SEP TecNM

TECNOLÓGICO NACIONAL DE MÉXICO

INSTITUTO TECNOLÓGICO DE ACAPULCO

TEMA:

VERIFICACIÓN AUTOMÁTICA DE IDENTIDAD A TRAVÉS DEL ANÁLISIS FACIAL

OPCIÓN I:

TESIS PROFESIONAL

QUE PARA OBTENER EL TÍTULO DE:

MAESTRO EN SISTEMAS COMPUTACIONALES

PRESENTA:

HONORIO CANDELARIO EMIGDIO

DIRECTOR DE TESIS:

DR. JOSÉ ANTONIO MONTERO VALVERDE

CO-DIRECTOR DE TESIS:

DRA. MIRIAM MARTÍNEZ ARROYO

Mayo de 2019

Dedicatoria

A mis padres.

Que sin ellos no hubiera logrado una meta más en vida profesional. **Rosalía** (**q.e.p.d**), gracias por haber estado a mi lado, tu apoyo moral y entusiasmo que me brindaste para seguir adelante en mis propósitos. **Laurentino** (**q.e.p.d**), por el tiempo que estuviste conmigo, compartiendo tus experiencias, conocimientos y consejos, por tu amor, Gracias.

A mi familia.

Lo dedico en especial a mi esposa **Esperanza** e hijas **Brisa**, **Mar** y **Azul**, quienes me brindaron su apoyo y cariño incondicional para hacer realidad este sueño.

A mis maestros.

Por el tiempo y esfuerzo que dedicaron a compartir sus conocimientos, sin su instrucción profesional no habría llegado a este nivel. Quienes brindaron dedicación al impartir su cátedra de tal forma que lo aprendido sea utilizado en la vida real, por el apoyo brindado, Gracias.

A Dios.

Por darme vida, salud y sabiduría a lo largo del estudio de la Maestría.

Agradecimientos

Este trabajo es el final de un periodo más de esfuerzos, sacrificios y emociones en mi vida, todo esto ha sido posible gracias a muchas personas que día tras día han estado conmigo y me apoyaron para poder culminar satisfactoriamente este proyecto.

En primer lugar, quiero agradecer al **Instituto Tecnológico de Acapulco** y en especial a la **División de Estudios de Posgrado e Investigación** por darme la oportunidad de realizar mis estudios de maestría.

A mi asesor **Dr. José Antonio Montero Valverde** por aconsejarme, apoyarme y brindarme su confianza en todo momento. A la Dra. **Miriam Martínez Arroyo** y al Dr. **Eduardo de la Cruz Gámez** por su ayuda durante este proyecto.

Al **CONACYT** por brindarme el soporte económico para poder realizar este trabajo de investigación.

Resumen

En la actualidad la tecnología está cada vez más presente en la vida cotidiana. Como consecuencia, la información privada de usuarios y empresas se ve expuesta a incontables amenazas. No es de extrañar, por lo tanto, que tecnologías como la biometría estén en pleno crecimiento y se presenten como una solución para proteger dicha información. El campo de la biometría tiene un gran número de áreas como el reconocimiento de huellas dactilares, el reconocimiento de voz o el reconocimiento facial.

En cuanto a la detección y reconocimiento de rostros la visión artificial ha alcanzado un gran crecimiento debido a las necesidades de aplicaciones orientadas hacia la seguridad y vigilancia en diferentes ámbitos como: control de acceso, proceso de investigación, criminalística, software comercial, entre otras, permitiendo la disminución del tiempo de realización de los procesos que conlleva la realización de los mismos.

En este trabajo, se muestran los resultados obtenidos del desarrollo de un sistema de verificación facial para la seguridad biométrica mediante la utilización de redes neuronales convolucionales. El sistema realiza un proceso de verificación facial, el cual se basa en detectar el rostro, alinearlo, extraer las características y realizar la verificación facial. Asimismo, se utilizan las redes neuronales convolucionales como método para la extracción y representación de vectores de características. Los vectores extraídos mediante la red son representativos de cada rostro y permiten realizar una comparación que devuelva una distancia entre las imágenes almacenadas en la base de datos. La red busca maximizar la distancia entre los vectores de rostros distintos y minimizar la distancia entre los de la misma persona. La distancia entre los vectores almacenados es evaluada tomando en cuenta un punto de corte para determinar si son de la misma persona.

Cada una de las fases realizadas dentro del presente proyecto ha permitido obtener los resultados necesarios para cumplir con cada uno de los objetivos planteados y así obtener la herramienta computacional basada en un dispositivo móvil que permitirá el reconocimiento y la verificación de la identidad de una persona por medio de su rostro.

Esta herramienta fue sometida a 6 pruebas diferentes, 3 en ambiente controlado (neutral, mediante un estado emocional y a una determinada distancia) y 3 en ambiente no controlado de la misma forma, obteniendo en promedio un porcentaje de confiabilidad de 97.7%.

Índice General

Índice de Figuras

Índice de Tablas

Capítulo 1 Introducción

La seguridad es un tema que ha sido de gran interés para la comunidad tecnológica y científica, el reconocimiento y la verificación de la identidad de las personas es uno de los aspectos fundamentales. Los ataques terroristas ocurridos en los últimos años han demostrado la necesidad de establecer métodos más confiables para verificar la identidad de las personas. Pero no sólo es eso, las aplicaciones en el control de acceso en el ámbito laboral o los controles en el tránsito de aduanas, entre otras aplicaciones abren un campo muy amplio en el desarrollo de aplicaciones. Los sistemas biométricos surgen como una solución real a los problemas de verificación. La biometría consiste de un conjunto de métodos automatizados para la autenticación de individuos mediante el uso de características físicas o del comportamiento de la persona [1]. Esta tecnología se basa en la premisa de que cada persona es única y posee rasgos distintivos que pueden ser utilizados para identificarla.

Los seres humanos poseen una alta capacidad para reconocer rostros, aún en escenarios donde existan altos niveles de variabilidad y ruido. Diseñar sistemas automáticos que emulen esta propiedad natural de los humanos, constituye una tarea compleja y con muchas limitaciones. Probablemente una de las primeras interrogantes sea ¿los rostros son diferenciables a través de medidas biométricas?. Afortunadamente en los últimos años se han realizado una gran cantidad de investigaciones que responden a esta interrogante, en especial el área de la biometría. La biometría busca obtener, clasificar y utilizar la información de características biológicas, para reconocer o verificar la identidad de las personas, restringir el acceso a sitios no permitidos, controlar horarios, autentificar información, y muchas otras aplicaciones para la identificación y seguridad de las

empresas. Para esto se utilizan equipos electrónicos que desarrollan las mediciones biométricas, y algoritmos que permiten digitalizar, clasificar y almacenar la información.

El procesamiento automático de imágenes para extraer contenido semántico es una tarea que ha ganado mucha importancia durante los últimos años, debido al número cada vez mayor de fotografías digitales en internet o que se almacenan en computadoras personales. La necesidad de organizarlos de forma inteligente utilizando técnicas de indexación y recuperación requiere un análisis efectivo, eficiente y algoritmos de reconocimiento de patrones que sean capaces de extraer información semántica relevante. Especialmente las de mayor costo, contienen una gran cantidad de información valiosa en comparación con otros objetos o elementos visuales en las imágenes. Por ejemplo, reconocer a una persona en una fotografía.

El objetivo principal del análisis de rostros es extraer información relevante del rostro, como su posición en la imagen, las características, las expresiones, el género o la identidad de la persona.

Brandon Amos y otros [2] plantean que la visión computacional se centra en la extracción de características del rostro humano para que estas sean entendidas por una computadora. Mediante el entendimiento de estas características, las computadoras pueden determinar la localización de ciertos objetos dentro de una imagen, reconocerlos, clasificarlos o descomponerlos.

Antes de empezar a analizar los procedimientos es necesario dejar claro el concepto de reconocimiento y verificación. En el reconocimiento el sistema no sabe quién es la persona de la cual han capturado los rasgos característicos (el rostro humano en este caso) por lo cual el sistema tiene que determinar a quién pertenecen los datos que acaba de procesar. En la verificación, la persona le informa al sistema cuál es su identidad ya sea presentando una

tarjeta de identificación o escribiendo alguna clave especial, el sistema captura el rasgo característico de la persona (el rostro humano en este caso), y lo procesa para crear una representación electrónica llamada modelo en vivo (“live template"en inglés). Por último, el sistema compara el modelo en vivo con el modelo de referencia de la persona. Si ambos modelos coinciden la verificación es exitosa. De otro modo, la verificación no es exitosa.

El presente trabajo se apoya en el campo de la visión computacional, en especial en los métodos de verificación facial para la creación de un sistema online con dichas funcionalidades (análisis de las características faciales del sujeto extraídas de la imagen). La verificación facial en este trabajo de investigación consiste en que se va a mostrar la identidad del rostro en el dispositivo móvil con el fin de tener una mayor aplicación con pocos recursos, además se plantea la utilización de redes neuronales convolucionales, ya que en la actualidad están dando mejores resultados [3].

1.1. Antecedentes

La identificación biométrica en teléfonos móviles/inteligentes es una de las áreas de investigación activa en sistemas de información seguros e inteligentes. Se han realizado diferentes estudios de investigación sobre las diferentes técnicas biométricas disponibles para teléfonos móviles/inteligentes. Estas técnicas incluyen, reconocimiento de huellas dactilares, rostro, geometría de la mano, iris, voz, firma y pulsación de teclas, entre otras.

M. Gargi y otros [4] propusieron un método para brindar seguridad a los teléfonos inteligentes Android que utilizan la función biométrica del iris. El estudio proporciona resultados prometedores en un dispositivo Android con procesador de 1 GHz y 4 Gb de memoria interna, y con un tiempo total de 80 a 90 segundos para autenticar a un solo

usuario móvil de la base de datos, de 75 personas que contienen 5 imágenes de iris de cada persona.

Santo Sierra y otros [5] propusieron un sistema biométrico basado en la geometría de la mano, que está orientado a dispositivos móviles. Los autores afirman que el sistema puede proporcionar resultados precisos en la identificación individual. La investigación de Santo Sierra muestra la implementación de la biométrica manual en una PC con 2.4 GHz y una plataforma móvil Android con procesador de 1 GHz y 576 Mb de RAM. El resultado de la investigación mostró un buen rendimiento con FAR[1] = 0.089% y FRR[2] = 5.89%. La implementación móvil tardó menos de 3 segundos en proporcionar una identificación de una base de datos de 120 individuos diferentes. Una de las limitaciones de la geometría de la mano es que no es única y no se puede utilizar para la identificación dentro de una gran población [14].

Guillaume Dave y otros [6] investigaron el rendimiento de diferentes algoritmos de reconocimiento facial en teléfonos inteligentes. Los autores analizan el rendimiento de los algoritmos aplicándolos a un teléfono Android con procesador de 600 MHz y 256 Mb de RAM. Las pruebas se realizaron con 134 imágenes de caras de 10 personas diferentes. Los resultados indican que logró una tasa de reconocimiento del 94% con algoritmos de *fisherface* y no tomó más de 1.6 segundos.

Vázquez-Fernández y otros [7] presentan una aplicación inteligente para compartir fotos en dispositivos móviles basada en el motor de reconocimiento facial. El sistema se implementa en la plataforma Android y se prueba en dos teléfonos inteligentes de diferentes fabricantes, HTC Desire con procesador de 1 GHz y 576 Mb de RAM y Samsung Galaxy Tab, con

[1] Tasa de Aceptación Falsa
[2] Tasa de Falso Rechazo

procesador de 1 GHz y 512 Mb de RAM. Las pruebas se realizaron para 50 contactos con 4 rostros por contacto. Los resultados mostraron que la aplicación tardó 0.35 segundos en HTC Desire y 0.47 segundos en el Samsung Galaxy para reconocer los rasgos faciales.

Estos estudios proporcionan la implementación de diferentes técnicas de identificación biométrica en los teléfonos inteligentes, pero no indican su rendimiento en grandes conjuntos de datos.

1.2. Planteamiento del Problema

La autenticación juega un papel importante ya que permite identificar a los usuarios antes de que estos accedan a los recursos y servicios de un sistema. Sin un mecanismo de autenticación apropiado, se puede suplantar fácilmente la identidad de los usuarios y realizar acciones en su nombre.

Los sistemas que habitualmente se utilizan para identificar a una persona, como el aspecto físico o la forma de hablar, son demasiado complejos para una computadora; el objetivo de los sistemas de identificación de usuarios no suele ser identificar a una persona, sino autenticar que esa persona es quien dice ser realmente. Aunque como humanos, seguramente ambos términos nos parecerán equivalentes, para una computadora existe una gran diferencia entre ellos: un potencial sistema de identificación estrictamente hablando, por ejemplo, uno biométrico basado en el reconocimiento de la retina; una persona miraría a través del dispositivo lector, y el sistema sería capaz de decidir si es un usuario válido, y en ese caso decir de quién se trata; esto es identificación. Sin embargo, lo que habitualmente hace el usuario es introducir su identidad (un número, un nombre de usuario...) además de mostrar sus retinas ante el lector; el sistema en este caso no tiene que

identificar a esa persona, sino autenticarlo: comprobar los parámetros de la retina que está leyendo con los guardados en una base de datos para el usuario que la persona dice ser: se está reduciendo el problema de una población potencialmente muy elevada a un grupo de usuarios más reducido, el grupo de usuarios del sistema que necesita autenticarlos.

Las técnicas de autenticación de usuario pueden clasificarse en tres factores, dependiendo de lo que se deba presentar para demostrar la identidad: (a) algo que se conoce (contraseña, PIN[3], etc.), (b) algo que se tiene (tarjeta inteligente, *token*, etc.), (c) algo que se es o característica biométrica (huella digital, voz, retina, iris, etc.) [8]. Si el usuario utiliza dos o más de estos factores para identificarse ante algún sistema, la autenticación es más fiable y si se utiliza sólo uno de ellos, la autenticación es débil.

Cada uno de los factores de autenticación tiene sus consideraciones para poder ser usado. El primero de ellos (algo que se conoce), es el más simple de implementar y el más usado (usuario/contraseña), pero también el más vulnerable, ya que, incluso haciendo uso de buenas prácticas, como contraseñas largas y complejas, no se evita que el usuario sea víctima de suplantación porque algún atacante cibernético obtuvo su contraseña [9]. El segundo de ellos (algo que se tiene), brinda una mejor protección al hacer uso de un elemento externo que identifica al usuario [10]. Su implementación es más compleja que el método anterior y es usado ampliamente en la actualidad, como en el caso de las tarjetas de crédito. Este se suele combinar con el primer factor, para ofrecer mayor seguridad. Sin embargo, los usuarios pueden perder este elemento externo, no cargarlo siempre con ellos o ser víctimas de algún robo o suplantación de identidad.

El último factor de autenticación (algo que se es) es el que presenta mayor complejidad de implementación y es usado principalmente para el control de acceso a nivel local, debido al

[3] Número de identificación personal

requerimiento de hardware especial para la captura de los datos biométricos y además, los datos biométricos no pueden viajar sin cifrar por la red, pues si un atacante obtuviera esta información, al ser estas características inherentes a cada ser humano muy particulares, no se podrían cambiar con facilidad y la autenticación ya no sería segura. Para evitar el problema de adquirir hardware muy especializado para la captura de datos biométricos, se ha pensado en el uso de dispositivos móviles actuales. La pregunta de investigación es: ¿Cómo implementar un mecanismo de autenticación biométrica por reconocimiento facial para dispositivos móviles, utilizando redes neuronales convolucionales como algoritmo de extracción de características?

En este proyecto se pretende implementar una herramienta de verificación (autenticación) robusta basada en rasgos faciales, donde los datos entre el cliente y el servidor viajen de manera cifrada, con el fin de evitar la suplantación de identidad en dispositivos móviles.

1.3. Justificación

La necesidad de incrementar la seguridad en algunos lugares por medio de un riguroso control de acceso es un campo de investigación activa durante los últimos años.

El uso de los patrones característicos de un rostro de una imagen para el reconocimiento o verificación de personas tiene algunas ventajas ante los demás tipos biométricos de identificación como son: huellas digitales, patrones de la palma de la mano, patrones del iris, firmas digitales, ya que no se necesita ningún equipo especial para la captura de el rasgo biométrico si no que con un simple dispositivo móvil se puede tomar la fotografía para procesarla (además que es un esquema no invasivo). Cualquier sistema de reconocimiento facial es uno de los métodos de identificación de personas, a partir de sus

características físicas, con una mayor aceptación entre los potenciales usuarios. Sin embargo, a pesar de que varios métodos han sido propuestos durante estos últimos años, su funcionamiento se debe mejorar aun de manera sustancial con el fin de reducir errores de identificación y verificación presentes. La identificación por medio del rostro tiene otra gran ventaja ante las demás características biométricas, ésta es que el reconocimiento de personas puede hacerse a larga distancia usando algún sistema o medio de comunicación tal como un dispositivo móvil.

Recientemente los estudios de investigaciones están enfocados en los sistemas de reconocimiento de personas por medio de sus rasgos faciales debido a la gran potencialidad en aplicaciones, tales como: Control de acceso a lugares restringidos o a información confidencial, apoyo a la ley, etc.

Este trabajo propone una herramienta que permita el reconocimiento de usuarios legítimos. Sin necesidad de conocer el nombre de la persona, sin tener que teclear o escribir datos, solo tomando una foto con nuestro teléfono móvil podremos obtener toda la información sobre la persona, es decir, que los usuarios sean realmente quienes dicen ser, por medio de la implementación biométrica de reconocimiento facial. Este tipo de autenticación o verificación facial (biométrica) es un método pasivo y no invasivo, ya que es una característica común de cualquier ser humano y no requiere de hardware especial para la captura de imágenes.

1.4. Objetivos

1.4.1 Objetivo general

Desarrollar una herramienta computacional que permita el reconocimiento y la verificación de la identidad de una persona por medio de su rostro a través de un dispositivo móvil. Dicho reconocimiento tiene que ser robusto a cambios de iluminación, expresiones faciales, y la de oclusión parcial, con un alto grado de acierto ante posibles ataques de robo de identidad.

1.4.2 Objetivos específicos

- Detección de la imagen (rostro) a analizar.
- Analizar los diferentes tipos de extracción de características a usar.
- Seleccionar características.
- Calcular la distancia de salida entre pares de rostros.
- Realizar evaluación al método de verificación facial.

1.5. Hipótesis

El diseño e implementación de una herramienta computacional basada en un dispositivo móvil, permitirá la verificación(autenticación) facial confiable, a fin de brindar una solución para minimizar el problema del robo y suplantación de identidad.

1.6. Metodología

La verificación facial se basará en seguir el proceso de detectar los rostros, alinearlos, representarlos y clasificarlos [3][2]. Asimismo, se utilizará como apoyo el desarrollo

OpenSource de OpenFace [1] el cual incluye un modelo pre-entrenado de red neuronal convolucional basado en FaceNet [11].

En la figura 1.1 se pueden apreciar las etapas necesarias que se deben seguir para poder hacer un procesamiento de una imagen. a) La obtención de la imagen del rostro se hace a través del dispositivo móvil y de manera sencilla dado que, al tomarse de forma vertical, se puede obtener la imagen rotada correctamente sin problemas. b) Pre-procesamiento: reducir el entorno que no es de interés para el problema fondo, ruido, etc. c) Extracción de características: seleccionar y extraer "características" apropiadas para la identificación de los objetos deseados. Las imágenes procesadas son enviadas una a la vez por la red neuronal convolucional la cual devuelve un vector de 128 dimensiones para cada imagen. d) En primer lugar, se obtienen las imágenes de una base de datos. Finalmente, se convierte cada archivo en matrices de pixeles para que puedan ser procesados por el resto del servicio. e) Mostrar información, una vez obtenidos los resultados de la red neuronal convolucional, se comparan ambos vectores de características obteniendo la distancia euclidiana entre ambos.

.

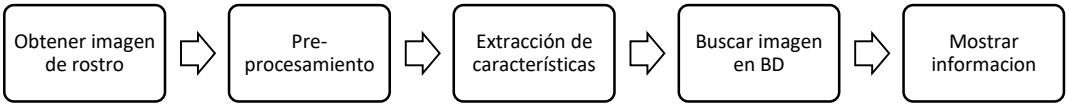

Figura 1.1 *Etapas del procesamiento de imágenes*

Una red neuronal convolucional se utiliza para la extracción de características. Es de gran importancia extraer características locales en vez de centrarse en los pixeles de manera específica en una imagen. Muchos objetos pueden aparecer distorsionados o en distintas posiciones haciendo que sea necesario que características generales describan la imagen en

su conjunto o por áreas. Por esta razón, se debe dar importancia a las regiones de la imagen para así detectar características en diversos tamaños y posiciones. Este comportamiento puede ser replicado en una red neuronal forzando a las capas ocultas a combinar fuentes de información local de la imagen. De esta forma, pueden aparecer distintas características especiales en distintos lugares de la imagen y ser detectados de igual manera [12].

Las redes neuronales convolucionales son una extensión de las redes neuronales clásicas, pero con más dimensiones al recibir valores matriciales de imágenes en más de un canal. Asimismo, estas redes tienen varias características que las diferencian entre las que resaltan el compartir pesos entre neuronas y el uso de pesos matriciales por cada neurona [13].

Las redes neuronales convolucionales están subdividas en capas. Las capas más comunes son las siguientes [2]:

- Capas de convolución que deslizan un filtro sobre los valores de características de entrada.
- Capas totalmente conectadas que hallan la sumatoria de los valores de entrada considerando ciertos pesos.
- Capas de submuestreo que suelen obtener el máximo o promedio de regiones espaciales del mapa de características.

La red planteada devuelve un vector de características por imagen. En la fase de verificación facial, la distancia entre ambos vectores se evalúa tomando en cuenta un punto de corte determinando para establecer si se trata de dos fotos de la misma persona.

1.7. Alcances y limitaciones

Se pretende abarcar los estudios relacionados con sistemas biométricos (verificación facial) y desarrollar una herramienta para el desarrollo de técnicas que permitan realizar de forma adecuada el proceso de verificación facial tomando en cuenta algunas variaciones como son la iluminación, las expresiones faciales y la oclusión parcial del rostro.

Cabe resaltar que entre las limitaciones presentes en el proyecto se encuentra el problema de la iluminación ya que, si bien se han implementado técnicas para reducir este inconveniente, existe un reducido porcentaje de veces en que, debido a esto las imágenes presentan errores durante el procesamiento de la imagen.

1.8. Estructura de la tesis

La organización de esta tesis está dada de la siguiente manera:

Capítulo 1: Introducción. Se trata del capítulo actual, en el cual se exponen las razones que han motivado a la realización del trabajo, así como los objetivos perseguidos con el mismo.

Capítulo 2: Trabajos relacionados. Muestra un compendio de los trabajos más sobresalientes sobre el reconocimiento de rostros.

Capítulo 3: Marco teórico. Aquí se detalla la teoría que va a fundamentar el presente trabajo con base al planteamiento del problema que se ha realizado.

Capítulo 4: Metodología. Se plantea el procedimiento seguido para la implementación de la herramienta computacional que permite el reconocimiento y verificación de la identidad de una persona.

Capítulo 5: Pruebas y Resultados. Muestra los experimentos que se llevan a cabo para evaluar las características del funcionamiento de la herramienta computacional.

Capítulo 6: Conclusiones y Trabajo futuro. Presenta las conclusiones que se obtuvieron al desarrollar este trabajo como también lo que sería tentativamente el trabajo futuro.

Capítulo 2 Trabajos relacionados.

2.1 Introducción

El reconocimiento de rostros humanos, es un tema actual que involucra diferentes disciplinas. Para la investigación en el ámbito de la visión computacional es una de las tareas más importantes en el área de reconocimiento de imágenes.

Esta importancia se deriva de su aplicación en diferentes ámbitos, tales como: seguridad, administración, industria, autenticación y etiquetado de imágenes en redes sociales, interacción robótica y biometría por mencionar algunos.

El reconocimiento de rostros humanos a partir de fotografías, se remonta a los años sesenta, mediante un sistema semiautomático diseñado por Woodrow Wilson Bledsoe fundador de Panoramic Research Inc [15].

En el presente trabajo se lleva a cabo la tarea de clasificación de rostros humanos, esta tarea va más allá que únicamente reconocer la existencia de rostros humanos en una imagen, se lleva a cabo la identificación de un sujeto específico. Para este efecto, partimos del hecho de que en la imagen ya se cuenta con un rostro humano, y lo que se propone es reconocer a un sujeto objetivo, sin importar las condiciones en las que la imagen se encuentre; por ejemplo, con variaciones en la pose, la iluminación, el ángulo en que se toma la imagen, la expresión facial (sonrisa, llanto, enojo, euforia etc.).

2.2 La clasificación de rostros humanos

La clasificación de rostros humanos por sujeto de estudio, partiendo de su identificación en una imagen. Se presenta también la revisión de la literatura respecto a los trabajos y métodos más relevantes respecto al reconocimiento y clasificación de rostros humanos en

imágenes desde el ámbito de la visión computacional, sin perder de vista el enfoque proporcionado por las neurociencias. Toda vez que algunas de las características que se consideran para la identificación de rasgos faciales, se consideran para proponer diversos modelos y métodos que se emplean en el reconocimiento de rostros en imágenes; considerando características como las siguientes:

- La unicidad de los rostros.
- El reconocimiento holístico o por características locales.
- Análisis y uso de expresiones faciales.
- El papel del hemisferio derecho del cerebro en la percepción de rostros.

Esta última característica, nos permite introducir el reconocimiento de rostros mediante el modelo de redes neuronales; a partir de la evidencia de que el cerebro humano dedica un sistema específico para el reconocimiento de rostros [17].

En la actualidad existe un gran número de escenarios en donde la identificación y/o reconocimiento de personas se hacen necesarios desde el punto de vista biométrico, para diferentes estudios como sistemas de seguridad y como una tarea fundamental en el reconocimiento de patrones desde el punto de vista de la visión computacional.

Uno de los problemas típicos que presentan los sistemas de seguridad, es el registro de personal que ingresa a un establecimiento. Esta actividad es llevada a cabo en la mayoría de los casos, haciendo uso de elementos como llaves, tarjetas o contraseñas para lograr la identificación de los sujetos que ingresan. Sin embargo, el uso de los elementos mencionados no es suficiente para garantizar la seguridad del sistema, puesto que es fácil suplantar identidades consiguiendo de esta manera acceso antes restringido [16].

Para llevar a cabo el reconocimiento de rostros, se han empleado diversos modelos y métodos entre los que encontramos: el uso de elementos geométricos en los rostros, el análisis estadístico, la lógica difusa, selección de componentes principales y las redes neuronales.

Algunos de los instrumentos aplicados al reconocimiento de rostros humanos, utilizan el algoritmo denominado Eigenfaces [18] y el método de análisis de características locales (Local Feature Analysis (LFA)) [19] ambos basados en el método de análisis de componentes principales.

A continuación, se mencionan algunos de los trabajos que se han desarrollado para la identificación automática de rostros.

2.2.1 Métodos y técnicas empleados para la identificación de rostros humanos.

En los últimos quince años se han desarrollado diversos métodos de detección de rostros, como describe G. Yang y otros en su revisión [20]. Este desarrollo se debe a las múltiples dificultades que existen en las imágenes, tales como:

Pose: Las imágenes del rostro varían según la pose frente a la cámara, con lo cual algunas características faciales pueden quedar ocluidas (por ejemplo, un solo ojo puede estar visible para una foto de perfil).

Presencia o ausencia de componentes estructurales: Características faciales tales como barba, bigote y lentes al estar o no estar presentes generan una gran variabilidad en el color, forma y tamaño de los conjuntos de rostros.

Expresión Facial: La apariencia del rostro es directamente afectada por la expresión en el mismo. Se pueden tener expresiones de tristeza, alegría, molestia, euforia. Que se expresan mediante cambios sutiles, en la posición de los labios, barbilla, ojos etc.

Oclusión: Los rostros pueden estar ocluidos parcialmente por distintos objetos, por ejemplo, por una bufanda. Pero también pueden estar ocluidos por otras caras, como en una imagen de una multitud de personas.

Condiciones de la Imagen: Cuando la imagen es adquirida, condiciones tales como la iluminación y características de la cámara (tipo de sensor, tipo de lente, obturador, etc.) afectan la apariencia del rostro.

A partir de estas características, podemos dividir a los métodos de detección de rostros en imágenes fijas en cuatro categorías principales de acuerdo con J. Yang y otros [21], sin embargo, existen también métodos donde hay una clara superposición en más de una categoría:

2.2.2 Métodos de detección de rostros en imágenes fijas.

a) Métodos basados en Conocimiento: Estos métodos se basan en codificar el conocimiento humano mediante reglas. Generalmente, dichas reglas capturan relaciones entre características faciales.

b) Métodos basados en Características Invariantes: Estos métodos buscan características que no son modificadas a pesar de que el rostro esté sometido a cambios de iluminación, pose y/o ubicación de la cámara, las cuales son utilizadas para la ubicación del rostro.

c) Métodos basados en Enmascaramiento: Estos métodos utilizan la correlación entre una imagen de entrada y patrones almacenados que describen un rostro.

d) Métodos basados en Apariencia: Estos métodos utilizan modelos obtenidos mediante un conjunto de entrenamiento de imágenes, los cuales capturan una representación variable de la apariencia facial.

A continuación, se explican en detalle cada una de estas categorías y se describen los métodos más característicos de cada una de ellas.

a) Métodos basados en Conocimiento.

Estos métodos de detección de cara se basan en reglas obtenidas del conocimiento que los investigadores tengan sobre cómo está constituido el rostro de una persona. Existen reglas simples que describen las características de una cara y sus relaciones. Por ejemplo, una cara aparece en una imagen con dos ojos que son simétricos entre ellos, una boca y una nariz. Las relaciones entre las características pueden representarse por las distancias relativas y posiciones.

Primero se extraen las características faciales de la imagen de entrada, y los rostros candidatos son obtenidos mediante un conjunto de reglas. Luego un proceso de verificación es usado para reducir las detecciones falsas.

El mayor problema de este enfoque está en la dificultad que existe en trasladar el conocimiento humano a reglas bien definidas. Si estas reglas son estrictas, es posible que no se detecte ninguna o muy pocas caras debido a que no se podrá satisfacer la totalidad de ellas. Y si las reglas son muy generales, existirá una gran cantidad de falsos positivos. Otra dificultad es extender esta metodología para detectar rostros en distintas poses, ya que es muy difícil definir todos los casos existentes.

b) Métodos basados en Características Invariantes.

A diferencia de los métodos basados en conocimiento, que buscan relaciones entre las características faciales, los investigadores han intentado buscar características invariantes de las caras para su detección. El supuesto se basa en que los seres humanos son capaces de detectar las caras u objetos en general en distintas poses y condiciones de iluminación, y por lo tanto deben existir propiedades o características invariantes sobre estas variabilidades. Muchos métodos se han propuesto con la intención de detectar primero los componentes faciales y luego inferir la presencia de una cara. Componentes faciales tales como cejas, ojos, boca, nariz y la línea del pelo son comúnmente extraídas por detectores de borde. Utilizando las componentes faciales extraídas, se construye un modelo estadístico para describir las relaciones y verificar la existencia de un rostro. El mayor problema de estos algoritmos es que las componentes faciales obtenidas de la imagen pueden estar seriamente dañadas por efectos de la iluminación, ruido y oclusión, por ejemplo, los bordes de las componentes faciales pueden estar debilitados, en cambio los bordes asociados a las sombras pueden ser más fuertes. Estos métodos se han aplicado principalmente a imágenes en escala de grises. Una imagen a color puede ser fácilmente transformada a la escala de grises.

C) Métodos basados en Enmascaramiento.

En los métodos basados en enmascaramiento, se predefine manualmente un patrón de un rostro estándar. Dada una imagen de entrada, se calcula la correlación entre la imagen y el patrón o máscara para el contorno de la cara, ojos, nariz y boca independientemente. La existencia de una cara es determinada utilizando los valores de la correlación. Este tipo de métodos tienen la cualidad de que son fáciles de implementar. Sin embargo, una sola

máscara no es capaz de manejar variaciones de pose, de escala y de forma. Para resolver los problemas de invariancia de escala y forma se proponen máscaras de múltiple-resolución, multiescala, y máscaras deformables.

d) Métodos basados en Apariencia.

Al contrario de los métodos basados en máscaras, los cuales deben ser predefinidos por expertos, las máscaras en los métodos basados en apariencia son aprendidas mediante un conjunto de patrones elegidos. Por ejemplo, la máscara al utilizar una red neuronal corresponderá a los pesos ajustados de ésta. En general, los métodos basados en apariencia se basan en técnicas de análisis estadístico y *machine learning* para encontrar las características relevantes de imágenes de cara y no-cara. Las características aprendidas corresponden a la forma de los modelos de distribución o las funciones discriminantes que son utilizadas para la detección del rostro. Mientras tanto, se lleva a cabo una reducción de dimensionalidad para mejorar la eficiencia computacional y la eficacia de la detección.

Además, Yang y otros [21], también llevaron a cabo sus experimentos en un conjunto de bases de datos estándar que se muestran en la tabla 2.1 y en la tabla 2.2, donde se ilustran los resultados de la tasa de detección y las tasas de detección falsas.

Tabla 2.1 *Conjunto de pruebas de base de datos estándar para la detección facial [21].*

Conjunto de datos	Localización	Descripción
MIT conjunto de prueba	http://www.cs.cmu.edu/~har	Dos conjuntos de imágenes en escala de grises de alta y baja resolución con múltiples caras en un fondo complejo
CMU conjunto de prueba	http://www.cs.cmu.edu/~har	130 imágenes en escala de grises de un total de 507 caras frontales.
CMU conjunto de prueba de perfil de cara	ftp://eyes.ius.cs.cmu.edu/usr20/ ftp://testing.face_images.tar.gz	208 imágenes en escala de grises con caras en vistas de perfil.
Kodak conjunto de datos	Eastman Kodak Corporation	Caras de tamaño múltiple, pose y bajo iluminación variable en imágenes en color. Diseñado para la detección y reconocimiento de rostros.

Tabla 2.2 *Resultados experimentales en imágenes del conjunto de pruebas 1(125 Imágenes con 483 rostros) y del conjunto de pruebas 2(23 Imágenes con 136 rostros) [21].*

Método	Conjunto de prueba 1		Conjunto de prueba 2	
	Tasa de detección	Detecciones falsas	Tasa de detección	Detecciones falsas
Basado en Distribución	N/A	N/A	81.9%	13
Redes Neuronales	92.5%	862	90.3%	42
Clasificador Naive Bayes	93.0%	88	91.2%	12
Información relativa de Kullback	98.0%	12758	N/A	N/A
Máquina de Soporte Vectorial	N/A	N/A	74.2%	20
Mezcla de análisis de factores.	92.3%	82	89.4%	3
Discriminante Lineal de Fisher	93.6%	74	91.5%	1
SNoW con características primitivas	94.2%	84	93.6%	3
SNoW con características multi-escalas	94.8%	78	94.1%	3
Aprendizaje Inductivo	90%	N/A	N/A	N/A

Como se resume en la tabla 2.2, los resultados experimentales muestran imágenes de diferentes conjuntos de entrenamiento con diferentes parámetros de ajuste que tienen un impacto directo en el rendimiento del entrenamiento. Por ejemplo, la reducción de la dimensionalidad se lleva a cabo para mejorar la eficiencia de cálculo y la eficacia de detección, con los patrones de imagen proyectados en un espacio dimensional inferior para formar una función discriminante para la clasificación. Además, el tiempo de

entrenamiento y ejecución y el número de ventanas de exploración en estos experimentos influyeron en el rendimiento de alguna manera. Hjelmås y Low [22], clasifican las metodologías de detección de rostros en dos categorías principales. Enfoques basados en imágenes, que se subdividen en métodos de subespacio lineal, redes neuronales y enfoques estadísticos.

Enfoques basados en imágenes; La mayoría de los intentos recientes basados en características en el mismo estudio de [22], han mejorado la capacidad de hacer frente a las variaciones, pero aún se limitan a la cabeza, los hombros y parte de las caras frontales. Por lo tanto, se necesitan técnicas para enfrentar situaciones hostiles, como detectar múltiples rostross en una escena desordenada, por ejemplo, fondo de desorden intensivo. Además, este método ignora el conocimiento básico del rostro en general y utiliza patrones de rasgos faciales de un conjunto dado de imágenes. Esto se conoce principalmente como la etapa de entrenamiento en el método de detección.

Desde esta etapa de entrenamiento, el sistema puede detectar patrones de rostros similares de una imagen de entrada. Ahora se establece una decisión de la existencia de un rostro por parte del sistema basándose en una comparación de la distancia entre el patrón de la imagen de entrada y la imagen de entrenamiento con una matriz de intensidad 2D extraída de la imagen de entrada. La mayoría de los enfoques basados en imágenes utilizan técnicas de escaneo de ventanas para la detección de rostros.

El algoritmo de escaneo de ventanas busca posibles ubicaciones de caras en todas las escalas. Este método depende de los algoritmos de escaneo de ventanas. En otra investigación realizada sobre este método que depende de los algoritmos de exploración de ventanas, Ryu y otros [23], en su estudio experimentaron las técnicas de la ventana de escaneo discutidas por [22] en su sistema. Van más lejos para experimentar su sistema,

basándose en una combinación de varios clasificadores para obtener un resultado más confiable en comparación con un solo clasificador. Diseñaron múltiples clasificadores faciales que pueden tomar diferentes representaciones de los patrones faciales. Utilizaron tres clasificadores, el clasificador de características de gradiente que contiene la información integral de la distribución de píxeles que devuelve cierta invariabilidad entre las características faciales.

El segundo clasificador es el de características de textura, que extrae las propias características por correlación (usa la ocurrencia de probabilidad conjunta de píxeles específicos), la varianza (mide la cantidad de variaciones locales en una imagen) y la entropía (mide el desorden de la imagen). El tercer clasificador utilizado aquí es la función de intensidad de píxeles, que extrae las características de intensidad de píxeles de la región de los ojos, nariz y boca para determinar el patrón del rostroa. Además, utilizaron el enfoque de clasificación *gruesa* a *fina* con sus clasificaciones para la eficiencia computacional. Sobre la base de 1056 imágenes que se obtuvieron del conjunto de datos de AT&T, BioID, [23], lograron los resultados presentados en la Tabla 2.3 y la Tabla 2.4. La clasificación de la primera cara de su experimento con respecto al cambio en ambas direcciones *x* e *y* logró una tasa de detección del 80% cuando las imágenes se desplazan dentro de los 10 píxeles en la dirección *x* y 4 píxeles en la dirección *y*. La segunda y tercera cara de su clasificación mostraron una tasa de detección de más del 80% cuando 2 píxeles se desplazaron en ambas direcciones *x* e *y* respectivamente.

Tabla 2.3 *Resultados que muestran el método de escaneo completo exhaustivo y el método de escaneo propuesto [23].*

Prueba DB	Resultados de Detección						Tasas de reducción del # de escaneos
	Método de escaneo completo exhaustivo			Método de escaneo propuesto			
	Tasa de detección	# de falsos	# de escaneos por imagen	Tasa de detección	# de falsos	# de escaneos por imagen	
IMM	96.2%	28	755.418	95.7%	8	72,273	90.4%
Caltech	94.5%	12	1,369.067	93.0%	10	176,674	87.1%
AR	95.7%	22	1,128.541	95.0%	6	142,136	87.3%
WWW	80.7%	46	4,312.203	83.7%	12	513,152	88.1%

Tabla 2.4 *Comparación de rendimiento por diferentes investigadores y sistema propuesto por Ryu y otros [23].*

	Tasa de detección	# de falsos
Método de Rowley	86.2%	23
Método de Froba	87.8%	120
Método de Feraud	86.0%	8
Método propuesto(Búsqueda ordinaria a fina)	86.6%	19
Método propuesto(Búsqueda completa)	89.1%	32

Como se ve en la Tabla 2.4, su sistema logró una tasa de detección entre 93.0% y 95.7%. Rowley y otros 1998 [24], en su estudio sobre detección de rostros basados en redes neuronales, experimentó en su sistema el cual aplica un conjunto de filtros basados en redes neuronales a una imagen y luego utiliza una referencia para combinar las salidas. Probaron su sistema contra dos bases de datos de imágenes. La base de datos de CMU, compuesta por 130 imágenes y la base de datos FERET, logró alcanzar una tasa de detección del 86.2% con 23 detecciones falsas. Feraud y otros [25] también experimentaron con una técnica de detección de rostros basada en redes neuronales. Utilizaron una combinación de diferentes componentes en su sistema (filtro de movimiento, filtro de color, filtro de red previa y red neuronal grande).

El filtro de red previa es un único perceptrón multicapa, con 300 entradas correspondientes a los tamaños extraídos de las subventanas, ocultas con 20 neuronas y genera una cara/no cara para un total de número de pesos [referencia]. Estos componentes, con una combinación de red neuronal, lograron una tasa de detección de 86.0% con 8 detecciones falsas, en base a una base de datos de 8000 imágenes de Sussex Face Database y CMU Database, que se subdivide en diferentes subconjuntos de igual tamaño correspondientes a diferentes vistas. Tabla 2.5 y Tabla 2.6 muestra los resultados experimentales llevados a cabo por estos investigadores.

Tabla 2.5 *Mostrando los resultados de la base de datos de Sussex Face [25].*

Resultados en la base de datos de Face Sussex						
Orientación (grado)	CGM1	CGM3	CGM5	Ensamble	Ensamble condicional	Mezcla condicional
0	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %
10	62.5 %	100.0 %	87.5 %	100.0 %	100.0 %	100.0 %
20	50.0 %	100.0 %	87.5 %	87.5 %	100.0 %	100.0 %
30	12.5 %	100.0 %	62.5 %	62.5 %	100.0 %	100.0 %
40	0.0 %	100.0 %	50.0 %	12.5 %	62.5 %	87.5 %
50	0.0 %	75.0 %	0.0 %	0.0 %	37.5 %	62.5 %
60	0.0 %	37.5 %	0.0 %	0.0 %	0.0 %	37.5 %
70	0.0 %	37.5 %	0.0 %	0.0 %	0.0 %	25.0 %

Tabla 2.6 *Mostrando los resultados del conjunto de prueba de CMU de A [25].*

Resultados en el conjunto de prueba A de CMU			
Modelo	Tasa de Detección	Tasa de falsa alarma	Falsa alarma
GM	84 %	$1000\ 10^{-6}$	≈20000
CGM1	77 %	$5.43\ 10^{-6}$	47
CGM3	85 %	$6.3\ 10^{-6}$	212
CGM5	85 %	$1.36\ 10^{-6}$	46
Un SWN(Rowley)	84 %	$8.13\ 10^{-6}$	179
Ensamble	74 %	$0.71\ 10^{-6}$	24
Ensamble Condicional	82 %	$0.77\ 10^{-6}$	26
Mezcla Condicional	87 %	$1.15\ 10^{-6}$	39

Wang y otros [26], en su estudio para admitir el detector facial de red neural utilizaron un detector facial basado en red neuronal convolucional multitarea, que se basa directamente en las características de aprendizaje de las imágenes en lugar de las características creadas a mano. De ahí su capacidad para diferenciar rostros en entornos no controlados. El sistema en el que se experimentó utilizó la red propuesta por la región y el detector basado en CNN para la salida de detección final. Experimentaron esto en base a 183,200 imágenes de su base de datos y utilizaron el conjunto de datos AFLW para la validación. Su sistema de detección de rostros se evaluó en conjuntos de datos de rostros humanos AFW, FDDB y Pascal, respectivamente, y logró una tasa de detección de rostros del 98.1%. Los autores no revelaron todos los hechos que llevaron al desarrollo del sistema y tiempo para implementar esto en OpenCV. 2.8, la figura 2.1 muestra las diferentes comparaciones de su sistema con otros estados de artes. Wang y otros [26], analizan su sistema (FaceHunter) con un mejor desempeño que todos los demás modelos estructurados. Sin embargo, esto no puede ser verificado independientemente ya que este sistema fue comercializado.

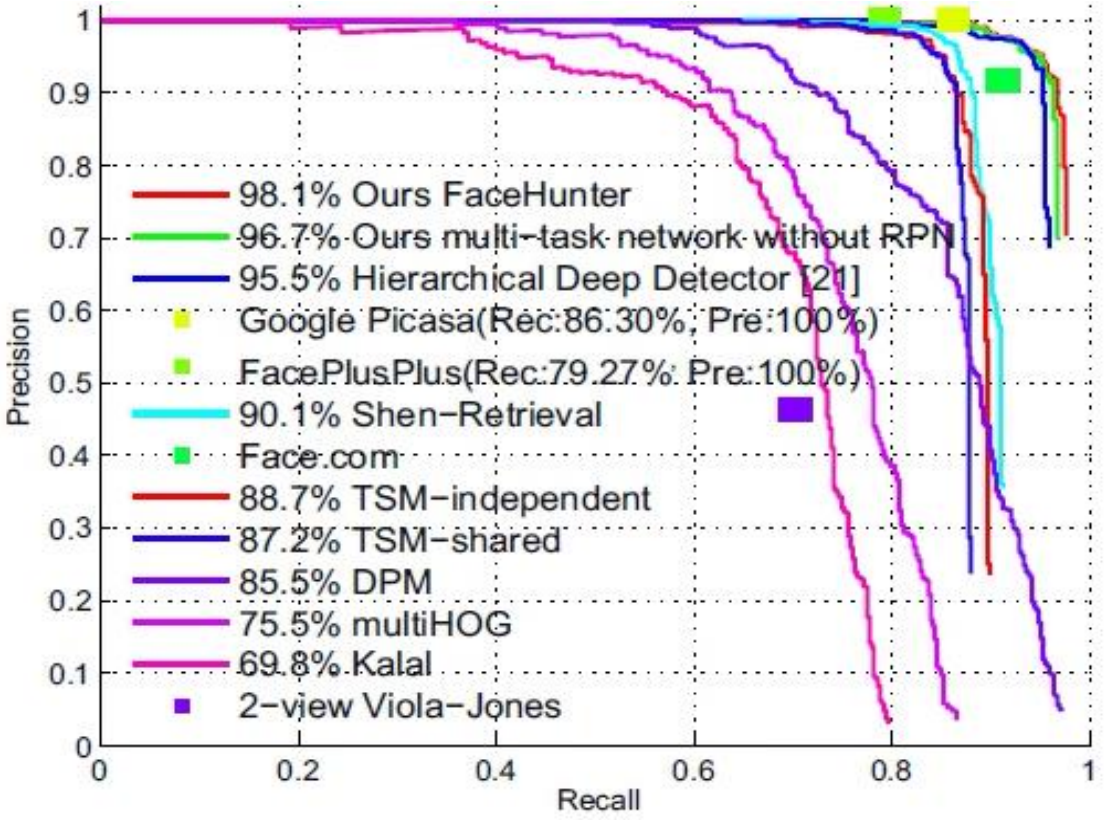

Figura 2.1 *Mostrando la curva PR en AFW [26].*

La otra categoría principal son los enfoques basados en características; depende de las propiedades extraídas que no se ven afectadas por las variaciones en las condiciones de iluminación y la pose. Los investigadores Hjelmås y Low [22], aclaran aún más que "las características visuales están organizadas en un concepto más global del rostro y rasgos faciales que utilizan información de la geometría facial". Esta técnica, será un poco difícil de usar para imágenes que contengan rasgos faciales de un ambiente no controlado. Esta técnica se basa en el análisis de propiedades y la derivación de características para obtener el conocimiento requerido sobre el rostro a detectar. Las propiedades extraídas son el color de la piel, la forma del rostro, los ojos, la nariz y la boca. Por otro lado, en otro estudio de

Mohamed y otros [27], sugiere que "el color de la piel humana es una particularidad muy efectiva usada para detectar rostros pues, a pesar de que las personas tienen diferente color de piel, varios estudios demuestran que la diferencia básica reside en la intensidad más que en su crominancia (color+luminosidad)".

La textura de la piel humana puede por lo tanto ser separada de diferentes objetos. Los métodos de atributos para la detección de rostros usan características para la detección de rostros. Algunos usuarios dependen de los bordes y luego agrupan los bordes para la detección de rostros. Además, sugieren que un buen proceso de extracción incluirá puntos de rasgos elegidos en términos de su fiabilidad de extracción automática e importancia para la representación de rostros. La mayoría de los enfoques basados en características geométricas utilizan el modelo de apariencia activa (AAM) como lo sugieren [28]. Esto permite la localización de puntos de referencia faciales de diferentes maneras para extraer la forma de los rasgos faciales y el movimiento de estas propiedades a medida que la expresión evoluciona Hjelmås y Low [22].

La Figura 2.2 muestra los diferentes enfoques para la detección facial como se informó en un estudio de Hjelmås y Low [22], que puede compararse con la figura 2.3 que muestra la misma clasificación exacta de Modi y Macwan [24].

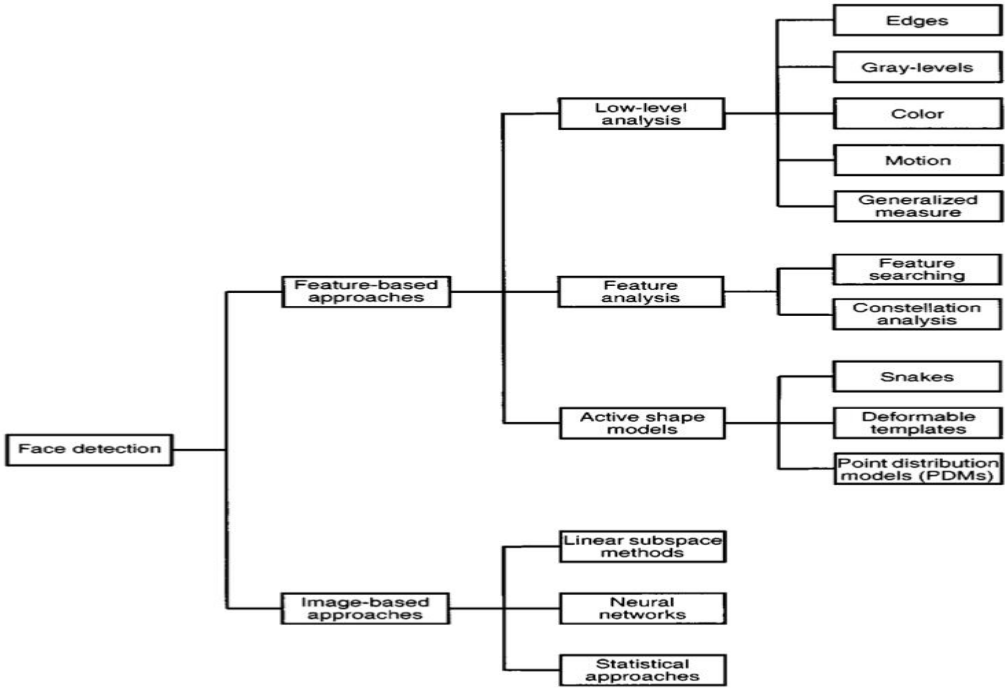

Figura 2.2 *Detección de rostros, clasificada en diferentes metodologías [22].*

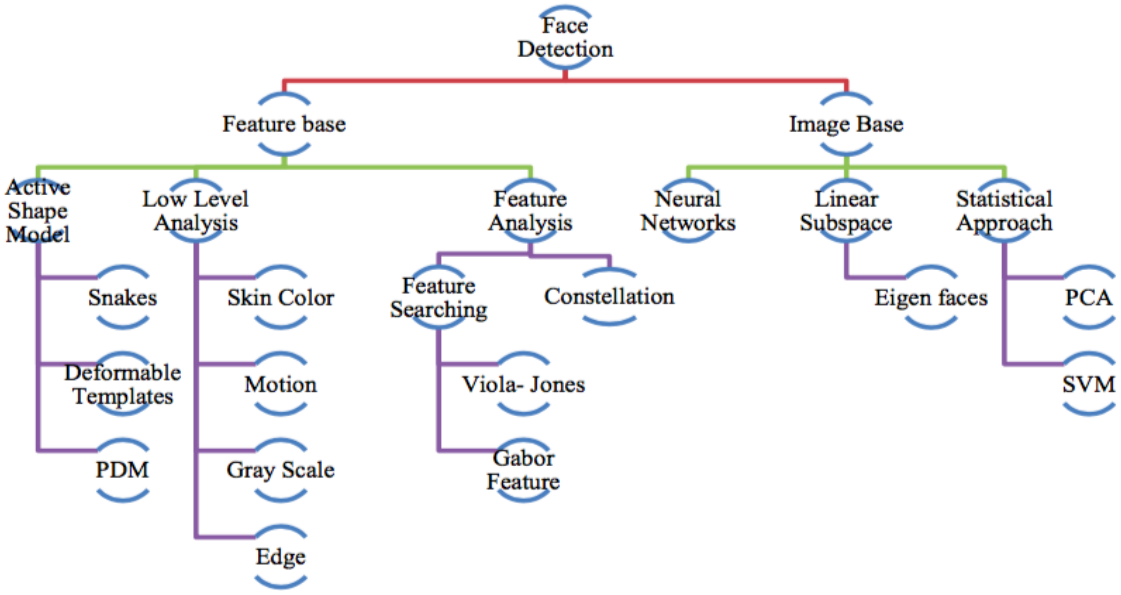

Figura 2.3 *Diversas metodologías de detección de rostros. Modi y Macwan [24].*

Hjelmås y Low [22], en su estudio, muestran un experimento basado en el enfoque de la detección de bordes para la detección de rostros, en un conjunto de 60 imágenes de 9 rostros, con fondos complejos y detectan correctamente el 76% de rostros con un promedio de dos falsas alarmas por imagen. Nehru y Padmavathi [30], en su estudio, experimentaron la detección de rostros basada en el algoritmo Viola-Jones en un conjunto de datos de hombres oscuros y de color para respaldar su afirmación que dice: "Es posible detectar varias partes del cuerpo humano basado en los rasgos faciales presentes", como los ojos, nariz y boca. En este caso, los sistemas como tales deberán ser entrenados adecuadamente para poder distinguir características como los ojos, la nariz, la boca, etc., cuando se usa un conjunto de datos en vivo. El algoritmo Viola-Jones para detectar rostros como se ve en las imágenes en la figura 2.4 que muestra rostros oscuros y de color de piel detectados con precisión.

Figura 2.4 *Detección de rostros en hombres oscuros [16].*

Además, en apoyo de la afirmación realizada por Nehru y Padmavathi [30], la investigación realizada por Viola-Jones para desarrollar el algoritmo Viola-Jones en la detección de rostros ha tenido el mayor impacto en la última década. Según lo sugerido por Mayank Chauhan y otros [31], Viola-Jones en la detección de rostros se usa ampliamente en

aplicaciones genuinas, como cámaras digitales y software de gestión de fotografías digitales. Esta afirmación se basa en un estudio de Viola y Jones [32]. La Tabla 2.7 proporciona un resumen de los resultados obtenidos por estos expertos, que muestran varios números de detecciones falsas y positivas basadas en el conjunto de bases de datos MIT y CMU con 130 imágenes y 507 caras.

Tabla 2.7 *Diversas tasas de detección por diferentes algoritmos que muestran tasas de detección positivas y falsas [32].*

False detections / Detector	10	31	50	65	78	95	110	167	422
Viola-Jones	78.3%	85.2%	88.8%	89.8%	90.1%	90.8%	91.1%	91.8%	93.7%
Rowley-Baluja-Kanade	83.2%	86.0%	-	-	-	89.2%	-	90.1%	89.9%
Schneiderman-Kanade	-	-	-	94.4%	-	-	-	-	-
Roth-Yang-Ahuja	-	-	-	-	(94.8%)	-	-	-	-

Wang y otros [33], afirman que "el proceso de búsqueda de una cara se llama detección facial. La detección facial consiste en buscar rostros con diferentes expresiones, tamaños y ángulos en imágenes que posean luz y fondo complicados y realimentar los parámetros del rostro". En su estudio, probaron la detección de rostros basada en dos módulos que demuestran que un módulo utiliza una combinación de dos algoritmos (PCA con SVM) y el otro módulo basado en tiempo real *field-programmable gate array* (FPGA). Concluyeron con resultados de precisión de detección de rostros del 89%. La Tabla 2.8 muestra los resultados experimentales de dos unidades combinadas para investigar la precisión del sistema.

Tabla 2.8 *Sistema de precisión de detección por Wang y otros [33].*

Número de pruebas	Detección correcta	Detección falsa	Precisión
1000	890	110	89%

Otro método son los **métodos basados en el aprendizaje**, que incluyen técnicas de aprendizaje automático que extraen características discriminativas de un conjunto de datos entrenados antes de la detección. Algunos clasificadores conocidos utilizados para la detección facial, basados en un estudio de Thai y otros [34] son Canny, Análisis de Componentes Principales (PCA), Máquinas de Soporte Vectorial (SVM) y Redes Neuronales Artificiales (ANN). Aunque se usan para la clasificación de la expresión facial, los algoritmos también se usan en la etapa inicial de su experimento, que es la fase de detección. Su experimento logró algunos resultados que se muestran en la Tabla 2.9.

Tabla 2.9 *Comparación de diferentes algoritmos sobre tasas de clasificación [34].*

Método	**Clasificación de Precisión**
Clasificación de expresión facial rápida utilizando redes neuronales artificiales	73.3%
Clasificación de expresión facial utilizando multi-redes neuronales artificiales	83.0%
Sistema propuesto (Canny_PCA_ANN)	85.7%

La tabla 2.10 muestra experimentos de diferentes investigadores. Han utilizado diferentes tamaños de *dataset* de imagen. Algunos han usado una combinación de diferentes algoritmos y han aplicado otros métodos como el filtrado de color, etc. y diferentes conjuntos de entrenamiento para obtener sus resultados. Sin embargo, podemos concluir que el algoritmo Viola-Jones, que es por sí solo, clasifica las imágenes basándose únicamente en las características locales y aún puede detectar con gran precisión y rapidez que los sistemas basados en píxeles [32].

Tabla 2.10 *Comparación de los resultados de diferentes investigadores que muestran la precisión de la detección facial y la detección falsa.*

Referencia	Método	Detección de Precisión	#Falsa Detección
Yang y otros (2002)	Método basado en el conocimiento	83.33%	28
Ryu y otros (2006)	Método basado en imágenes	89.1%	32
Feraud y otros (2001)	Método basado en red neuronal	86.0%	8
Rowley y otros (1998)	Método basado en red neuronal	86.2%	23
Wang y otros (2016)	Método basado en CNN	98.1%	16
Hjelmås y Low, (2001)	Método basado en detección de bordes	76%	30
Viola and Jones (2001).	Viola-Jones	88.84%	103
Wang y otros (2015)	Método basado en (PCA con SVM)	89%	110
Thai y otros (2011)	Método basado en Canny_PCA_ANN	85.7%	N/A

La revisión de los diferentes enfoques vistos en este capítulo ha sido evaluada por los expertos en diferentes bases de datos. Es probable que haya disparidades en la configuración experimental, lo que deja la conclusión de que no existe un enfoque de peor o mejor desempeño. Pero esto puede llevar a hacer una elección para realizar la aplicación que se discutirá en la fase de implementación. Por lo tanto, el propósito principal del reconocimiento facial relacionado con este trabajo es hacer coincidir una imagen facial determinada de un individuo capturado, a una base de datos de rostros conocidos, para identificar a la persona en la imagen de consulta.

Capítulo 3 Marco teórico

3.1. Sistemas Biométricos

Con la evolución de las tecnologías asociadas a la información, nuestra sociedad está cada día más conectada electrónicamente. Labores que eran realizadas por seres humanos son, gracias a las mejoras tecnológicas, realizadas por sistemas automatizados. Dentro de la amplia gama de posibles actividades que pueden automatizarse, aquella relacionada con la capacidad para establecer la identidad de los individuos ha cobrado importancia y como consecuencia directa, la biometría se ha transformado en una tarea relevante [41].

La biometría es la ciencia que se dedica a la identificación de individuos a partir de una característica anatómica o un rasgo de su comportamiento. Una característica anatómica tiene la cualidad de ser relativamente estable en el tiempo, tal como una huella dactilar, el rostro, patrones de la retina o el iris. Un rasgo del comportamiento es menos estable, pues depende de la disposición psicológica de la persona, por ejemplo, la firma.

Cualquier proceso de identificación personal puede ser comprendido mediante un modelo simplificado. Este postula la existencia de tres indicadores de identidad que definen el proceso de identificación:

i) **Conocimiento**. La persona tiene conocimiento del indicador (por ejemplo: un código o contraseña).

ii) **Posesión**. La persona posee un objeto para la identificación (por ejemplo: una tarjeta identificativa electrónica o de crédito).

iii) **Característica**. La persona tiene una característica que puede ser verificada (por ejemplo: su rostro, huella dactilar, iris o su propia voz).

La autenticación más robusta es una combinación del conjunto de los tres indicadores anteriormente descritos.

3.1.1. Propiedades y características

Usualmente se comparan los rasgos biométricos en función de propiedades inherentes al tipo de rasgo biométrico y las características dependen del nivel de eficacia de las tecnologías desarrolladas para el tipo de rasgo biométrico concreto.

Las propiedades de los rasgos biométricos dependen del tipo de rasgo biométrico concreto y son las siguientes cuatro:

1) **Universalidad**. Cualquier persona posee esa característica, huella dactilar.
2) **Unicidad**. La existencia de dos personas con una característica idéntica tiene una probabilidad muy pequeña, como puede ser el caso de dos personas que se parecen mucho en sus rasgos faciales.
3) **Permanencia**. La característica no cambia en el tiempo, la voz.
4) **Cuantificación**. La característica puede ser medida en forma cuantitativa. Como el caso de del ADN u otros rasgos químicos corporales.

Otras características aparecen a la hora de trabajar con determinados rasgos biométricos en función del desarrollo de las tecnologías relacionadas con la adquisición y el procesamiento de los rasgos, como son las siguientes:

a) **Precisión**: Que tan precisos son los sistemas y las tecnologías que trabajan con este tipo de rasgo en concreto, es decir, cómo de bien se identifican a los usuarios en una aplicación.

b) **Rendimiento**: Cómo de potentes y cómo de bien funcionan los sistemas en función del tipo de rasgo. Algunos requerirán procesos y características más complejas que otros.

c) **Usabilidad**: Que tan fáciles de utilizar son estas tecnologías por parte del usuario. Tiene relación con los factores socioculturales y con las aplicaciones de esta tecnología.

d) **Aceptación**: De qué modo está de acuerdo un usuario que se le adquiera un rasgo biométrico concreto. Tiene que ver mucho con aspectos socioculturales y legislativos, y con lo intrusivas que son las técnicas de adquisición de un determinado rasgo biométrico.

e) **Elusión**: Que tan fácil es eludir o burlar a un sistema que utilice determinado rasgo biométrico. Tiene que ver con la facilidad que tienen los usuarios impostores para ser aceptados por el sistema y la facilidad que puede existir para falsificar el rasgo.

Algunos tipos de biometría comúnmente utilizados son la huella dactilar, la retina, el iris, la geometría de la mano, el ADN, la voz, el rostro, la firma, dinámicas de ratón y teclado, etcétera. En la Tabla 3.1 se muestran en una escala cualitativa las propiedades y características de los rasgos biométricos utilizados en este trabajo de tesis [40, 42].

Tabla 3.1 *Características de la biometría facial [68].*

Biometría	Universalidad	Unicidad	Permanencia	Cuantificación
Facial	Alta	Alta	Media	Alta
Precisión	Rendimiento	Usabilidad	Aceptación	Elusión
Alta	Baja	Media	Alta	Alta

La tabla 3.1 se ha rellenado con los datos extraídos de dos estudios publicados en 2011 [68] y 2014 [42], debido a que no se encontraban todas las características tecnológicas: precisión, rendimiento, usabilidad, aceptación y elusión en un mismo estudio. Cabe destacar que las tecnologías de reconocimiento facial han evolucionado desde la publicación de estos estudios comparativos en diferentes factores, siendo estas más potentes, precisas y difíciles de eludir; además de que la biometría está cada vez más integrada en la sociedad y la aceptación y usabilidad por parte de los usuarios es cada vez mayor. También se puede observar que en determinadas características tecnológicas varían los datos presentados en función del autor del estudio como, por ejemplo, en [68] exponen que la elusión para el rasgo facial es baja, mientras que en [42] indica que es alta.

Lo interesante de la Tabla 3.1 respecto al estudio realizado es, sobre todo, la alta precisión y aceptación por parte de los usuarios de las tecnologías de rostro, y de que son rasgos altamente universales y fáciles de recoger, en especial con los sensores actualmente disponibles en los dispositivos móviles actuales.

3.1.2. Verificación biométrica

Es preciso resaltar, sobre todo en este contexto, las diferencias que existen entre verificación biométrica e identificación biométrica:

- **Verificación o autentificación biométrica**. Responde a la pregunta «¿Soy quien digo ser?». La respuesta del sistema puede ser «sí» o «no», y en ocasiones esta respuesta puede venir acompañada de cierto nivel medible de confianza (como puede ser un porcentaje o una puntuación). En verificación biométrica se enfrenta un modelo de autentificación contra otro modelo previamente enrolado en el

sistema. Requiere pues, como mínimo, una comparación de modelos, como se muestra más adelante.

- **Identificación biométrica**. En ocasiones también llamada reconocimiento biométrico responde a la pregunta «¿Quién soy?». La respuesta del sistema es una identidad o un rango de identidades, ordenadas por nivel de confianza. La identificación enfrenta un modelo de identificación contra *n* modelos previamente enrolados en el sistema. Este proceso requiere comparar contra todos los usuarios enrolados en el sistema si no se ha aplicado anteriormente un filtrado previo.

En ocasiones se le denomina *reconocimiento* biométrico al conjunto de técnicas que permiten comparar dos modelos de usuarios: el llamado modelo de *enrolamiento* previamente incorporado o enrolado al sistema contra el modelo del usuario que se desea verificar o identificar, independientemente del número de usuarios enrolados contra los que se realice la comparación de modelos. Por eso mismo, es preciso dejar claro que en este trabajo solamente se manejan técnicas de verificación biométrica, aunque en ocasiones se haga alusión a las técnicas utilizadas en *reconocimiento* biométrico.

En la verificación biométrica se distinguen dos etapas:

1. **Enrolamiento**, *registro o entrenamiento*, es la etapa en la que el usuario accede por primera vez al sistema y completa los datos de su perfil con un modelo biométrico. Este paso es fundamental y necesario para poder realizar todo el proceso de verificación.

2. **Verificación**, *acceso o autentificación*, es la etapa en la que el usuario desea autentificarse como quien dice ser, haciendo uso del mismo rasgo biométrico utilizado en la etapa de enrolamiento.

La etapa de registro debe realizarse al menos una vez, y dependiendo del sistema implementado pueden crearse uno o diferentes modelos de registro que contemplen variaciones ambientales, de entorno o temporales, haciendo así más robusta la verificación. Es importante que los modelos de entrenamiento se generen siempre en las mejores condiciones posibles (dentro de que estas pueden ser distintas) para conseguir los mejores resultados en la etapa de verificación. El entrenamiento se debe realizar, en la medida de lo posible, bajo condiciones de seguridad, para evitar que una persona con ánimo de intentar suplantar al usuario registrado genere un modelo de registro falso, y pueda así suplantarle en la etapa de verificación o corromper su modelo de registro, no permitiendo que el usuario genuino pueda acceder al sistema.

La etapa de verificación puede realizarse indefinidas veces. Para esta etapa es necesario proveer al sistema de una identidad contra la que verificarse y un modelo biométrico de la persona que se quiere verificar. Según el sistema implementado puede compararse el modelo de verificación con uno o varios modelos de entrenamiento, pero siempre del mismo usuario del que se provee la identidad. Es necesario haber generado previamente un modelo de registro para poder compararse con éste en la etapa de verificación.

Se entiende la verificación como un sistema de clasificación con dos clases: usuarios genuinos e impostores. Los usuarios genuinos son aquellos que acceden al sistema con intención de autentificarse como ellos mismos y los impostores como usuarios que intentan autentificarse como otro usuario distinto a sí mismo.

A la hora de comparar dos modelos se genera una puntuación que mide el grado de verosimilitud que existe entre el modelo de entrenamiento y el modelo de verificación. Generalmente esta puntuación es tanto más alta cuanto más parecidos sean los dos modelos entre sí. A la hora de tomar una decisión es necesario fijar un umbral donde si la puntuación

obtenida al comparar dos modelos es más alta que el umbral, el usuario es aceptado, y si ésta es más baja, el usuario es rechazado. Posteriormente se explicará el compromiso que existe a la hora de determinar el valor del umbral de decisión.

Se pueden presentar cuatro situaciones posibles a la hora de realizar una verificación biométrica en función de cuál es la clase del usuario genuino o impostor y de la decisión tomada por el sistema verificador dado un umbral de decisión aceptación o rechazo, como se muestra en la Tabla 3.2.

Tabla 3.2 *Situaciones posibles en la verificación biométrica.*

	Aceptación	Rechazo
Genuino	Verdadero Positivo	Falso Negativo
Impostor	Falso Positivo	Verdadero Negativo

- **Verdadero positivo**: esta situación se produce cuando un usuario genuino se quiere autenticar como sí mismo y el sistema determina que sí que es quien dice ser, aceptándolo. Es una situación deseable, pues obedece al propósito del sistema.
- **Falso negativo:** también llamada falso rechazo. Esta situación se da cuando un usuario genuino se quiere autenticar como él mismo y el sistema le rechaza. Es una situación que hay que evitar, pues puede causar insatisfacción por parte del usuario al intentar verificarse correctamente y no poder hacerlo.
- **Verdadero negativo:** esta situación aparece cuando un usuario impostor o no genuino desea autenticarse como alguien que no es y el sistema le rechaza. Esta situación también es deseable, pues evita que se cometa una suplantación de identidad.

- **Falso positivo:** también llamada falsa aceptación. Esta situación se da cuando un usuario impostor se desea autenticar como otro usuario que no es él mismo y el sistema le da acceso. Es una situación indeseable que además puede acarrear problemas de seguridad.

Normalmente se trabaja con la tasa de verdaderos positivos, número de usuarios genuinos aceptados entre el total de usuarios genuinos, la tasa de falso negativo número de usuarios genuinos clasificados como rechazados entre el total de usuarios genuinos, la tasa de verdaderos negativos número de usuarios impostores rechazados entre el total de usuarios impostores y la tasa de falso positivo número de usuarios impostores aceptados por el sistema entre el total de usuarios impostores.

Un sistema de verificación biométrica funciona correctamente cuando posee una *alta tasa de verdaderos positivos y verdaderos negativos*. Lo ideal es tener un sistema cuya tasa de falsos negativos y positivos sea prácticamente nula y que, al mismo tiempo, el sistema generalice bien en diversas condiciones ambientales y tenga una complejidad sencilla o moderada. Es aquí donde se encuentra el reto de los sistemas de verificación biométrica ya que si el sistema está bien diseñado se establece un compromiso entre generalización y precisión, por lo que existirá tasa de falso rechazo y tasa de falsa aceptación, por pequeña que sea.

Como se ha mencionado anteriormente, es necesario fijar un umbral de decisión para poder, hacer que el sistema tome la decisión de aceptar o rechazar a los usuarios, y así obtener el número de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos a partir de los datos de una base de datos utilizada para evaluar el sistema.

3.2. Biometría facial

La biometría facial es aquella que utiliza la información que se encuentra en el rostro del individuo con el fin de identificarlo de manera automática [41]. El reconocimiento basado en la información del rostro es un método tradicional de identificación de personas y es fácil encontrar una fotografía facial en numerosos documentos administrativos como el INE o pasaporte, con el fin de que la autoridad pertinente verifique que el portador de dicho documento es quien dice ser.

La biometría facial lleva desarrollándose desde finales del siglo XX, pero hasta principios del siglo XXI no se ha explotado en todo su potencial debido a las limitaciones de las computadoras para poder procesar con facilidad y en poco tiempo imágenes con la resolución suficiente para realizar la tarea eficientemente.

La biometría facial tiene dos principales vías de investigación: el reconocimiento facial 2D, que utiliza imágenes o secuencias de imágenes en blanco y negro o en color para reconocer al individuo; y, con la incorporación de cámaras 3D, el reconocimiento facial 3D, que además de captar imágenes o secuencias de imágenes con información de color añade información de profundidad. Estos sensores 3D pronto aparecerán de manera integrada en el mercado de los dispositivos móviles de manera masiva[4], con el ánimo de poder explotar la información de profundidad en imágenes para el desarrollo de diferentes aplicaciones de ocio, seguridad y accesibilidad.

La variación intraclase es aquella que ocurre al tomar diferentes muestras (imágenes 2D o 3D, o vídeos del rostro) de un mismo sujeto en diferentes condiciones o momentos, es el principal reto al que se enfrentan los algoritmos de reconocimiento facial. Algunos

[4] Ver Intel RealSense (Consultado 01/18).

ejemplos de variaciones que se pueden encontrar a la hora de realizar el reconocimiento utilizando biometría facial, y que dificultan esta tarea son los siguientes:

Expresión: cambio en la forma de algún elemento del rostro debido a que se realice un gesto o expresión facial, como pueden ser los ojos (cerrados o abiertos) o la boca (cerrada, abierta, sonrisa, mueca...), gestos con las cejas, etcétera.

Capilares: cambio tanto en el peinado del sujeto como en el vello facial.

Oclusiones: algunas partes del rostro están cubiertas o parcialmente cubiertas utilizando gafas graduadas, gafas de sol, bufandas, sombreros, etcétera.

Pose: el ángulo con el que se ha tomado la imagen del rostro presenta variación que incluso puede llegar a ocluir ciertos elementos de la cara, como puede suceder con las fotos tomadas de perfil.

Iluminación y calidad: la presencia de sombras o la falta de iluminación en la escena, así como el uso de cámaras que distorsionan (ojo de pez), con alta granularidad, ruido o baja resolución (imágenes de rostros por debajo de 60 píxeles de ancho o largo) es un reto para nada despreciable cuando se emplean algoritmos de reconocimiento en imágenes.

Distancia: otro reto al que se enfrenta el reconocimiento facial es diseñar sistemas que sean capaces de identificar personas por su rostro en imágenes con baja resolución o baja calidad [43].

Edad: la variación de la edad afecta en el reconocimiento facial de manera muy notoria en la época desde temprana edad hasta llegar a la juventud y a partir de la tercera edad, siendo más o menos estable durante la edad adulta del individuo.

Artificiales: cambios en el rostro de un individuo debido al uso de maquillaje, cirugía estética, tatuajes en el rostro, piercings o dilataciones en nariz y labios.

Genéticas y fenotípicas: en ocasiones determinadas etnias o diferenciar a dos gemelos con biometría facial puede convertirse en un reto debido a la poca variabilidad interclase —entre distintos sujetos—. También ocurre con individuos físicamente parecidos, aunque no compartan lazos familiares, como los denominados *sosias* o *doppelgangers*.

A continuación, se citan algunos algoritmos clásicos en la detección, el modelado y la comparación de modelos faciales que constituyen el estado del arte en el área del reconocimiento facial. Los algoritmos presentados están diseñados para trabajar con imágenes o secuencias de imágenes 2D, a no ser que se diga lo contrario.

3.2.1. Detección de rostros en imágenes

Como ya se mencionó con anterioridad, la detección facial consiste en encontrar rostros en imágenes donde puede o no haberlos.

Uno de los algoritmos de detección de rostros más extendidos y utilizados es el detector facial desarrollado por Viola-Jones [44], basado en características Haar, que son una familia de wavelets las cuales se escogen entrenando dichos filtros con imágenes de rostros y escogiendo aquellos que mejor detecten caras. En la figura 3.1 el algoritmo recorre la imagen con filtros Haar de diferentes formas y tamaños, para encontrar subimágenes de rostros en posiciones más o menos alejadas de la cámara. La eficacia del algoritmo Viola-Jones es que detecta rostros a gran velocidad, ya que utiliza los filtros Haar en cascada empleando el algoritmo AdaBoost; es decir, se hace una primera pasada donde se encuentran más subimágenes de rostros que las que en verdad existen y progresivamente se van descartando aquellas que sean rechazadas por los subsiguientes filtros más detallados.

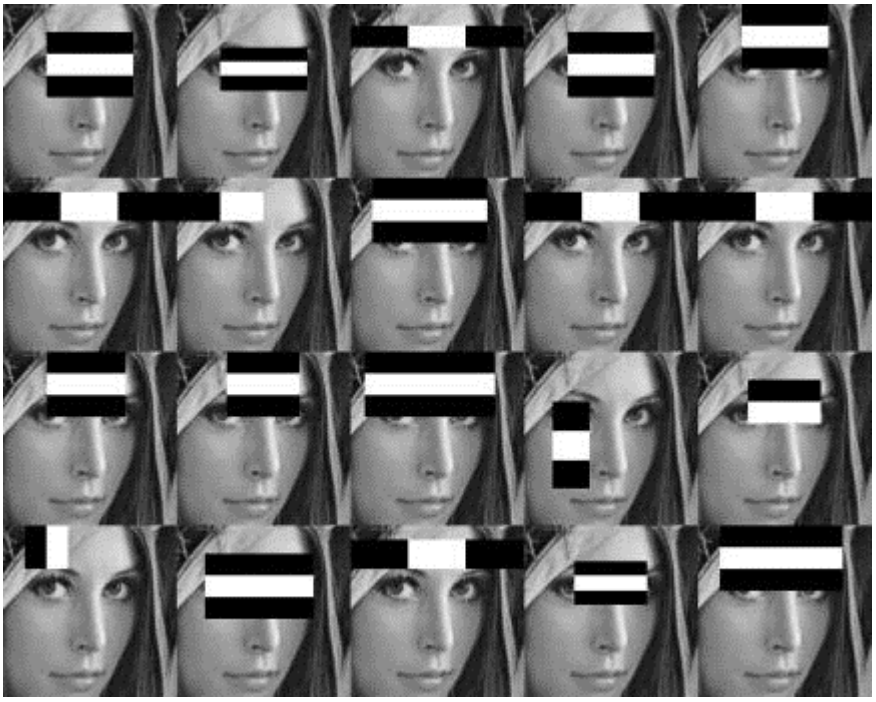

Figura 3.1 *Filtros Haar para detección facial [63].*

Se pueden entrenar también filtros Haar para encontrar distintos elementos del rostro, los cuales se pueden utilizar de forma conjunta filtros entrenados para detectar rostros con el fin de ganar robustez a la hora de comprobar que realmente un rostro detectado en una imagen contiene los diferentes elementos que la componen, como pueden ser ojos, nariz, labios, cejas, etcétera.

El algoritmo de Viola-Jones funciona bien en imágenes donde los rostros se encuentran con pose frontal, sin mucha rotación en el plano de la imagen haciendo este algoritmo más rápido y eficiente si se tienen conocimientos previos de la imagen donde se quiere encontrar rostros: cuántos rostros se desean detectar, posición de dónde se podrían encontrar, si existe un ángulo de rotación del rostro más frecuente que otros, etcétera. El resultado de este algoritmo es un rectángulo que encuadra el rostro y que tiene cuatro

atributos: las posiciones *x* e *y* del píxel de arriba a la izquierda del rectángulo y el ancho y alto del rectángulo en número de píxeles.

Con el desarrollo de algoritmos de Aprendizaje Profundo o *Deep Learning*, una forma más costosa, principalmente a la hora de entrenar el sistema, pero más eficaz de detectar rostros en una imagen es utilizando redes neuronales convolucionales multicapa [64], las cuales solucionan algunos problemas en la detección que podía tener el algoritmo de Viola-Jones, como puede ser rotaciones, cambios de poses y oclusiones, a costa de una elevada cantidad de datos de entrenamiento para la red.

La Figura 3.2 muestra un ejemplo del funcionamiento de una red neuronal convolucional para la detección de rostros en una imagen. La red se entrena utilizando imágenes de rostros, que es lo que se quiere detectar, y cada capa de la red profunda aprende características de diferentes niveles. Los primeros niveles aprenden características de más bajo nivel como esquinas, bordes o puntos, y los últimos niveles aprenden estructuras más complejas, como rostros completos o parciales.

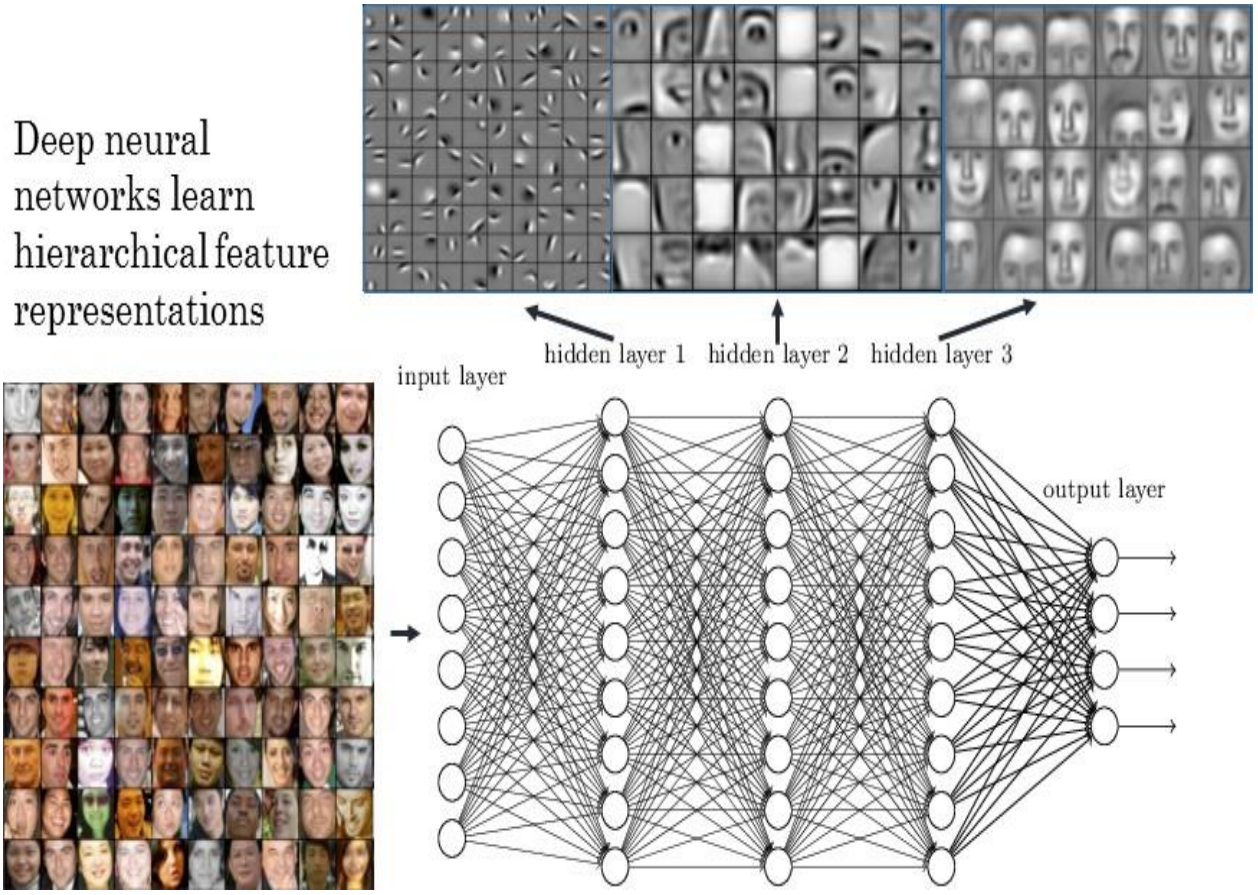

Figura 3.2 *Red neuronal convolucional profunda para la detección de rostros en imágenes [65].*

El resultado de aplicar Aprendizaje Profundo a una imagen para reconocer rostros es el que se presenta en la Figura 3.3. A la izquierda se ve la imagen que se quiere evaluar, donde aparecen los recuadros de los rostros detectados. A la derecha de la figura se ve una imagen que muestra el grado de confianza que tiene la red para asegurar que en ese subespacio de la imagen hay un rostro.

Si se quiere realizar una detección 3D, generalmente se suelen utilizar algoritmos de detección 2D y se aplica el recuadro resultante sobre el plano de profundidad para obtener la imagen del rostro en 3D, aunque también podrían entrenarse distintos filtros o redes que aprendan a detectar subimágenes de rostros utilizando imágenes de profundidad.

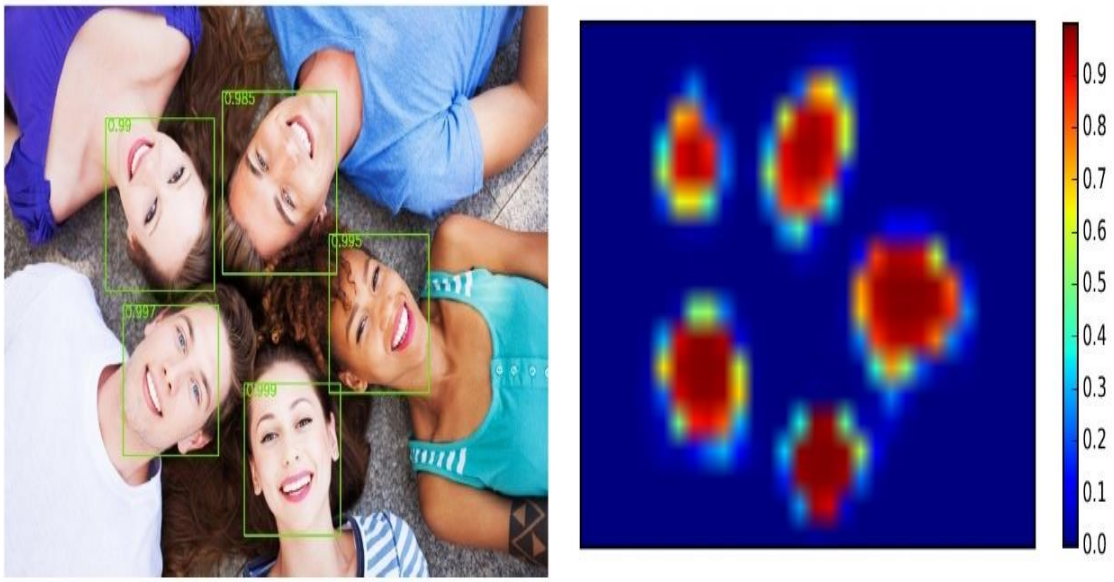

Figura 3.3 *Detección de rostros en imágenes aplicando la técnica de Aprendizaje Profundo [64].*

3.2.2. Características y modelado facial

De acuerdo con Imaoka [45] se pueden clasificar los algoritmos de reconocimiento facial en tres etapas, en función del tipo de características para modelar, su complejidad y cuándo se empezaron a utilizar.

Primera etapa

La primera etapa nace por los años 90, y utiliza métodos de modelado básicos a partir de modelos generativos con características simples y lineales, o bien descriptores de imágenes. Requiere conjuntos de datos de entrenamiento del orden de millares de imágenes de rostros para que el sistema pueda funcionar correctamente, y una capacidad de procesamiento de M-FLOPS[5]. La precisión de estos algoritmos típicamente alcanza valores de EER(Tasa de

[5] El FLOPS es una medida de rendimiento de una computadora que mide el número de operaciones de punto flotante por segundo que se requieren realizar. M-FLOPS indica que el orden es de millones de FLOPS.

Igual Error) entre el 10 y el 20 % [37, 38, 46]. La ventaja de utilizar este tipo de características en la actualidad es que tienen baja dimensionalidad, ocupan poco espacio a la hora de almacenar los modelos, no son computacionalmente costosas de desarrollar y existen publicadas numerosas investigaciones a partir de estos métodos clásicos que mejoran la precisión en la clasificación ya que han sido estudiados durante mucho tiempo.

Las autocaras o Eigenfaces [46] consiste en aplicar la técnica de Análisis de Componentes Principales (PCA) para obtener aquellas características que son las más discriminativas. Este método es un modelo generativo, es decir, cualquier imagen de un rostro puede reconstruirse como una combinación lineal de diferentes autocaras, por eso también puede utilizarse como un método de codificación. El espacio de las autocaras tiene una dimensionalidad equivalente al número de píxeles de la imagen —en una imagen de 100x100 píxeles se podría obtener 10,000 autocaras— pero no todas tienen una información igual de importante; por eso, usando PCA se pueden obtener los autovectores que generen aquellas autocaras que sean las más discriminativas y aporten mayor información. En la Figura 3.4 se observa el conjunto de las 10 principales autocaras para una misma imagen, donde se puede apreciar —sobre todo en los rostros número 4 y 5 de la fila de arriba— que la información que representan éstas es la iluminación en la escena.

Este método se utiliza en verificación facial cuando se tiene un conjunto de imágenes de un mismo usuario para realizar el entrenamiento y así extraer los autovectores y autovalores que generan las autocaras para crear el modelo de entrenamiento. Consecuentemente, el conjunto de autovectores de autocaras generadas para el modelo de verificación deben ser tanto más iguales cuanto más se parezcan las imágenes utilizadas en la etapa de registro y verificación entre sí.

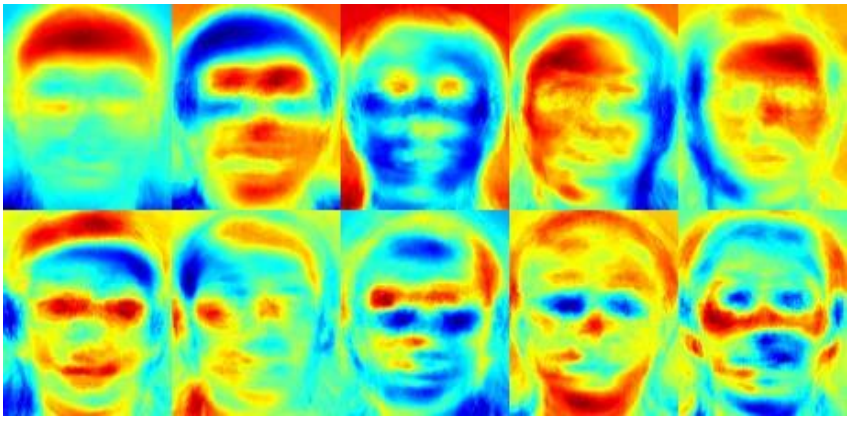

Figura 3.4 *Diez principales Eigenfaces para una imagen [63].*

Las características denominadas *Fisherfaces* [47] resuelven la limitación que pueden tener teóricamente las características Eigenfaces ya que PCA puede ser óptimo para la reconstrucción o codificación de imágenes de rostros, pero puede no ser óptimo para la discriminación entre ellas. Por ello, *Fisherfaces* se basa en Análisis Discriminante Lineal (LDA) —Ronald Fisher (1936)— buscando un subespacio de características que maximice la relación entre la matriz de dispersión interclase y la intraclase. Una vez computado el subespacio y halladas las matrices de dispersión que maximicen dicha proporción, debido a que la matriz de dispersión intraclase es singular [47] se reduce su dimensión utilizando PCA, con el fin de tomar solo las componentes principales que describen dicha matriz. En la Figura 3.5 se muestran las diez *Fisherfaces* más discriminativas para un usuario concreto.

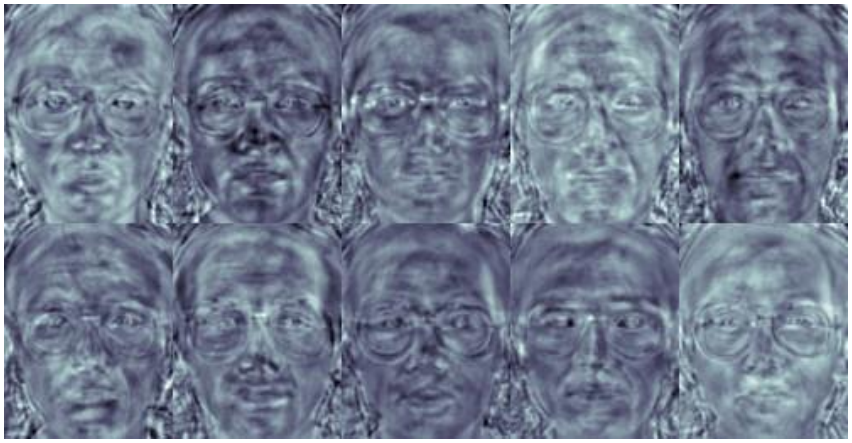

Figura 3.5 *Diez principales Fisherfaces para una imagen [63].*

Fisherfaces mejora con respecto a Eigenfaces a la hora de resolver el problema de reconocimiento y es más robusto a cambios de iluminación, si bien como modelo generativo, no es tan potente como Eigenfaces. Una desventaja de este método es que puede resultar complicado discriminar rostros que se encuentren en el límite entre una clase y otra.

Un descriptor de imágenes ampliamente utilizado cuyo objetivo es disminuir la variación de la iluminación en la imagen es el de los Patrones Binarios Locales (LBP) [65]. La Figura 3.6 ilustra cómo funciona la técnica para extraer los LBPs: en primer lugar, recorre la imagen con un filtro de un tamaño determinado en la figura 3.6.a se utiliza un filtro de conectividad 8 y se obtienen los valores de intensidad de los píxeles bajo el filtro si la imagen es monocromática (figura 3.6.b). Posteriormente se toma como umbral el valor del píxel en el centro del filtro y se umbralizan los valores del resto de los píxeles del filtro respecto a este valor (figura 3.6.c), valiendo 1 si son mayores o iguales y 0 si son menores (figura 3.6.d). Finalmente se escoge un orden en el que recorrer los valores binarios de los píxeles umbralizados, resultando un número binario que describe la textura en ese píxel concreto (figura 3.6.e).

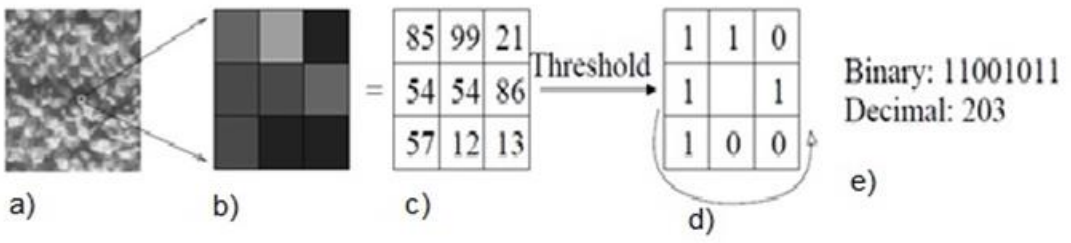

Figura 3.6 *Técnica de descripción de una textura utilizando descriptores LBP [65].*

Se pueden escoger diferentes parámetros del filtro, como puede ser cuántos valores se toman para describir un píxel y que tan alejados están los píxeles utilizados para la descripción. Como vector de características se suelen utilizar los histogramas de los LBPs. Para esto es necesario escoger la resolución (número de barras) del histograma y en cuantas regiones espaciales se divide la imagen resultante de aplicar el filtrado LBP, en ancho y alto, para describirlas con los histogramas.

En la Figura 3.7 se puede observar que los descriptores LBPs son prácticamente iguales para las cuatro imágenes originales, donde se han cambiado la intensidad de luz, robusto a cambios de intensidad.

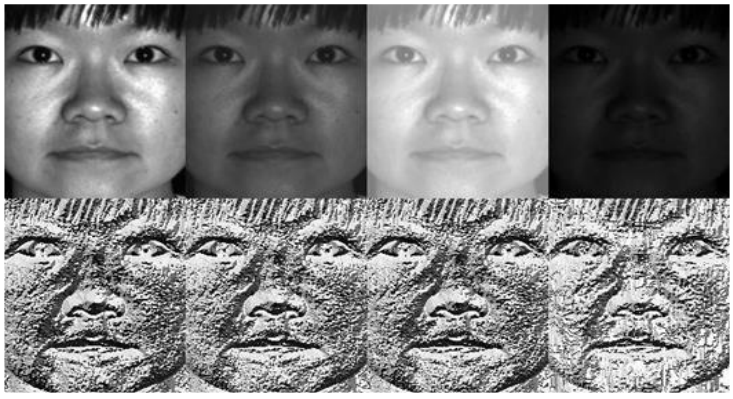

Figura 3.7 *Descriptores LBP para imágenes con diferentes intensidades de luz [63].*

Pese a ser descubiertos en el año 2011, el descriptor basado en la ley de Weber o *Weber Local Descriptor* (WLD) se le puede considerar de primera generación por ser un método de descripción de imágenes. Los descriptores WLD se utilizan para obtener características invariables a cambios de iluminación, como se puede observar en la figura 3.8a. Estas características se basan en la ley de Ernst Weber 1834 que postula que la relación entre un pequeño cambio perceptual en un estímulo y el nivel del estímulo es constante, es decir, los cambios apreciados son relativos, no absolutos. Esta ley se ha utilizado en diferentes campos de aplicación como clasificación de texturas, compensación de la iluminación y muestreo adaptativo de señal figura 3.8b [66].

Para calcular el descriptor de las Weber-faces, se calcula la diferencia entre el píxel central y los adyacentes, todo ello normalizado respecto al valor del píxel central. Con el ánimo de parametrizar la intensidad entre la razón descrita anteriormente para un píxel y su vecino, se ponderan por un factor α y se aplica la función arcotangente a modo de filtro, para reducir valores de alta magnitud positivos o negativos, aplicando la Ecuación 3.1, donde x_c es el valor del píxel central y x_i es el valor de cada uno de los p píxeles adyacentes.

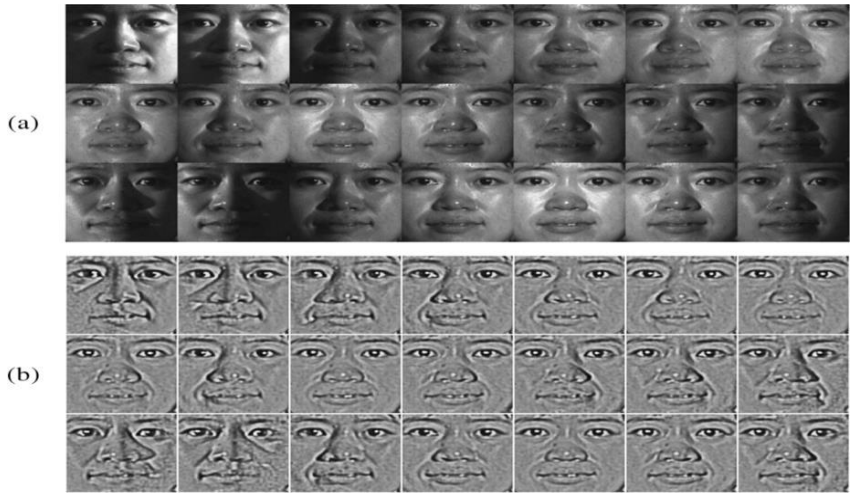

Figura 3.8 *(a) Muestras de la base de datos PIE. (b) Correspondientes Weber-faces [66].*

$$\varepsilon(x_c) = \arctan\left(\alpha \sum_{i=0}^{p-1} \frac{x_c - x_i}{x_c}\right) \quad (3.1)$$

Segunda etapa

La segunda etapa se desarrolla por la década del 2000, donde se siguen aplicando modelos lineales entrenando las transformaciones a partir de grandes conjuntos de datos de entrenamiento. Estos datos suelen ser del orden de millones de imágenes de rostros para que el sistema funcione correctamente, y requieren una capacidad de cómputo de G-FLOPS, y los modelos utilizados son bien generativos, bien discriminativos. La precisión de estos algoritmos típicamente alcanza valores de EER (Tasa de Igual Error) entre el 5 y el 10 % [39, 48, 69, 70].

Los dos métodos comúnmente utilizados son *Representación Dispersa* [67] y *Aprendizaje Métrico* [48]. Además de estos dos métodos se incluye las *Máquinas de Soporte Vectorial* (SVM) [49], debido a que es en esta etapa cuando se empiezan a utilizar vectores de características de elevada dimensionalidad y una técnica ampliamente utilizada consiste en entrenar hiperplanos de alta dimensionalidad para poder delimitar espacios que separen las características de un usuario y las del resto.

La Representación Dispersa [67] nace con el objetivo de disminuir problemas de oclusión, iluminación, ruido y corrupción para el reconocimiento facial, pero a cambio, requiere de un número de características más elevado para que obtenga buenos resultados. La figura 3.9a ilustra la idea: una imagen con oclusiones o una imagen ruidosa se puede representar como una combinación lineal entre el conjunto de imágenes de entrenamiento ponderada

por unos coeficientes figura 3.9b, más un ruido como puede ser las áreas ocluidas (a) o ruido aleatorio (b). De esta manera, aquel coeficiente de mayor valor se corresponde con la imagen de entrenamiento que proporciona mayor verosimilitud con la imagen con oclusiones o ruidosa.

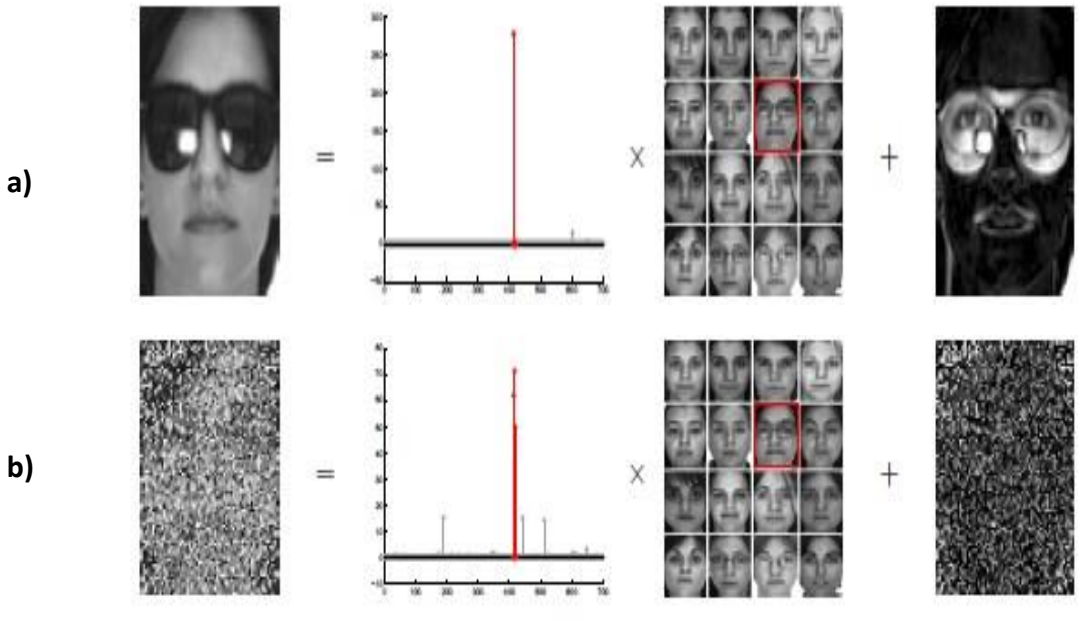

Figura 3.9 *Funcionamiento de Sparse Representation [67].*

Es sencillo visualizar la aplicación de identificación facial utilizando este método. Para aplicarlo en verificación facial, simplemente habría que establecer un umbral de forma que si un coeficiente supera dicho umbral el usuario es aceptado y si ningún coeficiente lo supera, es rechazado.

El *Aprendizaje Métrico* [48] consiste en optimizar una matriz M de forma que la distancia entre una pareja de características de entrenamiento x_e y de verificación x_v sea mínima si pertenecen al mismo sujeto y máxima si pertenecen a sujetos diferentes, como se muestra en la Ecuación 3.2.

$$d(x_e, x_v) = (x_e - x_v)^T M (x_e - x_v). \quad (3.2)$$

Con el fin de entrenar esta matriz se requiere maximizar un criterio de divergencia, de los cuales en la literatura recomiendan o bien el de *Log-Determinant* (LogDet) [48, 70] o bien el de *Kullbach-Leibler* (KL) [45]. Para realizar esto y conseguir buenos resultados, se requiere una gran cantidad de datos de entrenamiento y elegir correctamente los vectores de características y la dimensionalidad de la matriz M con el fin de poder alejar espacialmente características de usuarios similares a la vez que se mantienen cercanas características de un mismo usuario en diferentes condiciones ambientales.

Las *Máquinas de Soporte Vectorial* (SVM) [49] es una técnica consistente en generar un hiperplano en un espacio de alta dimensión con el fin de poder clasificar vectores de elevada dimensionalidad reduciendo el error. Se ha decidido incluir en este apartado la técnica de los SVMs por dos motivos: el primero es porque el hiperplano entrenado define en qué parte del espacio se encuentran los vectores de características de usuario genuino e impostor, luego éste define un modelo que utiliza aprendizaje automático para su adquisición; y el segundo es porque las técnicas de alta dimensionalidad se empiezan a utilizar más en esta generación que en la primera.

La idea tras este método es que al aumentar drásticamente la dimensionalidad no es necesario crear fronteras de clasificación complejas, simplemente con un hiperplano es suficiente, ya que, al estar en un espacio de elevada dimensión, es sencillo encontrar con más facilidad una frontera que separe, sin dejar apenas lugar a error, los vectores de características de usuarios genuinos e impostores. Es habitual utilizar SVMs con funciones de kernel a la hora de definir el espacio de alta dimensión. Existen tres problemas principales que aparecen al utilizar esta técnica: el primero es que se requieren un gran número de vectores de características de usuario impostor y aún más de genuino, para definir correctamente la frontera de decisión; el segundo es que las características

utilizadas deben ser lo suficientemente discriminativas es decir, que exista poca correlación entre usuarios distintos para poder encontrar dicho hiperplano y que sea eficaz; y el tercero es la necesidad de reentrenamiento del sistema completo en el caso en el que se desee incluir más datos tanto de usuarios genuinos o de usuarios impostores.

Los métodos de segunda etapa no se implementan en el trabajo, aunque sí que se consideran evaluar su eficacia para resolver problemas de verificación facial en entornos móviles en un futuro próximo, ya que se estima que no requieren de una computación excesiva que pueda presentar problemas de eficiencia en un *smartphone* o tablet actual.

Tercera etapa

Por último, la tercera etapa comienza a desarrollarse en torno al año 2010, y se dejan de lado las características lineales para pasar a métodos no lineales, con modelos discriminativos y características complejas. Se necesitan una inmensa cantidad de datos de entrenamiento, del orden de miles de millones de imágenes faciales para que el sistema de reconocimiento funcione correctamente, y necesitan una capacidad de procesamiento de T-FLOPS. La precisión de estos algoritmos típicamente alcanza valores de EER (Tasa de Igual Error) entre el 1 y el 5 %, llegando en ocasiones a ser incluso menor [36, 71, 72].

Los algoritmos aquí utilizados utilizan técnicas de *Aprendizaje Profundo*, como redes neuronales convolucionales multicapa donde se transforma el espacio de características de manera no lineal y requiere una enorme cantidad de datos para poder entrenar el modelo. Estos métodos son robustos ante variaciones de la posición del rostro, oclusiones e iluminación.

Estos métodos se están aproximando a la habilidad humana que se encuentran en torno al 97 % de precisión, por eso, ya se comienza a hablar de sistemas de reconocimiento facial que van más allá de la precisión humana [45].

3.2.3. Comparación basada en distancias

Como se ha mencionado anteriormente, para poder realizar operaciones de reconocimiento facial, sea tanto verificación como identificación, se pueden utilizar modelos basados en puntuaciones o en distancias (inverso de la puntuación) entre características. En este subapartado se enuncian medidas de distancia típicas usadas sobre vectores de características e histogramas. Como ya se dijo al principio de este apartado, una distancia devuelve un valor más alto cuanto más diferentes sean las características y es lo inverso a una puntuación o score.

Distancia Euclidiana

La distancia Euclidiana, según la definición del NIST es la distancia entre dos puntos unidos por una línea recta. Es una de las distancias más utilizadas ya que mide la distancia más corta entre dos puntos. La Ecuación 3.3 se utiliza para calcular esta distancia.

$$d_{Euclidean}(\boldsymbol{p}, \boldsymbol{q}) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \qquad (3.3)$$

Siendo **p** y **q** vectores y p_i y q_i, la componente *i-ésima* de cada vector.

Distancia Coseno

La distancia coseno se diferencia de las anteriores debido a que está normalizada en el rango [−1, 1]. Esta distancia mide el coseno del ángulo que forman dos vectores: si estos dos vectores son similares, el ángulo que forman tiende a cero, por lo que su coseno tiende a 1; en cambio, si ambos vectores son totalmente opuestos entre sí forman un ángulo llano, por lo que el valor del coseno es −1. La Ecuación 3.4 muestra la fórmula para calcular la distancia coseno entre dos vectores.

$$d_{Cosine}(\boldsymbol{p}, \boldsymbol{q}) = \frac{\sum_{i=1}^{n} p_i q_i}{\sqrt{\sum_{i=1}^{n} p_i^2}\sqrt{\sum_{i=1}^{n} q_i^2}}, \tag{3.4}$$

Siendo **p** y **q** vectores y p_i y q_i, la componente *i-ésima* de cada vector.

Distancia Chi-cuadrado

La distancia Chi-cuadrado o χ^2 es una medida de distancia utilizada en el ámbito de la visión artificial, ya que se suele utilizar para la medición de similitudes entre histogramas. Esta distancia está basada en el test estadístico con el mismo nombre, el cual evalúa la diferencia entre dos tablas de frecuencias o distribuciones de probabilidad. En la Ecuación 3.5 puede observarse como se calcula esta distancia.

$$d_{Chi-Square}(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{p_i + q_i}, \tag{3.5}$$

Siendo **p** y **q** vectores y p_i y q_i, la componente *i-ésima* de cada vector.

Distancia Bhattacharyya

La distancia *Bhattacharyya*, similar a la distancia Chi-cuadrado es bastante popular en el ámbito de la visión por computadora ya que mide la similitud entre dos distribuciones de probabilidades continuas o discretas. Esta medida está estrechamente relacionada con el coeficiente de *Bhattacharyya*, que mide el solape entre dos muestras o poblaciones estadísticas. La Ecuación 3.6 muestra el procedimiento para calcular esta distancia.

$$d_{Bhattacharyya}(\boldsymbol{p}, \boldsymbol{q}) = -\ln\left(\sum_{i=1}^{n}\sqrt{p_i q_i}\right) \tag{3.6}$$

Siendo **p** y **q** vectores y p_i y q_i, la componente *i-ésima* de cada vector.

Distancia Manhattan

La distancia *Manhattan*, según la definición del NIST[6] es la distancia entre dos puntos medida a lo largo de ejes con ángulos rectos. Se denomina Manhattan porque mide la distancia recorrida por un coche para llegar de un punto a otro en una ciudad de bloques cuadrados, como lo es la ciudad de Manhattan en Nueva York. Es una de las distancias más sencillas de calcular, ya que, como muestra la Ecuación 3.7, no requiere de ninguna operación de complejidad multiplicativa.

$$d_{Manhattan}(p, q) = \sum_{i=1}^{n} |p_i - q_i|\,, \tag{3.7}$$

Siendo **p** y **q** vectores y p_i y q_i, la componente *i-ésima* de cada vector.

[6] National Institute of Standards and Technology

Distancia Mahalanobis

La distancia Mahalanobis es una distancia estadística que mide la similitud que existe entre una muestra y una distribución, siendo ésta cero cuando la muestra coincide con la media. Si se considera que los vectores **p** y **q** de la Ecuación 3.8 tienen la misma distribución, S es la matriz de covarianza de dicha distribución.

$$d_{\text{Mahalanobis}}(p,q) = \sqrt{(p-q)^T S^{-1}(p,q)} = \sqrt{\sum_{i=1}^{n} \frac{(p_i - q_i)^2}{s_i^2}}, \tag{3.8}$$

Siendo **p** y **q** vectores y p_i y q_i, la componente *i-ésima* de cada vector.

Además de los modelos basados en distancias, con los vectores de características se podrían entrenar modelos basados en clasificadores "uno contra todos", teniendo que ser estos modelos actualizables con posteriores enrolamientos de usuarios nuevos en el sistema.

3.3. Reconocimiento facial

El reconocimiento facial ha sido un tema de investigación activo desde la década de 1970 [50]. Dada una imagen de entrada con rostros múltiples, los sistemas de reconocimiento facial suelen ejecutar primero la detección de rostros para aislar los rostros. Cada rostro es preprocesado y luego se obtiene una representación de baja dimensión. Una representación de baja dimensión es importante para una clasificación eficiente. Los desafíos en el reconocimiento facial surgen porque el rostro no es un objeto rígido y las imágenes se pueden tomar desde diferentes puntos de vista del rostro. Las representaciones faciales deben ser resistentes a las variaciones de imagen intrapersonal, como la edad, las

expresiones y el estilo, a la vez que se distinguen las variaciones de imagen interpersonal entre diferentes personas [51].

Jafri y Arabnia [52] proporcionan un listado exhaustivo de las técnicas de reconocimiento de rostros hasta 2009. Para resumir los regímenes, la investigación de reconocimiento facial se puede caracterizar por enfoques holísticos y basados en características. El primer trabajo en el reconocimiento de rostros se basó en las características y se intentó definir explícitamente una representación de rostros de baja dimensión basada en proporciones de distancias, áreas y ángulos [50]. Una representación del rostro explícitamente definido es deseable para un espacio y una técnica de características intuitivas. Sin embargo, en la práctica, las representaciones explícitamente definidas no son precisas. El trabajo posterior buscó utilizar enfoques holísticos derivados de las estadísticas y la Inteligencia Artificial (IA) que aprenden y funcionan bien en un conjunto de datos de imágenes faciales. Las técnicas estadísticas como el análisis de componentes principales (PCA) [53] representan rostros como una combinación de vectores propios [54]. *Eigenfaces* [55] y *fisherfaces* [56] son técnicas históricas en el reconocimiento facial basado en PCA. Lawrence y otros [57] presenta una técnica de IA que usa redes neuronales convolucionales para clasificar una imagen de un rostro.

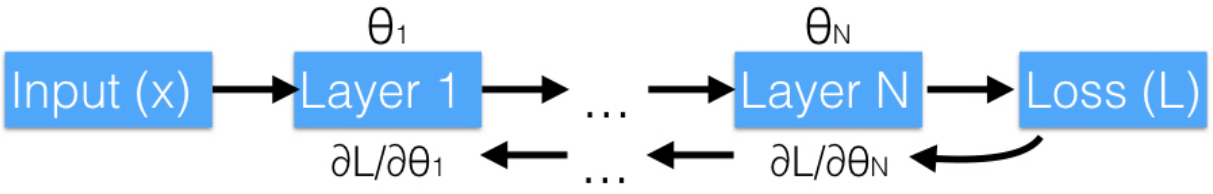

Figura 3.10 *Flujo de entrenamiento para una red neuronal feed-forward [66].*

Las técnicas actuales de reconocimiento facial de alto rendimiento se basan en redes neuronales convolucionales. Los sistemas *DeepFace* [58] de *Facebook* y *FaceNet* [59] de *Google* ofrecen la mayor precisión. Sin embargo, estas técnicas basadas en redes neuronales profundas están entrenadas con conjuntos de datos privados que contienen millones de imágenes de redes sociales con órdenes de magnitud mayores que los conjuntos de datos disponibles para la investigación.

3.3.1 Reconocimiento facial con redes neuronales

En este apartado proporciona una introducción más profunda al reconocimiento facial con redes neuronales a partir de las técnicas utilizadas en *DeepFace* [58] de *Facebook* y los sistemas *FaceNet* [59] de *Google* que se usan en *OpenFace*. Esta sección tiene como objetivo proporcionar una visión general de los dos trabajos más relevantes en este espacio y no pretende ser una descripción completa del campo del reconocimiento de rostros basado en redes neuronales. Otros esfuerzos notables en el reconocimiento de rostros con redes neuronales profundas incluyen el Descriptor de Rostros de Visual Geometry Group (VGG) [60] y las Redes Neurales Convolucionales Ligeras (CNN) [61].

Una red neuronal *feed-forward* consiste en muchas composiciones funcionales o capas, seguidas por una función L de pérdida como se muestra en la Figura 3.10. La función perdida mide qué tan bien la red neuronal modela los datos, por ejemplo, qué tan precisa la red neuronal clasifica una imagen. Cada capa *i* está parametrizada por θ_i, que puede ser un vector o matriz. Las operaciones de capa comunes son:

- **Convoluciones espaciales** que deslizan un núcleo sobre los mapas de características de entrada,

- **Capas lineales** o completamente conectadas que toman una suma ponderada de todas las unidades de entrada, y
- **Pooling** que toma el máximo, promedio o modelo euclidiana sobre regiones espaciales.

A menudo, estas operaciones son seguidas por una función de activación no lineal, como Unidades Lineales Rectificadas (ReLU), que se definen por $f(x) = max\{0, x\}$. El entrenamiento de la red neuronal es un problema de optimización (no convexo) que encuentra un θ que minimiza (o maximiza) L. Con capas diferenciables, $\partial L/\partial \theta_i$ se puede calcular con la retropropagación. El problema de optimización se resuelve con un método de primer orden, que avanza iterativamente hacia el valor óptimo basado en $\partial L/\partial \theta_i$. El trabajo desarrollado por Bengio [63], nos muestra una introducción más completa a las redes neuronales profundas modernas.

La Figura 3.11 muestra el flujo lógico para el reconocimiento facial con redes neuronales. Hay muchos métodos de detección de rostros para elegir, ya que es otro tema de investigación activo en visión artificial. Una vez que se detecta un rostro, los sistemas preprocesan cada rostro en la imagen para crear una entrada normalizada y de tamaño fijo a la red neuronal. Las imágenes preprocesadas tienen una muy alta dimensionalidad para que un clasificador pueda usarlas como entrada. La red neuronal se utiliza como un extractor de características para producir una representación de baja dimensión que caracteriza la cara de una persona. Una representación de baja dimensión es clave, por lo que se puede usar de manera eficiente en clasificadores o técnicas de agrupamiento.

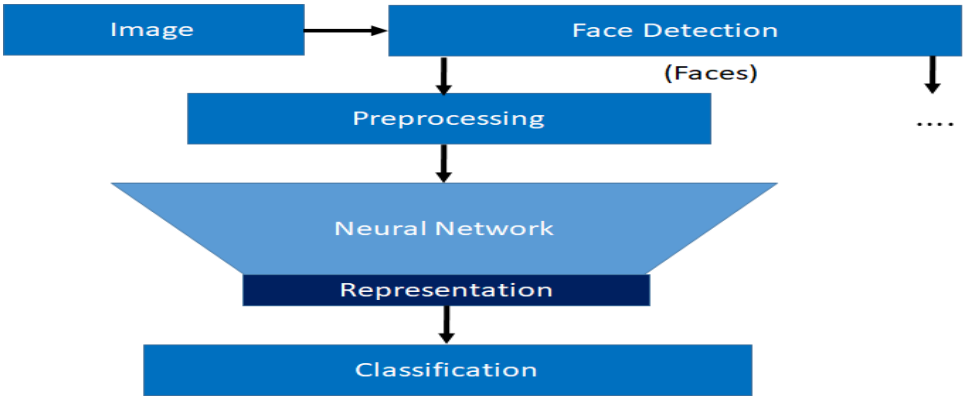

Figura 3.11 *Flujo lógico para el reconocimiento facial con una red neuronal [66].*

DeepFace primero preprocesa un rostro mediante el uso de modelado del rostro en 3D para normalizar la imagen de entrada para que aparezca como un rostro frontal, incluso si la imagen se tomó desde un ángulo diferente. Luego, *DeepFace* define la clasificación como una capa de red neuronal totalmente conectada con una función softmax, que hace que la salida de la red sea una distribución de probabilidad normalizada sobre las identidades. La red neuronal predice cierta distribución de probabilidad $\hat{p}$ y la función de pérdida L mide qué tan bien $\hat{p}$ predice la identidad real de la persona i. La innovación de *DeepFace* proviene de tres factores distintos: (a) la alineación 3D, (b) una estructura de red neuronal con 120 millones de parámetros, y (c) entrenamiento con 4.4 millones de rostros etiquetados. Una vez que la red neuronal se entrena en este gran conjunto de rostros, se elimina la capa de clasificación final y la salida de la capa anterior totalmente conectada se usa como una representación del rostro de baja dimensión.

A menudo, las aplicaciones de reconocimiento facial buscan una representación deseable de baja dimensión que generalice bien a los rostros nuevos en las que no se entrenó a la red neuronal. El enfoque de *DeepFace* para esto funciona, pero la representación es una consecuencia del entrenamiento de una red para una clasificación de alta precisión en sus

datos de entrenamiento. El inconveniente de este enfoque es que la representación es difícil de usar porque los rostros de la misma persona no están necesariamente agrupados, lo que los algoritmos de clasificación pueden aprovechar. La función de pérdida de triplete de FaceNet se define directamente en la representación.

La Figura 3.12 muestra como el procedimiento de entrenamiento de *FaceNet* aprende a agrupar las representaciones faciales de la misma persona. La unidad hiperesfera es una esfera de alta dimensión tal que cada punto tiene una distancia 1 desde el origen. Restringir la incrustación a la unidad hiperesfera proporciona una estructura a un espacio que de lo contrario no tiene límites. La innovación de *FaceNet* proviene de tres factores distintos: (a) la pérdida de triplete, (b) su procedimiento de selección de triplete y (c) entrenamiento con 100 a 200 millones de imágenes etiquetadas, experimentación a gran escala para encontrar una arquitectura de red como referencia.

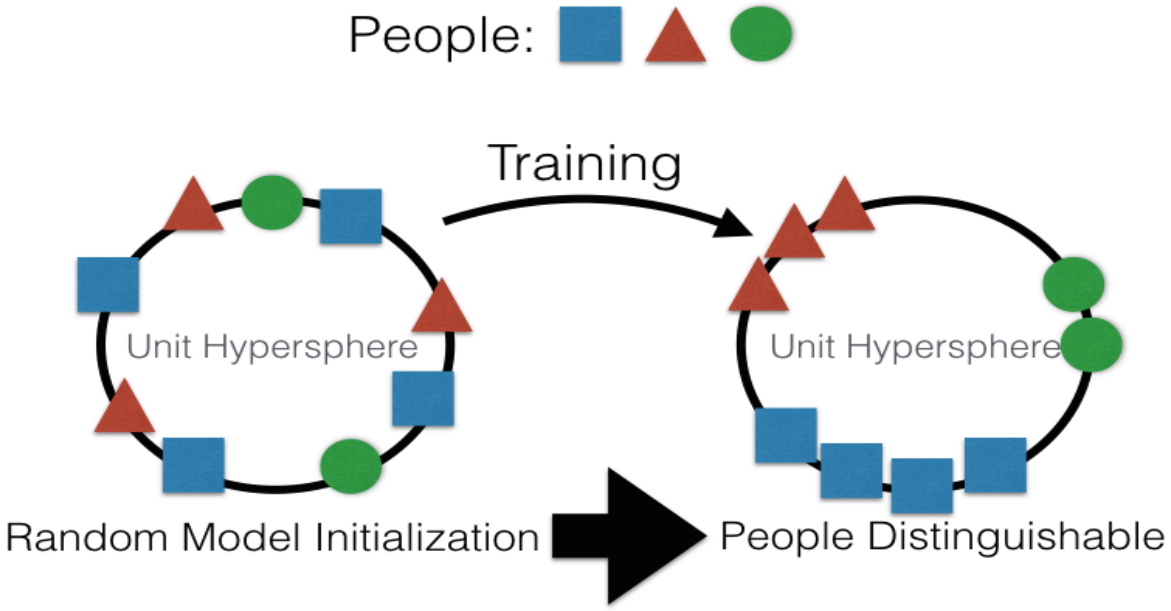

Figura 3.12 *Ilustración del procedimiento de entrenamiento de pérdida de triplete de FaceNet[66].*

3.4 Reconocimiento facial en computación móvil

El acceso a la información desde teléfonos móviles y tabletas se ha convertido en los últimos años en la corriente principal tanto en entornos empresariales como personales. El uso de estos dispositivos para acceder a servicios como las redes sociales, el correo o el comercio electrónico y la banca ha superado el acceso de las computadoras tradicionales [62], convirtiendo a los dispositivos móviles en herramientas esenciales en nuestra vida cotidiana. La movilidad y la ubicuidad del trabajo son herramientas poderosas para aumentar la eficiencia y la productividad en los negocios (y también en la vida personal). Sin embargo, sin el uso adecuado, las empresas y los usuarios pueden estar expuestos a riesgos y amenazas de seguridad.

La seguridad en el acceso a la información es una de las cuestiones más importantes a considerar en los escenarios de movilidad. Las contraseñas han sido el mecanismo habitual para la autenticación de usuarios durante muchos años. Sin embargo, hay muchas preocupaciones de usabilidad y seguridad que comprometen su efectividad. La gente usa contraseñas simples, las reutiliza en diferentes cuentas y servicios, las contraseñas pueden ser compartidas y agrietadas, etc. La cantidad de diferentes cuentas y contraseñas que tratamos en estos días contribuye a hacer más difícil el uso y mantenimiento adecuado. Como resultado, a menudo vemos noticias e informes que alerta de las cuentas y contraseñas robadas [64]. Este problema se vuelve crítico en los dispositivos móviles, ya que pueden ser fácilmente perdidos o robados. Sin embargo, los dispositivos móviles también pueden convertirse en parte de la solución, proporcionando mayores niveles de seguridad debido a sus nuevas opciones y capacidades de autenticación. El uso de la

biometría trae un método de autenticación más seguro y conveniente que las contraseñas tradicionales.

En el 2017 *Biometrics Institute Industry Survey* [67] “el uso de la biometría para el control de acceso móvil se ha establecido como el desarrollo más significativo en el mundo de la biometría en el último año. Además, la encuesta apunta a otras nuevas aplicaciones de biometría en dispositivos móviles, como los pagos móviles o la aplicación de la ley”.

Existen diferentes modalidades biométricas que se pueden integrar en dispositivos móviles: rostro, voz, iris, huella digital, etc. Todos tienen ventajas y desventajas, pero uno de los principales beneficios del reconocimiento facial (junto con el reconocimiento de voz) es que, los *smartphones* ya tienen cámaras integradas, no se requiere hardware adicional. Independientemente de la modalidad biométrica que se utilice, para lograr un sistema realmente efectivo se deben cumplir los siguientes requisitos:

- Usabilidad: la facilidad de uso es un factor clave para lograr bajas tasas de rechazo falso.
- Seguridad: es importante evitar que los impostores tengan acceso al sistema (es decir, baja tasa de aceptación falsa).
- Disponibilidad: el método de verificación debe ser utilizable en cualquier lugar y en cualquier momento.
- El reconocimiento facial cumple estos requisitos y aporta una poderosa solución de autenticación biométrica.
- Es fácil de usar, ya que el usuario ya está familiarizado con el uso de la cámara en el teléfono;

- Los sistemas actuales de reconocimiento facial logran altas tasas de identificación, adecuadas para la autenticación segura [68]; y
- Como se indicó anteriormente, el reconocimiento facial no necesita ningún hardware adicional en los dispositivos móviles. Se aprovecha de la cámara integrada para que esté disponible en la mayoría de los teléfonos inteligentes.

Sin embargo, hay algunos problemas relevantes para el reconocimiento facial en dispositivos móviles que permanecen sin resolver o no lo suficientemente estudiados. Estas preocupaciones deben abordarse en breve para que el reconocimiento facial sea un competidor líder en la autenticación de dispositivos móviles. Algunos de estos problemas son: métodos *anti-spoofing*, protección de plantilla, consumo de energía, disponibilidad bajo escenarios cambiantes y condiciones adversas o rendimiento entre dispositivos.

Capítulo 4 Metodología

4.1 Introducción

En este trabajo se presenta una arquitectura basada en la técnica de redes neuronales convolucionales para extraer las características de interés en una imagen y aplicando la distancia Euclidiana para comparar de manera apropiada estas características. La intención consiste en desarrollar una herramienta computacional que permita el reconocimiento y la verificación de la identidad de una persona por medio de su rostro en ambientes no controlados a través de un dispositivo móvil.

4.2 Metodología Utilizada

En la figura 4.1 se ilustra la metodología utilizada en este trabajo. Las fases mostradas pueden variar si el sistema, las tecnologías utilizadas, o la aplicación lo requiera. Un caso particular de verificación biométrica consta de dos etapas: 1) *detección facial*, donde el usuario crea un perfil biométrico con un modelo de enrolamiento y 2) *verificación,* donde el usuario además de generar un modelo de verificación se compara con el modelo de enrolamiento previamente creado y el sistema decide si el usuario es aceptado o rechazado.

A continuación, se describe cada etapa.

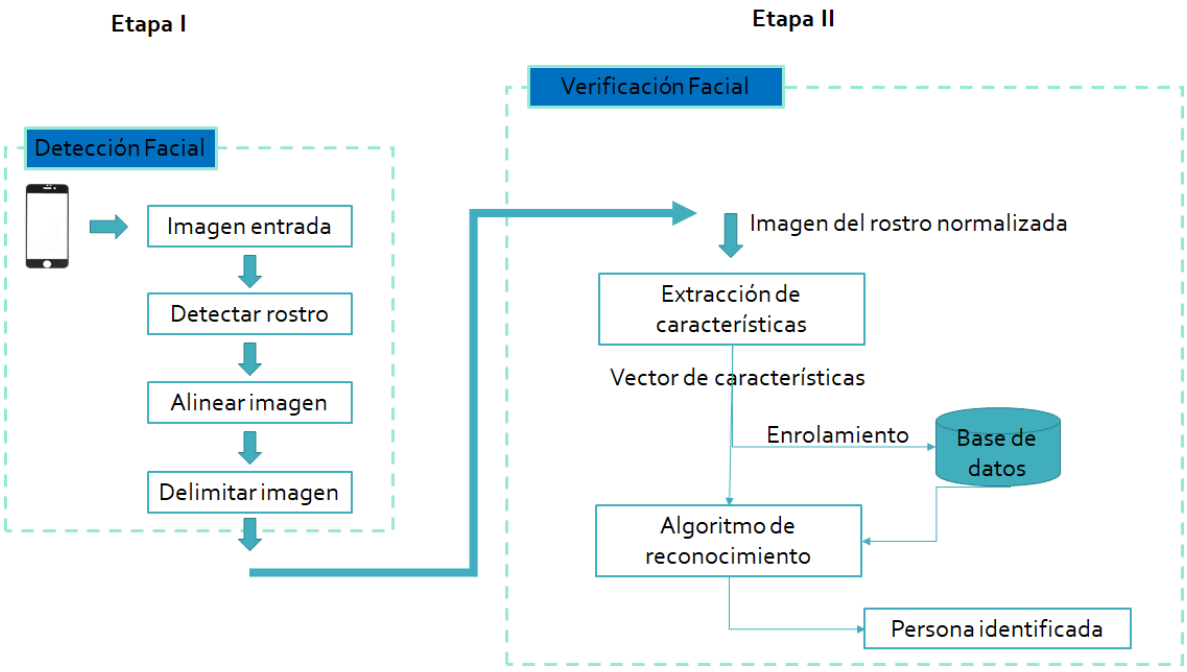

Figura 4.1 *Etapas para el reconocimiento facial.*

4.3 Etapa I: Detección facial

4.3.1 Imagen de Entrada

Se adquiere el rostro en formato RGB, usando la cámara del dispositivo móvil con sistema Android. Se almacenan con un tamaño de 1080 x 1920 pixeles (figura 4.2), se menciona esta resolución en base al dispositivo móvil[7] que se ha utilizado, ésta puede variar de acuerdo a la marca y modelo del dispositivo, en conclusión la resolución es dinámica. Estas imágenes son de manera inicial con el fin de resaltar ciertas características (bordes, orillas) para un procesamiento posterior. Ya que se obtuvo la imagen de entrada, se solicita una identidad, como puede ser un nombre de usuario, apellidos y número de control. En la etapa de registro se asocia la identidad con el modelo de enrolamiento en la fase de almacenamiento, mientras que en la etapa de verificación se provee una identidad, para que

[7] ZTE Blade V8

durante la fase de comparación se compare el modelo de verificación computado en esta etapa con el modelo de enrolamiento asociado a dicha identidad.

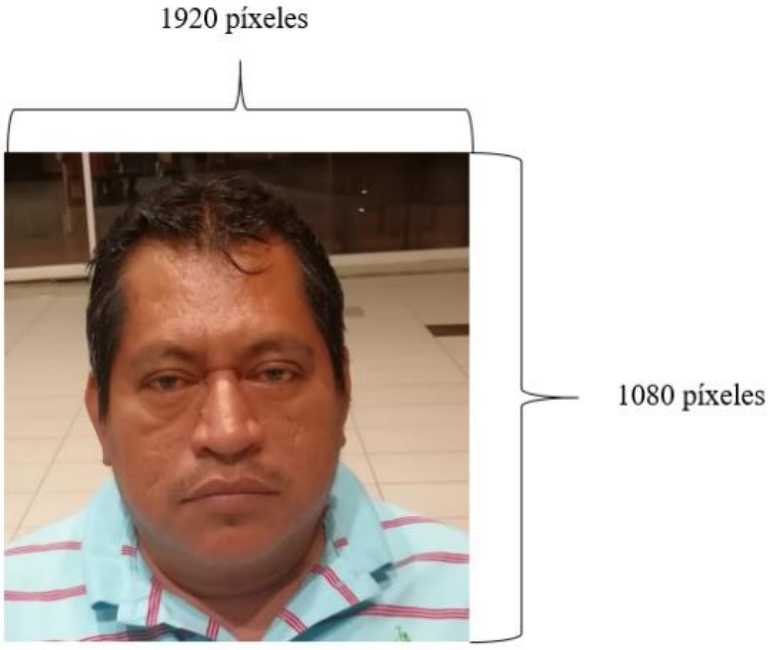

Figura 4.2 *Imagen de entrada.*

Se deben adquirir al menos 10 imágenes del rostro de una persona con el propósito de tener la información mínima requerida para el proceso de autenticación facial. Las imágenes se deben tomar frontal a la cámara del dispositivo móvil, bajo diferentes condiciones de iluminación, expresiones faciales normales (alegría, tristeza, miedo) y poca variación en la posición de la cabeza, tal como se ilustra en la figura 4.2.

4.3.2 Detectar rostro

4.3.2.1 Imagen Integral

La detección del rostro se lleva a cabo aplicando el concepto de imagen integral implementada en el algoritmo desarrollado por Viola y Jones [32]. Este algoritmo utiliza una imagen integral para extraer características de forma rápida y precisa, debido a que no trabaja directamente con los valores de intensidad de los pixeles, sino que lo hace a través de una imagen acumulativa que se va formando basada en las operaciones básicas que se realizan a medida que se recorre la imagen. La figura 4.3 ilustra la aplicación de este proceso con el fin de obtener la imagen integral (recuadro superior derecho) a partir de la imagen original I(x, y). La imagen integral se obtiene al realizar un desplazamiento de izquierda a derecha y de arriba hacia abajo realizando la suma de los pixeles a medida que se va desplazando en la localización de los puntos (x, y), como se muestra en la figura 4.3 aplicando la ecuación. 4.1.

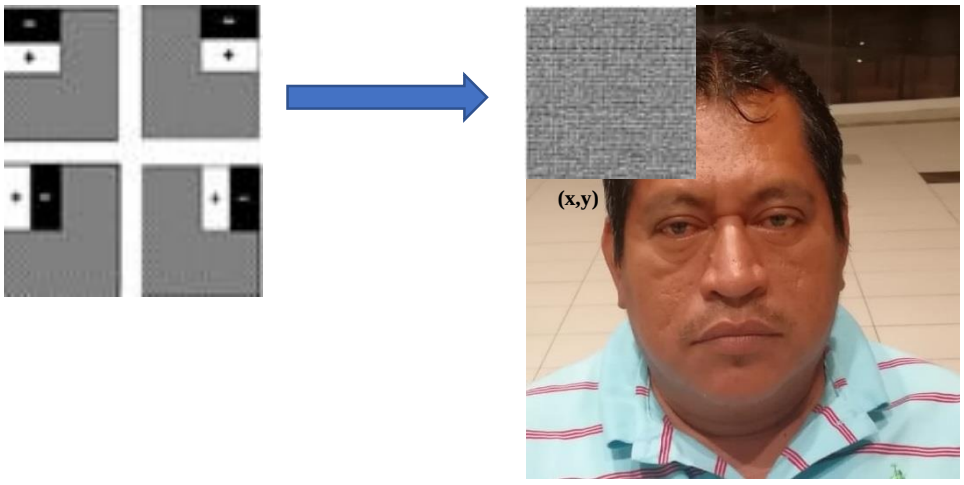

Figura 4.3 *Operaciones básicas aplicadas a la imagen original.*

En la figura 4.4 se muestran los filtros Haar que son utilizados para realizar la codificación de diferencia de intensidades en la imagen. La forma en que estas operaciones básicas se van aplicando se ilustran en la imagen, en la cual se muestran tres operaciones diferentes: suma y resta entre filas, suma y resta entre columnas, suma y resta en diagonal.

Estas operaciones se realizan a través de un proceso de filtrado (Filtros de Haar) o convolución sobre toda la imagen aplicando la ecuación. 4.1

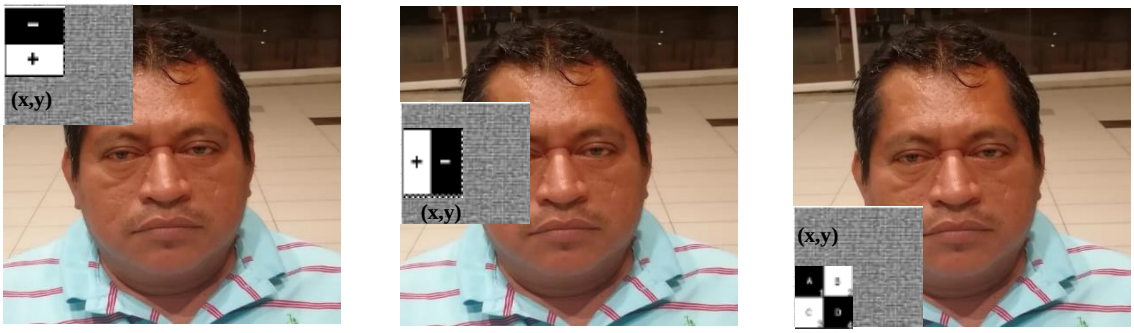

Figura 4.4 *Convolución de filtros para la detección del rostro.*

$$II(x,y) = \sum I(x',y') \qquad 4.1$$

$$x' \leq x\,;\; y' \leq y$$

Dónde:

II (x, y) → Representa la imagen integral

I (x', y') → Representa la imagen original

Condiciones:

x' → Menor o Igual a *x*

y' →Menor o Igual a *y*

4.3.2.2 Detección facial

La fase de detección consiste en encontrar áreas de la imagen que contengan un rostro para aislarlo del resto. Este es uno de los objetivos del trabajo, y una de las fases más importante de la aplicación ya que una mala detección conllevaría un error en el resto de las fases.

En la actualidad, existen distintos algoritmos que pueden ser implementados para una detección facial, muchos de ellos un tanto complejos o específicos según el área donde se desean emplear. Por ejemplo, gracias a la librería OpenCV es posible realizar la detección con el algoritmo de Viola&Jones basado en características Haar. El primer paso dentro de esta etapa consiste en cargar previamente ese algoritmo de un archivo XML.

Una vez cargado el algoritmo, como paso previo para la detección hay que convertir la imagen a escala de grises en caso de que no lo esté. Después, para llevar a cabo la detección se pueden modificar una serie de parámetros que ayudan a limitar y ajustar la detección como:

- Factor de escala: parámetro para determinar cuánto se reduce la imagen para que pueda detectar el rostro. A menor factor de escala mayor precisión, pero mayor retardo también. En este caso se emplea un factor de escala de 1,1, es decir, reducción del 10%.
- Número de vecinos: especifica el mínimo número de píxeles vecinos a tener en cuenta. A menor número menor precisión, pero más detecciones. En este caso se ha estimado que tres vecinos es el número óptimo.
- Tamaño mínimo y máximo: determinan cuales pueden ser el tamaño mínimo y máximo del rostro. En este caso se ha determinado que el tamaño mínimo tenga una altura al menos del 20% de la altura de la imagen.

Además, hay distintas versiones de los algoritmos. Se han comparado, tanto en eficiencia como en rapidez. En general todos proporcionan resultados favorables en cuanto a eficiencia, sin embargo, si se aprecian diferencias respecto a la rapidez. A continuación, en la Tabla 4.1 se encuentra un resumen de la rapidez de cada uno de los algoritmos bajo diferentes condiciones de iluminación y en la figura 4.5 se muestra una comparación de la detección del rostro proporcionado por algunos de los algoritmos probados.

Tabla 4.1 *Comparación de algoritmos de detección de rostro en función de la luminosidad y la velocidad de detección.*

Algoritmo	Luminosidad(lx[8])	Velocidad media (m)s
haarcascade_frontalface_alt_tree	40	36,4
	200	40,9
	450	37,3
haarcascade_frontalface_alt	40	35,3
	200	44,6
	450	34,2
haarcascade_frontalface_alt2	40	41,1
	200	45,9
	450	41,7
haarcascade_frontalface_default	40	56,3
	200	64,3
	450	59,2
lbpcascade_frontalface	40	14,7
	200	15,8
	450	16,3

(a) Original *(b) Haar_default* *(c) Haar_alt2*

Figura 4.5 *Ejemplo detección de rostro utilizando distintos algoritmos.*

[8] Lx: el lux es la unidad del Sistema Internacional para medir luminancia o nivel de iluminación

Como se puede comprobar cada uno de ellos proporciona una rapidez similar independientemente de la iluminación. El algoritmo elegido ha sido el Haar_alt2. En la figura 4.6 muestra un ejemplo de la detección realizada por la aplicación utilizando el método escogido.

Figura 4.6 *Detección facial.*

4.3.2.3 Cómo localizar los puntos de referencia.

Los resultados de detección y el rendimiento de las fases siguientes se verán afectados por la posición en la que se encuentre el rostro que se localizó, siendo la frontal la posición más ventajosa. La correlación entre dos rostros iguales que aparecen en ángulo y rotación diferentes es realmente baja. Por este motivo, y con el objetivo de facilitar las siguientes fases, se debe identificar los puntos clave del rostro para poder realizar el alineamiento facial.

Una de las técnicas más utilizadas actualmente para el alineamiento de rostros es la estimación de puntos de referencia en los rostros, propuesta por Vahid Kazemi y Josephine Sullivan [76,74]. Su método partía de la base como muchos otros de que el alineamiento facial se puede realizar mediante una cascada de funciones de regresión. La diferencia radica en que en su método incluían dos factores clave [73]:

1. En lugar de hacer una regresión sobre los valores de la forma del rostro en función de las características extraídas del sistema de coordenadas global de la imagen, en su método, la imagen se transforma en un sistema de coordenadas normalizado basado en una estimación actual de la forma del rostro. Posteriormente, las características se extraen para predecir un vector de actualización para los parámetros de la forma, y este proceso se repite hasta la convergencia.
2. La forma del rostro estimado debe encontrarse en un subespacio lineal, que puede descubrirse al encontrar los principales componentes de las formas de entrenamiento. Por lo tanto, emplean cierta clase de regresores que garantizan que la predicción se encuentre en un subespacio lineal definido por las formas del entrenamiento.

En esencia, este método consiste en la localización de las coordenadas de los puntos que componen la estructura de un rostro (figura 4.7). Partiendo de un conjunto de datos de entrenamiento de puntos de referencia faciales etiquetados y las “a priori”, es decir, la probabilidad de la distancia entre dos pares de píxeles, se entrena un conjunto de árboles de regresión para estimar los puntos de referencia a partir de la intensidad de los píxeles. Como resultado, se obtiene un detector de puntos de referencia faciales de gran precisión.

Figura 4.7 *Puntos de referencia del método de Vahid Kazemi y Josephine Sullivan [74].*

Una vez calculados los puntos de referencia permitirán, entre otras cosas, aplicar transformaciones afines al rostro para realizar el alineamiento facial, por ejemplo, rotando el rostro hasta que se encuentre en una posición frontal y, de esta forma, mejorar las siguientes fases en el reconocimiento.

Dlib ofrece la implementación de este modelo que permite extraer los 68 puntos de referencia del rostro a fin de conseguir el alineamiento. No obstante, los puntos clave para realizar dicho alineamiento son solo 3: relativos a la posición del ojo derecho, del ojo izquierdo y por último la punta de la nariz (figura 4.8).

Dlib ofrece tanto el modelo de los 68 puntos de referencia como una versión reducida en la que sólo se extraen los 3 puntos clave. Tras observar los resultados, se optó por emplear el modelo de los 3 puntos de referencia.

Figura 4.8 *Ejemplo de la extracción de los puntos de referencia.*

4.3.4 Alineación y normalización del rostro mediante puntos característicos

En los últimos años, con el rápido desarrollo de la biometría, la inteligencia artificial y la nueva generación de tecnología de interacción humano-computadora, las técnicas de procesamiento de imágenes relacionadas con el rostro, como el reconocimiento facial, análisis de la expresión facial, estimación de la pose del rostro, la codificación de la imagen del rostro, etc., han atraído la atención de muchos investigadores. Sin embargo, estas técnicas requieren la información de los puntos característicos de los rostros que se obtienen de la cámara o el vídeo como una condición previa. Es decir, primero se tiene que alinear el rostro a través de la localización de los puntos característicos y después extraer la información de las características del rostro. El tratamiento de los puntos característicos del rostro incluye los puntos del contorno de los ojos, la boca, nariz, cejas, barbilla y mejillas.

La exactitud de la alineación del rostro está afectada por muchos factores, como el tamaño del rostro, la posición, postura, expresión, edad, así como el cabello, las gafas, y los cambios de iluminación. La alineación del rostro es todavía un problema difícil de resolver.

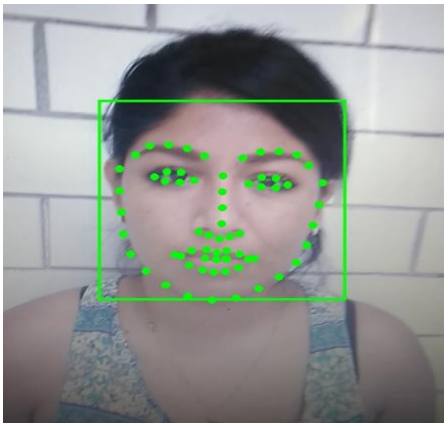

Figura 4.9 *Normalización del rostro.*

4.3.5 Delimitación de la imagen

Después de haber detectado el rostro en la imagen de entrada, el siguiente paso es delimitar el contorno facial de la imagen (rostro) para observar que están presentes todos y cada uno de los componentes faciales (ojos, boca, nariz, cejas, frente, barba). Esto se realiza con la finalidad de dejar solamente el rostro de la imagen, dejando a un lado los otros elementos componentes del rostro como: orejas, pelo, cuero cabelludo y cualquier otro objeto (aretes o algún tatuaje), que pueda estar presente en el rostro y que no sean relevantes para la aplicación.

Este punto es muy importante debido a que la imagen debe estar completamente despejada de cualquier objeto que pueda proporcionar información inadecuada que interfiera en el siguiente proceso. El rostro delimitado y alineado durante la fase de preprocesamiento se muestra en la figura 4.10.

Figura 4.10 *Imagen del rostro delimitado obtenido de la imagen de entrada.*

Hasta este punto se ha desarrollado la primera etapa de esta investigación, ya se tiene un rostro delimitado, detectado en la imagen de entrada.

4.4 Etapa II: Verificación facial

4.4.1 Extracción de características

En esta fase se van a generar los vectores característicos para cada una de los rostros que se pretenden reconocer (figura 4.11). Una vez que se tiene detectada la región de interés (Etapa I) el siguiente paso consiste en extraer las características de esta región con el fin de identificar el vector de patrones que mejor representan a cada una de los rostros que se pretenden reconocer. Este proceso consiste en encontrar el grupo de variables que mejor define los datos de entrada con el objetivo de clasificar los rostros encontrados y asignarlos a alguna de las clases predefinidas en el entrenamiento.

4.4.1.1 Redes neuronales convolucionales

Para el reconocimiento de imágenes se hace uso de las redes neuronales convolucionales (CNN)[9], las cuales son un modelo donde las neuronas corresponden a campos receptivos de una manera muy similar a las neuronas de la corteza visual primaria de un cerebro biológico [80]. La red se compone de múltiples capas. En el principio se encuentra la fase de extracción de características, compuesta de neuronas convolucionales y de reducción (*Pooling*).

A medida que se avanza en la red se disminuyen las dimensiones activando características cada vez más complejas. Al final se encuentran neuronas sencillas para realizar la clasificación.

[9] Neural Networs Convolutional

Existen cuatro operaciones principales que realizan las redes convolucionales.

Capa convolucional

Su principal propósito es extraer características de una imagen. Consiste de un conjunto de filtros entrenables que realizan producto punto con los valores de la capa precedente. En la práctica, los valores de los filtros son aprendidos para su activación al encontrar ciertas características. Al ser colocados en cascada se obtienen diferentes niveles de abstracción.

Rectificador Lineal de Unidad

Son utilizados después de cada convolución. Son una operación que reemplaza los valores negativos por cero y su propósito es agregar no linealidad al modelo, eliminando la relación proporcional entre la entrada y salida.

Pooling

Algoritmo utilizado para reducir las dimensiones, con el objetivo de disminuir los tiempos de procesado reteniendo la información más importante.

Capa totalmente conectada (Dense Layer)

Realiza la clasificación basado en las características extraídas por las capas de convolución y las reducidas por *pooling*. En esta capa todos los nodos están conectados con la capa precedente.

4.4.1.2 Funcionamiento de la Red Convolucional

Las capas de convolución y las de *pooling* se encargan de extraer características mientras que la capa totalmente conectada actúa como clasificador. Para el funcionamiento de este modelo debemos proceder al entrenamiento. Esto implica: 1) Inicializar todos los

parámetros o pesos con valores aleatorios; 2) Utilizar una imagen de entrenamiento y utilizarla en el modelo; 3) Calcular el error total de las probabilidades resultantes del modelo y finalmente; 4) Propagar hacia atrás para calcular el error de gradiente de todos los pesos en la red y utilizar gradiente descendiente para actualizar estos valores y minimizar el error de salida.

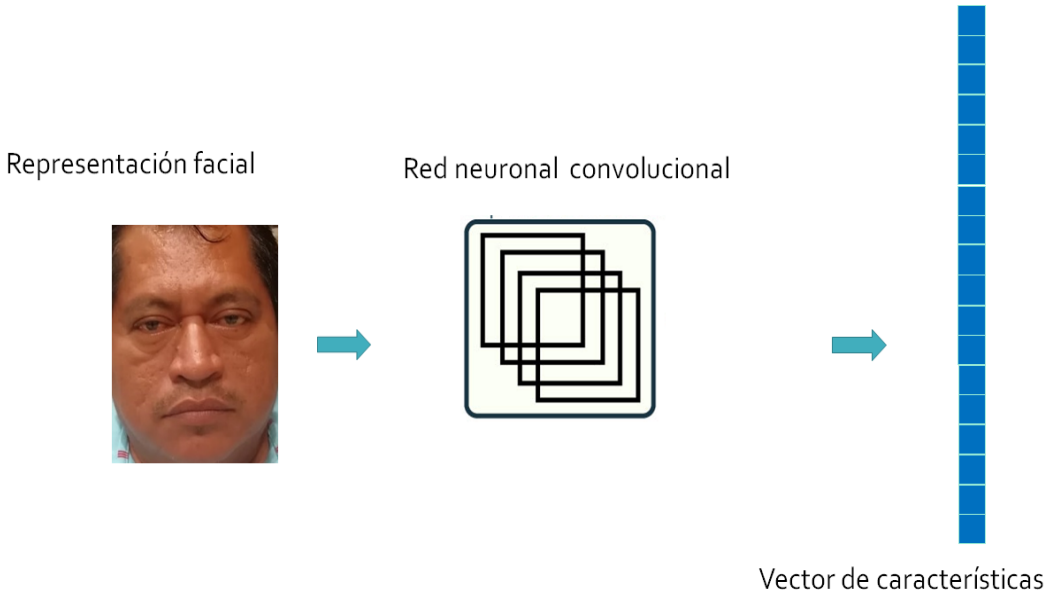

Figura 4.11 *Generación de vectores característicos.*

En la actualidad, y como muchos artículos sugieren, la mejor forma de extraer las características de una imagen es mediante redes neuronales que extraigan directamente las características. El hecho de entrenar una red neuronal para la identificación de rostros desde cero es algo realmente complejo, no solo por el gran costo computacional que ello conlleva, sino también debido a la dificultad de obtener *datasets* de un gran tamaño. Entre las mejores aproximaciones y más utilizadas, se encuentran las basadas en la extracción de vectores de 128 unidades dimensionales que representa cada rostro en 128 bytes, como se

muestra en el artículo publicado por *Google* [75] donde se expone una aproximación para el entrenamiento de una red neuronal de estas características.

En su artículo presentan su sistema *FaceNet*, el cual aprende directamente un mapeo de imágenes de rostros en un espacio euclídeo compacto, en el cual las distancias se corresponden directamente con una medida de similitud del rostro. Una vez que este espacio ha sido generado, es posible implementar con facilidad tareas como el reconocimiento facial o *clustering*, haciendo uso de técnicas básicas y usando las activaciones (incrustaciones) de *FaceNet* como vectores característicos.

La incrustación es una representación genérica para el rostro de cualquiera. A diferencia de otras representaciones de rostros, esta incrustación tiene la propiedad de que una distancia mayor entre dos incrustaciones de rostros significa que los rostros probablemente no sean de la misma persona. Esta propiedad hace que las tareas de agrupación, detección de similitudes y clasificación sean más fáciles que otras técnicas de reconocimiento facial en las que la distancia euclidiana entre características no es significativa.

En este punto en lugar de entrenar una red neuronal propia se optó por emplear un modelo ya entrenado. De entre todas las opciones disponibles se eligió la red neuronal proporcionada por Dlib. Esta red cuenta con un 99.38% de precisión en el test de Labeled Faces in the Wild[10]. La red es una red neuronal convolucional formada por 29 capas de convolución y es una versión de la red ResNet-34 expuesta por He, Zhang, Ren y Sun en su artículo Deep Residual Learning for Image Recognition, con la diferencia de que se han

[10] Base de datos de imágenes de carátulas etiquetadas destinadas al estudio del reconocimiento facial

empleado la mitad de los filtros en cada capa y se han eliminado algunas otras [76]. En su artículo, He, Zhang, Ren y Sun exponen el problema de que cuando las redes neuronales profundas comienzan a converger surge un problema de degradación, y es que cuanto más profunda se vuelve la red, es decir, más capas tiene, la precisión se satura y comienza a degradarse rápidamente [77].

Su solución pasa por emplear el aprendizaje profundo (*Deep Residual Learning*). En lugar de esperar que pocas capas no lineales encajen directamente con el mapeado subyacente deseado, se deja explícitamente que estas capas se adapten a un mapeo residual. Se define la función residual como F(x) = H(x) — x, donde F(x) y x representan a las capas no lineales y a la función identidad (entrada=salida) respectivamente y se sostiene que es más sencillo optimizar la función de mapeo residual que optimizar el mapeo original [78].

La red que diseñan emplea filtros de 3x3 en las capas convolucionales, una reducción de resolución mediante redes convolucionales con paso 2 y la red finaliza con una capa *pooling* de promedio global y una capa *SoftMax* con 1000 neuronas completamente conectada [78] (figura 4.12). Juntas, estas capas extraen las características útiles de las imágenes, introducen la no-linealidad en la red y reducen la dimensión de las matrices de características.

El entrenamiento de la red de Dlib se ha realizado utilizando varios *datasets* de más de 3 millones de rostros y siguiendo el modelo de la ResNet-34.

La implementación que Dlib hace de esta ResNet-34 proporciona en su salida el vector de 128 características del rostro que permitirá realizar la identificación de las personas.

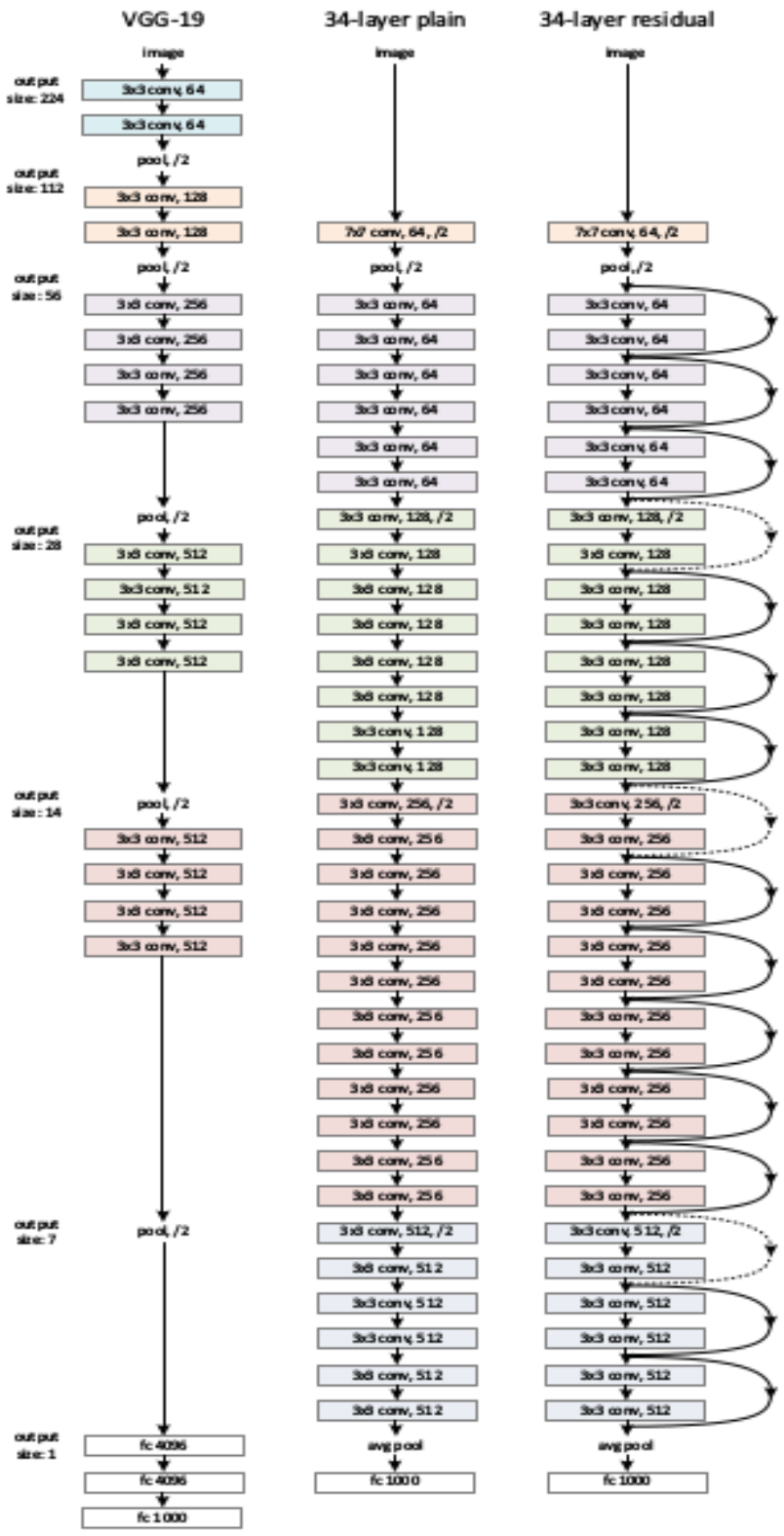

Figura 4.12 *Esquema de una red residual de 34 capas. [77]*

4.4.2 Clasificación de los vectores característicos del rostro.

En la última fase, será necesario comparar los vectores característicos de los rostros de las personas a las que se debe determinar la verificación, con los de aquellos rostros y datos que hay almacenados en la base de datos. Este paso puede realizarse empleando cualquier tipo de clasificador como podría ser uno tan sencillo como el de la distancia euclídea o, por ejemplo, un K-NN [75]. Para llevar a cabo un reconocimiento o identificación facial significa dar una imagen del rostro y se requiere que el sistema diga quién (si él o ella) es la más probable identificación. En este procedimiento se dice que la coincidencia es de 1:K, donde K representa el número de clases, es decir, se compara la imagen de entrada con las K existentes en la base de datos para concluir si se trata de una coincidencia o no (Figura 4.13).

En el caso de que el vector característico se asemeje lo suficiente, es decir, esté lo más cerca posible de uno de los vectores característicos de los rostros de nuestra base de datos, significa que pertenecen a la misma persona.

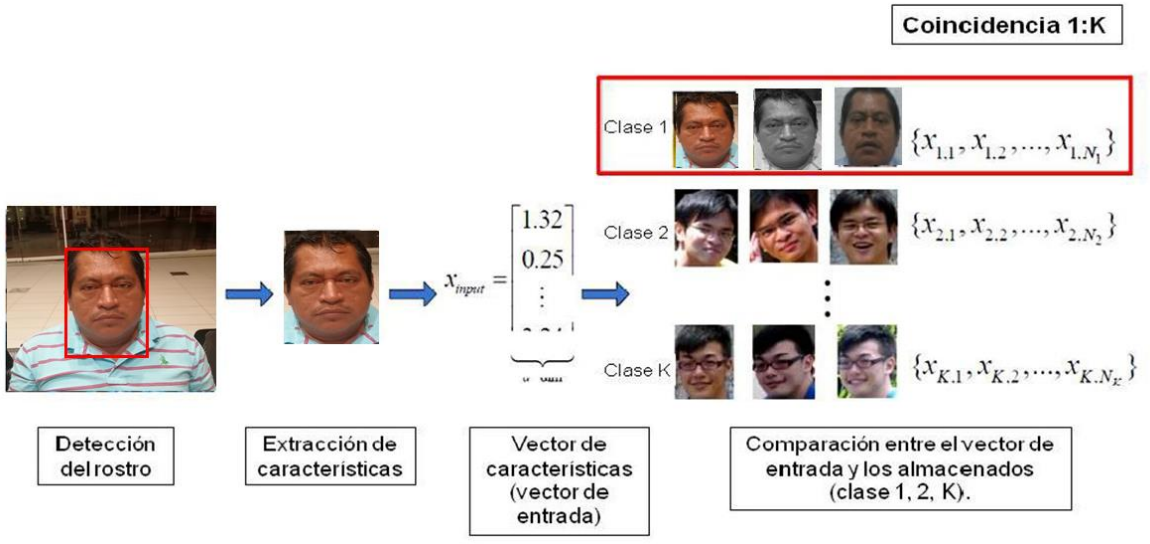

Figura 4.13 *Reconocimiento o Identificación facial.*

4.4.2.1 Algoritmo K-NN(K-Nearest Neighbor)

La clasificación en un sistema basado en reconocimiento facial parte de la base que rostros de un mismo individuo tienen características similares, por lo que se agrupan agrupan imágenes de individuos distintos en diferentes clases. Un sistema de clasificación debe ser capaz de asignar a cada nueva imagen de entrada una de estas clases, correspondiente al individuo que aparece en ella. De entre los distintos tipos de clasificadores existentes, se ha optado por KNN (vecino más cercano) [79] ha sido el método de clasificación utilizado en este proyecto. El principal motivo de nuestra elección es debido a que el algoritmo K-NN es un método de clasificación muy sencillo, pero a la vez muy potente.

El KNN es un sistema de clasificación automática y no supervisada. La idea principal de este clasificador es muy simple. Partiendo de los vectores de características de cada uno de los modelos realizados, se calcula la distancia del vector de características a testear a los vectores de cada uno de los modelos existentes.

La identificación final se obtiene a partir de las k distancias más pequeñas, que pueden corresponder todas al mismo modelo o a modelos diferentes. Así pues, una vez obtenidos los modelos cuya distancia es mínima a la imagen de entrada se procede a realizar una votación de manera que se asigna la imagen al modelo con el mayor número de votos.

Para explicar su funcionamiento se utiliza un ejemplo:

Suponiendo que se tiene un conjunto de entrenamiento de N vectores de características. Estos vectores están agrupados en distintas clases (C1, C2, C3,..). Suponiendo que también se introduce en el sistema un nuevo vector de características. El objetivo sería determinar la clase a la que pertenece. Para determinarla el algoritmo K-NN busca los K vecinos más

próximos al vector de entrada, y posteriormente realiza una valoración de las clases a las que pertenecen estos vectores vecinos. La valoración consiste en comparar las distancias existentes entre las clases. Los algoritmos aplicados en esta valoración pueden ser distintos, en el desarrollo del sistema se utilizó la Distancia Euclídea.

La ventaja de este sistema es la capacidad de modificación y creación de nuevas clases añadiendo nuevas muestras, así como la simplicidad de su sistema. Además, este sistema admite la incorporación de cualquier distancia métrica que se desee utilizar, añadiendo así cierta personalización.

Para realizar la comparación de los vectores característicos se optó por una de las soluciones propuestas en el artículo de *Google, FaceNet*[11], que consiste en hacer un cálculo de la distancia euclídea, por lo que el vector más próximo al vector del rostro que se pretende reconocer será el del rostro más similar (Figura 4.14).

En caso de no encontrarse ningún rostro similar, el rostro quedará marcado como desconocido. En este punto se encuentra uno de los parámetros que más se pueden optimizar, la distancia máxima a partir de la cual consideramos que un rostro es similar, el umbral. Se ha empleado para el sistema de este trabajo un umbral de 0.8, por lo que rostros con una distancia igual o inferior serán considerados similares.

[11] A Unified Embedding for Face Recognition and Clustering

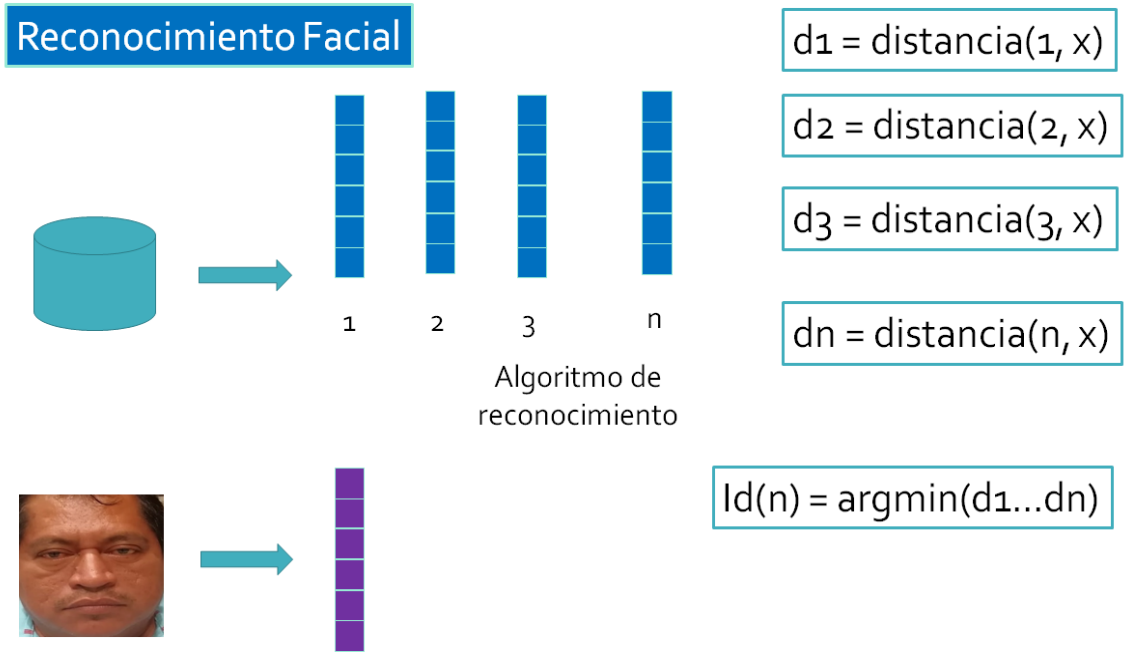

Figura 4.14 *Comparación de vectores característicos.*

Capítulo 5 Pruebas y Resultados

En este trabajo se presenta una arquitectura utilizada para diferenciar un individuo del resto que se tienen almacenados en la base de datos, por medio de una imagen facial. La arquitectura propuesta se basa en la utilización de dos técnicas que han demostrado ser eficientes cada una en su propósito, éstas se refieren a las redes neuronales convolucionales (CNN) y K-NN vecinos más cercanos con ayuda de la distancia Euclídea. Mientras que las CNN han sido muy utilizados como técnica para extraer y representar características obtenidas de imágenes digitales, los K-NN han demostrado ser una técnica de clasificación eficiente que proporciona resultados aceptables.

Las imágenes utilizadas fueron recopiladas con el consentimiento y colaboración de estudiantes del Instituto tecnológico de Acapulco (ITA). En la figura 5.1 se muestran algunas de ellas. Las imágenes se obtuvieron en ambientes controlados (interior) y ambientes no controlados (exterior). La resolución inicial de las mismas es de 1080 x 1920 píxeles. El tamaño inicial fue reducido posteriormente con fines de mejorar el proceso y una vez que se seleccionó de la imagen solamente el rostro a un tamaño de 96 x 96 píxeles. Las imágenes fueron capturadas utilizando el modelo de color estándar de 3 paletas (RGB), posteriormente y también con fines de optimizar el proceso se convirtieron a imágenes en escala de gris. Se almacenaron y utilizaron un total de 54 imágenes en ambientes controlados y 54 imágenes en ambientes no controlados.

Figura 5.1 *Muestra de algunas imágenes de rostros faciales con la participación de alumnos del ITA.*

Base de Datos de Rostros

La base de datos construida para las evaluaciones contiene los rostros de un total de 54 personas. Las imágenes fueron capturadas con un dispositivo móvil ZTE Blade V8 a una resolución de 1080 x 1920 píxeles. Esta base de datos fue diseñada para probar el funcionamiento específico del reconocedor de rostros diseñado para este proyecto. En vista de lo anterior, se tuvieron en cuenta ciertas condiciones para la adquisición de imágenes expuestas a continuación.

- Se asume que todos los rostros serán capturados en posición frontal y vertical sin rotaciones de ningún tipo.
- Las condiciones de iluminación pueden ser de tipos: controladas (luz artificial) y no controladas (luz natural). El total de muestras (fotografías) de los experimentos se tomaron de día con una luz clara del sol.
- Cada una de las imágenes contiene el rostro de solamente una persona.

El experimento realizado fue mediante pruebas para evaluar el funcionamiento del sistema, donde se pueden presentar cuatro situaciones posibles a la hora de realizar una verificación facial en función de cuál es la clase del usuario genuino o impostor.

VP es la cantidad de positivos que fueron clasificados correctamente como positivos

VN es la cantidad de negativos clasificados correctamente como negativos (No registrados en la BD y no identificados).

FN Es la cantidad de positivos que fueron clasificados incorrectamente como negativos

FP es la cantidad de negativos que fueron clasificados incorrectamente como positivos (No registrados en la BD, pero lo identifica como otra persona registrada)

Gracias a estas cuatro categorías podemos calcular métricas más elaboradas.

El objetivo de esta fase es evaluar el rendimiento que ofrece la aplicación. Los resultados obtenidos en estas pruebas se presentan a continuación:

5.1 Prueba 1: Reconocimiento bajo diferentes condiciones de iluminación

En el primer conjunto de pruebas, se midió la capacidad de reconocimiento de rostros bajo diferentes condiciones de iluminación. Para ésta se empleó la base de datos de rostros descrita con anterioridad. La población del estudio fue la estudiantil del Instituto Tecnológico mayor a 18 años. La prueba consistió en 270 comparaciones con 5 muestras diferentes de cada rostro, estos datos se muestran en las tablas 5.1 y 5.3. Estas fueron tomadas en dos ambientes diferentes de iluminación (tablas 5.2 y 5.4). Los resultados fueron los siguientes:

i) Ambiente controlado.

Tabla 5.1 *Datos de muestras totales tomados para la verificación.*

	Detección correcta	Detección Incorrecta	Totales
Personas	264	6	270
Totales	264	6	

Tabla 5.2 *Matriz de confusión de la clasificación Genuino e Impostor en ambiente controlado.*

Observación		**Detección**	
		Correcta	**Incorrecta**
	Genuino	261 (VP)	4 (FN)
	Impostor	2 (FP)	3 (VN)

Exactitud

Es el porcentaje total de los aciertos del modelo. Se estima como:

$$Exactitud = \frac{VP + VN}{Total} = \frac{261 + 3}{270} = 0.9777$$

Por lo tanto, se tiene una exactitud de:

Exactitud = 0.9777 X 100 = 97.7 %

Tasa de error

Es el porcentaje de errores del modelo. Se estima como:

$$Tasa\ de\ error = \frac{FP + FN}{Total} = \frac{2 + 4}{270} = 0.022$$

Por lo tanto, la Tasa de error es:

Tasa de error = 0.022 X 100 = 2.22 %

Sensibilidad

Es la probabilidad de que, dado que un individuo realmente está en la BD, la prueba lo detecte. Es decir, se estima esta probabilidad como:

$$Sensibilidad = \frac{VP}{Total\ Positivos(VP + FN)} = \frac{261}{261 + 4} = 0.9849$$

Por lo tanto, la Sensibilidad es:

Sensibilidad = 0.9849 X 100 = 98.49 %

Especificidad, tasa de verdaderos negativos.

Es la probabilidad de que, dado que un individuo no está realmente registrado en la BD. Se estima como:

$$Especificidad = \frac{VN}{Total\ Negativos(VN + FP)} = \frac{3}{3 + 2} = 0.6$$

Por lo tanto, la especificidad es:

Especificidad = 0.6 X 100 = 60 %

Precisión.

Es la probabilidad de que, dada una predicción positiva, la realidad sea positiva también. Se estima como:

$$Precisión = \frac{VP}{Total\ clasificados\ positivos(VP + FP)} = \frac{261}{261 + 2} = 0.992$$

Por lo tanto, la precisión es:

Precisión = 0.992 X 100 = 99.2 %

ii) Ambiente no controlado.

Tabla 5.3 *Datos de muestras totales tomados para la verificación.*

	Detección correcta	Detección Incorrecta	Totales
Personas	256	14	270
Totales	256	14	

Tabla 5.4 *Matriz de confusión de la clasificación Genuino e Impostor en ambiente no controlado.*

Observación		**Detección**	
		Correcta	**Incorrecta**
	Genuino	253 (VP)	10 (FN)
	Impostor	4 (FP)	3 (VN)

Exactitud

Es el porcentaje total de los aciertos del modelo. Se estima como:

$$Exactitud = \frac{VP + VN}{Total} = \frac{253 + 3}{270} = 0.948$$

Por lo tanto, se tiene una exactitud de:

Exactitud = 0.948 X 100 = 94.8 %

Tasa de error

Es el porcentaje de errores del modelo. Se estima como:

$$Tasa\ de\ error = \frac{FP + FN}{Total} = \frac{4 + 10}{270} = 0.0518$$

Por lo tanto, la Tasa de error es:

Tasa de error = 0.0518 X 100 = 5.18 %

Sensibilidad

Es la probabilidad de que, dado que un individuo realmente está en la BD, la prueba lo detecte. Es decir, se estima esta probabilidad como:

$$Sensibilidad = \frac{VP}{Total\ Positivos(VP + FN)} = \frac{253}{253 + 10} = 0.961$$

Por lo tanto, la Sensibilidad es:

Sensibilidad = 0.961 X 100 = 96.1 %

Especificidad, tasa de verdaderos negativos.

Es la probabilidad de que, dado que un individuo no está realmente registrado en la BD. Se estima como:

$$Especificidad = \frac{VN}{Total\ Negativos(VN + FP)} = \frac{3}{3 + 4} = 0.428$$

Por lo tanto, la especificidad es:

Especificidad = 0.428 X 100 = 42.8 %

Precisión.

Es la probabilidad de que, dada una predicción positiva, la realidad sea positiva también. Se estima como:

$$Precisión = \frac{VP}{Total\ clasificados\ positivos(VP + FP)} = \frac{253}{253 + 4} = 0.984$$

Por lo tanto, la precisión es:

Precisión = 0.984 X 100 = 98.4 %

5.2 Prueba 2: Reconocimiento mostrando alguna Emoción

En el segundo conjunto de pruebas, se midió la capacidad de reconocimiento de rostros con diferentes emociones. Para esta se empleó la base de datos con 26 rostros almacenados, donde se realizaron 78 comparaciones de 3 muestras de cada uno. Estas fueron tomadas en dos ambientes diferentes de iluminación. Los resultados fueron los siguientes:

i) Ambiente no controlado.

Tabla 5.5 *Matriz de confusión de la clasificación Genuino e Impostor mostrando una emoción en ambiente no controlado.*

Genuino/Impostor		**Emoción**		
		Feliz	**Triste**	**Enojado**
	VP	21	21	19
	VN	5	3	3
	FN	0	0	2
	FP	0	2	2

Exactitud

Es el porcentaje total de los aciertos del modelo. Se estima como:

$$Exactitud = \frac{VP + VN}{Total} = \frac{61 + 11}{78} = 0.923$$

Por lo tanto, se tiene una exactitud de:

Exactitud = 0.923 X 100 = 92.3 %

Tasa de error

Es el porcentaje de errores del modelo. Se estima como:

$$Tasa\ de\ error = \frac{FP + FN}{Total} = \frac{4 + 2}{78} = 0.0769$$

Por lo tanto, la Tasa de error es:

Tasa de error = 0.0769 X 100 = 7.69 %

Sensibilidad

Es la probabilidad de que, dado que un individuo realmente está en la BD, la prueba lo detecte. Es decir, se estima esta probabilidad como:

$$Sensibilidad = \frac{VP}{Total\ Positivos(VP + FN)} = \frac{61}{61 + 2} = 0.968$$

Por lo tanto, la Sensibilidad es:

Sensibilidad = 0.968 X 100 = 96.8 %

Especificidad, tasa de verdaderos negativos.

Es la probabilidad de que, dado que un individuo no está realmente registrado en la BD. Se estima como:

$$Especificidad = \frac{VN}{Total\ Negativos(VN + FP)} = \frac{11}{11 + 4} = 0.733$$

Por lo tanto, la especificidad es:

Especificidad = 0.733 X 100 = 73.3 %

Precisión.

Es la probabilidad de que, dada una predicción positiva, la realidad sea positiva también. Se estima como:

$$Precisión = \frac{VP}{Total\ clasificados\ positivos(VP + FP)} = \frac{61}{61 + 4} = 0.9384$$

Por lo tanto, la precisión es:

Precisión = 0.9384 X 100 = 93.84 %

ii) Ambiente controlado.

Tabla 5.6 *Matriz de confusión de la clasificación Genuino e Impostor mostrando una emoción en ambiente controlado.*

Genuino/Impostor		Emoción		
		Feliz	**Triste**	**Enojado**
	VP	21	21	20
	VN	4	3	4
	FN	0	0	1
	FP	1	2	1

Exactitud

Es el porcentaje total de los aciertos del modelo. Se estima como:

$$Exactitud = \frac{VP + VN}{Total} = \frac{62 + 11}{78} = 0.935$$

Por lo tanto, se tiene una exactitud de:

Exactitud = 0.935 X 100 = 93.5 %

Tasa de error

Es el porcentaje de errores del modelo. Se estima como:

$$Tasa\ de\ error = \frac{FP + FN}{Total} = \frac{4 + 1}{78} = 0.0641$$

Por lo tanto, la Tasa de error es:

Tasa de error = 0.0641 X 100 = 6.41 %

Sensibilidad

Es la probabilidad de que, dado que un individuo realmente está en la BD, la prueba lo detecte. Es decir, se estima esta probabilidad como:

$$Sensibilidad = \frac{VP}{Total\ Positivos(VP + FN)} = \frac{62}{62 + 1} = 0.984$$

Por lo tanto, la Sensibilidad es:

Sensibilidad = 0.984 X 100 = 98.4 %

Especificidad, tasa de verdaderos negativos.

Es la probabilidad de que, dado que un individuo no está realmente registrado en la BD. Se estima como:

$$Especificidad = \frac{VN}{Total\ Negativos(VN + FP)} = \frac{11}{11 + 4} = 0.733$$

Por lo tanto, la especificidad es:

Especificidad = 0.733 X 100 = 73.3 %

Precisión.

Es la probabilidad de que, dada una predicción positiva, la realidad sea positiva también. Se estima como:

$$Precisión = \frac{VP}{Total\ clasificados\ positivos(VP + FP)} = \frac{62}{62 + 4} = 0.9393$$

Por lo tanto, la precisión es:

Precisión = 0.9393 X 100 = 93.93 %

5.3 Prueba 3: Reconocimiento a diferentes distancias

En el tercer conjunto de pruebas, se midió la capacidad de reconocimiento de rostros a diferentes distancias. Para esta se empleó la misma base de datos de rostros con 26 rostros almacenados. Los resultados fueron los siguientes:

***i*) Ambiente no controlado.**

Tabla 5.7 *Matriz de confusión de la clasificación Genuino e Impostor a diferentes distancias en ambiente no controlado.*

Genuino/Impostor		Distancia			
		0.5 m	**1.0 m**	**1.5 m**	**2.0 m**
	VP	21	21	20	21
	VN	5	4	2	3
	FN	0	0	1	0
	FP	0	1	3	2

Exactitud

Es el porcentaje total de los aciertos del modelo. Se estima como:

$$Exactitud = \frac{VP + VN}{Total} = \frac{83 + 14}{104} = 0.9326$$

Por lo tanto, se tiene una exactitud de:

Exactitud = 0.9326 X 100 = 93.26 %

Tasa de error

Es el porcentaje de errores del modelo. Se estima como:

$$Tasa\ de\ error = \frac{FP + FN}{Total} = \frac{6 + 1}{104} = 0.0674$$

Por lo tanto, la Tasa de error es:

Tasa de error = 0.0674 X 100 = 6.74 %

Sensibilidad

Es la probabilidad de que, dado que un individuo realmente está en la BD, la prueba lo detecte. Es decir, se estima esta probabilidad como:

$$Sensibilidad = \frac{VP}{Total\ Positivos(VP + FN)} = \frac{83}{83 + 1} = 0.988$$

Por lo tanto, la Sensibilidad es:

Sensibilidad = 0.988 X 100 = 98.8 %

Especificidad, tasa de verdaderos negativos.

Es la probabilidad de que, dado que un individuo no está realmente registrado en la BD. Se estima como:

$$Especificidad = \frac{VN}{Total\ Negativos(VN + FP)} = \frac{14}{14 + 6} = 0.7$$

Por lo tanto, la especificidad es:

Especificidad = 0.7 X 100 = 70 %

Precisión.

Es la probabilidad de que, dada una predicción positiva, la realidad sea positiva también. Se estima como:

$$Precisión = \frac{VP}{Total\ clasificados\ positivos(VP + FP)} = \frac{83}{83 + 6} = 0.9325$$

Por lo tanto, la precisión es:

Precisión = 0.9325 X 100 = 93.25 %

ii) Ambiente controlado.

Tabla 5.8 *Matriz de confusión de la clasificación Genuino e Impostor a diferentes distancias en ambiente controlado.*

Genuino/Impostor		Distancia			
		0.5 m	**1.0 m**	**1.5 m**	**2.0 m**
	VP	21	21	21	20
	VN	5	4	1	3
	FN	0	0	0	1
	FP	0	1	4	2

Exactitud

Es el porcentaje total de los aciertos del modelo. Se estima como:

$$Exactitud = \frac{VP + VN}{Total} = \frac{83 + 13}{104} = 0.923$$

Por lo tanto, se tiene una exactitud de:

Exactitud = 0.923 X 100 = 92.3 %

Tasa de error

Es el porcentaje de errores del modelo. Se estima como:

$$Tasa\ de\ error = \frac{FP + FN}{Total} = \frac{7 + 1}{104} = 0.077$$

Por lo tanto, la Tasa de error es:

Tasa de error = 0.077 X 100 = 7.7 %

Sensibilidad

Es la probabilidad de que, dado que un individuo realmente está en la BD, la prueba lo detecte. Es decir, se estima esta probabilidad como:

$$Sensibilidad = \frac{VP}{Total\ Positivos(VP + FN)} = \frac{83}{83 + 1} = 0.988$$

Por lo tanto, la Sensibilidad es:

Sensibilidad = 0.988 X 100 = 98.8 %

Especificidad, tasa de verdaderos negativos.

Es la probabilidad de que, dado que un individuo no está realmente registrado en la BD. Se estima como:

$$Especificidad = \frac{VN}{Total\ Negativos(VN + FP)} = \frac{13}{13 + 7} = 0.65$$

Por lo tanto, la especificidad es:

Especificidad = 0.65 X 100 = 65 %

Precisión

Es la probabilidad de que, dada una predicción positiva, la realidad sea positiva también. Se estima como:

$$Precisión = \frac{VP}{Total\ clasificados\ positivos(VP + FP)} = \frac{83}{83 + 7} = 0.9222$$

Por lo tanto, la precisión es:

Precisión = 0.9222 X 100 = 92.2 %

5.4 Comparativa de los resultados de las pruebas realizadas

Tabla 5.9 *Tabla comparativa de las pruebas en ambiente controlado.*

Pruebas	**Exactitud**	**Tasa de error**	**Sensibilidad**	**Especificidad**	**Precisión**
Rostro normal	97.78 %	2.22%	98.49%	60 %	99.2 %
Emoción	93.5 %	6.41	98.4	73.3 %	93.93 %
Distancia	92.3 %	7.7 %	98.8 %	65 %	92.22 %

Tabla 5.10 *Tabla comparativa de las pruebas en ambiente no controlado.*

Pruebas	**Exactitud**	**Tasa de error**	**Sensibilidad**	**Especificidad**	**Precisión**
Rostro normal	94.8 %	5.2 %	96.1 %	42.8%	98.4 %
Emoción	92.3 %	7.7 %	96.8 %	73.3%	93.84 %
Distancias	93.26 %	6.74 %	98.8 %	70 %	93.25 %

Se puede observar que los resultados obtenidos utilizando las variantes propuestas son mejores en las obtenidas con las pruebas de reconocimiento de rostros en modo neutral bajo el ambiente controlado dado que se obtuvieron los porcentajes más altos en los valores de *exactitud* y *precisión*.

Equipo utilizado. La implementación y uso del sistema diseñado para la identificación automática a través del análisis facial requiere lo siguiente:

1) Una computadora personal portátil con las siguientes características mínimas:

    a) Disco duro de 1 Tb

    b) 12 Gb de memoria en RAM

    c) Procesador CORE I7

2) Un télefono móvil ZTE con las siguientes características:

- Pantalla: IPS de 5 pulgadas con resolución HD
- Procesador: MT6750 de 8 núcleos hasta 1.5 GHz
- RAM: 2 GB
- Memoria interna: 16 GB
- Cámara: 13 Mpx AF / 8 Mpx
- Versión de Android: 7.0 Nougat
- Batería: 2.500 mAh

3) Software:

    a) Windows 10.

    b) Lenguaje C++

    c) Android Studio Ver. 3.12

Capítulo 6 Conclusiones y Trabajo Futuro

Con este proyecto se ha logrado cumplir el objetivo impuesto al principio del desarrollo: a partir de una imagen, reconocer un rostro y saber si pertenece a alguna de las personas almacenadas en nuestra base de datos, además de mostrar el rostro almacenado con mayor coincidencia con el primero en una interfaz sencilla en el dispositivo móvil.

6.1 Conclusión

La motivación general que ha guiado esta tesis ha sido contribuir al estudio y al desarrollo de técnicas que permitan realizar de forma adecuada el proceso de verificación facial tomando en cuenta algunas variaciones como son la iluminación, las expresiones faciales y la oclusión parcial del rostro. En este trabajo se han estudiado las técnicas basadas en imágenes bidimensionales como son las redes neuronales convolucionales, además que se ha utilizado el clasificador: K-NN vecinos más cercanos utilizando la distancia Euclidiana.

El presente trabajo ha demostrado la efectividad de un servicio de verificación facial como parte de un sistema móvil online. La aplicación permite detectar y reconocer rostros en imágenes. No obstante, si no se trata de imágenes con caras frontales, su rendimiento no está totalmente optimizado y el rendimiento obtenido puede ser inferior.

Finalmente, se han evaluado los resultados de la herramienta implementada en una fase de pruebas. Se analizaron los experimentos realizados al sistema en dos ambientes de iluminación, el cual se observaron mejores resultados, en el reconocimiento de rostros en modo normal. En este caso, los resultados de *exactitud* y *precisión* tienen un resultado superior a las demás pruebas.

A nivel personal y como conclusión puedo decir que, en general, la experiencia obtenida tras el desarrollo de este proyecto ha sido muy positiva y enriquecedora. He aprendido a trabajar de manera autónoma en un área que, a pesar de no estar estrictamente relacionada, me despertaba bastante interés.

También he aprendido a utilizar nuevas herramientas y algoritmos. No solo he obtenido conocimientos a nivel práctico, también a nivel teórico gracias a la fase previa de documentación y búsqueda de información, así como los pasos para desarrollar un proyecto de estas características.

6.2 Trabajo futuro

El trabajo futuro podría centrarse inicialmente en abordar las limitaciones actuales. La aplicación móvil se puede implementar en dispositivos portátiles en el futuro para que el sistema sea más práctico en términos de peso y facilidad de uso. Los dispositivos como los *smartglasses* (gafas inteligentes) se pueden usar para adquirir y procesar las imágenes. Se pueden realizar otras optimizaciones a la aplicación y algoritmos para reducir el consumo de energía.

Para nuevas aplicaciones, los algoritmos y métodos utilizados en el sistema móvil pueden adaptarse a otras situaciones. La detección y el reconocimiento de objetos mediante redes neuronales convolucionales se pueden aplicar a otros escenarios, como los sistemas de seguridad y la navegación para personas con discapacidades visuales. Los sistemas de seguridad con cámaras de vigilancia pueden usar la detección de objetos para localizar intrusos y alertar a los propietarios mediante notificaciones en sus dispositivos móviles. Los sistemas de navegación para personas con discapacidades visuales implementados a través

de dispositivos móviles pueden guiar a los usuarios a sus destinos a la vez que los ayudan a evitar obstáculos y otros peligros.

También como trabajo futuro tenemos la elaboración de una base de datos de imágenes más amplia con diferentes condiciones de iluminación y algunas posiciones del rostro (ejemplo frente al dispositivo de captura y con algún ángulo de rotación de la cabeza). Estas bases de datos de imágenes servirán para ver y valorar la repetibilidad del algoritmo y si tiene alguna variación con los porcentajes de autenticación y reconocimiento facial.

Bibliografía

[**1**] Paul Reid. BIOMETRICS for Networks Security. Prentice Hall, 2004.

[**2**] Amos, B., Ludwiczuk, B., & Satyanarayanan, M. (2016). OpenFace: A general-purpose face recognition library with mobile applications. 1-18.

[**3**] Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). DeepFace: Closing the Gap to Human-Level Performance in Face Verification. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1701-1708.

[**4**] Gargi M, J. Jasmin Sylvia Rani, Madhu Ramiah, N. T. Naresh Babu, A. Annis Fathima and V. Vaidehi. "Mobile Authentication Using Iris Biometrics". Published by Springer Berlin Heidelberg, Networked Digital Technologies, Vol. 294. pp 332-341, 2012.

[**5**] De Santos Sierra. A, C. Sanchez Avila, A. MendazaOrmaza, J. Guerra Casanova. Towards Hand Biometrics in Mobile devices. In Proceeding of BIOSIG, Darmstadt, ISBN: 978-3-88579-285-7, 2011.

[**6**] Dave G, Chao, X., & Sriadibhatla, K. "Face Recognition in Mobile Phones". Department of Electrical Engineering Stanford University, USA, 2010.

[**7**] Vazquez-Fernandez, Esteban, y otros. "Built-in face recognition for smart photo sharing in mobile devices." Multimedia and Expo (ICME), 2011 IEEE International Conference on. IEEE.

[**8**] Red Académica y de Investigación Española - RedIRIS, "*Autenticación de Usuarios",* fecha de acceso: 05/07/2018, http://www.rediris.es/cert/doc/unixsec/node14.html

[**9**] F. Dinei, H. Cormac y C. Baris, *"Do Strong Passwords Accomplish Anything?"* en Actas del HOTSEC, Junio, 2007, fecha de acceso: 06/03/2018, http://static.usenix.org/event/hotsec07/tech/full_papers/florencio/florencio.pdf

[**10**] SANS Institute InfoSec Reading Room, *"Two-Factor Authentication: Can You Choose the Right One?",* fecha de acceso: 20/01/2017, http://www.sans.org/reading_room/whitepapers/authentication/two-factorauthentication-choose-one_33093 .

[**11**] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 815-823.

[**12**] LeCun, Y., Boser, B., Denker, J. S., Howard, R. E., Hubbard, W., & Jacket, L. D. (1989). Backpropagation Applied to Handwritten Zip Code Recognition. Neural Computation, 1(4), 541-551.

[**13**] Karpathy, A., Johnson, J. & Fei-Fei, L (2016). CS231n Convolutional Neural Networks for Visual Recognition. Stanford University. Extraído desde: http://cs231n.stanford.edu/

[**14**] Anil K. Jain and Arun Ross. Handbook of Biometrics.(2008). e-ISBN-13: 978-0-387-71041-9, 2008.

[**15**] C. Villegas Quezada, "Reconocimiento de rostos utilizando análisis de componentes principales", Universidad Iberoamericana, México D.F., 2005.

[**16**] P. Álvarez Corrales, "Prototipo de sistema piloto para control de acceso basado en el reconocimiento de rostros", Universidad Militar Nueva Granada, Granada, España, 2013.

[**17**] V. Bruce y A. W. Young, "Understanding face recognition", British Journal of Psychology, 1986, 77, 305--327., vol. 77, nº 1, pp. 305-307, 1986.

[**18**] L. Sirovik y M. Kirby, "Low-Dimensional Procedure for the Characterization of Human Face",Journal of the Optical Society of America, vol. 4, nº 1, pp. 519-524, 1987.

[**19**] P. S. Penev y J. J. Atick, "Local Feature Analysis: A general statistical theory for object representation", Computational Neuroscience Laboratory, Rockefeller University", 1996 .

[**20**] G. Yang y T. S. Huang, "Human face detection in a Scene", de IEEE Computer Society Conference on Comuter Vision and Pattern Recognition., N.Y., 1993.

[**21**] J. M. Yang, D. Kriegman y N. Ahuja, "Detecting faces in images: a survey", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, nº 1, pp. 34-58, 2002.

[**22**] Hjelmås, E. and Low, B. K. (2001) 'Face Detection: A Survey', Computer Vision and Image Understanding, 83 (3), pp. 236-274.

[**23**] Ryu, H., Chun, S. S. and Sull, S. (2006) 'Multiple classifiers approach for computational efficiency in multi-scale search based face detection', Advances in Natural Computation, Pt 1, 4221 pp. 483-492.

[**24**] Rowley, H. A., Baluja, S. and Kanade, T. (1998) 'Neural network-based face detection', Pattern Analysis and Machine Intelligence, IEEE Transactions on, 20 (1), pp. 23-38.

[**25**] Feraud, R., Bernier, O., Viallet, J. E. and Collobert, M. (2001) 'A Fast and Accurate Face Detector Based on Neural Netwok', IEEE Transactions on Pattern Analysis and Machine Intelligense, Vol. 23, No. 1.

[**26**] Dong Wang, Jing Yang, J., Deng, and Q., Liu. (2016) “FaceHunter: A multi-task convolutional neural network based face detector”. Signal Processing: Image Communication. Vol. 47. PP. 476-481.

[**27**] Mohamed, A. S. S., Ying Weng, S. S., Ipson, S. S. and Jianmin Jiang, S. S. (2007) 'Face detection based on skin color in image by neural networks', Intelligent and Advanced Systems, 2007.ICIAS 2007.International Conference on, pp. 779-783.

[**28**] Surayahani, S. and Masnani, M. (2010) 'Modeling Understanding Level Based Student Face Classification' Mathematical/Analytical Modelling and Computer Simulation (AMS), 2010 Fourth Asia International Conference on, pp. 271-275. 10.1109/AMS.2010.61.

[**29**] Modi, M. and Macwan , F. (2014) 'Face Detection Approaches: A Survey.', International Journal of Innovative Research in Science, Engineering and Technology, 3 (4), pp. 11107-11116.

[**30**] M. Nehru and S., Padmavathi. “Illumination invariant face detection using Viola Jones algorithm”. 2017 4th IEEE International Conference on Advanced Computing and Communication System.

[**31**] Mayank Chauhan , Mukesh Sakle. (2014) 'Study & Analysis of Different Face Detection Techniques', (IJCSIT) International Journal of Computer Science and Information Technologies, 5 (2), pp. 1615-1618.

[**32**] Viola, P. and Jones, M. (2001) 'Rapid object detection using a boosted cascade of simple features', Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1 pp. I511-I518.

[**33**] Wang, K., Song, Z., Sheng, M., He, P. and Tang, Z. (2015) 'Modular Real-Time Face Detection System', Annals of Data Science, 2 (3), pp. 317-333.

[**34**] Thai, L. H., Nguyen, N. D. T. and Hai, T. S. (2011) 'A Facial Expression Classification System Integrating Canny, Principal Component Analysis and Artificial Neural Network', International Journal of Machine Learning and Computing, 1 (4), pp. 388-393.

[**35**] Face Recognition with OpenCV | OpenCV 2.4.13.2 documentation.

[**36**] Sachin Sudhakar Farfade, Mohammad J. Saberian, and Li-Jia Li. Multi-view face detection using deep convolutional neural networks. In Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, pages 643-650. ACM, 2015.

[**37**] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen. Face description with local binary patterns: Application to face recognition. IEEE transactions on pattern analysis and machine intelligence, 28(12):2037-2041, 2006.

[**38**] Biao Wang, Weifeng Li, Wenming Yang, and Qingmin Liao. Illumination normalization based on weber's law with application to face recognition. IEEE Signal Processing Letters, 18(8):462-465, 2011.

[**39**] John Wright, Allen Y. Yang, Arvind Ganesh, S. Shankar Sastry, and Yi Ma. Robust face recognition via sparse representation. IEEE transactions on pattern analysis and machine intelligence, 31(2):210-227, 2009.

[**40**] K. P. Tripathi. A comparative study of biometric technologies with reference to human interface. International Journal of Computer Applications, 14(5):10-15, 2011.

[**41**] Anil Jain, Patrick Flynn, and Arun A. Ross. Handbook of biometrics. Springer Science & Business Media, 2007.

[**42**] Rupinder Saini and Narinder Rana. Comparison of various biometric methods. International Journal of Advances in Science and Technology (IJAST), 2(1):2, 2014.

[**43**] R. Vera-Rodriguez, J. Fierrez, P. Tome, and J. Ortega-Garcia. Face Recognition at a Distance: Scenario Analysis and Applications. In Andre Ponce de Leon F. de Carvalho, Sara Rodríguez-González, Juan F. De Paz Santana, and Juan M. Corchado Rodríguez, editors, Distributed Computing and Artificial Intelligence, number 79 in Advances in Intelligent and Soft Computing, pages 341{348. Springer Berlin Heidelberg, 2010. DOI: 10.1007/978-3-642-14883-5_44.

[**44**] Paul Viola and Michael J. Jones. Robust real-time face detection. International journal of computer vision, 57(2):137{154, 2004.

[**45**] Hitoshi Imaoka. Face Recognition: Beyond the Limit of Accuracy. In International Joint Conference on Biometrics. International Joint Conference on Biometrics, 2014.

[**46**] Matthew Turk and Alex Pentland. Eigenfaces for recognition. Journal of cognitive neuroscience, 3(1):71-86, 1991.

[**47**] Peter N. Belhumeur, Joao P. Hespanha, and David J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on pattern analysis and machine intelligence, 19(7):711-720, 1997.

[**48**] Jason V. Davis, Brian Kulis, Prateek Jain, Suvrit Sra, and Inderjit S. Dhillon. Information theoretic metric learning. In Proceedings of the 24th international conference on Machine learning, pages 209-216. ACM, 2007.

[**49**] Corinna Cortes and Vladimir Vapnik. Support-vector networks. Machine learning, 20(3):273-297, 1995.

[**50**] Takeo Kanade. Picture processing system by computer complex and recognition of human faces. Doctoral dissertation, Kyoto University, 3952:83–97, 1973.

[**51**] Tony S Jebara. 3D pose estimation and normalization for face recognition. PhD thesis, McGill University, 1995.

[**52**] Rabia Jafri and Hamid R Arabnia. A survey of face recognition techniques. JIPS, 5(2):41–68, 2009

[**53**] Harold Hotelling. Analysis of a complex of statistical variables into principal components. Journal of educational psychology, 24(6):417, 1933.

[**54**] Lawrence Sirovich and Michael Kirby. Low-dimensional procedure for the characterization of human faces. JOSA A, 4(3):519–524, 1987.

[**55**] Matthew Turk and Alex Pentland. Eigenfaces for recognition. Journal of cognitive neuroscience, 3(1):71–86, 1991.

[**56**] Peter N Belhumeur, Jo˜ao P Hespanha, and David J Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 19(7):711–720, 1997.

[**57**] Steve Lawrence, C Lee Giles, Ah Chung Tsoi, and Andrew D Back. Face recognition: A convolutional neural-network approach. Neural Networks, IEEE Transactions on, 8(1):98–113, 1997.

[**58**] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, pages 1701–1708. IEEE, 2014.

[**59**] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 815–823, 2015.

[**60**] Omkar M Parkhi, Andrea Vedaldi, and Andrew Zisserman. Deep face recognition. Proceedings of the British Machine Vision, 1(3):6, 2015.

[**61**] Xiang Wu, Ran He, and Zhenan Sun. A lightened cnn for deep face representation. arXiv preprint arXiv:1511.02683, 2015.

[**62**] The U.S. mobile app report | Tech. Rep., comScore (August 2014)

[**63**] Yoshua Bengio, Ian J. Goodfellow, and Aaron Courville. Deep learning. Book in preparation for MIT Press, 2015.

[**64**] A. Pascual | 2014 Identity Fraud Report: card data breaches and inadequate consumer password habits fuel disturbing fraud trends | Tech. Rep., Javelin (February 2014)

[**65**] https://www.rsipvision.com/exploring-deep-learning/

[**66**] Brandon Amos, Bartosz Ludwiczuk, and Mahadev Satyanarayanan. OpenFace: A general-purpose face recognition library with mobile applications, School of Computer Science Carnegie Mellon University Pittsburg, PA 15213, 2016.

[**67**] B. Institute | Biometrics Institute Industry Survey 2017 |Tech. Rep., Biometrics Institute (July 2017). https://www.biometricupdate.com/201707/biometrics-institute-releases-results-from-2017-industry-survey

[**68**] A.K. Jain, K. Nandakumar, A. Ross | 50 years of biometric research: accomplishments, challenges, and opportunities | Pattern Recogn. Lett. (2016)

[**69**] Yaniv Taigman, Lior Wolf, Tal Hassner, and others. Multiple One-Shots for Utilizing Class Label Information. In BMVC, volume 2, pages 1-12, 2009.

[**70**] Chang Huang, Shenghuo Zhu, and Kai Yu. Large scale strongly supervised ensemble metric learning, with applications to face veri_cation and retrieval. arXiv preprint arXiv:1212.6094, 2012.

[**71**] Yi Sun, Yuheng Chen, XiaogangWang, and Xiaoou Tang. Deep learning face representation by joint identification-verification. In Advances in neural information processing systems, pages 1988-1996, 2014.

[**72**] Yaniv Taigman, Ming Yang, Marc_ Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1701{1708, 2014.

[**73**] Openaccess.thecvf.com, 2018. [Online]. Available: http://openaccess.thecvf.com/content_cvpr_2014/papers/Kazemi_One_Millisecond_Face_2014_CVPR_paper.pdf. [Accessed: Jun - 2018].

[**74**] V. Kazemi, "Face Alignment", Csc.kth.se, 2018. [Online]. Available: http://www.csc.kth.se/~vahidk/face_ert.html . [Accessed: 04- Jun- 2018].

[**75**] Cv-foundation.org, 2018. [Online]. Available: https://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Schroff_FaceNet_A_Unified_2015_CVPR_paper.pdf . [Accessed: Jun - 2018].

[**76**] D. King, "High Quality Face Recognition with Deep Metric Learning", Blog.dlib.net, 2018. [Online]. Available: http://blog.dlib.net/2017/02/high-quality-face-recognition-with-deep.html . [Accessed: 05- Jun- 2018].

[**77**] "Understanding and Implementing Architectures of ResNet and ResNeXt for stateof-the-art Image…", Medium, 2018. [Online]. Available: https://medium.com/@14prakash/understanding-and-implementingarchitectures-of-resnet-and-resnext-for-state-of-the-art-image-cf51669e1624 . [Accessed: Jun - 2018].

[**78**] Openaccess.thecvf.com, 2018. [Online]. Available: http://openaccess.thecvf.com/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf . [Accessed: Jun - 2018].

[**79**] Belur V. Dasarathy, “Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques”, 1991

[**80**] Goodfellow, Ian, Bengio, Yoshua and Courville, Aaron. (2016, December). Deep Learning. MIT Press.