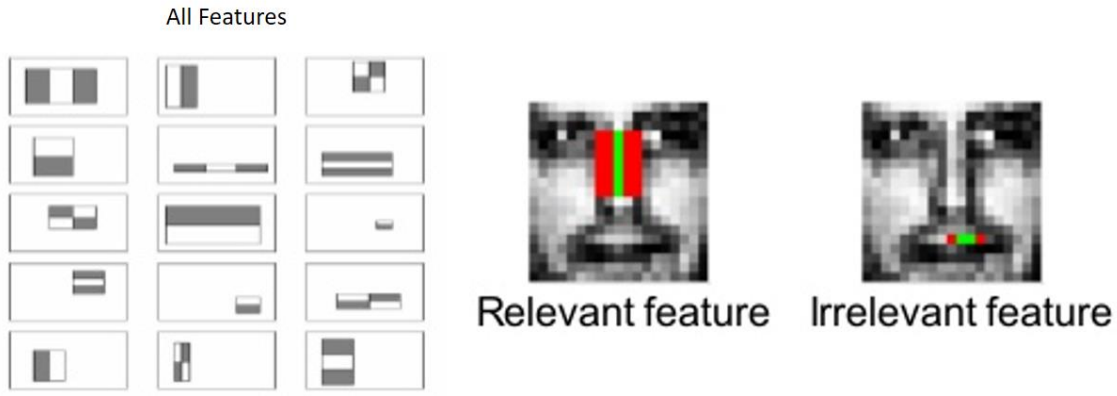
3.5.1.3 *Adaboost*

Existe un problema de *performance* cuando se hacen todas las combinaciones posibles al aplicar *Haar Features*, es decir, tener que calcular más de 160 000 *features* por cada 24 x 24 pixeles. *Adaboost* elimina los *features* redundantes o los *features* que son inútiles para él cálculo y conserva los útiles, es decir los relevantes.

La relevancia es determinada por *Adaboost*, en este caso las características encontradas en la parte de la nariz son relevantes para la detección de la misma, en cambio las características encontradas en los labios no lo son, como se muestra en la Figura 3.9

Figura 3.9

El algoritmo Adaboost determina que características son relevantes para la detección de una facción de la cara en particular y cual es irrelevante.



Fuente: Face Reco. Recuperado de http://siret.ms.mff.cuni.cz/facereco/method.

Adaboost identifica cierto número de características de las 160 000 y después se le da un peso a cada una de las características y hace una combinación linear de todas las características para decidir cuál es un rostro y cual no es.

$$F(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x) + \alpha_3 f_3(x) + \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (4)$$

Donde:

- *F(x)* = Strong classifier.
- *X*= Weak classifier.

Weak classifier es una característica relevante, que por lo menos tiene una mejor performance que una aleatoria, es decir si se le pasa como *input* más de cien imágenes de rostros, será capaz de reconocer más de cincuenta caras. El resultado que se obtiene del *weak classifier* es "1" cuando tuvo una buena *performance* e identificó la característica o "0" porque no tuvo éxito al identificar la característica.

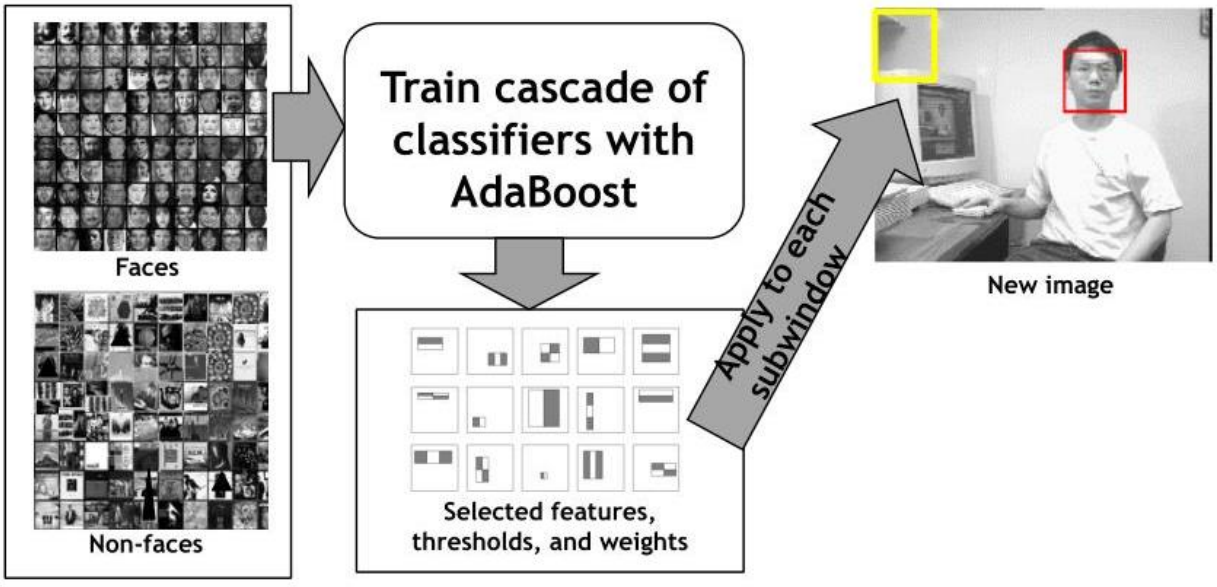
Finalmente, para cada sub ventana de 24 x 24 pixeles, se tiene que evaluar 2500 *features* que fueron obtenidos de *Adaboost* y ver si supera un cierto umbral para decidir si se trata de una cara o no.

3.5.1.4 *Cascading*

El algoritmo se concentra en descartar las partes de la imagen que no son rostros y centrarse en las partes que si podría ser un rostro. Para no calcular las 2500 características en cada sub-ventana de 24 x 24 pixeles, se usan las cascadas, estas clasifican las 10 primeras características en un escenario después las siguientes 10 en el siguiente escenario y así hasta llegar a las 2500, ver Figura 3.10. La ventaja es que se podrá predecir si la sub-ventana mostrada es un rostro o no, rápidamente. Entonces se podrá rechazar una imagen (una sub-ventana de la imagen) en muy poco tiempo. En vez de evaluar todas las 2500 características, se dividen en varios escenarios para establecer una jerarquía en los clasificadores para ver si podría ser un rostro o no, si una imagen entra como input y pasa el primer escenario es posible que sea una imagen, pero sino pasa él escenario es porque definitivamente no es una imagen.

Figura 3.10

A partir de una Base de Datos de imágenes de caras e imágenes de no-caras se entrena al clasificador para que pueda determinar, en una imagen de entrada que parte de esta podría ser una cara.



Fuente: Recuperado de https://www.youtube.com/watch?v=_QZLbR67fUU.

3.5.2 *Support Vector Machine* (SVM)

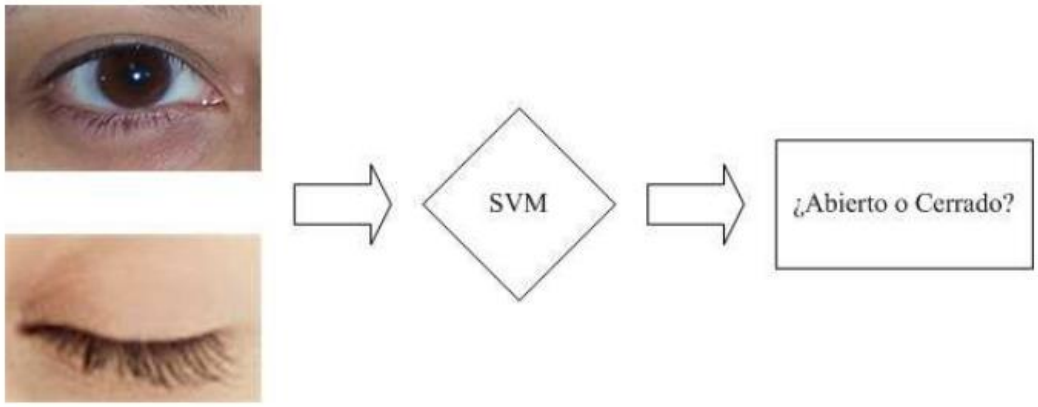
El *Support Vector Machine* (SVM) es un conjunto de algoritmos de aprendizaje supervisado desarrollados por Vladimir Vapnik y su equipo en los laboratorios AT&T

(Flores Calero, 2009). Es un modelo lineal para problemas de clasificación y regresión. Puede resolver problemas lineales y no lineales y funciona para muchos problemas prácticos. La idea de SVM es: el algoritmo crea una línea o un "hiperplano" que separa los datos en clases (Pupale, 2018).

Se tiene un conjunto de datos y se necesita clasificarlos en dos grupos, entonces, la tarea del SVM es encontrar una línea ideal que separe los conjuntos de datos en dos clases. De acuerdo con el algoritmo, primero encontramos los puntos más cercanos a la línea de ambas clases, estos puntos se denominan vectores de soporte. Ahora, se calcula la distancia entre la línea y los vectores de soporte. Esta distancia se llama el margen, el objetivo principal es maximizarlo. Finalmente, el hiperplano para el cual el margen es máximo, es el hiperplano óptimo. Por lo tanto, SVM intenta tomar un límite de decisión de tal manera que la separación entre las dos clases sea lo más amplia posible (Pupale, 2018).

Figura 3.11

El SVM hace una clasificación y determinación del estado de los ojos del usuario, es decir si se encuentran abiertos o cerrados.



Fuente: Flores Calero (2009).

De acuerdo a la Figura 3.11 se percibe que SVM procesa la información para entregar un resultado sin conocer cuál es el procedimiento empleado, por tal motivo, es importante resaltar las características que mejor identifican a los objetos que se desea estudiar (Flores Calero, 2009).

3.5.2.1 Histograma de Gradientes Orientados *(HOG)*

El *Histogram of oriented gradients* (HOG) es un descriptor característico utilizado en el procesamiento de imágenes y la visión por ordenador para poder detectar los objetos. Esta técnica cuenta con ocurrencias de orientación de gradiente en porciones localizadas de una imagen, es similar a la de los histogramas de orientación de borde, pero se diferencia en que se calcula en una densa retícula de células uniformemente espaciados y utiliza la superposición normalización para el contraste local para mejorar la precisión (Harb, Desuky, Mohammed, y Jennane, 2017).

La idea esencial es que la apariencia y la forma del objeto local dentro de una imagen se pueden describir mediante la distribución de gradientes de intensidad o direcciones de borde. La imagen se divide en pequeñas regiones conectadas llamadas células, y para los píxeles dentro de cada célula, se compila un histograma de direcciones de gradiente. El descriptor es la concatenación de estos histogramas. Para una mejor precisión, los histogramas locales pueden normalizarse por contraste calculando una medida de la intensidad a través de una región más grande de la imagen, denominada bloque, y luego utilizando este valor para normalizar todas las células dentro del bloque. Esta normalización resulta en una mejor invariancia a los cambios en la iluminación y la sombra (Harb, Desuky, Mohammed, y Jennane, 2017).

El descriptor de HOG tiene algunas ventajas sobre otros descriptores, ya que opera en las células locales, es invariante a transformaciones geométricas y fotométricas, a excepción de la orientación a objetos. Por otra parte, el muestreo espacial grueso, el muestreo de orientación fina y la normalización local fotométrica permiten que el movimiento corporal individual de los peatones sea ignorado siempre y cuando mantengan una posición aproximadamente erguida. Por tanto, el descriptor HOG es particularmente adecuado para la detección humana en imágenes.

Los pasos para el cálculo de la función descriptor HOG son:

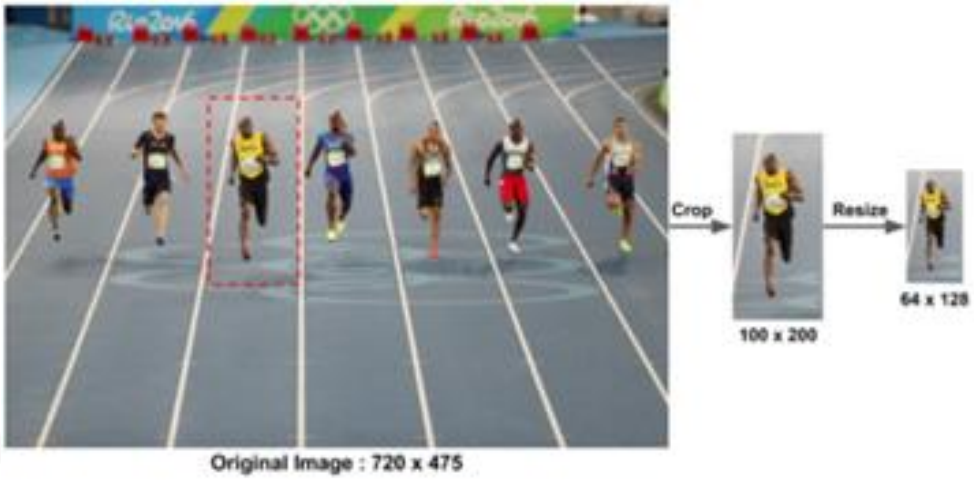
1. Pre-procesamiento:

- El descriptor HOG es utilizado para la detección de objetos, se calcula en un parche de 64 x 128 de una imagen. Por supuesto, una imagen puede ser de cualquier tamaño. Normalmente, los parches en varias escalas se analizan en muchas ubicaciones de imágenes. La única restricción es que los parches que se analizan tienen una relación de aspecto fija, deben tener una relación

de aspecto de 1: 2, ver Figura 3.12. Por ejemplo, pueden ser 100 x 200, 128 x 256 o 1000 x 2000, pero no 101 x 205 (Mallick, 2016).

Figura 3.12

Etapa pre procesamiento: Relación de 1:2 de los pixeles para poder crear un parche de una imagen.



Fuente: Mallick (2016).

2. Calcular las imágenes de degradado

- Para el cálculo de un descriptor de HOG, tenemos que calcular primero los gradientes horizontales y verticales, como se muestra en la Figura 3.13. Esto se logra fácilmente mediante la filtración de la imagen con los siguientes Kernels: -1,0,1 (Horizontal o Vertical, dependiendo en qué posición se quiere calcular el gradiente) (Mallick, 2016).

Figura 3.13

Calculo del descriptor HOG Núcleos de una filtración de imagen.

Fuente: Mallick (2016).

- A continuación, podemos encontrar la magnitud y dirección del gradiente usando la siguiente fórmula (4) y (5):

$$g = \sqrt{g_x^2 + g_y^2} \quad \text{.....(5)}$$

$$\theta = \arctan \frac{g_y}{g_x} \quad \text{.....(6)}$$

Donde:

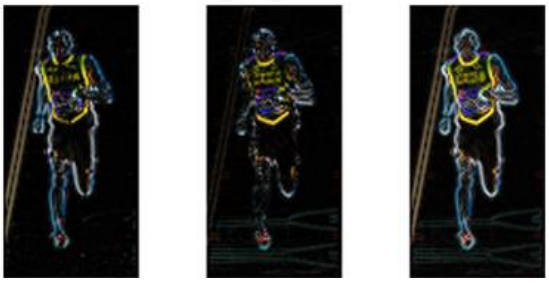
- g = dirección del gradiente
- g_x = gradiente en dirección vertical
- g_y = gradiente en dirección horizontal
- θ = radio del gradiente

- Se observa que el gradiente "x" se dirige en líneas verticales y el gradiente "y" se dirige en líneas horizontales. La imagen de degradado eliminó información no esencial (por ejemplo, fondo de color constante), pero resaltó los contornos. En otras palabras, se puede observar en la Figura 3.14 de degradado que hay una persona en la imagen (Mallick, 2016).
- En cada píxel, el gradiente tiene una magnitud y una dirección. Para las imágenes en color, se evalúan los gradientes de los tres canales. La magnitud del gradiente en un píxel es el máximo de la magnitud de los gradientes de los tres canales, y el ángulo es el correspondiente al gradiente máximo (Mallick, 2016).

Figura 3.14

Izquierda: valor absoluto del gradiente x. Centro: valor absoluto del gradiente y.

Derecha: magnitud del gradiente.



Fuente: Mallick (2016).

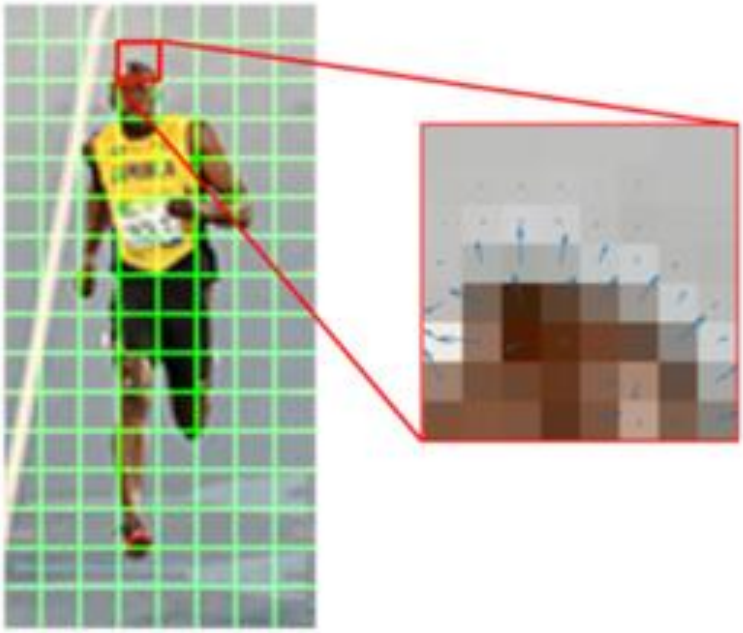
3. Cálculo del histograma de gradientes en 8 x 8 cédulas.

- En este paso, la imagen se divide en 8 x 8 células y un histograma de gradientes se calcula para cada 8 x 8 células. Una de las razones importantes para utilizar un descriptor de función para describir un parche de una

imagen es que proporciona una representación compacta. Un parche de imagen de 8 x 8 contiene 8 x 8 x 3 = 192 valores de píxel. El gradiente de este parche contiene 2 valores (magnitud y dirección) por píxel que se suma a 8 x 8 x 2 = 128 números. Al final estos 128 números se representan utilizando un histograma 9-bin que se puede almacenar como una matriz de 9 números. El histograma es esencialmente un vector (o una matriz) de 9 cubos (números) que corresponden a ángulos de 0, 20, 40, 60...160 (Mallick, 2016).

Figura 3.15

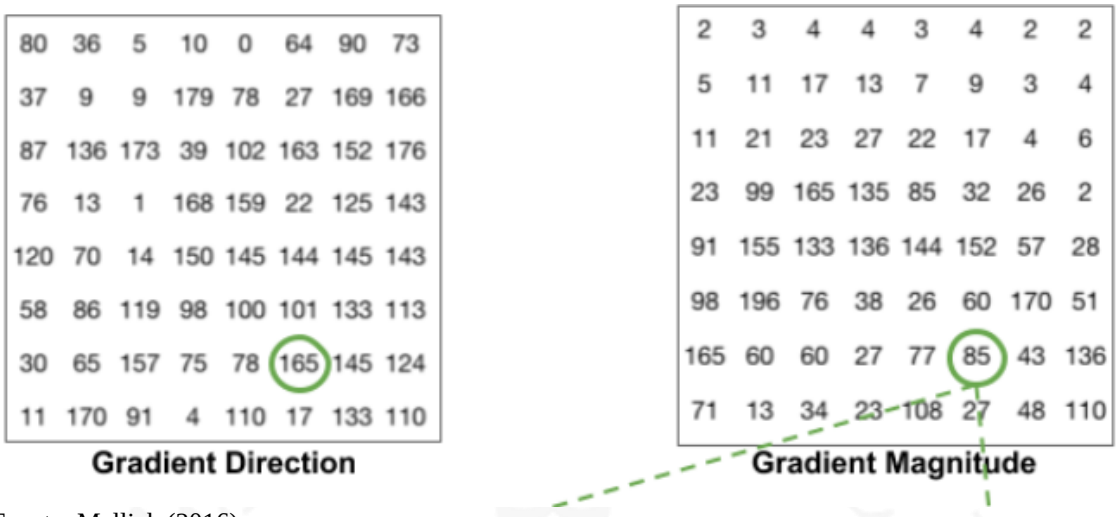
Cálculo del histograma de gradientes de 8 x 8 cedulas, El parche RGB y gradientes se representan mediante flechas.



Fuente: Mallick (2016).

Figura 3.16

Representación de gradientes en cédulas de 8 x 8.



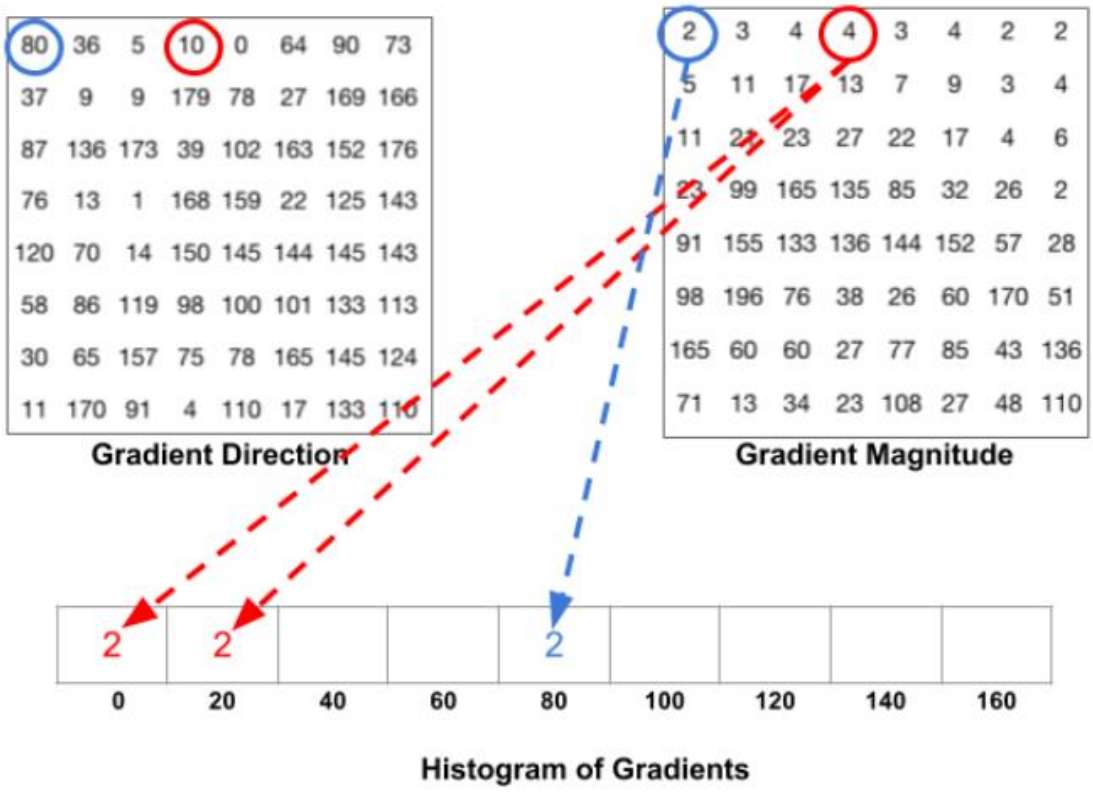
Fuente: Mallick (2016).

- La Figura 3.15 muestra el parche de la imagen superpuesta con flechas que muestran el gradiente, estas flechas muestra la dirección del gradiente y su longitud muestra la magnitud. Se observa cómo el sentido de las flechas indica la dirección del cambio en la intensidad y la magnitud muestra cuán grande es la diferencia.
- En la Figura 3.16 se ve los números en bruto que representan los gradientes en las células de 8 x 8 (ángulos entre 0 y 180 grados). Estos son denominados gradientes "sin signo" porque un gradiente y su negativo están representados por los mismos números. En otras palabras, una flecha de gradiente y la de 180 grados opuesta a ella se consideran iguales. (Mallick, 2016).
- El siguiente paso es crear un histograma de gradientes en este 8 x 8 células. El histograma contiene 9 contenedores correspondientes a ángulos de 0, 20, 40...160. En la Figura 3.17 se ilustra el proceso. Buscamos en magnitud y dirección de la pendiente de la misma parcela 8 x 8. Un *bin* se selecciona basándose en la dirección, y se selecciona el voto (el valor que va en el *bin*) basado en la magnitud (Mallick, 2016). Primero vamos a centrarse en el píxel rodeado en azul. Tiene un ángulo (dirección) de 80 grados y la magnitud de 2. Por lo tanto, añade 2 a la quinta *bin*. El gradiente en el píxel rodeado usando rojo tiene un ángulo de 10 grados y la magnitud de 4.

Puesto que 10 grados es la mitad de camino entre 0 y 20, el voto por el pixel se divide uniformemente en los dos compartimientos.

Figura 3.17

Creación del Histograma en cédulas de 8x8.

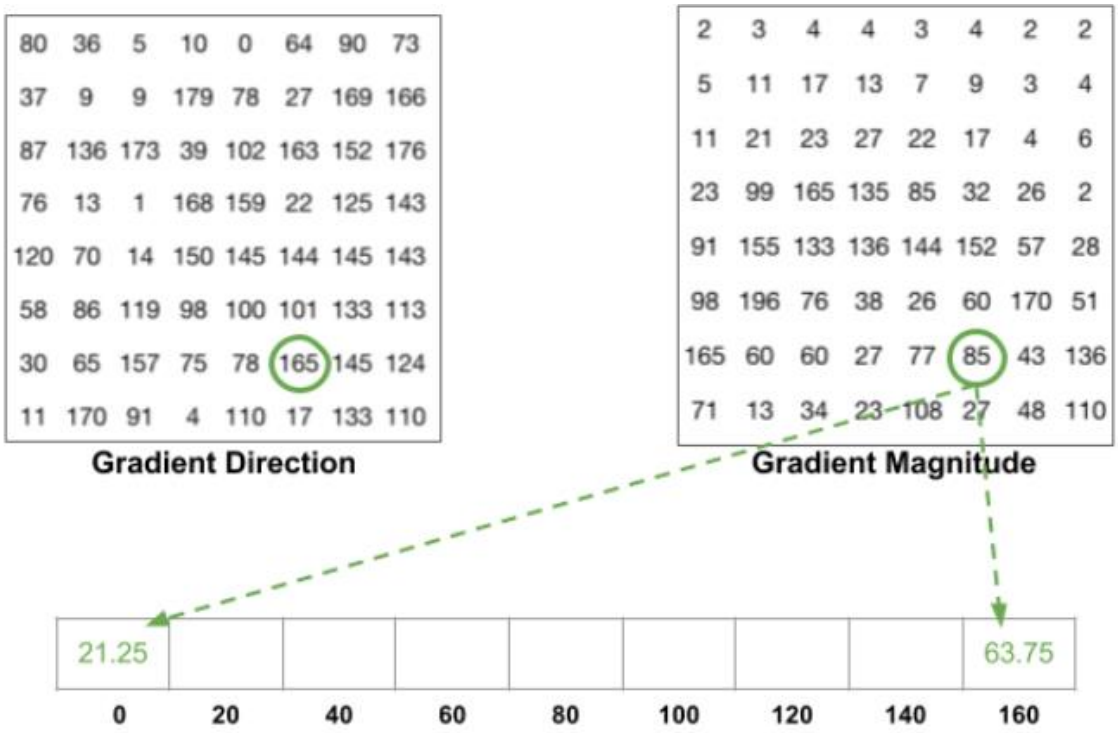


Fuente: Mallick (2016).

- Hay un detalle más a tener en cuenta, si el ángulo es mayor a 160 grados y menor a 180, y sabemos que el ángulo se envuelve formando 0 y 180 equivalentes. Por lo tanto, en la Figura 3.18 se muestra un ejemplo, el píxel con ángulo de 165 grados contribuye proporcionalmente al contenedor de 0 grados y al contenedor de 160 grados.

Figura 3.18

Creación del Histograma en cédulas de 8 x 8.

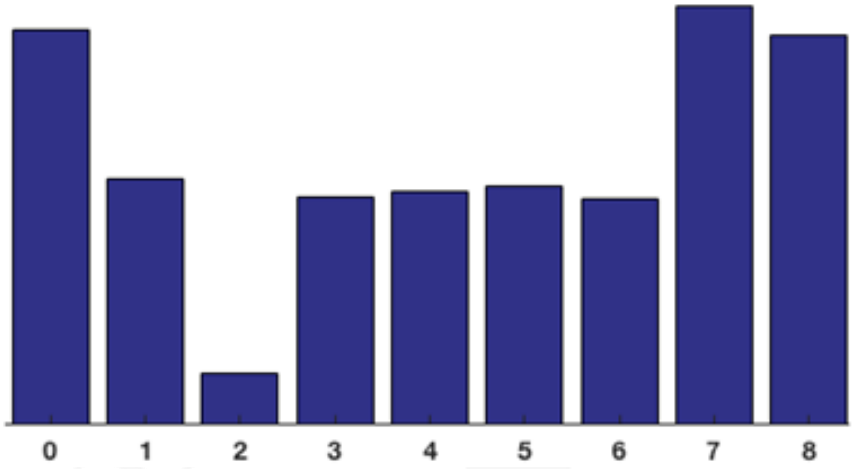


Fuente: Mallick (2016).

- Las contribuciones de todos los píxeles de las células 8 x 8 se suman para crear el histograma 9-*bin* como se muestra en la Figura 3.19 .Para el parche anterior, parece que este es nuestra representación, el eje y es de 0 grados. Se puede ver el histograma tiene mucho peso cerca de 0 y 180 grados, lo que es más que otra forma de decir que en el parche gradientes están apuntando hacia arriba o hacia abajo.

Figura 3.19

Creación del histograma 9-bin. El eje y indica las contribuciones de todos los pixeles de las células 8 x 8.



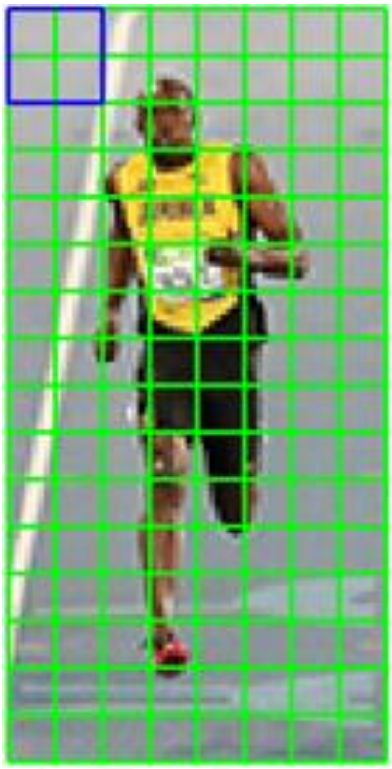
Fuente: Mallick (2016).

1. Bloque de normalización 16 x 16:

- Los gradientes de una imagen son sensibles a la iluminación general. Si la imagen se hace más oscura dividiendo todos los valores de los píxeles por 2, la magnitud del gradiente cambiará a la mitad, y por lo tanto los valores del histograma cambiarán a la mitad. Lo ideal es que el descriptor debe ser independiente de las variaciones de iluminación. En otras palabras, se procedería a normalizar el histograma para que no se ven afectados por las variaciones de iluminación.
- Un bloque de 16 x 16 tiene 4 histogramas que se pueden concatenar para formar un vector 36 x 1 elemento. La ventana se mueve entonces por 8 y un normalizado 36 x 1 vector, se calcula sobre esta ventana y el proceso se repite, ver Figura 3.20.

Figura 3.20

Bloque de normalización 16 x 16.



Fuente: Mallick (2016).

2. Cálculo del vector de característica HOG.

- Para calcular el vector de características final para todo el parche de imagen, los 36 x 1 vectores se concatenan en un vector gigante (Mallick, 2016).
    - Hay 7 horizontal y 15 posiciones verticales que hacen un total de 7 x 15 = 105 posiciones.
    - Cada bloque de 16 x 16 está representada por un vector 36 x 1.
    - Así que cuando concatenamos un vector obtenemos un 36 x 105 = 3780 vector de dimensiones.

3. Visualización del histograma de gradientes orientados.

- El descriptor de HOG de un parche de imagen generalmente se visualizó mediante el trazado de los 9 x 1 histogramas normalizados en las células 8 x 8. Se notará que la dirección dominante del histograma captura la forma de la persona, especialmente alrededor del torso y las piernas como se muestra en la Figura 3.21 (Mallick, 2016).

Figura 3.21

Visualización del histograma de gradientes orientados.

Fuente: Mallick (2016).

3.5.2.2 *Facial LandMarks.*

Los puntos de referencia del rostro (*facial landmarks*) se definen como un subconjunto de puntos característicos en el rostro, los cuales tienen un impacto en tareas como reconocimiento de rostros, detección de mirada, seguimiento facial, reconocimiento de expresiones y gestos, etc. Los *facial landmarks* más comunes son la punta de la nariz, los bordes de los ojos, el mentón, las comisuras de la boca, el arco de la ceja y los lóbulos de las orejas. Para facilitar el análisis, la mayoría de los algoritmos prefieren una región semántica entera como la zona de la boca, la nariz, los ojos, las cejas, las mejillas o la barbilla. Los puntos de referencia se pueden clasificar en dos grupos, primario y secundario, esta distinción es basada en la fiabilidad de las técnicas de detección de características en una imagen. Generalmente en un rostro son reconocidos entre 60 y 80 puntos como se muestra en la Figura 3.22. La ventaja de la técnica de seguimiento basada en puntos faciales es que las deformaciones del rostro y el movimiento de la cabeza son estimadas en forma conjunta.

Figura 3.22

Facial Landmarks Detection. Puntos fáciles en el rostro. Aproximado de unos 80 puntos.

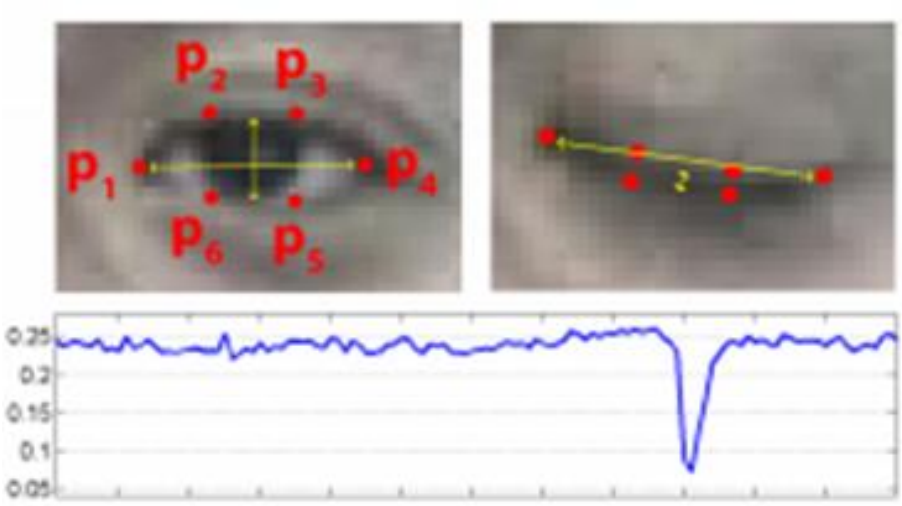
Fuente: Soukupova y Cech (2016).

3.5.2.3 *Eye Aspect Ratio (EAR).*

En términos de detección de somnolencia, la parte fundamental que nos interesa es la zona de los ojos (lado izquierdo y lado derecho). Está representado por coordenadas en un plano bidimensional, ver Figura 3.23 tomando como inicio el ojo izquierdo y luego trabajando en sentido horario (Soukupova y Cech, 2016).

Figura 3.23

Coordenadas bidimensionales del ojo (formados por 6 puntos).



Fuente: Soukupova y Cech (2016).

Según Soukupová y Cech (2016), existe una relación entre el ancho y la altura de estas coordenadas. A partir de ello, se puede derivar una ecuación que refleja esta relación, se le conoce como *Eye aspect ratio* (EAR). Las distancias entre los puntos verticales del ojo se calculan en la parte del denominador, mientras que calcular la distancia entre los puntos horizontales del ojo se calculan en el numerador. El denominador se debe ponderar, ya que existe un solo sistema de puntos horizontales, pero 2 sistemas de puntos verticales; por ello, el denominador se multiplica por dos.

$$EAR = \frac{||p_2 - p_6|| + ||p_3 - p_5||}{2||p_1 - p_4||} \qquad (6)$$

Donde:

- $p_1, p_2, p_3, p_4, p_5, p_6$: Coordenadas bidimensionales del ojo.

Si el EAR resulta mayor a la apertura definida constantemente, no sucede nada y se sigue transmitiendo el siguiente *frame*. Sin embargo, si el EAR resulta menos, sonará la alarma indicando que se encontró somnolencia.

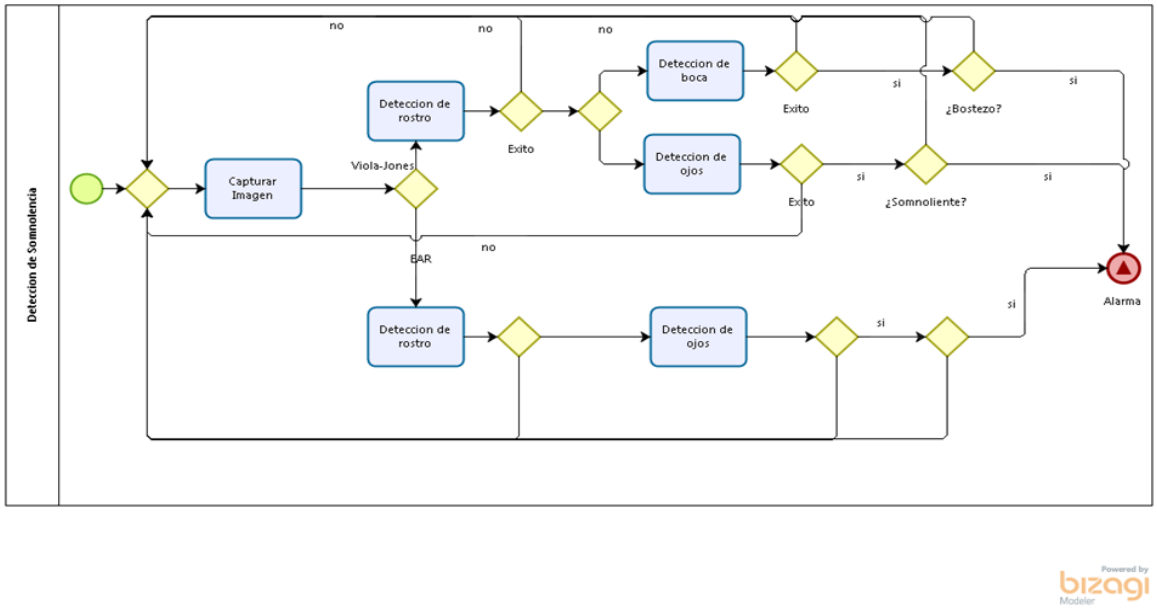
CAPÍTULO IV: DESARROLLO DE LA SOLUCIÓN PROPUESTA

En este capítulo se presenta el desarrollo de la solución propuesta, detallando la secuencia de pasos para la detección de somnolencia de las dos metodologías mencionadas anteriormente.

4.1 Diagrama de bloque.

Figura 4.1

Método de Reconocimiento facial- Diagrama de Bloques.



Fuente: Elaboración propia.

4.2 Desarrollo de la experimentación.

En la Figura 4.1 se muestra los pasos a seguir de las metodologías escogidas desde la captura de la imagen hasta el disparo de la alarma por parte de cada uno de los sistemas, si es que se detecta somnolencia en los participantes. Después de tomar resultados de las pruebas, los compararemos para determinar cuál de ellos es el más eficaz y el que mejor sobrepasa los obstáculos de oclusión, luminosidad y rotación del rostro. Para el desarrollo de la experimentación nos basamos en los códigos abiertos: para la Metodología Viola-Jones, se usó de base el código fuente proporcionado en el artículo *Facial Features Monitoring for Real Time Drowsiness Detection* (B.N, 2016) y para la Metodología *Regression based facial Landmark Detection* se usó de base el código fuente proporcionado por el autor Adrian Rosebrock (Rosebrock, pyimagesearch, 2017) que originalmente eran algoritmos de reconocimiento facial y que posteriormente fueron modificados para motivos del desarrollo de la prueba de concepto y validación.

4.2.1 Captura de la imagen.

La imagen de entrada es obtenida a partir de 3 fuentes: la primera es extraída de los videos de la base de datos "*NTHU Drowsy Driver Detection* (NTHU-DDD)" estos se encuentran en formato de video ".avi", con una resolución de "640 x 480 pixeles" y formato "RGB24"; la segunda fuente, son videos grabados desde una laptop "*HD Webcam* integrada Laptop Toshiba Satellite S55T-A5136" en formato de video ".avi", con una resolución de "640 x 480 pixeles"; y la tercera fuente, son videos grabados desde una Cámara IP en formato de video ".avi", con una resolución de "640 x 360 pixeles".

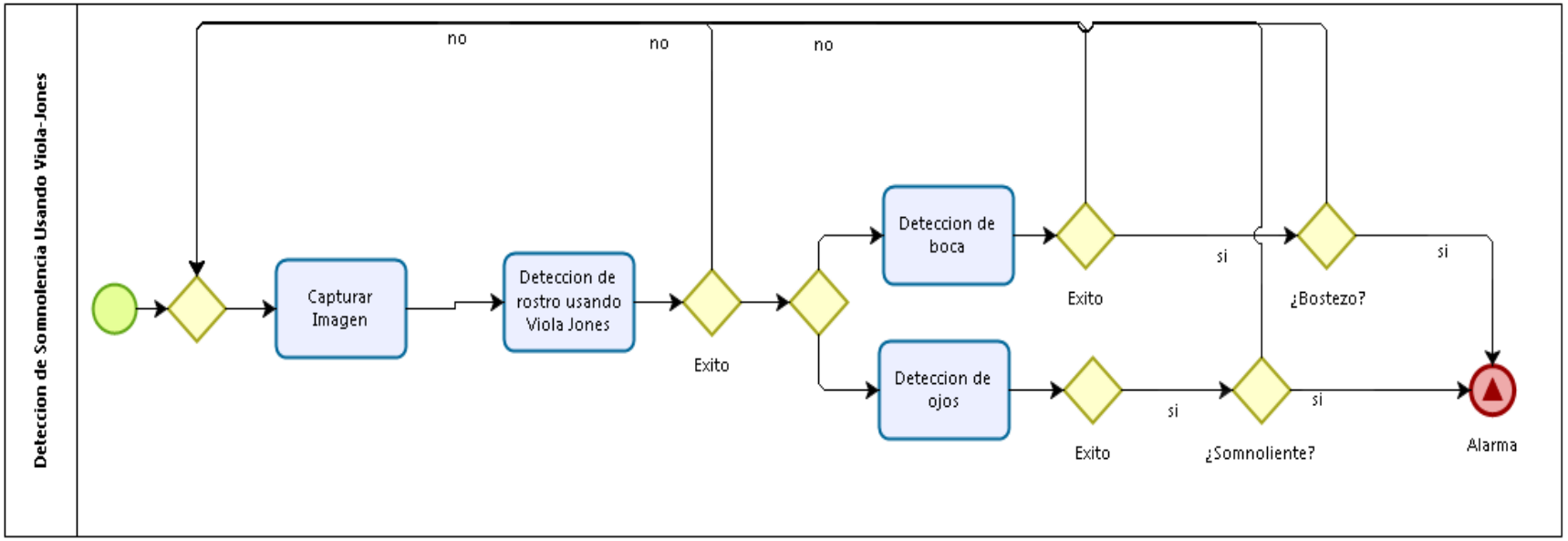
4.2.2 Metodología Viola-Jones.

La implementación del sistema de detección de somnolencia seguirá el siguiente diagrama como se muestra en la Figura 4.2.El proceso comienza con la captura de imagen *frame* por *frame*, hasta que la entrada deje de transmitir, se detectará el rostro, a partir de ello se detectará ojos y boca, y finalmente se disparará la alarma por parte del sistema si es que se detecta somnolencia en los participantes. Para la adaptación de Viola-Jones para que detecte la somnolencia en imágenes se usó de base el código

fuente proporcionado en el artículo *Facial Features Monitoring for Real Time Drowsiness Detection* (B.N, 2016).

Figura 4.2

Diagrama de detección de somnolencia con metodología Viola-Jones.



Fuente: Elaboración propia.

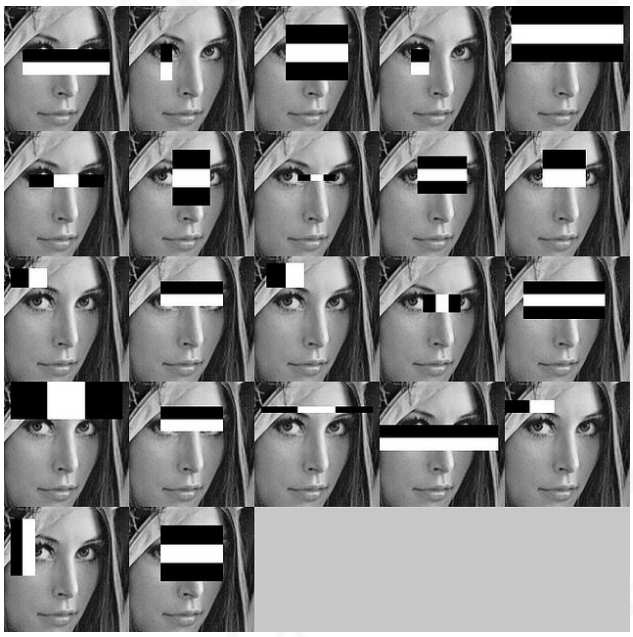
4.2.2.1 Detección del rostro.

La detección del rostro es realizada con el marco de trabajo de Viola-Jones. El objetivo final de la detección de rostros es minimizar las falsas detecciones en la identificación de las expresiones faciales. El primer punto del *framework* de Viola-Jones son los *Haar Features*, explicados anteriormente, que se utilizaran para detectar la presencia de una característica para cada sub-ventana que va a ir recorriendo toda la imagen.

Así que, los *Haar features* son básicamente rectángulos en blanco y negro, ver Figura 4.3. Cada uno de ellos representa una característica que el algoritmo busca en la imagen.

Figura 4.3

Haar Features aplicados a una imagen.

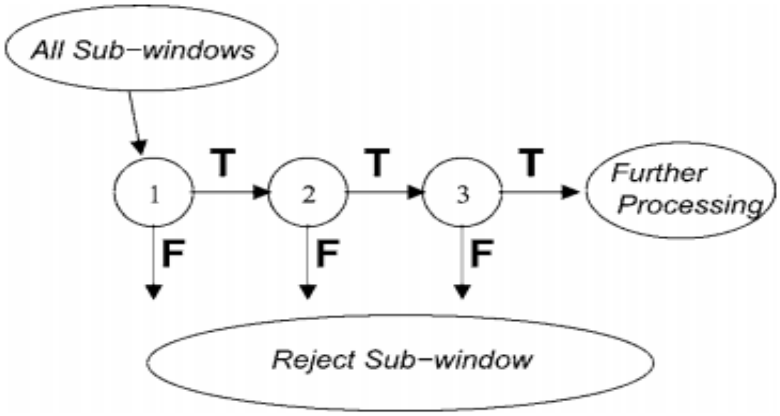


Fuente: Borenstein (2012).

Finalmente se aplica el método de cascada, como se muestra en la Figura 4.4, el clasificador de cascada se compone de etapas que contienen un clasificador fuerte de *AdaBoost*. El trabajo de cada etapa es determinar si una sub-ventana dada no es definitivamente una cara o quizá será una cara. Cuando una sub-ventana se clasifica como no-cara en una etapa determinada, se descarta inmediatamente. A la inversa, una sub-ventana clasificada como una posible cara pasa a la siguiente etapa de la cascada. A partir de ello se puede determinar que cuantas más etapas pasa una sub-ventana, mayor es la probabilidad de que la esta contenga una cara.

Figura 4.4

Representación esquemática del método cascada, todas las sub-ventanas entran al método de cascada para determinar si esa sub-ventana en específico podría ser una cara o no, en caso que, si podría serlo, pasa al siguiente estado, caso contrario se rechaza.



Fuente: Viola y Jones (2004).

4.2.2.2 Detección de somnolencia – Sonido de la alarma.

Luego de haber realizado el reconocimiento facial, de los ojos y de la boca. Se procederá a la definición de algunas reglas para poder reconocer si es que existe somnolencia en el conductor o no.

Los casos donde se detectará somnolencia y sonará la alarma son:

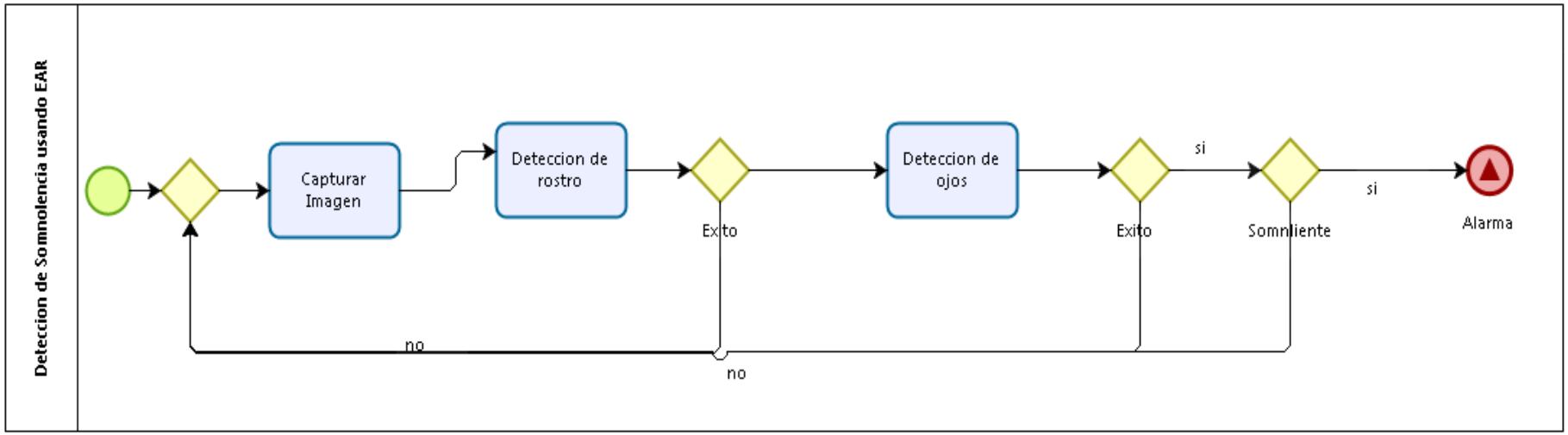
- Para la detección de la boca del usuario, si esta empieza a abrirse en señal de bostezo se activará la alarma, caso contrario, volverá a empezar el flujo, capturando la imagen del siguiente *frame*.
- Para la detección ocular, si se detectan los ojos correctamente se procede a discriminar el estado de los ojos, sino volverá a empezar el flujo, capturando la imagen del siguiente *frame*. Si al discriminar el estado de los ojos, se determina que están cerrados se disparará la alarma, caso contrario, volverá a empezar el flujo, capturando la imagen del siguiente *frame*.

4.2.3 Metodología *Regression Based Facial Landmark Detection.*

La implementación del sistema de detección de somnolencia seguirá el siguiente diagrama como se muestra en la Figura 4.5. El proceso comienza con la captura de imagen frame por frame, hasta que la entrada deje de transmitir, se detectará el rostro, luego se detectará los ojos, y finalmente se disparará la alarma por parte del sistema si es que se detecta somnolencia en los participantes. Para la adaptación de esta metodología para que detecte la somnolencia en imágenes se usó de base el código fuente proporcionado en el artículo "*Real-Time Eye Blink Detection using Facial Landmarks*" (Soukupova y Cech, 2016).

Figura 4.5

Diagrama de detección de somnolencia con metodología *Regression Based Facial Landmark Detection.*



Fuente: Elaboración propia.

A cada *frame* del video se le aplicará el algoritmo de *Histogram of oriented gradients* (HOG), este es un descriptor característico utilizado para poder detectar los objetos, en este caso para la detección de rostro. La ventaja de este método (HOG) es que la apariencia y la forma del objeto se puede describir mediante la distribución de gradientes de intensidad o direcciones de borde.

Este algoritmo se basa en 2 librerías: Open CV que se utiliza para la detección del rostro y Dlib para la obtención de la zona de interés, es decir, los ojos (lado izquierdo y derecho).

Open CV (*Open Source Computer Vision Library*) es una librería de visón artificial por computador de código abierto). Una ventaja de esta librería es que utiliza procesadores con múltiples núcleos para minimizar los gastos de recursos computacionales y dar un enfoque hacia las aplicaciones de tiempo real (Open CV *team,* 2018).

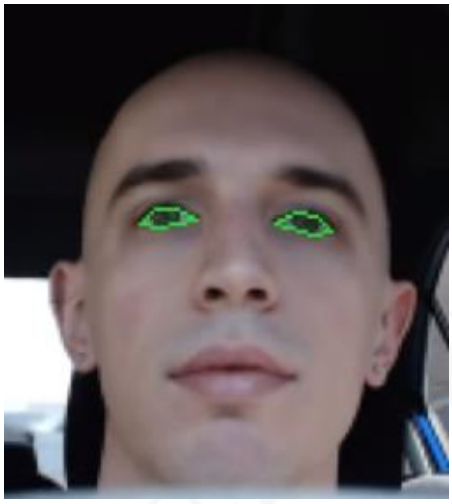
Dlib es una biblioteca de software multiplataforma de uso general que contiene algoritmos de aprendizaje de máquina, compresión, análisis de imágenes, entre otras. Esta librería también es de código abierto y se compila para poder hacer uso de los facial Landmark (Marin, 2017).

4.2.3.1 Detección de ojos.

Luego del reconocimiento del rostro, el paso siguiente es detectar los ojos. Para hacer esto posible, se hace uso de los *Facial Landmark,* que son un subconjunto de puntos característicos en el rostro. En nuestro caso, solo delimitaremos a que marquen las zonas de interés que son: el ojo derecho y el ojo izquierdo como se muestra en la Figura 4.6.

Figura 4.6

Facial Landmark delimitados en la zona de los ojos (izquierdo y derecho).



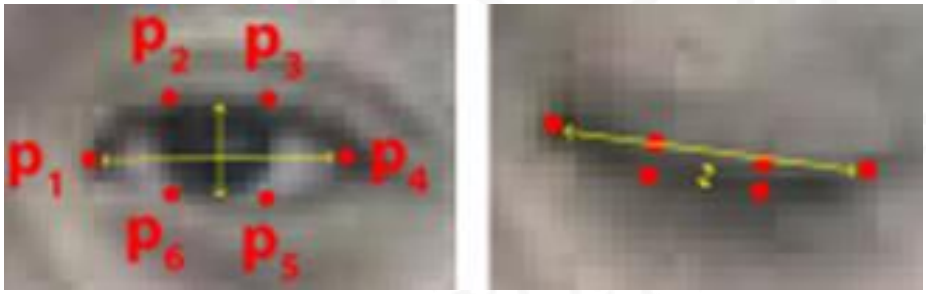
Fuente: Soukupova y Cech (2016).

4.2.3.2 Apertura de los ojos.

Para detectar la somnolencia, la parte fundamental es la zona de los ojos. Estos están representados por coordenadas en un plano bidimensional y existe una relación entre el ancho y alto de estas coordenadas, ver Figura 4.7. A partir de ello, se hace uso de una ecuación que refleja esta relación y se le conoce como *Eye aspect ratio* (EAR).

Figura 4.7

Coordenadas en un plano bidimensional de los ojos.



Fuente: Soukupova y Cech (2016).

4.2.3.3 Detección de somnolencia – Sonido de la alarma.

Luego de haber realizado el reconocimiento facial y de los ojos. Se procederá a la definición de algunas reglas para reconocer si es que existe somnolencia en el conductor.

Los casos donde se detectará somnolencia y sonará la alarma son:

- Si se detectan los ojos correctamente se procede a discriminar el estado de los ojos, sino volverá a empezar el flujo, capturando la imagen del siguiente *frame*. Si al discriminar el estado de los ojos, se determina que la apertura de los ojos (EAR) se encuentra debajo del rango definido y ese comportamiento continua en 30 fotogramas, se llega a la conclusión de que los ojos del conductor están cerrados y se procederá a dispararse la alarma, caso contrario, volverá a empezar el flujo, capturando la imagen del siguiente *frame*.

4.3 Implementación de la propuesta.

4.3.1 Implementación de la metodología Viola-Jones.

En primera instancia, se captura el *frame* del video o de la cámara Web integrada a la laptop, se reconoce el rostro de la imagen gracias al algoritmo Viola-Jones y lo encierra en un recuadro rojo como se muestra en la Figura 4.8.

Figura 4.8

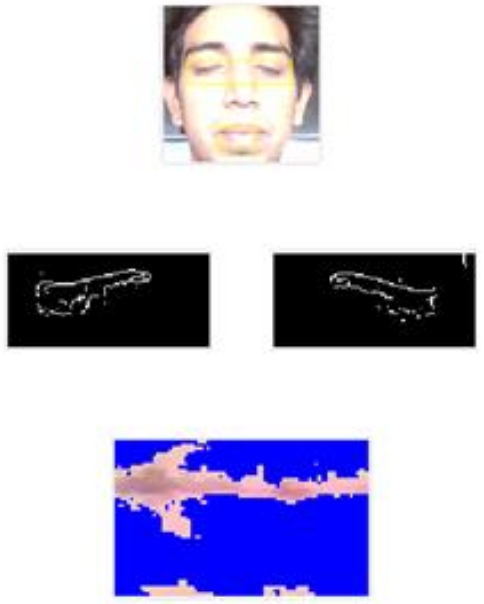
Primeras pruebas con el reconocimiento de rostro Viola-Jones, El algoritmo reconoce el rostro y lo encasilla en un recuadro de color rojo.

Fuente: Elaboración propia.

Luego de la detección facial, se procede a la detección de ojos y boca de la imagen del rostro adquirida con el algoritmo mencionado, teniendo así un recuadro para cada ojo y más abajo un recuadro para la boca como se muestra en la Figura 4.9.

Figura 4.9

A partir del reconocimiento del rostro, empieza el reconocimiento de ojos y boca del usuario para identificar si estos se encuentran abiertos o cerrados.



Fuente: Elaboración propia.

Finalmente, el SVM, previamente entrenado, discrimina si los ojos de la persona están abiertos o cerrados, para determinar la somnolencia del usuario, al igual que la boca si se encuentra abierta, existe la presencia que un bostezo. PERCLOS es el índice más utilizado en la construcción de este tipo de sistemas, por tal motivo ha sido implementado en el presente trabajo. Para calcular el índice, es imprescindible conocer el estado del ojo en cada instante, se manejan dos estados: abierto y cerrado.

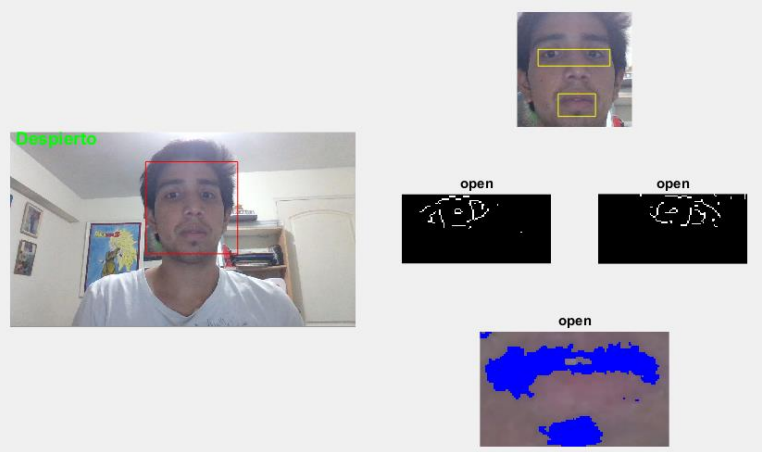
4.3.1.1 Pruebas mediante cámara web integrada.

Este experimento se realizó con una HD Webcam integrada Laptop Toshiba Satellite S55T-A5136 para el reconocimiento del rostro, los ojos y boca para discriminar el estado del usuario y alertarlo en tiempo real. Se tiene una pantalla con la *WebCam* transmitiéndose en tiempo real, a su derecha la ventana del rostro enfocado, luego se muestran los ojos, por separado, convertidos de RGB a *Grayscale* y luego usando *Edge detection* se enfatiza el contraste. En pocas palabras, el sistema detecta el rostro de la persona, luego sus ojos y boca en tiempo real para luego determinar si los ojos están abiertos o cerrados.

En el primer caso, en la Figura 4.10, los ojos están abiertos y la boca está cerrada (el sistema detecta como abierto, pero por el ángulo de abertura no lo clasifica como somnolencia). Dando como resultado el estado “despierto”.

Figura 4.10

La implementación de la propuesta hecha a partir de la metodología Viola-Jones, detecta el rostro y arroja como resultado el estado “despierto”. Reconocimiento de somnolencia webcam. Estado: despierto.

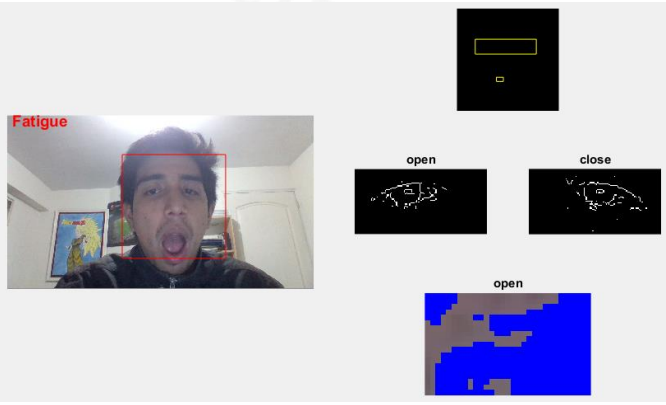


Fuente: Elaboración propia.

Luego, se presenta el caso del bostezo en la Figura 4.11, síntoma de una persona próxima a quedarse dormida. Por eso, el detector señala como estado “fatigado” cuando detecta un bostezo.

Figura 4.11

Se detecta en este caso la presencia de “fatiga” por parte del participante al Bostezar, el programa lo reconoce y lo alerta. Reconocimiento de somnolencia webcam. Bostezo, Estado: somnoliento.

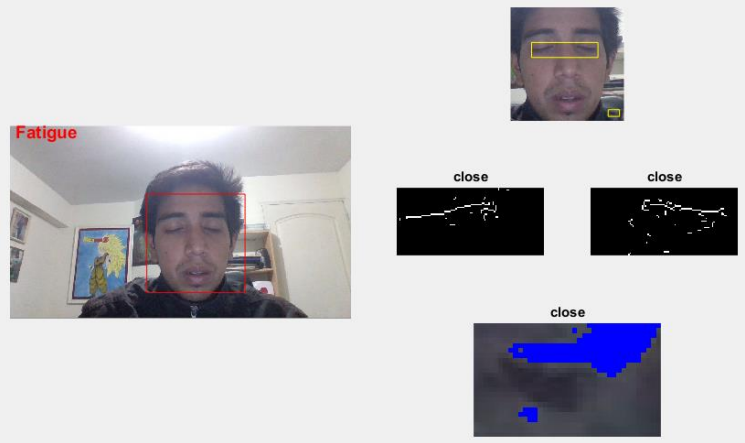


Fuente: Elaboración propia

Por último, para el caso cierre de los ojos como se muestra en la Figura 4.12, da como resultado el estado “fatigado”, los dos ojos están cerrados, la alarma se dispara e inmediatamente el conductor abre sus ojos para seguir con su viaje o para a descansar.

Figura 4.12

En este caso el participante cierra los ojos por un tiempo determinado por eso el programa reconoce su “fatiga” y lanza la alarma. Reconocimiento de somnolencia webcam. Estado: somnoliento.

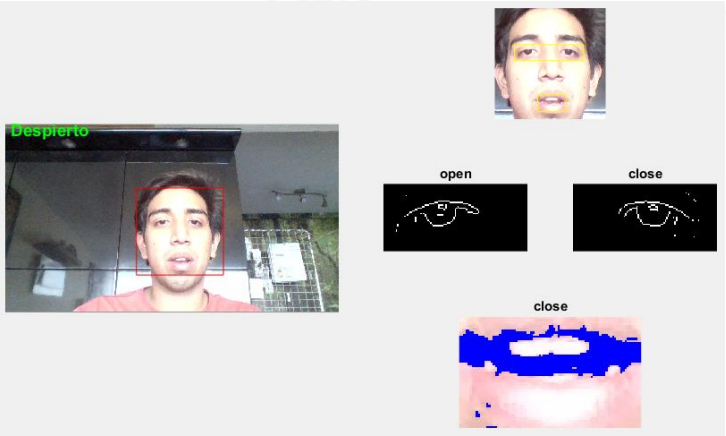


Fuente: Elaboración propia.

El segundo experimento se hizo a la luz del día, ver Figura 4.13, el brillo del primer experimento, iluminado solo con un foco de 65 *Watts*, es mucho menor que la iluminación generada por la luz solar. En esta ocasión el algoritmo también detectó correctamente la cara del participante al igual que sus ojos y boca.

Figura 4.13

Segundo Experimento, el participante se presenta en un escenario con alta luminosidad. Reconocimiento de somnolencia webcam. Estado: despierto, luz solar.

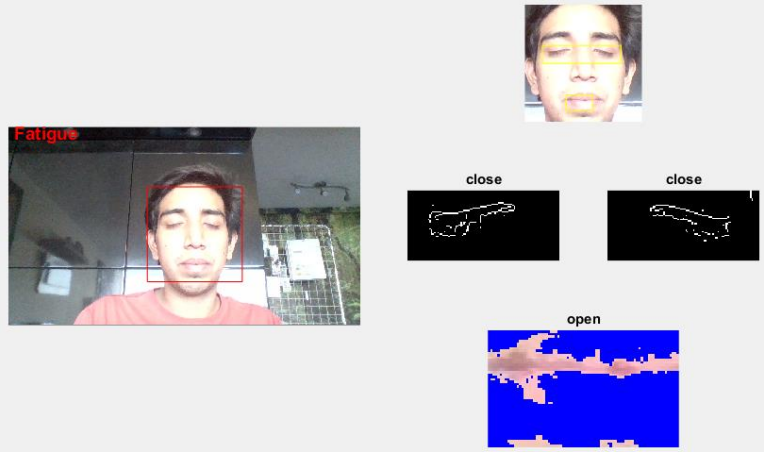


Fuente: Elaboración propia.

Se detecta correctamente el estado de fatiga en el participante, en la Figura 4.14, dado que los ojos del mismo se encuentran cerrados.

Figura 4.14

Segundo experimento, el programa reconoce el estado de somnolencia del usuario ante la presencia de los ojos cerrados. Reconocimiento de somnolencia webcam. Estado: fatigado, luz solar.

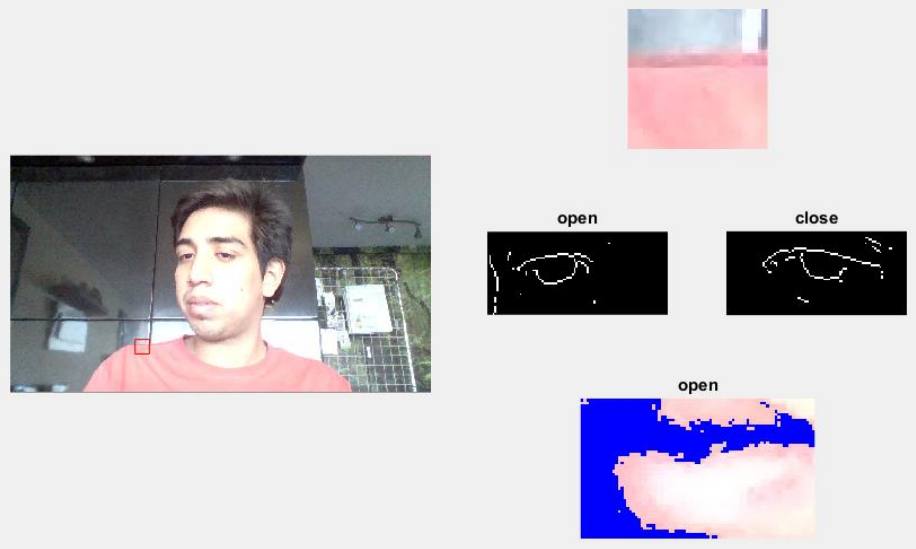


Fuente: Elaboración propia.

El algoritmo Viola-Jones no detecta correctamente el rostro cuando en el *frame* a reconocer se produce una rotación del rostro, ver Figura 4.15 .En el momento de la experimentación no se detectó correctamente el rostro en estas condiciones.

Figura 4.15

Se presenta la limitación de Viola-Jones al momento de la presencia de rotación del rostro, el participante al girar su cara más de 45 grados, la metodología dicha no detecta el rostro de una manera eficiente.



Fuente: Elaboración propia

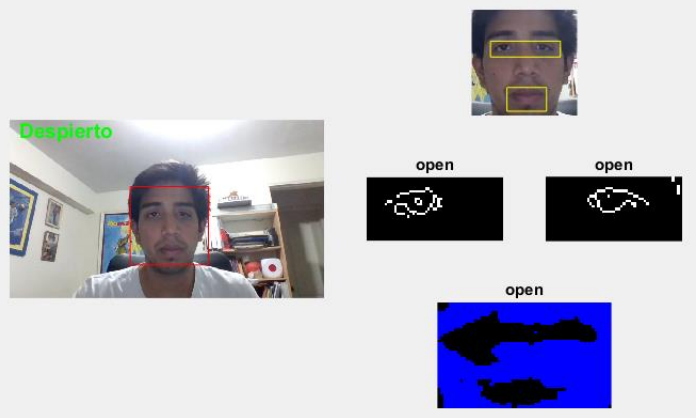
Finalmente, el detector reconoce el estado de "fatigado" cuando se produce cualquiera de estas dos opciones, la primera es cerrando los ojos por un periodo más largo que el de un simple pestañeo, una señal clara de que alguien se está quedando dormido. La segunda se da por el movimiento de la boca al bostezar, otro síntoma de la somnolencia. Además, de las limitaciones de rotación de rostro que posee el algoritmo.

4.3.1.2 Pruebas mediante un video.

En esta ocasión, se le proporciona al sistema un video como *input* para que pueda reconocer el rostro, los ojos y la boca del usuario, para saber si la persona se encuentra somnolienta o no. El video utilizado está en formato de video ".avi", la resolución del mismo es "640 x 480" y formato "RGB24". Según el esquema se puede observar que se cumple, mientras los ojos están abiertos el sistema lo detecta como estado "despierto", tal como sucede en la Figura 4.16.

Figura 4.16

Experimento video, se presenta el usuario en un escenario de alta luminosidad. Reconocimiento de somnolencia video. Estado: despierto.



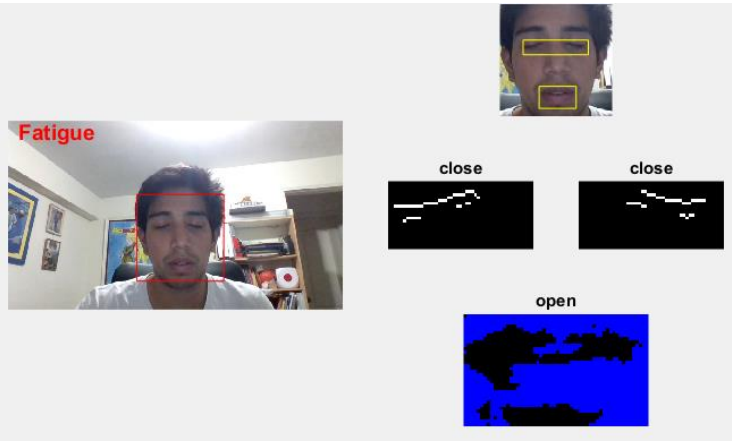
Fuente: Elaboración propia.

Se aprecia la detección de los ojos y la boca gracias al algoritmo Viola-Jones (vemos los dos ojos en estado: *open* y la boca en estado: *open*, pero este último al no tener un ángulo tan pronunciado no llega a ser calificado como estado de "fatiga". El SVM discrimina si la abertura de los ojos es la suficiente para determinar el estado de despierto/ fatigado, al igual que en la boca, el ángulo de abertura para indicar un bostezo.

Para hacer la comprobación del sistema, después de un momento se cierran los ojos y se inclina la cabeza un poco para simular el estado de somnolencia. El sistema lo detecta y dispara la alarma para que despierte el conductor.

Figura 4.17

Experimento Video, con una luminosidad alta, el programa reconoce el estado de fatiga del participante. Reconocimiento de somnolencia video. Estado: somnoliento.



Fuente: Elaboración propia.

Como se aprecia en la Figura 4.17 los ojos están en estado "*close* (cerrado)", el sistema lo detecta como tal y muestra el mensaje de "fatiga".

4.3.2 Implementación de la metodología *Regression Based Facial Landmark Detection.*

4.3.2.1 Pruebas mediante un video.

Se le proporciona al sistema un video como *input* para saber si la persona se encuentra somnolienta o no. El video está en formato AVI, la resolución del mismo es "640 x 480" y se encuentra en formato "RGB24".

Se aprecia en la Figura 4.18, la detección de los ojos gracias a que se aplica los *Facial Landmark* y en la esquina superior derecha se observa el valor del *Eye aspect ratio* (EAR) que sería la medida de la apertura de los ojos.

Figura 4.18

Detección del rostro gracias al algoritmo *Histogram of oriented gradients* (HOG), luego se detecta los ojos gracias a que se aplica los *Facial Landmark*.

Fuente: Elaboración propia.

En este *frame* se puede visualizar a la persona frente a la cámara, con una luminosidad alta y se puede ver que la metodología detecta ambos ojos, ver Figura 4.19.

Figura 4.19

Aplicación de la metodología *Regression Based Facial Landmark Detection y Eye aspect ratio* (EAR). Primera prueba: rostro frontal a la cámara con los ojos abiertos y luminosidad alta. Resultado: Despierto.

Fuente: Elaboración propia.

Luego de detectar la zona de ambos ojos, se podrá calcular el *Eye Aspect Ratio* (EAR) para determinar si los ojos están cerrados o abiertos.

En este *frame*, la persona se encuentra frente a la cámara con los ojos abiertos, por ende, el sistema seguirá tomando *frames* hasta que suceda lo contrario, ver Figura 4.20

Figura 4.20

Aplicación de la metodología *Regression Based Facial Landmark Detection* y *eye aspect ratio*. El EAR determina que ambos ojos se encuentran abiertos y el sistema seguirá pasando los *frames* hasta que vea un comportamiento contrario.

Fuente: Elaboración propia.

En este *frame*, la persona se encuentra frente a la cámara, pero con los ojos cerrados, es ahí donde el sistema de detección empieza a contar el tiempo de cuánto dura este estado, ver Figura 4.21.

Figura 4.21

Aplicación de la metodología *Regression Based Facial Landmark Detection and eye aspect ratio*. El EAR determina que ambos ojos se encuentran cerrados.

Fuente: Elaboración propia.

Si el EAR indica que los ojos se encuentran cerrados durante 36 fotogramas consecutivos, el sistema de detección de somnolencia mandará una alerta de prevención indicando que la persona se encuentra dormida, ver Figura 4.22.

Figura 4.22

Aplicación de la metodología *Regression Based Facial Landmark Detection and eye aspect ratio*. Detección de somnolencia, si los ojos se mantienen cerrados por un determinado tiempo, el sistema reconocerá que existe somnolencia y votará una alerta.

Fuente: Elaboración propia.

CAPÍTULO V: PRUEBAS Y RESULTADOS

En este capítulo se describen las pruebas realizadas con las metodologías de detección de rostro: Viola-Jones y *Regression Based Facial Landmark Detection* adaptados para la detección de somnolencia.

5.1 Metodología de la experimentación.

Para poder evaluar los resultados obtenidos con las metodologías de reconocimiento de somnolencia mencionadas anteriormente, se analizó cada fotograma de cada uno de los videos proporcionados por la base de datos NTHU-DDD y cada fotograma de los videos de nuestra propia base de datos. Las dos metodologías disparan la alarma al momento de detectar la presencia de somnolencia y mediante una verificación manual contamos la cantidad de fotogramas que fueron detectados como somnolientos y también los que fueron clasificados como no-somnolientos.

Para la validación de la propuesta de solución se harán uso de indicadores como: la precisión media (AAC) y la tasa de detección (DR) de acuerdo con las ecuaciones (7) y (8) respectivamente. Estas dos medidas, nos permitirán evaluar la exactitud e indicar que algoritmo propuesto es más aceptable en la detección de somnolencia en la cara de los conductores al momento de manejar su vehículo. Para ello, haremos uso de una matriz de confusión donde:

- NTP = Si tiene fatiga y si ha detectado fatiga correctamente.
- NFN = No tuvo fatiga y si ha detectado fatiga **por error.**
- NTN = No tuvo fatiga y no detecto fatiga.
- NFP = Si tiene fatiga y no ha detectado fatiga **por error.**

$$AAC = (N_{TP} + N_{TN})/(N_{TP} + N_{TN} + N_{FP} + N_{FN}) \quad \text{(7)}$$

$$DR = Verdaderos\ Positivo/Drowsy\ Frames\ = N_{TP}/(N_{TP} + N_{FN}) \quad \text{(8)}$$

5.2 Características de los videos de la base de datos NTHU-DDD.

Se ha hecho el experimento sobre los videos de la base de datos "NTHU *Drowsy Driver Detection* (NTHU-DDD)" en los que se les pidió a los participantes que simulen efectos de somnolencia es decir pestañear, bostezar cuando se tiene sueño; en pequeños intervalos de tiempo.

Características de los videos de la base de datos:

- Duración: entre 00:01:29 y 00:01:34.
- Ancho de fotograma: 640.
- Alto de fotograma:480.
- Velocidad fotograma: 15/30 fotogramas/segundo.
- Luminosidad: Alta/Baja.

Los videos con el escenario de luminosidad baja presentan 15 fotogramas/segundo y los demás escenarios presentan 30 fotogramas/segundo.

Para las pruebas realizadas se cuentan con cuatro tipos de escenarios, cada uno de los escenarios con ocho videos de diferentes personas que poseen características distintas, ya sea por etnia, color de piel, tamaño de ojo, etc. Los cuatro escenarios de prueba son los siguientes:

1. Sin oclusión y luminosidad alta.
2. Sin oclusión y luminosidad baja.
3. Con oclusión y luminosidad alta.
4. Con oclusión y luminosidad baja.

Consideramos los siguientes criterios:

- Sin oclusión: escenario donde los participantes no utilizan anteojos.
- Con oclusión: escenario donde los participantes usan anteojos.
- Luminosidad alta: escenario donde el participante se encuentra en un espacio donde exista luz natural (día/tarde) o luz artificial.
- Luminosidad baja: escenario donde el participante se encuentra en un espacio donde la existencia de luz sea mínima (noche) o nula (habitación sin luz).

Figura 5.1

Capturas de imágenes de los tipos de escenarios existentes en la base de datos puestos a prueba (Imágenes capturadas de los videos proporcionados por en la base de datos NTHU DDD).

Fuente: Base de datos NTHU DDD.

5.3 Características de los videos de nuestra propia base de datos.

Se realizó el experimento sobre los videos que hemos grabado para la base de datos propia en los que se les pidió a los participantes que simulen efectos de somnolencia es decir pestañear, bostezar cuando se tiene sueño; en pequeños intervalos de tiempo, con el fin de poder determinar qué tan optimo son las metodologías de *Viola-Jones* y *Regression Based Facial Landmark Detection* en el reconocimiento de somnolencia con rostros latinos. Imitando los 4 escenarios que utilizaremos de la base de datos NTHU-DDD.

Características de los videos de la base de datos:

- Duración: entre 00:01:40 y 00:02:05.
- Ancho de fotograma: 640.
- Alto de fotograma: 360.
- Velocidad fotograma: 10/15 fotogramas/segundo.
- Luminosidad: Alta/Baja.

Los videos con el escenario de luminosidad baja presentan 10 fotogramas/segundo y los demás escenarios presentan 15 fotogramas/segundo.

Para las pruebas realizadas se cuentan con cuatro tipos de escenarios, cada uno de los escenarios con ocho videos de diferentes personas que poseen características distintas: color de piel, tamaño de ojo, etc. Sin embargo, todos los videos de la base de datos son rostros latinos. Los cuatro escenarios de prueba son los siguientes:

1. Sin oclusión y luminosidad alta.
2. Sin oclusión y luminosidad baja.
3. Con oclusión y luminosidad alta.
4. Con oclusión y luminosidad baja.

Consideramos los siguientes criterios:

- Sin oclusión: escenario donde los participantes no utilizan anteojos.
- Con oclusión: escenario donde los participantes usan anteojos.
- Luminosidad alta: escenario donde el participante se encuentra en un espacio donde exista luz natural (día/tarde) o luz artificial.

- Luminosidad baja: escenario donde el participante se encuentra en un espacio donde la existencia de luz sea mínima (noche) o nula (habitación sin luz).

Figura 5.2

Capturas de imágenes de los tipos de escenarios existentes en nuestra base de datos puestos a prueba (elaboración propia con nuestra base de datos).

Fuente: Elaboración propia.

5.4 Experimentación usando videos de la base de datos NTHU-DDD.

5.4.1 Situación 1: Sin oclusión y luminosidad alta.

Para el primer caso "Sin oclusión y luminosidad alta", se presentan 8 pruebas con participantes sin anteojos y con una luminosidad alta dando como resultado:

- Método: *Regression Based Facial Landmark Detection.*

Tabla 5.1

Resultados - Situación 1: Sin oclusión y luminosidad alta - Método: *Regression Based Facial Landmark Detection.*

Intento	Número de *Frames*	Número de *Frames* con Fatiga		Número de *Frames* sin Fatiga		AAC	DR
		NTP	NFN	NTN	NFP		
1	1000	36	0	898	66	93.4%	100%
2	1000	29	0	864	107	89.3%	100%
3	1000	264	0	657	79	92.1%	100%
4	1000	0	0	823	177	82.3%	0%
5	1000	66	0	839	95	90.5%	100%
6	1000	36	0	904	60	94%	100%
7	1000	76	0	777	147	85.3%	100%
8	1000	7	0	816	177	82.3%	100%
Total	8000	514	0	6578	908	88.65%	100%

Fuente: Elaboración propia.

- Comentarios:
  - Se obtuvo un poco más del 88 % de precisión media para el algoritmo de *Regression Based Facial Landmark Detection*, y un 100% en la tasa de detección de somnolencia. Esto quiere decir que el algoritmo, bajo estas condiciones, tiene un alto porcentaje de reconocimiento, ya que se puede observar que en todas las pruebas realizadas se detectó la zona de interés, y no se encontró ningún falso negativo dentro de las pruebas.
  - A pesar de que en los videos son grabados con un cierto grado de rotación del rostro, el algoritmo supera este obstáculo, detectando los ojos de manera satisfactoria. Sin embargo, si el rostro gira 90° a los costados, o bajamos demasiado el rostro, esta perderá el

enfoque haciendo que el algoritmo no reconozca la zona de interés.

- Método: Viola-Jones.

Tabla 5.2

Resultados - Situación 1: Sin oclusión y luminosidad alta - Método: Viola-Jones.

Intento	Número de *Frames*	Número de *Frames* con Fatiga		Número de *Frames* sin Fatiga		AAC	DR
		NTP	NFN	NTN	NFP		
1	1000	30	127	768	75	79.8%	19.11%
2	1000	8	154	822	16	83.0%	4.94%
3	1000	38	174	733	55	77.1%	17.92%
4	1000	202	502	296	0	49.8%	28.69%
5	1000	139	609	228	24	36.7%	18.58%
6	1000	44	492	464	0	50.8%	8.2%
7	1000	272	438	287	3	55.9%	38.31%
8	1000	150	581	269	0	41.9%	20.52%
Total	8000	883	3077	3867	173	59.38%	22.30%

Fuente: Elaboración propia.

- Comentarios:
  - En este caso, para el algoritmo de Viola-Jones se obtuvo una precisión media menor a 60%, este resultado se da porque los participantes en los videos no presentan el rostro de frente, sino que se tiene un ángulo de inclinación del rostro muy elevado y esta metodología no es invariante a la rotación
  - Tasa de detección de 22.3%, resultado bajo por la misma razón.

5.4.2 Situación 2: Sin oclusión y luminosidad baja.

Para el segundo caso "Sin Oclusión y luminosidad baja" igual se hicieron 8 pruebas, para las dos metodologías propuestas, con participantes sin anteojos y con una luminosidad baja dando como resultado:

- Método: *Regression Based Facial Landmark Detection.*

Tabla 5.3

Resultados - Situación 2: Sin oclusión y luminosidad baja - Método: *Regression Based Facial Landmark Detection.*

Intento	Número de *Frames*	Número de *Frames* con Fatiga		Número de *Frames* sin Fatiga		AAC	DR
		NTP	NFN	NTN	NFP		
1	1000	83	338	521	58	60.4%	19.71%
2	1000	231	212	495	62	72.6%	52.14%
3	1000	214	158	346	282	56%	57.53%
4	1000	282	0	617	101	89.9%	100%
5	1000	0	0	869	131	86.9%	0%
6	1000	544	69	229	158	77.3%	88.74%
7	1000	453	161	265	121	71.8%	73.78%
8	1000	210	24	631	135	84.1%	89.74%
Total	8000	2017	962	3973	1048	74.88%	67.71%

Fuente: Elaboración propia.

- Comentarios:
  - Se obtuvo un poco más del 74 % de precisión media para el algoritmo de *Regression Based Facial Landmark Detection,* y un 67.71% en la tasa de detección de somnolencia. Esto quiere decir que el algoritmo, bajo estas condiciones, tiene un porcentaje menor de reconocimiento que a la "Situación 1", ya que al realizar las pruebas hubieron casos donde se encontraron *frames* en que la persona no estaba fatigado, sin embargo, el algoritmo indicaba lo contrario haciendo sonar la alarma, debido a que no fueron grabados de manera frontal sino de manera diagonal de abajo hacia arriba, haciendo que la apertura de los ojos no se distinga de manera adecuada, dándole una apertura menor a 0.30 definida en el algoritmo.
  - A pesar de que en los videos son grabados con un cierto grado de rotación del rostro, el algoritmo supera este obstáculo, detectando los ojos de manera satisfactoria. Sin embargo, si el rostro gira 90° a los costados, o bajamos demasiado el rostro, este perderá el

enfoque haciendo que el algoritmo no reconozca la zona de interés.

- Método: Viola-Jones.

Tabla 5.4

Resultados - Situación 2: Sin oclusión y luminosidad baja - Método: Viola-Jones.

Intento	Número de *Frames*	Número de *Frames* con Fatiga		Número de *Frames* sin Fatiga		AAC	DR
		NTP	NFN	NTN	NFP		
1	1000	4	70	926	0	93.00%	5.41%
2	1000	185	42	773	0	95.80%	81.50%
3	1000	34	135	831	0	86.50%	20.12%
4	1000	62	107	582	249	64.40%	36.69%
5	1000	130	104	739	27	86.90%	55.56%
6	1000	204	119	586	91	79.00%	63.16%
7	1000	264	195	522	19	78.60%	57.52%
8	1000	166	295	528	11	69.40%	36.01%
Total	8000	1049	1067	5487	397	81.70%	49.57%

Fuente: Elaboración propia.

- Comentarios:
  - En el segundo escenario Viola - Jones aumenta el porcentaje de precisión a 81.70% ya que se adecua mejor a la luminosidad que la metodología de *Regression Based Facial Landmark Detection*. Además de aumentar también la tasa de detección a 49.57% por la misma razón que se adecua mejor a la luminosidad y que los estos 8 videos pudieron haber estado en un ángulo menor de inclinación del rostro.
  - Este resultado si se compara con la "Situación 1" llega a tener una precisión mayor por un poco más de 20%, pero sigue siendo baja la tasa de detección que debería mejorar cuando el rostro se encuentra de manera frontal hacia la cámara al momento de grabar el video.

5.4.3 Situación 3: Con oclusión y luminosidad alta.

Para el tercer caso "Con Oclusión y luminosidad alta" se realizaron 8 pruebas para las dos metodologías propuestas, con participantes con anteojos y con una luminosidad alta dando como resultado:

- Método: *Regression Based Facial Landmark Detection.*

Tabla 5.5

Resultados - Situación 3: Con oclusión y luminosidad alta - Método: *Regression Based Facial Landmark Detection.*

Intento	Número de *Frames*	Número de *Frames* con Fatiga		Número de F*rames* sin Fatiga		AAC	DR
		NTP	NFN	NTN	NFP		
1	1000	18	26	911	45	92.90%	40.91%
2	1000	29	56	707	208	73.60%	34.12%
3	1000	0	0	965	35	96.50%	0%
4	1000	61	39	751	149	81.20%	61.00%
5	1000	0	0	887	113	88.70%	0%
6	1000	532	0	200	268	73.20%	100.00%
7	1000	290	15	535	160	82.50%	95.08%
8	1000	246	0	688	66	93.40%	100.00%
Total	8000	1176	136	5644	1044	85.25%	89.63%

Fuente: Elaboración propia.

- Comentarios:
  - Para la tercera situación, se obtuvo una precisión media de 85.25% y un 89.63% de detección de somnolencia en promedio. Esto quiere decir que, bajo estas condiciones, el algoritmo tiene un alto porcentaje de reconocimiento y es de mayor precisión que el algoritmo de Viola-Jones, a pesar del obstáculo de los lentes.
  - Dentro de las pruebas, se encontraron *frames* donde si bien se detectaba la zona de interés, en algunos casos, se perdía y detectaba la zona del parpado inferior/superior, debido al uso de los lentes, provocando que, si la persona estaba fatigada, no detectara el algoritmo ya que tomaba la apertura del parpado.

- A pesar de que en los videos son grabados con un cierto grado de rotación del rostro, el algoritmo supera este obstáculo, detectando los ojos de manera satisfactoria. Sin embargo, si el rostro gira 90° a los costados, o bajamos demasiado el rostro, este perderá el enfoque haciendo que el algoritmo no reconozca la zona de interés.

- Método: Viola-Jones.

Tabla 5.6

Resultados - Situación 3: Con oclusión y luminosidad alta - Método: Viola-Jones.

Intento	Número de *Frames*	Número de *Frames* con Fatiga		Número de *Frames* sin Fatiga		AAC	DR
		NTP	NFN	NTN	NFP		
1	1000	0	183	817	0	81.70%	0%
2	1000	-	-	-	-	-	-
3	1000	0	176	824	0	82.40%	0%
4	1000	-	-	-	-	-	-
5	1000	-	-	-	-	-	-
6	1000	-	-	-	-	-	-
7	1000	-	-	-	-	-	-
8	1000	396	442	162	0	55.80%	47.26%
Total	8000	396	801	1803	0	73.30%	33.08%

Fuente: Elaboración propia.

- Comentarios:
  - Para este tercer escenario, la metodología Viola-Jones no detectó el rostro ni los ojos del participante en varios videos, por dos factores: la rotación del rostro de las personas en algunos videos y la oclusión que se generó por los lentes usados. Por lo que se obtuvo los resultados solo de los videos que pudieron ser evaluados.
  - Los primeros escenarios no tuvieron muchos problemas al momento de la detección del rostro a pesar de la rotación del mismo, presente en todos los videos, pero en esta situación con

participantes con anteojos (es decir con una oclusión en el rostro) se tornó más complejo detectar el rostro y por consiguiente los ojos.

5.4.4 Situación 4: Con oclusión y luminosidad baja.

Para el cuarto caso "Con oclusión y luminosidad baja" se realizaron 8 pruebas para las dos metodologías propuestas, con participantes con anteojos y con una luminosidad baja dando como resultado:

- Método: *Regression Based Facial Landmark Detection.*

Tabla 5.7

Resultados - Situación 4: Con oclusión y luminosidad baja - Método: *Regression Based Facial Landmark Detection.*

Intento	Número de *Frames*	Número de *Frames* con Fatiga		Número de *Frames* sin Fatiga		AAC	DR
		NTP	NFN	NTN	NFP		
1	1000	109	0	776	115	88.5%	100%
2	1000	0	0	891	109	89.1%	0%
3	1000	75	528	222	175	29.7%	12.44%
4	1000	115	65	642	178	75.7%	63.89%
5	1000	0	0	767	233	76.7%	0%
6	1000	135	40	616	209	75.1%	77.14%
7	1000	157	0	602	241	75.9%	100%
8	1000	112	15	691	182	80.3%	88.19%
Total	8000	703	648	5207	1442	73.88%	52.04%

Fuente: Elaboración propia.

- Comentarios:
  - Para la cuarta situación, se obtuvo una precisión media de 73.88% y una tasa de detección de 52.04% en promedio. Bajo estas condiciones de pruebas, el algoritmo tiene un porcentaje cercano que la metodología de Viola-Jones, a pesar de la condición de luminosidad y oclusión. Sin embargo, el porcentaje de precisión y la tasa de detección es menor a los 3 escenarios anteriores.

- Dentro de las pruebas, se encontraron *frames* donde si bien se detectaban la zona de interés, en otras, se perdía debido al reflejo de la luz en los lentes, detectando así el parpado inferior/superior, provocando que, si la persona estaba fatigada, no detectara el algoritmo, ya que tomaba la apertura del parpado.
- A pesar de que en los videos son grabados con un cierto grado de rotación del rostro, el algoritmo supera este obstáculo, detectando los ojos de manera satisfactoria en la mayoría de los frames. Sin embargo, si el rostro gira 90° a los costados, o bajamos demasiado el rostro, este perderá el enfoque haciendo que el algoritmo no reconozca la zona de interés.

- Método: Viola-Jones.

Tabla 5.8

Resultados - Situación 4: Con oclusión y luminosidad baja - Método: Viola-Jones.

Intento	Número de *Frames*	Número de *Frames* con Fatiga		Número de *Frames* sin Fatiga		AAC	DR
		NTP	NFN	NTN	NFP		
1	1000	-	-	-	-	-	-
2	1000	-	-	-	-	-	-
3	1000	100	75	785	40	88.50%	57.14%
4	1000	112	453	547	0	59.26%	19.82%
5	1000	-	-	-	-	-	-
6	1000	-	-	-	-	-	-
7	1000	-	-	-	-	-	-
8	1000	336	182	482	0	81.80%	64.86%
Total	8000	548	710	1814	40	75.90%	43.56%

Fuente: Elaboración propia.

- Comentarios:
  - En este último escenario de pruebas, al igual que el anterior, algunos intentos tampoco se pudieron realizar bien porque el algoritmo no detectó el rostro (lo hizo por unos pocos *frames*, insuficientes para realizar la prueba optima). Los ojos se perdían

con el reflejo de la luz en los anteojos del mismo participante, por eso no se pudo realizar la prueba completa y solo se obtuvo los resultados de los videos que si se pudieron evaluar.

- Los ojos y los rostros de los usuarios con lentes y además con una luminosidad baja no pueden ser detectados con facilidad por el algoritmo de Viola-Jones, el cual solo pudo captar tres de los ocho videos de prueba para este escenario.

5.5 Experimentación usando videos de nuestra propia base de datos.

5.5.1 Situación 1: Sin oclusión y luminosidad alta.

Para el primer caso "Sin oclusión y luminosidad alta", se presentan 8 pruebas con participantes sin anteojos y con una luminosidad alta dando como resultado:

- Método: *Regression Based Facial Landmark Detection.*

Tabla 5.9

Resultados - Situación 1: Sin oclusión y luminosidad alta - Método: *Regression Based Facial Landmark Detection.*

Intento	Número de *Frames*	Número de *Frames* con Fatiga		Número de *Frames* sin Fatiga		AAC	DR
		NTP	NFN	NTN	NFP		
1	1000	209	0	643	148	85,20%	100%
2	1000	235	31	593	141	80,20%	100%
3	1000	264	0	587	149	85,10%	100%
4	1000	430	0	434	136	86,40%	100%
5	1000	50	0	797	153	84,70%	100%
6	1000	229	0	704	67	93,30%	100%
7	1000	127	62	725	86	85,20%	67%
8	1000	177	0	758	65	93,50%	100%
Total	8000	1721	93	5241	945	87,03%	95%

Fuente: Elaboración propia.

- Comentarios:
  - Se obtuvo un poco más del 87.03 % de precisión media para el algoritmo de *Regression Based Facial Landmark Detection*, y un 95% en la tasa de detección de somnolencia. Esto quiere decir que el algoritmo, bajo estas condiciones, tiene un alto porcentaje

de reconocimiento, ya que se puede observar que en todas las pruebas realizadas se detectó la zona de interés; se detectaron 93 falsos negativos dentro de la prueba, ya que en participantes tendieron a sonreír en la grabación y el algoritmo los detecto como somnolientos, ya que se redujo la apertura de los ojos.

- A pesar de que en los videos son grabados con un cierto grado de rotación del rostro, el algoritmo supera este obstáculo, detectando los ojos de manera satisfactoria. Sin embargo, si el rostro gira 90° a los costados, o bajamos demasiado el rostro, esta perderá el enfoque haciendo que el algoritmo no reconozca la zona de interés.

- Método: Viola-Jones.

Tabla 5.10

Resultados - Situación 1: Sin oclusión y luminosidad alta - Método: Viola-Jones.

Intento	Número de *Frames*	Número de *Frames* con Fatiga		Número de *Frames* sin Fatiga		AAC	DR
		NTP	NFN	NTN	NFP		
1	1000	255	339	406	0	66,10%	43%
2	1000	204	427	369	0	57,30%	32%
3	1000	291	467	242	0	53,30%	38%
4	1000	341	200	459	0	80,00%	63%
5	1000	16	0	847	147	85,45%	100%
6	1000	382	592	26	0	40,80%	39%
7	1000	201	499	249	51	45,00%	29%
8	1000	379	435	167	19	54,60%	47%
Total	8000	2069	2959	2765	217	60,35%	53,45%

Fuente: Elaboración propia.

- Comentarios:
  - Se obtuvo una precisión media de 60.35% para la metodología Viola-Jones siendo un numero bajo con respecto al de *Regression Based Facial Landmark Detection* que tiene un 87% y una tasa de detección de 53.54%, esta última aumenta considerablemente

cuando se compara con las pruebas realizadas a los videos de la Base de datos NTHU-DDD, ver Tabla 5.2.

- A pesar que los videos que se grabaron no tienen un alto grado de rotación de rostro, el algoritmo sigue presentando algunos problemas de identificación de la somnolencia en la mayoría de pruebas que puede deberse a diferentes variables tales como el ángulo en que la cámara debe encontrarse al momento de grabar los videos o la intensidad de luz que se encuentra en el fondo del video, por ejemplo.

5.5.2 Situación 2: Sin oclusión y luminosidad baja.

Para el segundo caso "Sin Oclusión y luminosidad baja" igual se hicieron 8 pruebas, para las dos metodologías propuestas, con participantes sin anteojos y con una luminosidad baja dando como resultado:

- Método: *Regression Based Facial Landmark Detection.*

Tabla 5.11

Resultados - Situación 2: Sin oclusión y luminosidad baja - Método: *Regression Based Facial Landmark Detection.*

Intento	Número de *Frames*	Número de *Frames* con Fatiga		Número de *Frames* sin Fatiga		AAC	DR
		NTP	NFN	NTN	NFP		
1	1000	237	0	668	95	90,50%	100%
2	1000	201	18	657	124	85,80%	92%
3	1000	259	14	657	70	91,60%	95%
4	1000	368	21	504	107	87,20%	95%
5	1000	128	0	660	212	78,80%	100%
6	1000	182	3	661	154	84,30%	98%
7	1000	237	0	670	93	90,70%	100%
8	1000	64	41	745	150	80,90%	61%
Total	8000	1676	97	5222	1005	86,23%	95%

Fuente: Elaboración propia.

- Comentarios:

- Se obtuvo un poco más del 86 % de precisión media para el algoritmo de *Regression Based Facial Landmark Detection*, y un 95% en la tasa de detección de somnolencia. Esto quiere decir que el algoritmo, bajo estas condiciones, tiene un porcentaje de reconocimiento similar que a la "Situación 1".
- A pesar de que en los videos son grabados con un cierto grado de rotación del rostro, el algoritmo supera este obstáculo, detectando los ojos de manera satisfactoria. Sin embargo, si el rostro gira 90° a los costados, o se baja demasiado el rostro, este perderá el enfoque haciendo que el algoritmo no reconozca la zona de interés.
- Hubo casos donde al momento de bajar el rostro, el sistema detectaba como región de interés los parpados; haciendo que, en algunos casos, el algoritmo no detecte la somnolencia en los participantes.
- Al realizar las pruebas, hubo casos donde se encontraron *frames* donde la persona no se encontraba fatigada, sin embargo, el algoritmo indicaba lo contrario haciendo sonar la alarma, debido a que algunos videos no fueron grabados de manera frontal al participante, haciendo que la apertura de los ojos no se distinga de manera adecuada.

- Método: Viola-Jones.

Tabla 5.12

Resultados - Situación 2: Sin oclusión y luminosidad baja - Método: Viola-Jones.

Intento	Número de *Frames*	Número de *Frames* con Fatiga		Número de *Frames* sin Fatiga		AAC	DR
		NTP	NFN	NTN	NFP		
1	1000	212	283	501	4	71,30%	43%
2	1000	172	189	638	1	81,00%	48%
3	1000	212	107	601	80	81,30%	66%
4	1000	221	292	442	45	66,30%	43%
5	1000	212	467	191	130	40,30%	31%
6	1000	319	300	381	0	70,00%	52%
7	1000	265	317	247	171	51,20%	46%
8	1000	297	222	457	24	75,40%	57%
Total	8000	1910	2177	3458	455	67,10%	47%

Fuente: Elaboración propia.

- Comentarios:
  - En el segundo escenario Viola-Jones aumenta el porcentaje de precisión media a 67.10% con respecto a la anterior situación, es decir que la metodología se adecua de una mejor forma al cambio de luminosidad, la tasa de detección a 47% es casi la misma entre los dos escenarios: Luminosidad Alta y Luminosidad Baja, esta es una razón más de remarcar que Viola-Jones es invariante ante la luminosidad.
  - Este resultado si se compara con Tabla 5.11 estas pruebas llegaron a tener una precisión mayor por un poco más de 7%, pero sigue siendo baja al igual que la tasa de detección que debería haber mejorado cuando el rostro se encuentra de manera frontal hacia la cámara al momento de grabar el video.

5.5.3 Situación 3: Con oclusión y luminosidad alta.

Para el tercer caso "Con oclusión y luminosidad alta" se realizaron 8 pruebas para las dos metodologías propuestas, con participantes con anteojos y con una luminosidad alta dando como resultado:

- Método: *Regression Based Facial Landmark Detection.*

Tabla 5.13

Resultados - Situación 3: Con oclusión y luminosidad alta - Método: *Regression Based Facial Landmark Detection.*

Intento	Número de *Frames*	Número de *Frames* con Fatiga		Número de F*rames* sin Fatiga		AAC	DR
		NTP	NFN	NTN	NFP		
1	1000	350	25	545	80	89,50%	93%
2	1000	230	36	592	142	82,20%	86%
3	1000	264	27	592	117	85,60%	91%
4	1000	98	255	556	91	65,40%	28%
5	1000	227	0	670	103	89,70%	100%
6	1000	205	39	678	78	88,30%	84%
7	1000	201	33	692	74	89,30%	86%
8	1000	66	32	719	183	78,50%	67%
Total	8000	1641	447	5044	868	83,56%	79%

Fuente: Elaboración propia.

- Comentarios:
  - Para la tercera situación, se obtuvo una precisión media de 83.56% y un 79% de detección de somnolencia en promedio. Esto quiere decir que, bajo estas condiciones, el algoritmo tiene una tasa de detección aceptable y es mayor que el algoritmo de Viola-Jones, a pesar del obstáculo de los lentes.
  - Dentro de las pruebas, se encontraron *frames* en los que se detectaba la zona del parpado inferior/superior, debido al uso de lentes, provocando que el sistema no reconociera cuando el participante presentaba signos de fatiga de manera correcta.
  - Además, se detectaron algunos falsos negativos dentro de la prueba, ya que algunos participantes tendieron a sonreír en la

grabación, haciendo que el sistema los detectara como somnolientos, debido a la reducción de la apertura de los ojos.

- A pesar de que en los videos son grabados con un cierto grado de rotación del rostro, el algoritmo supera este obstáculo, detectando los ojos de manera satisfactoria. Sin embargo, si el rostro gira 90° a los costados, o bajamos demasiado el rostro, esta perderá el enfoque haciendo que el algoritmo no reconozca la zona de interés.

- Método: Viola-Jones.

Tabla 5.14

Resultados - Situación 3: Con oclusión y luminosidad alta - Método: Viola-Jones.

Intento	Número de *Frames*	Número de *Frames* con Fatiga		Número de *Frames* sin Fatiga		AAC	DR
		NTP	NFN	NTN	NFP		
1	1000	242	359	304	95	54,60%	40%
2	1000	194	423	345	38	53,90%	31%
3	1000	242	476	271	11	51,30%	34%
4	1000	321	243	423	13	74,40%	57%
5	1000	87	0	762	151	84,90%	100%
6	1000	285	167	515	33	80,00%	63%
7	1000	365	311	324	0	68,90%	54%
8	1000	115	68	664	153	77,90%	63%
Total	8000	1851	2047	3608	494	68,24%	47%

Fuente: Elaboración propia.

- Comentarios:
  - En este caso la precisión media llego a 68,24%, Viola – Jones no supera la barrera de la oclusión porque en estos casos en algunos *frames* no pudieron ser detectados sus ojos ni su boca cuando el participante bosteza o cuando inclina su cabeza hacia abajo en señal de fatiga, es decir pierde el foco, pero no solo sucede al momento de la somnolencia.

- En estas pruebas solo se observó que, en algunos videos al inclinar el rostro hacia abajo, el algoritmo detectaba las cejas del participante como si fuera sus ojos realmente y ya que la ceja es un lineal, el mismo algoritmo detecta que los ojos estaban cerrados.

5.5.4 Situación 4: Con oclusión y luminosidad baja.

Para el cuarto caso "Con oclusión lentes y luminosidad baja" se realizaron 8 pruebas para las dos metodologías propuestas, con participantes con anteojos y con una luminosidad baja dando como resultado:

- Método: *Regression Based Facial Landmark Detection.*

Tabla 5.15

Resultados - Situación 4: Con oclusión y luminosidad baja - Método: *Regression Based Facial Landmark Detection.*

Intento	Número de *Frames*	Número de *Frames* con Fatiga		Número de *Frames* sin Fatiga		AAC	DR
		NTP	NFN	NTN	NFP		
1	1000	218	0	617	165	83,50%	100%
2	1000	153	29	675	143	82,80%	84%
3	1000	231	0	698	71	92,90%	100%
4	1000	366	21	492	121	85,80%	95%
5	1000	60	5	853	82	91,30%	92%
6	1000	207	35	568	190	95,68%	86%
7	1000	302	0	627	71	92,90%	100%
8	1000	151	0	727	122	87,80%	100%
Total	8000	1688	90	5257	775	88,92%	95%

Fuente: Elaboración propia.

- Comentarios:
  - Para la cuarta situación, se obtuvo una precisión media de 88.92% y una tasa de detección de 95% en promedio. Bajo estas condiciones de pruebas, el algoritmo tiene un similar porcentaje a la Situación 1 y 2, a pesar de la condición de luminosidad y oclusión, superando al algoritmo de Viola-Jones.

- Hubo casos donde al momento de bajar el rostro, el sistema detectaba como región de interés los parpados; haciendo que, en algunos casos, el algoritmo no detecte la somnolencia en los participantes.
- A pesar de que en los videos son grabados con un cierto grado de rotación del rostro, el algoritmo supera este obstáculo, detectando los ojos de manera satisfactoria. Sin embargo, si el rostro gira 90° a los costados, o bajamos demasiado el rostro, este perderá el enfoque haciendo que el algoritmo no reconozca la zona de interés.

- Método: Viola-Jones.

Tabla 5.16

Resultados - Situación 4: Con oclusión y luminosidad baja - Método: Viola-Jones.

Intento	Número de *Frames*	Número de *Frames* con Fatiga		Número de *Frames* sin Fatiga		AAC	DR
		NTP	NFN	NTN	NFP		
1	1000	129	229	529	113	65,80%	36%
2	1000	120	149	648	83	76,80%	45%
3	1000	95	176	601	128	69,60%	35%
4	1000	105	252	622	21	72,70%	29%
5	1000	68	103	596	233	66,40%	40%
6	1000	334	215	333	118	66,70%	61%
7	1000	121	361	208	310	32,90%	25%
8	1000	185	99	597	119	78,20%	65%
Total	8000	1157	1584	4134	1125	66,14%	42%

Fuente: Elaboración propia.

- Comentarios:
  - Para la cuarta situación "Oclusión con Luminosidad Baja" se observa una precisión media de 66.14% que es uniforme entre casi todas las pruebas realizadas en el experimento con nuestra propia base de datos.
  - Con respecto al experimento anterior de la Base de datos NTHU-DDD, esta vez en todos los videos siempre se detectaron los ojos

y la boca del participante en condiciones normales y por eso se pudieron obtener pruebas más uniformes y no se descartaron algunos videos.

- La tasa de detección disminuye en un 5 % frente a la situación anterior lo cual nos puede indicar que se tuvo un poco de problemas con los anteojos en esta situación, pero nada fuera del rango de aceptación.

5.6 Resultados.

5.6.1 Resultados con la base de datos NTHU-DDD.

Finalmente, contrastando los resultados obtenidos de las dos metodologías, en base a los videos de la base de datos "NTHU-DDD", en la Tabla 5.17 se muestran los resultados.

Tabla 5.17

Resultados de AAC finales de cada situación.

	Metodología	
	Regression Based Facial Landmark Detection	Viola-Jones
Situación 1	88.65%	59.38%
Situación 2	74.88%	81.70%
Situación 3	85.25%	73.30%
Situación 4	73.88%	75.90%

Fuente: Elaboración propia.

En el tema de la precisión media (AAC) se puede apreciar que la metodología *Regression Based Facial Landmark Detection* supera mínimamente a la metodología Viola-Jones, en algunas situaciones, debido a la rotación de rostros que presentan los participantes en los videos.

Una de las ventajas que se pudo observar fue que el algoritmo de Viola-Jones es que se adecuaba mejor en las situaciones con luminosidad baja que el algoritmo de *Regression Based Facial Landmark Detection,* por ello en las situaciones 2 y 4 tuvo un mayor porcentaje que la otra metodología.

Una ventaja que se pudo observar del algoritmo *Regression Based Facial Landmark Detection*, es que no tenía problemas en cuanto a la rotación del rostro, ya

que en la mayoría de *frames* se detectó la zona de interés. Sin embargo, la metodología de Viola-Jones, se vieron dificultades con respecto a este ítem por la rotación de rostro.

Tabla 5.18

Resultados de DR finales de cada situación.

	Metodología	
	Regression Based Facial Landmark Detection	Viola-Jones
Situación 1	100.00%	22.30%
Situación 2	67.71%	49.57%
Situación 3	89.63%	33.08%
Situación 4	52.04%	43.56%

Fuente: Elaboración propia.

Al igual que la precisión media, la tasa de detección de *Regression Based Facial Landmark Detection* supera a Viola-Jones en este caso por un porcentaje alto en comparación a la del AAC. Esto sucedió por el mismo motivo, la rotación del rostro, Viola-Jones detectaba un número muy alto de "Falsos Negativos (NFN)" en todos los videos, en cambio el *Facial Landmark Detection* no, ver Tabla 5.18.

Se puede ver que ninguna de las dos metodologías tuvo un porcentaje de detección alta en la última situación probada, las dos se encuentran aproximadamente por el 50% de detección, por ejemplo, en una situación real, un conductor se quedara dormido, hay una probabilidad del 50% que no sea detectado, y por ende no sea alarmado que esté quedándose dormido, lo que podría ocasionar un accidente.

5.6.2 Resultados con nuestra propia base de datos.

Finalmente, contrastando los resultados obtenidos de las dos metodologías, en base a los videos de nuestra base de datos se muestran los siguientes resultados, ver Tabla 5.19.

Tabla 5.19

Resultados de AAC finales de cada situación con Base de datos Propia

	Metodología	
	Regression Based Facial Landmark Detection	Viola-Jones
Situación 1	87,03%	60,35%
Situación 2	86,23%	67,10%
Situación 3	83,56%	68,24%
Situación 4	88,92%	66,14%

Fuente: Elaboración propia.

En el tema de la precisión media (AAC) se puede apreciar que la metodología *Regression Based Facial Landmark Detection* supera en un 20% a la metodología Viola-Jones, en todas las situaciones.

La metodología Viola-Jones al momento de procesar los videos daba como resultado vario *frames* como "Falsos Negativos" por tal motivo se da la superioridad de *Regression Based Landmark,* ya que este en la mayoría de veces detectó la somnolencia cuando si había.

Cabe resaltar que frente al experimento de la base de datos NTHU-DDD, Viola-Jones no tuvo problemas al detectar los ojos del participante ya que en esta ocasión no se encontró una rotación del rostro de los mismos y la cámara se encontraba en una posición frontal al rostro del participante.

Una ventaja que se pudo observar del algoritmo *Regression Based Facial Landmark Detection*, es que no tenía problemas en cuanto a la rotación del rostro, ya que en la mayoría de *frames* se detectó la zona de interés que en este caso son los ojos.

Tabla 5.20

Resultados de DR finales de cada situación con Base de datos Propia

	Metodología	
	Regression Based Facial Landmark Detection	Viola-Jones
Situación 1	95%	53,45%
Situación 2	95%	47%
Situación 3	79%	47%
Situación 4	95%	42%

Fuente: Elaboración propia.

Al igual que la precisión media (AAC), la tasa de detección de *Regression Based Facial Landmark Detection* supera a Viola-Jones en este caso por un porcentaje alto en comparación a la del AAC al igual que en el experimento con la base de datos NTHU-DDD. Esto sucedió porque Viola-Jones detectaba un número muy alto de "Falsos Negativos (NFN)" en todos los videos al igual que el otro experimento, en cambio el *Facial Landmark Detection* no, ver Tabla 5.20.

Hubo momentos en que los algoritmos no detectaban ciertas facciones del rostro del participante correctamente, por ejemplo, cuando este bajaba mucho la cabeza al momento de simular la somnolencia o cuando giraba su cabeza más de 60 grados hacia sus laterales, no detectaba los ojos correctamente. Por ello, en la metodología de *Regression Based Facial Landmark Detection* tuvo un porcentaje de detección alta en todas las situaciones, excepto en la tercera, debido a que se detectaba la zona del parpado inferior/superior en vez de los ojos, por el uso de los lentes, provocando que, si la persona estaba fatigada, no detectara el algoritmo, ya que tomaba la apertura del parpado.

En cuanto a la metodología de Viola-Jones, además de detectar los ojos, también detecta la boca y su bostezo. Cabe resaltar que en muchos casos esta detección causaba más problemas que aciertos, cuando un participante bosteza, la mayoría de las veces, mueve la cabeza hacia arriba dificultando la detección de la boca y en su totalidad la de su rostro

Para trabajos futuros se podría incluir que el algoritmo de Viola-Jones sea invariante a la rotación del rostro, porque, en este momento, si el participante mueve el rostro en más de 45 grados hacia uno de sus lados laterales, el algoritmo no detecta sus ojos correctamente. Al incluir esta mejora, aumentará la tasa de precisión media, al igual que la tasa de detección, en cualquier escenario, en especial en un escenario más realista, donde el participante no esté mirando a la cámara todo el momento de la grabación.

Además, se podrá hacer una prueba en un escenario real, el cual podría ser que un participante esté en un auto conduciendo por las calles para ver la efectividad resultante durante el día y la noche.

CONCLUSIONES

Para realizar nuestra comparación de metodologías usamos una base de datos NTHU-DDD proporcionada por la universidad *National Tsing Hua*, donde encontramos videos de personas de etnias distintas, simulando el estado somnoliento, en 4 diferentes situaciones:

1. Con luminosidad alta y sin oclusión.
2. Con luminosidad baja y sin oclusión.
3. Con luminosidad alta y con oclusión.
4. Con luminosidad baja y con oclusión.

Con respecto a la precisión media (ACC), el algoritmo de *Regression Based Facial Landmark Detection* obtuvo, en las diferentes situaciones 88.65%, 74.88%, 85.25% y 73,88% respectivamente; en comparación con el algoritmo Viola-Jones que obtuvo un porcentaje de 59.38%, 81.7'%, 73.30% y 75.90% respectivamente. Estos resultados se dieron porque el algoritmo de Viola-Jones, en algunas situaciones, no detectaba correctamente los *frames*, debido a la rotación del rostro que el participante presentaba. Sin embargo, este mismo algoritmo respondía mejor a las situaciones de luminosidad baja. Con respecto a la tasa de detección de somnolencia (DR), el algoritmo de *Regression Based Facial Landmark Detection* obtuvo, en las diferentes situaciones, un 100%, 67.71%, 89.63% y 52.04% respectivamente; en comparación con el algoritmo Viola-Jones que obtuvo un porcentaje de 22.30%, 49.57%, 33.08% y 43.56% respectivamente. Estos resultados se dieron por el mismo motivo mencionado anteriormente, la rotación del rostro fue un problema para el algoritmo de Viola-Jones captando un número alto de falsos negativos.

Para profundizar mejor los resultados, nos planteamos realizar nuestra propia base de datos y analizar cómo se comportan ambos algoritmos frente a rostros latinos. Se grabaron 32 videos con ayuda de alumnos de la universidad. Para poder comparar los resultados obtenidos de nuestra base de datos frente a la base de datos NTHU-DDD, se plantearon las mismas situaciones definidas inicialmente:

1. Con luminosidad alta y sin oclusión.
2. Con luminosidad baja y sin oclusión.
3. Con luminosidad alta y con oclusión.
4. Con luminosidad baja y con oclusión.

Con respecto a la precisión media (ACC), el algoritmo de *Regression Based Facial Landmark Detection* obtuvo, en las diferentes situaciones, un 87.03%, 86.23%, 83.56% y 88,92% respectivamente; en comparación con el algoritmo Viola Jones que obtuvo un porcentaje de 60.35%, 67.10'%, 68.24% y 66.14% respectivamente. Estos resultados se dieron porque el algoritmo de Viola Jones al procesar los videos captaba varios *frames* como falsos negativos, a pesar de que el algoritmo no tuviera el problema de rotación de rostro y se detectaban los ojos de los participantes. Con respecto a la tasa de detección de somnolencia (DR), el algoritmo de *Regression Based Facial Landmark Detection* obtuvo, en las diferentes situaciones, un 95%, 95%,79% y 95% respectivamente; en comparación con el algoritmo Viola Jones que obtuvo un porcentaje de 53.45%, 47%, 47% y 42% respectivamente. Estos resultados se dieron por la misma razón mencionada anteriormente, Viola Jones al procesar los videos captaba varios *frames* como falsos negativos, a pesar de que el algoritmo no tuviera el problema de rotación de rostro y se detectaban los ojos de los participantes.

En base a las pruebas realizadas, para el caso de la metodología Viola-Jones, la inclusión de la detección del bostezo para detectar la somnolencia, en vez de mejorar la efectividad, empeoraba la precisión de la detección de la somnolencia, en cambio la metodología *Regression Based Facial Landmark Detection* al solo concentrarse en la apertura de los ojos, obtuvo un mejor resultado que Viola-Jones, como se muestra en los porcentajes. Finalmente, de acuerdo con los resultados obtenidos en la mayoría de los escenarios, la metodología *Regression Based Facial Landmark Detection* detecta mejor la somnolencia que Viola-Jones superando así las condiciones variables de oclusión, luminosidad y rotación del rostro.

En base a los trabajos realizados en este campo de investigación, se puede apreciar el aporte en el tema de validación, con una cantidad mayor de *frames*, que a su vez refleja una mayor cantidad de videos. Además, se armó una base de datos propia con el objetivo de poder analizar cómo actúan los algoritmos frente a rostros latinos, en específico rostros peruanos.

Por último, para la adaptación de Viola-Jones, para que detecte la somnolencia, se usó de base el código fuente proporcionado en el artículo *Facial Features Monitoring for Real Time Drowsiness Detection* (B.N, 2016) se tuvo que modificar en algunos aspectos, por ejemplo, para que aceptara videos de input en vez de que utilice una cámara en tiempo real, se agregó un contador de *frames* para automatizar la

validación y ver en qué *frame* del video se perciben los distintos escenarios de la matriz de confusión. En el caso del algoritmo de *Regression based LandMark Detection,* se usó de base el código fuente proporcionado por el autor Adrian Rosebrock y se modificaron ciertos parámetros, por ejemplo para que aceptara de input videos en vez de que utilice la cámara en tiempo real, se agregó un contador de *frames* para automatizar la validación y se agregó al código la fórmula de la apertura de los ojos proporcionado en el artículo *Real-Time Eye Blink Detection using Facial Landmarks* (Soukupova y Cech, 2016).

TRABAJOS FUTUROS

Para trabajos futuros se podría incluir que el algoritmo de Viola-Jones sea invariante a la rotación del rostro porque en este momento si el participante mueve el rostro en más de 45 grados hacia uno de sus lados laterales el algoritmo no detecta sus ojos correctamente. Al incluir esta mejora, aumentará la tasa de precisión media al igual que la tasa de detección en cualquier escenario, en especial un escenario más realista, donde el participante no esté mirando a la cámara todo el momento de la grabación.

Además, se podría hacer una prueba en un escenario real, el cual podría ser que un participante esté un auto conduciendo por las calles de la ciudad para ver la efectividad resultante durante el día y durante la noche.

Adicionalmente, se recomienda que el programa pueda aceptar otros formatos de videos, ya que actualmente las dos metodologías, Viola-Jones como *Regression Based Facial Landmark Detection* solo aceptan videos en formato AVI.

Finalmente, de acuerdo a las investigaciones de Vicente, Laguna, Bartra, & Bailón, sobre el ritmo cardiaco y el sueño, se podría agregar un sensor de ritmo cardiaco con el fin de reducir la cantidad de falsos negativos en los resultados, logrando una mayor precisión.

GLOSARIO DE TÉRMINOS

- *Principal Component Analysis* (PCA): Es un método matemático que transforma un número de variables correlacionadas en un menor número de variables no correlacionadas llamadas "componentes principales".
- Electroencefalograma (EEG): Es un examen que sirve para medir la actividad eléctrica del cerebro mediante el registro de las ondas cerebrales.
- Electrocardiograma (EKG): Es un examen que registra la actividad eléctrica del corazón, ya que este emite señales eléctricas que empieza desde la aurícula derecha y se expande por vías específicas.
- *Active Shape Model* (ASM): Son modelos estadísticos de la forma de un objeto que se deforman iterativamente para ajustarse a un ejemplo del objeto en una nueva imagen. Este algoritmo tiene como objetivo hacer coincidir el modelo con una nueva imagen.
- Imágenes *Self-quotient*: Es una representación de inspiración biológica que se ha propuesto como una característica invariante de la iluminación para el reconocimiento facial automático. Debido a la falta de fuertes supuestos específicos de dominio que subyacen a esta representación, se puede extraer fácilmente de imágenes en bruto, independientemente de la postura de las personas, la expresión facial, etc. Lo que hace que la imagen de auto-cociente sea además atractiva es que puede calcularse rápidamente y en una forma cerrada utilizando simples operaciones de imagen de bajo nivel.
- Redes neuronales convolucionales: Considerado uno de los tipos más populares de redes neuronales profundas. Una CNN convoluciona las características aprendidas con los datos de entrada haciendo uso de capas convolucionales 2D, lo cual hace que este diseño resulte adecuado para el procesamiento de imágenes. Las CNN eliminan la necesidad de extraer características manualmente, ya que las extrae directamente de las imágenes. Las características relevantes no se entrenan previamente; sino se aprenden mientras la red se entrena con una colección de imágenes. Esta extracción de características automatizada hace que los modelos de Deep Learning sean precisos para temas de visión artificial, tales como la clasificación de objetos.

- La fotopletismografía: Técnica donde se utiliza un haz de luz para determinar el volumen que ocupa un órgano. El fotopletismógrafo es un instrumento utilizado para la adquisición de signos vitales de un paciente. Con este se monitorea la presión sanguínea; su funcionamiento se basa en la absorción diferencial de la luz reflejada desde los capilares del dedo hacia el fototransistor.
- Algoritmo genético (*Genetic algorithm* GA): Los Algoritmos Genéticos son métodos adaptativos que pueden usarse para resolver problemas de búsqueda y optimización. Están basados en el proceso genético de los organismos vivos. Estos son capaces de ir creando soluciones para problemas del mundo real. La evolución de dichas soluciones hacia valores óptimos del problema depende en buena medida de una adecuada codificación de las mismas.
- *Electrooculogram*(EOG): Es un método que nos permite medir las variaciones eléctricas al detectar el movimiento ocular, basándose en el registro de la diferencia de potencial entre la córnea y la retina.
- Viola-Jones *Object Detection Framework:* Es un marco de trabajo que utiliza una serie de diferentes algoritmos e ideas para poder obtener una robusta y rápida detección visual.
- *Adaboost:* Algoritmo para el uso de detección de objetos que utiliza la clasificación en "cascada" y una representación nueva de la imagen denominada "imagen integral".
- *Haar Features:* Son usados para detectar la presencia de alguna característica en una imagen.
- Imagen integral*:* Es un método para la suma de pixeles dentro de una región en concreto que mejora la *performance.*
- *Support Vector Machine (SVM):* Es un conjunto de algoritmos de aprendizaje supervisado. Estos están relacionados con problemas de clasificación y regresión.
- *Histogram of oriented gradient (HOG):* Es un descriptor característico utilizado en el procesamiento de imágenes y la visión por ordenador para poder detectar los objetos.

- *Facial LandMarks:* Subconjunto de puntos característicos en el rostro, los cuales tienen un impacto en tareas como reconocimiento de rostros, detección de mirada, seguimiento facial, reconocimiento de expresiones y gestos, etc.
- *Eye Aspect Ratio (*EAR*):* Es la relación entre el ancho y la altura de las distancias entre los puntos verticales del ojo y el punto horizontal para poder medir la apertura del ojo.
- PERCLOS: Es el índice más utilizado en la construcción de sistemas donde se discrimina si los ojos de la persona se encuentran abiertos o cerrados para determinar la somnolencia en conductores.
- RGB*:* Denominado a la composición de color en términos de la intensidad de los colores primarios de la luz.
- Precisión Media *(*AAC): Es la suma de los verdaderos positivos con los verdaderos negativos entre todos los frames.
- Tasa de detección *(*DR*):* Son los verdaderos positivos entre los frames que ha detectado que esta somnoliento.
- Matriz de confusión*:* Herramienta que permite la visualización del desempeño de un algoritmo que se emplea en aprendizaje supervisado.

REFERENCIAS

Aidman, E., Chadunow, C., Johnson, K., y Reece, J. (2015). Real-time driver drowsiness feedback improves driver alertness and self-reported driving performance. *Accident Analysis & Prevention, 81*, 08-13.

Akhtar, Z., y Rattani, A. (2017). A Face in any Form: New Challenges and Opportunities for Face Recognition Technology. IEEE, 80 - 90. doi:10.1109/MC.2017.119

Alshamsi, H., y Meng, H. (2016). Real time facial expression recognition app development on mobile phones. *12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD).*

B.N, M. (29 de Marzo de 2017). mathworks. Obtenido de mathworks: https://la.mathworks.com/matlabcentral/fileexchange/55152-drowsiness-detection-using-a-binary-svm-classifier.

Bledsoe, W. W. (1964). Facial recognition project report. *Panoramic research inc*.

Bruner, J. S., y Tagiuri, R. (1954). The perception of people. *Harvard univ cambridge ma lab of social relations.*

Cen, K. (2016). Study of Viola-Jones Real Time Face Detector. Web Stanford, 5.

Cheng, B., Zhang, W., y Lin, Y. (2012). Driver drowsiness recognition based on computer vision technology. *Tsinghua Science and Technology, 17*, 354-362.

Chua, E. C.-P., Tan, W.-Q., Yeo, S.-C., Lau, P., Lee, I., Mien, I. H.,. . . Gooley, J. J. (2012*). Heart Rate Variability Can Be Used to Estimate Sleepiness-related Decrements in Psychomotor Vigilance during Total Sleep Deprivation*. Sleep. doi:http://dx.doi.org/10.5665/sleep.1688

Dwivedi, K., Biswaranjan, K., y Sethi, A. (2014). Drowsy driver detection using representation learning. *Advance Computing Conference (IACC), IEEE International*.

Flores Calero, M. J. (2009). Sistema avanzado de asistencia a la conducción mediante visión por computador para la detección de la somnolencia (tesis doctoral, Madrid).

García, J., y Crespín, L. (2014). *Sistema detector de somnolencia en secuencias de video de conductores manejando usando visión computacional* (tesis de pregrado). Universidad nacional de Trujillo.

Hamada, T., Ito, T., Adachi, K., Nakano, T., y Yamamoto, S. (2003). Detecting method for drivers' drowsiness applicable to individual features. *IEEE*. doi:10.1109/ITSC.2003.1252715

Harb, H. M, Desuky A., Mohammed A., y Jennane R. (2017) Histogram of Oriented Gradients and Texture Features for Bone Texture Characterization. International Journal of Computer Applications, 165, 0975 – 8887.

Huang, P., Gao, G., Qian, C., Yang, G., y Yang, Z. (2017). Fuzzy Linear Regression Discriminant Projection for Face Recognition. *IEEE Access, 5*, 4340-4349.

Hwang, K.-A., y Yang, C.-H. (2009). Attentiveness assessment in learning based on fuzzy logic analysis. *Expert Systems with Applications, 36*, 6261-6265.

King, D. E. (2009). Dlib-ml: A Machine Learning Toolkit. *The Journal of Machine Learning Research*, 1755-1758.

Lee, B.-G., Jung, S.-J., y Chung, W.-Y. (2011). Real-time physiological and vision monitoring of vehicle driver for non-intrusive drowsiness detection. *IET Communications, 5*, 2461-2469.

Mallick, S. (19 de Marzo de 2018). Learn OpenCV. Recuperado de https://www.learnopencv.com/facemark-facial-landmark-detection-using-opencv/

Marin, C. (2017). Tutor de Programación. Recuperado de http://acodigo.blogspot.com/2017/11/face-landmarks-detector-con-dlib-y.html

Marks, P. (2013). Drowsiness detector wakes sleepy miners if they're dozing off. *New Scientist, 218*, 22.

MTC - Secretaría Técnica del Consejo Nacional de Seguridad Vial. (2018). MTC. Recuperado de https://www.mtc.gob.pe/cnsv/documentos/heridosAccidenteTransito_2006-2017.pdf

Mundial, G. B. (2017). *Población urbana (% del total)*. Recuperado de http://datos.bancomundial.org/indicator/SP.URB.TOTL.IN.ZS?end=2015ystart=1985&view=chart

Open CV team. (2018). Open CV. Recuperado de https://opencv.org/

Pereyra, P. A. (s.f.). Reconocimiento Facial (tesis de pregrado). *Universidad de Buenos Aires*.

Pupale, R. (16 de Junio de 2018). *Support Vector Machines(SVM)—An Overview* [versión PDF]. Recuperado de https://towardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989

Riveros, O. A. (2008). Análisis y Detección de Características Faciales Usando (tesis de postgrado). *Universidad de Chile.*

Rosebrock, A. (24 de Abril de 2017). pyimagesearch. Recuperado de https://www.pyimagesearch.com/2017/04/24/eye-blink-detection-opencv-python-dlib/

Soukupova, T., y Cech, J. (2016). Real-Time Eye Blink Detection using Facial Landmarks. *21st Computer Vision Winter Workshop*, 8.

Triantafyllidou, D., y Tefas, A. (2016). Face detection based on deep convolutional neural networks exploiting incremental facial part learning. *23rd International Conference on Pattern Recognition (ICPR)*.

Vicente, J., Laguna, P., Bartra, A., y Bailón, R. (2016). *Drowsiness detection using heart rate variability*. Med Biol Eng Comput, 927–937. DOI 10.1007/s11517-015-1448-7

Viola, P., y Jones, m. J. (2004). Robust Real-Time Face Detection. *International Journal of Computer Vision 57(2)*, 137–154.

Vojt, B. J. (2016). Deep neural networks and their (tesis de postgrado). *Charles University in Prague*.

Wang, X., y Xu, C. (2015). Driver drowsiness detection based on non-intrusive metrics considering individual specifics. *ScienceDirect*. https://doi.org/10.1016/j.aap.2015.09.002