

**SISTEMA DE DETECCIÓN Y RECONOCIMIENTO DE
ROSTRO PARA IMÁGENES QUE CONTIENEN
MÚLTIPLES PERSONAS**

Laura Natalia Valdez Galindo

**UNIVERSIDAD SANTO TOMÁS
FACULTAD DE INGENIERÍA ELECTRÓNICA
BOGOTÁ D.C.**

2018

**SISTEMA DE DETECCIÓN Y RECONOCIMIENTO DE
ROSTRO PARA IMÁGENES QUE CONTIENEN
MÚLTIPLES PERSONAS**

Laura Natalia Valdez Galindo

Trabajo de grado

Director:

Fabian Eduardo Perez Gordillo

Codirector:

Andrea Katherín Pérez Hernández

**UNIVERSIDAD SANTO TOMÁS
FACULTAD DE INGENIERÍA ELECTRÓNICA
BOGOTÁ D.C.**

2018

Agradecimientos

Me gustaría que estas líneas sirvieran para expresar mis más profundos y sinceros agradecimientos a todas aquellas personas que con su ayuda directa o indirectamente ha colaborado en la realización del presente proyecto.

A mi madre Maria Claudia, por su colaboración e inspiración en todo momento. A mis abuelas Dora y Alba y mi tía Marcela quienes me ayudaron en todo aspecto posible.

Debo agradecer de manera especial y sincera al los profesores Andrea Katerin Pérez Hernández y Fabian Eduardo Perez Gordillo por su apoyo y confianza en mi proyecto y su capacidad de guiar mis objetivos ha sido invaluable no solo en esta tesis si no a lo largo de mi carrera.

Agradezco a todos los integrantes del laboratorio de Robotica, por compartir sus conocimientos experiencias y apoyo en todos los proyectos realizados. Y a mis amigos y compañeros que coincidimos en espacio y tiempo, de corazón agradezco su amistad y apoyo.

CONTENIDO

	pág.
INTRODUCCIÓN	13
1. PROBLEMA	15
2. JUSTIFICACIÓN	17
3. OBJETIVOS	18
3.1. OBJETIVO GENERAL	18
3.2. OBJETIVOS ESPECÍFICOS	18
4. ANTECEDENTES	19
5. MARCO CONCEPTUAL	23
5.1. Detección Facial	23
5.1.1. Métodos basados en reglas	24
5.1.2. Métodos estadísticos	35
5.2. Reconocimiento Facial	42
5.2.1. Análisis de componentes principales (PCA)	43
5.2.2. Análisis lineal discriminante (LDA)	46
5.2.3. Eigenfaces	48
5.2.4. Fisherfaces	49

5.2.5.	Local Binary Pattern Histogram	50
5.3.	Aprendizaje de Maquina	52
5.3.1.	Redes Nauronales	53
5.3.2.	Aprendizaje supervisado para Redes Neuronales	55
5.3.3.	Aprendizaje no supervisado para Redes Neuronales	57
6.	DISEÑO METODOLÓGICO	58
6.1.	Software de procesamiento de información	60
6.2.	Técnicas de análisis de datos	62
7.	DISEÑO DE LA SOLUCIÓN	64
7.1.	Selección de la Base de Datos	64
7.2.	Selección del algoritmo de Detección	67
7.3.	Selección del algoritmo de Reconocimiento	70
7.4.	Implementación del sistema de detección y reconocimiento de rostro	72
7.4.1.	Sistema de Detección y Localización	72
7.4.2.	Modelo de detección y localización	74
7.4.3.	Sistema de reconocimiento e identificación	79
7.4.4.	Modelo de reconocimiento e identificación	81
8.	RESULTADOS	86
8.1.	Sistema de detección y localización	86
8.1.1.	Algoritmo de entrenamiento de clasificadores	86
8.1.2.	Algoritmo de Detección y localización	88

8.2. Sistema de Reconocimiento e Identificación	98
8.2.1. Rotación del Rostro	99
8.2.2. Cambios de iluminación	101
8.2.3. Rotación y cambios de iluminación al rostro	103
8.3. Sistema de detección y reconocimiento	104
9. CONCLUSIONES Y TRABAJO FUTURO	111
BIBLIOGRAFÍA	114

Indice de Figuras

pág.

1.	Ejemplos de los problemas al detectar un rostro (tomadas de [25]) . . .	24
2.	Resultado del Skin color a) imagen original b) imagen con el modelo de skin color (tomada de [2]).	25
3.	Representación espacial del modelo RGB (tomada de [26]).	26
4.	Representación espacial del modelo HSV (tomada de [26]).	27
5.	histogramas bimodal representando una imagen con 2 regiones. (Tomada de [26])	30
6.	Crecimiento de regiones: a) pixel semilla, b) primera iteración c) iteración final.	33
7.	Detector de bordes de un rostro por el algoritmo de Canny (tomada de [2]).	34
8.	Plantilla del rostro humano a) imagen original b) imagen de la plantilla del rostro (tomada de [47]).	35
9.	Características tipo haar de dos rectángulos A) y B). la Figura C) muestra característica de 3 rectángulos y D) una característica de cuatro rectángulos (tomada de [43]).	37
10.	Proceso de obtención de patrones binarios locales.	39
11.	Convertir imagen a imagen Integral.(Tomada de [12])	39
12.	Calculo de un rectángulo a partir de una imagen integral. Para calcular el valor del rectángulo D podría realizarse con cuatro referencias de arreglos. Donde el valor de la imagen integral en la posición 1 es la suma de los pixeles en el rectángulo A. El valor en la posición 2 es A+B, en la posición 3 es A+C y en la posición 4 es A+B+C+D. Para calcular el valor del rectángulo D es $4+1-(2+3)$.(tomada de [36])	40

13.	Adaboost clasificador débil. (Tomada de [38])	41
14.	Descripción esquemática de la detección en cascada (tomada de [36]). .	41
15.	Imágenes de la misma persona pero con cambios de iluminación (tomada de [49]).	42
16.	Ejemplo de reducción de dimensional aplicada por PCA (tomada de [10])	44
17.	Análisis de LBP [48].	51
18.	Histograma por cada region de la imagen [48].	52
19.	Capas y proceso de la red neuronal (tomada de [30]).	55
20.	Diagrama del Desarrollo del sistema	59
21.	Sistema de detección y reconocimiento	72
22.	Esquema general del Sistema de detección y localización del rostro . . .	73
23.	Resultados de detección a) detección de rostros que contiene la imagen y b) lista de coordenadas de los rostros detectados	73
24.	Esquema del Sistema de detección y localización del rostro	76
25.	Estructura de clasificador en Cascada	77
26.	Las características Haar usadas para la extracción de las características, una de dos rectángulos que abarca la características de los ojos y una de tres rectángulos que no tiene características relevantes para identificar el rostro	78
27.	Imágenes usadas para el entrenamiento positivas y negativas	79
28.	Estructura de sistema de reconocimiento e identificación	80
29.	Sistema de reconocimiento a) imagen de entrada b) numero se predijo c) nombre de la persona	80
30.	Esquema de la estructura del sistema de reconocimiento e identificación	82
31.	Tipos de imágenes a entrenar para el sistema de reconocimiento	83

32.	Reconstrucción de la imagen de entrada con Eigenface	85
33.	Represnetacion de las 6 primeras Eigenfaces	85
34.	Imagen con detección y localización de rostro	89
35.	Recorrido en la detección a) primer recorrido b) segundo recorrido c) tercer recorrido d) suma de todos los recorridos e) resultado de toda la detección	90
36.	Variación de componentes para el algoritmo Eigenfaces	99
37.	Comparación del desempeño de los algoritmos en función al cambio de dimension sobre las imagenes de la base de datos Head Posepor	100
38.	Comparación del desempeño de los algoritmos con imágenes de con variaciones de iluminación	102
39.	Comparación del desempeño de los algoritmos con 115 personas	103
40.	Tipo de imagen modificada con diferentes objetos diferentes y 4 personas para detectar y reconocer	105
41.	Resultados del sistema de detección y reconocimiento	106
42.	Resultados del sistema de detección y reconocimiento	107
43.	la primera imagen es el resultado del sistema de deteccion y reconocimiento sin las modificaciones en la detección y la segunda es el resultado con las modificación realizadas	109
44.	la primera imagen es el resultado del sistema de deteccion y reconocimiento sin las modificaciones en la detección y la segunda es el resultado con las modificación realizadas	110

Indice de Tablas

	pág.
1. Características de Bases De Datos para la detección de rostros	65
2. Bases De Datos para el reconocimiento de rostros	66
3. Características de Técnicas de Detección	69
4. Características de Técnicas de Reconocimiento	71
5. Parámetros de evaluación para el sistema de detección y localizacion . .	74
6. Tiempos de entrenamiento	88
7. Imágenes a evaluar	88
8. Evaluación al sistema de detección entrenamiento con características Haar, tamaño de imagen 38 por 38 pixeles y 20 etapas del clasificador	91
9. Gráficas de Precisión y recall de entrenamiento con características Haar	93
10. Mejores resultados de los sistemas de detección	95
11. Resultados de la evaluación con la imágenes que contiene 2 personas a los sistemas seleccionados	95
12. Diagramas PR de los sistemas seleccionados con cada grupo de imágenes	96
13. Tiempos promedios que demora los algoritmos en reconocer los rostros con variación de posición	101
14. Tiempos promedios que demora los algoritmos en reconocer los rostros con variación de iluminación	102
15. Resultado del sistema de detección y reconocimiento con las 12 imágenes creadas	107

GLOSARIO

Adaboost: es un algoritmo cuya finalidad es encontrar una hipótesis fuerte a partir de utilizar hipótesis simples y débiles.

Algoritmo de Viola y Jones: es un método de detección de objetos que destaca por su bajo coste computacional, lo que permite que sea empleado en tiempo real.

Características Haar: son imágenes digitales características utilizadas en el reconocimiento de objetos. Deben su nombre a su similitud intuitiva con las ondas de Haar y se usaron en el primer detector de rostros en tiempo real.

Detección facial: encontrar la ubicación arbitraria de un número indefinido de rostros en una imagen o video.

Eigenfaces: es el nombre dado a un conjunto de vectores propios (o eigenvectors) cuando se usan en el problema de visión artificial del reconocimiento de rostros humanos .

Fisherfaces: es un método que se encarga del reconocimiento de rostros, teniendo en cuenta como se refleja la luz y las expresiones faciales.

Local Binary Pattern (LBP): es un operador de textura simple y eficiente que estudia si el nivel de gris de cada píxel supera un determinado umbral y codificando dicha comparación mediante un número binario.

Linear Discriminant Analysis(LDA): es una generalización del discriminante lineal de Fisher, un método utilizado en estadística, reconocimiento de patrones y aprendizaje de máquinas para encontrar una combinación lineal de rasgos que caracterizan o separan dos o más clases de objetos o eventos.

Principal component analysis(PCA): es una técnica utilizada para describir un set de datos en términos de nuevas variables no correlacionadas.

INTRODUCCIÓN

En las últimas décadas, muchas aplicaciones requieren interactuar automáticamente con los usuarios, y la necesidad de que este tipo de aplicaciones tengan la capacidad de obtener la información acerca del usuario es imprescindible. Motivo por lo cual se han realizado investigaciones en el campo de la visión por computador, de forma genérica, se enmarca en ella los sistemas de procesamiento de imágenes, reconocimiento de patrones y percepción artificial, cuyo objeto de estudio es el ser humano. El estudio del rostro humano es una de las ramas más investigadas debido a las cantidad de aplicaciones en las que se puede usar, por ejemplo: el desarrollo de nuevos métodos de interacción hombre-máquina; la codificación y etiquetado de vídeo para videoconferencias; vídeo vigilancia; el reconocimiento automático de personas, entre otras.

En el presente proyecto aborda grandes problemas del procesamiento visual de rostros humanos, con el objetivo de clasificar las características faciales y reconocerlas, usando algoritmos de aprendizaje estadístico, en otras palabras se busca poder detectar y reconocer rostros en una imagen a través de la morfología facial.

Estos dos problemas son estudiados en detalle, y se proponen posibles métodos para detectar rostros y reconocerlos, cabe destacar que para reconocer los rostros, previamente tuvieron que ser localizados en la imagen. Razón por la cual, se implemento primero el sistema de detección y localización de rostro, con el método de Viola y Jones, obteniendo como resultado una tasa de detección mayor al 80% al evaluar 100 imágenes con 277 personas en ellas, es importante resaltar que las pruebas se realizaron a imágenes de la base de datos *Face Detection Data Set and Benchmark* (FDDB) que contenían rostros en diferentes posiciones, cambios de iluminación, con oclusiones en el rostro como lentes, turbantes, bufandas, gorros o cachuchas, entre otros. Con el rostro detectado y localizado se emplea el sistema de reconocimiento e identificación, según las investigaciones realizadas se presenta tres algoritmos que cumplen con los objetivos propuestos, motivo por el cual se emplearon: Eigenfaces Fisherfaces y LBPH, al compararlos se presento que el sistema Eigenfaces, es él adecuado para usarlo en el sistema de detección y reconocimiento, debido a que con las pruebas realizadas presento una tasa de reconocimiento mayor al 70% al evaluarlo con 115 personas de diferentes edades, razas, culturas, cambios de iluminación, poses y expresiones faciales.

El sistema propuesto en este proyecto para la detección y reconocimiento son los méto-

dos de Viola y Jones y Eigenfaces respectivamente. Los experimentos llevados a cabo demuestran que se puede detectar y reconocer con cambios de iluminación y posiciones del rostro y es invariante frente a oclusiones y expresiones faciales. Además el coste computacional al momento de reconocer el rostro es menor a 8 ms, por lo tanto se puede usar para la detección y el reconocimiento en tiempo real.

1. PROBLEMA

En los últimos años se han presentado avances significativos en el área de sistemas automáticos, debido a que con esta área se puede dar soluciones a problemas que actualmente se tienen en el mundo de hoy, como por ejemplo en los campos de industrialización, medicina, seguridad, entre otros. Para el campo de seguridad se han implementado diseños y aplicaciones con diferentes sistemas, a causa de, el aumento a un 80% del nivel de terrorismo en el mundo según los estudios publicados en *Global Terrorism Index* en el 2014 [37], por tal razón varios países invierten en soluciones para reducir este nivel. Estados Unidos es uno de los países que ha empleado diferentes sistemas para resolver esta problemática, por ejemplo; el diseño de un automóvil inteligente para uso policial [17], su estructura contiene diferentes sensores para detectar radiación o escanear patentes, además posee múltiples cámaras con un sistema de reconocimiento facial con la capacidad de buscar e identificar los fugitivos de la ley. Otro ejemplo que se está implementando en Nueva York es reconocer a los terroristas que se encuentren en áreas concurridas con un sistema de reconocimiento facial [8]. Es indispensable explicar que un área concurrida según la Real Academia Española es un lugar que tiene gran afluencia de personas, como las calles, los centros comerciales o los aeropuertos. En este último lugar el Buró Federal de investigaciones (o FBI, de *Federal Bureau of Investigation*) está usando desde el 2014 un sistema de reconocimiento de rostro denominado “*Next Generation Identification System*” (o en sus siglas en inglés NGI), que alberga más de 51 millones de fotografías en una base de datos,[35], como resultado detectó a Mohamed Abrini el tercer terrorista que participo en los atentados contra el aeropuerto de Bruselas haciendo que lo detuvieran y juzgaran por sus actos [5].

Con los avances tecnológicos que se han mencionado, se detectó que un sistema de reconocimiento facial, o también denominado como biometría facial, tiene como finalidad identificar rostros basado en unas características morfológicas del individuo para detectarlo y reconocerlo. Para elaborar la aplicación de biometría facial, primero es necesario detectar el rostro en una imagen o vídeo, usando diferentes métodos, pero los más usados son los métodos de aprendizaje con redes neuronales. Con el rostro detectado se busca reconocerlo utilizando métodos de clasificación o redes neuronales con el fin de identificar las características morfológicas más relevantes del rostro del individuo y así compararlas con una base de datos previamente realizada para identificar a la persona. El sistema de reconocimiento facial anteriormente explicado se ha implementado

en áreas concurridas como una herramienta, que ayuda a la policía o al personal de seguridad a identificar los individuos mas buscados, como por ejemplo, en un centro comercial se encuentra una persona identificada como peligrosa, y el sistema de reconocimiento facial la detecta, éste dará aviso a las autoridades para que realicen el debido procedimiento de captura. Otro ejemplo es, si una personas se pierde en un área que es monitoreada por el sistemas de reconocimiento facial se puede buscar y detectar su ubicación.

Teniendo en cuenta la importancia que tiene las aplicaciones de detección y reconocimiento de rostros en zonas concurridas, se pretende en este proyecto realizar el diseño e implementación de un sistema que pueda detectar y reconocer rostros humanos a través de de imágenes en las que contengan todas variaciones que se pueden presentar al monitorear un área concurrida. Para esa implementación primero se busca dar solución la siguiente pregunta: ¿Que sistema de detección y reconocimiento es el mas apropiado para el monitoreo de un área concurrida?

2. JUSTIFICACIÓN

El monitoreo en un área concurrida es uno de los aspectos más importantes para la protección de una población civil, ante aquello que pueda perturbar o alterar su integridad física o moral, si se asegura que la zona este libre de riesgo a las personas que la frecuentan, podrán incentivar a otros individuos que asistan a esta área y aumente su producción, por eso muchos lugares buscan constantemente elevar la seguridad con nuevas tecnologías y estrategias de anti riesgos.

Una de esas nuevas tecnologías para el monitoreo de un área concurrida es la detección y reconocimiento de rostros, esté tipo de aplicación se realiza con el fin de identificar a los terroristas o posibles individuos que perturben la seguridad de otras personas, y así el personal de seguridad podrá detenerlo y efectuar el procedimiento necesario antes que cumplan con su objetivo, por consiguiente los algoritmos de visión artificial para este tipo de aplicaciones ayudan disminuir el índice de riesgos que se presenta en áreas concurridas como pueden ser un aeropuerto, centro comercial o parques de diversiones.

Los sistemas de detección y reconocimiento de rostros, es uno de los pilares de la investigación que se usan para la robótica social, como se presenta en una de las ligas de RoboCup. RoboCup es una iniciativa científica internacional con el objetivo de avanzar en el estado del arte de los robots inteligentes, en el cual una de sus ligas se enfoca en otorgar a los robot una interacción activa con el usuario, para eso es necesario una deteccion y reconocimiento de las personas, dicha liga tiene como nombre *the Social Standard Platform League* (SSPL), en la cual realizan competencias con robots, y muestran los avances realizados cada año [44]. Cabe señalar que en cada año muestra nuevas investigaciones y desarrollos de los métodos para que la detección y reconocimiento con el fin de que sea mas rápido, eficiente y confiables que el año anterior. En los antecedentes (sección 4), describe algunas investigaciones realizadas para el sistema de detección y reconocimiento que se usan en la actualidad.

Si se implementara esté tipo de aplicaciones facilitaría la búsqueda de personas peligrosas y aumentaría la seguridad en el área monitoreo, como se muestra en los articulos [35] y [5], constatando la utilidad de esas aplicaciones, al dar aviso a las autoridades de que se identifico y ubico a un fugitivos de la ley en un área que es monitoreada por un sistema de detección y reconocimiento.

3. OBJETIVOS

3.1. OBJETIVO GENERAL

Diseñar e implementar un sistema para la de detección y reconocimiento de rostro en imágenes que contienen múltiples personas.

3.2. OBJETIVOS ESPECÍFICOS

- Estudiar técnicas y algoritmos para detección y reconocimiento de rostros.
- Desarrollar un algoritmo para la detección de rostros en una imagen digital.
- Desarrollar un algoritmo para el reconocimiento de rostros a partir de características faciales.
- Evaluar el sistema propuesto para la detección y reconocimiento de rostros.

4. ANTECEDENTES

La detección y reconocimiento facial es un campo que actualmente se ha investigado para el diseño de diferentes tecnologías, con el fin de adquirir alguna información necesaria del rostro de la persona como por ejemplo el género, la edad aproximada, si tiene gafas o no, el nombre de dicha persona, entre otros. Para realizar este tipo de aplicaciones muchos investigadores han realizado diferentes algoritmos para la detección de rostros y para su reconocimiento. Donde la detección de rostro se enfoca en encontrar en una imagen o vídeo los múltiples rostros que está tiene, despreciando cualquier otro elemento que no lo sea, por ejemplo los árboles, carros, sillas etc. . . . Mientras que el reconocimiento de rostro consiste en identificar a la persona que se encuentra en dicha imagen o vídeo. Es indispensable que para reconocer correctamente el rostro es necesario de una buena detección. Por lo tanto es importante identificar qué tipos de algoritmos y métodos se han implementado en las aplicaciones para la detección de rostros.

Para implementar los sistemas de detección del rostro se identificó que la mayoría de los algoritmos son basados en Adaboost como se presenta en el artículo [20] que realizan un algoritmo multclasificador en cascada llamado AM-CC (por sus siglas en inglés ó an automatically cascade classifier), basado en los métodos Feature Subspace Method y el AdaBoost, éste evita un sobre entrenamiento a comparación del AdaBoost , tiene una tasa de detección de rostros mayor al 10 % con una velocidad superior al AdaBoost. Uno de los problemas que se presentan en este algoritmo base, es la alta tasa de falsos negativos (o tasa de rostros que no han sido detectados) cuando el rostro en la imagen se encuentra con una inclinación , para esto se presenta una posible solución en [49] donde combinan el algoritmo Adaboost con dos clasificadores (clasificador de rostros frontal y clasificador de rostros en perfil), denominado DP-Adaboost, con este método analizaron que la detección de múltiples rostros con diferentes ángulos, tienen un alta tasa de detección correcta y una tasa baja de falsos en comparación al algoritmo base. Otra solución se presenta en [38], con la integración de Integrales Difusas con el algoritmo base, obteniendo una tasa de detección parecida al Adaboost, pero éste posee una ventaja con el clasificador, ya que reduce el número de características a calcular y la complejidad computacional.

En los artículos [13] y [23] proponen la unión del filtro PFMPF (*Probability-based Face Mask Pre-Filtering*) y el método PBHFA (*the Pixel-Based Hierarchical-Feature Ada-*

boosting), el cual las dos etapas proporciona menos tiempo en el proceso de entrenamiento y una reducción en la complejidad computacional en la detección de rostro, esto se debe a que el PFMPF puede filtrar que no es un rostro en una imagen a un 85 % según [13] y los posibles rostros que se tienen en la imagen se filtra con el clasificador PBHFA obteniendo los rostros de la imagen, además con esta unión es posible detectar el rostro con baja resolución [23] y el tiempo de entrenamiento es corto en comparación del algoritmo de Adaboost. Por otra parte un algoritmo en el cual se reduce la tasa de fallos y además trabaja en tiempo real es el de Viola and Jones, este presenta la unión de los rasgos de Haar-like y el clasificador Adaboost en cascada dando así un algoritmo para la detección de rostros, en el artículo [41] se encuentra un análisis en el cual utiliza el algoritmo de Viola and Jones y diferentes algoritmos de escalamiento como Nearest Neighbor, Bilinear, Bicubic, Extended Linear y Piece-wise Extended Linear, comparándolos entre sí, concluyeron que el algoritmo de Piece-wise Extended Linear tiene mejores resultados que los otros métodos de escalamiento, además la unión de este algoritmo y el de Viola and Jones pueden trabajar en tiempo real y reduce la tasa de falsos positivos y negativos. También con el método de Viola and Jones se puede presentar la detección de rostro enfocado en identificar la zona de los ojos la cual es la característica más representativa de cada rostro humano. En el artículo [43] realizan una aplicación en donde al usar el método de Viola and Jones para detectar el rostro en una imagen para así con la normalización geométrica poder localizar los ojos de las personas, otra forma de detectar los ojos de las personas es utilizando landmark [18] que detecten las esquinas de los ojos y con algoritmos de Geometrical Measurement puede ubicar la pupila de las personas y así poder reconocerlo.

Igualmente se han implementado otros tipos de algoritmos para la detección de rostros que no se basan en el clasificador Adaboost, como por ejemplo la detección de rostro por medio de una plantilla que tiene modelado geoméricamente el rostro humano, esta plantilla se evalúa con la imagen a analizar identificando la correlación entre ellas, y así detectar los rostros de dicha imagen. Este método posee diferentes problemas al momento de detectar rostros cuando están inclinados o cuando la imagen tiene baja luminosidad. Una de las posibles soluciones se realiza en el artículo [47] donde hacen una comparación de 2 tipos de plantillas, una tiene las características de todo el rostro y la otra presenta la mitad de los rasgos, el análisis se realizó con varias imágenes faciales en diferentes posiciones concluyendo que la segunda plantilla elimina la problemática al detectar rostros con inclinaciones además tiene un 95 % de tasa de detección.

Una metodología diferente a los mencionados es con la tonalidad del rostro como [2] [46] es decir elimina con una umbralización todo lo que no se encuentre en el matiz facial, con esa región identifican si es un rostro o no, por ejemplo en el artículo [2] realizan una combinación de segmentación con diferentes tipos de modelos de color (RGB, hsv y YYCbCr) y el algoritmo de detección Canny, evaluándolos con diferentes imágenes de rostros inclinados obteniendo la detección de rostro, pero esta combinación no elimina otra zona que tenga la misma tonalidad de la piel como las manos o cuello, para eliminar este problema en el artículo [46] se combina la segmentación y *Local Interated Conditional Model* el cual ayuda a eliminar cualquier elemento que no sea el rostro y así la unión de estos 2 elementos hacen que la detección sea más precisa en comparación al diseño de [2]. Aun así los sistemas de detección de rostros por los métodos de plantilla y tonalidad no son los más confiables y viables como se analizó en [25] donde compara el algoritmo de Adaboost, la tonalidad de la piel y el método geometría facial o plantilla concluyendo que cada uno tiene diferentes problemas pero los dos últimos métodos su tasa de detección es baja en condiciones de poca luminosidad, inclinación del rostro o con un cubrimiento parcial del rostro por lo tanto la mayoría de aplicaciones de detección de rostros se usa el Adaboost .

Con el rostro detectado en una imagen se puede reconocer dicho rostro, realizando los siguientes pasos primero se extraen las características morfológicas faciales para realizar una base de datos de los individuos conocidos y segundo se compara el rostro detectado con todos los rostros de la base de datos para finalmente reconocerlos. Esta metodología usualmente utiliza la técnica del Principal Component Analysis (PCA) [34], dicha técnica brinda la posibilidad de reducir la dimensionalidad de los datos por lo que, investigadores lo implementan con otros algoritmos de clasificación, por ejemplo en el artículo [32] se presenta una combinación de los métodos Eigenfaces, PCA y KNN (EPK) donde el primero extrae las características del rostro, el segundo clasifica estas características y el tercero clasifica las imágenes guardadas (dataset) y así reconocer al individuo, también en este artículo se analiza la combinación los métodos de Eigenfaces, LDA y KNN (ELK) que tiene la misma finalidad que la anterior combinación, lo que las difiere es la tasa de éxito donde el método EPK es alrededor del 97.50% y para el ELK es alrededor 94.55% y su tasa de fallos es de 2.5% y 5.55% respectivamente, para este análisis usaron una base de datos en específico y una medida de similitud, por lo tanto concluyeron que al realizar la unión de estos métodos se puede obtener un mejor resultado que trabajando con cada uno por separado, brindando un

reconocimiento facial adecuado. Con el PCA tambien se pude combinar con el metoso SOM (self- Organizing Maps), que fue analizado en el artículo [3] usado para reconocer a los estudiantes de la universidad de Essex haciendo uso de su base de datos de 7900 imágenes de rostros tomadas a 395 individuos donde el PCA se emplea para encontrar las características del rostro y el SOM para su agrupación dado una mejor clasificación y mayor índice de reconocimiento.

Considerando ahora la combinación del PCA con redes neuronales para la clasificación de las características [24] mostrando una alta precisión con una tasa de reconocimiento del 91 % con baja tasa de error y baja complejidad computacional. En el artículo [4] presentan el PCA con LDA (Linear Discriminant Analysis) en el cual se presenta una tasa de reconocimiento del 95 % demostrando que este sistema tiene menor probabilidad de error en comparación con el PCA que tiene 91 % y el LDA con un 94 % de tasa de reconocimiento. Por lo tanto se puede concluir que al combinar el PCA con otro método se obtiene mejor resultado en el momento de reconocer rostros.

5. MARCO CONCEPTUAL

En esta sección se presenta, las características generales y métodos con los cuales es posible diseñar e implementar un sistema de detección y reconocimiento de rostro para el monitoreo de un área concurrida. En primer lugar se definirá y se proporcionara una breve explicación de algunos métodos que se pueden usar para la elaboración de detectar un rostro en una imagen o vídeo, y en segundo lugar se explicara que es el reconocimiento de rostros y algunos métodos para su implementación.

5.1. Detección Facial

La detección facial o detección de rostros es un proceso computacional el cual tiene como finalidad encontrar la ubicación arbitraria de un número indefinido de rostros en una imagen o video [41]. Esta ha sufrido un vertiginoso avance debido a la amplia investigación e importancia para la visión computacional y patrones de reconocimiento [20][47] [9] dado que la detección de rostros es la base de varias aplicaciones que utilizan el análisis facial entre los que estan; el reconocimiento de rostros, la interacción máquina – hombre en una interfaz, realidad virtual, gestión de archivos de seguridad pública, monitoreo de seguridad en un área concurrida, cajeros automáticos, entre otras [38][41]. Dicho hasta aquí supone que el objetivo de la detección de rostros es determinar si existe uno o varios rostros en una imagen o vídeo otorgando su ubicación espacial.

Al realizar la detección de rostros se puede presentar tres tipos de problemas: el primero definir el rostro para localizarlo, esto es un problema dado que presenta una alta complejidad debido a que el rostro humano tiene varias variaciones como por ejemplo la tonalidad de la piel, las dimensiones del rostro, si tiene el rostro con una bufanda o gorra o gafas. El segundo problema es la ambientación en donde se está evaluando el rostro, dicho de otra manera la imagen puede presentar baja iluminación, fondos de diferentes tonos o borrosa y el tercer problema es espacial, en otras palabras cuando el rostro presenta diferentes rotaciones las cuales se pueden clasificar en dos tipos: rotación vertical (persona mirando hacia arriba o abajo) y rotación horizontal (persona mirando a la izquierda o derecha), para una mejor ilustración de los 3 problemas que se pueden presentar, ver la Figura 1. [25]

En consecuencia de los posibles problemas que se pueden presentar en la detección facial,



Figura 1. Ejemplos de los problemas al detectar un rostro (tomadas de [25])

se pretende encontrar un método en el cual extraiga las características más relevantes del rostro humano, en otras palabras obtener un modelo del rostro robusto y confiable para así detectar los rostros en una imagen. Con esto quiere decir, que si se desarrolla un método que extraiga las características que define un rostro con exactitud, y usar esas características al detectar se podrá tener una alta tasa de detección de rostros en una imagen.

Los métodos que se presentan actualmente para la elaboración de una aplicación de detección de rostros se pueden clasificar en dos grupos los cuales se presentan a continuación; el primero son métodos basados en reglas (en la sección 5.1.1) y el segundo son métodos estadísticos (en la sección 5.1.2) [38].

5.1.1. Métodos basados en reglas

Los métodos basados en reglas son métodos en los se establecen un conocimiento previo de las características principales del rostro, es decir, que basado en diferentes parámetros se sabe si es un rostro o no, por ejemplo la tonalidad de la piel o la simetría del rostro como la distancia de los ojos, la nariz y la boca. [46],[47]. A continuación se presenta algunos métodos basados en reglas.

Skin Color

El modelo de detección de rostros por medio del color de la piel es un enfoque común ya que ayuda evitar una búsqueda exhaustiva del rostro en una imagen o video [2], debido a que este método busca identificar si los pixel de la imagen tienen la tonalidad de la piel humana, en el caso de que un grupo de pixeles tenga esa tonalidad se asume que es el rostro de la persona.

Es necesario recalcar que la piel humana posee el mismo croma, pero se distingue debido a la intensidad de iluminación que esta tiene [25]. Al realizar la correcta selección de los rangos en donde se encuentra la tonalidad de la piel en los espacios de color, se garantizaría que es posible eliminar los elementos que no se encuentren en ese rango como, la ropa el pelo o el fondo de la imagen, dando así el rostro de la persona como se puede ver en la Figura 2.

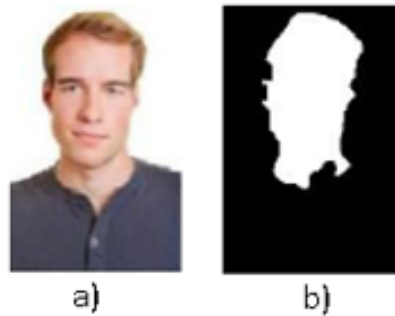


Figura 2. Resultado del Skin color a) imagen original b) imagen con el modelo de skin color (tomada de [2]).

Como se observa en la Figura 2 este método tiende a tener diferentes problemas en la detección del cuello o el pelo de la persona asumiendo que eso es parte del rostro, además si la imagen tiene poca iluminación es mucho más difícil detectar el rostro [25].

Para solucionar los diferentes problemas mencionados y realizar este método, primero hay que escoger que tipo de espacio del color va a tratar la imagen RGB, HSV o YCrCb, hay que aclarar que estos no son los únicos espacios de color que existe pero para detección de rostros son los más usados actualmente, segundo paso es escoger el sistema que se va a usar para establecer el rango correcto para detectar la piel y efectuar el cambio en la imagen. Los más usados en este tipo de metodología son: segmentación

por umbralización paramétricos y no paramétricos y segmentación por crecimiento de regiones.

Espacios del color

Los espacios de color son modelos, que buscan dar una interpretación de la percepción humana que se tiene como color, a través de ecuaciones matemáticas tridimensional basadas en tres coordenadas que describen características del color. Es decir que hay una relación entre la percepción del color con la asignación de puntos en las tres coordenadas, las cuales varían según el modelo de espacio de color en que se esté trabajando.

1. El modelo RGB (Red, Green, Blue) es un modelo cromático que se basa en la combinación de tres colores el rojo, verde y azul para crear una variedad de colores. Éste modelo se representa en un espacio cubico 3D Figura 3, que se rige con tres coordenadas cartesianas donde cada vértice tienen un valor de los colores RGB, hay que recalcar que se asumen que el cubo es normalizado es decir que trabaja en un rango de 0 a 1 [2].

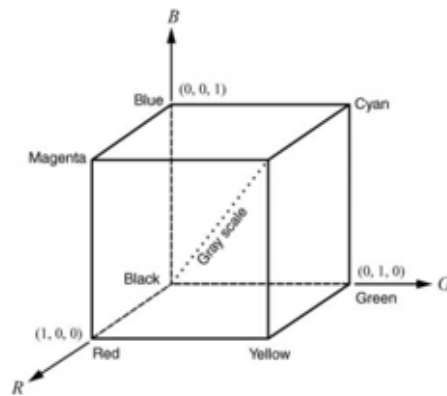


Figura 3. Representación espacial del modelo RGB (tomada de [26]).

Este modelo es uno de los más usados para el almacenamiento de dispositivos gráficos, pero no es muy adecuado en la implementación de varios algoritmos de procesamiento de imágenes digitales, debido a que es muy sensible a la luz y los colores rojo verde y azul están estrechamente relacionados.

2. El modelo HSV (Hue, Saturation, Value o matiz, saturación, valor), es un espacio de color que tiene una representación espacial en forma de cono o pirámide hexagonal (Figura 4), donde la matriz o H es el componente de tonalidad que define el color como por ejemplo el rojo, verde, azul magenta entre otros. Los valores en que varía H es de 0 a 1 en una escala circular, esto es que cuando $H=0$ y $H=1$ representan el mismo color. La saturación o S es la cantidad de luz blanca que tiene el color en otras palabras la pureza del color, y se evalúa entre 0 y 1 siendo $S = 1$ un color con una pureza de 100%. El valor o V es la intensidad del color dicho de otra manera en la cantidad de luz o oscuridad que puede tener el color. El valor varía entre 0 a 1. Para tener mayor claridad de como varia el color en el modelo HSV se presenta el siguiente ejemplo; si ya se tiene el H definido y un valor de S diferente a 0 y se le disminuye V, a el color que se tiene va a presentar un oscurecimiento en otras palabras si es un azul claro al disminuirle V se va a convertir en azul oscuro.

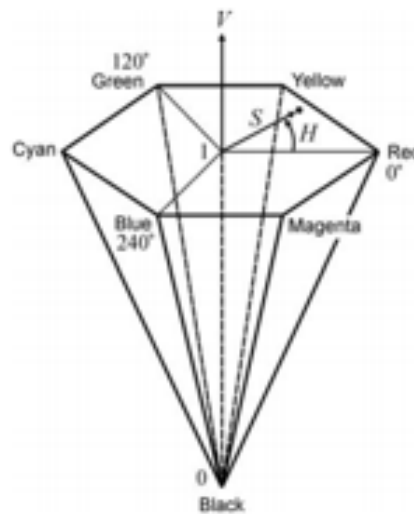


Figura 4. Representación espacial del modelo HSV (tomada de [26]).

EL modelo HSV es bastante usado en los algoritmos de procesamiento, pero como en la mayoría de las aplicaciones se obtiene las imágenes en el modelo RGB es necesario transformar del formato RGB a HSV para eso se utilizan la ecuación 1,2 y 3 o si se está utilizando Matlab con el comando “*rgb2hsv*” para convertir de

RGB a HSV o si es de HSV a RGB se usa el comando “*hsv2rgb*” [26].

$$H \begin{cases} (6 + \frac{G-B}{Max-Min} \times 60^\circ), Si R = Max \\ (2 + \frac{B-R}{Max-Min} \times 60^\circ), Si G = Max \\ (4 + \frac{R-G}{Max-Min} \times 60^\circ), Si B = Max \end{cases} \quad (1)$$

$$S = MAX - Min \quad (2)$$

$$V = Max \quad (3)$$

Donde MAX = max(R, G, B), MIN =min(R, G, B)

3. El modelo YCbCr es un espacio de color, generalmente usado para representar la información de videos digitales y realizar diferentes procesamientos de imágenes. Este modelo YCbCr está representado por la luminancia y dos señales que compone la información del color. Donde Y representa la luminancia y se obtiene con la suma ponderada de los valores RGB, Cb y Cr almacenan el componente del color, el cual Cb es la que contiene componente crominancia azul y se calcula con la diferencia del componente azul y el valor de referencia y por ultimo Cr (o cromancia roja) se calcula con la diferencia del componente rojo y el valor de referencia.

La conversión de RGB a YCbCr se establece con la matriz (ecuación 4) o si se está utilizando Matlab con el comando “*rgb2ycbcr*” para convertir de RGB a YCbCr o si es de YCbCr a RGB se usa el comando “*ycbcr2rgb*” [26].

$$\begin{pmatrix} Y \\ Cb \\ Cr \end{pmatrix} = \begin{pmatrix} 0,299 & 0,587 & 0,114 \\ -0,169 & -0,332 & 0,500 \\ 0,500 & -0,419 & -0,081 \end{pmatrix} + \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (4)$$

Debido a que la luminancia del modelo YCbCr es independiente a la información del color, es adecuado implementarlo en algoritmos de procesamientos de imágenes digitales como la detección de rostros.

Segmentación

La segmentación es una técnica muy usada en visión computacional y procesamiento

de imágenes digitales, que se ha implementado en los últimos 40 años a diferentes aplicaciones que actualmente se están usando [26]. El objetivo principal de la segmentación es reconocer objetos en una imagen, para eso es necesario hacer procesos de partición o división de la imagen en varias zonas o regiones que tengan características similares, a partir de su contorno, textura, color, gradiente, geometría o intensidad de los píxeles. Conviene subrayar que cada una de las regiones o zonas al unir las corresponde a la imagen original, además estas zonas hacen referencia a los objetos y partes significativas como el fondo en dicha imagen.

Para la aplicación de este tipo de técnica se utilizan algoritmos de segmentación los cuales se basan en dos propiedades básicas de los valores de píxel [26]:

1. Discontinuidad: es cuando la imagen presenta cambios bruscos a nivel de intensidad para así detectar los puntos aislados, líneas o bordes.
2. Similitud: se presenta en una imagen cuando esta se puede dividir basándose en regiones que tengan alguna semejanza, por medio de crecimiento o umbralización.

Teniendo las dos propiedades en que se basan los algoritmos de segmentación, es necesario mencionar los métodos que usan para elaborar la segmentación, los cuales se pueden clasificar en tres categorías:

1. Métodos basados en intensidad (también conocido como no contextual): son métodos que se basan en la distribución de píxeles en el histograma de la imagen como por ejemplo la técnica de segmentación por umbrales.
2. Métodos basados en regiones (también conocido como contextual): se basan en un criterio de similitud entre un píxel y sus vecinos, el método que comúnmente se utiliza es el de crecimiento de regiones.
3. Otros métodos como por ejemplo la segmentación basada en texturas, bordes entre otros.

Métodos basados en intensidad

Los métodos basados en intensidad se enfocan en estudiar las estadísticas de los píxeles a través del histograma para así poder determinar los objetos de la imagen. En estos

métodos se pueden encontrar muchas técnicas pero las más usadas en la detección de rostro son: segmentación por umbralización y el método de Otsu.

1. **Umbralización:** Este método se utiliza cuando la imagen a analizar presenta una clara distribución de intensidad permitiendo visualizar una diferencia en los píxeles, es decir que se considera que la imagen tenga dos regiones diferentes (como por ejemplo un objeto y el fondo uniforme), de modo que el histograma de esa imagen sea de una forma bimodal (ver Figura 5), esto es que se presenta 2 picos en el histograma representando cada región, por lo cual se pueda tener un único valor de umbralización (o T), en el que se evalué toda la imagen. Cabe señalar que las imágenes a procesar por segmentación es necesario que se presenten en escala de grises debido a su facilidad de análisis [26].

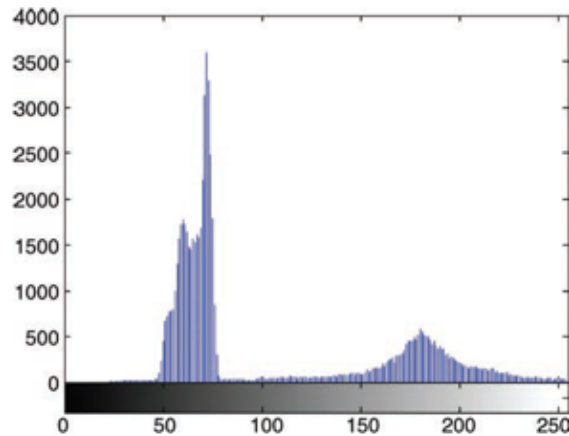


Figura 5. histogramas bimodal representando una imagen con 2 regiones. (Tomada de [26]) .

Para los casos en donde la imagen presenta un histograma bimodal el valor de T se puede obtener de forma manual con los siguientes pasos:

- a) Observar el histograma de la imagen (Matlab *imhist*)
- b) Seleccionar el umbral apropiado usualmente corresponde al mínimo local que está entre los 2 picos del histograma.
- c) Aplique el valor de T a la imagen (Matlab *im2bw*).
- d) Verificar el resultado; si es lo esperado guardar la imagen, en cambio si no es lo esperado repita el punto b y c.

No obstante en ocasiones las imágenes no presentan solo dos regiones razón por la cual en el histograma no se distingue las zonas con los dos picos, para esos casos es necesario implementar modelos que obtengan el valor umbral de manera óptima ya que el método manual explicado presenta muchas limitaciones como:

- Es difícil identificar correctamente los mínimos en el histograma.
- Se presenta problemas si la variación de intensidad es muy leve.

2. **Metodo de otsu:** El método otsu inventado en 1979 por Nobuyuki Otsu quien planteo un metodo en donde usa técnicas estadísticas con el objetivo de encontrar un umbral optimo sin que tenga una supervisión humana en el procesamiento de imagen.

El método de otsu utiliza una distribución de probabilidad, media y varianza para definir el umbral optimo, con esto quiere decir que el método de otsu calcula el valor del umbral de una imagen a través de la dispersión de los niveles de grises, buscando que la media de la dispersión dentro de cada clase sea lo más pequeña, pero que la dispersión entre las clases sea la más alta [45].

Los pasos para realizar este método son los siguientes:

- a) Se obtiene el histograma normalizado de una imagen que contiene N pixeles y un nivel de gris ente 1 y L niveles.
- b) Se evalúa la probabilidad con la ecuacion de ocurrencias del nivel de gris en la imagen.

$$P_i = \frac{F_i}{N} \quad (5)$$

Donde i es el nivel de gris F_i es la frecuencia de repetición de nivel de gris con $i= 1, 2,3, \dots L$

- c) En el caso de umbralización en 2 niveles de una imagen (o binarización), los pixeles se dividen en 2 clases C1 y C2 con nivel de gris de $[1, 2, 3,t]$ y $[t+1,t+2, L]$ respectivamente, por lo tanto la distribución de probabilidad en ambas clases.

$$C1 = \frac{P1}{W_1(t)}, \dots, \frac{Pt}{W_t(t)} \quad (6)$$

$$C1 = \frac{P_{t+1}}{W_2(t)}, \frac{P_{t+2}}{W_2(t)}, \dots, \frac{P1}{W_2(t)} \quad (7)$$

Donde

$$W_1(t) = \sum_{i=1}^t P_i \quad W_2(t) = \sum_{i=t+1}^L P_i \quad (8)$$

d) Se calcula las medias para una de las clases con las ecuaciones 9.

$$\mu_1 = \sum_{i=1}^t \frac{i*P_i}{W_1(t)} \quad \mu_2 = \sum_{i=t+1}^L \frac{i*P_i}{W_2(t)} \quad (9)$$

e) Se calcula la intensidad media total de la imagen con la ecuación 10.

$$\begin{aligned} W_1 * \mu_1 + W_2 * \mu_2 &= \mu_T \\ W_1 + W_2 &= 1 \end{aligned} \quad (10)$$

f) Se halla la varianza entre clases de una imagen con un análisis discriminante o la dispersión total.(ecuación 11).

$$\sigma_B^2 = W_1 * (\mu_1 - \mu_T)^2 + W_2 * (\mu_2 - \mu_T)^2 \quad (11)$$

g) Se encuentra el umbral según el método otsu que define, que para una umbralización de 2 niveles, el umbral óptimo se elige con el valor máximo de la varianza.

$$\begin{aligned} t^* &= Max(\sigma_B^2(t)) \\ 1 &\leq t \leq L \end{aligned} \quad (12)$$

Al utilizar el método de otsu se obtiene muchas ventajas como en encontrar un umbral óptimo y automático, sin embargo si se aumenta el número de las regiones o clases en la imagen este método tiende a utilizar más tiempo computacional para seleccionar un umbral de multinivel óptimo, además este método no tiene en cuenta los bordes, sombras, luminosidad, regiones entre otros aspectos perceptuales que afectan a la adecuada decisión del umbral [26].

Métodos basados en regiones

Los métodos de segmentación por medio de crecimiento de regiones son procesos en los cuales buscan agrupar los píxeles, o conjuntos de píxeles, para encontrar varias regiones (u objetos) basándose en un criterio predeterminado [26]. Matemáticamente

este método se puede describir como un proceso de partición de una imagen en n regiones R_1, R_2, \dots, R_n , tal que mantengan con las siguientes propiedades:

1. La unión de las n regiones será la imagen completa.
2. Requiere que cada región sea conexa, es decir que todos los pixeles que estén en una región estén conectados entre sí.
3. Las regiones no se pueden superponer.
4. Los pixeles deben satisfacer el criterio establecido para pertenecer a esa región.
5. Asegurarse que 2 regiones adyacentes son diferentes entre sí.

El crecimiento de las regiones tiene un procedimiento en el que consiste en ir formando regiones de pixeles incorporando sucesivamente los pixeles a los grupos que cumplan con el criterio predefinido.

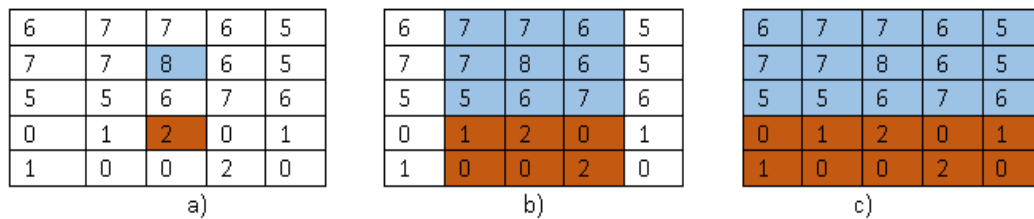


Figura 6. Crecimiento de regiones: a) pixel semilla, b) primera iteración c) iteración final.

El proceso del crecimiento de regiones con que se comienza con un conjunto de pixeles llamados semillas, que son las que representan las distintas regiones de la imagen. Teniendo las semillas se van incorporando los nuevo pixeles a la región, utilizando un mecanismo de crecimiento detectado si el pixel pertenece a esa región. Si el pixel detectado cumple con la regla de homogeneidad, es decir que la región añadiéndole ese pixel sigue siendo homogénea, hace que dicha región se amplié con el nuevo pixel. En la Figura 6 representa el proceso de crecimiento de regiones explicado anteriormente [46] [26].

Los factores claves para que el procedimiento de crecimiento de regiones se lleve a cabo con éxito es necesario tener en cuenta los siguientes criterios.

- La elección de criterios de similitud.
- La selección de los puntos semillas, estas puede determinarse de forma interactiva o basarse en un análisis preliminar de agrupamiento de la imagen.
- Una región debe dejar de crecer cuando no hay más pixeles que satisfagan el criterio de homogeneidad.

Métodos de detección de bordes Canny

Canny propuso un método muy popular, potente y efectivo para la detección de bordes en una imagen digital. Su algoritmo se puede describir con los siguientes pasos:

1. A la imagen original se suaviza con un filtro para bajo gaussiano con parámetros de escala.
2. El gradiente local se calcula a cada punto de la imagen suavizada.
3. Se busca que el grosor de los bordes sea considerablemente delgado a este proceso se le denomina supresión no máxima.
4. Al tener los bordes delgados, se busca limitarlos utilizando 2 umbrales T_{low} y T_{high} , en otras palabras se establecen unos límites donde los bordes con valores mayores que T_{high} son pixeles con bordes fuertes y los que tienen valores menores a T_{low} son pixeles con bordes débiles. A este proceso se le conoce como umbral de histéresis.
5. Se vinculan los bordes, agregando los pixeles de bordes débiles a los fuertes.



Figura 7. Detector de bordes de un rostro por el algoritmo de Canny (tomada de [2]).

Al realizar estos pasos se obtiene los bordes de la imagen como se muestra en la Figura 7 que detecta los bordes del rostro humano. El comando de Matlab para realizar Canny en una imagen es $j = \text{edge}(I, 'canny', T, \text{sigma})$; donde $T = [T_{low} \ T_{high}]$ [26].

Método de construcción de plantilla del rostro humano

El método de construcción de plantillas para la detección de rostros humanos, es una forma para identificar el rostro de una persona en una imagen digital. Este método se basa en reconocer las características más representativas del rostro humano como son; los ojos, nariz, boca, orejas, mejillas entre otros. Para así construir una plantilla que tenga todas las características como se ve en la Figura 8 a), hay que tener en cuenta que pueden haber diferentes tipos de plantilla como una de rostro completo o medio rostro, es necesario recalcar que las facciones de la zona izquierda y la derecha son similares como se ven en la Figura 8 b) por lo tanto usualmente para detectar el rostro se utiliza la plantilla de medio rostro[47].

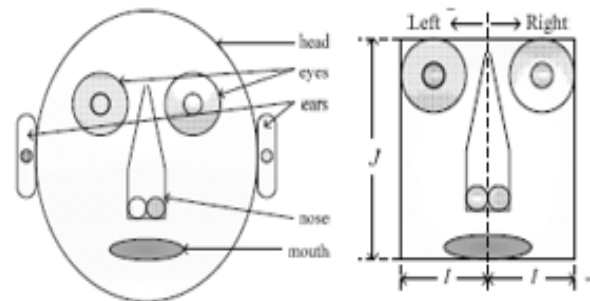


Figura 8. Plantilla del rostro humano a) imagen original b) imagen de la plantilla del rostro (tomada de [47]).

Para el diseño de estas plantillas se establece el siguiente proceso.

1. Elegir las imágenes del rostro.
2. Determinar el tamaño del área del rostro y seleccionar el área del rostro.
3. Normalizar áreas del rostro seleccionado en el mismo tamaño.
4. Calcular el valor promedio de cada pixel correspondiente del área del rostro .

5.1.2. Métodos estadísticos

Los modelos basados en estadísticas comúnmente llamados aprendizaje automático (o

machine learning), son aquellos algoritmos que no asumen ningún tipo de información previa de la topología de un rostro, al contrario a partir de un conjunto de muestras (imágenes de rostros y no rostros) de entrenamiento, se extraen la información relevante que diferencia un rostro y un no rostro. De modo que el algoritmo tendría la capacidad de clasificar donde se encuentra ubicado espacialmente un rostro en una imagen [38].

Actualmente se tienen diferentes métodos estadísticos, pero el más usado para la detección de rostro es el método de Viola - Jones creado como su nombre lo dice por Paul Viola y Michael Jones en el 2004, éste presenta varias ventajas respecto a otros algoritmos que se utilizan en la detección de rostros, como por ejemplo su eficiencia, rapidez y el bajo coste computacional[41].

Detector de viola-jones

Paul Viola y Michael Jones crearon un algoritmo de detección de rostros eficiente y rápido denominado Viola-Jones, éste fue uno de los primeros algoritmos de detección capaz de tener un rendimiento apto para aplicaciones en tiempo real. Por lo tanto, actualmente se usa en diversas aplicaciones que necesiten detectar un rostro en una imagen digital[41].

El algoritmo de Paul Viola y Michael Jones tiene tres contribuciones importantes para la detección de rostros los cuales son: primero, elegir las características que describen el rostro, esa elección es fundamental para los resultados de la detección, debido a que el clasificador necesita de esas características para discriminar rostro y no rostro en una imagen. El objetivo es encontrar máscaras simples para permitir una clasificación rápida pero con suficientes características para poder clasificar el rostro. Esta selección se obtiene mediante el uso de características Haar. La respuesta de las características Haar no es más que la diferencia de dos, tres o cuatro regiones rectangulares en diferentes escalas y formas. Para mejorar el tiempo de cálculo de esas características se emplea una nueva representación de la imagen denominada imagen integral, la cual permite calcular un área de rectángulo con operaciones elementales. Segunda contribución es la realización del algoritmo AdaBoost el cual busca que con las características del rostro, proporcionar un algoritmo de aprendizaje efectivo. Finalmente, la tercera contribución es la implantación en cascada del clasificador, enfocado en la detección de regiones, en otras palabras al descender en el proceso de cascada los clasificadores son más específicos

en la detección de rostros, dando como resultado la detección del rostro en una imagen [43].

Con las contribuciones del método de viola-jones se puede establecer que éste se divide en 4 partes:

- Las características tipo Haar
- Imagen Integral
- Algoritmo de AdaBoost
- Cascada de clasificadores

Características tipo Haar

Una de las propuesta de Viola y Jones es la utilización de las características tipo Haar que son las mismas utilizadas por Papageorgiou [43]. Las características tipo Haar se define sobre regiones de imágenes en escala de grises, el cual utiliza características muy sencillas del rostro, es decir que cada sub-imagen (o región rectangular) representa una zona del rostro como los ojos la nariz, la boca entre otros.

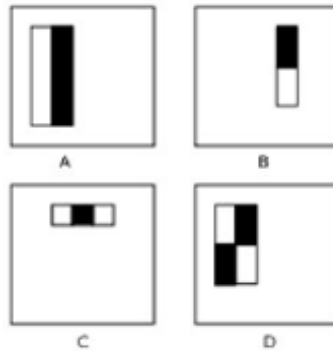


Figura 9. Características tipo haar de dos rectángulos A) y B). la Figura C) muestra característica de 3 rectángulos y D) una característica de cuatro rectángulos (tomada de [43]).

Viola y jones limitaron el espacio de las características con regiones que tienen el mismo tamaño y forma y están adyacente horizontal o verticalmente, permitiéndose solo 2, 3 o 4 rectángulos (ver Figura 9). El valor de una característica se obtiene sumando todos los pixeles del rectángulo blanco y restándole todos los pixeles del rectángulo negro.

Por ejemplo una característica de tres rectángulos se calcula con la suma dentro de dos rectángulos exteriores sustraídos de la suma en un rectángulo central, o con una función de cuatro rectángulos se calcula la diferencia entre los pares diagonales de rectángulos. Cada característica se puede rotar 90° para obtener nuevas características.

El número de características diferentes que se puede generar para una imagen de resolución de 24×24 pixeles es de más de 160.000.

Patrones Binarios Locales (LBP)

El método de patrones binarios locales fue diseñado para la descripción texturas, ya que es altamente discriminativo e invariante a cambios constantes de iluminación. Sin embargo, Ahonen [19], consideran que un rostro es una composición de micro-patrones, y que por lo tanto puede ser descrito efectivamente usando este método. Los autores [16] demuestran en su trabajo que el uso de LBP para la detección de rostros puede tener buenos resultados. El método de LBP asigna etiquetas a cada uno de los pixeles de la imagen tomando en cuenta la distribución de sus pixeles vecinos. Es decir, cada pixel es comparado con los pixeles alrededor para establecer que etiqueta debe ser asignada a dicho pixel. El histograma de las etiquetas de todos los pixeles es posteriormente utilizado como una descripción de la textura de la imagen. El proceso que se realiza en el siguiente:

- Una máscara de tamaño determinado recorre la imagen de manera iterativa seleccionando cada vez un pixel central y sus vecinos.
- Este pixel central se compara con cada uno de sus vecinos de forma ordenada. Se asigna un 1 cada vez que el pixel central sea menor que el pixel comparado y un 0 en el caso contrario.
- El número binario resultante se convierte en un número decimal que es contado en el histograma para formar la descripción.

En la Figura 10 muestra gráficamente el proceso de obtención de los patrones con una máscara de 3×3 ; Sin embargo, diferentes tamaños de máscara son utilizados para lograr la descripción invariante a escala de texturas de diferentes tamaños[16].

En 2002, Ojala y Pietikäinen (2002) propusieron una mejora del método, incorporando vecindarios circulares, donde los vecinos se ubicaban igualmente espaciados, realizando

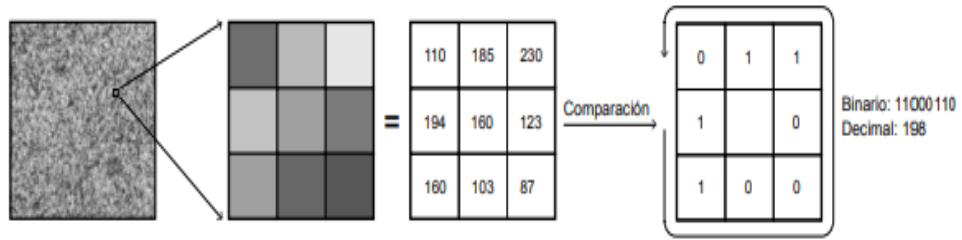


Figura 10. Proceso de obtención de patrones binarios locales.

una interpolación bilineal en caso de que éstos no coincidieran con el centro de un píxel concreto de la imagen. Dado que este método puede describir regiones locales del rostro que son posteriormente combinadas para formar una descripción global del rostro, esa descripción global se puede usar para la detección de rostros.

Imagen integral

La imagen integral es una representación alternativa para la imagen que se puede



Figura 11. Convertir imagen a imagen Integral.(Tomada de [12])

calcular muy rápido. Un valor de la imagen integral $ii(x,y)$ en las coordenadas (x,y) será igual al valor del píxel $i(x,y)$ de la imagen sumando todos los píxeles que estén arriba y a la izquierda de la posición incluyendo este último punto(como se ve en la Figura 11). Se puede definir la imagen integral mediante la ecuación 13.

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \quad (13)$$

Sabiendo que ii es la imagen integral de la imagen inicial i y $ii(x,y)$ es el valor de

la imagen integral en el punto (x,y) . Gracias a la utilización de esta representación cualquier suma rectangular en la imagen original se puede calcular en cuatro referencias de arreglos en la imagen integral.(ver Figura 12). Claramente la diferencia entre dos sumas rectangulares se puede calcular en ocho referencias. Puesto que las características de dos rectángulos definidas anteriormente implican sumas rectangulares adyacentes, pueden calcularse en seis referencias de arreglo, ocho en el caso de las características de tres rectángulos y nueve para cuatro características de rectángulo [36].

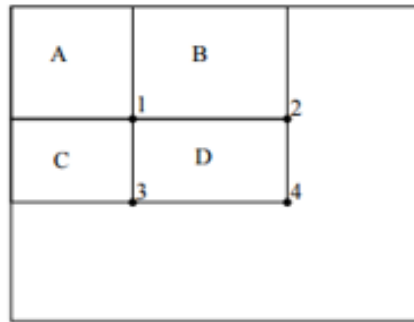


Figura 12. Calculo de un rectángulo a partir de una imagen integral. Para calcular el valor del rectángulo D podría realizarse con cuatro referencias de arreglos. Donde el valor de la imagen integral en la posición 1 es la suma de los pixeles en el rectángulo A. El valor en la posición 2 es $A+B$, en la posición 3 es $A+C$ y en la posición 4 es $A+B+C+D$. Para calcular el valor del rectángulo D es $4+1-(2+3)$.(tomada de [36])

Algoritmo de AdaBoost

El método de Boosting adaptativo o Adaboost, es un algoritmo que busca crear un clasificador “fuerte” con la combinación lineal de clasificadores “débiles”. En otras palabras el AdaBoost entrena una serie de clasificadores “débiles” de manera iterativa, de modo que cada nuevo clasificador o “weak learner” se enfoque en los datos que fueron erróneamente clasificados por su predecesor, por esa razón el algoritmo se adapta con el fin de obtener mejores resultados [20].

Viola y Jones propusieron combinar varias características sencillas en un solo clasificador entrenado por el algoritmo de Adaboost para detectar rostros. Es decir, las características obtenidas por el proceso de Haar son evaluadas por un “Clasificador débil” si la sub- imagen corresponde a un rostro o no como se muestra en la Figura 13.

El criterio de clasificación es dependiendo de que la característica este por encima del peso que se le asignó al clasificador.

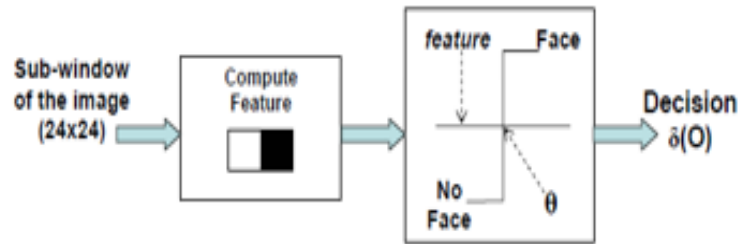


Figura 13. Adaboost clasificador débil. (Tomada de [38])

En vista de que un “clasificador débil” suelen entregar resultados muy pobres, se combinan varios clasificadores débiles para generar un “clasificador fuerte” cuya tasa de detección crece exponencialmente [38]. Uno de los problemas que tiene el Adaboost es que no minimiza directamente los falsos negativos si no que minimiza el error total.

Cascada de clasificadores

Debido a que al construir un detector de rostro con el algoritmo de Adaboost, presenta la desventaja de que la tasa de falsos negativos (detectar un rostro donde no lo es) es muy alta. Para resolver ese problema Viola y Jones propusieron un esquema basado en una cascada de “clasificadores fuertes” como se ve en la Figura 18.

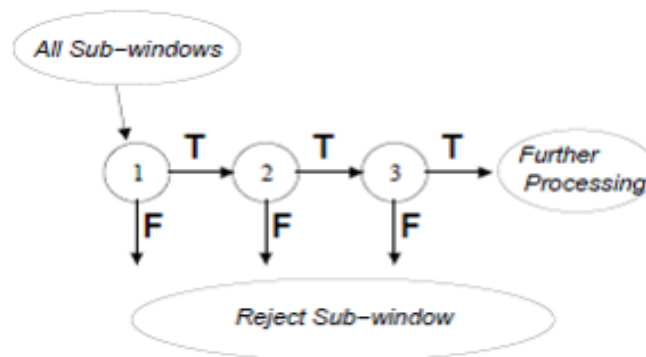


Figura 14. Descripción esquemática de la detección en cascada (tomada de [36]).

El detector de rostros consiste en una serie de clasificadores fuertes (etapas) colocadas en cascadas. Una sub-imagen se clasifica con la primera etapa, si la etapa considera que

es un rostro, se pasa a la siguiente etapa. Por el contrario si la etapa lo clasifica como no rostro, la sub-imagen se desecha. La sub-imagen atravesara uno tras otro clasificadores más complicados y solo las que lleguen al final se consideran como rostros.

5.2. Reconocimiento Facial

El reconocimiento facial o reconocimiento del rostro es una aplicación ampliamente explorada en las últimas décadas[49]. Debido a que se ha convertido en una de las aplicaciones mayor usadas de la visión computacional, ya que ayuda a garantizar la identidad de un individuo a través de los rasgos fisiológicos del rostro de la persona[24].

El reconocimiento facial se puede presentar principalmente en dos tareas verificación e identificación. La verificación se centra en verificar si dos imágenes de un rostro coincide o no, es decir que comparan las características faciales de las personas conocidas con la persona a reconocer. La identificación consiste en reconocer a la persona, como ya se tiene verificada a la persona con esa imagen se busca en una base de datos su nombre o número de identificación.

Aunque actualmente para el reconocimiento facial se han propuesto muchos métodos, todavía se investiga para mejorar la precisión de reconocer ante diferentes variaciones, tales como la orientación facial, iluminación y expresión facial como se ve en la Figura 13. Entre esas investigaciones se presenta la extracción de características ya que juega un papel muy importante para mejorar la tasa de reconocimiento dado que si se extrae unas buenas característica la capacidad de discriminación aumenta [14].



Figura 15. Imágenes de la misma persona pero con cambios de iluminación (tomada de [49]).

Entre las soluciones para esos problemas se presentaron algoritmos y clasificadores, que se describen a continuación:

5.2.1. Análisis de componentes principales (PCA)

Cuando se recoge la información de una muestra de datos, lo más frecuente es tomar el mayor número de variables posibles para tener mejores resultados. Sin embargo, si se toma demasiadas variables sobre un conjunto de datos el coste computacional es alto, por ejemplo si toma 5 imágenes del rostro de una persona en diferentes posiciones e iluminación y son 20 personas, se tendría que analizar como mínimo 100 imágenes para reconocer el rostro de cada persona. Evidentemente, en este escenario se presenta un problema de dimensionalidad y de coste computacional, debido a que el problema es la cantidad de datos que se tiene que analizar por cada imagen.

El análisis de componentes principales (o en sus siglas en inglés PCA) tiene como objetivo reducir la dimensionalidad de un conjunto de datos, conservando al mismo tiempo la mayor variación posible en el conjunto de datos [24]. En otras palabras es un proceso de reducción dimensional que permite descubrir los vectores que mejor simbolizan la repartición de un grupo de imágenes[6].

PCA está basada en la Transformada de Karhunen-Loeve [34], que consiste en representar un proceso estocástico no periódico a través de una base de vectores obtenidos por el proceso, lo que significa que la PCA permite representar una imagen de un rostro usando una base que se obtuvo a partir de la observación de diferentes rostros.

La finalidad de este método es representar una imagen en términos de un sistema de coordenadas óptimo reduciendo el número final de componentes que se tendrá una imagen. Hay que tener en cuenta que la base de la transformación depende de las observaciones y se forman a partir de los d (tamaño de la imagen proyectada) auto-valores más significativos de la matriz de covarianza. Los auto-valores representan las componentes principales que son comúnmente las imágenes de diferentes rostros, lo anterior supone que, los auto-vectores son los ejes del sub-espacio donde proyectamos las imágenes de los rostros humanos los cuales se corresponden a las direcciones de máxima varianza de los datos[10].

La reducción dimensional realizada por PCA es equivalente al número de auto-vectores que se utilicen. Por lo tanto la imagen proyectada por PCA tendrá una dimensión de valor d como se ve en la Figura 16.

Hay que recalcar otra de las cualidades de PCA, es que realiza una discriminación entre

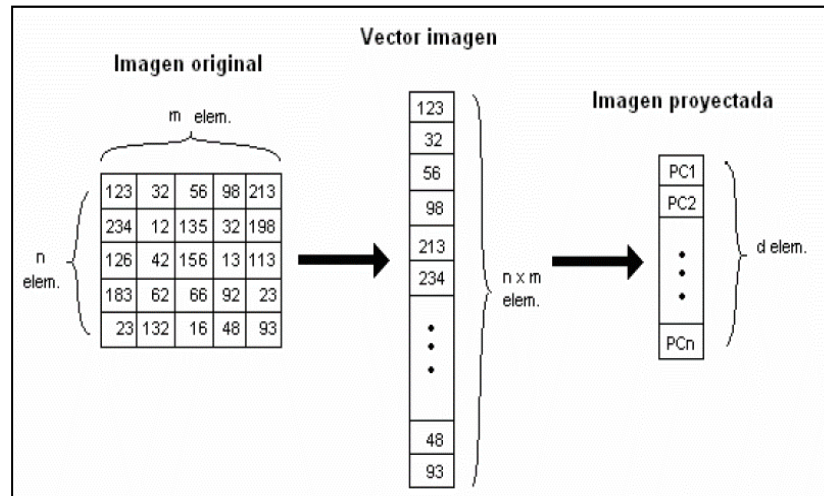


Figura 16. Ejemplo de reducción de dimensional aplicada por PCA (tomada de [10])

clases, donde todas las imágenes del rostro de un individuo están cerca entre sí, mientras que las imágenes del rostro correspondiente a individuos diferentes están más alejadas.

En resumen la PCA puede reducir dimensionalmente una imagen a un punto de sub-espacio de características. Con este sub-espacio de características, es posible proyectar cualquier imagen del rostro y obtener un conjunto de coeficiente de coordenadas. Este conjunto de coeficiente puede utilizarse como base para el reconocimiento del rostro [34]. Los pasos para la realización del algoritmo PCA son [32]:

1. Tratamiento de la base de datos:
 - a) Adquirir el conjunto de M imágenes (T_i) de entrenamiento a través de una cámara web, que son de diferentes personas con varias vistas (frontales, de perfil, con cambio de iluminación, con diferentes gestos faciales, con y sin gafas ...). Las imágenes de rostros $I(x,y)$ que se puede representar en dos dimensiones $N \times N$ (matriz de valores de intensidad de 8-bits). Cabe subrayar que N imagen se puede considerar como un vector de dimensiones N^2 . Por ejemplo una imagen de 256×256 pixeles puede ser un vector de una dimension de 65.536.
 - b) Se normaliza cada imagen y eliminar el ruido.
 - c) Obtiene la imagen media \bar{x} del conjunto de imágenes normalizadas, y restar

el rostro media del conjunto de imágenes para obtener la matriz A.

$$\begin{aligned}\bar{x} &= \frac{1}{M} \sum_{i=1}^M T_i \\ \phi_i &= T_i - \bar{x}\end{aligned}\tag{14}$$

d) Se calcula la matriz de covarianza C (ecuación 15) de $N^2 \times N^2$

$$C = \frac{1}{M} \sum_{K=1}^M (\phi_k \phi_k^T)\tag{15}$$

Donde $C = AA^T$ y $A = \{\phi_1, \phi_2, \phi_3, \dots, \phi_n\}$

En este punto se ha encontrado los vectores y valores propios de la matriz de covarianza (llamados Eigenvectores y Eigenvalores respectivamente) pero esto es una tarea demasiado compleja debido al inmenso tamaño de esta matriz. Es por esto que se adopta una forma simplificada para calcular estos vectores y valores propios. Esta forma simplificada se fundamenta en que el número de imágenes de entrenamiento es usualmente mucho menor que el número de píxeles de la imagen, tan solo existirán M en lugar de N^2 vectores propios significativos, los demás son asociados a valores propios igual a cero.

e) Se calcula los vectores y valores propios de la matriz L que tiene un tamaño M x M y es igual a:

$$L = AA^T\tag{16}$$

f) Una vez que se obtiene el vector V, que contiene los M vectores propios (EigenVectores) $v_i (i = 1, 2, \dots, M)$ de la matriz se pueden obtener los vectores propios de la matriz de covarianza mediante una combinación lineal de los vectores diferencia ϕ_i y los vectores propios de la matriz L. Para esto se utiliza la siguiente ecuación:

$$u = [u_1, u_2, \dots, u_M] = [\phi_1, \phi_2, \dots, \phi_M][v_1, v_2, \dots, v_M] = AV\tag{17}$$

g) se obtiene el patrón (Ecuacion 18)

$$\Omega_i^T = [w_1, w_2, \dots, w_M] \text{ con } i = 1, 2, \dots, M\tag{18}$$

donde

$$w_k = u_k^T (T_i - \bar{x}) \text{ con } k = 1, 2, \dots, M'\tag{19}$$

2. Reconocimiento de la imagen:

- a) Ingresar al sistema una imagen del rostro esta se trasforma a escala de grises, y se convierte a un vector, se pre-procesa
- b) Se calcula la matriz de pesos de las imágenes del conjunto de entrenamiento mediante la expresión.

$$\Omega^T = [w_1, w_2, \dots, w_M] \quad (20)$$

el cual describe la contribución de cada Eigenfaces, en representación de la imagen de entrada.

- c) Transformar la imagen de prueba en sus componentes de Eigenface.

$$\Omega_{in} = U^T \text{EntradaNormalizada} \quad (21)$$

- d) Calcular la distancia euclidiana entre la imagen de entrada y cada imagen almacenada en la base de datos(ecuación 22).

$$E_K^2 = \| \Omega_{in} - \Omega_k \| \quad (22)$$

Si la distancia media euclidiana de todos los vectores es mayor que un valor umbral, entonces la imagen no se reconoce. Si la distancia media euclidiana es menor que el valor de umbral dado, entonces se dice que la imagen se identifica con la base de datos dada.

5.2.2. Análisis lineal discriminante (LDA)

El análisis lineal discriminante (o en sus siglas en ingles LDA) tiene como objetivo convertir un problema de alta dimensionalidad en uno de baja [10]. Para ello maximiza la proporción de dispersión entre clases en lugar de maximizar la dispersión global, con el propósito de encontrar una combinación de características que separen las clases de forma única. En si, la idea principal es que las mismas clases se agrupen entre sí, mientras que las diferentes clases se encuentran más alejadas unas de las otras, de ese modo garantiza una máxima discriminación entre clases. Es necesario subrayar que el LDA lo utiliza Fisherfaces para el reconocimiento de personas; Fisherfaces a diferencia

al Eigenfaces utiliza LDA como técnica para el reconocimiento facial. Con este sistema se busca maximizar la distancia entre clases y minimizar la distancia dentro de la misma clase. La descripción del algoritmo se describe de la siguiente manera [32]:

Sea X un vector aleatorio con muestras tomadas de m clases, tales que $X = X_1, X_2, X_3, \dots, X_m$ y $X_i = x_1, x_2, x_3, \dots, x_i$. Las matrices de dispersión S_b y S_w están dadas por (ecuación 23): Donde μ es la media total

$$\mu = \frac{1}{N} \sum_{i=1}^n x_i \quad (23)$$

Y μ_i la media de la clase $i \in 1, 2, \dots, m$

$$\mu_i = \frac{1}{|X_i|} \sum_{x_j \in X_i} x_j \quad (24)$$

El algoritmo de Fischer busca una proyección W , que maximiza la distancia entre clases:

$$W_{opt} = \operatorname{argmax}_w \frac{|W^T S_b W|}{|W^T S_w W|} \quad (25)$$

La solución para este problema de optimización se da resolviendo el problema de valor Eigen General (ecuación 26) (o General Eigen-value Problem, i.e.)

$$\begin{aligned} S_b v_i &= \lambda_i S_w V_i \\ S_w^{-1} S_b &= \lambda_i V_i \end{aligned} \quad (26)$$

El rango de S_w es de $(N-m)$, con N muestras y m clases. En problemas de visión por ordenador, el número de muestras N es siempre menor que la dimensión de los datos de entrada (el número de píxeles). Por lo tanto la matriz de dispersión S_w se convierte en singular. La solución del problema puede obtenerse realizando un análisis de componentes principales sobre los datos y proyectando las muestras en el espacio $(N-m)$ -dimensional. Pero ya que, S_w ya no es singular, la LDA se realizó sobre los datos reducidos.

El problema de optimización puede ser escrito como (ecuación 27):

$$\begin{aligned} W_{pca} &= \underset{w}{\operatorname{argmax}} |W^T S_t W| \\ W_{fld} &= \underset{w}{\operatorname{argmax}} \frac{|W^T W_{pca}^T S_b W_{pca} W|}{|W^T W_{pca}^T S_w W_{pca} W|} \end{aligned} \quad (27)$$

La matriz de transformación W , que proyecta una muestra en el espacio dimensional ($c-1$), se da entonces con la ecuación 28:

$$W = W_{fld}^T W_{pca}^T \quad (28)$$

Donde W es una matriz de proyección. La W_{fld} es una matriz de proyección que maximiza la división de imágenes de entrenamiento en diferentes grupos. La W_{pca} es una matriz de proyección que maximiza los vectores característicos de todas las clases.

El LDA busca proyectar los datos de manera que su nueva dispersión sea la clasificación óptima. Mientras PCA busca los vectores que mejor describe los datos, LDA busca los vectores que proporcionan mejor discriminación entre clases después de la proyección.

5.2.3. Eigenfaces

El Eigenfaces es un algoritmo que tiene como objetivo reconocer los rostros de una persona en una imagen digital, para ello el algoritmo extrae la información pertinente de la imagen del rostro, para hacer la codificación tan eficiente como sea posible y así compara un rostro codificado con los rostros codificados de una base de datos. Usando La información extraída de una imagen de rostro es de alguna manera las variaciones en un conjunto de imágenes similares (o imágenes con el mismo rostro, pueden ser capturadas y luego usadas para codificar y comparar imágenes de rostros individuales[32]. En otros términos se busca encontrar los componentes principales de la distribución de los rostros o de los eigenvectores de la matriz de covarianza del grupo de características, las cuales en un conjunto hace representación a la variación entre las imágenes de rostro[3]. Donde cada imagen facial se representa como un subconjunto de sus eigenvectores y por lo tanto, llamados Eigenfaces [24].

La cantidad de Eigenfaces que se tiene es la misma cantidad que de imágenes faciales

del grupo de entrenamiento. No obstante, los rostros pueden ser aproximados mediante la utilización de los mejores eigenfaces, es decir los que tienen mayor eigenvalores y por consiguiente cuentan para la mayoría de las varianzas dentro del conjunto de imágenes de rostro. La principal razón para utilizar menos eigenfaces es por la eficiencia computacional, debido a que el mejor Eigenface expande un subespacio dimensional de todas las imágenes posibles. Por lo tanto es usual que se utilice algoritmos de reducción de dimensionalidad como lo es el PCA o LDA, al usar los mejores Eigenface.

Dada una imagen con rostro desconocido y (centrada y de tamaño como las caras de entrenamiento) los pasos del Eigenface son:

1. Normalizar (ecuación 29):

$$y = \phi - \mu \quad (29)$$

Donde μ es la imagen media y ϕ es la diferencia entre la imagen de entrada y la imagen media

2. Proyectarlo en el espacio Eigenspace con la ecuación 30.

$$\sigma = \sum_{i=1}^K w_i u_i \quad (30)$$

Donde $w_i = w_i^T \phi$ y $w = [w_1, w_2, \dots, w_m]^T$

3. Encuentre la distancia entre las dos imágenes faciales ver la ecuación 31.

$$e_r = \min_i \|w - w^i\| \quad (31)$$

4. Si $e_r < T_r$ entonces se reconoce como un rostro del conjunto de entrenamiento. Aquí e_r es la distancia euclidiana entre los eigenfaces de la imagen de entrada (w) y la base de datos (w^l).

5.2.4. Fisherfaces

Fisherface es una técnica de reconocimiento de rostros, que tiene en cuenta la luz y expresiones faciales. Este se encarga de clasificar y reducir las dimensiones de los

rostros usando el método FLD (Discriminant Lineal Fisher). En este sentido, el análisis discriminante de Fisher intenta proyectar los datos de manera que su nueva dispersión sea óptima para la clasificación. Mientras PCA busca los vectores que mejor describen los datos, LDA (Discriminant Lineal Analysis) busca los vectores que proporcionan mejor discriminación entre clases después de la proyección [32]. Fisherfaces realiza un LDA, donde se busca aprovechar la información disponible, sobre la clasificación de las imágenes de entrenamiento, para buscar una proyección que maximice la separación entre imágenes de diferentes personas (o clases) y minimice la distancia entre imágenes de una misma clase. Así logra concentrar las imágenes mejorando, en forma importante, la tasa de reconocimiento [10].

5.2.5. Local Binary Pattern Histogram

A diferencia de los algoritmos Eigenfaces y Fisherfaces que se basan en la comparación directa de las imágenes teniendo en cuenta sus vectores y su dimensionalidad, el algoritmo patrones binarios locales (o en sus siglas en inglés LBPH), se basa en la extracción de características geométricas o fotométricas (contraste, brillo, intensidad, etc.) de cada imagen asignándoles un nivel de gris determinado que posteriormente es representado en un histograma, (el histograma de una imagen es una función discreta de la frecuencia de ocurrencia de los píxeles de la imagen en función de los niveles de intensidad o luminosidad)[48]. El histograma es representado como un gráfico en el cual se muestran distribuidos los valores de intensidad de cada píxel de la imagen (niveles de gris), con el fin de dar a conocer información sobre el brillo, contraste, intensidad, etc. de la misma.

El método de LBPH asigna etiquetas a cada uno de los píxeles de la imagen tomando en cuenta la distribución de los vecinos. Estos son los pasos que el LBPH realiza para su respectivo reconocimiento de imágenes.

- Una máscara de tamaño determinado (8×8), recorre la imagen de manera iterativa seleccionando cada vez un píxel central y sus vecinos
- Este píxel central se compara con cada uno de sus vecinos de forma ordenada. Se asigna un 1 cada vez que el píxel central sea menor que el píxel comparado y un 0 en el caso contrario, como se muestra en la ecuación 32 y se visualiza en la

Figura 2.

$$LBP = \sum_{p=0}^7 s(g_p - g_c)2^p \quad s(x) = \begin{cases} 1, & \text{si } x \geq 0 \\ 0, & \text{otro valor} \end{cases} \quad (32)$$

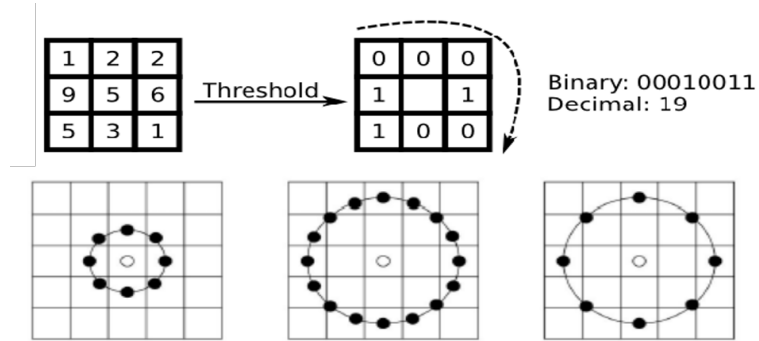


Figura 17. Analisis de LBP [48].

- El número binario resultante se convierte en un número decimal que es contado en el histograma para formar la descripción. El histograma de las etiquetas de todos los pixeles es posteriormente utilizado como una descripción de la textura de la imagen, que se indica en la ecuación 33.

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

$$I(x) = \begin{cases} 1, & \text{si } x \text{ es verdadero} \\ 0, & \text{otro valor} \end{cases} \quad (33)$$

- Por ultimo, para hacer el reconocimiento facial se divide la imagen procesada con LBP en X regiones locales para extraer un histograma de cada región, como lo muestra la Figura 18. El vector de características necesarias para el reconocimiento facial se obtiene concatenando los histogramas locales, es por ello que a esta variación del algoritmo LBP se le da el nombre de LBPH.

Este método resulta ser tolerante a las variaciones de iluminación, es por ello que se utiliza principalmente en el reconocimiento facial, aunque también se ha incursionado en otros enfoques como el reconocimiento del iris, huellas dactilares, clasificación de edad facial.

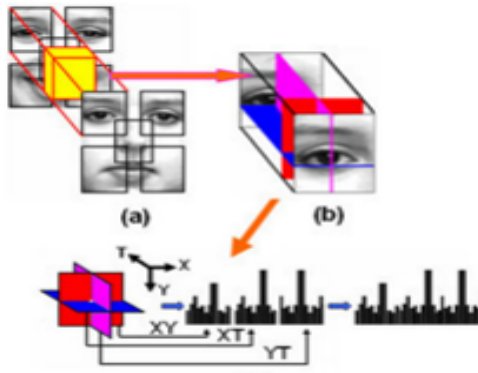


Figura 18. Histograma por cada region de la imagen [48].

5.3. Aprendizaje de Maquina

El aprendizaje de Maquina (o en ingles Machine learning) es una de las áreas de más rápido crecimiento de la informática con aplicaciones de gran alcance. Para dar la definición de ¿qué es? y ¿para se usa? el aprendizaje de maquina, se tiene que remontar al aprendizaje animal, como por ejemplo, cuando las ratas encuentran un alimento con un nuevo olor o aspecto, primero comerán cantidades muy pequeñas. Si el alimento produce un mal efecto, el nuevo alimento a menudo se asociará como una enfermedad y, posteriormente, las ratas no la comerán. Es evidente que hay un mecanismo de aprendizaje, el animal usó la experiencia pasada con algún alimento para adquirir experiencia en la detección de la seguridad del ese alimento. Si la experiencia pasada con el alimento fue etiquetada negativamente, el animal predice que también tendrá un efecto negativo cuando se encuentre en el futuro ese alimento [42].

Teniendo el ejemplo anterior de un aprendizaje exitoso para el animal, se puede demostrar una típica tarea de aprendizaje automático. La programación de una máquina que aprenda a detectar correo electrónicos no deseados. Una posible solución es escanear los mensajes de correo electrónicos previamente vistos y extraer un conjunto de palabras que den la indicativa de que es un correo no deseado. Tal sistema potencialmente podría predecir correctamente la etiqueta de los correos no deseados.

Con los ejemplos anteriormente descritos se puede concluir que la incorporación de conocimientos previos, influye en el éxito o fracaso de los algoritmos de aprendizaje. Pero cuanto más fuerte es el conocimiento previo (o suposiciones anteriores) que inicia el proceso de aprendizaje, más fácil es aprender de otros ejemplos, pero cuanto más fuertes

son estos conocimientos previos, menos flexible es el aprendizaje [42]. Respondiendo las dos primeras preguntas el Aprendizaje de maquina es el proceso que permitan a las maquinas aprender por si solas a partir de un conjunto de datos. Donde la entrada a un algoritmo de aprendizaje es la formación de datos, que representa la experiencia, y la salida es alguna experiencia, que por lo general toma la forma de otro algoritmo que puede realizar alguna tarea.

Unos de los aprendizajes automáticos que se estudian en la actualidad son las redes neuronales (seccion 5.3.1).

5.3.1. Redes Nauronales

Una red neuronal artificial es un modelo matemático inspirado en la estructura de las redes neuronales en el cerebro. En los modelos simplificados del cerebro, se compone de un gran número de dispositivos básicos de cálculo (neuronas) que están conectados entre sí en una compleja red de comunicación, a través del cual el cerebro es capaz de llevar a cabo complejos cálculos. Las redes neuronales artificiales son redes interconectadas masivamente en paralelo de elementos simples (usualmente adaptativos) y con organización jerárquica, las cuales intentan interactuar con los objetos del mundo real del mismo modo que lo hace el sistema nervioso biológico. [42]

Las redes neuronales artificiales tienen gran número de características semejantes a las del cerebro. Por ejemplo, son capaces de aprender de la experiencia, de generalizar de casos anteriores a nuevos casos, de abstraer características esenciales a partir de entradas que representan información irrelevante, etc. Esto hace que ofrezcan numerosas ventajas como por ejemplo:

- Aprendizaje Adaptativo. Capacidad de aprender a realizar tareas basadas en un entrenamiento o en una experiencia inicial.
- Auto-organización. programas cuyo comportamiento se adapta a sus datos de entrada - ofrecen una solución a estos problemas; Son, por naturaleza, adaptables a los cambios en el ambiente con el que interactúan.
- Tolerancia a fallos. La destrucción parcial de una red conduce a una degradación de su estructura; sin embargo, algunas capacidades de la red se pueden retener,

incluso sufriendo un gran daño.

- Operación en tiempo real. realiza procesos con datos de forma muy rápida.
- Fácil inserción dentro de la tecnología existente. Se pueden obtener chips especializados para redes neuronales que mejoran su capacidad en ciertas tareas. Ello facilitará la integración modular en los sistemas existentes.

Topología de una red neuronal

La distribución de neuronas dentro de la red se realiza formando niveles o capas, con un número determinado de dichas neuronas en cada una de ellas como se aprecia en la Figura 19. Además se puede apreciar la estructura de una posible red multicapa, en la que cada nodo o neurona únicamente está conectada con neuronas de un nivel superior. Se debe notar que hay más conexiones que neuronas en sí; en este sentido, se dice que una red es totalmente conectada si todas las salidas desde un nivel llegan a todos y cada uno de los nodos del nivel siguiente. que se pueden distinguir tres tipos de capas las cuales son:

- Entrada: es la capa que recibe directamente la información proveniente de las fuentes externas de la red.
- Ocultas: son internas a la red y no tienen contacto directo con el entorno exterior. El número de niveles ocultos puede estar entre cero y un número elevado. Las neuronas de las capas ocultas pueden estar interconectadas de distintas maneras, lo que determina, junto con su número, las distintas topologías de redes neuronales.
- Salidas: transfieren información de la red hacia el exterior.

Si la topología de la red y las diferentes funciones de cada neurona (entrada, activación y salida) no pueden cambiar durante el aprendizaje, mientras que los pesos sobre cada una de las conexiones si pueden hacerlo; Se dice que el aprendizaje de la red neuronal es: *adaptación de los pesos*.

En otras palabras el aprendizaje es el proceso por el cual una red neuronal modifica sus pesos en respuesta al error en la predicción dada una información de entrada. Los cambios que se producen durante el mismo se reducen a la destrucción, modificación y creación de conexiones entre las neuronas trabajando técnicas de Pruning. Por lo tanto

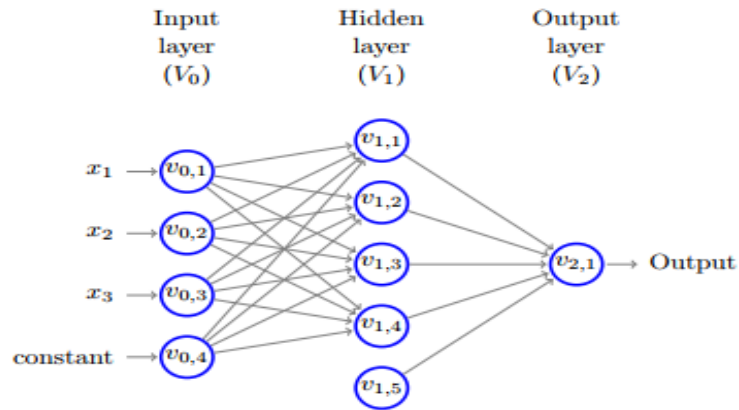


Figura 19. Capas y proceso de la red neuronal (tomada de [30]).

es imperativo conocer cómo se modifican los valores de los pesos, es decir, cuáles son los criterios que se siguen para cambiar el valor asignado a las conexiones cuando se pretende que la red aprenda una nueva información.

Hay dos métodos de aprendizaje en las redes neuronales importantes que pueden distinguirse:

- Aprendizaje supervisado.
- Aprendizaje no supervisado.

5.3.2. Aprendizaje supervisado para Redes Neuronales

El aprendizaje supervisado se caracteriza por que a partir de un conjunto de ejemplos clasificados (conjunto de entrenamiento), intenta asignar una clasificación a un segundo conjunto de ejemplos.[42] En otras palabras en el proceso de aprendizaje se realiza mediante un entrenamiento controlado por un agente externo (supervisor, maestro) que determina la respuesta que debería generar la red a partir de una entrada determinada.

En este tipo de aprendizaje se suelen considerar, a su vez, tres formas de llevarlo a cabo, que dan lugar a los siguientes aprendizajes supervisados:

1. **Aprendizaje por corrección de error** : Consiste en ajustar los pesos de las conexiones de la red en función de la diferencia entre los valores deseados y los

obtenidos a la salida de la red, es decir, en función del error cometido en la salida. Este tipo de aprendizaje tiene varios algoritmos pero los más conocidos son regla de aprendizaje del Perceptron, y regla de aprendizaje Delta o regla del mínimo error cuadrado (LSM). En este trabajo se va a explicar el Perceptron dado que se ha aplicado en los sistemas de reconocimiento de rostros [24][42].

Perceptron tiene como una regla muy simple, para cada neurona en la capa de salida se le calcula la desviación a la salida objetivo como el error δ , el cual, luego se utiliza para cambiar los pesos sobre la conexión de la neurona precedente. El cambio de los pesos por medio de la regla de aprendizaje del Perceptron se realiza según la siguiente regla (ver ecuación 34):

$$\Delta W_{ij} = \sigma * out_j * (a_{qi} - out) \quad (34)$$

Donde a_{qi} es la salida deseada o objetivo de la neurona de salida N_i , $\delta_i = (a_{qi} - out_{qi})$ la desviación objetivo de la neurona N_i y σ tasa de aprendizaje. La salida de la neurona N_j (out_j) se utiliza, porque este valor influye en la entrada global y, por ende, en la activación y luego en la salida de la neurona N_i . Esto es semejante a un “efecto en cadena”.

2. **Aprendizaje por refuerzo:** trata de un aprendizaje supervisado, más lento que Aprendizaje por corrección de error dado que los ejemplos de entrenamiento contienen más información que los ejemplos de prueba, el alumno debe predecir aún más información para los ejemplos de prueba. Por ejemplo, uno puede intentar aprender una estrategia que describa cada posible movimiento en un tablero de ajedrez es decir el grado por el cual la posición de las fichas blancas es mejor que las fichas negras. Sin embargo, la única información disponible para el alumno en el momento del entrenamiento son las posiciones que ocurrieron a lo largo de los juegos de ajedrez reales, etiquetados por quién finalmente ganó ese juego [42].

La tarea de supervisor en este aprendizaje es indicar, mediante una señal de refuerzo (éxito = +1 o fracaso = -1), si la salida obtenida por la red se ajusta o no a lo deseado, y en función de ello se ajustan los pesos basándose en un mecanismo de probabilidades.

3. **Aprendizaje estocástico:** se basa en la introducción de cambios aleatorios en los valores de los pesos de la red, evaluando su efecto a partir de la salida deseada y de una distribución de probabilidad.

5.3.3. Aprendizaje no supervisado para Redes Neuronales

Las redes con aprendizaje no supervisado (también conocido como autosupervisado) no requieren influencia externa para ajustar los pesos de las conexiones entre sus neuronas. La red no recibe ninguna información por parte del entorno que le indique si la salida generada en respuesta a una determinada entrada es o no correcta. En el aprendizaje no supervisado se presenta a la red un conjunto de patrones sin adjuntar la respuesta deseada por lo que la red, por medio de algoritmo de aprendizaje estima la función de densidad probabilística que describen la distribución de patrones.

En cuanto a los algoritmos de aprendizaje no supervisado, en general se suelen considerar dos tipos, que dan lugar a los siguientes aprendizajes:

1. **Aprendizaje hebbiano:** Se basa en el siguiente postulado, formulado por Donal O, hebb en 1949: “Cuando un axón de una celda A esta suficientemente cerca como para conseguir excitar una celda B y repetida o persistentemente toma parte en su activación, algún proceso de crecimiento tiene lugar en una o ambas celdas, de tal forma que la eficiencia de A, cuando la celda a activar es B, aumenta”. En otras palabras se puede decir que el aprendizaje Hebbiano consiste básicamente en el ajuste de los pesos de las conexiones, de acuerdo con la correlación de los valores de activación (salidas) de las dos neuronas conectadas.
2. **Aprendizaje competitivo y comparativo:** Se enfoca a la clusterización o clasificación de los datos de entrada. La característica principal del aprendizaje competitivo es si un patrón nuevo se determina que pertenece a una clase reconocida previamente, entonces la inclusión de este nuevo patrón a esta clase matizará la representación de la misma. Al contrario si el patrón de entrada se determinó que no pertenece a ninguna de las clases reconocidas anteriormente, entonces la estructura y los pesos de la red neuronal serán ajustados para reconocer la nueva clase.

6. DISEÑO METODOLÓGICO

Según el problema propuesto, los objetivos planteados y la investigación realizada se puede determinar que el tipo de estudio que se empleará en este trabajo, será investigativo y de desarrollo cuantitativo-descriptivo, puesto que se busca profundizar en los algoritmos y métodos para la detección y reconocimiento del rostro humano en un área concurrida para así construir un sistema que cumpla con los objetivos planteados.

Para la elaboración de un sistema de detección y reconocimiento del rostro es necesario cumplir principalmente con dos etapas, una de investigación y otra de desarrollo. La etapa de investigación se inicia con el estudio de los antecedentes (sección 4) en los temas relacionados con este trabajo, para ello se realiza la búsqueda de múltiples artículos que reportan resultados de propuestas para detección y reconocimiento de rostros. Por otra parte, para la etapa de desarrollo se plantea seguir los pasos descritos en el diagrama de la Figura 20. Se puede observar que este trabajo cuenta con dos algoritmos principales; el primero es el de detección y localización de rostro, el cual busca encontrar en una imagen la ubicación de rostros humanos y así extraerlo de los demás elementos que tiene la imagen; el segundo es el algoritmo de reconocimiento e identificación del rostro humano, el cual busca identificar a las personas que están presentes en la imagen digital. De acuerdo a los pasos descritos, para la implementación del sistema de detección y reconocimiento es necesario establecer, seleccionar y analizar algunos parámetros:

- Población
- Instrumentos
- Software
- Técnicas de análisis de datos

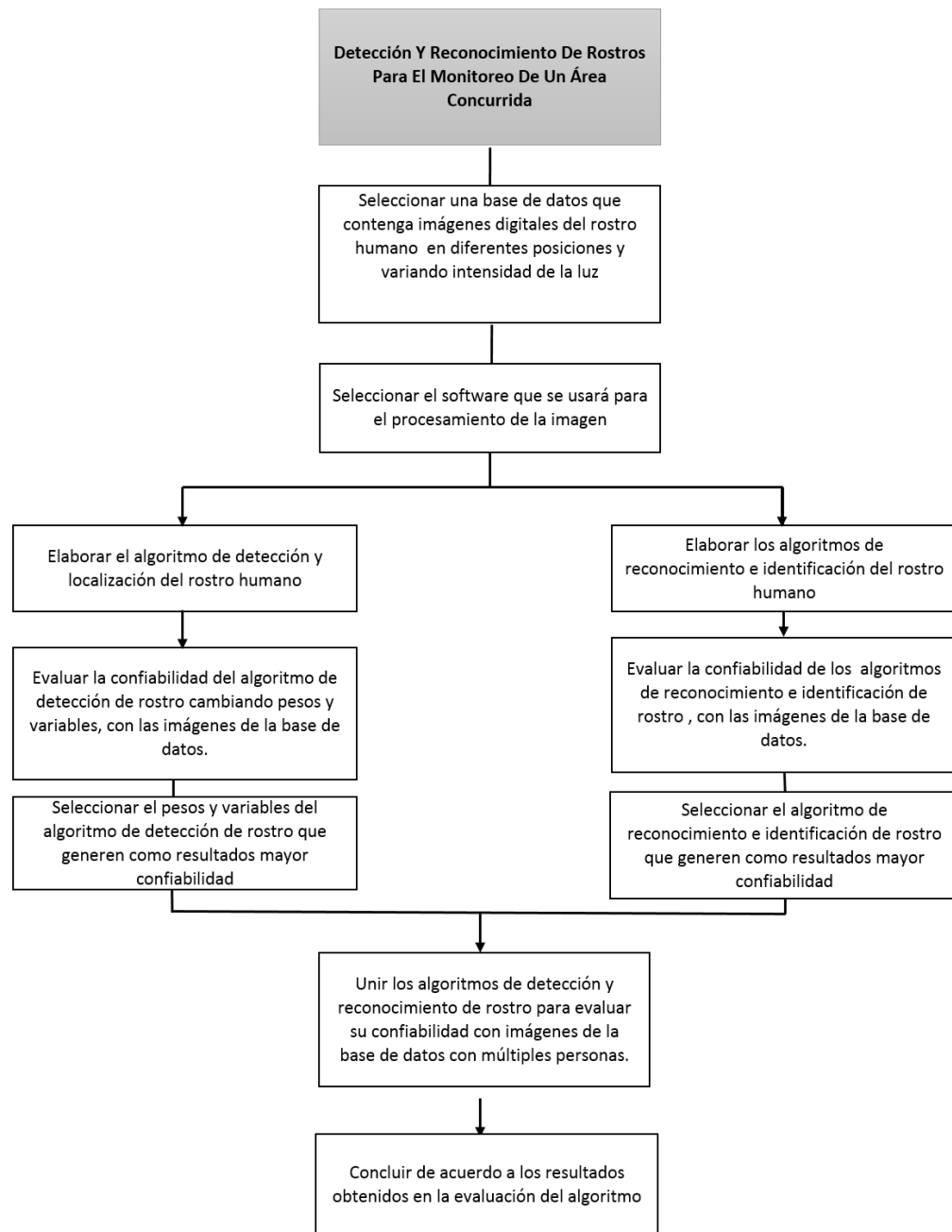


Figura 20. Diagrama del Desarrollo del sistema

El sistema de detección y reconocimiento de rostro que se desarrolló tiene como finalidad identificar todos los rostros que estén en una imagen, por lo tanto primero se seleccionó las bases de datos que contengan imágenes, que cumpla con los siguientes requisitos; primero para el sistema de detección y localización se necesita imágenes con cambios

de iluminación, posición del rostro, y lo mas importante que presenten múltiples personas en dichas imágenes; segundo para el sistema de reconocimiento e identificación se necesitan imágenes con una persona pero con cambios de postura en el rostro, cambios de iluminación y cambios de expresión facial.

El segundo paso realizado en este proyecto fue la seleccionar los métodos para la implementación el sistema de detección y reconocimiento, basándose en los resultados obtenidos de otros investigadores (ver los antecedentes sección 4). Para el sistema de detección y localización del rostro el método que cumple con los objetivos propuestos es el algoritmo de Viola y Jones. Mientras que para el sistema de reconocimiento e identificación se implemento tres métodos que se usan actualmente y cumplen con la propuesta de este proyecto, que son: Eigenfaces, Fisherfaces y LBPH.

Al sistema de detección y localización de rostros se realizaron diferentes pruebas con imágenes con múltiples personas, cambiando los pesos y la variables necesarias para encontrar el sistema que tenga una tasa alta de detección y pocos falsos positivos, y así elegir el sistema que obtenga los resultados deseados.

Para la selección del sistema de reconocimiento e identificación se comparó los resultados obtenidos de los tres métodos mencionados a través de 3 pruebas: la primera con imágenes que contienen cambios de posición de rostros, la segunda con imágenes con cambios de iluminación y la ultima con imágenes que contienen cambios de posición e iluminación, teniendo en cuenta la tasa de reconocimiento y el tiempo de computo.

Para finalizar se unieron los dos sistemas, desarrollando el sistema de detección y reconocimiento de rostros, para así concluir como se desempeña los métodos usados para detectar y reconocer rostros en imágenes que contiene múltiples personas, con variación de iluminación y cambios de posición del rostro.

6.1. Software de procesamiento de información

Para la elaboración del sistema de detección y reconocimiento de rostros es necesario seleccionar el software que se empleará en el procesamiento de las imágenes, algunos de los programas planteados a utilizar son:

1. **Matlab:** es una herramienta de software matemático que combina un entorno de

escritorio perfeccionado para el análisis interactivo y los procesos de diseño con un lenguaje de programación que expresa las matemáticas de matrices y arrays directamente. Es decir que Matlab es un entorno de computación y desarrollo de aplicaciones totalmente integrado orientado para llevar a cabo proyectos en donde se encuentren implicados elevados cálculos matemáticos y la visualización gráfica de los mismos.

Matlab dispone de un amplio abanico de programas de apoyo especializados, denominados Toolboxes, que extienden significativamente el número de funciones incorporadas en el programa principal. Estos Toolboxes cubren prácticamente casi todas las áreas principales en el mundo de la ingeniería y la simulación, destacando entre ellos el 'toolbox' de proceso de imágenes, control robusto, estadística, análisis financiero, redes neuronales lógica difusa, simulación de sistemas dinámicos etc. es un entorno de cálculo técnico, que se ha convertido en estándar de la industria, con capacidades no superadas en computación y visualización numérica.[29]

Esta herramienta está disponible para las plataformas Unix, Windows, Mac OS X y Linux.

2. **Opencv:** (Open Source Computer Vision) es una librería Software open-source de visión artificial y machine learning. Esta librería provee una infraestructura para aplicaciones de visión artificial que ha sido diseñada para ser eficiente en cuanto a gasto de recursos computacionales y con un enfoque hacia las aplicaciones de tiempo real [7]. La librería tiene más de 2500 algoritmos, que incluye algoritmos de machine learning y de visión artificial para usar. Estos algoritmos permiten identificar objetos, rostros, clasificar acciones humanas en vídeo, hacer tracking de movimientos de objetos, extraer modelos 3D, encontrar imágenes similares, eliminar ojos rojos, seguir el movimiento de los ojos, reconocer escenarios, entre otros.

La librería está escrita en los lenguajes C y C++ y es compatible con Linux, Windows y Mac OS X. Cuenta con un desarrollo activo en interfaces para Python, Ruby, Matlab y otros lenguajes.

3. **Visual Studio:** es un conjunto de herramientas y otras tecnologías de desarrollo de software basado en componentes para crear aplicaciones eficaces y de alto rendimiento, permitiendo a los desarrolladores crear sitios y aplicaciones web, así como otros servicios web en cualquier entorno que soporte la plataforma.

En otras palabras Visual Studio es un conjunto completo de herramientas de desarrollo para la generación de aplicaciones web ASP.NET, Servicios Web XML, aplicaciones de escritorio y aplicaciones móviles. Visual Basic, Visual C# y Visual C++ utilizan todos el mismo entorno de desarrollo integrado (IDE), que habilita el uso compartido de herramientas y facilita la creación de soluciones en varios lenguajes [30].

4. **Qt**: es mucho más que un simple SDK multiplataforma. Es una estrategia tecnológica que te permite crear rápidamente y con bajo coste diseñar, desarrollar, desplegar y mantener software mientras provees una experiencia de usuario constante en todos los dispositivos. Además es un framework de C++ abierto, extensible y modular que soporta unos ciclos de vida de desarrollo de bajo coste [33].

Estos sistemas se emplearan en un portátil marca Lenovo con un procesador Intel Core(TM)i7 de 2,10GHz con memoria Ram de 6GB, para realizar el sistema de detección y reconocimiento.

6.2. Técnicas de análisis de datos

Como se ha mencionado anteriormente este trabajo presenta una metodología cuantitativa, buscando validar el sistema de detección y reconocimiento propuesto a través de métricas de evaluación como confiabilidad. Conviene subrayar que la confiabilidad en un software tiene que ser medible, de acuerdo con la IEEE que define el manejo de confiabilidad en el software como “el proceso de optimización de la confiabilidad en el software por un programa que se enfatiza en la prevención de errores en el software, detección y eliminación de las averías que se pueden presentar en el uso de mediciones para maximizar la confiabilidad en la restricción de proyectos, calendarios y su propios funcionamiento”.

Usando esta definición se establecen algunas métricas para evaluar el sistema:

- Taza de ocurrencia de error (ROCOF): esta medida menciona la frecuencia con la cual ocurre un comportamiento inesperado del sistema.

- Métricas de calidad: Son todas las métricas de software que definen de una u otra forma la calidad del software; Tales como exactitud, estructuración o modularidad, pruebas, mantenimiento, reusabilidad, cohesión del módulo, acoplamiento del módulo, etc. Como la matriz de confusión que es una herramienta que permite la visualización del desempeño de un algoritmo.
- Métrica de recuperación de precisión para evaluar la calidad de salida del clasificador: Precision-recall es una medida útil del éxito de la predicción cuando las clases están muy desequilibradas. la precisión es la fracción de instancias recuperadas que son relevantes, mientras que el recall es la fracción de instancias recuperadas para todas las instancias relevantes. En el contexto de Recuperación de información, la curva de Precision-recal se vuelve muy útil.

7. DISEÑO DE LA SOLUCIÓN

En esta sección se presentara el diseño empleado para la elaboración del sistema de detección y reconocimiento. Además la selección de las bases de datos usadas para el entrenamiento y la evaluación de los sistemas.

7.1. Selección de la Base de Datos

En esta sección se describe brevemente diversas bases de datos experimentales que contienen imágenes con rostros, para comparar sus características y poder elegir la base de datos acorde con las necesidades del trabajo.

Para la selección de la bases de datos se establecieron parámetros acorde a la problemática propuesta (ver sección 1). Determinando la necesidad de emplear dos tipos de bases de datos, una para el algoritmo de detección y la segunda para el algoritmo de reconocimiento de rostro. Para el algoritmo de detección se necesita entrenar y evaluar con imágenes donde se encuentren más de un rostro en una imagen y dichos rostros se encuentren en diferentes posiciones, variación de iluminación y con oclusiones. Mientras que el algoritmo de reconocimiento facial necesita imágenes de múltiples personas, donde por persona se tenga un conjunto de imágenes con su rostros en diferentes posiciones, variación de iluminación o oclusiones, esas imágenes deben solo contener el rostro de la persona. Dicho hasta aquí, supone que la diferencia entre la base de datos del algoritmo de detección con la de reconocimiento es que, el primero necesita imágenes con varios rostros en cada imagen y el segundo con imágenes de varios rostros pero en cada imagen el rostro de una sola persona para identificarlo.

Con la explicación anteriormente establecida, se buscó bases de datos para el algoritmo de detección de rostro en la cual se encontraron dos bases de datos que cumplen con los parámetros establecidos, se pueden ver sus características en la Tabla 1.

Considerando la Tabla 1, se seleccionó la base de datos *Face Detection Data Set and Benchmark* (FDDDB) [15], para hacer el entrenamiento y la evaluación del algoritmo de detección de rostros, dado que contiene capturas en diferentes ángulos del rostros, con cambios de iluminación y oclusiones, además tiene imágenes con 2 a 5 rostros por imagen. Por el contrario de la base de datos *Annotated Facial Landmarks in the Wild*

Tabla 1: Características de Bases De Datos para la detección de rostros

	Base De Datos Nombre	Fddb	AFLW
Información	Total de imágenes	2845	25000
	Total de rostros	5171	25010
	Numero de rostros por imagen	2 a 5	2
Características	Color	Si	Si
	Posiciones	Si	Si
	Iluminación	Si	Si
	Oclusiones	Si	Si

(AFLW) [27] presenta pocas imágenes con solo 2 rostros haciendo la base de datos no apta para las necesidades planteadas, pero dado que para el entrenamiento del algoritmo solo se necesitan imágenes que contengan un rostro se uso unas imágenes de esta base de datos.

Para el algoritmo de reconocimiento e identificación de rostros se busco, otras bases de datos diferentes a las mostradas en la Tabla 1, que contengan imágenes que se usaran en el entrenamiento y la evaluación del sistema. Esas imágenes deben contener una persona por imagen, con cambios de iluminación, posición y expresiones faciales, además para las pruebas es necesario que se tengan muchas personas de diferentes razas y edades, con ese conjunto de imágenes. Con estas características se encontraron seis bases de datos, que se presentan en la Tabla 2.

La base de datos Yale face [21] contiene 5850 imágenes de una sola fuente de luz de 10 personas que realizaron 9 posiciones y a esas con 64 condiciones de iluminación. Cada archivo en esta base de datos se puede separar fácilmente, debido a una clara descripción de los nombres de los archivos de las fotos frontales y los otros. Las imágenes son de buena calidad y en alta resolución (640×480 píxeles), como se ve en la Tabla 2.

La base de datos MUCT Face Database [31] de la Universidad de Ciudad del Cabo. Consta de 3.755 imágenes faciales de 276 sujetos. Cada rostro fue fotografiado con el uso de cinco cámaras diferentes al mismo tiempo. Con eso se obtuvieron cinco imágenes faciales con diferentes posiciones. Además, cada persona fue fotografiada con 10 juegos de iluminación diferentes.

La base de datos Specs on Faces (SoF) [1] es una colección de 42.592 ($2,662 \times 16$) imágenes para 112 personas (66 varones y 46 mujeres) que usan gafas en diferentes

Tabla 2: Bases De Datos para el reconocimiento de rostros

	Base de datos nombre	The Yale Face B	MUCT	SOF	Head Pose	UFI	AR
Información	Numero de personas	10	276	112	15	605	126
	Imágenes por persona	576	10 a 15	380	93	9	19 a 36
	Total de imágenes	5850	3755	42592	1395	5445	4000
	Tamaño de la imagen	640x480	480x640	640x480	384x288	128x128	768x576
	Formato de imagen	PGM	JPG	JPG	JPG	PGM	PGM
Características	color	No	Si	Si	Si	No	Si
	variación iluminación	Si	Si	Si	No	No	Si
	Posiciones	Si	Si	No	Si	Si	No
	Oclusiones	No	No	Si	Si	Si	Si
	Variación Fondos	No	No	No	Si	No	No

condiciones de iluminación. El conjunto de datos presenta un nuevo reto en cuanto a detección y reconocimiento de rostros. Se dedica a dos problemas que afectan la detección de rostros, el reconocimiento y la clasificación, que son los ambientes de iluminación duras y las oclusiones faciales. Además las imágenes originales (sin filtros o oclusiones sintéticas) fueron capturadas en diferentes países durante un largo período haciendo que se presente una alta gama de personas de diferentes culturas.

Una de las bases de datos mejor construidas en términos de uso, facilidad de la clasificación de archivos en términos de características es The head pose database [11] de la universidad de Cambridge, que consta de 15 conjuntos de imágenes. Cada conjunto contiene de 2 series de 93 imágenes de la misma persona en diferentes poses. Hay 15 personas en la base de datos, usando gafas o no y teniendo varios colores de la piel. La postura o orientación de la cabeza está determinada por 2 ángulos (h, v), que varía de -90 grados a +90 grados.

La base de datos Unconstrained Facial Images (UFI) [22], tiene dos tipos de conjuntos de imágenes. El primero contienen solo el rostro de las personas donde el tamaño del

rostro es así casi igual en todas las imágenes. El segundo conjunto de imágenes el rostro no está localizado y las imágenes tienen objetos que no son rostros.

La última base de datos investigada es The AR Face Database, creada en The Ohio State University, USA: contiene más de 4.000 imágenes en color correspondientes a los rostros de 126 personas (70 hombres y 56 mujeres). Las imágenes presentan rostros frontales con diferentes expresiones faciales, condiciones de iluminación y oclusiones (gafas de sol y bufanda). Las imágenes fueron tomadas en el CVC bajo condiciones estrictamente controladas [28].

De todas las bases de datos disponibles, la que contiene mejores especificaciones es la MUCT debido a que se tienen imágenes con 10 diferentes enfoques de iluminación, además posee rostros en diversas rotaciones y a diferencia de bases de datos como SOF y AR las imágenes están ordenadas de forma que se diferencia el conjunto de imágenes de una persona al de otra. Y la base de datos Head Pose por que tienen diferentes ángulos de captura del rostro, los rostros tienen oclusiones y diferentes posiciones, además las imágenes están ordenadas por persona y por el ángulo en que se encuentra el rostro. Como las dos bases de datos contienen los datos necesarios para el entrenamiento y la evaluación del algoritmo de reconocimiento se va a trabajar con la información de las dos bases de datos lo cual generaría una base de datos con 5150 imágenes, 291 individuos y con rostros con diferentes ángulos de captura y variación de iluminación.

7.2. Selección del algoritmo de Detección

Un algoritmo de detección de rostros en imágenes se enfoca en localizar la ubicación arbitraria de un número indefinido de rostros en una imagen. El problema de detectar rostros en imágenes que contienen múltiples objetos, es la capacidad de clasificar correctamente los objetos que no son rostros y los rostros. Actualmente existen diversas técnicas que han sido presentadas en proyectos donde solucionaron esta problemática, a través de técnicas basadas en reglas o en técnicas probabilísticas.

Luego de estudiar el estado del arte respecto al problema de detección en la Tabla 3 se resumen algunas características principales de cada una de dichas técnicas.

Analizando la información encontrada para el problema de este trabajo se desearía aplicar una técnica que no tenga análisis de color porque sería muy susceptible a cambios

de iluminación y en este caso no se va a tener una condición de la luz controlada, técnicas que no se basen en la geometría del rostro puesto que es sensible a los cambios de posición del rostro y en el trabajo no se tendrá un ambiente controlado en el ángulo en que se encuentre el rostro. Por lo cual ninguna técnica que se use la información del pixel o la geometría del rostro cumple con las condiciones propuestas

En las técnicas probabilísticas se tiene la combinación de algoritmo de Adaboost con dos tipos de clasificadores [49] es muy usado para detectar rostros en perfil y rostros frontales pero no detecta rostros con ángulos de 30 o 45 o con oclusiones por tal razón no es apta para este trabajo. Pero con una de las técnicas probabilísticas más importantes en la detección de objetos es la de Viola and Jones debido a que su tasa de detección es alta, tienen la capacidad de detectar rostros en diferentes posiciones y con cambios de iluminación además este trabaja en tiempo real y su costo computacional es menor que el de otros algoritmos mostrados. Por consiguiente para solucionar la problemática de este trabajo se seleccionó la técnica de Viola and Jones para la detección de rostros en una imagen, sin embargo como se muestra en la Tabla 3 se muestra algunas derivaciones de esta técnica en donde se presenta algunas mejoras en el tiempo de procesamiento o en clasificación pero ninguno de estos se trabajó en tiempo real.

Tabla 3: Características de Técnicas de Detección

Técnicas	Características	Datos
Segmentación: skin color y Canny [2]	Detección de rostros a través del color de la piel y el detector de bordes llamado Canny	No elimina otra zona que tenga la misma tonalidad de la piel como las manos o cuello
Skin color y Local Iterated Conditional Modes [46]	Detección por el color y crecimiento de regiones	Elimina cualquier elemento que no sea rostro además el tiempo de coste es 18.14 s y la tasa de detección 90.27 %
Geométrico [25]	Detección de rostro a través de la simetría del rostro con plantillas que contengan características del rostro completo y medio rostro	La tasa de detección de plantillas de rostro completo 22.5 % y plantilla medio rostro 95 % para rostro con ángulo de 30°
Adaboost (AM-CC) [20]	Generación de cascada de clasificación, evitando el sobreentrenado.	Tasa de detección 91.6 % y un tiempo medio de 1.2s
DP-Adaboost [49]	Combinan el algoritmo Adaboost con dos clasificadores (clasificador de rostros frontal y clasificador de rostros en perfil)	Pueden detectar muchos rostros de perfil y también se puede mejorar la tasa de éxito para la detección de rostros del 68,35 % al 71,87 %.
Adaboost integrales difusas [38]	Clasificador basado en Integrales Difusas. Minimiza el número de datos clasificadores	La tasa de detección de rostros obtenida es superior al 92 %, y se rechazan casi todas las subimágenes (99 %)
Viola and Jones [43]	Trabaja en tiempo real y al emplear una cascada de clasificadores débiles de Adaboost para reducir el tiempo de computación	La tasa de detección con imágenes normalizadas es del 97 %

7.3. Selección del algoritmo de Reconocimiento

En la actualidad existen diferentes técnicas para identificar la persona en una imagen, estas técnicas buscan verificar si dos imágenes de un rostro coinciden o no, es decir que comparan las características faciales de una persona desconocida con personas ya identificada y así se reconoce a la persona en la imagen. Algunas de las técnicas que realizan dicha identificación son: Eigenfaces, Fisherfaces, LBPH, Redes Neuronales o Support Vector Machines (SVM). Estas técnicas fueron empleadas en diferentes proyectos en los cuales solucionaron problemáticas como el coste computacional, el tiempo de procesamiento o la tasa de eficiencia. Al estudiar el estado del arte respecto al problema de reconocimiento e identificación de rostros se encontraron diferentes soluciones que se presenta en la Tabla 4 que resume algunas características de dichas técnicas.

Considerando las investigaciones y los resultados de los métodos usados actualmente para el reconocimiento de rostro (ver Tabla 4), se seleccionó tres métodos que según se vio en el estado del arte son los mas usados para solucionar el problema de reconocimiento de rostros, los cuales, presentan una tasa de reconocimiento alto en ambientes no controlados (como se ve en la Tabla 4), y el coste computacional bajo, que son Eigenfaces (con PCA), Fisherfaces (o LDA) y Local Binary Patterns Histograms (LBPH). A éstos tres metodos se realizaran pruebas para comparar su rendimiento y escoger el método que cumpla con los objetivos planteados en este proyecto y tenga un mejor desempeño en las pruebas realizadas.

En la Tabla 4 se puede ver resultados de los tres metodos escojidos para este proyecto el eigenfaces con el Análisis de Componentes principales (PCA) y el eigenfaces con el Análisis Discriminante Lineal (LDA) [40], con variantes para mejorar su tasa de eficiencia y su tiempo de procesamiento, esto lo realizan con redes neuronales o con SVM, poseen una tasa de reconocimiento mayor al 90 % y el LBPH tiene la tasa de reconocimiento del 96 %.

Tabla 4: Características de Técnicas de Reconocimiento

Técnicas	Características	Datos
Eigenfaces PCA KNN (EPK) Eigenfaces LDA KNN (ELK) [32]	Minimizar el error entre la imagen de prueba y las imágenes de base de datos. Las variaciones de postura causan vulnerabilidad en la identificación	La tasa de éxito donde el método EPK es de un 97.50% y para el ELK es de un 94.55% y su tasa de fallos son de 2.5% y 5.55% respectivamente
La fusión del PCA y LDA [4]	A diferencia de PCA, LDA minimiza la dispersión dentro de la clase y maximiza la dispersión entre clases en el espacio proyectado	Tasa de reconocimiento (91,25%) en comparación con PCA (89%) y LDA (90,25%)
PCA para extracción de características, Redes neuronales (NN) y Support Vector Machines (SVM) para la clasificación [24]	La buena capacidad de aprendizaje del NN y del SVM, aumenta la precisión en reconocer, reducir la complejidad en términos de los tamaños necesarios para dispositivos de memoria y en términos de cálculo de tiempo.	La precisión de reconocimiento del 91,33% para 120 imágenes
LBPH (Local Binary Patterns Histograms) [48]	Para el LBPH asigna etiquetas a cada uno de los píxeles de la imagen tomando en cuenta la distribución de los vecinos. Estos son los pasos que el LBPH realiza para su respectivo reconocimiento de imágenes.	La tasa de reconocimiento es de 96%
The multi-task learning in deep convolutional neural networks(ConvNet)[39]	El multi-task learning in deep ConvNet es para una representación de características discriminativa y pose-robusta. introduciendo cuatro tareas multiples	Al usar todas las tareas para aprender ConvNet, la tasa de reconocimiento es de 81,2%
Linear Discriminant Regression Classification (LDRC) y Deep Learning[40]	La simplicidad, la generalidad y la buena capacidad de aprendizaje del NN y del SVM, aumenta la precisión en reconocer, reducir la complejidad en términos de los tamaños necesarios para dispositivos de memoria y en términos de cálculo de tiempo.	La precisión de reconocimiento del 92,8% para las imágenes de la base de datos YAEL

7.4. Implementación del sistema de detección y reconocimiento de rostro

En esta sección se explicara la estructura que se implementó para el sistema de detección y reconocimiento de rostros según los objetivos planteados. En primer lugar se define el sistema de detección y localización, como se ve en la Figura 21; la entrada del sistema es: una imagen con uno o varios rostros, con el fin de detectar y ubicar todos los rostros encontrados. En segundo lugar con el rostro detectado, se identifica con el sistema de reconocimiento de rostro, el rostro que fue detectado como se ve en la salida del sistema en la Figura 21.

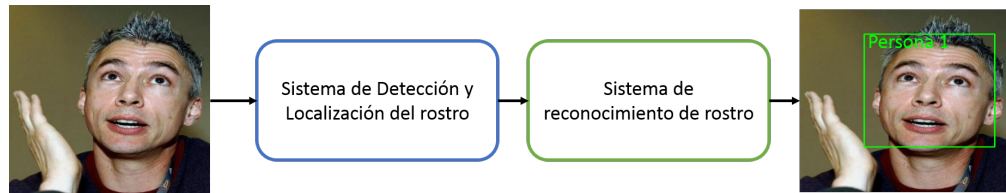


Figura 21. Sistema de detección y reconocimiento

7.4.1. Sistema de Detección y Localización

Un Sistema de Detección y Localización es un proceso que a partir de una imagen se pueda detectar y localizar arbitrariamente un número indefinido de rostros que se encuentre en la imagen. En la Figura 22 se puede ver el esquema de funcionamiento que tiene el sistema de detección y localización, la entrada es una imagen que puede contener uno o múltiples rostro. A la imagen se le realizar un proceso de escaneo el cual detecta los posibles rostros que este contenga, y así extraer la ubicación de los rostros detectados para finalmente señalarlos en la imagen. En otras palabras la salida del sistema de detección y localización es una lista con las coordenadas de la esquina superior, el ancho y alto del rostro detectado por el sistema. Por ejemplo en la Figura 23, muestra una lista con la coordenada de los rostro detectados donde la primera columna es la cantidad de rostros encontrados (“numero de rostros detectados”) la segunda y tercera columna son las coordenadas en x y y de la esquina superior izquierda del rostro, y por ultimo esta el largo (h) y ancho del rostro (w), con las anteriores indicaciones se dibujar un rectángulo en la imagen para ubicar los rostros en esta como se ve en la Figura 23 a).

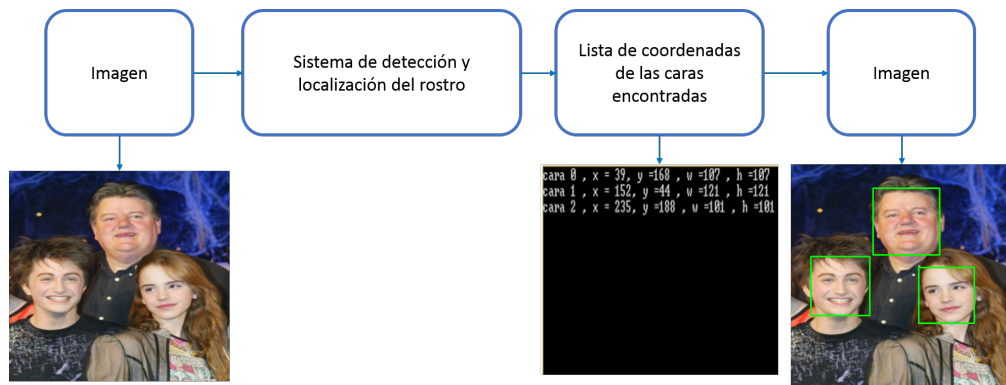


Figura 22. Esquema general del Sistema de detección y localización del rostro

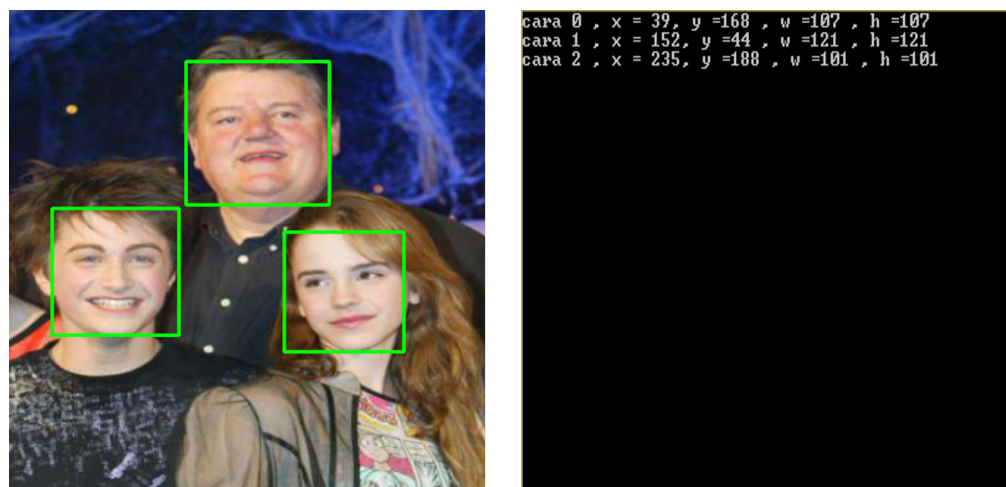


Figura 23. Resultados de detección a) detección de rostros que contiene la imagen y b) lista de coordenadas de los rostros detectados

Existen diversos sistemas de detección y localización de rostros humanos como se ve en los antecedentes (sección 4), pero según la investigación realizada el método que cumple con los objetivos propuestos en este proyecto es el de Viola and Jones, debido a que con este se puede trabajar en tiempo real y el gasto computacional es mínimo comparado con otros métodos, además según los resultados de los trabajos investigados es un sistema confiable y con una tasa alta de detección.

Para poder evaluar el sistema se necesita entender el termino de tasa de detección (*Detection Rate* en inglés) que se define como el porcentaje de rostros correctos detectados respecto del total de rostros de la imagen. Para calcular la tasa de detección y la calidad de un sistema de detección y localización es necesario usar los siguientes parámetros, ver la Tabla 5.

Tabla 5: Parámetros de evaluación para el sistema de detección y localización

Parametro	Definición
Verdaderos Positivos (VP)	Indica la cantidad de rostros que el clasificador detecto correctamente.
Falsos Negativos (FN)	Indica la cantidad de rostros que no han sido correctamente detectadas.
Falsos Positivos (FP)	Indica la cantidad de regiones que ha marcado como rostro cuando realmente no lo hay.

Para tener un clasificador con un rendimiento adecuado es necesario realizar un análisis con el método de evaluación, *precision-recall*. El *precision-recall* o PR son diagramas comúnmente utilizados por investigadores para evaluar la calidad de salida de un clasificador, que se caracteriza por dos medidas la *precision* y *recall*. Donde la precisión mide la calidad de las respuestas positivas del clasificador, definido como la proporción entre los verdaderos positivos y el total de todas las detecciones realizadas (ver ecuación 35). *Recall*, mide la eficiencia en la clasificación de todos los elementos, se define por la ecuación 36.

$$Precision = \frac{VP}{VP + FP} \times 100 \quad (35)$$

$$Recall = \frac{VP}{VP + FN} \times 100 \quad (36)$$

La curva de PR muestra la compensación entre precision y eficiencia para diferentes clasificadores. El clasificador ideal es cuando la precision y el recall son 100 %. Cuando el recall es alto y el precision bajo , el clasificador da como resultado muchas detecciones incorrectas (no hay rostro). En cambio si precision es alto y bajo el recall, entrega pocos resultados pero la mayoría de las detecciones son correctas (hay rostros). Por lo tanto se busca un sistema el cual tenga alto el recall y también la precision.

7.4.2. Modelo de detección y localización

En esta sección se describe la estructura y el funcionamiento del sistemas de detección y localización realizado en este proyecto, con base al método diseñado por Viola y Jones

[36]. El cual es un algoritmo reconocido y usado en la actualidad debido a que es un método que se distingue por su velocidad y precisión. El detector de Viola y Jones toma como base imágenes de 384 x 288 píxeles, capaz de procesar 15 frames por segundo, operado en un procesador Intel Pentium III de 700MHz convencional [36].

Según los autores tiene tres contribuciones principales que caracterizan a el detector de Viola y Jones como unos de los mas eficientes algoritmos de detección de rostros, por su bajo coste computacional, velocidad y alta tasa de detección. La primera contribución es la nueva representación de la imagen de entrada, a través de cálculos simples de los píxeles, a esto se le denomina imagen integral; está permite evaluar de forma rápida las características Haar utilizadas en la detección. No se trabaja con la intensidad de la imagen, si no con la representación de la misma. Una de las ventajas de usar la imagen integral es que se puede calcular las características utilizadas en cualquier lugar de la imagen, a cualquier escala, en la misma velocidad y eficiencia desde el punto de vista del gasto computacional.

La segunda contribución es un método de clasificación eficiente, con el fin de garantizar una rápida clasificación. El proceso de aprendizaje debe excluir la mayoría de características y centrarse en las mas representativas que contiene las sub-ventanas analizadas. Se logra mediante modificaciones del clasificador Adaboost: combinado clasificadores “débiles” para la construcción de un clasificador “fuerte”. El Adaboost es utilizado tanto para seleccionar las características, como para entrenar los clasificadores.

Para finalizar la tercera contribución es el método de combinación de los clasificadores; los clasificadores mas sencillos de bajo costo computacional descartan la mayoría de las sub-ventanas que no tiene rostros dejando que los clasificadores mas complejos clasifiquen las zonas donde hay mas probabilidad de que haya un rostro.

Con las contribuciones planteadas por Viola y Jones se estableció la estructura de sistema de detección realizada en este proyecto que se divide en cuatro etapas como se muestra en la Figura 24.

Pre-procesamiento. En esta etapa, la imagen de entrada se somete a procesos de escalamiento y de escala de grises.

1. La escala es el proceso de cambiar el tamaño de la imagen digital

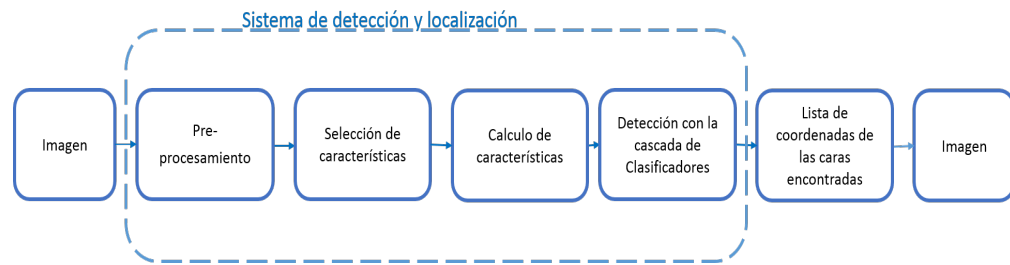


Figura 24. Esquema del Sistema de detección y localización del rostro

- Convertir la imagen digital de color a escala de gris.

Selección de características Después de escalar y convertir a escala de gris la imagen de entrada, el algoritmo de detección y localización de rostro busca detectar y localizar los rostros que contiene dichas imágenes. Esto es llevado a cabo mediante la selección de características que tiene un alto grado de información de un rostro. Las características usadas en este proyecto son las Haar básicas (ver Figura 9). Dado que las características permite hacer una mejor representación de la información que contiene la imagen, además el procesamiento con características es más rápido que, con los píxeles.

Calculo de Características Para el cálculo de todas las características se crea una nueva representación de la imagen de entrada llamada imagen integral. La cual permite extraer de forma rápida características a diferentes escalas y en cualquier lugar de la sub-ventana, ya que no se trabaja directamente con los valores de intensidad sino con una imagen acumulativa de la imagen de entrada. Para realizar la representación de la imagen integral es necesario realizar cálculos básicos con los valores de luminancia de los píxeles que tiene la imagen de entrada, los cálculos están explicados en la sección 5.1.2.

Detección con la cascada de Clasificadores Esta etapa se describe un algoritmo que asigna una sub-ventana a los conjuntos de sub-ventanas que contengan o no un rostro, con la clase que tenga mayor similitud, de acuerdo al modelo al que se ha entrenado. En otras palabras se busca una clasificación de sub-ventanas con rostro o sin rostro. En el algoritmo de Viola y Jones buscan rechazar sub-ventanas negativas (no rostros) con el menor costo computacional y en el menor tiempo posible, con esto hace

referencia a que el objetivo no es buscar los rostros sino descartar las sub-ventanas que contengan no rostros.

Basados en la idea de descartar primero las sub-ventanas que contengan no rostros, y que se tenga un alto grado de detección a un bajo coste computacional, se usa las contribuciones de Viola y Jones, basado en tener en cascada varios clasificadores sencillos, con el fin que los primeros sean capaces de rechazar los no rostros, para que los últimos clasificadores puedan detectar los posibles rostros.

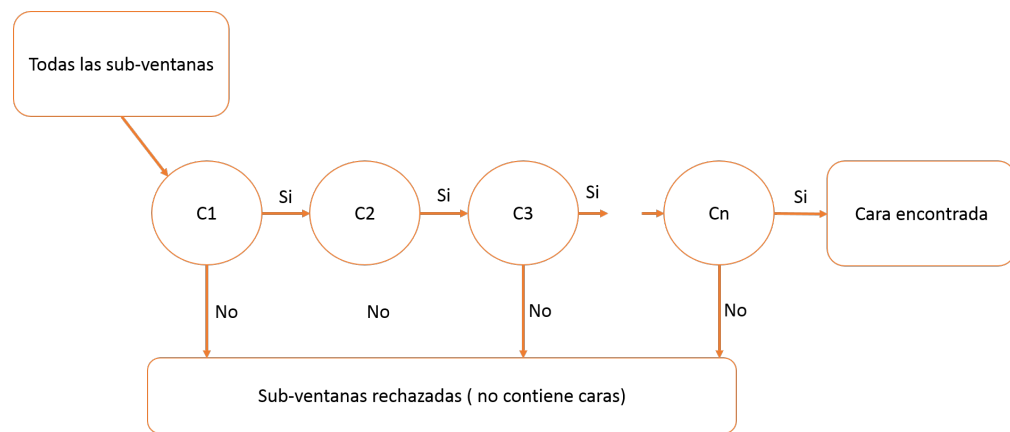


Figura 25. Estructura de clasificador en Cascada

la forma general del proceso de detección que planteo Viola y Jones, y usada en este proyecto, es la de un árbol binario de clasificadores que deciden si una sub-ventana tiene o no un rostro. Llamado también cascada de clasificadores (ver Figura 25) , construido por clasificadores débiles entrenados con Adaboost.

El trabajo de cada etapa de los clasificadores consiste en determinar si la sub-ventana que se analiza no es un rostro o podría ser un rostro. Cuando una sub-ventana es clasificada como no rostro en cualquier etapa inmediatamente se descarta como un no rostro. Por el contrario, si se clasifica como un posible rostro, pasa esa sub-ventana a la siguiente etapa. Si esa sub-ventana pasa por todos los clasificadores, se clasificara como un rostro.

Entrenamiento de clasificadores Para la detección y localización del rostro se implementa características que tengan la capacidad de representar y describir un rostro en una imagen, y así procesarlas y asignarlas como un rostro o no rostro. Razón por

la cual es necesario utilizar un aprendizaje supervisado (o entrenamiento supervisado) que sea capaz de diferenciar rostros a no rostros.

Hay que recordar que en este proyecto se planteo usar las características Haar para la extracción de características. Estas características se pueden calcular de manera rápida y eficiente en una imagen integral como se ha mencionado, además se pueden escalar y orientar con facilidad como se muestra en la Figura 26 en una imagen o en la Figura 9. Viola y Jones utilizan en el descriptor con una resolución de 24 por 24 píxeles, que conlleva a un total de 160000 características Haar básicas [36]. Dado que esa cantidad de características conlleva un coste computacional alto se busca seleccionar de las mismas características las más representativas. La dificultad es una elección lo bastante robusta como para detectar un rostro.



Figura 26. Las características Haar usadas para la extracción de las características, una de dos rectángulos que abarca la características de los ojos y una de tres rectángulos que no tiene características relevantes para identificar el rostro

Por lo tanto es necesario utilizar la segunda contribución de Viola y Jones, la modificación del Adaboost. Como se mencionó anteriormente Adaboost usa un conjunto de clasificadores débiles para formar un clasificador fuerte, pero a los clasificadores débiles se limitan, de modo que cada clasificador débil puede depender de una sola característica Haar. Como resultado, cada etapa de los clasificadores débiles se selecciona las características Haar más relevantes para la detección, en la Figura 26 se ve un ejemplo de una característica relevante y no relevante. La ventaja de usar este método es que el algoritmo de aprendizaje es efectivo y tiene fuertes límites de rendimiento. Cabe recalcar que el Adaboost se utiliza para seleccionar las características Haar más relevantes y entrena el algoritmo.

En el proceso de entrenamiento se tienen en cuenta las limitaciones a las que se ve sometido. En la mayoría de los casos, los clasificadores que utilizan más características

logran tener una mayor tasa de detección y tasa de falsos positivos mas baja. Al mismo tiempo, los clasificadores con más características a calcular necesitan mas tiempo de computo, para determinar si una sub-ventana contiene un rostro o no. De modo que se puede decir para tener un entrenamiento optimo se tiene que elegir: primero el numero de etapas de clasificación, numero de características, el umbral del cada etapa y el tamaño del descriptor (sub-ventana de detección). Sin embargo realizar esta selección de forma optima es un trabajo muy costoso. Por tal razón Viola y Jones especificó que para entrenar de manera efectiva la cascada de clasificadores se elige la tasa de falsos positivos (o falsas detecciones del rostro) y tasa de detección de cada etapa así como la tasa de detección referente al el detector.

También para el entrenamiento es imprescindible escoger las imágenes a entrenar es decir las imágenes positivas Figura 27 (imágenes que contiene un rostro por imagen), imágenes negativas (imágenes que no contiene rostros como fondos o objetos que se pueden encontrar en las imágenes que contiene rostros).



Figura 27. Imágenes usadas para el entrenamiento positivas y negativas

7.4.3. Sistema de reconocimiento e identificación

Un sistema de reconocimiento e identificación forma parte de un área de estudio denominado Biometria, cuyo propósito consiste en desarrollar métodos automatizados para reconocer e identificar las personas mediante características físicas, como por ejemplo la forma de las orejas, el patrón del iris, la huella dactilares o el rostro, entre otros. En este proyecto se usó las características morfológicas del rostro que definen a cada individuo, como la forma de los ojos, los pómulos, la boca, la nariz, entre otros.

El sistema implementado para realizar el reconocimiento de rostros, se basó en el esquema que se ve en la Figura 28; al sistema le ingresa el rostro de una persona en

una imagen digital (obtenida por el sistema de detección y localización), dicha imagen entra en el sistema de reconocimiento e identificación con el fin de determinar si es una persona conocida o desconocida. En el primer caso si reconoce el rostro de la imagen, el sistema determina el numero que le fue asignado a esa persona (como se ve en la Figura 29 que presenta la imagen de entrada y los datos de salida, que son: el numero de la persona, y con el numero de la persona se busca en la base de datos de conocidos el nombre). Y el segundo caso es cuando no reconoce el rostro de la persona en otras palabras es un desconocido para el sistema, el numero asignado para las personas desconocidas es -1.

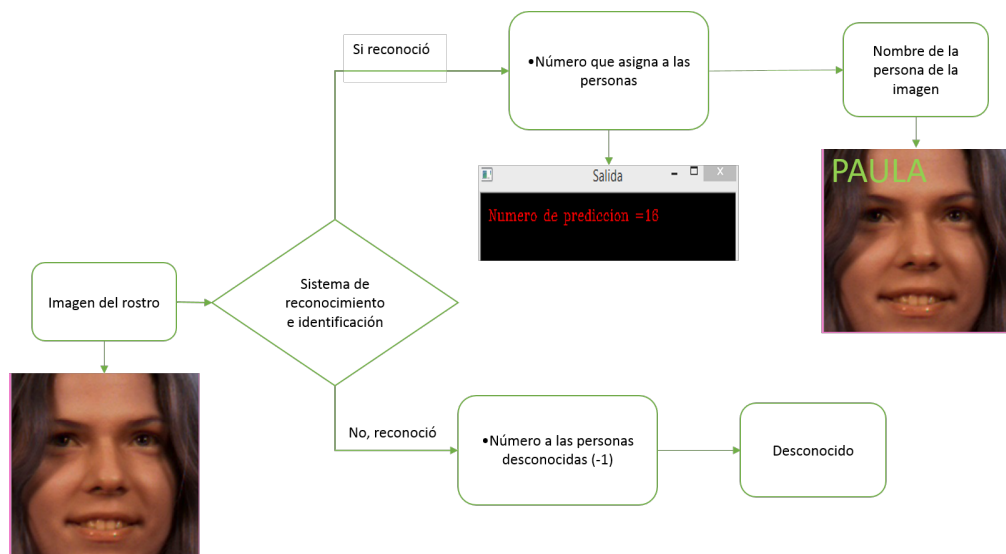


Figura 28. Estructura de sistema de reconocimiento e identificación

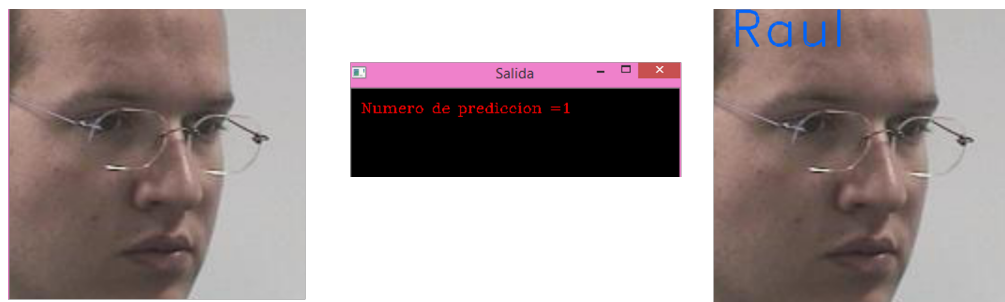


Figura 29. Sistema de reconocimiento a) imagen de entrada b) numero se predijo c) nombre de la persona

Para el desarrollo del sistema de reconocimiento e identificación se investigaron diversos métodos que se emplean para reconocer rostros como se ve en los antecedentes (sección

4), pero según el estudio y las pruebas realizadas el sistema que mejor se adapta a los objetivos propuestos es el Eigenfaces, debido a que puede reconocer rostros con cambios de iluminación y en diferentes posiciones, además el tiempo de reconocimiento es mínimo y se puede emplear en tiempo real.

7.4.4. Modelo de reconocimiento e identificación

En esta sección se explicará la estructura y el funcionamiento del sistema de reconocimiento e identificación realizado en este proyecto, basándose en el método Eigenfaces [32]. Como se ha mencionado en el marco conceptual (sección 5.2) el Eigenfaces es un método que consiste en tomar una imagen digital, de “a” filas y “b” columnas, a lo que se transforma en un vector unitario contenido en un espacio de imágenes de n-dimensional ($n = a \times b$) (ver la Figura 16). Luego se sustrae la imagen promedio y se proyecta el vector resultante en un sub-espacio de menor dimensión utilizando el método de reducción de dimensión PCA. Esa proyección se realiza mediante la imagen de un rostro en el sub-espacio formado por los Eigenfaces y compara con los rostros conocidos.

Los Eigenfaces son conjuntos de vectores representados gráficamente, convirtiéndose en una variación de la imagen. Estos vectores se obtiene al utilizar el PCA a la matriz de covarianza de un conjunto de imágenes de rostros, denominados eigen-vector, siendo una imagen tratada como un vector en un espacio multidimensional. A su vez, cada rostro individual puede ser representada exactamente en términos de una combinación lineal de las Eigenfaces, la cual puede ser aproximada usando solamente los “mejores” Eigenfaces que son las que tienen mayores Eigenvalores. Los M mejores Eigenfaces conforman un sub-espacio M-dimensional de los rostros de todas las posibles imágenes. Lo anterior quiere decir que el objetivo de PCA es encontrar los vectores que mejor almacenen la distribución de las imágenes de rostros en el espacio completo de imágenes. Estos vectores definen el sub-espacio de imágenes.

Una explicación mas detallada de de los pasos realizados para la elaboración del sistema de reconocimiento e identificación como se ve en la Figura 30.

Pre-procesamiento Como se ha mencionado con anterioridad, la eficiencia de los Eigenfaces reside en la capacidad que tiene para caracterizar con un mínimo de error una imagen del rostro. Esa capacidad es alta cuando es con el conjunto de imágenes

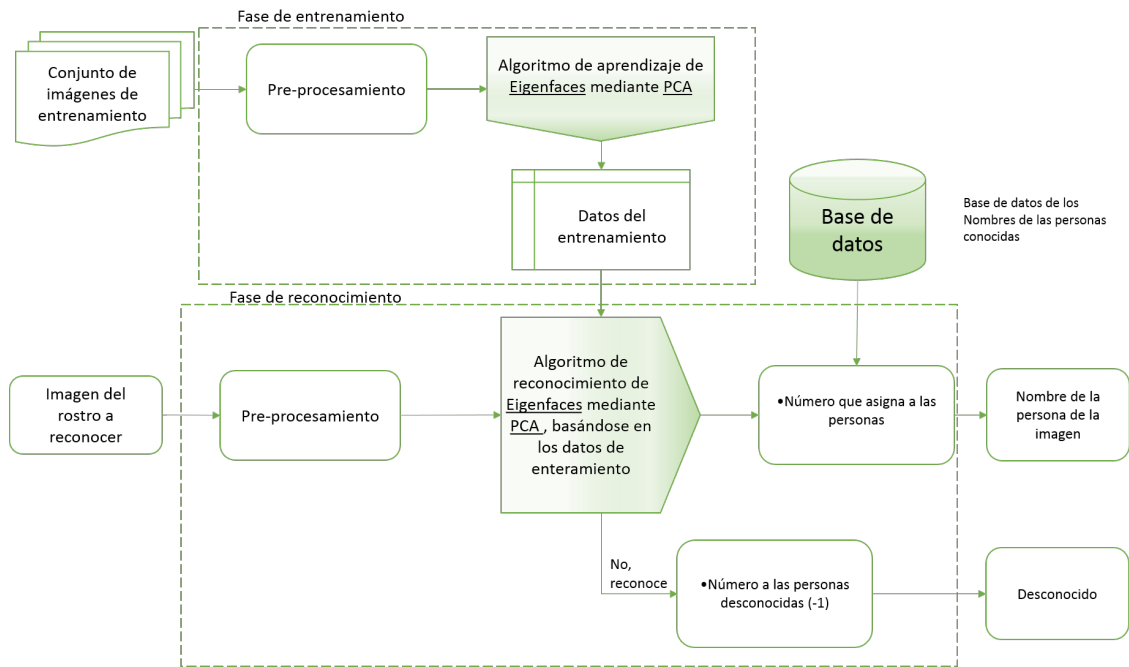


Figura 30. Esquema de la estructura del sistema de reconocimiento e identificación

entrenadas, pero con las otras imágenes a reconocer depende del procesado que se le aplique a la imagen entrada (que es solo el rostro sin elementos externos para eso fue detectada por el sistema de detección y localización). En este proyecto se pre-procesa la imagen en los siguientes aspectos.

1. Escalar la imagen: se normaliza la altura y el ancho del rostro, es preferible que se convierta al mismo tamaño de las imágenes de los rostros entrenados. En este proyecto el tamaño de la imagen entrenadas es de 168x192 pixeles.
2. Contraste y nivel de iluminación: se expande el rango de niveles de gris de la imagen al máximo, en otras palabras hace que el valor mínimo que puede tener los pixeles es de cero(0) y el máximo 255.

Algoritmo de aprendizaje de Eigenfaces mediante PCA El algoritmo de aprendizaje es aquel que le enseña al sistema los rostros de las personas conocidas, y con estás se puedan comparar el rostro a reconocer, el conjunto de imágenes conocidas se le denominara conjunto de entrenamiento. Es necesario recalcar que un buen reconocimiento se basa en que se tenga un abundante numero de imágenes de entrenamiento

por persona, esas imágenes tiene que tener cambios de iluminación, posición del rostro, expresiones faciales etc. En el conjunto de entrenamiento para este proyecto contiene por persona 24 imágenes diferentes. Como se ve en la Figura 31 se encuentra imágenes de entrenamiento de los sujeto 1 y 16, presentando al primero con cambios de posición de rostro y expresiones faciales, mientras que el sujeto 16 tiene cambios de iluminación y expresiones faciales.



Figura 31. Tipos de imágenes a entrenar para el sistema de reconocimiento

En este proyecto se plantea evaluar el sistema de reconocimiento e identificación, con rostros rotados y cambios de iluminación por lo tanto se seleccionó 15 personas con rotación del rostro y 15 con cambios de iluminación, de las 115 personas elegidas de las bases de datos MUCT y Head Posepor (ver la sección 7.1). Las 115 personas son de diferentes edades, razas, culturas además pueden tener oclusiones en el rostro como gafas o turbantes, etc. Cabe destacar que que las imágenes entrenadas y evaluadas son totalmente diferentes.

Teniendo el conjunto de imágenes de rostros con los que se va a entrenar (T_i), se pasa a encontrar los vectores que mejor describan la distribución de la información del rostro usando el Análisis de Componentes Principales (PCA). Realizando los siguientes pasos (ver la sección 5.2.1).

1. Se adquiere el promedio de conjunto de imágenes de entrenamiento y luego se cal-

cula la diferencia, entre el vector que representa cada rostro con el rostro promedio(ver la ecuación 14).

2. Se encuentran los Eigenvectores y Eigenvalores con la matriz de covarianza ver la ecuación 15
3. Dado que en la practica no es necesario utilizar los M vectores principales para construir las Eigenfaces sino un conjunto un poco menor $M' < M$. Para esto se debe seleccionar los M' vectores propios más significativos, por lo cual se ordena los vectores propios de mayor a menor y se utiliza tan solo los M' primeros vectores propios que son los que forman el espacio de rostros. Estos vectores pueden ser redimensionados al tamaño original de la imagen $N \times N$ y es lo que se conoce como Eigenfaces.
4. Se obtiene unas nuevas imágenes proyectándolas en el espacio del rostro con la ecuación 19.

Ya teniendo los Eigenfaces de todas las imágenes y los datos necesarios en el entrenamiento se puede realizar el reconocimiento de cualquier persona entrenada.

Algoritmo de reconocimiento de Eigenfaces mediante PCA Para el proceso de reconocimiento es necesario primero tener el conjunto de imágenes a reconocer que sean de las mismas personas entrenadas, pero que sean diferentes a las entradas. Para este proyecto se evaluarán por persona entre 12 a 16 imágenes con diferentes características.

Los pasos que el sistema de reconocimiento e identificación tiene que hacer para reconocer a las personas que se encuentran en la imagen es:

- 1 paso Ejecutar el algoritmo de entrenamiento anteriormente descrito.
- 2 paso Con la imagen de entrada se realiza el pre-procesamiento y se calcula los Eigen-vectores mediante la matriz de covarianza con las ecuaciones explicadas en la sección 5.2.1. Es decir que el proceso realizado con las imágenes de entrenamiento también se realiza con la imagen de entrada.
- 3 paso Una vez obtenidos los vectores característicos se comparan las distancias entre el vector que representa a la imagen original con las imágenes de entrenamiento.

4 paso Establecido un umbral de reconocimiento (UR) a priori, si el menor valor del paso anterior, es menor que el UR, la imagen del rostro de entrada es considerada como conocida, si es mayor, se considera desconocida. Este umbral se selecciona dependiendo de cada sistemas y cantidad de personas entrenadas.



Figura 32. Reconstrucción de la imagen de entrada con Eigenface

Al realizar el proceso anterior se puede observar que el resultado de los Eigenfaces en las imágenes es que aparecerán áreas claras y oscuras que disponen de un patrón específico (como se ve en la Figura 33 las 6 primeras Eigenface), y en la Figura 32 se ve la reconstrucción de la imagen de entrada con el Eigenfaces. El patrón son las característica mas destacadas del rostro para ser evaluados y almacenados.



Figura 33. Representación de las 6 primeras Eigenfaces

8. RESULTADOS

En esta sección se analizará y seleccionará el sistema de detección y reconocimiento de rostro que cumpla con los objetivos propuestos. Para el análisis de este tipo de sistema se realiza a través de pruebas y evaluaciones con diferentes parámetros y ambientes, los cuales ayudan a identificar el sistema mas adecuado. El criterio para seleccionar el sistema se basará en los resultados obtenidos de las evaluaciones realizadas.

8.1. Sistema de detección y localización

El modelo usado para la elaboración del sistema de detección y localización es el de Viola y Jones como se menciona en el modelo de detección y localización (Sección 7.4.2). Dado que el algoritmo contiene diferentes parámetros los cuales son imprescindibles para una buena detección se busca seleccionarlos dependiendo los resultados obtenidos en los diagramas de *precision-recall* (PR) y las evaluaciones a realizar.

El sistema de detección se realizó en dos partes, la primera en el entrenamiento y la segunda el clasificador de rostros, considerando que uno depende del otro, se evalúa como un sistema y no como secciones independientes.

8.1.1. Algoritmo de entrenamiento de clasificadores

El aprendizaje supervisado realizado en este proyecto para la detección y localización de rostros, se realiza basados en los estudios de Viola y Jones de clasificadores en cascada de Adaboost (como se explica en el la sección 7.4.2).

Como se sabe el clasificador en cascada en cada etapa selecciona las características Haar mas relevantes y ademas descartan las imágenes negativas, para esto se definió una tasa de fallo y una tasa de detección. Para la tasa de falsos positivos en cada etapa se definió de un 0.5, considerando que al tener una tasa baja de falsos positivos, aumentara la tasa de detección y reducirá la cantidad de detecciones erróneas. La tasa de detección se definió por el porcentaje establecido en el algoritmo de Viola y Jones, el cual recomienda que sea un 95 % para que cumpla con la detección deseada. Con estos parámetros el clasificador tiene un criterio el cual entrena a partir de que se cumpla con el objetivo

del falso positivo y tasa de detección. Por ejemplo en la primera etapa del clasificador se selecciona 2 características mas relevantes (que es la de los ojos y la boca) y descarta toda las imágenes negativas que no cumplan con esas dos características, para pasar a el siguiente clasificador el primero tiene que cumplir con el 50 % de falsos positivos en el caso de que no lo cumpla se añade mas características hasta que cumpla con el objetivo o no se tenga mas características, esto realiza en cada etapa del clasificador. En el caso que cumpla los requisitos en el siguiente clasificador descarta las imágenes que el primer clasificador dejo pasar, pero ademas se le añade mas características que representen el rostro y así sucesivamente hasta que se alcance en numero de etapas determinados. Razón por la cual es necesario determinar la cantidad de etapas para seleccionar las características y poder descartar el total de las sub-ventanas negativas. En este proyecto se estableció tres valores para las etapas y así compararlas, se espera que con mayor etapas en el entrenamiento mas calidad en la detección. Al realizar el entrenamiento se observo que si es mayor de 20 etapas se presenta un sobre-entrenamiento, razón por la cual se establecieron tres etapas menores a 20 para analizar su desempeño y establecer si la hipótesis antes mencionada es correcta, las tres etapas son 10, 15 y 20 etapas.

Como se menciona con anterioridad para la elaboración del entrenamiento se usa un conjunto de imágenes negativas (4000 imagenes negativas) y un conjunto de imágenes positivas (con rostros) (ver Figura 27), este conjunto consintió en 3000 imágenes de las bases de datos FDDB y AFLW, donde se seleccionaron 1813 y 1187 imágenes respectivamente, teniendo en cuenta que solo se tenga el rostro de una persona, señalado, escalado y alineado con base de 10×10 pixeles, 24×24 pixeles y 38×38 pixeles. Según el método de Viola y Jones se usa la escala de 24×24 pixeles usualmente, pero en este proyecto se evaluara los tres tipos de escala con el objetivo de analizar el comportamiento del sistema al cambiar el tamaño de la imagen a entrenar. Otro parámetro que se evaluó es el cambio de las características Haar por las características LBP dado que según Zhang en [19] el numero de características es menor que las características Haar así que el proceso de entrenamiento es menor. Con todos los parámetros establecidos se busca analizar el tiempo de procesamiento como se presenta en la Tabla 6.

Se observa que la diferencia en el tiempo que toma el entrenamiento con características Haar y LBP es muy mínima, a pesar de que las LBP utilice menos características. También se observa que al aumentar las etapas del clasificador y el tamaño de las imágenes aumenta considerablemente el tiempo de entrenamiento.

Tabla 6: Tiempos de entrenamiento

Características Haar			Características LBP			Diferencia de tiempo (%)
Tamaño Imagen	Etapas del clasificador	Tiempo (d:h:m)	Tamaño Imagen	Etapas del clasificador	Tiempo (d:h:m)	
10x10	10	00:00:16	10x10	10	00:00:11	31.25
	15	00:00:21		15	00:00:22	-4.76
	20	00:03:40		20	00:03:10	13.63
24x24	10	00:08:08	24x24	10	00:9:00	-10.65
	15	00:17:02		15	00:16:46	1.56
	20	00:30:05		20	00:30:02	0.17
38x38	10	01:02:28	38x38	10	00:24:33	7.24
	15	03:15:02		15	03:16:45	-1.97
	20	05:05:26		20	04:20:10	7.38

8.1.2. Algoritmo de Detección y localización

Al realizar el modelo diseñado para el sistema de detección y localización de rostros se determinó que es necesario establecer la cantidad y el tipo de imágenes a evaluar para poder seleccionar el sistema y cumplir con los objetivos propuestos. Razón por la cual con la bases de datos *Face Detection Data Set and Benchmark* (FDDDB) se escogió estratégicamente imágenes que contengan una cantidad especifica de personas para evaluar (ver Tabla 7), cabe mencionar que las imágenes usada para entrenar y para evaluar son totalmente diferentes, considerando que para el entrenamiento se usaron 1813 imágenes de la base de datos FDDDB , en la prueba se usaran 1032 imágenes. Además a cada grupo de imágenes por cantidad de personas se le incluyeron imágenes sin rostros para evaluar si el sistema detecta un falso positivo (como se ve en la Tabla 7).

Tabla 7: Imágenes a evaluar

numero de de personas	numero de rostros	Imágenes con rostro	Imágenes sin rostro
1	332	332	32
2	980	490	40
3	435	145	21
4	200	50	7
5	76	15	12
total	2023	1032	112

Dado que evaluar todas las opciones del sistema de detección y localización con las 1032 imágenes seleccionada tendría un coste computacional alto y sería ineficiente, por lo tanto por cada conjunto de imágenes por personas se selecciona aleatoriamente una cantidad determinada para que se obtuviera 100 imágenes con 277 personas en ellas, para evaluar la calidad del sistema. A las 100 imágenes es necesario establecer en que lugar se encuentran los rostros, en otras palabras se debe dibujar una margen en el rostro manualmente, con el fin de compararlo con el rectángulo que se genere con la detección, por ejemplo en la Figura 34 el recuadro de color azul oscuro es el que se realizo manualmente, mientras que el recuadro verde es realizado por el detector, por lo tanto para esta imagen se presenta 2 VP, 1 FN y 1FP. los 2 VP son aquellos rectángulos verdes que se traslapa sobre el azul y que el solapamiento sea mayor al 50 %, por lo tanto el primer recuadro verde (de izquierda a derecha) solo se traslapa un 10 % al recuadro azul, dando así un falso negativo por no detectar el rostro y un falso positivo por detectar algo que no es un rostro.

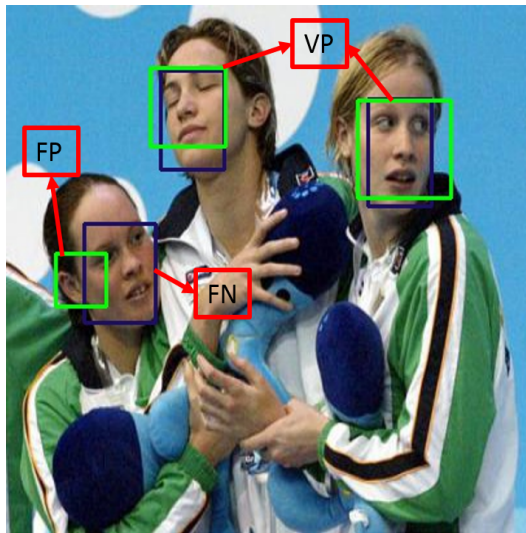


Figura 34. Imagen con detección y localización de rostro

En las 100 imágenes seleccionadas se puede presentar rostros en diferentes tamaños y posiciones, es razón por la cual al detector se le establecen dos parámetros importantes que al modificar los se tiene mas posibilidad de detectar esas derivaciones que tienen los rostros de las imágenes, el primero es el tamaño de la sub-ventana descriptor, como se menciono está se desplaza por toda la imagen detectando rostros y dado que se puede presentar en diferentes tamaños el rostro se establece el factor escalar el cual



Figura 35. Recorrido en la detección a) primer recorrido b) segundo recorrido c) tercer recorrido d) suma de todos los recorridos e) resultado de toda la detección

indique en que porcentaje se modifica el tamaño de la sub-ventana que evalúa los rostros. El segundo parámetro depende de cuantas veces una sub-ventana se clasifico como rostro, dado que con el anterior parámetro cambia de tamaño el descriptor hay una alta posibilidad de que se obtenga múltiples detecciones de un solo rostro o falsos positivos, por ejemplo en una imagen que contiene un rostro (ver la Figura 35), en el primer recorrido se estipulo que en la imagen contenía tres rostros, en el segundo recorrido se detecto dos rostro y el tercero se detecto uno, dado que termino de hacer todos los recorridos se suman las sub-ventanas que detectan el mismo objeto y si es igual o mayor al valor establecido, se clasifica como rostro y así saber cuantos rostros contiene la imagen y en que ubicación, como se ve en la Figura 35. Este parámetro denominado como Vecinos nos ayuda a disminuir los falsos positivos que se puedan presentar en la detección.

Teniendo en cuenta que asignar valores a los parámetros mencionados no se asegura que el sistema sea el mejor, por tal razón se evaluó con un conjunto de numero los cuales para, el Factor de escala son 1.03, 1.04, 1.08, 1.1, 1.2, 1.3, 1.5 y 1.8 y para Vecinos son 3, 5, 7, 9 y 11. A priori a establecer los valores de los parámetros se realizo una prueba en el algoritmo de detección para saber los limites del Factor de escala y los Vecinos, en que se desempeña correctamente, con lo limites se establecieron los valores con los cuales se pueden realizar un análisis completo de como se desempeña el sistema .

Con todos los parámetros establecidos se combinan para así analizar y seleccionar el sistema óptimo para la detección y localización de rostros a través del método de evaluación PR (ver ecuaciones 35 y 36), con las 100 imágenes seleccionadas. En la Tabla 8, se presenta un resultado del sistema con un entrenamiento con características Haar, el

tamaño de la imagen 38×38 pixeles y con 20 etapas, a este entrenamiento se evaluó con diferentes factores de escala y vecinos, además se presenta por cada sistema el resultado de VP, FP y FN y el cálculo *recall* y precisión.

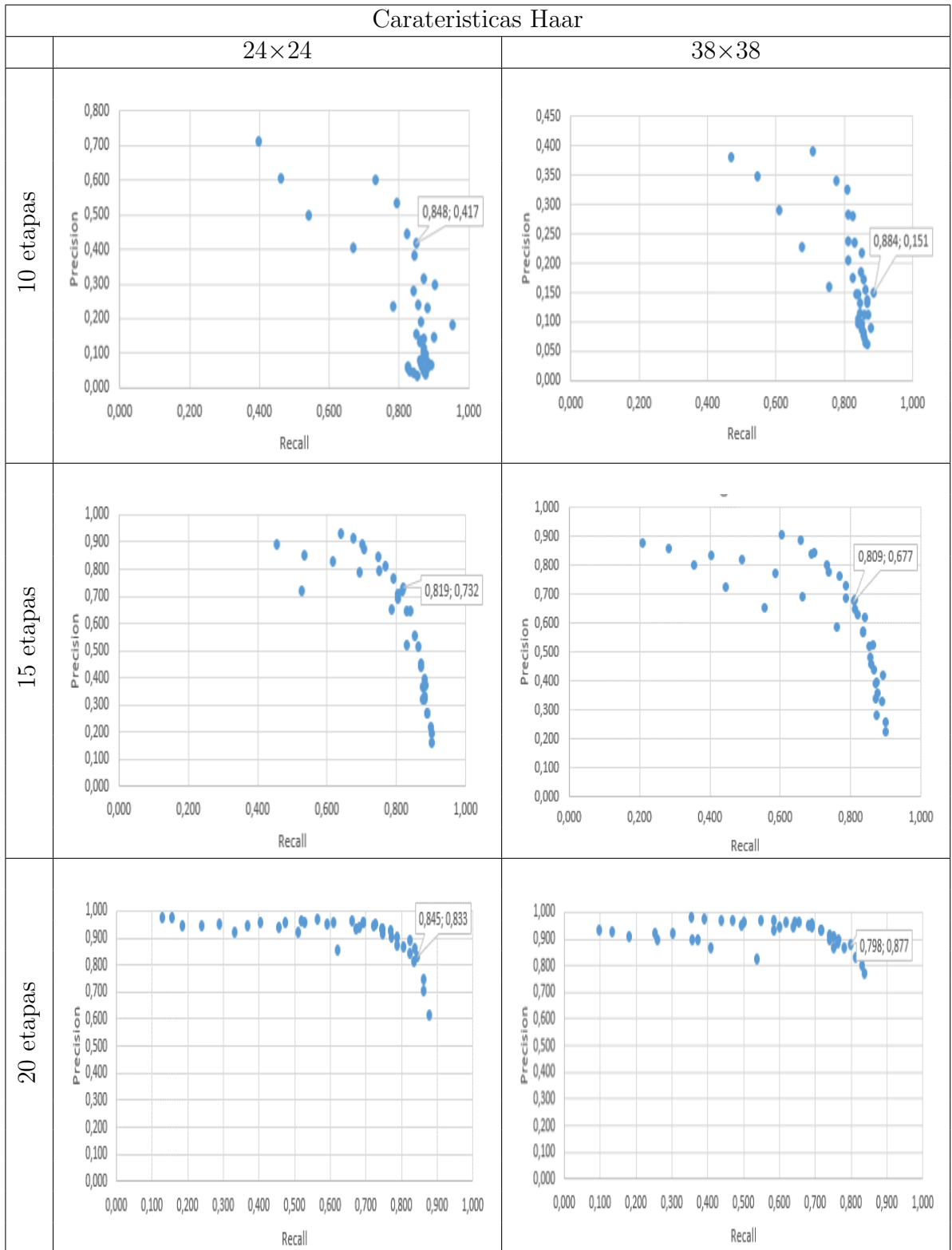
Tabla 8: Evaluación al sistema de detección entrenamiento con características Haar, tamaño de imagen 38 por 38 pixeles y 20 etapas del clasificador

factor de escala	Vecinos	VP	FP	FN	Recall	Precision
1.03	3	232	69	45	0,838	0,771
1.03	5	225	46	52	0,812	0,830
1.03	7	216	33	61	0,780	0,867
1.03	9	211	27	66	0,762	0,887
1.03	11	208	20	69	0,751	0,912
1.04	3	230	58	47	0,830	0,799
1.04	5	221	31	56	0,798	0,877
1.04	7	212	24	65	0,765	0,898
1.04	9	205	19	72	0,740	0,915
1.04	11	199	14	78	0,718	0,934
1.08	3	208	32	69	0,751	0,867
1.08	5	198	14	79	0,715	0,934
1.08	7	189	10	88	0,682	0,950
1.08	9	181	7	96	0,653	0,963
1.08	11	178	7	99	0,643	0,962
1.1	3	205	24	72	0,740	0,895
1.1	5	191	11	86	0,690	0,946
1.1	7	177	10	100	0,639	0,947
1.1	9	171	7	106	0,617	0,961
1.1	11	162	5	115	0,585	0,970
1.2	3	191	9	86	0,690	0,955
1.2	5	166	9	111	0,599	0,949
1.2	7	152	5	125	0,549	0,968
1.2	9	139	5	138	0,502	0,965
1.2	11	130	4	147	0,469	0,970
Sigue en la página siguiente.						

factor de escala	Vecinos	VP	FP	FN	Recall	Precision
1.3	3	162	12	115	0,585	0,931
1.3	5	137	7	140	0,495	0,951
1.3	7	121	4	156	0,437	0,968
1.3	9	108	3	169	0,390	0,973
1.3	11	98	2	179	0,354	0,980
1.5	3	149	31	128	0,538	0,828
1.5	5	113	17	164	0,408	0,869
1.5	7	99	11	178	0,357	0,900
1.5	9	84	7	193	0,303	0,923
1.5	11	70	6	207	0,253	0,921
1.8	3	103	12	174	0,372	0,896
1.8	5	72	8	205	0,260	0,900
1.8	7	50	5	227	0,181	0,909
1.8	9	37	3	240	0,134	0,925
1.8	11	27	2	250	0,097	0,931

Cada entrenamiento se evaluó con los parámetros de Vecinos y Factor escalar, para obtener en cada uno los resultados y los cálculos de PR como se ve en la Tabla 8, se gráfico precision vs recall para establecer que sistema es el mejor, los sistemas a seleccionar son aquellos mas cercanos a la esquina superior derecha como se ve en la en las gráficas de la Tabla 9. Por ejemplo en la gráfica de entrenamiento con característica Haar, de 38×38 pixeles y de 20 etapas, el punto mas alto es el de precision de 0,877 y recall de 0,7978, al tener ese punto se busca en la Tabla 8, para obtener esos resultados el sistema trabaja con un factor de escala de 1.04 y vecinos de 5 (como esta resaltado en la Tabla 8), dado que hay mas puntos cercanos a el valor ideal (precision = 1 y recall = 1) se buscan otros dos mejores para tener mas opciones a elegir. Estos pasos se realizan a cada uno de los sistemas evaluados.

Tabla 9: Gráficas de Precisión y recall de entrenamiento con características Haar



Teniendo todas las gráficas se pudo observar que al aumentar el tamaño de la imagen de entrenamiento se presenta una detección con mayor precisión, en vista de que al entrenar con tamaños grandes se tiene mayor detalles el rostro que con pequeñas, por ejemplo los entrenamientos con características Haar y 20 etapas, se presenta un tasa de detección máxima del 55 %, cuando el tamaño de las imágenes de entrenamiento es 10×10 pixeles, imágenes de 24×24 pixeles la tasa de detección es de 86 %, pero la detección de FP es alta, en cambio las imágenes de 38×38 pixeles la tasa de detección tiene mayor precisión dado que FP son bajos y la tasa de detección es del 80 %. Cabe subrayar que estos datos son con la máxima detección no con el mejor sistema según los diagramas PR.

Con la Tabla 9 también se puede concluir que a mayor etapas el entrenamiento mejor es la detección, pero si se realiza con muchas etapas tendrían un problema de sobre-entrenamiento, ya que al tener mas etapas en el clasificador se están evaluando mas características Haar que las específicas para definir el rostro humano, obteniendo como resultado una mala detección. Por lo tanto es esencial saber cual es el límite, para este proyecto fueron 20 etapas. Además se presenta que al aumentar el factor de escala y los vecinos, se disminuye la tasa de detección y los falsos positivos, como se ve en la Tabla 8, dicho hasta aquí, el sistema que se busca debe tener una factor de escala con un valor bajo, considerando que al cambio de tamaño de la sub-ventana sea mínimo para poder detectar en todos los tamaños los rostros, igualmente si el parámetro de Vecinos es de un valor bajo la detección tiene menos restricciones al determinar que es un rostro, no obstante también se debe tomar en cuenta la curva del precision recall, dado que con ella se puede saber si el sistema presenta el mejor desempeño respecto a las detecciones realizadas.

Con el análisis anterior se seleccionó 8 sistemas que cumplen con los criterios establecidos, como se ve en la Tabla 10. Con los resultados mostrados en la Tabla 10 de los 8 sistemas, no se puede asegurar que alguno de esos sistemas es el mejor al detectar múltiples rostros en una imagen digital, por lo tanto se realizaron otras pruebas con cada conjunto de imágenes que contienen de uno a cinco rostros por imagen (ver Tabla 7), para analizar si al variar la cantidad de rostros por imagen tienen una alta tasa de detección.

Al realizar la prueba con los 8 sistemas seleccionados con todas las imágenes se presentó resultados de cada grupo de personas como se observa en la Tabla 11, que tiene el

Tabla 10: Mejores resultados de los sistemas de detección

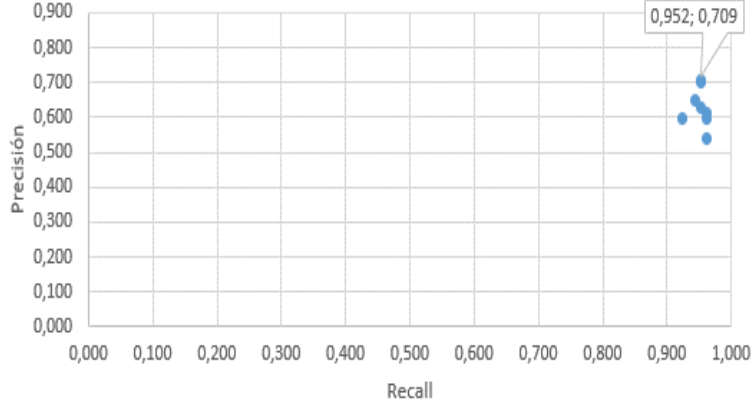
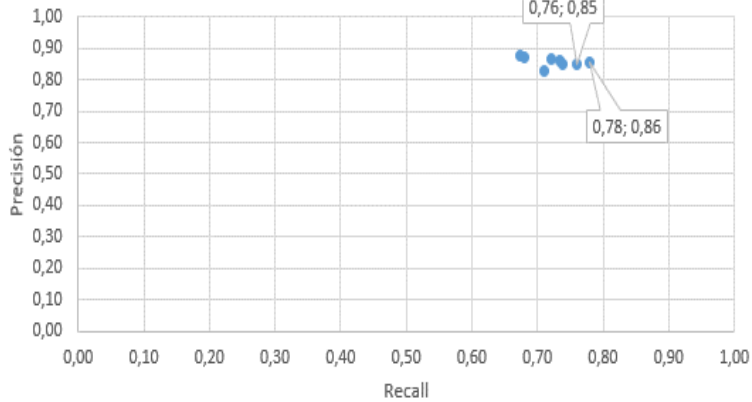
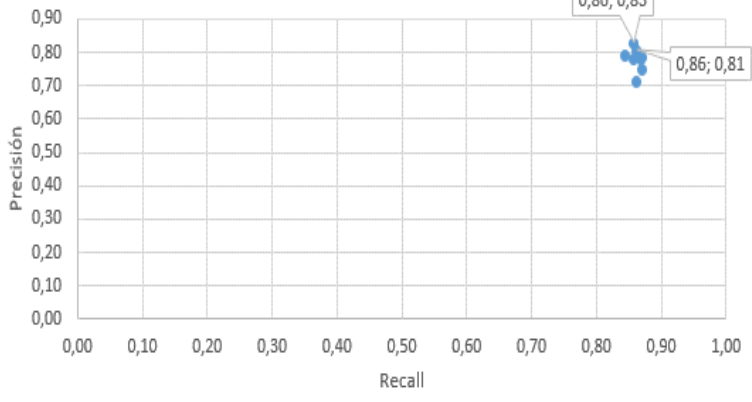
Características	tamaño de imagen	etapas	factor de escala	vecinos	PP	FP	FN	Recall	precision
Haar	24	20	1.03	5	239	80	38	0.863	0.749
Haar	24	20	1.03	7	234	47	43	0.845	0,833
Haar	38	20	1.03	3	232	69	45	0.838	0.771
Haar	38	20	1.04	3	230	58	47	0.830	0.799
LBP	24	20	1.03	3	229	54	48	0.827	0,809
LBP	24	15	1.08	9	233	63	44	0.841	0.787
LBP	38	20	1.03	3	236	93	41	0.852	0.717
LBP	38	20	1.04	3	232	74	45	0.838	0,758

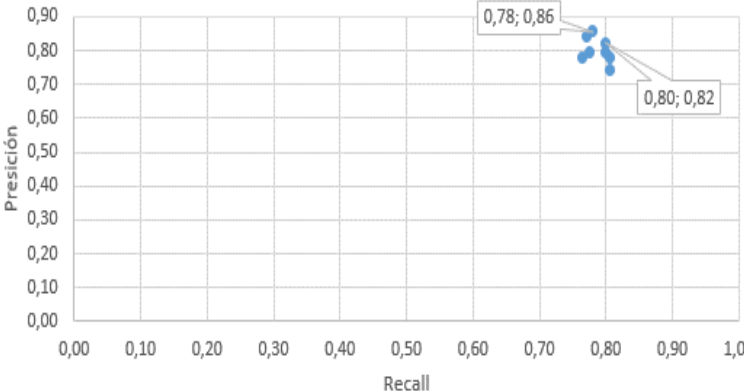
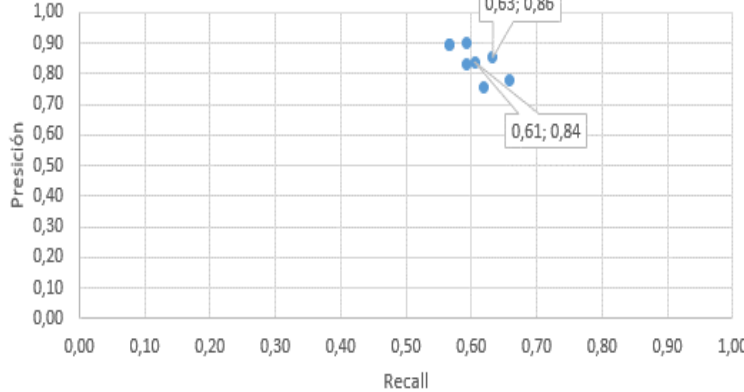
Tabla 11: Resultados de la evaluación con la imágenes que contiene 2 personas a los sistemas seleccionados

	tamaño de imagen	etapas	factor de escala	vecinos	PP	FP	FN	IN	Recall	precision
Haar	24	20	1,03	5	853	401	127	36	0,68	0,87
Haar	24	20	1,03	7	834	264	146	37	0,76	0,85
Haar	38	20	1,03	3	843	307	137	38	0,73	0,86
Haar	38	20	1,04	3	838	238	142	39	0,78	0,86
LBP	24	20	1,03	3	831	293	149	38	0,74	0,85
LBP	24	15	1,08	9	813	332	167	38	0,71	0,83
LBP	38	20	1,03	3	858	414	122	38	0,67	0,88
LBP	38	20	1,04	3	850	330	130	39	0,72	0,87

resultado de detección al grupo de imágenes que contiene 2 personas por imagen, dado que con la Tabla solo podemos observar todas las detecciones, PP, FP, FN y imágenes negativas (IN imágenes que no tienen rostro), con esta se puede establecer cual es el mejor sistema para este proyecto, en contraste con la Tabla se realizo las gráficas de PR para asegurarse de elegir el mejor sistema.(ver la Tabla12).

Tabla 12: Diagramas PR de los sistemas seleccionados con cada grupo de imágenes

Cantidad de personas por imagen	Diagrama PR
1 Persona	 <p>Diagrama PR para 1 Persona. El eje horizontal es Recall (0,00 a 1,00) y el eje vertical es Precisión (0,00 a 0,90). Se muestra un grupo de puntos azules en la parte superior derecha, indicando un alto rendimiento. Un punto específico está etiquetado con el par de valores (0,952; 0,709).</p>
2 Persona	 <p>Diagrama PR para 2 Personas. El eje horizontal es Recall (0,00 a 1,00) y el eje vertical es Precisión (0,00 a 1,00). Se muestra un grupo de puntos azules en la parte superior derecha. Dos puntos están etiquetados con los pares de valores (0,76; 0,85) y (0,78; 0,86).</p>
3 Persona	 <p>Diagrama PR para 3 Personas. El eje horizontal es Recall (0,00 a 1,00) y el eje vertical es Precisión (0,00 a 0,90). Se muestra un grupo de puntos azules en la parte superior derecha. Dos puntos están etiquetados con los pares de valores (0,86; 0,83) y (0,86; 0,81).</p>
<p>Sigue en la página siguiente.</p>	

Cantidad de personas por imagen	Diagrama PR
4 Persona	
5 Persona	

Con las gráficas de PR de todas las imágenes deducimos que el sistema que cumple con los objetivos es el entrenado características Haar, el tamaño de la imagen de entrenamiento de 24×24 píxeles y de 20 etapas, además el factor de escala es de 1,03 y con 7 vecinos. Este sistema se seleccionó puesto que al evaluarlo con todos los grupos de imágenes tiene los mejores resultados a comparación de los otros sistemas, aunque con 4 personas no fue uno de los mejores, pero cumple con el objetivo de detectar en imágenes rostros una y varias personas con diferente iluminación y posición.

Se puede concluir que al tener una precisión alta se baja la tasa de falsos positivos, y la tasa de detección aumenta cuando el recall aumenta. Además se evidenció que el método de Viola y Jones es efectivo al momento de detectar múltiples personas en diferentes posiciones e iluminación diferente considerando que se evaluó con otros

parámetros diferentes.

8.2. Sistema de Reconocimiento e Identificación

En esta sección se mostraran los resultados del rendimiento de los métodos de reconocimiento e identificación mas utilizados actualmente, que se han estudiado a lo largo del proyecto (ver secciones 7.4.4 y 5.2.1). Con los resultados de los diferentes métodos se busca determinar cual es el que tiene mejor desempeño. Considerando que se busca un sistema que tenga una alta tasa de reconocimiento con imágenes que contengan rostros de personas en diferente posiciones, con cambios de iluminación y diferentes expresiones faciales, ademas que se busca tener el menor tiempo posible al reconocer las diferentes personas.

Los métodos implementados para evaluar son: Eigenface, Fisherface y LBPH. Como se ha mencionado anteriormente Eigenface es un método que usan la reducción de dimensionalidad para encontrar vectores que mejor representen la distribución de un grupo de imágenes, Fisherface se basa en el método de análisis lineal discriminante de Fisher (LDA) para maximizar la relación inter/intra clase donde una clase corresponde a una persona. Para estos dos últimos se les establecen la cantidad de componentes para la comparación en el instante del reconocimiento, para este proyecto se seleccionaron por medio de pruebas realizadas a priori a las pruebas de selección, puesto que se observo en esas pruebas a priori que con otros valores de componentes la tasa de reconocimiento es menor como se ve en la Figura 36, que para el algoritmo Eigenfaces con el componente 50 tiene mejor tasa de reconocimiento . El método LBPH se basa en la extracción de características geométricas, de cada imagen asignándoles un nivel de gris determinado que posteriormente es representado en un histograma.

Teniendo claro los métodos a implementar, se establecen los parámetros para evaluarlos: el primer parámetro es seleccionar a 115 personas de diferentes edades, razas, culturas, que en sus imágenes contengan cambios de iluminación, poses y expresiones. Seleccionado de las base de datos Head Posepor y MUCT hay que tener en cuenta que las dos bases de datos no tienen la misma cantidad de imágenes por personas por lo tanto se selecciono por persona una cantidad determinada de imágenes y así establecer un criterio para el entrenamiento y evaluacion del sistema .

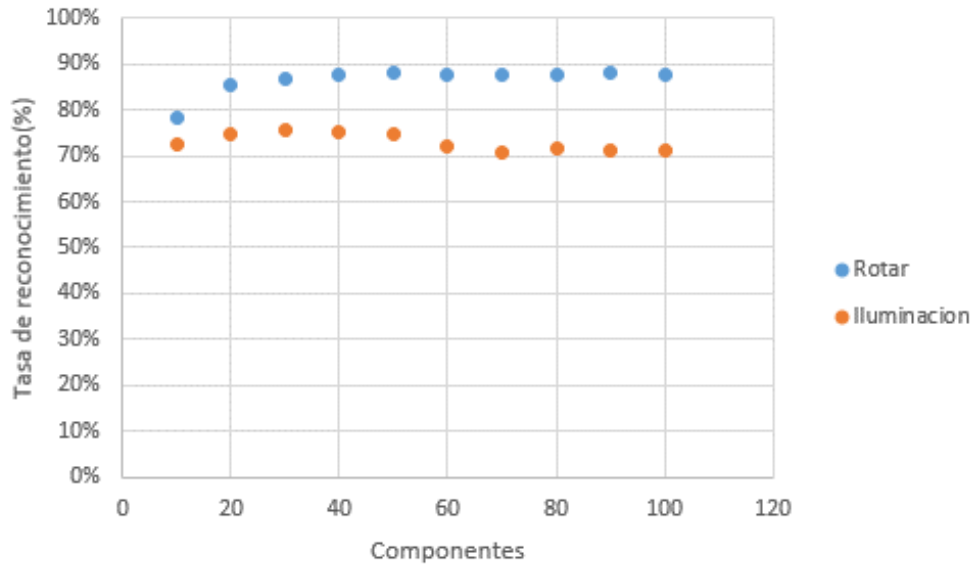


Figura 36. Variación de componentes para el algoritmo Eigenfaces

Por persona se eligen aleatoriamente 24 imágenes para realizar el entrenamiento y entre 12 y 16 imágenes para evaluar, teniendo en cuenta que aproximadamente 65 % de los datos es para entrenar y el 35 % es para evaluar. En vista de que con las imágenes de 115 personas no se puede analizar como se desempeña el sistema con cambios de posición en el rostro, y con cambios de iluminación, por separado, por lo cual, primero se evaluara con 15 personas que contengan imágenes con cambios de posición y 15 personas con imágenes con diferente iluminación. El siguiente parámetro a establecer son las dimensiones de las imágenes a reconocer, teniendo presente que el sistema tenga la capacidad de reconocer rostros en diferentes dimensiones. Se fijaron catorce dimensiones que usualmente se usa para videos y fotografías en las cuales se encuentran: 1280×960 pixeles, 1216×912 pixeles, 1152×864 pixeles, 1024×768 pixeles, 896×672 pixeles, 832×624 pixeles, 768×576 pixeles, 704×528 pixeles, 640×480 pixeles, 512×384 pixeles, 384×288 pixeles, 256×192 pixeles, 192×144 pixeles y 128×96 pixeles. Es necesario recalcar que a todas las imágenes se evaluara con cambio de dimension y con los tres métodos antes mencionados.

8.2.1. Rotación del Rostro

Para evaluar el desempeño de los algoritmos de reconocimiento utilizando imágenes de caras en distintas posiciones, se realiza una prueba en el cual se evalúa la canti-

dad de rostros reconocidos (verdaderos positivos) y los rostros no reconocidos (Falsos positivos), para obtener la tasa de reconocimiento. Se tienen a 15 personas con 360 imágenes para entrenar y 240 imágenes para evaluar teniendo en cuenta el criterio de entrenamiento y evaluación anteriormente establecido, en el cual por persona se usan 24 imágenes para entrenar y para la evaluación 16 imágenes. A esas imágenes se le cambian la dimension y se obtiene el tiempo de reconocimiento. Los resultados se muestran en la Figura 37.

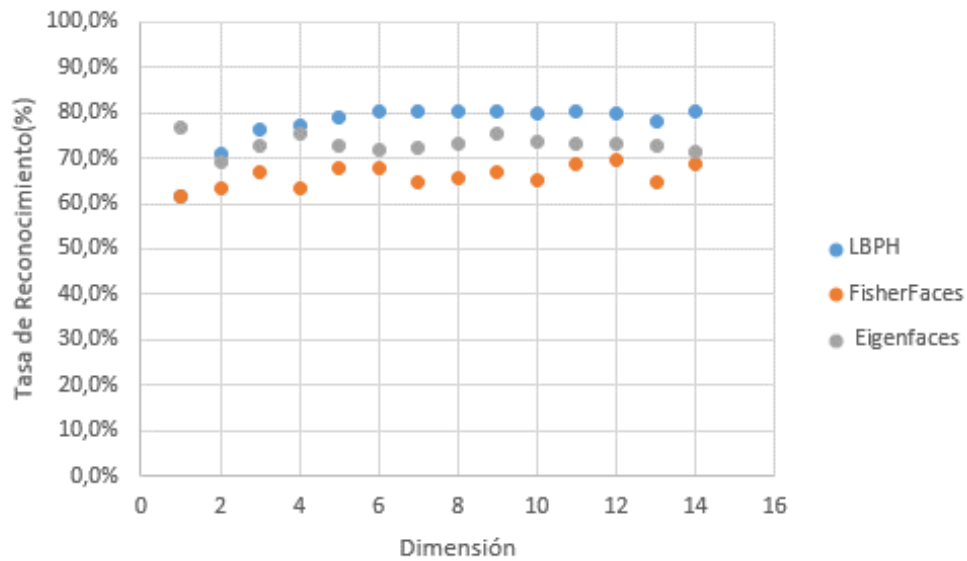


Figura 37. Comparación del desempeño de los algoritmos en función al cambio de dimension sobre las imagenes de la base de datos Head Posepor

Se observa que en la Figura 37 que el método LBPH tiene una tasa de reconocimiento entre el 70 % al 80 %, pero que a medida que se disminuye la dimension de la imagen de entrada también se disminuye la tasa de reconocimiento. El método Eigenfaces obtuvo una tasa de reconocimiento mayor al 70 % este como el Fisherfaces no tuvieron problemas al reconocer los rostros cuando cambiaba las dimensiones. Hay que mencionar que el reconocimiento de LBPH toma tiempo considerable para reconocer el rostro, como se observa en la Tabla 13, por el contrario es el caso de Fisherface y Eigenfaces que tardaron entre 4 ms a 3 ms y 7ms a 8ms respectivamente .

De los resultados se infiere que LBPH tiene mejor tasa de reconocimiento aunque tome más tiempo en reconocerlos que los otros dos métodos. Pero Eigenfaces tiene una tasa de reconocimiento aceptable dado el caso de que la cantidad de imágenes de entrenamiento es muy escasa, a la cantidad de posibles posiciones que puede tener el rostro,

Tabla 13: Tiempos promedios que demora los algoritmos en reconocer los rostros con variación de posición

Tiempo de reconocimiento				Diferencia de tiempos (%)			
Tamaño de Imagen		LBPH (ms)	Eigenface (ms)	Fisherface (ms)	LBPH-Eigenfaces	LBPH-Fisherface	Eigenface-Fisherface
1280	960	191,59	7,68	4,74	95,99	97,52	38,25
1216	912	199,07	7,88	4,26	96,04	97,86	45,89
1152	864	210,49	7,63	3,82	96,38	98,18	49,89
1024	768	202,38	7,68	4,64	96,21	97,71	39,56
896	672	207,21	7,57	3,52	96,35	98,30	53,52
832	624	177,11	7,98	4,25	95,49	97,60	46,75
768	576	201,15	8,37	4,68	95,84	97,67	44,09
704	528	192,03	8,32	3,62	95,67	98,12	56,52
640	480	178,59	7,77	3,54	95,65	98,02	54,40
512	384	203,26	7,54	4,50	96,29	97,79	40,31
384	288	178,92	7,56	4,12	95,77	97,70	45,46
256	192	202,31	7,69	3,39	96,20	98,32	55,89
192	144	73,40	7,52	4,24	89,76	94,23	43,63
128	96	72,12	7,53	4,31	89,56	94,03	42,79

ademas cabe mencionar que el grupo de imágenes a evaluar también presenta cambios de expresiones faciales. Con lo anterior podemos concluir que es necesario implementar una mejor conjunto de imágenes de entrenamiento, con mas imágenes por persona y mas variación de posiciones.

8.2.2. Cambios de iluminación

Se realizo una evaluación a los métodos de reconocimiento con respecto a la variación de iluminación. Dado que Fisherfaces y LBPH son tolerantes a la variación de iluminación se esperaría que estos dos métodos tengan altos niveles de reconocimiento, además el conjunto de imágenes de entrenamiento para este tipo de análisis esta con varios cambios en el brillo y el contraste como se observa en la Figura 31. En Figura 38 se presenta el resultado de los algoritmos con las imágenes con cambios de iluminación.

Como se mencionó anteriormente los dos métodos que presentan una tasa de reconocimiento alta son LBPH y Fisherface mientras que Eigenfaces presenta una tasa de reconocimiento entre 85 % a 90 %. El tiempo de reconocimiento para LBPH es aproxi-

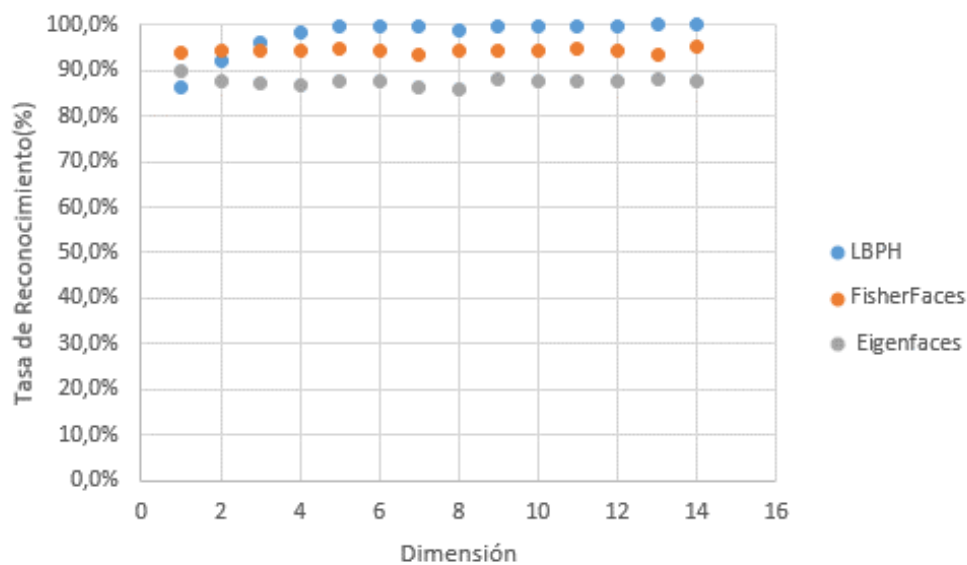


Figura 38. Comparación del desempeño de los algoritmos con imágenes de con variaciones de iluminación

madamente 187 ms y Fisherface tiene un tiempo de 4 ms siendo el método que toma menos tiempo en reconocer los rostros como se ve en la Tabla 14.

Tabla 14: Tiempos promedios que demora los algoritmos en reconocer los rostros con variación de iluminación

		Tiempo de reconocimiento			Diferencia de tiempos (%)		
Tamaño Imagen		LBPH (ms)	Eigenface (ms)	Fisherface (ms)	LBPH-Eigenfaces	LBPH-Fisherface	Eigenface-Fisherface
1280	960	225,29	8,42	3,71	96,26	98,35	55,92
1216	912	234,88	8,93	3,95	96,20	98,32	55,83
1152	864	218,69	8,19	4,07	96,25	98,14	50,31
1024	768	217,63	8,88	3,99	95,92	98,17	55,06
896	672	214,29	10,93	3,93	94,90	98,16	64,00
832	624	185,73	11,07	3,59	94,04	98,07	67,58
768	576	187,49	10,65	3,58	94,32	98,09	66,43
704	528	186,43	8,41	4,36	95,49	97,66	48,12
640	480	208,85	8,15	4,27	96,10	97,95	47,58
512	384	187,19	8,08	3,50	95,69	98,13	56,69
384	288	182,88	8,34	3,54	95,44	98,06	57,51
256	192	183,71	9,65	3,50	94,74	98,10	63,76
192	144	114,27	5,33	1,75	95,34	98,46	67,06
128	96	77,37	5,98	3,47	92,27	95,52	42,02

Con la Figura 38 y la Tabla 14 se puede concluir que el LBPH es el que tiene mejor tasa de reconocimiento, aunque la diferencia de reconocimiento entre Fisherface y LBPH es solamente del 5 %.

8.2.3. Rotación y cambios de iluminación al rostro

Ya teniendo los análisis de como se desempeña los tres métodos con la problemática de reconocer los rostros con cambios de posición e iluminación, se evaluara con las imágenes de las 115 persona, debido a la cantidad de imágenes a entrenar y que el sistema no contiene tanta memoria para almacenar todas las imágenes, se divide 3 grupos de 33 personas y uno de 16 personas para realizar el análisis. Con esos grupos se evalúa y se suman la cantidad de rostros conocidos entregando el resultado que se presenta en la Figura 39.

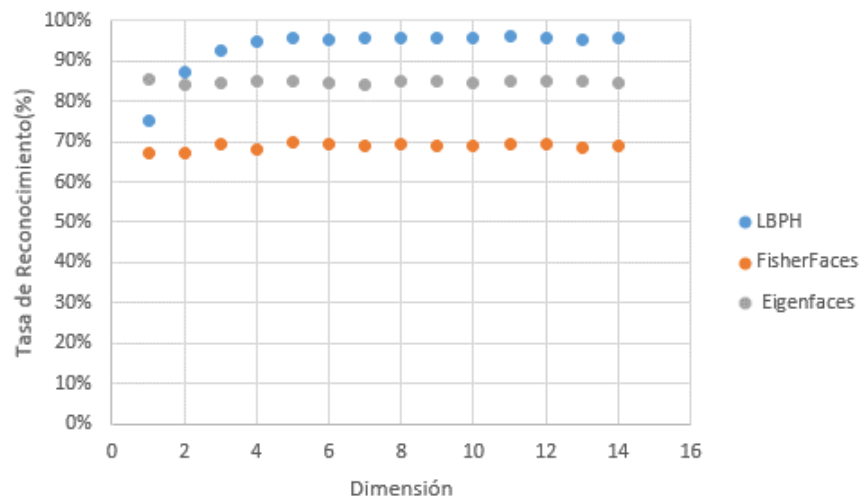


Figura 39. Comparación del desempeño de los algoritmos con 115 personas

Con la gráfica de la Figura 39 se puede concluir que el mejor método de reconocimiento es el LBPH, aunque los resultados de Eigenfaces son aceptables ya que la tasa de reconocimiento es mayor al 80 % y no tiene cambios en reconocer por la dimension de la imagen de entrada.

Con los resultados y el análisis de los tres anteriores evaluaciones se selecciono Eigenfaces dado que la tasas de reconocimiento en las 3 pruebas fueron mayor a 70 %, no obstante LBPH en las tres pruebas sus resultados fueron excelentes pero el tiempo de

ejecución es alto a comparación de Eigenface, y uno de los objetivos para seleccionar el método de reconocimiento es la rapidez de reconocimiento. Aun así los 3 métodos de reconocimientos tiene sus fortalezas y debilidades, dependiendo de los requerimientos cada uno de los métodos puede ser el adecuado, en un escenario en el cual se requiera una alta tasa de reconocimiento sin importar el tiempo de procesamiento el método más adecuado será LBPH, en el escenario en el cual se requiera una alta velocidad de reconocimiento y que tenga tolerancia con la variación de luminosidad se puede seleccionar Fisherfaces, por ultimo Eigenfaces es el tiempo de reconocimiento es menor y su tasa de reconocimiento es bueno. Cabe mencionar que Eigenfaces y Fisherfaces son sistemas que se pueden mejorar teniendo un umbral dinámico y con un entrenamiento e imágenes mas robusto.

8.3. Sistema de detección y reconocimiento

El sistema de detección con el método de Viola y Jones y reconocimiento con el método Eigenfaces se escogieron por los resultados de las pruebas antes mencionado.

Para el sistema de detección se establece un entrenamiento de 20 etapas con un tamaño de la imagen de 24×24 pixeles, el factor de escala es de 1,03 y con 7 vecinos, de acuerdo a lo concluido en la sección 8.1.2 . Para el sistema de reconocimiento se encuentra Eigenfaces con 24 imágenes por persona de entrenamiento y el pre-procesado necesario para su implementación.

Considerando que el objetivo del sistema es detectar y reconocer múltiples rostros en una imagen, se modificaron imágenes, en las cuales contengan múltiples personas para detectar y reconocer, en vista de que las imágenes usadas anteriormente no son adecuadas para evaluar el sistema de detección y reconocimiento, se selecciono fondos y objetos que no son rostros y se unieron con rostros conocidos para formar las imágenes a evaluar (ver la Figura 40). Se modificaron 20 imágenes que contiene rostros de 10 personas tomadas del conjunto de personas evaluadas en el sistema de reconocimiento, a esas 10 personas se seleccionaron 34 imágenes aleatorias para entrenar con el 70% (24 imágenes) y para evaluar el 30% (10 imágenes). Las 20 imágenes modificadas se dividen en 4 grupos de 5 imágenes en los cuales cada grupo de imágenes contienen entre 5 y 2 personas. Para ilustrar en la Figura 40 se ven 4 personas donde 3 son conocidas (recuadro verde) y una desconocida (recuadro rojo) además en la imagen

contiene objetos como juguetes, sillas entre otras.

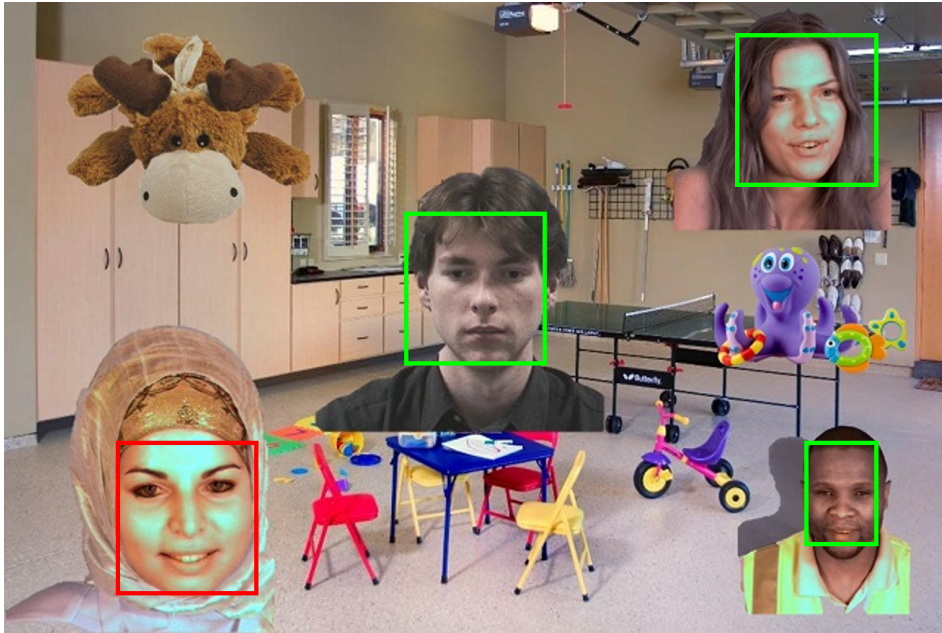


Figura 40. Tipo de imagen modificada con diferentes objetos diferentes y 4 personas para detectar y reconocer

Al tener las 20 imágenes, se evaluaron con el sistema de detección y reconocimiento en conjunto, teniendo en cuenta que se puede presentar los siguientes casos: primer caso el sistema detecte el rostro de la persona y la reconozca, segundo caso que se detecte y no reconozca, y el tercer caso, que no detecte el rostro y por consiguiente no puede reconocer. Teniendo eso casos definidos, se puede ver los resultados en la Figura 41 donde presenta a 5 personas, 4 son conocidas y una desconocida, el sistema detecto a las 5 personas sin falsos positivos, reconoció las 4 personas, como se ve en la figura donde “VP” (verdadero positivo) de la detección, “R” (reconoció) y “NO R” (no reconoció) pero a la persona desconocida le asigno un nombre (Julio) en vez de establecerlo como desconocido. En la Figura 42 se observan 3 personas donde todas fueron identificadas pero el detector presenta un falso positivo (FP), además el sistema reconoció a 2 individuos correctamente y la falsa detección la reconoció como una persona conocida como se ve en la Figura 42.

Al realizar las primeras evaluaciones se pudo observar que el reconocimiento no era el deseado como el obtenido en las pruebas en las que se evaluó el sistema de reconocimiento por separado,, debido a que el sistema de detección es muy selectivo, es decir que solo se establece como rostro los ojos la nariz y la boca, pero para reconocer es

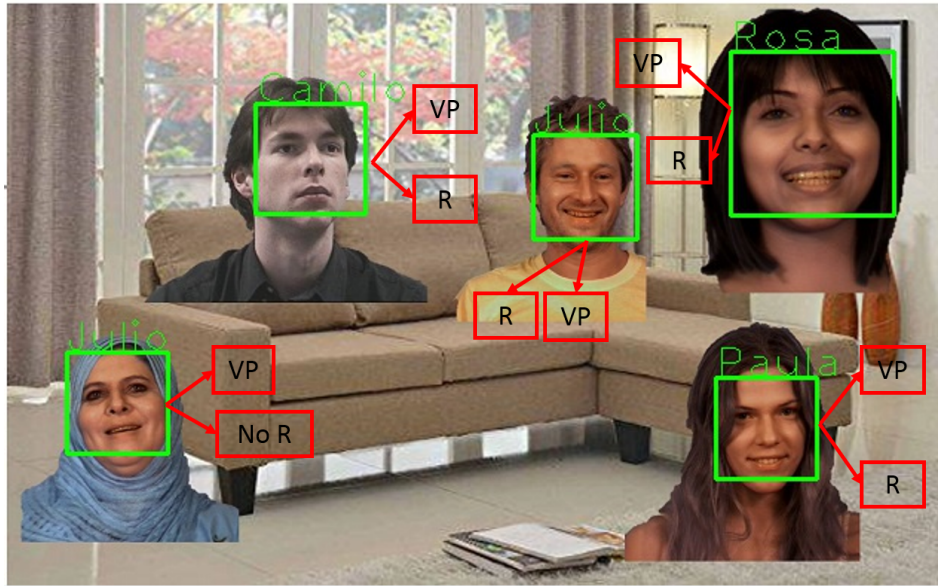


Figura 41. Resultados del sistema de detección y reconocimiento

necesario tener mas características como la frente y la barbilla, por lo tanto se modifico la dimension de la detección aumentando de tamaño aproximadamente 20 pixeles de ancho y alto, para esa imagen del rostro de mayor tamaño sea la entrada al sistema de reconocimiento. Con esas especificaciones se evaluó el sistema y se obtuvo un mejor resultado como se ve en la Figura 43 donde la imagen superior es el resultado sin modificar el tamaño de la imagen de entrada al sistema de reconocimiento, esté reconoció a las 3 personas erróneamente , mientras que la segunda imagen al modificar la detección reconoció a 4 individuos de la imagen correctamente. En la Figura 44 se puede observar que detecto todos los rostros de la imagen pero reconoció de las 5 personas 4 correctamente después de realizar el cambio de tamaño, la persona que reconoció erróneamente es una persona desconocida.

Con las modificaciones establecidas al sistema de detección y reconocimiento se obtienen los resultados que se presenta en la Tabla 15, evaluando las 20 imágenes, que contiene 70 rostros que pertenecen a 10 conocidos y a 4 desconocidos, esos rostros se presentan en las imágenes con variación de posición e iluminación como se puede apreciar en la Figura 40 y 41. Además por el grupo de personas se modificaron 5 imágenes. En la Tabla 15 se divide los resultados por la cantidad de rostros por imágenes, para analizar como se comporta el sistema con imágenes de diferente cantidad de rostros, además se mostrara cuantos rostros conocidos (C) y no conocidos (No C) hay en cada

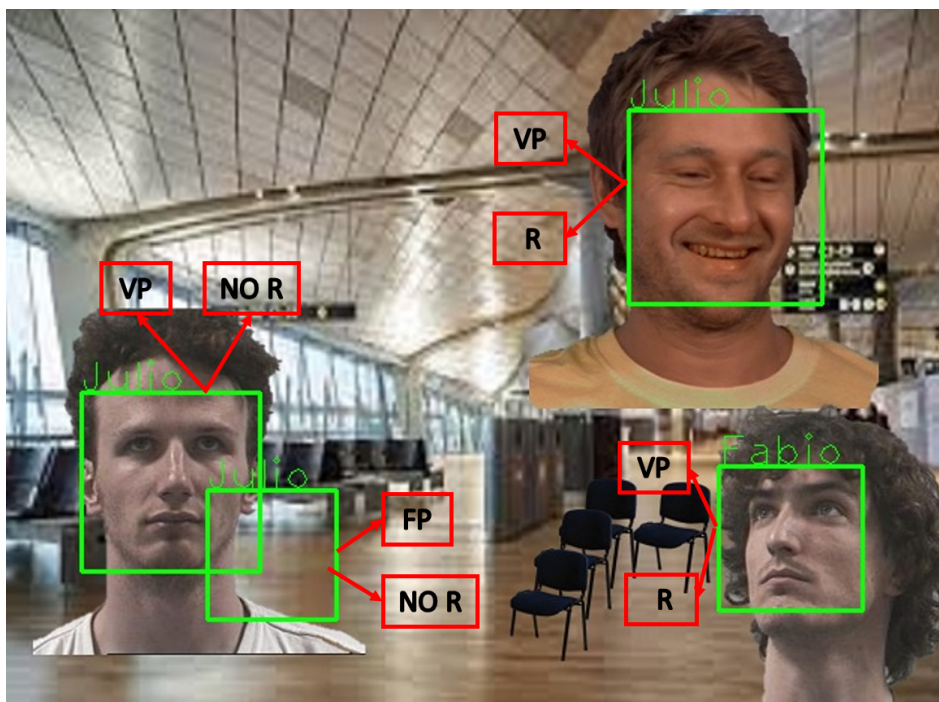


Figura 42. Resultados del sistema de detección y reconocimiento

conjunto de imágenes, y se ve los resultados de la detección y el reconocimiento de cada conjunto. Hay que tener en cuenta que aunque se encuentre personas no conocidas si este lo detecta como desconocido sera un resultado correcto.

Tabla 15: Resultado del sistema de detección y reconocimiento con las 12 imágenes creadas

cantidad personas	cantidad		Detección			Reconocimiento	
	C	No C	VP	FP	FN	Reconoció	No Reconoció
2 personas	10	0	10	1	0	8	3
3 personas	15	0	15	2	0	11	6
4 personas	17	3	20	0	0	12	8
5 personas	23	2	25	0	0	16	9

En la Tabla 15 y en la Figuras 40 y 41 se observa que el sistema aunque no reconozca todas los rostros y tenga falsos positivos en la detección de rostros, tiene una tasa de detección de 100 % pero con falsos positivos y una tasa de reconocimiento de entre 60 % a 70 % en la prueba con las 20 imágenes. Por lo tanto se puede concluir que el sistema de detección y reconocimiento que se elaboro en este proyecto es un sistema confiable al momento de detectar y al reconocer puede que presente problemas al reconocer

quien es una persona desconocida pero como se observo en la Tabla 15, la tasa de reconocimiento es alto, ademas con este se puede usar en tiempo real que es uno de los puntos importantes al momento de realizar una aplicación.

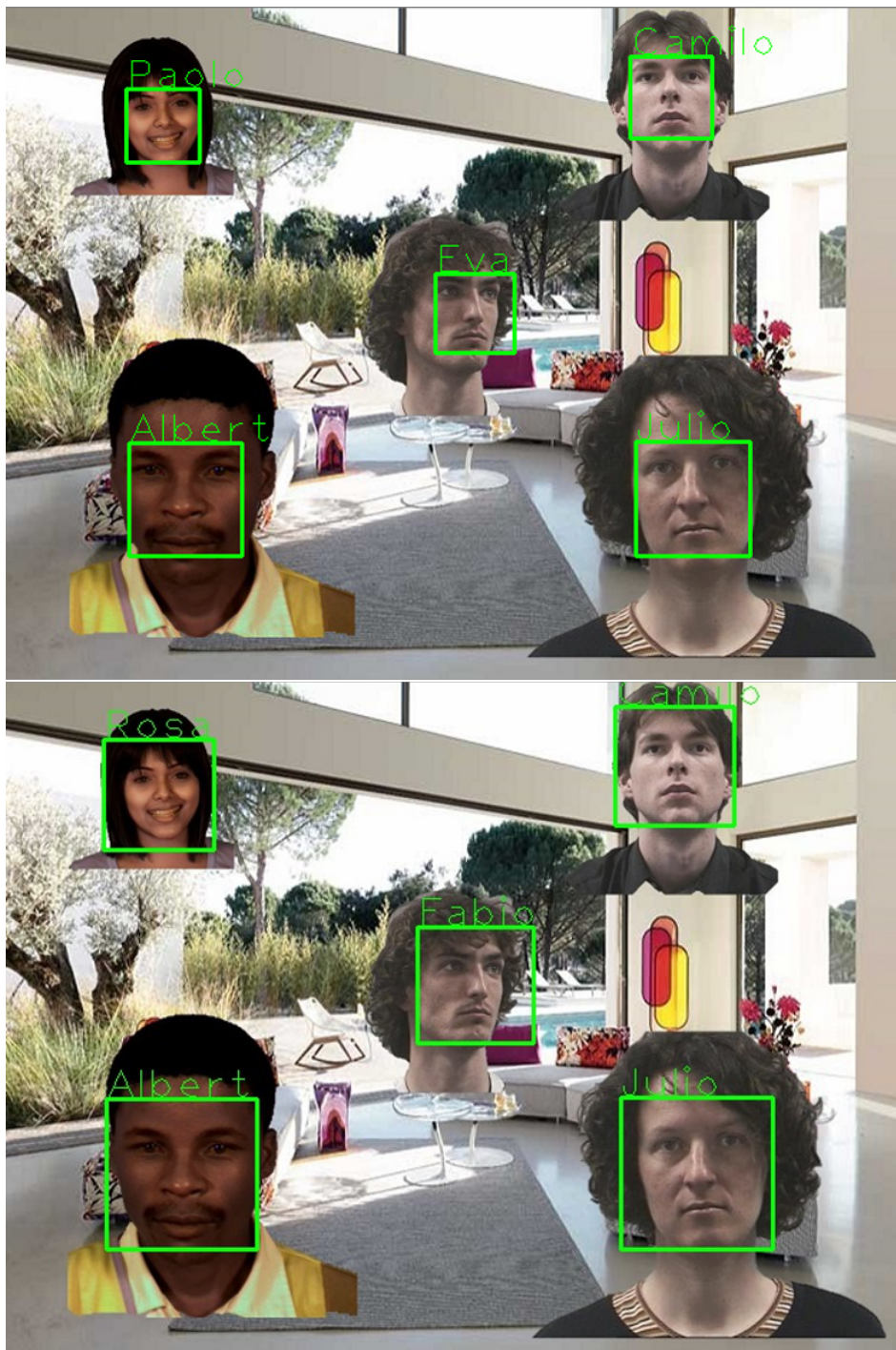


Figura 43. la primera imagen es el resultado del sistema de detección y reconocimiento sin las modificaciones en la detección y la segunda es el resultado con las modificación realizadas

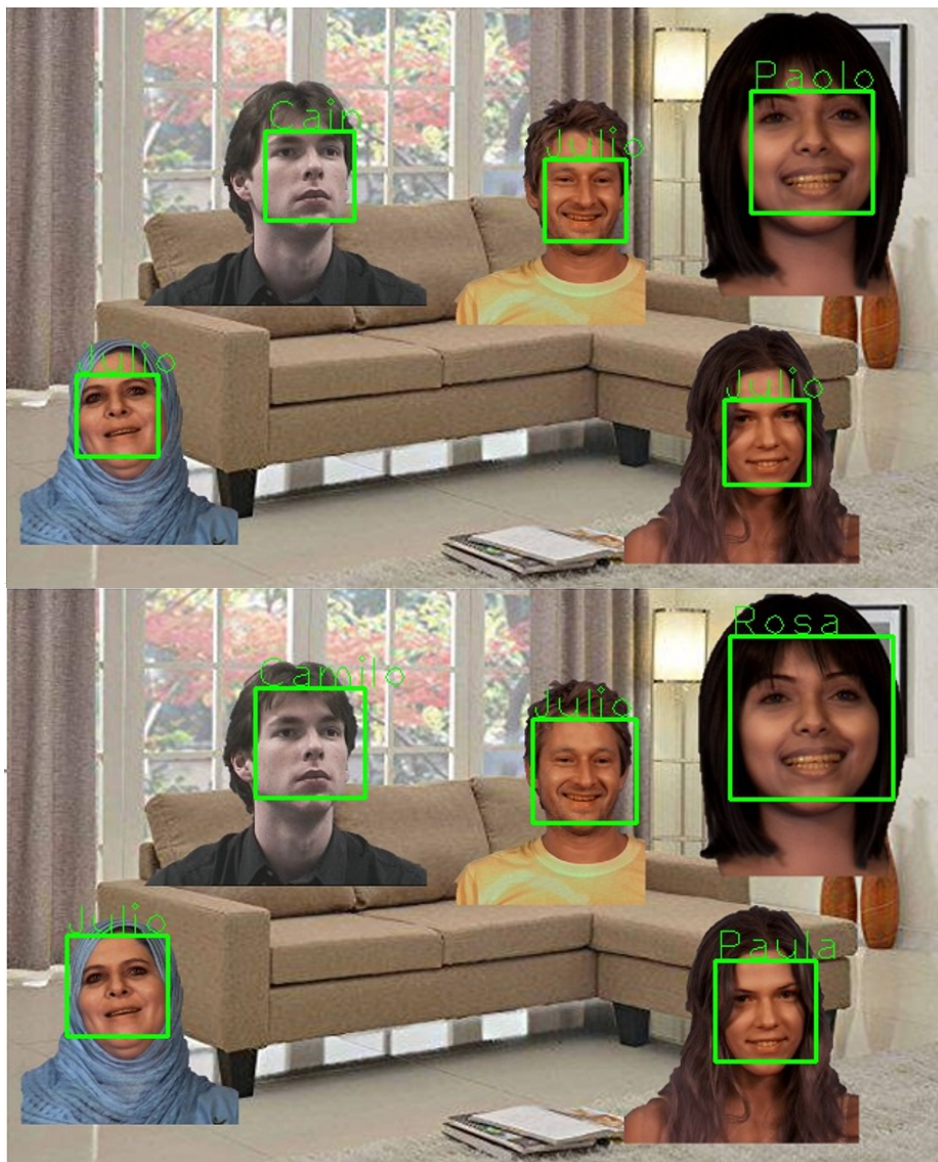


Figura 44. la primera imagen es el resultado del sistema de detección y reconocimiento sin las modificaciones en la detección y la segunda es el resultado con las modificación realizadas

9. CONCLUSIONES Y TRABAJO FUTURO

En la actualidad, el procesamiento y análisis de los rostros humanos es una de las áreas más activas y usadas en el campo de la visión artificial. Como se ha descrito a lo largo del este proyecto, es un área que se encuentra en un estado de madurez, por todos los nuevos métodos y diseños que se emplean día a día y el creciente uso de nuevas aplicaciones que se surgen continuamente. En el presente proyecto se aborda la elaboración de un sistema que detecte cualquier rostro en diferentes posición y dimensiones, y reconozca dicho rostro.

Partiendo de una sólida investigación plasmada en la sección 4, se ha estudiado exhaustivamente la viabilidad del diseño e implementación de manera fiable, precisa y eficiente, en todos los problemas que conlleva el procesamiento visual de rostros humano: la detección, localización de las componentes faciales, reconocimiento de personas y la extracción de información facial.

Con el estudio investigado sobre como afrontar el problema de detectar rostros en una imagen digital, se encontró métodos con diferentes enfoques para solucionarlo, pueden ser métodos basados en reglas como con la tonalidad de la piel o la simetría del rostro, o métodos basados en estadísticas como Viola y Jones o redes neuronales. La contribución de este proyecto para solucionar el sistema de detección y localización de rostros es la evaluación de los diferentes métodos, concluyendo que el método de Viola y Jones presenta el mejor desempeño. El método se desempeña satisfactoriamente en ambientes con cambios de luminosidad, diferentes poses del rostro y cuando el rostro presenta oclusiones. Al usar el método de Viola y Jones es imprescindible realizar un entrenamiento supervisado, en el cual afecta considerablemente la detección de los rostros. Por lo tanto es importante que tener en cuenta las medidas pertinente para un buen entrenamiento. Como primera medida es necesario contar con una base de datos grande de imágenes de rostros con diferentes variaciones; la segunda medida es el tamaño de las imágenes a entrenar, un entrenamiento con imágenes grandes otorga mejores resultados en la tasa de detección, pero eso implica más tiempo de entrenamiento, además, se debe tener cuidado con el tamaño ya que se puede tender a sobre-entrenar. El tercer parámetro es el uso de las características Haar, para este punto es necesario tener a consideración la importancia de la información del rostro, en concreto si se busca tener una información robusta y con toda la morfología del rostro se puede usar la características Haar con

rotación, pero si bien se busca generalizar el rostro se usa las características Haar básicas. El último parámetro a considerar, es que tan profundo debe estar el clasificador de características Adaboost para que pueda discriminar correctamente todos los rostros y no rostros, es lógico pensar que al tener más etapas en el clasificador es mejor para la toma de decisiones, y es correcto, sin embargo a mayor número de etapas la cascada de clasificación tendrá un mayor costo computacional y un sobre-entrenamiento, razón por la cual es pertinente tener una cantidad de etapas según las muestras de entrada.

En el proceso de detección se presentó un desafío debido a los objetivos propuestos en este proyecto los cuales son detectar todo tipo de rostro humano a diferente tamaño y posición, por tal motivo al factor escalar se asignó un valor bajo, considerando que al cambiar el tamaño de la sub-ventana de detección este detecte todo tipo de rostros en diferentes tamaños, esta decisión implicó que se demorara más el sistema de detección. La otra variable a considerar son los vecinos, si se tiene un valor menor en el momento de detectar no se tienen tantas restricciones al determinar si es un rostro o no, sin embargo dado que no hay tantas restricciones se puede presentar una alta tasa de falsos positivos. Para este proyecto se seleccionó un valor medio el cual no sea tan restrictivo con la cantidad de detecciones que debe tener para clasificar un rostro. Con esas variables establecidas el algoritmo propuesto presentó un porcentaje próximo al 90% para 2 falsos positivos por cada 5 imágenes. En imágenes típicas de 640×480 píxeles, el proceso tarda aproximadamente 0,9 s.

Para el sistema de reconocimiento e identificación se establecieron tres métodos usados en la actualidad, para reconocer rostros en una imagen digital, esos métodos fueron sometidos a estudios frente a diversas situaciones, con esos resultados se obtuvo la información necesaria para saber cómo se desempeñan en el reconocimiento de rostros y el tipo de entrenamiento que requieren. Como se ha explicado a lo largo de este documento los métodos Eigenfaces, Fisherfaces y LBPH requieren un entrenamiento previo al reconocimiento. La manera de realizar el entrenamiento como se explica en la sección 7.4.4 afecta de manera contundente el reconocimiento. En el caso de Eigenfaces y Fisherfaces es muy importante el número de imágenes a utilizar aunque sea ligeramente diferente a las que se usarán en el reconocimiento o totalmente diferentes, es necesario que sea numerosas y variada. Es por esto que en un futuro es necesario construir una amplia base de datos de entrenamiento para obtener mejores resultados que los obtenidos en este proyecto. Mientras que para el método LBPH no es tan imprescindible la

cantidad de las imágenes si no la calidad de estas.

Con los sistemas realizados y los parámetros establecidos se realiza unas pruebas las cuales presentan casos con condiciones muy concretas y conocidas, como con variación de iluminación y posición del rostro. En los resultados obtenidos se llego a la conclusión de que el sistema en un entorno con cambios de iluminación el método que presenta una mejor solución es LBPH con una tasa de reconocimiento mayor al 90 % con un tiempo aproximado de 187 ms, en la prueba de reconocer imágenes que contengan cambios de posición el método con mejor resultados fue Eigenfaces dado que el tiempo de ejecución fue de 7 ms y su tasas de detección fue mayor al 70% aun así el LBPH presento una tasa de reconocimiento mayor a comparación del Eigenfaces pero su ejecución es de aproximadamente 204 ms. Por ultimo se realizo con 115 personas una prueba de reconocimiento en el cual el método que se destaca es Eigenfaces y LBPH. Concluyendo que LBPH es un método en el cual se puede confiar al proceso de reconocer rostros con cambios de iluminosidad y rotación, pero para usar este método en tiempo real no es recomendable dado que tarda en realizar el reconocimiento. Mientras que Eigenfaces tiene un tasa de reconocimiento alto y el tiempo empleado para reconocer es menor que los otros métodos por tal motivo en este proyecto se selecciono este sistema.

Para trabajo a futuro se puede usar en un vídeo en tiempo real, situando una cámara en un área concurrida, monitoreando las personas que recorren esa zona. Para esta aplicación es necesario realizar mejoras en el entrenamiento del sistema de reconocimiento facial y realizar pruebas al combinar estos métodos para poder analizar si se aumenta la tasa de reconocimiento, igualmente estudiar otro tipos de metodologías como el Deep-learning para este tipo de problemáticas.

BIBLIOGRAFÍA

- [1] AFIFI, M., AND ABDELHAMED, A. Afif4: Deep gender classification based on adaboost-based fusion of isolated facial features and foggy faces. *arXiv preprint arXiv:1706.04277* (2017).
- [2] ALABBASI, A., AND MOLDOVEANU, F. Human face detection from images, based on skin color. *IEEE 18th ICSTCC* (Oct. 2014), 532–537.
- [3] ANGGRAINI, D. Face recognition using principal component analysis and self organizing maps. *IEEE ICT-ISPC2014* (2014), 91–94.
- [4] BORADE, S., DESHMUKH, R., AND RAMU, S. Face recognition using fusion of pca and lda: Borda count approach. *IEEE 24th MED* (June 2016).
- [5] CHEFFOU, F. Un programa de reconocimiento facial del fbi permitió identificar al 'hombre del sombrero', europa press, Disponible en internet. URL: [urlhttp://www.europapress.es/internacional/noticia-programa-reconocimiento-facial-fbi-ermitio-identificar-hombre-sombrero-20160415100158.htm](http://www.europapress.es/internacional/noticia-programa-reconocimiento-facial-fbi-ermitio-identificar-hombre-sombrero-20160415100158.htm), Abr 2016. Accedido 23-08-2017.
- [6] CHENGYUAN, L., TING, Z., DONGSHENG, D., AND CHONGSHAN, L. Design and application of compound kernel-pca algorithm in face recognition. *35th Chinese Control Conference* (July 2016), 4122–4126.
- [7] CORPORATION, I. *Open Source Computer Vision Library*. opencv, 2017. <https://opencv.org/>.
- [8] EDIZIONES, P. Nueva york utilizará el reconocimiento facial para atrapar terroristas, Portaltic europa press, Disponible en internet. URL: [urlhttp://http://www.europapress.es/portaltic/sector/noticia-nueva-york-utilizara-reconocimiento-facialatrapar-terroristas-20161007171657.html](http://http://www.europapress.es/portaltic/sector/noticia-nueva-york-utilizara-reconocimiento-facialatrapar-terroristas-20161007171657.html) , Oct 2016. Accedido 23-08-2017.
- [9] FARRUGIA, N., MAMALET, F., ROUX, S., YANG, F., AND PAINDAVOINE, M. Design of real-time face detection parallel architecture using high level systhensis. *Hindawi Publishing Corporation 8* (Nov. 2008).

- [10] GIMENO, R. Estudio de técnicas de reconocimiento facial. *Un. Politécnica, Cat, Barc, Es* (May 2010).
- [11] GOURIER, N., HALL, D., AND CROWLEY, J. L. Estimating face orientation from robust detection of salient facial features. *ICPR, International Workshop on Visual Observation of Deictic Gestures* (2004).
- [12] GUEVARA, M., AND ARDILA, W. Deteccion de rostros en imagenes digitales usando clasificadores en cascada. *Scientia et Technica Año XIV* (Jun 2008).
- [13] GUO, J., LIN, C., NAD C. CHANG, M. W., AND LEE, H. Complexity reduced face detection using probability-based face mask prefiltering and pixel-based hierarchical-feature adaboosting. *IEEE Signal Processing Letters 18* (Aug. 2011), 447–450.
- [14] HUANG, Y., AND CHEN, S. A geometrical-model-based face recognition. *IEEE ICIP* (2015), 3106–3110.
- [15] JAIN, V., AND LEARNED-MILLER, E. Fddb: A benchmark for face detection in unconstrained settings. Tech. Rep. UM-CS-2010-009, University of Massachusetts, Amherst, 2010.
- [16] JUN, Z., JIZHAO, H., ZHENGLAN, T., AND FENG, W. Face detection based on lbp. *2017 13th IEEE International Conference on Electronic Measurement Instruments (ICEMI)* (Oct 2017), 421–425.
- [17] KEITH BEDFARO. Policía de nueva york utilizará automóviles inteligentes para optimizar su trabajo Grupogreek, Disponible en internet. URL: [urlhttp://http://grupogreek.com/policia-de-nueva-york-utilizara-automoviles-inteligentes-para-optimizar-su-trabajo](http://http://grupogreek.com/policia-de-nueva-york-utilizara-automoviles-inteligentes-para-optimizar-su-trabajo), 2013. Accedido 20-08-2017.
- [18] KUMAR, R. Eye center localization using cascaded corner detection and geometrical measuments algorithm. *IEEE* (2015).
- [19] L. ZHANG, R. CHU, S. X., AND LIAO, S. Face detection based on multi-block lbp representation. *in Advance in Biometrics: Lecture Notes in Computer ScienceSpringer Berlin Heidelberg* (2007).
- [20] LANG, L., AND GU, W. Study of face detection algorithm for real-time face detection system. *IEEE 2nd ISECS* (2009), 129–132.

- [21] LEE, K., HO, J., AND KRIEGMAN, D. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. Pattern Anal. Mach. Intelligence* 27, 5 (2005), 684–698.
- [22] LENC, L., AND KRÁL, P. Unconstrained Facial Images: Database for face recognition under real-world conditions.
- [23] LIN, C., AND LIN, S. Efficient face detection method with eye region judgment. *EURASIP Int J Image Video Process* (2013), 1–14.
- [24] MAHDI, S., MENHAJ, M., AND HORMAT, M. A new feature extraction based on advanced pca for real time face recognition. *IEEE 13th IFSC* (2013).
- [25] MARCINIAK, T., CHMIELEWSKA, A., WEYCHAN, R., PARZYCH, M., AND DABROWSKI, A. Influence of low resolution of images on reliability of face detection and recognition. *Multimed Tools Appl* 74 (July 2013), 4329–4349.
- [26] MARQUES, O. *Practical Image and Video Processing Using Matlab*. Pres A John Wiley and Sons.INC., Publications, Floria, Atlantic, 2011.
- [27] MARTIN KOESTINGER, PAUL WOHLHART, P. M. R., AND BISCHOF, H. Annotated Facial Landmarks in the Wild: A Large-scale, Real-world Database for Facial Landmark Localization. Tech. rep., 2011.
- [28] MARTINEZ, A., AND BENAVENTE, R. The ar face database. *CVC Technical Report* (1998).
- [29] MATLAB. *MATLAB 2017 version 2.1*. The MathWorks Inc, 2017. <https://www.mathworks.com>.
- [30] MICROSOFT. *Microsoft Visual Studio*. Itaniu, 2017. <https://www.visualstudio.com/es/>.
- [31] MILBORROW, S., MORKEL, J., AND NICOLLS, F. The MUCT Landmarked Face Database. *Pattern Recognition Association of South Africa* (2010). <http://www.milbo.org/muct>.
- [32] MISHRA1, A., KUMAR, K., RAI, S., AND MITTAL, K. Multi-stage face recognition for biometric access. *IEEE INDICON* (2011), 1–6.

- [33] OYJ, G. *QT*. Qt Development Frameworks, 2017. <https://www1.qt.io/es/>.
- [34] PADMAPRIYA, S., AND KALAJAMES, E. Real time smart car lock security system using face detection and recognition. *ICCCI* (Jan. 2012).
- [35] PAGLIERI, J. El fbi lanza un poderoso sistema de reconocimiento facial, CNNEspañol, Disponible en internet. URL: <http://cnnespanol.cnn.com/2014/09/17/el-fbi-lanza-un-poderoso-sistema-de-reconocimiento-facial> , Sep 2014. Accedido 23-08-2017.
- [36] PAUL, V., AND JONES, M. Robust real-time face detection. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2001).
- [37] PISA, K. El terrorismo mundial aumentó en un 80% en 2014, según informe CNNEspañol Disponible en internet. URL: <http://cnnespanol.cnn.com/2015/11/18/elterrorismomundialaumentoen-un-80-en-2014-segun-informe/ash>, Nov 2015. Accedido 22-08-2017.
- [38] RAMA, A., AND TARRÉS, F. Un nuevo método para la detección de caras basado en integrales difusas. *Un. Politècnica, Cat, Barc, Es* (2016).
- [39] SEO, J. J., KIM, H. I., AND RO, Y. M. Pose-robust and discriminative feature representation by multi-task deep learning for multi-view face recognition. *2015 IEEE International Symposium on Multimedia (ISM)* (Dec 2015), 166–171.
- [40] SHAILAJA, K., AND ANURADHA, B. Effective face recognition using deep learning based linear discriminant classification. *2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)* (Dec 2016), 1–6.
- [41] SHARMA, H., SAURAV, S., SINGH, S., SAINI, A., AND SAINI, R. Analyzing impact of image scaling algorithms on viola-jones face detection framework. *IEEE ICACCI* (2005), 1715–1718.
- [42] S.SHWARTZ, AND DAVIDS, S. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, US, 2014.
- [43] .TALELE, K., AND KADAM, S. Face detection and geometric face normalization. *IEEE TENCON* (2015).

- [44] VAN BEEK, L., HOLZ, D., MATAMOROS, M., RASCON, C., , AND WACHSMUTH, S. Robocup@home 2017: Rules and regulations. http://www.robocupathome.org/rules/2017_rulebook.pdf, 2017.
- [45] WANG, B., CHANG, X., AND LIU, C. A robust method for skin detection and segmentation of human face. *IEEE 2nd ICINISs* (2009), 290–293.
- [46] WANG, X., WANG, L., LEI, T., AND WANG, C. Face detection based on improved skin model and local iterated conditional modes. *IEEE 11th ICNC* (2015), 964–970.
- [47] W.CHEN, SUN, T., YANG, X., AND WANG, L. Face detection based on half face-template. *IEEE ICEMI* (2009), 54–58.
- [48] ZHAO, X., AND WEI, C. A real-time face recognition system based on the improved lbph algorithm. *2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP)* (2017), 72–76.
- [49] ZHENG, Y., AND YAO, J. Multi-angle face detection base on dp-adaboost. *IJAC 12* (Aug. 2015), 421–431.