



**Benemérita Universidad Autónoma de Puebla**

**Facultad de Ciencias de la Computación**

---

**Método de extracción de rasgos faciales para la autenticación de personas**

---

**Tesis presentada para obtener el grado de:  
Doctor en Ingeniería del Lenguaje y del Conocimiento**

**Presenta:  
Aida Anai Aparicio Arroyo**

**Director de Tesis:  
Dr. Ivan Olmos Pineda**

**Mayo 2022**

*En memoria de todas esas personas (familia, amigos y conocidos)  
que perdieron la batalla durante y por la pandemia.*

# Agradecimientos

En primer lugar, quiero agradecer a Dios por darme la oportunidad de haber llegado hasta este punto de mi vida y poder culminar una etapa más.

En segundo lugar, le agradezco a mi mamá, que me ha apoyado en cada momento y me ha alentado a seguir adelante.

Agradezco a mis asesores Dr. Ivan Olmos Pineda y Dr. José Arturo Olvera López, por la paciencia, las observaciones y consejos, los conocimientos compartidos y las largas horas de trabajo a lo largo de todo el doctorado.

De igual manera, quisiera agradecer a mi comité doctoral: Dr. Antonio Benitez Ruiz, Dr. Aldrin Barreto Flores y Dr. Juan Manuel González Calleros, por sus comentarios y consejos, que gracias a ellos se pudo mejorar el trabajo de investigación.

De igual forma, quiero agradecer al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico brindado a lo largo de los estudios doctorales, a través de la beca no. 481147.

Igualmente, quiero agradecer a mis amistades, que siempre me apoyaron en todo momento y estuvieron ayudándome directa e indirectamente.

Por último, pero no menos importante, quiero dar gracias a esas personas, que, con el transcurso de los días, se convirtieron en mis amigos.

*"On ne voit bien qu'avec le cœur. L'essentiel est invisible pour les yeux"*  
*Antoine de Saint-Exupéry.*

# Resumen

Dentro de la visión por computadora, existen diferentes áreas de investigación, como: identificación, reconocimiento y autenticación. Como aclaración, estas áreas son totalmente diferentes. Actualmente, la autenticación de personas se ha convertido en una forma de mantener la seguridad en diferentes ámbitos tanto en el sector público como en el privado, en la telefonía móvil, entre otros. Existen diversas maneras de autenticar personas, como utilizando su voz, su huella dactilar, rasgos faciales o incluso su iris.

En los últimos años, se han desarrollado un sin fin de algoritmos de seguridad, que son capaces de autenticar personas a través de sus rasgos faciales, pero estos algoritmos o sistemas de seguridad, tienen un buen rendimiento bajo ciertas circunstancias, como que la persona esté de frente a la cámara de vigilancia, que haya una buena iluminación o simplemente, que se cuente con una buena cámara.

Tomando en cuenta las especificaciones de los sistemas existentes y sus pequeñas ineficiencias, en este trabajo de tesis se presenta el diseño de un método de extracción de rasgos faciales, el cual, por sus características y precisión, servirá para el proceso de autenticación de personas. Este método fue diseñado, tratando de abordar las características de los algoritmos existentes y que, además, como aportación, abarque la opción de que la persona sea autenticada no solo cuando esté de frente a la cámara, sino que también se pueda autenticar estando de perfil.

Antes de diseñar el método de extracción de rasgos faciales, fue necesario realizar una revisión al estado del arte, con el fin de buscar las técnicas más relevantes en el proceso de la autenticación de personas, para que, a partir de ahí se pudiera proponer un método que fuera preciso y eficaz al mismo tiempo.

De manera que se fue desarrollando el trabajo de tesis, se efectuaron algunas modificaciones al método de extracción de rasgos faciales, ya que en primeros experimentos era muy ineficiente y, por ende, no podía ser comparado con trabajos relacionados. Después de haber realizado un análisis y algunas correcciones al algoritmo, se obtuvo como resultado un método que tuvo un buen rendimiento y pudo compararse con el estado del arte.

Hechos los ajustes necesarios al método, los resultados de los experimentos realizados en este

trabajo de investigación son superiores al 96 %, este porcentaje de precisión es comparativo con trabajos relacionados, pero un punto a tratar es el tiempo de ejecución, ya que está por arriba de los reportados en los trabajos relacionados y como trabajo a futuro, es reducir el tiempo de respuesta, para que el método propuesto sea comparado con el estado del arte.

# Abstract

Within computer vision, there are different areas of research, such as: identification, recognition and authentication. As a clarification, these areas are totally different. Currently, the people authentication has become a way of maintaining security in different limits both in the public and private sectors, in mobile telephony, among others. There are several ways to authenticate people, such as using their voice, fingerprint, facial features, or even their iris. In recent years, endless security algorithms have been developed that are capable of authenticating people through their facial features, but these algorithms or security systems perform well under certain circumstances, such as when the person is facing the surveillance camera, that there is good lighting or simply that there is a good camera.

Taking into account the specifications of the existing systems and their small inefficiencies, this thesis work presents the design of a facial features extraction method, which, due to its characteristics and precision, will serve for the person authentication process.

This method was designed trying to address the characteristics of existing algorithms and also, as a contribution, encompasses the option for the person to be authenticated not only when facing the camera, but also when they are in profile.

Before designing the facial features extraction method, it was necessary to review the state of the art, in order to find the most relevant techniques in the people authentication process, so that, from there, a new method could be proposed. Method that was precise and efficient at the same time.

As the thesis work was developed, some modifications were made to the method of extracting facial features, since in the first experiments it was very inefficient and therefore, it could not be compared with related works. After having carried out an analysis and some corrections to the algorithm, a method was obtained that had a good performance and could be compared with the state of the art.

After making the necessary adjustments to the method, the results of the experiments carried out in this research work are greater than 96 %, this precision percentage is comparative with related works, but a point to be discussed is the execution time, since it is above those reported

in related works and as future work, it is to reduce the response time, so that the proposed method is compared with the state of the art.

# Índice general

<b>Índice de figuras</b>	<b>xI</b>
<b>Índice de tablas</b>	<b>xII</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Problema de investigación .....	1
1.2. Justificación .....	2
1.3. Objetivos .....	3
1.3.1. Objetivo general .....	3
1.3.2. Objetivos específicos .....	3
1.4. Preguntas de investigación .....	4
1.5. Hipótesis .....	4
1.6. Alcances y límites .....	4
1.7. Organización del trabajo .....	5
<b>2. Marco Teórico</b>	<b>6</b>
2.1. Procesamiento digital de imágenes .....	6
2.2. Aplicaciones del PDI .....	9
2.3. Procesamiento digital de imágenes faciales .....	10
2.3.1. Extracción de rasgos faciales .....	11
2.3.2. Clasificadores .....	20
2.4. Métricas de evaluación .....	22
2.4.1. Validación cruzada .....	23
<b>3. Estado del Arte</b>	<b>25</b>
3.1. Detección de rostros .....	26
3.2. Preprocesamiento .....	27
3.3. Técnicas de extracción de rasgos faciales .....	29

3.4. Clasificadores .....	31
3.5. Discusión .....	32
<b>4. Propuesta Metodológica</b>	<b>34</b>
4.1. Etapa de entrenamiento del modelo.....	37
4.1.1. Detección de rostros.....	37
4.1.2. Detección de orientación.....	38
4.1.3. Etapa de preprocesamiento .....	40
4.1.4. Etapa de extracción de rasgos faciales .....	41
4.1.5. Etapa de entrenamiento del clasificador .....	43
4.2. Etapa de prueba del modelo entrenado .....	43
<b>5. Resultados de la experimentación</b>	<b>44</b>
5.1. Base de imágenes.....	44
5.1.1. Base de datos generada .....	44
5.1.2. Base de datos CFPW .....	45
5.2. Detección de rostros .....	46
5.3. Detección de orientación .....	48
5.4. Etapa de preprocesamiento .....	50
5.5. Extracción de rasgos faciales.....	51
5.6. Etapa de clasificación .....	55
5.6.1. Proceso metodológico para experimentación .....	55
5.6.2. Resultados experimentales.....	55
5.6.3. Experimentos finales.....	58
5.6.4. Evaluación del método propuesto .....	59
5.7. Discusión .....	61
<b>6. Conclusiones</b>	<b>63</b>
6.1. Trabajo a futuro .....	64
6.2. Publicaciones .....	64
<b>Bibliografía</b>	<b>66</b>

# Índice de figuras

2.1. Mejora de imagen mediante la eliminación de desenfoque provocado por el movimiento (McAndrew, 2004) .....	7
2.2. Operaciones lógicas entre imágenes (Passariello, 1995) .....	9
2.3. Puntos de referencia faciales .....	12
2.4. Pixel central y pixeles vecinos .....	15
2.5. Ejemplo de una imagen original y su matriz de co-ocurrencia correspondiente	15
2.6. Imagen de entrada y su espectro frecuencial .....	17
2.7. Red Neuronal Artificial (Kohonen et al., 2001).....	21
2.8. Ejemplo de la arquitectura de una CNN (O’Shea & Nash, 2015).....	21
2.9. Validación cruzada (Developers, 2021a) .....	23
2.10. Comportamiento de la validación cruzada (Developers, 2021c) .....	24
3.1. Diagrama del proceso general de autenticación de personas .....	26
4.1. Proceso general de la autenticación de personas.....	34
4.2. Etapa de entrenamiento.....	35
4.3. Etapa de prueba.....	35
4.4. Metodología propuesta de manera detallada.....	36
4.5. Detección de rostros.....	38
4.6. Resultados de detección de rostros .....	38
4.7. Detección de orientación .....	39
4.8. Resultados de detección de orientación .....	39
4.9. Corrección de pose y escalamiento.....	40
4.10. Resultados de corrección de pose .....	41
5.1. Conjunto de imágenes de la base de datos generada.....	45
5.2. Conjunto de imágenes de la base de datos CFPW .....	45
5.3. Etapas del procesamiento del algoritmo Mask R-CNN .....	47

5.4. Detección de personas e imagen segmentada .....	48
5.5. Detección de orientación (Frente o Perfil).....	49
5.6. Corrección de pose en el eje z y escalamiento.....	51
5.7. Resultados de segmentación de imágenes .....	52
5.8. Distancias a medir en imágenes de frente.....	54
5.9. Distancias a medir en imágenes de perfil .....	54
5.10. Oclusión de los ojos.....	60

# Índice de tablas

3.1. Tabla de comparación de las diferentes técnicas de extracción de rasgos faciales	32
5.1. Resultados de proceso de autenticación utilizando la base de datos generada	56
5.2. Resultados de proceso de autenticación utilizando la base de datos CFPW	56
5.3. Resultados de proceso de autenticación: base de datos generada, estratificación de la base, cv=5, kernel:linear	57
5.4. Resultados de proceso de autenticación: base de datos CFPW, estratificación de la base, cv=5, kernel:linear	58
5.5. Resultados de proceso de autenticación: base de datos generada, estratificación de la base, cv=10, kernel:Poly-6	58
5.6. Resultados de proceso de autenticación: base de datos CFPW, estratificación de la base, cv=10, kernel:Poly-6	59
5.7. Precisión de la clasificación frente a la oclusión de ojos, obtenida por SVM con diferentes kernels (Biswas et al., 2021)	60
5.8. Comparación con trabajo relacionado	61

# Capítulo 1

## Introducción

Gracias al avance tecnológico se han desarrollado diferentes dispositivos como: teléfonos celulares, computadoras de escritorio y portátiles, tablets, diferentes tipos de robots o humanoides e incluso diferentes herramientas diseñadas para realizar actividades específicas. Estos dispositivos cuentan con una parte tanto de hardware como de software y en su mayoría, tienen una interfaz para interactuar con el usuario (persona). Algunos de estos dispositivos, actualmente cuentan con diferentes capacidades que son similares a las que puede realizar una persona de manera cotidiana, como por ejemplo los robots o humanoides. Como resultado del progreso tecnológico, un humanoide puede jugar fútbol, ajedrez, caminar e incluso hasta imitar las expresiones faciales. Para que estas tareas se puedan realizar, es necesario desarrollar una tecnología avanzada en diferentes áreas de investigación, como el área de visión por computadora, la inteligencia artificial, la robótica, el área del lenguaje natural, entre otras.

### 1.1. Problema de investigación

Hoy en día, la visión por computadora y el aprendizaje automático se han convertido en una herramienta importante para el desarrollo de la tecnología y así llegar a nuevos alcances dentro de la inteligencia artificial. El ser humano es capaz de identificar a una persona mediante características como lo son su rostro, su voz, su firma, sus huellas, entre otras.

Para entrar en materia, la autenticación es el acto o proceso de confirmar que alguien es quien dice ser. A la parte que se identifica se le llama probador, mientras que, a la parte que verifica la identidad se le llama verificador. Durante el proceso de identificación, los rasgos característicos se comparan con los conjuntos de patrones almacenados en una base de datos y el resultado de este proceso es la identidad del individuo.

En la actualidad, la autenticación se desarrolla en el área de seguridad a través del análisis

de información en distintos ámbitos, tales como: videos de vigilancia para la identificación de personas sospechosas, el robo de identidad a través de páginas web o en redes sociales e incluso, se ha implementado la autenticación en algunos aeropuertos internacionales para así evitar largas filas al momento del abordaje.

Otro ámbito donde se ha implementado la autenticación de personas, pero de un modo limitado es en la interacción humano-humanoide. Actualmente, existen humanoides que puedan llevar a cabo tareas tanto dentro del hogar, como en algunas empresas o en el ámbito militar. Sin embargo, en esta interacción (humano-humanoide) existen algunos problemas por resolver, por ejemplo: cómo es que los humanoides puedan aprender a identificar a las personas por su nombre, lo cual permitiría un trato más personalizado y generaría un ambiente de confianza para el usuario.

## 1.2. Justificación

En el área de visión por computadora, existen áreas de investigación de especial interés como lo son: la identificación, el reconocimiento y la autenticación de personas. Cabe mencionar que estas áreas son diferentes. El concepto de identificar consiste en localizar o ubicar a una persona dentro de una fotografía o video, mientras que el concepto de reconocimiento responde a la pregunta ¿quién es usted?, aquí se hace una comparación de 1:N, donde N es un grupo de personas infinitas. Mientras que la autenticación responde a la pregunta ¿es usted quién realmente dice ser?, y en este caso se hace una comparación 1:1. Una de las tecnologías de identificación es la biometría, la cual permite que la autenticación de una persona se certifique mediante la comparación de la información obtenida en las imágenes de entrada con las imágenes previamente registradas para la persona que dice ser (Jain et al., 2011).

Dentro de la biometría existen características que se utilizan para la autenticación como la huella dactilar, la retina, las ondas de voz o el rostro de la persona. Sin embargo, en los últimos años los rasgos faciales se han convertido en los principales atributos en el proceso de la autenticación. Como se menciona, técnicas de visión por computadora en conjunto con técnicas de aprendizaje automático, deben ser tomados en cuenta para crear un método de extracción de rasgos faciales, el cual debe contar con ciertas características como: tener una respuesta de procesamiento corta, ser un método comparativo y diferente en relación con los propuestos dentro de la literatura y, por último, tener un buen porcentaje de precisión al momento de la autenticación.

Cuando se habla de ser un método comparativo y diferente, es en relación con el conjunto de técnicas o algoritmos utilizados para la extracción de rasgos faciales. Cabe mencionar que, el método de extracción de rasgos faciales propuesto en este trabajo de investigación contará

con un preprocesamiento de las imágenes, esto con el fin de corregir ciertas características dentro de la imagen, como: iluminación, contraste, eliminar ruido, segmentar las imágenes, usar diferentes filtros para resaltar los bordes en la imagen, entre otras tareas.

Actualmente, existen sistemas de seguridad que son capaces de autenticar personas, que utilizan diferentes técnicas de visión por computadora y aprendizaje automático, pero una aplicación importante es en el área de visión artificial en robots, donde aún se está investigando no solo como identificar a las personas, sino como autenticarlas en la interacción humano-humanoide. En el área de robótica, es de especial interés dotar a los humanoides con la capacidad de poder no solo identificar en donde hay una persona, sino además de verificar de quien se trata, ya que esto permitirá que el humanoide pueda interactuar con las personas de una forma personalizada (dirigirse a las personas por su nombre) bajo condiciones normales en el trato de personas, es decir, que las personas se ubiquen centímetros o pocos metros del humanoide, que las condiciones de luz sean favorables para ver con claridad a la persona, entre otras condiciones. Y es gracias a varias técnicas de visión por computadora y aprendizaje automático que el humanoide pueda lograr la tarea de autenticación.

### **1.3. Objetivos**

Los objetivos para el desarrollo de la tesis están conformados de la siguiente manera:

#### **1.3.1. Objetivo general**

Diseñar un método para la extracción de rasgos faciales en imágenes de rostros humanos mediante la utilización de técnicas de visión por computadora y aprendizaje automático para la autenticación de personas.

#### **1.3.2. Objetivos específicos**

1. Crear un método de extracción de rasgos faciales que permita extraer la información más relevante de los rostros humanos para ser implementado en la autenticación de personas.
2. Seleccionar un clasificador supervisado que por sus características de operación pueda tener un mejor funcionamiento de precisión al momento de autenticar personas.
3. Determinar la precisión del método propuesto comparándola con las técnicas representativas de extracción de rasgos faciales dentro de la autenticación de personas.

## 1.4. Preguntas de investigación

- ¿Cuál algoritmo de extracción de rasgos faciales con un alto grado de precisión en un lapso aceptable puede servir para tener una autenticación de personas favorable dentro de la interacción humano-humanoide?
- ¿Cuáles son los rasgos faciales más relevantes que se deben extraer para favorecer el proceso de autenticación de personas?
- A partir de las características faciales extraídas, ¿cuál algoritmo de clasificación de rostros, de igual manera, con un alto grado de precisión y dentro de un tiempo de respuesta aceptable, servirá para autenticar personas en el momento de la interacción humano-humanoide?

## 1.5. Hipótesis

Partiendo de la visión por computador, utilizando técnicas para la extracción de características faciales en conjunto con un clasificador supervisado es posible llevar a cabo la autenticación de personas dentro de la interacción humano-humanoide.

## 1.6. Alcances y límites

Para alcanzar los objetivos que se plantean, debe de investigarse y analizarse cada una de las etapas que conforman el proceso general de autenticación de personas, que van desde la recolección de un conjunto de imágenes, la etapa de detección de rostros y su orientación, corrección de pose y escalamiento, extracción de rasgos faciales y la etapa de clasificación, para después de analizada cada etapa, se tenga como resultado el diseño de un método de extracción de rasgos faciales que sea preciso y además, comparativo con trabajos relacionados. Para el diseño del método de extracción de rasgos faciales, se deben tomar en consideración diferentes aspectos, como que al momento de capturar el conjunto de imágenes que se utilizarán durante todo el proceso de autenticación, las personas deben estar a no más de 1 metro de distancia con respecto al lugar donde se encuentra la cámara, con el fin de que al momento de hacer el preprocesamiento a las imágenes (escalamiento), los rostros no queden distorsionados y los valores de los rasgos faciales no varíen entre imagen e imagen correspondiente a cada persona.

Otra característica a tomar en cuenta en el momento de la captura de las imágenes es que, haya una buena iluminación, debido que al igual que la distancia a la que se toman las imágenes, la iluminación es un factor que puede ayudar o perjudicar al momento de la aplicación del método de extracción de rasgos faciales, ya que pueden variar los valores de los rasgos.

Otra acotación es que, debido al límite de tiempo, para el desarrollo del trabajo de investigación, solo se analizará un escenario con un máximo de 10 personas, proponiendo como trabajo a futuro, experimentar en un escenario con un mayor número de personas.

Por las características del equipo en el que se realizaron los diferentes experimentos, el tiempo de ejecución sobrepasa el reportado en trabajos relacionados, por ende, otro trabajo a futuro es el reducir ese tiempo de ejecución, para que el método propuesto no solo sea comparativo por su precisión, sino también por su eficacia.

## **1.7. Organización del trabajo**

En el Capítulo 1, se plantea el problema de investigación a resolver, de igual manera se presenta una justificación del porqué es importante resolver este problema. Se formula una serie de preguntas relacionadas a los objetivos específicos, se introduce una hipótesis y, por último, se exponen tanto los alcances como las limitaciones para el desarrollo de este trabajo. En el Capítulo 2, se presenta el marco teórico. En este, se aborda la base teórica para el desarrollo del trabajo. Esta base teórica consiste en la descripción de algunas de las técnicas o algoritmos que se implementarán posteriormente. En el Capítulo 3, se muestran los trabajos más relevantes y recientes encontrados dentro de la literatura. En tanto que, al final del capítulo se presenta una discusión. El Capítulo 4, se presenta la metodología que se va a implementar a lo largo del desarrollo de la tesis, con el fin de lograr el objetivo general que se plantea en el Capítulo 1. Mientras que en el Capítulo 5, aparece la etapa de los resultados obtenidos después de la implementación de la metodología propuesta. Estos resultados pueden ser tanto gráficos, numéricos y comparativos. Por último, en el Capítulo 6 se someten los resultados obtenidos para su análisis y así obtener la sección de conclusiones. De igual manera, en este último capítulo se presenta una breve introducción a un trabajo a futuro y se presentan las publicaciones obtenidas a lo largo del desarrollo de este trabajo de tesis.

# Capítulo 2

## Marco Teórico

En este capítulo, se presenta la base teórica y fundamental de las diferentes técnicas y algoritmos que se utilizaron para el desarrollo del trabajo de tesis. Dando un preámbulo, se abordarán las técnicas para la extracción de rasgos faciales, al igual que, los clasificadores que se utilizan para el proceso de la autenticación de personas.

### 2.1. Procesamiento digital de imágenes

El procesamiento digital de imágenes (PDI) se aplica tanto para resaltar características como para mejorar la visualización de ellas, también se usa para modificar el contraste o, en la aplicación de diversos filtros y así destacar elementos dentro de la imagen. Los datos de la imagen son almacenados de manera digital y esto hace que se puedan manejar de una manera más fácil (Mehl & Peinado, 1997).

El procesamiento de imágenes puede ser visto como la transformación de una imagen en otra mediante diferentes algoritmos aplicados a esta, dando como resultado una imagen mejorada o diferente (Pajares Martinsanz & de la Cruz García, 2002). En la figura 2.1, se muestra un ejemplo de la mejora de la visualización mediante la eliminación de desenfoque provocado por el movimiento.

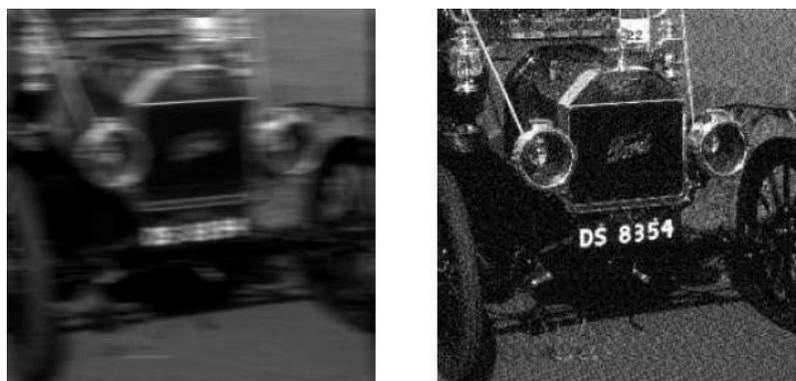


Figura 2.1: Mejora de imagen mediante la eliminación de desenfoque provocado por el movimiento (McAndrew, 2004)

Una imagen es una función de dos dimensiones, que representa una medida de algunas características tales como son el brillo o el color de una escena. Esta es una proyección de una escena 3D dentro de una proyección de plano en 2D. Esto puede ser definido como una función de dos variables como se muestra en la ecuación 2.1:

$$f(x, y) \quad (2.1)$$

donde para cada posición  $(x, y)$  en la proyección del plano, la función define la intensidad de luz en ese punto (González et al., 2009). A continuación, se enlistan algunos tipos de formatos de imágenes digitales:

- Mapa de bits. Es una imagen conformada por un conjunto de píxeles que cada uno cuenta con características como son color y textura. La imagen se forma uniendo miles de píxeles con diferentes características.
- Imágenes vectoriales. Difieren significativamente con los mapas de bits, en dos aspectos:
  - Tienen diferente composición, una imagen vectorial, se conforma de vectores.
  - El tamaño de la imagen es diferente.

Los espacios de color son determinantes a la hora de representar una imagen, existen diferentes los cuales son:

- Mapa de Bits: Ofrece una imagen monocromática sin tonos intermedios: blanco y negro. Solo precisa 1 bit por cada pixel.

- Escala de grises: Ofrece una imagen monocromática de 256 tonos entre el blanco y el negro.
- Color RGB: Ofrece una imagen tricromática, compuesta por los colores primarios en la luz: Rojo, Verde y Azul.
- Color CMYK: Ofrece una imagen cuatricromática, compuesta de los 4 colores primarios para impresión: Cyan, Magenta, Amarillo y Negro.
- Color LAB: Consiste en tres canales: un canal de Luminosidad (L) y dos canales cromáticos, A (que oscila entre verde y rojo) y B (que oscila entre azul y amarillo).
- Color HSV: Define un modelo de color basado en sus componentes: Hue, Saturation y Value (Matriz, Saturación y Valor).

En un video, la resolución de la imagen es uno de los factores fundamentales para definir la calidad de reproducción: 240p, 360p, 480p, 720p, 1080p, etc. La resolución indica la cantidad de pixeles que hay en una determinada medida de longitud. Pixel es la abreviatura de Picture Element, la cual es la unidad más pequeña que se encuentra en las imágenes.

La información de una imagen se puede representar en forma de un histograma, en el que los valores se presentan de manera gráfica en forma de barras, donde cada barra es proporcional a la frecuencia de los valores representados. Este permite constatar la homogeneidad de intensidad de gris de la imagen (Grau, 2003). El histograma de una imagen digital en niveles de grises dentro de un rango de  $[0, L-1]$  es una función discreta. Es común normalizar un histograma, esto se realiza dividiendo cada uno de los valores sobre el total del número de pixeles de la imagen. La suma de todos los componentes de la normalización del histograma debe ser igual a 1 (González & Woods, 2012).

Aparte del histograma, existen diferentes operaciones que se pueden realizar con las imágenes, pueden ser operaciones aritméticas u operaciones lógicas. Un ejemplo de una operación aritmética es la substracción de dos imágenes, dando como resultado una nueva imagen, otros ejemplos pueden ser: la suma y la multiplicación de imágenes, mientras que para las operaciones lógicas es la implementación de una AND, OR y NOT (González & Woods, 2012). En la figura 2.2, se pueden observar unos ejemplos de las operaciones lógicas entre imágenes.

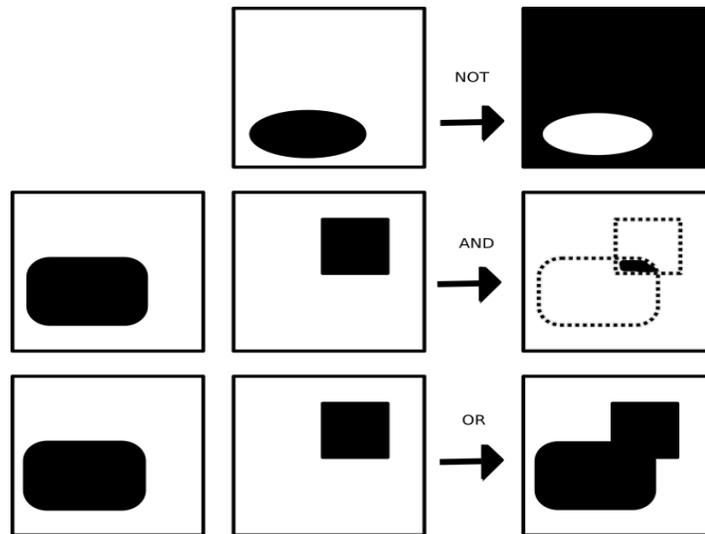


Figura 2.2: Operaciones lógicas entre imágenes (Passariello, 1995)

## 2.2. Aplicaciones del PDI

El procesamiento digital de imágenes está presente en diferentes aplicaciones como en el ámbito industrial, médico, en la agricultura, ganadería, así como dentro de la biometría e identificación y entre otros ámbitos sociales. Como tema a tratar en este trabajo de investigación, al ámbito que se abordará es el relacionado a la autenticación facial (biometría).

### Autenticación facial

La autenticación se puede definir como el proceso de demostración que algo o alguien es realmente que o quien asegura ser. Para que este proceso se lleve a cabo, debe de haber una serie de pasos previos a la autenticación, como se ha mencionado anteriormente, primeramente, puede haber un preprocesamiento de las imágenes, posteriormente una extracción de rasgos faciales mediante diferentes técnicas y finalmente, los rasgos extraídos son almacenados a un clasificador y tener como resultado la autenticación, para el caso de trabajo de tesis, la autenticación de personas.

Existe la biometría, que es un estudio que se dedica al análisis de la autenticación de personas basado en uno o más aspectos ya sea físicos o conductuales. A continuación, se dará una breve explicación de en qué consiste la biometría.

## **Biometría**

Jain et al. (2011) usa el término "biometría", en el sentido más reciente de individualización, como el reconocimiento automático de seres humanos en función de sus rasgos biológicos y de su comportamiento.

Los sistemas para la identificación automática de personas se han basado en datos biométricos duros (imágenes del iris, huellas dactilares y ADN (Ácido desoxirribonucleico)) que requieren la cooperación individual para ser adquiridos. De igual manera, existe la necesidad de identificar a las personas a distancia, esto ha motivado la investigación en la biometría suave, que son atributos físicos y de comportamiento que pueden usarse para identificar a las personas.

La biometría suave permite la identificación de humanos en bases de datos basadas únicamente en descripciones semánticas que abordan cuestiones de invarianza y subjetividad en descripciones de rostros humanos. El rostro humano es rico en características que se pueden usar para identificar personas a distancia, aunque en ocasiones difieren en la medida en que semánticamente puede ser descrita una persona (Almudhahka et al., 2016).

La extracción de los rasgos biométricos de la cara tiene que ver con la identificación de los rostros humanos y las partes faciales en las secuencias de imágenes. Después del procedimiento de detección de rostros, los rasgos faciales como los ojos, la nariz y la boca se extraen del perfil de la cara en función de su disposición geométrica en el esqueleto facial (Chowdhury et al., 2016).

Aparte de usar la biometría para la autenticación de personas mediante el análisis de rasgos faciales, se puede medir mediante la conducta, este es el proceso de medir las tendencias conductuales de un usuario resultantes de las diferencias psicológicas y fisiológicas de persona a persona (Bailey et al., 2014).

### **2.3. Procesamiento digital de imágenes faciales**

Dentro del área de visión por computadora existe una amplia variedad de técnicas para el análisis o el tratamiento de las imágenes. Mientras que para el tema particular que se va a desarrollar a lo largo de este trabajo de tesis, las técnicas a analizar son las utilizadas para la extracción de rasgos en imágenes faciales y los diferentes clasificadores utilizados para la autenticación de personas. Por tal motivo, esta sección se divide en dos subsecciones:

- Técnicas para la extracción de rasgos faciales
- Clasificadores para la autenticación de personas

### 2.3.1. Extracción de rasgos faciales

Como se mencionó anteriormente, el tema a desarrollar durante este trabajo de tesis es la autenticación de personas y una de las partes involucradas en este tema es la extracción de rasgos faciales. La extracción de rasgos en imágenes faciales consiste en extraer información asociada con la activación de los diferentes músculos del rostro, donde se seleccionan regiones de interés del rostro como: los ojos, las cejas y la boca (Castrillon et al., 2008). Existen diferentes categorías en las que se pueden dividir las técnicas para la extracción de rasgos faciales (Kumar et al., 2017):

- Métodos basados en el conocimiento. Son los métodos basados en la geometría de la cara.
- Enfoque de características invariables. Su propósito es encontrar rasgos en la cara que incluso con la iluminación no cambien.
- Métodos basados en plantillas. Estos modelos son sensibles a la variación de la postura y la forma del cuerpo.
- Métodos basados en la apariencia. Se escanea toda la imagen y las regiones de la imagen se identifican como cara o como no cara. (PCA (Principal Component Analysis), Eigenfaces, LBP (Local Binary Pattern), LDA (Linear Discriminant Analysis)).

A continuación, siguiendo con el orden de las técnicas para la extracción de rasgos faciales, se describen algunas de ellas.

#### **Landmarks (Puntos de referencia)**

El entrenamiento de este sistema (Landmarks) está basado en diferentes técnicas, (Le et al., 2012) utilizan la ASM (Active Shape Model) global clásica, el modelo se representa de manera holística como un conjunto de puntos predefinidos llamados puntos de referencia, a lo largo del contorno de la forma del rostro y se describe mediante un vector que concatena todas las coordenadas x de los puntos de referencia ordenados seguidos por todas las coordenadas y. (Kazemi & Sullivan, 2014) muestran cómo se puede usar un conjunto de árboles de regresión para estimar las posiciones de hito de la cara directamente desde un subconjunto disperso de intensidades de píxeles, logrando un rendimiento en tiempo real con predicciones de alta calidad. Estos puntos sirven para medir distancias entre diferentes zonas de interés. La Anthropometry Landmarks marca 68 puntos de referencia, que fueron extraídos del análisis

de una base de datos utilizada para crear este sistema. En la figura 2.10 se muestran los 68 puntos de referencia.

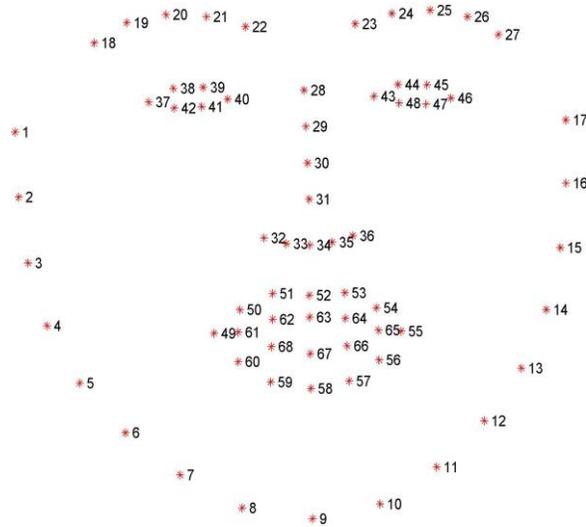


Figura 2.3: Puntos de referencia faciales  
(Sagonas et al., 2013)

### Transformación afín

Las transformadas más usadas para pasar de un sistema de coordenadas a un sistema de fotocoordenadas son: la transformación bidimensional de semejanza, la transformación afín, la transformación bilineal, la proyectiva y la expresada mediante polinomio de orden  $n$ .

La transformación afín, necesita de 6 parámetros, al menos 3 puntos identificables en la imagen y conocidos en ambos sistemas. Se compone de una traslación del origen de coordenadas, un factor de escala para cada eje, una corrección de ortogonalidad y una rotación (Buill Pozuelo et al., 2003).

Para representar las transformadas afines de manera matricial, se necesita usar coordenadas proyectadas, es decir, poder pasar de un plano  $(x, y)$  al plano  $(x, y, 1)$ . La transformada afín queda de la siguiente manera (Rodríguez, 2016):

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = T * \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2.2)$$

### Traslación

La traslación se puede realizar tanto en el eje  $x$  como en el eje  $y$ . De manera algebraica,

es equivalente a agregar una constante a la primera o segunda componente, según sea el movimiento en el eje  $x$  o el eje  $y$ . Utilizando la notación matricial, se puede expresar de la siguiente manera:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & a \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} x + a \\ y + b \\ 1 \end{bmatrix} \quad (2.3)$$

### Rotación

Si se desea rotar una imagen alrededor de su centro, se necesita trasladar la imagen, luego rotarla y, por último, trasladarla nuevamente. Haciendo la multiplicación de esas tres matrices, se obtiene la siguiente ecuación:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & -w/2\cos\theta + h/2\sin\theta + w/2 \\ \sin\theta & \cos\theta & -w/2\sin\theta + h/2\cos\theta + h/2 \\ 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2.4)$$

$w$  y  $h$  son el ancho y la altura de la imagen.

### Escala

Esta transformación expande o contrae una imagen, según las necesidades.  $\alpha$  y  $\beta$  son los factores de escala sobre el plano  $(x, y)$ .

$$T = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.5)$$

### Transvección

Cuando se inclina una imagen, la operación se realiza alrededor del origen de coordenadas. La matriz de transvección en el eje  $x$ :

$$T = \begin{bmatrix} 1 & \tan\alpha & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.6)$$

La matriz de transvección en el eje  $y$ :

$$T = \begin{bmatrix} 1 & 0 & 0 \\ \tan\alpha & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.7)$$

### Métricas estadísticas

Dentro de las métricas más empleadas en el análisis de imágenes son la desviación estándar, varianza y la entropía, por tal motivo, son las que se extraen en el conjunto de rasgos faciales. La desviación estándar es una medida de la propagación de una distribución de los elementos de la matriz. Es la raíz cuadrada de la varianza y está dada por:

$$\sigma = \sqrt{\frac{\sum(x - \mu)^2}{n}} \quad (2.8)$$

La varianza, es una medida de la variabilidad de los datos y está constituida por la siguiente ecuación:

$$\sigma^2 = \frac{\sum(x - \mu)^2}{n} \quad (2.9)$$

Mientras que, la entropía está relacionada con la complejidad de un vecindario estructurado y está dada de la siguiente manera (Python Software Foundation, 2021):

$$- \sum_{x=0}^n p(x) \log_2 p(x) \quad (2.10)$$

### Distancia Euclidiana

La distancia o transformada de distancia, proporciona una medición de la separación existente entre dos puntos dentro de una imagen. La distancia euclidiana está dada por la ecuación 2.11:

$$d_E(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (2.11)$$

donde  $x_1$ ,  $x_2$ ,  $y_1$  y  $y_2$  son las coordenadas de dos puntos dentro de la imagen. La distancia euclidiana es comúnmente usada para medir, de manera puntual en el rostro, el grosor de la cara, el ancho de los ojos, la nariz, la boca y la distancia que hay entre cada elemento anteriormente mencionado (Elizondo & Maestre, 2002).

### Matriz de co-ocurrencia de niveles de grises (GLCM)

La matriz de co-ocurrencia, mide la frecuencia con la que un nivel de gris aparece en una relación espacial entre un pixel llamado central y sus pixeles vecinos, dentro del área de una ventana determinada. La matriz de co-ocurrencia tiene diferentes medidas de textura, las cuales son: homogeneidad, contraste, entropía, energía, correlación, entre otras (Presutti,

2004). En la figura 2.4 se muestra un diagrama de la representación del pixel central y sus vecinos, y de las diferentes direcciones en las que se puede realizar el análisis correspondiente.

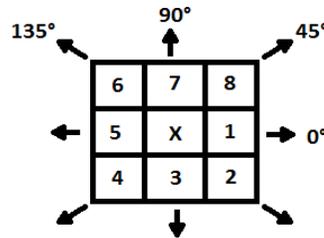


Figura 2.4: Pixel central y pixeles vecinos

Mientras que en la figura 2.5, se muestra un ejemplo de cómo se obtiene la matriz de co-ocurrencia, primero se toma un pixel de referencia, que en este caso es el número 2, el pixel vecino más cercano en horizontal es el 3, ya se tiene un par de pixeles, es momento de ver cuántas veces se repite esta combinación a lo largo de la imagen de izquierda a derecha. Como se observa en el ejemplo, solamente se encontró una vez esa combinación, es momento de poner el número de repeticiones en la matriz de co-ocurrencia en la casilla correspondiente a la combinación encontrada.

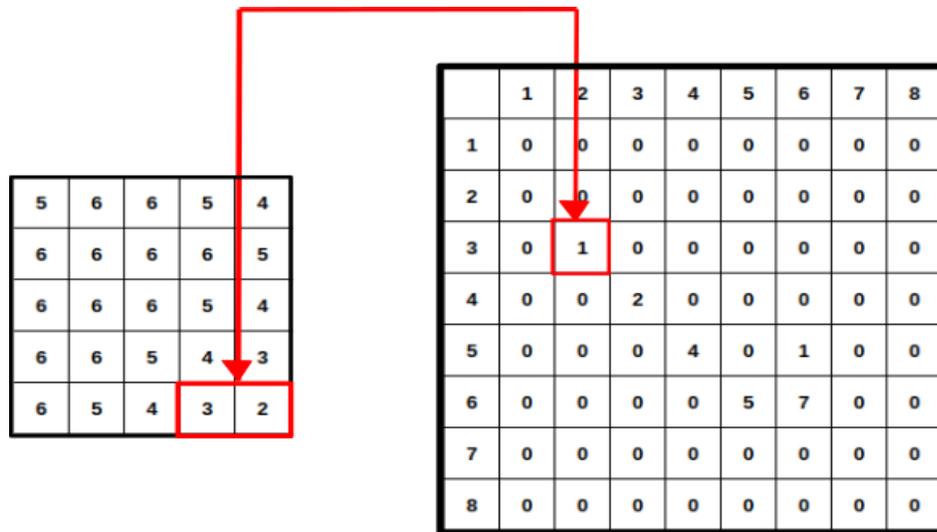


Figura 2.5: Ejemplo de una imagen original y su matriz de co-ocurrencia correspondiente

La matriz de co-ocurrencia  $p(m, n)$ , corresponde a una función de probabilidad de densidad, la cual se obtiene dividiendo todos los elementos sobre el número total de pares de pixeles considerados como se muestra en la ecuación 2.12:

$$p(m, n) = \frac{1}{\text{Todos los pares de píxeles usados}(C_d)(m, n)} \quad (2.12)$$

A partir de esta ecuación, se pueden desglosar diferentes propiedades de la matriz de co-ocurrencia, las cuales van desde la ecuación 2.13 hasta la ecuación 2.21:

- Energía

$$\sum_{m=0}^{G-1} \sum_{n=0}^{G-1} p(m, n)^2 \quad (2.13)$$

- Entropía

$$- \sum_{m=0}^{G-1} \sum_{n=0}^{G-1} p(m, n) \log_2 p(m, n) \quad (2.14)$$

- Contraste

$$\frac{1}{(G-1)^2} \sum_{m=0}^{G-1} \sum_{n=0}^{G-1} (m, n)^2 p(m, n) \quad (2.15)$$

- Correlación

$$\frac{\sum_{m=0}^{G-1} \sum_{n=0}^{G-1} mnp(m, n) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (2.16)$$

donde:

$$\mu_x = \sum_{m=0}^{G-1} m \sum_{n=0}^{G-1} p(m, n) \quad (2.17)$$

$$\mu_y = \sum_{n=0}^{G-1} n \sum_{m=0}^{G-1} p(m, n) \quad (2.18)$$

$$\sigma_x = \sum_{m=0}^{G-1} (m - \mu_x)^2 \sum_{n=0}^{G-1} p(m, n) \quad (2.19)$$

$$\sigma_y = \sum_{n=0}^{G-1} (n - \mu_y)^2 \sum_{m=0}^{G-1} p(m, n) \quad (2.20)$$

- Homogeneidad

$$\sum_{m=0}^{G-1} \sum_{n=0}^{G-1} \frac{p(m, n)}{1 + |m - n|} \quad (2.21)$$

### Transformada de Fourier

La transformada de Fourier es una representación de una imagen como una suma de exponenciales complejas de diferentes magnitudes, frecuencias y fases. La transformada de Fourier juega un papel crítico en una amplia gama de aplicaciones de procesamiento de imágenes, incluyendo la mejora, el análisis y la restauración de imágenes (Gilat, 2006).

La transformada de Fourier es una extensión de las series de Fourier a señales no periódicas. La fase obtenida mediante la transformada de Fourier contiene información esencial sobre la estructura de la imagen. La amplitud por sí sola implica únicamente que existe una estructura periódica dentro de la imagen, pero no especifica dónde se encuentra. Es decir, si no se conoce la fase de la transformada de Fourier, se puede determinar qué objetos hay en la imagen, pero no su posición (Esqueda & Palafox, 2005).

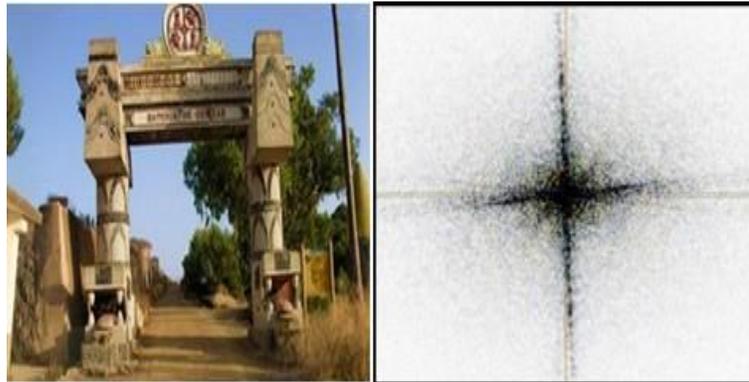


Figura 2.6: Imagen de entrada y su espectro frecuencial

Como se observa en la figura 2.6, se observa una imagen de entrada a la cual se le aplicó la transformada de Fourier y posteriormente, se obtuvo su espectro en frecuencia.

La transformada de Fourier  $F(u, v)$  de una imagen  $f(x, y)$  está dada de la siguiente manera:

$$F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \exp(-2\pi i(ux + vy)) dx dy \quad (2.22)$$

Dejando a un lado las cuestiones matemáticas y las integrales, la Transformada de Fourier se puede ver como la suma generalizada de exponenciales complejas. El análisis de Fourier nos dice que se pueden construir funciones imagen  $f(x, y)$  más complejas incluyendo más términos en la suma generalizada de exponenciales, considerando más pesos específicos de la forma  $F(u, v)$ , de igual manera incluyendo más componentes de frecuencia espacial Pajares Martinsanz & de la Cruz García (2002). En el caso discreto, la transformada de Fourier para una imagen  $f(x, y)$  de dimensión  $M \times N$ , está descrita de la siguiente manera:

$$F(u, v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \exp(-2\pi i(ux/M + vy/N)) \quad (2.23)$$

para  $u = 0, 1, 2, \dots, M - 1$  y  $v = 0, 1, 2, \dots, N - 1$ .

Existen algunas propiedades de la transformada de Fourier, las cuales se van a explicar a continuación.

- Separabilidad. El par de transformadas discretas de Fourier se pueden expresar en la forma separable.

$$F(u, v) = \frac{1}{N} \sum_{x=0}^{N-1} \exp(-i2\pi ux/N) \sum_{y=0}^{N-1} f(x, y) \exp(-i2\pi vy/N) \quad (2.24)$$

La principal ventaja de la propiedad de separabilidad es que  $F(u, v)$  pueden obtenerse en dos pasos por aplicaciones sucesivas de la transformada de Fourier unidimensional.

- Traslación. Las propiedades de traslación del par de transformada de Fourier son:

$$f(x, y) \exp(2\pi i(u_0 x + v_0 y)/N) \Leftrightarrow F(u - u_0, v - v_0) \quad (2.25)$$

$$f(x - x_0, y - y_0) \Leftrightarrow F(u, v) \exp(-2\pi i(ux_0 + vy_0)/N) \quad (2.26)$$

donde la flecha indica la correspondencia entre la función y su transformada de Fourier.

- Periodicidad y simetría conjugada. La transformada de Fourier es una función periódica de período  $N$ .

$$F(u, v) = F(u + N, v) = F(u, v + N) = F(u + N, v + N) \quad (2.27)$$

Lo que indica que  $F(u, v)$  se repita así misma infinitamente para muchos valores  $u, v$ . Con el caso de las imágenes  $f(x, y)$  es real, la transformada de Fourier también exhibe simetría conjugada.

$$F(u, v) = F^* (-u, -v) \quad (2.28)$$

donde  $F^* (-u, -v)$  es la compleja conjugada de  $F(u, v)$

- Rotación. Si introducimos coordenadas polares tendremos:

$$x = r \cos \theta \quad (2.29)$$

$$y = r \sin \theta \quad (2.30)$$

$$u = w \cos \phi \quad (2.31)$$

$$v = w \sin \phi \quad (2.32)$$

donde  $f(x, y)$  y  $F(u, v)$  son ahora  $f(r, \theta)$  y  $F(w, \phi)$ . La situación directa:

$$f(r, \theta + \theta_0) \Leftrightarrow F(w, \phi + \theta_0) \quad (2.33)$$

- Distributividad y escalado. A partir de la definición del par de transformada de Fourier:

$$\mathfrak{F}\{f_1(x, y) + f_2(x, y)\} = \mathfrak{F}\{f_1(x, y)\} + \mathfrak{F}\{f_2(x, y)\} \quad (2.34)$$

La transformada de Fourier es distributiva sobre la adición, pero no en la multiplicación. Y para los escalares

$$af(x, y) \Leftrightarrow aF(u, v) \quad (2.35)$$

$$f(ax, by) \Leftrightarrow \frac{1}{|ab|} F(u/a, v/b) \quad (2.36)$$

- Valor medio. Un valor importante en funciones discretas 2-D, es el valor medio  $m$  dado de la siguiente manera:

$$m = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \quad (2.37)$$

Sustituyéndola  $u = v = 0$  en la transformada de Fourier.

$$F(0,0) = \frac{1}{N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \quad (2.38)$$

La relación entre el valor medio y la transformada queda de la siguiente manera:

$$m = \frac{1}{N} F(0,0) \quad (2.39)$$

- Laplaciana. La laplaciana de una función imagen  $f(x, y)$  está dada por:

$$\nabla^2 f(x, y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \quad (2.40)$$

Aplicada en la transformada de Fourier:

$$\mathfrak{F}\{\nabla^2 f(x, y)\} \Leftrightarrow -(2\pi)^2 (u^2 + v^2) F(u, v) \quad (2.41)$$

El operador Laplaciano es de uso útil, para la extracción de bordes (Pajares Martinsanz & de la Cruz García, 2002).

### Gabor Wavelet

Esta técnica es adecuada para la representación de información de textura, ofrece un análisis de imágenes considerando diferentes escalas o resoluciones y orientaciones (Patel & Shah, 2017). La propiedad importante de la Wavelet es que minimiza el producto de sus desviaciones estándar en el dominio de tiempo y frecuencia. Dicho de otra manera, la incertidumbre en la información transmitida por esta wavelet se minimiza (Soares et al., 2006). La ecuación de Gabor Wavelet en 2D es la 2.42:

$$\Phi_{u,v}(x) = \frac{\|k_{u,v}\|^2}{\sigma^2} \exp\left(\frac{-\|k_{u,v}\|^2 \|x\|^2}{2\sigma^2}\right) \left[ \exp(ik_{u,v} \cdot x) - \exp\left(-\frac{\sigma^2}{2}\right) \right] \quad (2.42)$$

donde  $u$ ,  $v$  son la escala y la orientación del filtro de Gabor y el vector de onda está dado por:

$$k_{u,v} = k_v e^{i\Phi_u} \quad (2.43)$$

Los kernels son aplicados en 8 orientaciones  $u = 0, 1, 2, 3, 4, 5, 6, 7$  y 5 escalas  $v = 0, 1, 2, 3, 4$  (Karthika & Parameswaran, 2016).

### 2.3.2. Clasificadores

Otro elemento importante para el desarrollo de este trabajo es la selección de una técnica o técnicas de clasificación que permita asociar los rasgos de los rostros humanos con el nombre de la persona. La información obtenida del proceso de extracción de rasgos faciales se le da al clasificador como vector de entrada. El clasificador procesa esa entrada y proporciona la salida más adecuada con respecto a la entrada. Existen diferentes clasificadores como los son:

- Redes neuronales
- Máquinas de vectores de soporte
- Métodos basados en geometría
- Clasificadores basados en reglas (Kantharia & Prajapati, 2015)

### Redes Neuronales

Una Red Neuronal es considerada como un sistema de procesadores paralelos conectados entre sí en forma de grafo dirigido. Esquemáticamente cada elemento de procesamiento de la red se representa como un nodo. Estas conexiones establecen una estructura jerárquica que trata de emular la fisiología del cerebro (Sotolongo-Aguilar & Guzmán-Sánchez, 2001).

Una Red Neuronal Artificial (RNA), es una categoría de estos clasificadores, está formada por un conjunto de procesadores elementales, que a partir de estímulos externos generan una salida única (López & Fernandez, 2008). En la 2.7 se muestra el esquema de una RNA.

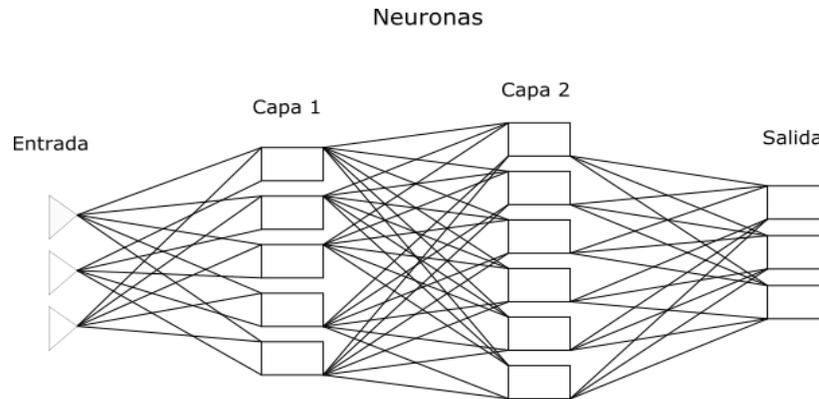


Figura 2.7: Red Neuronal Artificial (Kohonen et al., 2001)

Otra categoría perteneciente a estos clasificadores se encuentran las Redes Neuronales Convolucionales (CNN). Este método de aprendizaje profundo se utiliza ampliamente como parte del reconocimiento de imágenes y el análisis del habla. Combina campos de funciones locales junto con pesos compartidos. La Red Neuronal Convolutiva contiene grupos de capas que se componen de muchos planos de neuronas. Cada unidad en el plano está asociada con una capa pasada usando un vecino local y la salida de una capa es la entrada de la siguiente capa (Patel & Shah, 2017).

Las CNN se componen de 3 capas: capas convolucionales, capas agrupadas y capas completamente conectadas (convolutional, pooling and fully-connected layers). En la figura 2.8, se muestra un ejemplo de una arquitectura simple de una CNN.

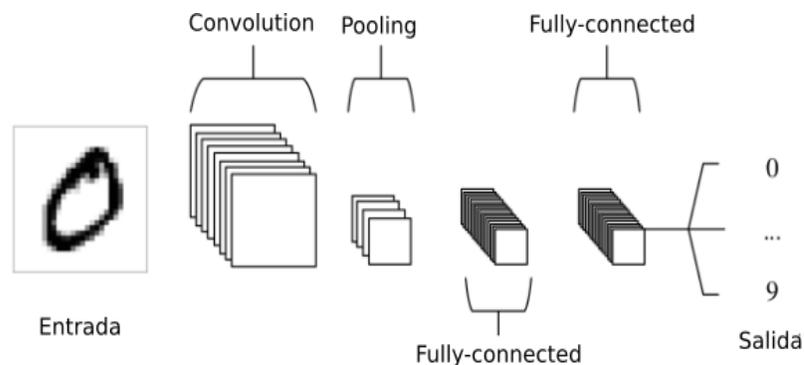


Figura 2.8: Ejemplo de la arquitectura de una CNN (O'Shea & Nash, 2015)

### SVM (Support Vector Machines)

Es un sistema para entrenar de manera eficiente un sistema lineal. Se utiliza como una poderosa técnica de aprendizaje automático para la clasificación de datos. Se realiza un mapeo implícito de datos en un espacio de características dimensionales superiores, y luego encuentra un hiperplano de separación lineal con el margen máximo para separar los datos en este espacio de dimensiones superiores. Este clasificador trabaja sobre la base de planos de decisión usando límites de decisión (Kausar & Sharma, 2016), la ecuación general es 2.44.

$$P(\vec{x}, y) = P(\vec{x})P(\vec{x}|y) \quad (2.44)$$

donde  $\vec{x}$  es un vector de entrenamiento n-dimensional, que además indica a la clase a la que pertenece. Supongamos que tenemos un hiperplano que separa a dos clases diferentes, y está dado por las ecuaciones 2.45 y 2.46:

$$\vec{w} \cdot \vec{x}_i + b \geq 1, \forall \vec{x}_i \in Class1 \quad (2.45)$$

$$\vec{w} \cdot \vec{x}_i + b \leq -1, \forall \vec{x}_i \in Class2 \quad (2.46)$$

Los puntos  $\vec{x}$  que se encuentran en el hiperplano satisfacen  $\vec{w} \cdot \vec{x} + b = 0$  donde  $\vec{w}$  es la normal del hiperplano (Amayri & Bouguila, 2010). Este clasificador trabaja mediante kernels, los cuales son:

- RBF (Radial Basis Function)
- Linear
- Polynomial
- Sigmoid
- Precomputed

## 2.4. Métricas de evaluación

En esta sección, se describirá a grandes rasgos una de las métricas que se utiliza para la evaluación de los resultados obtenidos. Uno de los objetivos de este trabajo de investigación es poder determinar la precisión del método propuesto, por tal motivo es fundamental tener

métricas que puedan evaluar esa eficiencia. La precisión es la capacidad del clasificador de no etiquetar como positiva una muestra que es negativa y está dado de la siguiente manera:

$$precision = \frac{tp}{tp + fp} \quad (2.47)$$

donde  $tp$  son los verdaderos positivos y  $fp$  falsos positivos (Developers, 2021b).

### 2.4.1. Validación cruzada

La validación cruzada (cross-validation) se basa en eliminar de la muestra original ( $Q$ ) una submuestra de datos de tamaño  $n$ . El clasificador se entrena con los  $Q - n$  datos restantes, haciendo el proceso de pruebas con la submuestra inicial. El proceso se realiza hasta que todas las submuestras se eliminan.

La validación cruzada por grupos es más compleja, se trata de evitar la pérdida de datos y se obtienen resultados más robustos en presencia de muestras pequeñas. Se suele considerar 10 grupos distintos de validación, con el objetivo de garantizar una estimación fiable del error de generalización.

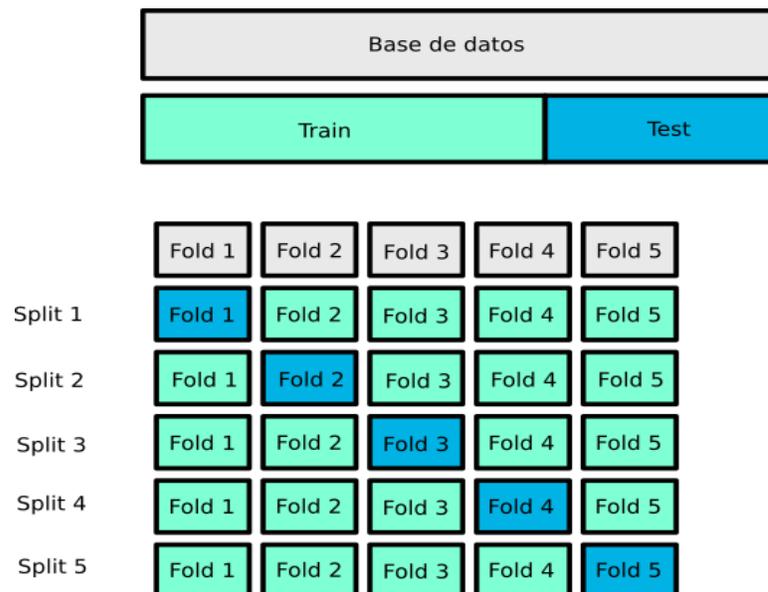


Figura 2.9: Validación cruzada (Developers, 2021a)

Existen distintas variantes de la validación cruzada, pero en este trabajo de investigación se utiliza la variante *Stratified*.

Esta variante, obtiene submuestreos estratificados a partir de la muestra original, de tal manera que, cada grupo contiene aproximadamente la misma proporción de individuos de cada clase que contiene la muestra original (López & Fernandez, 2008).

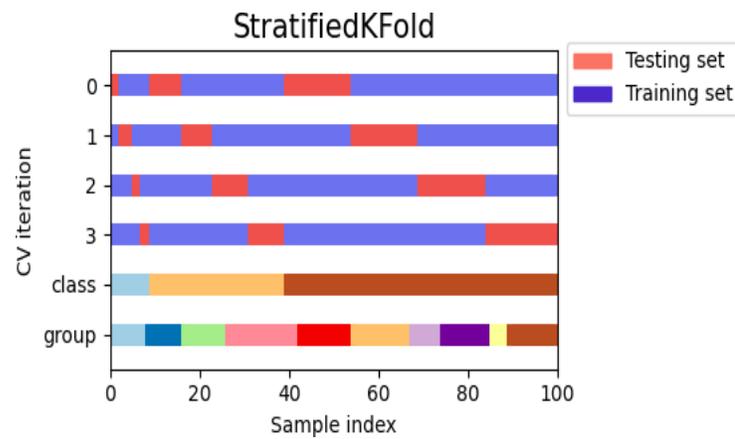


Figura 2.10: Comportamiento de la validación cruzada (Developers, 2021c)

# Capítulo 3

## Estado del Arte

La autenticación de personas involucra conceptos empleados en la visión por computadora, como son: técnicas para la extracción de rasgos faciales y clasificadores para la evaluación de los rasgos faciales extraídos. Dentro de la literatura se reportan trabajos donde emplean diferentes técnicas para extraer los principales rasgos faciales (cejas, ojos, nariz y boca). Dentro de las principales técnicas que son comúnmente usadas se encuentran: PCA, LBP, LDA, Gabor Wavelet, Matriz de co-ocurrencia en niveles de grises, entre otras. Mientras que, en la parte de clasificadores, los mayormente usados dentro de la literatura son: SVM, Redes Neuronales, K-NN y AdaBoost.

En la figura 3.1, se muestra a grandes rasgos el proceso general de la autenticación de personas. Este proceso se divide en varias etapas, la primera es la adquisición de las imágenes, que, para el problema a resolver, se tratan de imágenes faciales. A estas imágenes, en caso dado, se hace un preprocesamiento. Como etapa siguiente, se hace la extracción de rasgos faciales mediante la aplicación de diferentes técnicas. En esta etapa es donde se emplea el método de extracción de rasgos faciales que se propone diseñar, cuyo objetivo del método es la aceleración del proceso de la extracción de rasgos faciales, para así tener una respuesta en un tiempo determinado. Después de haber extraído los rasgos faciales, se entrena el clasificador para tener como resultado un modelo de entrenamiento. Y finalmente, después de haber evaluado el modelo de entrenamiento, se tiene como respuesta la autenticación de las personas.

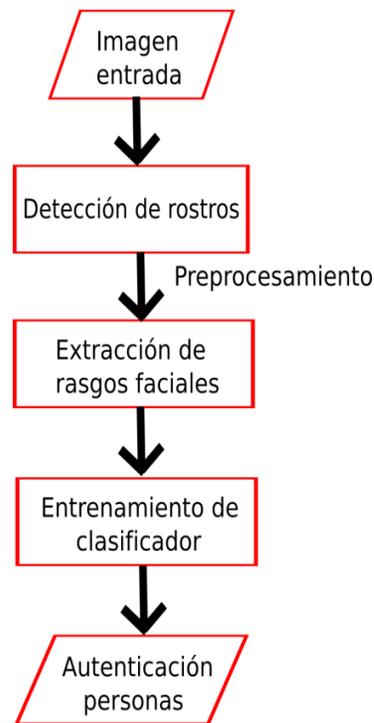


Figura 3.1: Diagrama del proceso general de autenticación de personas

En este capítulo, se presentarán algunos de los trabajos relacionados a la etapa de detección de rostros, la etapa de preprocesamiento, así como de las técnicas utilizadas para la extracción de rasgos faciales y de los diferentes clasificadores utilizados para el proceso de la autenticación de personas.

### 3.1. Detección de rostros

Uno de los detectores más reconocidos e utilizados dentro de la literatura es el desarrollado por (Viola & Jones, 2001), este algoritmo cuenta con 4 etapas: selección de las características Haar, la creación de una imagen integral, el entrenamiento de un clasificador AdaBoost y la toma de la decisión a partir de clasificadores en cascada. Este algoritmo en (Bolotnikova et al., 2017) es utilizado. Los autores, después de haber detectados las caras, las imágenes son transformadas al espacio de color YUV. Se obtiene el histograma de cada canal, de cada resultado se obtiene la función de probabilidad de densidad. Estas funciones se comparan, con el objetivo de tener una métrica de tasa de reconocimiento. Como resultados se tiene una tasa de reconocimiento entre el 91.43 % y el 96.88 %.

Cabe mencionar que, hay otros trabajos en los que en lugar de implementar el algoritmo de Viola-Jones usan algunas otras técnicas para ubicar dónde se encuentra algún rostro. Algunos

ejemplos son (Deng & Pei, 2008), (Nair et al., 2011), (Farfade et al., 2015), (Rikhtegar et al., 2016) y (Bong et al., 2018). Algunos de estos trabajos utilizan Redes Neuronales Convolucionales (CNN, por sus siglas en inglés Convolutional Neural Network), las cuales son entrenadas para poder detectar donde hay un rostro dentro de una escena (imagen o video). En primer lugar, extraen los rasgos faciales de diferentes imágenes y en diferentes posiciones de la cara y posteriormente, se entrena un clasificador y después de la evaluación del modelo se obtiene un detector de rostros. Como resultado se obtuvo un 97 % de exactitud.

En el trabajo de (Zhan & Ando, 2006), utilizan un sensor de intensidad de correlación para obtener una imagen de correlación. A esta imagen, se le aplica la curvatura de Gauss y la curvatura media, con el objetivo de determinar la posición del tabique y la punta de la nariz. Se obtuvo una tasa de detección de 88 %.

(Bormann et al., 2013), presentan en su trabajo una modalidad de detección de rostros, utilizando Viola-Jones y sensores RGB-D (sensores de profundidad), mientras que para la etapa de la clasificación de las caras se usan los métodos de K-Nearest Neighbors y SVM, éstos en conjunto son implementados en un paquete ROS (Robot Operating System) de código abierto. Los autores obtuvieron como resultado un porcentaje de reconocimiento arriba del 90 %.

## 3.2. Preprocesamiento

Basándose en la figura 3.1, una etapa previa a la extracción de rasgos faciales es la etapa de preprocesamiento. Esta etapa es de suma importancia, debido a que se deben hacer todas las correcciones necesarias, para que, al momento de la extracción de rasgos faciales, todas las imágenes de la base de datos tengan las mismas propiedades en cuanto a tamaño (escala), contraste, iluminación, el rostro en una misma posición, entre otras correcciones. Una de las consideraciones a tomar en cuenta es la corrección de pose de la cabeza, es por eso por lo que, en esta sección se presentan algunos trabajos relacionados a este tipo de corrección.

### Corrección de pose

Una parte importante de preprocesamiento es la corrección de la pose de la cara. (Beymer, 1994) propone localizar los ojos y la nariz. Utiliza plantillas para lo que es la alineación geométrica, para medir la eficiencia de la estimación de la postura, hace la correlación entre la plantilla, que usó como base y, la imagen de entrada.

(Wiskott et al., 1997) propone utilizar los coeficientes complejos de Gabor Wavelets obtenidos de una imagen como nodos para la graficación de una malla. A cada pose capturada, se obtiene

una malla diferente. Para la etapa de reconocimiento, se obtiene la gráfica de la imagen de entrada y se compara con todos los modelos y, se selecciona la malla que tenga mayor similitud.

(Rowley et al., 1998) proponen utilizar 2 clasificadores de redes neuronales para la detección de caras frontales sujetas a rotación en el plano. La primera red neuronal está capacitada para estimar la orientación de una cara. Las entradas a la red son valores de intensidad de la imagen. Con la estimación de orientación, una subventana se rota para que la cara potencial quede vertical. La segunda red es un detector facial frontal, el cual tiene la capacidad de descartar algunas partes de la imagen de entrada, con el objetivo de tener una detección favorable.

(González-Jiménez & Alba-Castro, 2007), estos autores usan un conjunto de entrenamiento de mallas faciales de diferentes poses. Se establecen valores típicos que corresponden a una cara frontal. Una malla adopta los parámetros de pose de la otra y, con base en esos valores se hace la corrección de la pose. Ya corregida la imagen con respecto a la pose, para la etapa de extracción de rasgos faciales se utiliza el algoritmo de Gabor Wavelets.

(Murphy-Chutorian & Trivedi, 2008) realizan un análisis de los diferentes enfoques conceptuales para estimar la postura de la cabeza. Existen los métodos basados en plantilla de apariencia, estos comparan una imagen de entrada con un conjunto de ejemplos para encontrar la imagen más similar. Otro enfoque es el basado en la matriz de detectores, aquí se entrena una serie de detectores para diferentes posiciones de la cabeza. Otro enfoque, utiliza la ubicación de regiones de interés (ojos nariz y boca) para determinar la postura. O incluso, hay enfoques que están basados en métodos de seguimiento a partir del análisis en video.

(Kong & Mbouna, 2015) presentan la estimación de los ángulos de postura de la cabeza a partir de una imagen 2D, utilizando un modelo de cara en 3D a partir de un modelo de cara de referencia. Se utiliza una matriz de proyección, la cual relaciona las coordenadas de un punto en el espacio 3D con su proyección en el plano 2D.

(Dahmane et al., 2015) proponen la selección de áreas simétricas en la cara en relación a la intensidad de los píxeles de la piel. En primer lugar, se detecta la cara utilizando el algoritmo de Viola-Jones. Luego, se busca el eje de simetría en el área de la cara, posteriormente, se deduce el ángulo de balanceo a partir de la orientación del eje de simetría.

(Moeini & Moeini, 2015) proponen un modelo elástico genérico de reconocimiento de expresión facial probabilístico 3D para reconstruir un modelo 3D a partir de una imagen frontal 2D. Se genera una matriz de características para cada una de las poses de las diferentes personas de la base de datos. Se obtiene una estimación inicial de los ángulos de triplete a partir de la postura de la cara. Por último, se hace una comparación entre las matrices resultantes mediante la clasificación utilizando una máquina de vectores de soporte.

En (Samad & Iftekharuddin, 2016), proponen un método de representación de curva de espacio

generalizado basado en cuadros Frenet. Las curvas faciales en 3D se extraen de los datos faciales 3D frontales. A través de un marco matemático, se mide la invarianza de la pose, para posteriormente, extraer los rasgos faciales.

### 3.3. Técnicas de extracción de rasgos faciales

En relación con las diferentes técnicas para la extracción de rasgos faciales, dentro de las más usadas en el estado del arte se encuentra la de PCA como se reporta en los siguientes trabajos (Pang et al., 2002), (Kim et al., 2005), (Travieso et al., 2007), (Tsai et al., 2009), (Ou, 2012) y (Wu et al., 2016). Estos autores, utilizan esta técnica para la extracción de los rasgos faciales más relevantes del rostro. Después de aplicar otras técnicas de procesamiento digital de imágenes y de almacenar la información extraída en un clasificador, el algoritmo resultante puede ser utilizado en un sistema de video vigilancia, con el objetivo de detectar personas o incluso rastrearlas. Esta técnica puede implementarse sola o en combinación con otras y de igual manera, en la mayoría de los casos aparecerá de la mano con alguno de los diferentes clasificadores.

(De Jong et al., 2015) utilizan algoritmos genéticos en combinación con las técnicas PCA, LDA y Gabor Wavelets para la extracción de rasgos faciales, estos rasgos son introducidas a un clasificador Bayesiano y una máquina de vectores de soporte. Como resultado se obtuvieron un 75.33 % y un 94.48 % de precisión para cada clasificador respectivamente.

(Wu et al., 2016) realizan un análisis en subespacios, utilizan las técnicas BPCA, LDA y gradiente de Fisherfaces. Se generan 4 subbandas, cada una representa una región de interés: ojo izquierdo, ojo derecho, nariz y boca. Se aplica la métrica de mínimos cuadrados no lineales. Como resultados se obtuvo una tasa de reconocimiento del 81.7 %.

Otra de las técnicas que más se frecuente para la extracción de rasgos faciales es LBP como se muestra en (Tao & Veldhuis, 2010), (Choi et al., 2012), (Guo et al., 2014), (Vipparthi et al., 2016) y (Samangouei et al., 2017). Este operador o técnica se implementa para la extracción de rasgos faciales, al igual que, para el reconocimiento de expresiones faciales. De igual manera, esta técnica es considerada para el análisis de personas con o sin maquillaje, ya que proporciona un buen análisis en el área de textura. Esta técnica divide la imagen del rostro en pequeñas regiones y en cada una de estas regiones se analiza su textura.

Otro trabajo en el que analizan la textura es el que (Liu et al., 2017) presentan. Aquí utilizan la técnica LBP, al cual se obtiene el histograma. Los valores del histograma son introducidos a un clasificador de vecinos más cercanos (k-NN) y se reporta que se obtuvo un 95.94 % de exactitud al momento del reconocimiento facial.

En (Zheng et al., 2016), se utilizan propiedades de textura de la imagen en escala de grises

para detectar los rasgos faciales. Se obtiene el histograma de la imagen para determinar la posición tanto de la cara como de la boca y así poder identificar a una persona.

En (Rikhtegar et al., 2016), los autores utilizan una red neuronal convolucional de 5 capas para la extracción de rasgos faciales, estos rasgos se introducen a un clasificador de máquina de vectores de soporte, para obtener una tasa de reconocimiento del 94.67 %.

Por otro lado, hay trabajos relacionados que extraen otro tipo de rasgos del rostro, como los rasgos basados en la simetría. Estos rasgos pueden ser: medir el contorno del rostro, el ancho y el alto de la cara, medir separación de los ojos, entre otros rasgos. (Li et al., 2018) y (Wu & Ji, 2019) son algunos ejemplos de trabajos relacionados en los que se obtienen métricas del rostro.

### **Otras técnicas de extracción de rasgos faciales**

Además de las diferentes técnicas para la extracción de rasgos faciales anteriormente mencionados, existen algunos otros operadores para extraer rasgos faciales. En (Tong et al., 2014), se optimiza el operador considerando el gradiente horizontal y el diagonal para extraer la mayor información contenida en los cambios de textura de las expresiones faciales. Debido a que al analizar la textura utilizando el gradiente, solo se toma en cuenta el gradiente vertical y de esta manera se puede perder información relevante al momento de analizar los rostros.

Se presenta un enfoque de detección de rasgos faciales, usando características de intensidad en (Park & Lee, 2010). Se utilizan propiedades de textura de la imagen en escala de grises para detectar los rasgos faciales. Se obtiene el histograma de la imagen para determinar la posición tanto de la cara como de la boca y así poder identificar a una persona.

Otra técnica usada es SURF, en (Kumar Shukla et al., 2016) se usa para encontrar rasgos que coincidan en las diferentes imágenes de rostros. Se propone un método que identifica rasgos importantes para la detección de rostros y el reconocimiento facial.

(Choi et al., 2012) realizan la transformación al espacio de color YCbCr, se hace una segmentación local a cada uno de los canales. Para el análisis de las imágenes en relación con la textura, se emplean las técnicas LBP y Gabor Wavelets. Después de haber extraído los rasgos faciales basadas en la textura, estos datos se utilizan para el entrenamiento de una red neuronal para la clasificación. Como resultados del proceso de reconocimiento facial se obtuvo un 88.87 % de precisión.

(Goswami et al., 2016) a partir de imágenes capturadas por un Kinect, se obtiene un mapa de profundidad, un mapa de prominencia y un mapa de entropía. Se aplica la entropía a cada uno de los mapas, se saca el HOG y se concatenan para obtener un descriptor, el cual se mete al clasificador Random Forest. Se obtiene un porcentaje de exactitud del 95.8 %.

(Farfade et al., 2015) presentan un método que obtienen un holograma utilizando diferentes sensores, estos sensores analizan la proyección emitida por la luz. Estas proyecciones generan un patrón único, el cual servirá para determinar la posición en la que se encuentra la cabeza en imágenes en 3D.

### 3.4. Clasificadores

Con respecto a los clasificadores más implementados al momento de la autenticación de personas dentro del área de visión por computadora, se encuentra SVM como uno de los más usados. Algunos de los trabajos reportados son (Wu et al., 2016), (Pang et al., 2002), (Xi & Lee, 2002), (Kaur et al., 2017), (Wagner & Sinzinger, 2008) y en (Kim et al., 2007). En gran parte de estos trabajos siempre se trabaja a la par de una técnica de extracción de rasgos faciales y de un clasificador para así tener como resultado un reconocimiento facial o incluso hasta una autenticación de personas.

Este clasificador se puede usar tanto para la clasificación de castras en imágenes como en video. No solamente se usa para reconocimiento facial, sino también puede servir para clasificar videos, y así poder determinar qué video es el verdadero y cuál es el falso, extrayendo ciertas características de ambos.

Otros de los clasificadores encontrados dentro de la literatura es el algoritmo AdaBoost en (Tsai et al., 2009) y (Kim et al., 2011), lo utilizan para la detección de caras, el primer autor lo implementa con otras técnicas como se mencionó en el párrafo anterior para transformar imágenes con expresiones faciales a imágenes neutrales, a comparación del segundo autor, que presenta un algoritmo de detección falsa basado en el movimiento y la similitud para sistemas biométricos de reconocimiento de rostros. Se tiene un video de entrada el cual se segmenta, se extrae el rostro del fondo y se compara con cada uno de los frames (fotogramas). Un trabajo donde se implementan en conjunto los clasificadores SVM y AdaBoost es en (Chowdhury et al., 2016), donde se presenta un esquema robusto y eficiente para extraer los rasgos biométricos faciales de las imágenes usando reglas difusas. Se usa un espacio de color combinado que consiste en modelos HSV y YCbCr para extraer los componentes de color. Después de la segmentación del color, los componentes extraídos de Y, Cr y Cb se analizan para detectar la región de la cara usando reglas de inferencia difusa. Se emplea la inferencia difusa de estilo Mamdani para categorizar los pixeles como faciales o no faciales. Y los resultados se comparan con los clasificadores SVM y Adaboost.

### 3.5. Discusión

Los trabajos relacionados presentados en este capítulo muestran las distintas técnicas de extracción de rasgos faciales y los diferentes clasificados utilizados para el proceso de reconocimiento o autenticación de personas. En la mayoría de estos artículos, trabajan con bases de imágenes de personas que están totalmente de frente a la cámara y que además, están bajo circunstancias controladas. En la tabla 3.1, se muestran algunos de los trabajos relacionados con las técnicas de extracción de rasgos faciales mostradas en este capítulo.

Trabajo	Técnicas	Base de datos	Precisión
Pang et al., 2002	PCA	Propia	98.5 %
Kim et al., 2005	PCA	Inha univ. y ETRI-A	95.48 %
Travieso et al. 2007	PCA	ORL y Yale	99.33 %
Wu et al., 2016	PCA	ORL, Extended YaleB y Propia	91.7 %
De Jong et al., 2015	PCA, LDA y Gabor Wavelets	TwinsUK y MeIn 3D	94.48 %
Tao y Veldhuis, 2010	LBP	BioID, FERET y YaleB	EER: 2 %
Choi et al., 2012	LBP	FRGC 2.0	91.01 %
Guo et al., 2014	LBP	Propia	80 %
Vipparthi et al., 2016	LBP	Extended YaleB	91.56 %
Liu et al., 2017	LBP	JAFFE	95.94 %
Lekdioui et al., 2017	LBP, CLBP y LTP	CK, FEED y KDEF	96.06 %
Zheng et al., 2016	Histograma	Cohn-Kanade, Bu-EDFE Facial expression y Oulu-CASIA VIS	96.11 %
Rikhtegar et al., 2016	Redes Neuronales Convolucionales	AT&T y Yale	94.67 %
Li et al., 2018	Simetría	CelebA, LFWA y MTFL	86.55 %

Tabla 3.1: Tabla de comparación de las diferentes técnicas de extracción de rasgos faciales

Como se puede observar, tanto a lo largo del capítulo como en la tabla 3.1, las diferentes técnicas que se utilizan en los trabajos relacionados se basan en su mayoría en extraer rasgos faciales basados en textura y no se explora a detalle la combinación de rasgos faciales basadas

en diferentes criterios. Por ejemplo, si en un trabajo relacionado solo se analiza la simetría del rostro, los autores solo se enfocan en ese tipo de rasgos faciales y no utilizan otras técnicas de extracción. Haciendo un análisis más profundo, dentro de la literatura que se ha revisado, no se ha encontrado alguna investigación que utilice varias técnicas de diferentes ámbitos, es decir, que aparte de utilizar las técnicas convencionales también pueda medir la textura de manera local y que, además, pueda medir la simetría del rostro, por tal motivo, en el método propuesto como tema de investigación, se determina que se utilicen distintas técnicas para extraer un conjunto de rasgos faciales que abarquen distintas áreas de análisis como: color, textura y simetría.

# Capítulo 4

## Propuesta Metodológica

Con base en los artículos revisados, se conoce que existe un conjunto de etapas que conforman el proceso de autenticación de personas, en la figura 4.1 se muestra un diagrama general de este proceso.

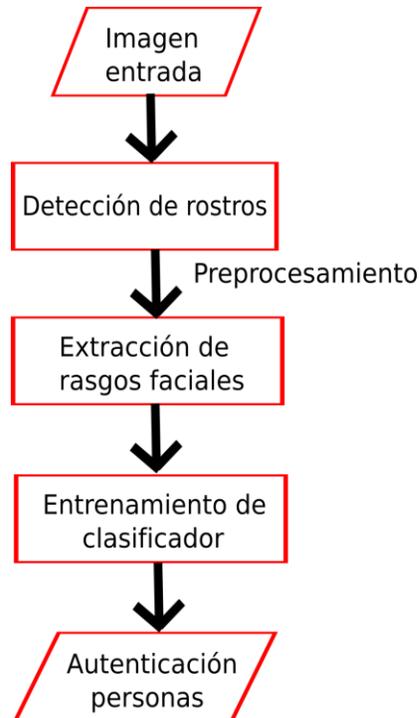


Figura 4.1: Proceso general de la autenticación de personas

Cabe mencionar que este proceso consta de dos etapas: la etapa de entrenamiento y la etapa de prueba. En la primera etapa, se tiene un conjunto de imágenes previamente almacenadas, éstas pasan a la sección de detección de rostros, posteriormente, se le hace un preprocesamiento a

las imágenes, con el fin de corregir varios factores como: iluminación, contraste, corregir la orientación del rostro, hacer un escalamiento de las imágenes, entre otras cosas. Acto seguido, se implementa el método de extracción de rasgos faciales, este método genera un descriptor, el cual sirve para obtener un modelo de entrenamiento generado por un clasificador. Por último, sigue la etapa de pruebas, donde el modelo se evalúa y como resultado final se obtiene la autenticación de la persona. En la figura 4.2, se muestra la etapa de entrenamiento, mientras que en la figura 4.3, se muestra la etapa de prueba.

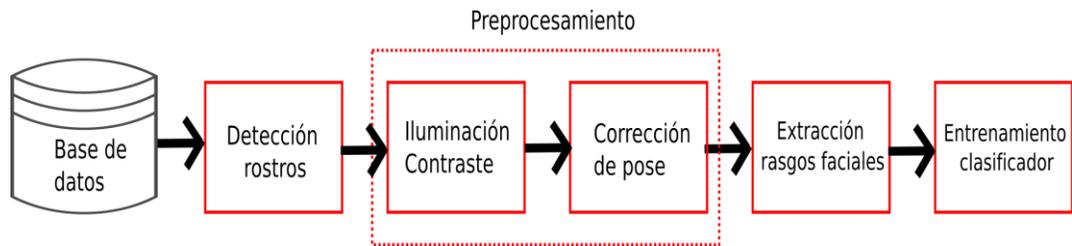


Figura 4.2: Etapa de entrenamiento

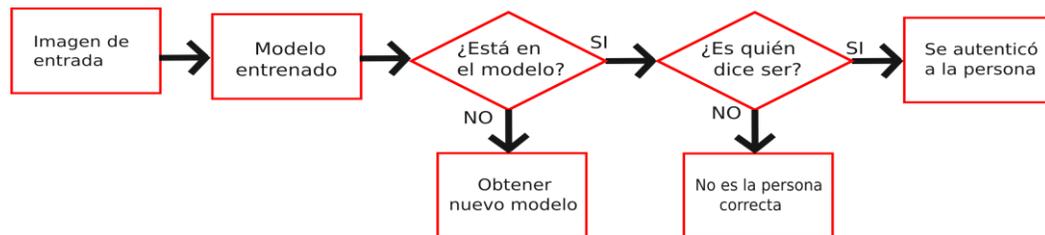


Figura 4.3: Etapa de prueba

En la figura 4.4, se muestra de manera detallada cada una de las etapas y las técnicas que se proponen utilizar. Para tener una idea ordenada de todo el proceso, la descripción de cada etapa se dividirá por secciones.

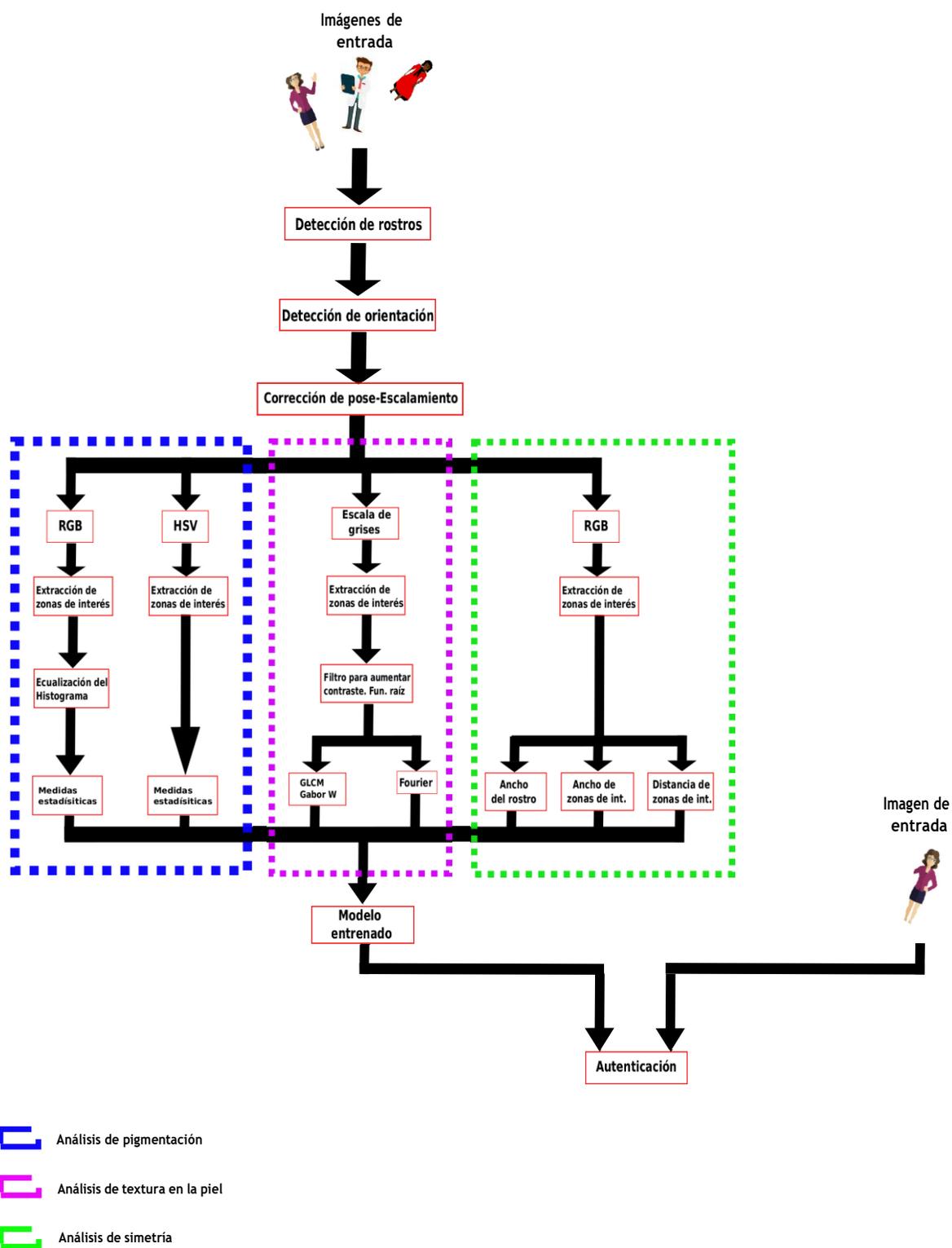


Figura 4.4: Metodología propuesta de manera detallada

Para ilustrar cada una de las etapas que se describirán en este capítulo, se mostrarán algunos de los resultados obtenidos de los experimentos realizados con la base de datos CFPW (Celebrities Frontal-Profile in the Wild), la cual está conformada por 500 personas con un total de 5000 imágenes.

## **4.1. Etapa de entrenamiento del modelo**

A continuación, se presentan los diferentes procesos que conforman esta etapa: detección de rostros, detección de orientación, preprocesamiento, extracción de rasgos faciales y entrenamiento del clasificador.

### **4.1.1. Detección de rostros**

Se tiene un conjunto de imágenes, las cuales pertenecen a diferentes personas que se encuentran dentro de un salón en un ambiente social, los rostros de las personas pueden estar frente o de perfil viendo hacia la cámara. Estas imágenes se utilizan como primera instancia para la sección de detección de rostros. En la figura 4.5 se muestra el proceso general de esta etapa. La detección de rostros se realiza utilizando una Red Neuronal Convolutiva (CNN) ya entrenada. Esta fue entrenada con la base de datos COCO (Common Objects in Context), la cual contiene 91 clases y más de 300000 imágenes (Lin et al., 2014).

Se seleccionó esta CNN, ya que, gracias a sus características, tiene un alto porcentaje de detección sin importar si la persona está a diferentes distancias o su rostro está de frente o de perfil. Cabe recordar que, dentro de la literatura una de las técnicas más implementadas es la de Viola-Jones, pero esta técnica tiene buenos resultados debido al tipo de imágenes con las que trabaja (imágenes de personas que están de frente). Por otro lado, utilizando una CNN se extrae un conjunto de características mucho más ricas y además la CNN puede ser escalable.



Figura 4.5: Detección de rostros

En la figura 4.6, se muestran algunos de los resultados obtenidos para esta etapa.



Figura 4.6: Resultados de detección de rostros

#### 4.1.2. Detección de orientación

Ya con los rostros detectados, sigue la detección de orientación, en la figura 4.7 se muestra el diagrama de los pasos a implementar. Se tiene como entrada las imágenes de los rostros detectados que se obtuvieron en el paso anterior, estas fueron separadas en dos grupos: las imágenes con rostro de frente y las imágenes con rostro de perfil, con el objetivo de utilizar

estos grupos de imágenes para entrenar una CNN y así tener un modelo que determinará la orientación del rostro en una nueva imagen de entrada.

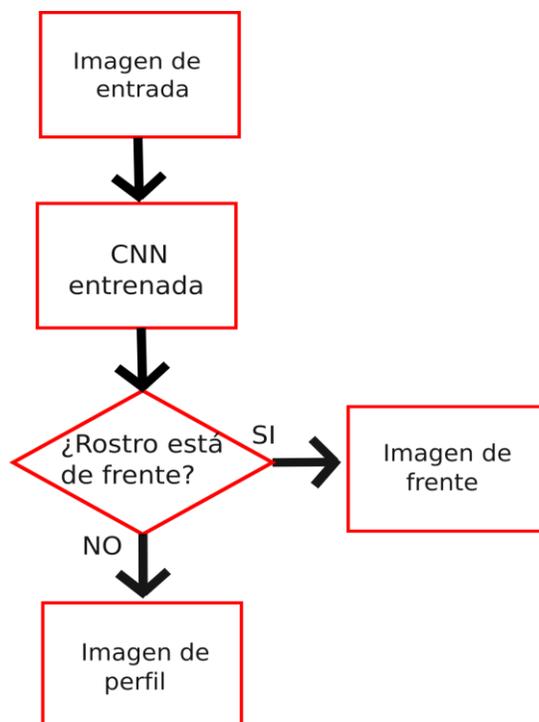


Figura 4.7: Detección de orientación

Algunos de los resultados de la detección de orientación se pueden observar en la figura 4.8.



Figura 4.8: Resultados de detección de orientación

### 4.1.3. Etapa de preprocesamiento

Una etapa importante es la corrección de pose y escalamiento, esto con el propósito de que, al momento de la extracción de rasgos faciales, los valores de estos no varíen entre imagen e imagen de cada persona. En la figura 4.9, se muestra el conjunto de pasos para la corrección de pose y escalamiento. El proceso que a continuación se describe, se realiza a todas las imágenes previamente almacenadas. En la imagen con el rostro detectado, se localizan los ojos y la nariz utilizando la plantilla de Landmarks, se extraen las coordenadas y se calcula el centro de cada ojo y el ángulo entre el centroide de cada ojo. De igual manera, se calcula el punto medio entre los ojos (este punto se localiza arriba de la nariz). Con los datos del ángulo entre los centroides, el punto medio entre los ojos y la escala a la que se reducirá y/o modificará la imagen, se puede determinar una transformación afín, la cual servirá para corregir la imagen en el eje z y escalarla. En la figura 4.10, se muestran algunos resultados de este preprocesamiento, en la fila superior son las imágenes originales, mientras que, la fila inferior son las imágenes ya corregidas en pose.

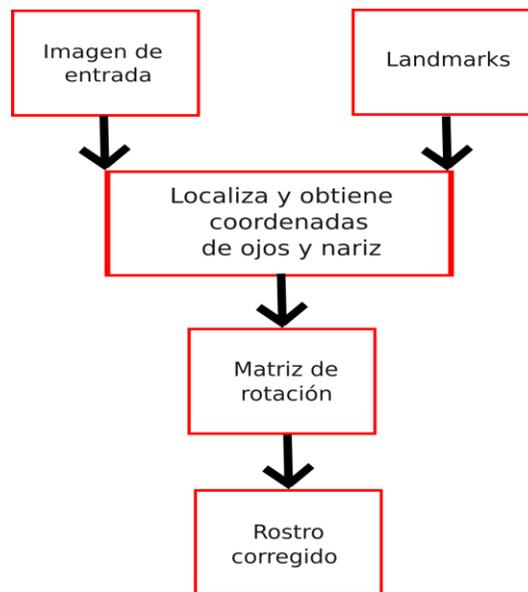


Figura 4.9: Corrección de pose y escalamiento



Figura 4.10: Resultados de corrección de pose

#### 4.1.4. Etapa de extracción de rasgos faciales

Ya hechas las correcciones tanto de pose como en el escalamiento, es momento de la etapa de extracción de rasgos faciales. Los rasgos por extraer se dividen en tres áreas: análisis de pigmentación en zonas de la piel, análisis de textura en la piel y análisis de simetría. Para cada una de estas áreas, las imágenes serán segmentadas y a cada región se le conocerá como ROIs (zonas de interés).

En el área de análisis de pigmentación, a cada imagen correspondiente a las ROIs, se ecualizará el histograma de cada uno de los canales (RGB), con el fin de modificar los valores de los píxeles y así tener una imagen más uniforme. En esta área, los rasgos que se pueden extraer son medidas estadísticas, como: varianza, desviación estándar y entropía. Siguiendo con el análisis de pigmentación, se hace una transformación de color a HSV (Hue, Saturation, Value), de cada uno de estos canales se extraen las medidas estadísticas anteriormente mencionadas. En la segunda área (textura), se aplica un filtro para aumentar el contraste de las ROIs. Dentro de los filtros más utilizados, se encuentran: función gamma, función logarítmica, función exponencial, entre otras. El proceso de la selección del filtro/función se hará de manera automática, implementando algún algoritmo de búsqueda y de igual manera, dependerá de la imagen de entrada, es decir, si la imagen está oscura, necesitará aclararse, o viceversa. La textura se puede analizar tanto en el dominio espacial como de la frecuencia. Para el dominio espacial, se implementarán las técnicas de Matriz de Co-ocurrencia (GLCM, por sus siglas en inglés) y Filtro de Gabor. Mientras que, en el dominio de la frecuencia se aplicará la Transformada de Fourier.

Por último, en el área de simetría se analizará todo el rostro. Utilizando la plantilla de los Landmarks (previamente utilizada), se puede medir la simetría del rostro, es decir, con la obtención de las coordenadas de cada ROI, se puede determinar la longitud, el ancho y la distancia que hay entre cada una de ellas, mediante la distancia euclidiana.

De manera específica, las distancias que se tomaron en cuenta para los casos de frente son:

- Ancho del rostro
- Ancho de los ojos
- Ancho de las cejas
- Ancho de la nariz
- Ancho de la boca
- Distancia de separación de ojos
- Distancia de la línea de las pestañas a la aleta de la nariz
- Distancia de la línea de las pestañas a la comisura
- Distancia de la aleta de la nariz a comisura

Mientras que para los casos de imágenes de perfil:

- Ancho del ojo (derecho o izquierdo)
- Ancho de la ceja (derecha o izquierda)
- Distancia de la aleta de la nariz a la punta de la nariz
- Distancia de la comisura al arco de cupido
- Distancia de la línea de las pestañas a la aleta de la nariz
- Distancia de la aleta de la nariz a comisura
- Distancia de la línea de las pestañas a la comisura
- Distancia del lóbulo de la oreja al mentón

### **4.1.5. Etapa de entrenamiento del clasificador**

Los diferentes rasgos faciales extraídos de las tres áreas son almacenados en un descriptor, el cual se servirá para entrenar el clasificador y así obtener un modelo de entrenamiento. Éste último se evaluará en la siguiente etapa. El clasificador que se utilizará en una SVM (Support Vector Machine). Este clasificador trabaja con diferentes kernels como: rbf, linear, polynomial y sigmoid. Además, se realizará una validación cruzada como medida de evaluación, al igual que, se estratificarán los datos, con el objetivo de que cada clase esté en el submuestreo que se realiza en la validación cruzada.

## **4.2. Etapa de prueba del modelo entrenado**

En esta sección, se describe la última etapa de la metodología propuesta, la que consiste en la evaluación del modelo previamente entrenado.

### **Autenticación de personas**

Para esta etapa, se tiene una nueva imagen de entrada, la cual puede ser una imagen capturada por una cámara fotográfica o una cámara de video. A esta imagen, se le detectará el rostro con la misma CNN que se utilizó en la etapa de entrenamiento. Con el modelo obtenido en la sección de detección de orientación, se podrá determinar si esta nueva imagen está de frente o de perfil. Ya determinada su orientación, se procede a la parte de preprocesamiento, para tener una imagen (corregida de pose y escalada) lo más parecida a las imágenes que se utilizaron en la etapa de entrenamiento. Ya realizada esta corrección, se hará una comparación, entre la nueva imagen de entrada y la información almacenada en el modelo entrenado y así poder obtener como resultado, la autenticación de la persona.

# Capítulo 5

## Resultados de la experimentación

En este capítulo, se presentan los resultados obtenidos a lo largo del desarrollo de este tema de investigación. Este capítulo se divide en 7 secciones, en la primera se mencionan a grandes rasgos, las bases de datos que se utilizan en los experimentos realizados. En secciones posteriores, se muestran los resultados de cada una de las etapas descritas en el capítulo 4: detección de rostros, detección de orientación, corrección de pose y escalamiento, extracción de rasgos faciales y la etapa de clasificación. Mientras que, en la última sección, se muestra una discusión de los resultados obtenidos.

### 5.1. Base de imágenes

Para la realización de los experimentos que conforman este trabajo de investigación, se utilizaron 2 bases de datos: una base de imágenes que se generó y Celebrities in Frontal-Profile in the Wild (CFPW) obtenida de un repositorio. A continuación, se presentarán a grandes rasgos cada una de las bases de imágenes.

#### 5.1.1. Base de datos generada

Esta base de datos cuenta con 300 imágenes de 10 personas diferentes (5 mujeres y 5 hombres), 30 imágenes por persona. En la figura 5.1, se muestran algunas de las imágenes obtenidas. Estas imágenes fueron tomadas por la misma persona a diferentes distancias y circunstancias. El tamaño de la imagen varía y van desde 250x250 a 800x900, estas variaciones son debido al dispositivo con el que se hayan tomado las imágenes, ya que no se utilizó el mismo dispositivo.



Figura 5.1: Conjunto de imágenes de la base de datos generada

### 5.1.2. Base de datos CFPW

La base de datos Celebrities in Frontal-Profile in the Wild (CFPW), cuenta con aproximadamente 5000 imágenes de 500 celebridades (deportistas, políticos, cantantes, entre otros). De las 5000 imágenes, 3000 son frontales y las otras 2000 son de perfil, con un promedio de 14 imágenes por persona, el tamaño de las imágenes es de 250x250. En la figura 5.2, se muestran algunas de las imágenes que contiene esta base de datos.



Figura 5.2: Conjunto de imágenes de la base de datos CFPW

## 5.2. Detección de rostros

Dentro de la literatura, uno de los algoritmos más implementados para la detección de rostros es el de Viola-Jones. Pero como se ha comprobado, en su mayoría trabaja bien con imágenes de frente, para el caso de cuando la persona se encuentra de perfil o a un ángulo diferente, el algoritmo no detecta los rostros. Por tal motivo, fue necesario buscar otros algoritmos que pudieran detectar a las personas sin importar la posición en la que se encontraran.

Como se menciona en el capítulo de metodología propuesta, para esta etapa se utiliza una Red Neuronal Convolutiva ya entrenada. Esta red se entrenó con la base de datos COCO, la cual es un conjunto de datos de detección, segmentación y subtítulo de objetos a gran escala, esta base de datos cuenta con un total de 91 clases, las cuales son: persona, perro, gato, coche, mesa, silla, libro, corbata, etc. Este algoritmo consta de diferentes procesos y parámetros de entrada, los cuales son:

- Imagen de entrada
- Modelo: Mask R-CNN
- Valor de confianza: 0.9
- Valor de umbral: 0.018
- Clases entrenadas: 91

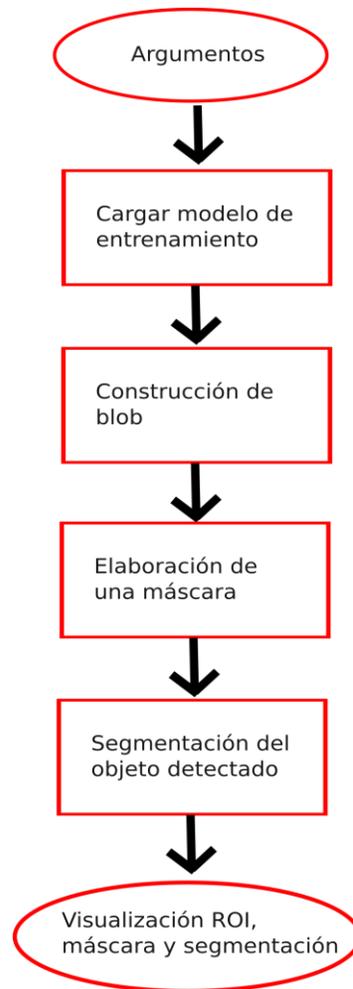


Figura 5.3: Etapas del procesamiento del algoritmo Mask R-CNN

En la figura 5.3, se observan las diferentes etapas que constituyen el algoritmo. En primer lugar, se introducen los parámetros anteriormente en listados, luego, se genera un recuadro, el cual proporciona las coordenadas de la ubicación del objeto detectado. Dentro de un proceso interno, se extrae la clase del objeto detectado (persona, perro, gato, etc.) y se hace una comparación con el modelo de entrenamiento de la CNN, posteriormente, se hace una clonación de la imagen de entrada con la que se construye una máscara, esta máscara sirve para realizar la segmentación del objeto y se aplica un umbral para obtener una imagen binaria. Como etapa final, se extrae la región de interés (ROI por sus siglas en inglés) y se visualizan las imágenes finales (ROI, máscara y la imagen segmentada). La imagen final (ROI), presenta la ubicación de la persona u objeto y un porcentaje de precisión. En la figura 5.4, se muestran algunas de las imágenes resultantes después de la implementación del programa Mask R-CNN.



Figura 5.4: Detección de personas e imagen segmentada

Cabe mencionar que este algoritmo, solamente detecta que clase de objeto se encuentra en la imagen, es decir, determina si hay una persona, un perro, un gato, etc. Pero para el trabajo de investigación, basta con que detecte la ubicación de la persona.

### 5.3. Detección de orientación

Tomando como referencias las bases de datos que se menciona en la sección 5.1, se buscó algún algoritmo que fuera capaz de detectar la orientación del rostro de la persona.

Una de las opciones más viables era utilizar alguna máscara o entrenar una Red Neuronal y se optó por la segunda opción. Se encontró una Red Neuronal Convolutiva (CNN, por sus siglas en inglés), la cual tiene como arquitectura una variante más pequeña y compacta de la red VGGNet (Visual Geometry Group Network). El entrenamiento de la CNN fue a través de Keras y Deep Learning, todo esto en Python. A continuación, se mencionan algunas de las características para el entrenamiento de la CNN.

- Para la etapa de entrenamiento, se debe contar con mínimo 3 clases (personas, cosas, etc.)
- Por cada clase, se necesitan como mínimo 100 imágenes.
- Tiene una iteración de 100 épocas.
- Para el entrenamiento se utiliza el 75 % de las imágenes y el 25 % restante se usa para pruebas.

Una característica importante a tomar en cuenta de esta CNN es que para el entrenamiento se debe tener como mínimo 3 clases. Para los experimentos de este trabajo de tesis, se pusieron como clases: imágenes de frente, imágenes de perfil e imágenes de un objeto. Después de haber entrenado la CNN, se realizaron pruebas para ver la eficiencia de la clasificación. En la figura 5.5, se muestran algunos de los resultados de detección de orientación.

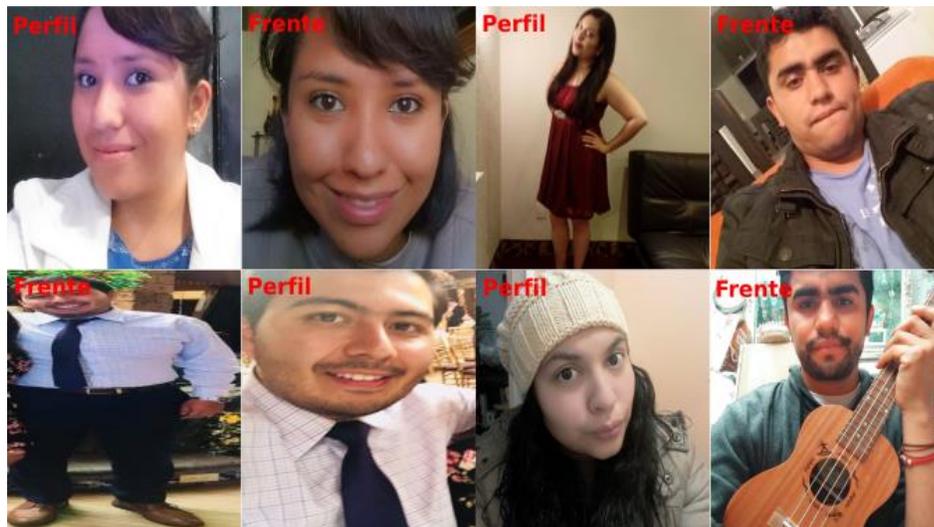


Figura 5.5: Detección de orientación (Frente o Perfil)

## 5.4. Etapa de preprocesamiento

Una etapa previa a la extracción de rasgos faciales es la etapa de preprocesamiento. Esta etapa es importante, debido a que se hacen las correcciones o mejoras necesarias a las imágenes, con el objetivo de que, en el momento de la extracción de rasgos faciales, las imágenes estén en condiciones similares (iluminación, contraste, el rostro ubicado en una misma posición, entre otras).

### Corrección de pose

Dentro de la literatura, se encontraron algunos algoritmos utilizados para la corrección de pose, tanto para el eje  $y$  como para el eje  $z$ . Para este trabajo de investigación, se utilizó una plantilla de Landmarks para detectar los ojos, la nariz y la boca. Esta plantilla detecta 68 puntos de referencia, la cual fue generada con la base de datos iBUG-300-W. Después de detectar los puntos de referencia, se hace una comparación de las coordenadas de las zonas de interés entre la imagen de entrada (imagen no corregida) y la posición deseada. Si las coordenadas son las deseadas, no se hace ninguna corrección de pose y ningún escalamiento, pero si las coordenadas no coinciden, se aplica una matriz de transformación o transformación afín, la cual incluye un proceso de rotación, traslación y escalamiento. Después de aplicar la transformación, se hace una nueva comparación para ver si la imagen de entrada ya se encuentra en la posición deseada, si es así, la imagen fue corregida tanto en el eje deseado y escalada al tamaño deseado. En la figura 5.6 se muestran algunos de los resultados obtenidos de la aplicación del algoritmo para la corrección de pose en el eje  $z$  y el escalamiento.

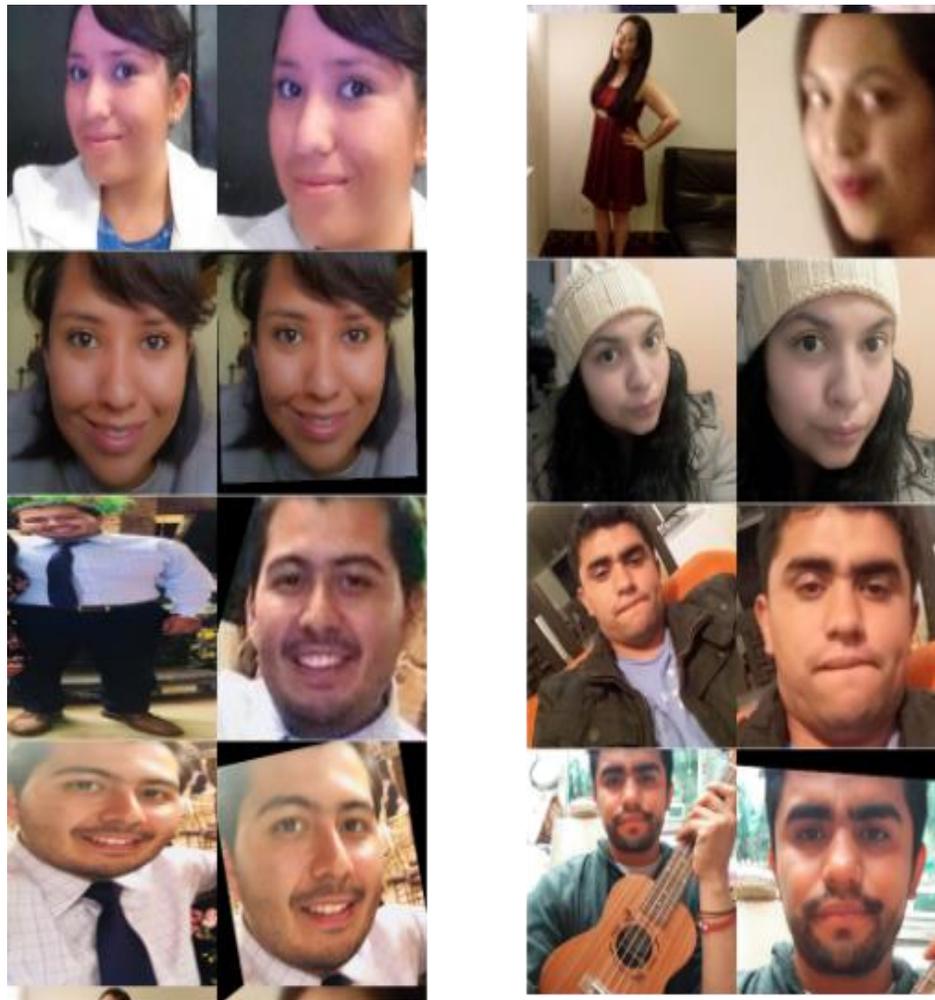


Figura 5.6: Corrección de pose en el eje z y escalamiento

## 5.5. Extracción de rasgos faciales

Después de haber realizado la corrección de pose en las imágenes, se prosigue con la etapa de la extracción los rasgos faciales, pero con el fin de no analizar la imagen completa, se segmenta el rostro.

### Segmentación de las zonas de interés

El proceso de segmentación se realiza utilizando la plantilla de Landmarks con los 68 puntos de referencia. Ya detectados estos puntos de referencia en el rostro, se hace un recorte a cada una de las ROIs, las cuales son delimitadas por las coordenadas obtenidas con la plantilla

de los Landmarks y cada ROI es guardada como una imagen independiente. Las zonas de interés a segmentar son: ojo izquierdo, ojo derecho, ceja izquierda, ceja derecha, nariz y boca. En la figura 5.7, se muestran algunos rostros segmentados.



Figura 5.7: Resultados de segmentación de imágenes

Después de segmentar los rostros y tener las ROIs, se extraen los rasgos faciales de cada una de esas regiones. A continuación, se presentan los diferentes rasgos faciales que se extraen:

- Desviación estándar (R,G,B)
- Varianza (R,G,B)
- Valor máximo (H,S,V)
- Valor mínimo (H,S,V)
- Desviación estándar (H,S,V)
- Varianza (H,S,V)
- Entropía (H,S,V)
- GLCM

- Contraste
- Energía
- Correlación
- Homogeneidad
- Disimilitud
- Entropía
- Fourier
  - Media de la magnitud del espectro
  - Varianza de la magnitud del espectro
- Gabor
  - Media del filtro de Gabor
  - Varianza del filtro de Gabor
- Distancias de los ROI's
- Distancia entre los ROI's

Detallando los rasgos faciales relacionadas a simetría, la obtención de estas distancias es a través de la herramienta de la distancia euclidiana, dentro del algoritmo se hace de manera automática al momento de la segmentación. Las distancias que se midieron en las imágenes de frente se muestran en la figura 5.8:

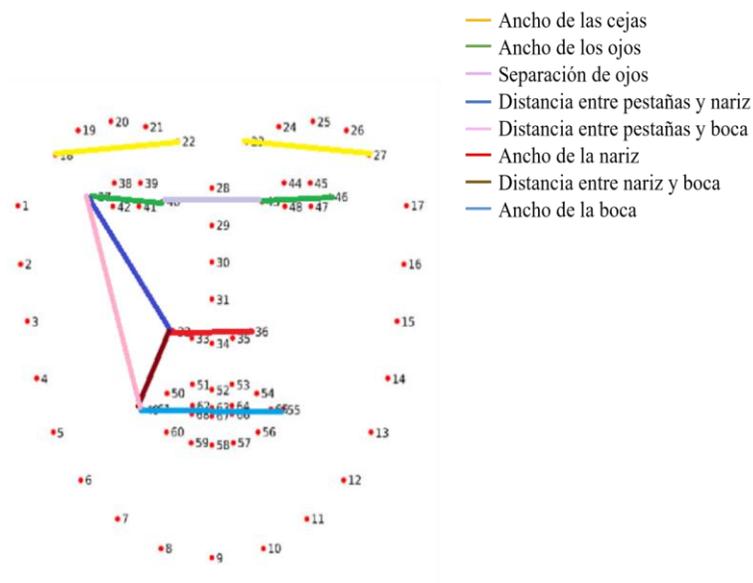


Figura 5.8: Distancias a medir en imágenes de frente

Por otro lado, las distancias que se tomaron en cuenta para los casos de perfil se ilustran en la figura 5.9:

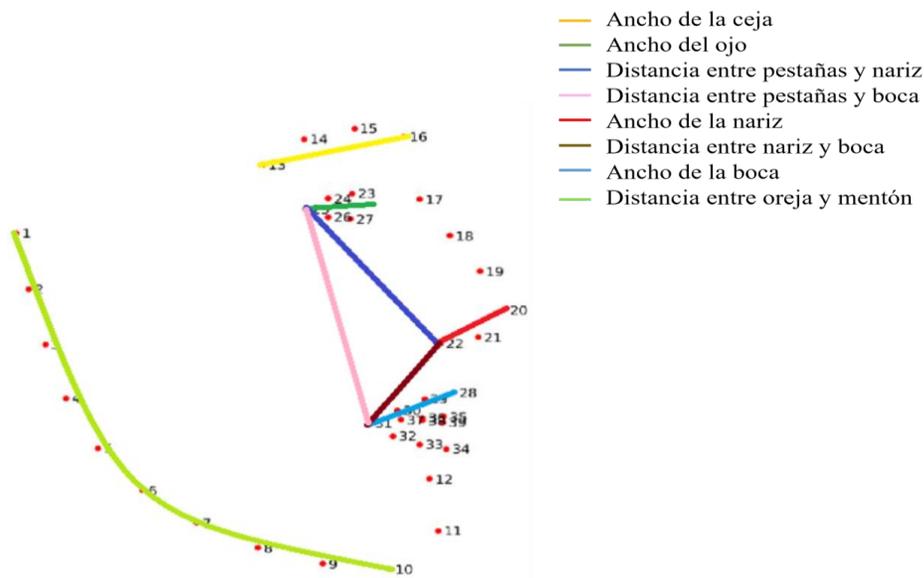


Figura 5.9: Distancias a medir en imágenes de perfil

## 5.6. Etapa de clasificación

En esta sección se presentarán los resultados obtenidos de los diferentes experimentos realizados. Los resultados, son utilizando las 2 bases de datos que se mencionan al principio de este capítulo. Además, como se mencionó en el capítulo anterior, el clasificador que se utiliza es una SVM. En los distintos experimentos que se realizaron, se obtuvieron diferentes porcentajes de precisión, pero conforme se hicieron algunos ajustes en la obtención de los rasgos faciales y enriqueciendo el algoritmo, se fue mejorando el porcentaje de precisión.

### 5.6.1. Proceso metodológico para experimentación

Los experimentos realizados, se efectuaron considerando los siguientes elementos:

- Seleccionar de base de datos.
- Determinar el número de personas para cada experimento (2-10).
- Definir el número de imágenes a utilizar por cada persona.
- Elegir la distribución de las muestras para el entrenamiento y las pruebas.
- Determinar el kernel a utilizar de SVM.
- Medir la precisión mediante la métrica de evaluación.

Se tiene una base de datos, la cual está constituida por 10 personas (base de datos generada), estas a su vez, cuentan con un promedio de 30 de imágenes por persona. Para la etapa de la experimentación, el conjunto de rasgos faciales extraídos de cada una de las imágenes se divide en muestras tanto para el entrenamiento del clasificador como para las pruebas correspondientes. En una etapa inicial, los últimos 3 puntos se tomaron con sus valores por default, estos elementos conforme se fue avanzando en la investigación, se fueron ajustado hasta obtener un método que fuera comparativo con trabajos relacionados.

### 5.6.2. Resultados experimentales

Considerando el proceso metodológico descrito en la sección 5.6.1, en este apartado se presentan los resultados obtenidos de la experimentación. El número de imágenes por cada persona fue reducido (6 imágenes por persona), con el objetivo de ver el comportamiento del método ante la escasez de imágenes y tratando de que el tiempo de respuesta fuera corto. Como

recordatorio, a todas las imágenes de cada persona, se segmentaron las ROIs y a cada una de estas, se extrajeron los rasgos faciales correspondientes a las diferentes áreas de análisis. El muestreo se hizo de la siguiente manera: del total de las ROIs y los rasgos faciales resultantes de cada una de éstas, se dividió un 75 % para la etapa de entrenamiento, mientras que, el 25 % restante se utilizó para la etapa de pruebas. Cabe mencionar que, este muestreo se hizo de manera aleatoria.

### Base de datos generada

En este apartado, se presentan los resultados obtenidos de la experimentación utilizando la base de datos generada. En la tabla 5.1, se muestran los resultados obtenidos de los diferentes experimentos.

Número de personas	Precisión Global
2 personas	81 %
4 personas	69 %
6 personas	52 %
8 personas	54 %
10 personas	47 %

Tabla 5.1: Resultados de proceso de autenticación utilizando la base de datos generada

### Base de datos CFPW

Al igual que con la base de datos generada, para estos experimentos solamente se utilizaron 6 imágenes por cada persona. En la tabla 5.2, se muestran los resultados obtenidos de los diferentes experimentos realizados.

Número de personas	Precisión Global
2 personas	90 %
4 personas	51 %
6 personas	47 %
8 personas	43 %
10 personas	42 %

Tabla 5.2: Resultados de proceso de autenticación utilizando la base de datos CFPW

Haciendo un análisis de estas pruebas iniciales, cuando se tiene dos sujetos, la clasificación se comporta de manera aceptable, sin embargo, es importante notar que cuando se fue aumentando el número de sujetos, la precisión global bajó, esto debido a que el número de imágenes por persona no es suficiente para tener un buen conjunto de elementos para el muestreo, también que en la etapa de entrenamiento del clasificador, por hacerse un muestreo aleatorio, no se garantiza que por lo menos una imagen de cada clase se encuentre en el modelo de entrenamiento. Tomando esto en cuenta, se procedió a realizar una segunda etapa de experimentos, considerando las siguientes correcciones:

- Se aumentó el número de imágenes por persona.
- Se estratificó la base de datos.
- Se realizó una validación cruzada.

### Base de datos generada

Como se menciona en el párrafo de anterior, para esta segunda etapa de experimentos, se hicieron algunos ajustes a nuestras variables iniciales, como pasando de 6 imágenes por persona a 30 imágenes por persona (en el caso de esta base de datos). Otra modificación que se realizó fue en la etapa del entrenamiento del clasificador, de tener un muestreo aleatorio, se hace una estratificación de la base de datos, esto con el objetivo de que cada clase (persona) esté en el submuestreo y, además, se realizó una validación cruzada de 5 como métrica de evaluación. En la tabla 5.3, se observan los resultados que se obtuvieron.

Datos	Precisión	Std	Tiempo (seg)
2 personas	100 %	0	19.34
4 personas	94.27 %	1.0	36.94
6 personas	90.07 %	1.41	55.02
8 personas	75.81 %	2.23	76.62
10 personas	73.92 %	2.82	86.71

Tabla 5.3: Resultados de proceso de autenticación: base de datos generada, estratificación de la base, cv=5, kernel:linear

### Base de datos CFPW

Se hace el mismo procedimiento que en la base de datos generada, pero en este caso, se utilizan las 14 imágenes por persona (son las que tiene la base de datos original), se estratifica la bases y se realiza una validación cruzada de 5. En la tabla 5.4, se muestran los resultados.

Datos	Precisión	Std	Tiempo (seg)
2 personas	90.62 %	0	6.83
4 personas	78.94 %	1.0	14.29
6 personas	58.92 %	1.41	21.65
8 personas	63.33 %	2.0	28.47
10 personas	46.15 %	2.23	35.56

Tabla 5.4: Resultados de proceso de autenticación: base de datos CFPW, estratificación de la base, cv=5, kernel:linear

### 5.6.3. Experimentos finales

Revisando trabajos relacionados publicados en los últimos años, algunos autores utilizan un kernel diferente al que se usó para la realización de los experimentos, por tal motivo, se hicieron unos últimos experimentos, utilizando el kernel Poly-6 y, además, se hizo una validación cruzada de 10. En las tablas 5.5 y 5.6 se muestran los resultados obtenidos con las 2 bases de datos.

Datos	Precisión	Std	Tiempo (seg)
2 personas	100 %	0	60.70
4 personas	97.82 %	1.0	123.35
6 personas	96.12 %	1.73	169.15
8 personas	98.30 %	2.23	238.47
10 personas	99.11 %	2.82	300.19

Tabla 5.5: Resultados de proceso de autenticación: base de datos generada, estratificación de la base, cv=10, kernel:Poly-6

Datos	Precisión	Std	Tiempo (seg)
2 personas	100 %	0	18.56
4 personas	98.58 %	1.0	52.10
6 personas	97.74 %	1.41	69.49
8 personas	96.25 %	2.23	94.24
10 personas	91.66 %	2.64	114.46

Tabla 5.6: Resultados de proceso de autenticación: base de datos CFPW, estratificación de la base, cv=10, kernel:Poly-6

#### 5.6.4. Evaluación del método propuesto

Haciendo una comparación entre un trabajo relacionado y la propuesta metodológica, (Biswas et al., 2021) proponen una técnica para la extracción de rasgos faciales y la denominan OSF-DNS (One-Shot Frequency Dominant Neighborhood Structure). En este trabajo, hacen una comparación de diferentes técnicas utilizadas para la extracción de rasgos faciales dentro del estado del arte, como son: Dlib, VGG16, OpenFace, FaceNet, ArcFace, SS-S-SF (Salient Structure Features), RP-IVD (Ring Partition and Invariant Vector Distance), pHash (Percentual Image Hash) y OSF-DNS y como clasificador se utiliza una SVM, realiza diferentes experimentos con los diferentes kernels con los que cuenta el clasificador.

La idea general de la propuesta de los autores es que, de una base de datos extraen los rasgos faciales utilizando su propuesta, las cuales son almacenadas en un vector para entrenar un clasificador SVM y así obtener un modelo. Por otra parte, a las imágenes de las bases de datos, les hacen la oclusión de los ojos o la boca. Posteriormente, a las imágenes modificadas, les extraen los rasgos faciales que ellos proponen y también son almacenadas. El modelo entrenado de las imágenes sin modificaciones es utilizado para evaluar los nuevos rasgos (de las imágenes modificadas) y como resultado se obtiene la clasificación de las personas.

Base	Kernels	Dlib	VGG16	OpenFace	FaceNet	ArcFace	SS-S-SF	RP-IVD	pHash	OSF-DNS
LFW	Linear	4.00	12.67	5.3	13.45	6.58	28.01	23.81	77.08	82.74
	RBF	9.36	4.81	11.2	4.53	4.01	8.22	17.31	77.66	45.34
	Poly-2	19.25	4.81	21.5	14.23	6.82	16.28	12.42	74.34	89.53
	Poly-4	27.39	4.89	9.5	20.02	4.33	14.98	7.02	72.80	88.14
CFPW	Poly-6	23.56	4.54	5.3	20.02	2.78	12.86	7.08	72.16	86.06
	Linear	48.78	14.56	37.42	50.94	3.23	51.06	19.30	28.98	26.00
	RBF	50.94	7.82	2.0	54.14	2.0	43.74	13.78	41.32	18.18
	Poly-2	46.82	6.82	29.38	49.26	3.75	32.52	19.16	52.32	38.52
	Poly-4	31.02	3.01	10.56	22.06	2.33	28.86	19.32	55.10	58.60
	Poly-6	17.28	1.29	3.92	6.68	2.05	24.80	19.32	61.54	63.24

Tabla 5.7: Precisión de la clasificación frente a la oclusión de ojos, obtenida por SVM con diferentes kernels (Biswas et al., 2021)

Para poder comparar la propuesta con el trabajo relacionado, en la figura 5.10 se muestran las modificaciones que se hicieron a las imágenes, mientras que, en la tabla 5.8 se presentan los resultados que se obtuvieron.



Figura 5.10: Oclusión de los ojos

Base	Kernels	OSF-DNS	Propuesta
CFPW	Linear	26.00 %	24.19 %
	Poly-6	63.24 %	65.32 %

Tabla 5.8: Comparación con trabajo relacionado

En esta tabla, solo se presentan los resultados obtenidos utilizando 2 kernels diferentes: linear y Poly-6. Estos fueron los dos kernels que se utilizaron en los experimentos realizados a lo largo del desarrollo de este trabajo de tesis.

## 5.7. Discusión

En este capítulo, se presentaron todos los resultados gráficos y numéricos de cada una de las etapas del proceso general de la autenticación de personas, desde: detección de rostros, detección de orientación, corrección de pose y escalamiento, extracción de rasgos faciales y la etapa de clasificación. En todas estas etapas, se utilizaron 2 bases de datos totalmente diferentes, en cuanto a tamaños, iluminación, calidad de la imagen, diferente cantidad de personas e imágenes por persona, entre otras cosas. La parte de suma importancia, es la etapa de clasificación, donde se mide la eficiencia del método propuesto que se desarrolló a lo largo del trabajo de investigación.

Haciendo una comparación de todos los resultados obtenidos, desde la primera etapa de experimentación hasta los resultados de experimentos finales, se notó una mejoría en relación con el porcentaje de precisión. Esto debido a los ajustes que se realizaron a las variables mencionadas en el proceso metodológico (subsección 5.6.1), al igual que, el último cambio que se realizó fue pasar de un kernel linear a un kernel Poly-6, esto, gracias al análisis de trabajos publicados recientemente.

Analizando los resultados obtenidos con la base de datos generada, se muestra un buen rendimiento para el caso de 10 personas, con un porcentaje de precisión pasando de un 73.92 % a un 99.11 %, utilizando un kernel linear y un kernel Poly-6 respectivamente.

Considerando este cambio de kernel y viendo que se mejoró el porcentaje de precisión, el método de extracción de rasgos faciales propuesto en este trabajo de investigación se pudo comparar con el estado del arte, utilizando la base de datos CFPW. Tomando como comparación los resultados mostrados en (Biswas et al., 2021) (artículo que utiliza la base de datos CFPW), con la propuesta se obtuvo un 24.19 % con el kernel linear y el trabajo relacionado un porcentaje de 26.00 %. Por otro lado, con el método propuesto se obtuvo un

porcentaje de 65.32 %, mientras que en el trabajo relacionado se obtuvo un 63.24 %, en ambos casos utilizando el kernel Poly-6. Empleando este último kernel (Poly-6), se obtuvo un mejor porcentaje de precisión, esto haciendo al método propuesto comparable con el estado del arte.

# Capítulo 6

## Conclusiones

El objetivo general de este trabajo de tesis fue diseñar un método para la extracción de rasgos faciales, el cual fue comparado con trabajos relacionados reportados hasta el momento, midiendo la precisión y el tiempo de ejecución.

Para el diseño de este método, se comenzó con la revisión de la literatura, enfocándose en los diferentes métodos de extracción de rasgos faciales que utilizan para el proceso de autenticación de personas. Con base en esta revisión, se propuso un conjunto de rasgos faciales, que abarcan diferentes aspectos faciales como: el color de piel, su textura y simetría. Conforme se fueron haciendo los diferentes experimentos, se fue observando que había rasgos faciales que eran sensibles a la iluminación y a la escala, por tal motivo, dentro de las limitaciones o consideraciones al momento de utilizar el método propuesto, es que las imágenes tengan una iluminación equilibrada y que, de ser posible, sean tomadas a la mismadistancia.

Otro punto que también se observó durante la experimentación, fue que conforme se aumentaba el número de personas, el porcentaje de precisión disminuía, esto se debe a que, dentro del conjunto de datos, habrá personas con rasgos faciales similares o que incluso, utilizan accesorios parecidos (lentes, maquillaje, entre otras), por tal motivo, al momento de segmentar el rostro, el clasificador toma algunas imágenes similares como si pertenecieran a la misma persona.

Pasando al análisis de los resultados obtenidos, con la base de datos generada, el método propuesto muestra un buen rendimiento para el caso de 10 personas, con un porcentaje de precisión del 99.11 % utilizando una SVM y el kernel Poly-6, considerando que en la primera etapa de experimentación se tenía un porcentaje de precisión del 47 % con una SVM y un kernel lineal.

Mientras que, los experimentos que se realizaron con la base de datos CFPW, se pasó de

un porcentaje de precisión del 42 % a un 91.66 %, con un kernel linear y un kernel Poly-6 respectivamente. Para ambas bases de datos, los resultados obtenidos fueron favorables. Por otro lado, si se hace una comparación con respecto a los resultados mostrados en (Biswas et al., 2021) (donde se utiliza la base de datos CPFW) y el método propuesto, con el kernel linear, el trabajo relacionado obtuvo un porcentaje de 26.00 % y con el método propuesto se obtuvo un 24.19 %. Ahora, con el kernel Poly-6, en el trabajo relacionado obtuvieron un 63.24 %, mientras que con el método propuesto se obtuvo un porcentaje de 65.32 %. Como se puede observar, cuando se utiliza el kernel Poly-6, se obtiene un mejor porcentaje de precisión y esta acotación es tanto para esta base de datos (CFPW) como para la base de datos generada.

## 6.1. Trabajo a futuro

Una aplicación inicial del método de extracción de rasgos faciales propuesto era ser utilizado en la interacción humano-humanoide, por tiempo, no fue posible llegar a este punto. Por lo tanto, como trabajo a futuro, además de probar el algoritmo con otras bases de datos y con una mayor cantidad de personas, también sería aplicar el algoritmo resultante en un humanoide y ver su funcionamiento y precisión dentro de la interacción humano-humanoide. Al igual que, un punto a tratar sería el reducir el tiempo de ejecución, ya que, a pesar de que no se tarda más de 6 minutos analizando a 10 personas, lo ideal sería que lo hiciera en cuestión de segundos, esto, con el objetivo de su posible aplicación.

## 6.2. Publicaciones

En esta sección se enlistan los trabajos obtenidos a lo largo del desarrollo de este trabajo de tesis.

- Aparicio Arroyo, A. A., Olmos Pineda, I., & Olvera Lopez, J. A. (2019). Extracción de rasgos faciales para la autenticación de personas. En *Tópicos actuales en la Ingeniería del lenguaje y del conocimiento* (pp. 22-28). Montiel & Soriano Editores.
- Aparicio Arroyo, A. A., Olmos Pineda, I., & Olvera López, J. A. (2019). Etapa de detección de rostros para la autenticación de personas. En *Avances en tecnología del lenguaje y el conocimiento* (pp. 77-84). Montiel & Soriano Editores.
- Aparicio Arroyo, A. A., Olmos Pineda, I., & Olvera López, J. A. (2020). Preprocesamiento de imágenes para la autenticación de personas. En *Progresos actuales*

*en la Ingeniería del Lenguaje y del Conocimiento* (pp. 1-10). Montiel & Soriano Editores.

- Aparicio Arroyo, A. A., Olmos Pineda, I., & Olvera López, J. A. (2021). Entrenamiento de SVM para la autenticación de personas. En *Avances de ingeniería del lenguaje, del conocimiento y la interacción humano máquina* (pp. 15-22). United Academic Journals.
- Aparicio Arroyo, A. A., Olmos Pineda, I., & Olvera López, J. A. (2021, 1 diciembre). Autenticación de personas utilizando un clasificador SVM. *Revista Colombiana de Computación* (pp. 48-57).

# Bibliografía

- Almudhahka, N. Y., Nixon, M. S., & Hare, J. S. (2016). Unconstrained human identification using comparative facial soft biometrics. In *Biometrics Theory, Applications and Systems (BTAS), 2016 IEEE 8th International Conference on*, (pp. 1–6). IEEE.
- Amayri, O. & Bouguila, N. (2010). A study of spam filtering using support vector machines. *Artificial Intelligence Review*, 34(1), 73–108.
- Bailey, K. O., Okolica, J. S., & Peterson, G. L. (2014). User identification and authentication using multi-modal behavioral biometrics. *Computers & Security*, 43, 77–89.
- Beymer, D. (1994). Face recognition under varying pose,". In *Proceedings of 23rd Image Understanding Workshop*, volume 2, (pp. 837–842).
- Biswas, R., González-Castro, V., Fidalgo, E., & Alegre, E. (2021). A new perceptual hashing method for verification and identity classification of occluded faces. *Image and Vision Computing*, 113, 104245.
- Bolotnikova, A., Demirel, H., & Anbarjafari, G. (2017). Real-time ensemble based face recognition system for nao humanoids using local binary pattern. *Analog Integrated Circuits and Signal Processing*, 92(3), 467–475.
- Bong, K., Choi, S., Kim, C., Han, D., & Yoo, H.-J. (2018). A low-power convolutional neural network face recognition processor and a cis integrated with always-on face detector. *IEEE Journal of Solid-State Circuits*, 53(1), 115–123.
- Bormann, R., Zwölfer, T., Fischer, J., Hampp, J., & Hägele, M. (2013). Person recognition for service robotics applications. In *Humanoid Robots (Humanoids), 2013 13th IEEE-RAS International Conference on*, (pp. 260–267). IEEE.
- Buill Pozuelo, F., Núñez Andrés, M. A., & Rodríguez, J. J. (2003). Fotogrametría analítica.

- Castrillon, W. A., Alvarez, D. A., & López, A. F. (2008). Técnicas de extracción de características en imágenes para el reconocimiento de expresiones faciales. *Scientia et Technica*, 14(38), 7–12.
- Choi, J. Y., Ro, Y. M., & Plataniotis, K. N. (2012). Color local texture features for color face recognition. *IEEE transactions on image processing*, 21(3), 1366–1380.
- Chowdhury, M., Gao, J., & Islam, R. (2016). Fuzzy rule based approach for face and facial feature extraction in biometric authentication. In *Image and Vision Computing New Zealand (IVCNZ), 2016 International Conference on*, (pp. 1–5). IEEE.
- Dahmane, A., Larabi, S., Bilasco, I. M., & Djeraba, C. (2015). Head pose estimation based on face symmetry analysis. *Signal, Image and Video Processing*, 9(8), 1871–1880.
- De Jong, M. A., Wollstein, A., Ruff, C., Dunaway, D., Hysi, P., Spector, T., Liu, F., Niessen, W., Koudstaal, M. J., Kayser, M., et al. (2015). An automatic 3d facial landmarking algorithm using 2d gabor wavelets. *IEEE Transactions on Image Processing*, 25(2), 580–588.
- Deng, P. & Pei, M. (2008). Multi-view face detection based on adaboost and skin color. In *Intelligent Networks and Intelligent Systems, 2008. ICINIS'08. First International Conference on*, (pp. 457–460). IEEE.
- Developers, S. (2021a). 3.1. Cross-validation: evaluating estimator performance.
- Developers, S. (2021b). Sklearn metrics: precision score.
- Developers, S. (2021c). Visualizing cross-validation behavior in.
- Elizondo, J. E. & Maestre, L. P. (2002). Fundamentos de procesamiento de imágenes. *Documentación Universidad Autónoma de Baja California, Unidad Tijuana*.
- Esqueda, J. & Palafox, L. (2005). Fundamentos de procesamiento de imágenes. *Baja California, México: Universidad Autónoma de Baja California*.
- Farfadi, S. S., Saberian, M. J., & Li, L.-J. (2015). Multi-view face detection using deep convolutional neural networks. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, (pp. 643–650). ACM.
- Gilat, A. (2006). *Matlab: Una introducción con ejemplos prácticos*. Reverté.
- González, R. C. & Woods, R. E. (2012). Digital image processing.

- González, R. C., Woods, R. E., & Eddins, S. L. (2009). *Digital Image Processing Using MATLAB®*. Gatesmark Publishing.
- González-Jiménez, D. & Alba-Castro, J. L. (2007). Toward pose-invariant 2-d face recognition through point distribution models and facial symmetry. *IEEE Transactions on Information Forensics and Security*, 2(3), 413–429.
- Goswami, K., Lee, J.-H., Jang, K.-S., Kim, B.-G., & Kwon, K.-K. (2016). Entropy difference-based early skip detection technique for high-efficiency video coding. *Journal of Real-Time Image Processing*, 12(2), 237–245.
- Grau, J. F. P. (2003). *Técnicas de análisis de imagen: Aplicaciones en Biología*, volume 65. Universitat de València.
- Guo, G., Wen, L., & Yan, S. (2014). Face authentication with makeup changes. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(5), 814–825.
- Jain, A. K., Ross, A. A., & Nandakumar, K. (2011). *Introduction to biometrics*. Springer Science & Business Media.
- Kantharia, K. J. & Prajapati, G. I. (2015). Facial behavior recognition using soft computing techniques: A survey. In *Advanced Computing & Communication Technologies (ACCT), 2015 Fifth International Conference on*, (pp. 30–34). IEEE.
- Karthika, R. & Parameswaran, L. (2016). Study of gabor wavelet for face recognition invariant to pose and orientation. In *Proceedings of the international conference on soft computing systems*, (pp. 501–509). Springer.
- Kaur, J., Upadhyay, S., & Sharma, A. (2017). A video database for intelligent video authentication. In *2017 International Conference on Computing, Communication and Automation (ICCCA)*, (pp. 1081–1085). IEEE.
- Kauser, N. & Sharma, J. (2016). Automatic facial expression recognition: A survey based on feature extraction and classification techniques. In *ICT in Business Industry & Government (ICTBIG), International Conference on*, (pp. 1–4). IEEE.
- Kazemi, V. & Sullivan, J. (2014). One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, (pp. 1867–1874).

- Kim, D., Lee, J., Yoon, H.-S., & Cha, E.-Y. (2007). A non-cooperative user authentication system in robot environments. *IEEE Transactions on Consumer electronics*, 53(2).
- Kim, D.-H., Lee, J.-Y., Cha, E.-Y., & Cho, Y.-J. (2005). Face identification using multiple combination strategy for human robot interaction. *IFAC Proceedings Volumes*, 38(1), 1053–1058.
- Kim, Y., Yoo, J.-H., & Choi, K. (2011). A motion and similarity-based fake detection method for biometric face recognition systems. *IEEE Transactions on Consumer Electronics*, 57(2).
- Kohonen, T., Schroeder, M., Huang, T., & Maps, S.-O. (2001). Springer-verlag new york. *Inc., Secaucus, NJ*, 43, 2.
- Kong, S. G. & Mbouna, R. O. (2015). Head pose estimation from a 2d face image using 3d face morphing with depth parameters. *IEEE Transactions on Image Processing*, 24(6), 1801–1808.
- Kumar, S., Singh, S., & Kumar, J. (2017). A study on face recognition techniques with age and gender classification. In *IEEE International Conference on Computing, Communication and Automation (ICCCA)*, 5th-6th May.
- Kumar Shukla, R., Das, D., & Agarwal, A. (2016). A novel method for identification and performance improvement of blurred and noisy images using modified facial deblur inference (fadein) algorithms. In *Electrical, Electronics and Computer Science (SCEECS), 2016 IEEE Students' Conference on*, (pp. 1–7). IEEE.
- Le, V., Brandt, J., Lin, Z., Bourdev, L., & Huang, T. S. (2012). Interactive facial feature localization. In *European conference on computer vision*, (pp. 679–692). Springer.
- Lekdioui, K., Messoussi, R., Ruichek, Y., Chaabi, Y., & Touahni, R. (2017). Facial decomposition for expression recognition using texture/shape descriptors and svm classifier. *Signal Processing: Image Communication*, 58, 300–312.
- Li, J., Zhao, F., Feng, J., Roy, S., Yan, S., & Sim, T. (2018). Landmark free face attribute prediction. *IEEE Transactions on Image Processing*, 27(9), 4651–4662.
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *European conference on computer vision*, (pp. 740–755). Springer.

- Liu, Z., Wu, M., Cao, W., Chen, L., Xu, J., Zhang, R., Zhou, M., & Mao, J. (2017). A facial expression emotion recognition based human-robot interaction system. *IEEE/CAA Journal of Automatica Sinica*, 4(4), 668–676.
- López, R. F. & Fernandez, J. M. F. (2008). *Las redes neuronales artificiales*. Netbiblo.
- McAndrew, A. (2004). An introduction to digital image processing with matlab notes for scm2511 image processing. *School of Computer Science and Mathematics, Victoria University of Technology*, 264.
- Mehl, H. & Peinado, O. (1997). Fundamentos del procesamiento digital de imágenes. *Aplicación de la teledetección y de los sistemas de información geográfica en la gestión de recursos naturales. Parte I: Fundamentos teóricos y prácticos*.
- Moeini, A. & Moeini, H. (2015). Real-world and rapid face recognition toward pose and expression variations via feature library matrix. *IEEE Transactions on Information Forensics and security*, 10(5), 969–984.
- Murphy-Chutorian, E. & Trivedi, M. M. (2008). Head pose estimation in computer vision: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 31(4), 607–626.
- Nair, B. M., Foytik, J., Tompkins, R., Diskin, Y., Aspiras, T., & Asari, V. (2011). Multi-pose face recognition and tracking system. *Procedia Computer Science*, 6, 381–386.
- O’Shea, K. & Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.
- Ou, J. (2012). Classification algorithms research on facial expression recognition. *Physics Procedia*, 25, 1241–1244.
- Pajares Martinsanz, G. & de la Cruz García, J. (2002). *Visión por computador: Imágenes digitales y aplicaciones*.
- Pang, S., Kim, D., & Bang, S. Y. (2002). Membership authentication in dynamic face groups. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 2, (pp. 374–377). IEEE.
- Park, C.-W. & Lee, T. (2010). A robust facial feature detection on mobile robot platform. *Machine Vision and Applications*, 21(6), 981–988.
- Passariello, G. (1995). *Imágenes médicas. Adquisición, Análisis*. Equinoccio.

- Patel, T. & Shah, B. (2017). A survey on facial feature extraction techniques for automatic face annotation. In *Innovative Mechanisms for Industry Applications (ICIMIA), 2017 International Conference on*, (pp. 224–228). IEEE.
- Presutti, M. (2004). La matriz de co-ocurrencia en la clasificación multispectral: tutorial para la enseñanza de medidas texturales en cursos de grado universitario. *4ª Jornada de Educação em Sensoriamento Remoto no Âmbito do Mercosul*.
- Python Software Foundation (18 de Noviembre de 2021). <https://docs.python.org/es/3/library/statistics.html>.
- Rikhtegar, A., Pooyan, M., & Manzuri-Shalmani, M. T. (2016). Genetic algorithm-optimised structure of convolutional neural network for face recognition applications. *IET Computer Vision*, 10(6), 559–566.
- Rodriguez, A. (2016). El álgebra lineal y el procesamiento digital de imágenes. Parte III. Transformaciones afines.
- Rowley, H. A., Baluja, S., & Kanade, T. (1998). Rotation invariant neural network-based face detection. In *Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No. 98CB36231)*, (pp. 38–44). IEEE.
- Sagonas, C., Tzimiropoulos, G., Zafeiriou, S., & Pantic, M. (2013). 300 faces in-the-wild challenge: The first facial landmark localization challenge. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, (pp. 397–403).
- Samad, M. D. & Iftekharuddin, K. M. (2016). Frenet frame-based generalized space curve representation for pose-invariant classification and recognition of 3-d face. *IEEE Transactions on Human-Machine Systems*, 46(4), 522–533.
- Samangouei, P., Patel, V. M., & Chellappa, R. (2017). Facial attributes for active authentication on mobile devices. *Image and Vision Computing*, 58, 181–192.
- Soares, J. V., Leandro, J. J., Cesar, R. M., Jelinek, H. F., & Cree, M. J. (2006). Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification. *IEEE Transactions on medical Imaging*, 25(9), 1214–1222.
- Sotolongo-Aguilar, G. & Guzmán-Sánchez, M. V. (2001). Aplicaciones de las redes neuronales. el caso de la bibliometría. *Ciencias de la Información*, 32(1), 27–34.

- Tao, Q. & Veldhuis, R. (2010). Biometric authentication system on mobile personal devices. *IEEE Transactions on Instrumentation and Measurement*, 59(4), 763–773.
- Tong, Y., Chen, R., & Cheng, Y. (2014). Facial expression recognition algorithm using lgc based on horizontal and diagonal prior principle. *Optik-International Journal for Light and Electron Optics*, 125(16), 4186–4189.
- Travieso, C. M., Alonso, J. B., & Ferrer, M. A. (2007). Arbitrary illumination conditions for facial identification. In *Security Technology, 2007 41st Annual IEEE International Carnahan Conference on*, (pp. 93–98). IEEE.
- Tsai, P., Tran, T. P., & Cao, L. (2009). Expression-invariant facial identification. In *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*, (pp. 5151–5155). IEEE.
- Viola, P. & Jones, M. (2001). Robust real-time object detection. *International journal of computer vision*, 4(34-47), 4.
- Vipparthi, S. K., Murala, S., Gonde, A. B., & Wu, Q. J. (2016). Local directional mask maximum edge patterns for image retrieval and face recognition. *IET Computer Vision*, 10(3), 182–192.
- Wagner, G. M. & Sinzinger, E. D. (2008). Cascading trilinear tensors for face authentication. In *Applications of Computer Vision, 2008. WACV 2008. IEEE Workshop on*, (pp. 1–6). IEEE.
- Wiskott, L., Fellous, J.-M., Krüger, N., & Von Der Malsburg, C. (1997). Face recognition by elastic bunch graph matching. In *International Conference on Computer Analysis of Images and Patterns*, (pp. 456–463). Springer.
- Wu, Y. & Ji, Q. (2019). Facial landmark detection: A literature survey. *International Journal of Computer Vision*, 127(2), 115–142.
- Wu, Z., Yuan, J., Zhang, J., & Huang, H. (2016). A hierarchical face recognition algorithm based on humanoid nonlinear least-squares computation. *Journal of Ambient Intelligence and Humanized Computing*, 7(2), 229–238.
- Xi, D. & Lee, S.-W. (2002). Face detection and facial feature extraction using support vector machines. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 4, (pp. 209–212). IEEE.

- 
- Zhan, S. & Ando, S. (2006). Facial authentication system based on real-time 3d facial imaging by using correlation image sensor. In *Intelligent Systems Design and Applications, 2006. ISDA'06. Sixth International Conference on*, volume 2, (pp. 396–400). IEEE.
- Zheng, H., Geng, X., Tao, D., & Jin, Z. (2016). A multi-task model for simultaneous face identification and facial expression recognition. *Neurocomputing*, *171*, 515–523.