



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO  
INSTITUTO DE CIENCIAS BÁSICAS E INGENIERÍA  
ÁREA ACADÉMICA DE COMPUTACIÓN Y ELECTRÓNICA  
LICENCIATURA EN CIENCIAS COMPUTACIONALES



---

# DETECCIÓN DE ROSTROS CON REDES NEURONALES ARTIFICIALES, INVARIANTES DE HU Y OTRAS TÉCNICAS

---

*Tesis que para obtener el grado de Licenciada en Ciencias Computacionales*

*PRESENTA*

**Arisbeth Mejía Mejía**

*Asesor de tesis:*

**Dr. Joel Suárez Cansino**

*Coasesor de tesis:*

**Dr. Vigilio López Morales**

Mineral de la Reforma, Hidalgo, México, Jueves 30 de Noviembre de 2023





UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO  
 Instituto de Ciencias Básicas e Ingeniería  
*School of Engineering and Basic Sciences*

Mineral de la Reforma, Hgo., a 30 de noviembre de 2023

Número de control: ICBI-D/1850/2023  
 Asunto: Autorización de impresión.

**MTRA. OJUKY DEL ROCÍO ISLAS MALDONADO**  
**DIRECTORA DE ADMINISTRACIÓN ESCOLAR DE LA UAEH**

Con fundamento en lo dispuesto en el Título Tercero, Capítulo I, Artículo 18 Fracción IV; Título Quinto, Capítulo II, Capítulo V, Artículo 51 Fracción IX del Estatuto General de nuestra Institución, por este medio le comunico que el Jurado asignado a la Pasante de la Licenciatura en Ciencias Computacionales **Arisbeth Mejía Mejía**, quien presenta el trabajo de titulación "**Detección de rostros con redes neuronales artificiales invariantes de hu y otras técnicas**", después de revisar el trabajo en reunión de Sinodales ha decidido autorizar la impresión del mismo, hechas las correcciones que fueron acordadas.

A continuación, firman de conformidad los integrantes del Jurado:

**Presidente** Dr. Virgilio López Morales

**Secretario:** Dra. Anilú Franco Arcega

**Vocal:** Dr. Joel Suárez Cansino

**Suplente:** Dr. Edgar Olguín Guzmán

Sin otro particular por el momento, reciba un cordial saludo.

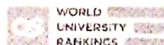
Atentamente  
 "Amor, Orden y Progreso"

Dr. Otilio Arturo Acevedo Santolucito  
 Director del ICBI



OAAS/YCC

Ciudad del Conocimiento  
 Carretera Pachuca-Tulancingo km 4.5 Colonia  
 Carboneras, Mineral de la Reforma, Hidalgo,  
 México. C.P. 42184  
 Teléfono: 771 71 720 00 ext. 2231 Fax 2109  
 direccion\_icbi@uaeh.edu.mx



*«Nunca consideres el estudio como una obligación, sino como una oportunidad para penetrar en el bello y maravilloso mundo del saber.»*

Albert Einstein

*«... Para concluir, les pido por favor que apaguen las cámaras... Me gustaría decir lo siguiente... Comprendemos que ciertos decretos... Quiero decirlo con respecto a este caso, pero se aplica a todo. Es un trabajo difícil, ¿ustedes creen que no lo sé? ¡Claro que lo sé! Los problemas son difíciles de solucionar pero son solucionables. Pero nunca los resolveremos si trabajamos así. No es casualidad que nos encontremos hoy en este momento, no es casualidad que haya pedido esta reunión. Porque como les mencioné, la calidad de nuestro trabajo es simplemente desagradable. Hacen todo en la superficie... Si continuamos así, ¡nunca vamos a lograr un resultado! Si comenzamos a trabajar en una manera más profesional con el entendimiento de que “debemos lograr esto”, entonces lo lograremos. Si no podemos lograr estos resultados, debemos llegar a la conclusión que yo o ustedes no somos buenos en este trabajo... Si son ustedes, entonces deben renunciar a sus posiciones... déjenme remarcarles algo hoy. Creo que es lo último. Pienso que ustedes comprenden esto. Solo para no tener mal entendidos y que hablamos de forma sincera.»*

Vladimir Putin

*...Y entonces se le impone el deber de no limitar su acción inmediata y cierta al pequeño número de hombres que la casualidad le acerca, sino de extenderla a la humanidad, un fin de que pueda alcanzar las excepciones, los mejores, los elegidos. Pero el único órgano por el que se dirige a la humanidad es la escritura... Y esto se efectúa no por la tradición, falsificada paso a paso, sino únicamente por la escritura, la sola conservadora fiel de las ideas. Además, cada espíritu que piensa con profundidad, tiene, necesariamente la tendencia, para su propia satisfacción de retener sus pensamientos y llevarlos a la mayor claridad y concreción posibles, es decir, a encarnarlos en palabras. Pero esto no se obtiene en plenitud sino a través de la escritura. La palabra escrita es, en efecto, esencialmente distinta a la oral, porque ella sola permite la mayor precisión y concisión, la brevedad por excelencia, lo que hace de ella el puro ektypos del pensamiento. De acuerdo con todo lo que acabamos de decir, sería una petulancia rara en un pensador no querer aprovechar la invención más importante del género humano. Por esta razón me es difícil creer en la inteligencia realmente grande de quienes no han escrito. Más bien me siento inclinado a tenerlos por héroes sobre todo prácticos, que han actuado más por su carácter que por su cerebro...*

Arthur Schopenhauer

*«La irracionalidad y la confusión predominan en el mundo, dentro del cual pocos logran percibir algunos oasis de claridad que les den la oportunidad de crear microcosmos en los que, por definición, están ausentes la ansiedad y la angustia.»*

Joel Suárez Cansino

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO

## *Resumen*

Instituto de Ciencias Básicas e Ingeniería  
Área Académica de Computación y Electrónica

Licenciatura en Ciencias Computacionales

### **DETECCIÓN DE ROSTROS CON REDES NEURONALES ARTIFICIALES, INVARIANTES DE HU Y OTRAS TÉCNICAS**

por  
Arisbeth Mejía Mejía

La detección de rostro en imágenes o en video es un problema que ha sido estudiado por diferentes autores desde hace varios años. El algoritmo de Viola–Jones es uno que ha tenido bastante éxito en parte de la solución de este problema, sin contar las incursiones recientes que se han tenido dentro del campo de las Redes Neuronales Artificiales con aprendizaje profundo. Sin embargo, en esta tesis se llevó a cabo el análisis acerca de la eficiencia que Redes Neuronales Artificiales, combinadas con Filtros de Gabor o bien con Invariantes de Hu, pueden tener en la detección de rostro. Los resultados experimentales muestran que el algoritmo de Viola–Jones proporciona mejores resultados y que, posiblemente, bajo un análisis más profundo para buscar la arquitectura óptima de las Redes Neuronales Artificiales, se logre mejorar la eficiencia de éstas para resolver el problema de detección.

#### *Abstract*

Face detection in images or videos is a problem that has been studied by different authors for several years. The Viola-Jones algorithm has been quite successful in solving this problem, not counting recent incursions that have been made within the field of Artificial Neural Networks with deep learning. However, this thesis carried out an analysis of the efficiency that Artificial Neural Networks, combined with Gabor Filters or Hu Invariants, can have in face detection. Experimental results show that the Viola-Jones algorithm provides better results and that, possibly, under a deeper analysis to find the optimal architecture of Artificial Neural Networks, their efficiency can be improved to solve the detection problem.



## *Agradecimientos*

Dr. Joel Suárez Cansino y Dr. Virgilio López Morales, sin ustedes y sus virtudes, su paciencia y constancia este trabajo no lo hubiese logrado tan fácil, sus consejos fueron siempre útiles cuando no salían de mi pensamiento las ideas para escribir lo que hoy he logrado.

Ustedes formaron parte importante de esta historia con sus aportes profesionales que los caracterizan, muchas gracias por sus múltiples palabras de aliento, cuando más las necesite; por estar allí cuando mis horas de trabajo se hacían confusas, gracias por sus orientaciones.





# Contenido

<b>Agradecimientos</b>	<b>VII</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Planteamiento del problema	2
1.1.1. Problemática	2
1.1.2. Originalidad del problema	2
1.1.3. Impacto	2
1.2. Hipótesis	3
1.3. Objetivos	3
1.3.1. General	3
1.3.2. Objetivos específicos	3
1.4. Metodología	5
1.5. Estructura del documento	5
<b>2. Marco Teórico</b>	<b>7</b>
2.1. Imágenes a color y tonos de gris	7
2.2. Detección, Reconocimiento y Análisis de Rostros	8
2.3. Criterios Falso-Positivo, Falso-Negativo, Verdadero-Positivo, Verdadero-Negativo	9
2.4. Viola-Jones	10
2.4.1. Características de Haar	10
2.4.2. Imagen integral	11
2.4.3. El algoritmo de impulso (AdaBoost)	11
2.4.4. Arquitectura en cascada	12
2.5. Redes Neuronales Artificiales	13
2.6. Invariantes de Hu	13
2.7. Componentes Conexas	14
2.8. Filtro de Gabor	15
<b>3. Estado del Arte</b>	<b>17</b>
<b>4. Desarrollo Experimental</b>	<b>25</b>
4.1. Herramientas utilizadas	25
4.1.1. Software	25
4.1.2. Hardware	25
4.2. Algoritmos	26
4.3. Experimento 1: Viola-Jones versus RNA-Filtro de Gabor	28
4.4. Experimento 2: Invariantes de Hu antecediendo a redes neuronales artificiales para detectar rostros según los invariantes	30
4.5. Experimento 3: Viola-Jones versus RNA-Filtro de Gabor versus Invariantes de Hu con tratamiento de imagen previo usando Igualización de histogramas adaptable (adapthisteq), igualización de histograma (histeq) y ajuste de contraste (imajust)	34

<b>5. Conclusiones y Trabajo Futuro</b>	<b>37</b>
<b>Bibliografía</b>	<b>41</b>

# Lista de Figuras

2.1. Ejemplo de las características similares a Haar (Skelly y Latifi, 2019). . . . .	10
2.2. Ejemplo de una imagen integral (Skelly y Latifi, 2019). . . . .	11
2.3. Estructura específica de una red neuronal artificial . . . . .	13
4.1. Diagrama de bloques del algoritmo Viola–Jones . . . . .	26
4.2. Diagrama de bloques de RNA con filtro de Gabor . . . . .	26
4.3. Diagrama de bloques de RNA con Invariantes de Hu. . . . .	27
4.4. Diagrama de flujo de la iterfaz de la detección de rostros . . . . .	27
4.5. Número de rostros detectados por Viola-Jones y RNA. . . . .	29
4.6. Interfaz gráfica utilizando el algoritmo de Viola-Jones . . . . .	30
4.7. Interfaz gráfica utilizando la Red Neuronal Artificial . . . . .	30
4.8. Interfaz gráfica utilizando el algoritmo de Invariantes de Hu. . . . .	32
4.9. Mejores métodos de entrenamiento de una red neuronal artificial cuando se combinan con Invariantes de Hu. . . . .	34
4.10. Comparación de algoritmos empleados para la detección de rostros. . . . .	35
4.11. Número de Falso–Positivo en relación con de Viola-Jones, RNA e Invariantes de Hu. . . . .	36
5.1. Distribución de resultados posibles en la detección de rostro y su posible impacto en el cálculo de la eficiencia de un algoritmo de detección de acuerdo a lo que se comenta en el documento. . . . .	38



# Lista de Tablas

1.1. Diferentes técnicas empleadas para detectar, reconocer y analizar rostros. . .	1
4.1. Resultados de la prueba número 1. . . . .	29
4.2. Técnica Adaphisteq con 7 10 90 1 neuronas . . . . .	32
4.3. Técnica Adaphisteq con 7 10 50 1 neuronas . . . . .	33
4.4. Técnica Histeq con 7 10 90 1 neuronas . . . . .	33
4.5. Técnica Histeq con 7 10 50 1 neuronas . . . . .	33
4.6. Técnica Imagejust con 7 10 90 1 neuronas . . . . .	33
4.7. Técnica Imagejust con 7 10 50 1 neuronas . . . . .	33
4.8. Resultados del número de rostros detectados . . . . .	34
4.9. Resultados de los Falso-Positivo para cada uno de los tres algoritmos anali- zados. . . . .	35



# Lista de Acrónimos

**RNA** – Red Neuronal Artificial  
**RNAs** – Redes Neuronales Artificiales  
**RGB** – Red Green Blue  
**MLP** – Multi- Layer Perceptron  
**FFL** – Feed Forward Learning  
**GWT** – Gabor Wavelet Transform  
**CNN** – Convolutional Neural Network  
**CFNN** – Conventional Feedforward Neural Network  
**HNN** – Hybrid Neural Network  
**PNN** – Polynomial Neural Network  
**LM** – Levenberg- Marquardt  
**BFG** – BFGS Quasi Newton  
**RP** – Resilient Backpropagation  
**SCG** – Scaled Conjugate Gradient  
**CGB** – Conjugate Gradient with Powell/ Beale Restarts  
**CGF** – Fletcher- Powell Conjugate Gradient  
**CGP** – Polak- Ribière Conjugate Gradient  
**OSS** – One- Step Secant  
**GDX** – Variable Learning Rate Backpropagation  
**GD** – Basic Gradient Descent  
**GDM** – Gradient Descent with Momentum  
**BR** – Bayesian Regularization  
**GPU** – Graphics Processing Unit  
**CPU** – Central Processing Unit  
**LSTM** – Long Short Term Memory  
**TN** – True Negative  
**TP** – True Positive  
**FP** – False Positive  
**FN** – False Negative  
**NRD** – Número de Rostros Detectados  
**NRF** – Número de Rostros Falsos





# Lista de Símbolos

$\phi$  – Momento

$\mu_{pq}$  – Invariante a la traslación y varición en la escala de una imagen

$\sigma$  – Varianza de la función gaussiana

$\lambda$  – Ancho de banda de la función sinusoidal

$\theta$  – Orientación de la normal

$\gamma$  – Elipicidad de la función de Gabor



*Dedicada a mi familia quien siempre me apoyo y aliento a seguir  
adelante para realizar mis metas.*



## Capítulo 1

# Introducción

La autenticación biométrica es un área que incluye diferentes disciplinas, entre las que se encuentra el procesamiento de imágenes. En particular, existen tres problemáticas referentes al tratamiento biométrico del rostro humano (Fernández, 2014; Bouzalmat et al., 2011), la primera de ellas se relaciona con la detección del rostro y se refiere al hecho de poder encontrar la ubicación del mismo en un ambiente real, en una imagen o en un video que contiene a una persona.

La segunda de estas problemáticas considera el reconocimiento de rostros previamente detectado, y esto se refiere a no sólo detectar la ubicación de éste obviamente, sino también poder decir a qué persona le corresponde el rostro identificado. El tercer y último problema consiste en el análisis facial, que está enfocado a reconocer los estados de ánimo en que se encuentra la persona, e incluso proponer una aproximación de su edad, entre otros datos biológicos.

Este trabajo aborda al primer problema que, a pesar de ser la primera etapa para llegar a una de las otras dos opciones, no significa que este sea un problema sencillo de resolver o que ya esté resuelto en su totalidad, ya que aún existen muchos aspectos que se deben trabajar, lo cual explica la presencia de diferentes técnicas y sus correspondientes algoritmos, algunos de los cuales son mostrados por la Tabla 1.1, en donde se aprecian algunas de las técnicas existentes y su posible aplicación en la detección, reconocimiento y/o análisis del rostro.

TABLA 1.1: Diferentes técnicas empleadas para detectar, reconocer y analizar rostros.

Técnicas	Áreas		
	Detección	Reconocimiento	Análisis
PCA	X	X	X
RNA	AdaBoost Viola–Jones Feedforward Invariantes de Hu	MLP y FFL GWT y FNN CNN CFNN HNN PNN	CNN
Probabilísticas	X	X	X
Árboles de Decisión	X	X	X
Eigenrostros	X	X	X

PCA	=	Principal Component Analysis
RNA	=	Redes Neuronales Artificiales
MLP	=	Multilayer Perceptron
FFL	=	Feedforward Learning
GWT	=	Gabor Wavelet Transform
FNN	=	Feedforward Neural Network
CNN	=	Convolutional Neural Network
CFNN	=	Conventional Feedforward Neural Network
HNN	=	Hybrid Neural Network
PNN	=	Polynomial Neural Network

Adicionalmente, debe decirse que algunas de las problemáticas que se atacan están relacionadas con el cambio de apariencia, cambio de iluminación, expresión y ocultamiento facial, cambio de posición, rotación y escala, algoritmos poco adecuados por su velocidad, y presencia de varios rostros (existe un límite). Finalmente, las técnicas se han aplicado en colores RGB y en escala de grises.

En particular, para la detección de rostros resulta interesante estudiar la técnica de redes neuronales artificiales y su posible combinación con métodos de tratamiento de imágenes tales como AdaBoost, Viola–Jones e invariantes de Hu.

## **1.1. Planteamiento del problema**

El enunciado del problema se divide en tres partes. La primera de ellas bosqueja brevemente el contexto en el cual se ubica la problemática, tratando de explicar incluso alguna terminología necesaria para el planteamiento.

La segunda parte es el planteamiento del problema en sí, en términos de una proposición que puede ser incluso compuesta, pero que contiene la presencia de un obstáculo que debe ser removido.

Finalmente, la tercera parte establece el impacto que se tiene una vez que se ha encontrado la solución al problema planteado.

### **1.1.1. Problemática**

Redes Neuronales Artificiales han venido dando buenos resultados cuando se aplican al problema de detección de rostros, cuyas soluciones se sabe tienen grandes aplicaciones en cuestiones de seguridad, domótica, entre otras posibles áreas de interés.

Sin embargo, existen algunas dificultades que requieren propuestas e implantación de soluciones. La versión que utiliza aprendizaje supervisado generalmente requiere de un tamaño grande de las muestras para entrenamiento, validación y prueba, ocasionando que el tiempo que se emplea para entrenar a la red sea extremadamente grande.

### **1.1.2. Originalidad del problema**

Una de las características en la detección de rostro es que la eficiencia de los algoritmos de detección sufre una disminución cuando aumenta o disminuye la resolución de la imagen. Adicionalmente, para ciertos algoritmos, la eficiencia en la detección se deteriora cuando aparece una imagen con demasiada iluminación, o bien contiene oclusiones o es adverso el entorno en el cual se tomó la fotografía; adverso en el sentido de que existen malas condiciones climáticas o ambientales.

En este sentido, no se tiene una base de comparación de la eficiencia de algoritmos, tales como Redes Neuronales Artificiales con Viola – Jones y combinaciones interesantes con Invariantes de Hu. En este sentido, interesa mucho analizar la eficiencia en la detección de rostros que presentan estas técnicas y algunas de sus combinaciones.

### **1.1.3. Impacto**

Existen muchas aplicaciones en el mundo real en las que una mejoría en la detección de rostros tendría enormes implicaciones en la calidad del reconocimiento y del análisis de rostros. Por ejemplo, en el problema de identificación de usuarios de cajeros automáticos, videovigilancia, en el problema de identificación de estudiantes que harán un examen en línea (un problema que interesa resolver en Evaluación Basada en Computadoras o en Evaluación Adaptable Computarizada), etc.

## 1.2. Hipótesis

Si la alta resolución de una imagen hace que el recorrido necesario para detectar rostros sea lento, mientras que la baja resolución no contiene el suficiente número de píxeles con los cuales definir apropiadamente el rostro; además, demasiada iluminación hace también que se pierda información acerca del rostro y lo mismo sucede cuando aparecen oclusiones en la imagen, o cuando el entorno en el cual se tomó ésta es adverso (se tomó bajo la lluvia, estuvieron presentes tolvaneras, etc.), y se construyen algoritmos para detección de rostros basándose en combinaciones de técnicas de Redes Neuronales Artificiales, Viola-Jones e Invariantes de Hu, entonces la detección de rostros se hará más eficiente usando solamente el algoritmo de Viola-Jones o bien al combinar Redes Neuronales Artificiales con Invariantes de Hu, o bien con Filtro de Gabor.

## 1.3. Objetivos

Los objetivos de este trabajo se dividen en un objetivo general y en objetivos específicos orientados a lograr el objetivo general, y todos ellos se enuncian como sigue,

### 1.3.1. General

Evaluar y comparar la eficiencia de las técnicas basadas en Redes Neuronales Artificiales combinadas con Invariantes de Hu o con Filtro de Gabor, y los resultados que aporta Viola-Jones, en la detección de rostros humanos. Se busca determinar cuál de estos métodos es el más óptimo. Para ello, se utilizará una interfaz que muestra la imagen seleccionada (de un total de 432 imágenes) y las características que contiene cada una.

### 1.3.2. Objetivos específicos

1. Crear un banco de imágenes, haciendo uso de bancos ya existentes o bien considerando uno propio que contiene fotografías de personas en grupo, o de individuos solos, y empleando un manejador de bases de datos conveniente y relativamente simple, el cual acepte conexión con MatLab, para utilizarlo durante la prueba, validación y, en su caso, entrenamiento de los modelos de detección de rostros que se empleen.
2. Seleccionar los algoritmos adecuados de detección de rostro, tomando en cuenta las ventajas y desventajas de ellos reportadas en la literatura existente sobre el tema y proponiendo modificaciones propias en los mismos, para obtener el mejor funcionamiento posible durante el proceso de detección de rostro.
3. Diseñar la parte experimental del trabajo que se propone, considerando las diferentes variables dependientes e independientes que afectan el óptimo desempeño de las técnicas involucradas, para contar con elementos de referencia con los cuales sea posible llevar a cabo una comparación de resultados.
4. Seleccionar las librerías de MatLab apropiadas para la extracción de características y detección de rostros. Es importante fijarse en el tipo de técnicas que emplean estas librerías y verificar si coinciden con las ya propuestas. Esto permitirá evitar desarrollar desde cero toda la plataforma propuesta. Finalmente, se podrán conectar estas librerías a los algoritmos para lograr la detección de rostros en una imagen.
5. Realizar el experimento con el prototipo de plataforma logrado hasta este punto, siguiendo los pasos especificados en el diseño experimental, para obtener resultados



relacionados con el tiempo de detección de rostros en sus partes de preprocesamiento, procesamiento y posprocesamiento.

6. Llevar a cabo un análisis de la eficiencia en la detección de rostros usando los conceptos de falso-positivo y falso – negativo y ajustar los resultados obtenidos, usando técnicas de estadística descriptiva bastante conocidas y repitiendo experimentalmente aquellos resultados que muestren comportamiento poco aceptable. De esta manera, se contará con mejor información para una buena toma de decisiones.

## 1.4. Metodología

En este trabajo se utilizará la metodología benchmarking; la cual nos permite medir la eficiencia de los algoritmos. Primero se determinó hacer un estudio de lo que implica la detección de rostros, por lo cual el primer paso fue buscar referencias para determinar el software y lenguaje que se empleó en el detector de rostros, posteriormente se determinaron los algoritmos utilizados, teniendo como objetivo conocer cuál de ellos es más efectivo.

Se diseñó la interfaz del detector usando el ambiente de desarrollo de interfaces de Matlab y empleando el algoritmo de Viola–Jones implementado en esta misma plataforma, además de usar el paquete de Redes Neuronales Artificiales del mismo entorno Matlab, implementando en conjunto con lo anterior los algoritmos relacionados con los Invariantes de Hu y el Filtro de Gabor, para posteriormente examinar los factores que afectan y favorecen la detección de rostros, y finalmente determinar cuál de estos es más eficiente.

De acuerdo al análisis de la información recabada en trabajos y artículos de investigación referente al tema, se determinaron los factores que intervienen en la detección de rostros los cuales son: la iluminación, oclusión (obstrucción del rostro), el entorno (si llueve o hay humo), la distancia, la resolución de la imagen y si es a color o escala de grises.

Para las primeras pruebas, se creó un banco de imágenes, siendo éstas de un equipo de fútbol, a color y escala de grises de diferente resolución. De acuerdo a los primeros resultados obtenidos, se procedió a realizar nuevas pruebas, manejando imágenes de 6 personas siendo 2 mujeres y los demás hombres, en donde se manejó la combinación de los factores que intervienen en la detección de rostros.

Los resultados obtenidos se muestran en una tabla y se graficaron, logrando así ver la eficiencia de los algoritmos empleados y deduciendo de qué manera se puede mejorar el algoritmo para tener resultados más óptimos, hasta poder llegar al grado de minimizar alguna deficiencia.

## 1.5. Estructura del documento

Además de esta introducción, el documento se estructura en otras cuatro secciones, las cuales se comentan a continuación. La Sección 2, denominada Marco Teórico, proporciona los argumentos teóricos que fundamentan el proyecto. La Sección 3, conocida como Estado del Arte, muestra una revisión crítica propositiva respecto a los trabajos relacionados con el tema.

La Sección 4, llamada Desarrollo Experimental y Resultados, muestra la interfaz utilizada, así como las características tomadas en cuenta para determinar la eficiencia de los algoritmos empleados. Finalmente, la Sección 5, que lleva por nombre Conclusiones y Trabajo Futuro, muestra tablas y gráficas de los resultados obtenidos, asimismo comenta cuál de los algoritmos empleados tiene una mejor eficiencia de acuerdo a los factores tomados en cuenta, estipulando además los trabajos a futuro.



## Capítulo 2

# Marco Teórico

Es importante resaltar que los elementos básicos que comprenden el estudio realizado en este trabajo parten de la suposición de que se analizan imágenes digitales en cuyo campo aparecen conceptos tales como imágenes a color y tonos de grises. Aunado a esto, y dependiendo de si la imagen es a color o en tonos de grises, se emplean diferentes algoritmos para el procesamiento de la imagen. Estos son algunos de los temas que se tratan en este capítulo.

### 2.1. Imágenes a color y tonos de gris

De acuerdo a la referencia (Alcalá y Navarro, 2008) La imagen de mapa de bits, también llamada bitmap o gráfico rasterizado, consiste en una estructura que representa una rejilla rectangular de píxeles o puntos de color. Esta rejilla, o ráster, representa una imagen en cualquier soporte indicado para la representación (monitor, pantalla, etc.). La representación de una imagen cualquiera se obtiene a través de la asignación y almacenamiento de un único color en cada uno de los píxeles.

El color de cada píxel, en el caso que la imagen esté en modo RGB, se definirá por asignación de un valor para cada uno de los tres colores del modo, esto es, un valor para el rojo, otro para el verde y otro para el azul.

La imagen bitmap, a diferencia de la imagen vectorial, no puede escalarse sin consecuencias que alteren su aspecto. A la hora de escalar, tanto a mayor como a menor escala, hay que tener en cuenta muchos factores, entre los cuales se encuentra la resolución, el modo de color, la profundidad de bits y el formato de compresión.

Los modos de color son representaciones del espectro visible, que se pueden considerar como el contenedor en el que se coloca la información sobre cada píxel de una imagen, esto permite describir el color de cada píxel utilizando valores numéricos. En este trabajo se utilizaron dos modos de color, los cuales se describen a continuación, basándose en la referencia (Alcalá y Navarro, 2008).

#### 1. Escala de grises

Este modo de color de la imagen está constituida por píxeles que pueden adoptar distintas tonalidades de un mismo color, por ejemplo, desde el blanco (0% de negro) hasta el negro (100% de negro). Este espectro de tono se gradúa, normalmente en una escala que tiene 256 niveles. Esto hace que las imágenes en escala de grises sean las más apropiadas para la impresión de imágenes en blanco y negro tramadas.

#### 2. RGB

RGB es el acrónimo de Red (rojo), Green (verde) y Blue (azul), que también puede encontrarse representado por las siglas RVA. El RGB es el modo de color más usual para las imágenes bitmap que han de verse en pantallas, monitores y otros soportes electrónicos.

En modo RGB, el color de cada pixel está compuesto por una determinada cantidad de rojo, verde y azul. A la combinación de determinadas cantidades de luz roja, verde y azul se le conoce como mezcla aditiva de colores.

En la mezcla aditiva de colores de luz a su máxima intensidad, con la suma de rojo + verde se obtiene amarillo, de la suma de verde + azul el color cian, de la suma de azul +rojo el magenta, y de la suma de los tres colores primarios da como resultado el color blanco.

Por lo tanto, combinando los colores primarios con los resultantes de sus combinaciones se pueden obtener todos los colores que se aprecian en una pantalla. Cada color RGB (rojo, verde o azul) tiene una intensidad que se mide en una escala de 0 a 255, siendo 0 la mínima intensidad de color y 255 la máxima.

Los colores en RGB se expresan del siguiente modo: un paréntesis con tres cifras separadas por comas. Cada cifra corresponde a la intensidad de cada color. La primera cifra corresponde al color rojo, la segunda al verde y la tercera al azul, de la forma (R, G, B). Así pues, se puede encontrar un color expresado de la siguiente forma: (255, 0, 0) entonces se obtiene el color puramente rojo.

Se tiene en cuenta que las imágenes en modo RGB sólo son válidas para su visionado en soportes electrónicos, nunca para su utilización en medios impresos.

También se tomó en cuenta la resolución de las imágenes utilizadas para la detección de rostro. La resolución en una imagen digital de mapa de bits es la característica que le permite tener mayor o menor nitidez o calidad visual, y apreciar mayor o menor detalle en la imagen.

Las dimensiones del área que ocupa una imagen digital, alto y ancho, ya que siempre son cuadrangulares, se miden en píxeles. El pixel es la unidad mínima de color homogéneo que constituye una imagen digital bitmap, sea ésta una fotografía, un gráfico o un fotograma de video digital. Las dimensiones de una imagen están íntimamente relacionadas con la resolución y peso de la imagen.

Como se ha escrito líneas arriba, una vez que se cuenta con una imagen en la que los objetos son rostros y otros tipos de objetos, se procede a realizar operaciones sobre estas imágenes en las que los primeros tipos de objetos son los de mayor interés. Así por ejemplo se puede estar interesado en detectar rostros, reconocer éstos, y/o analizarlos. Cada una de estas operaciones tiene sus detalles y a continuación se describirán algunos aspectos básicos de ellos.

## 2.2. Detección, Reconocimiento y Análisis de Rostros

La **detección** de rostros es una tecnología informática que determina la ubicación y el tamaño de un rostro humano en la imagen digital (Kumar, Kaur y Kumar, 2018). Se detectan las características faciales y cualquier otro objeto como árboles, edificios y cuerpos se ignoran de la imagen digital.

Puede considerarse como un caso específico de detección de clases de objetos, donde la tarea es encontrar la ubicación y los tamaños de todos los objetos en una imagen que pertenece a una clase determinada.

La detección de rostros puede verse como un caso más general de localización de rostros. En la localización de rostros, la tarea es identificar las ubicaciones y los tamaños de un número conocido de rostros (generalmente uno) (Kumar, Kaur y Kumar, 2018). En particular, en la detección de rostros se encuentran algoritmos muy conocidos en el campo, algunos de los cuales se comentan a continuación.

## 2.3. Criterios Falso–Positivo, Falso–Negativo, Verdadero–Positivo, Verdadero–Negativo

Los conceptos más comunmente usados para especificar la eficiencia de un algoritmo de detección de rostros, aunque también en otros tipos de aplicaciones, son los de “Falso–Positivo” y “Falso–Negativo”(Shetty et al., 2021). Sin embargo, como se observa más adelante, resultan también importantes los conceptos de “Verdadero–Positivo” y “Verdadero–Negativo”. Todos estos conceptos se explican a continuación.

1. Falso–Positivo. La palabra “falso” se refiere al hecho de que es verdad que “no hay un rostro en la imagen”, y la palabra “positivo” se refiere al resultado de que se detecta un rostro en la imagen, suponiendo que en ésta no hay rostros en realidad.

En otras palabras, “habiendo no rostro en la imagen”, se detecta rostro (Falso–Positivo):  $\sim(\text{no hay rostro en la imagen}) = \text{sí detecta rostro en la imagen}$ , donde el símbolo “ $\sim$ ” significa “negación” y es la razón por la que se falsea la proposición “no hay rostro en la imagen”.

La falsedad radica en la proposición  $\sim(\text{no hay rostro en la imagen})$  (se niega que no haya rostro en la imagen y esto es una falsedad porque efectivamente no hay rostro en la imagen), mientras que la positividad se manifiesta en el hecho de que se detecta rostro en la imagen.

2. Falso–Negativo. La palabra “falso” se refiere al hecho de que no es verdad que “no hay rostro en la imagen” y la palabra “negativo” se refiere al resultado de la detección de un rostro en la imagen, suponiendo que en la imagen hay un rostro realmente. En otras palabras, “habiendo rostro en la imagen”, no se detecta rostro (Falso–Negativo):  $\sim(\text{sí hay rostro en la imagen}) = \text{no detecta rostro en la imagen}$ , donde el símbolo “ $\sim$ ” significa “negación” y es la razón por la que se falsea la proposición “sí hay rostro en la imagen”. La falsedad radica en la proposición  $\sim(\text{sí hay rostro en la imagen})$  (se niega que haya rostro en la imagen y esto es una falsedad porque efectivamente sí hay rostro en la imagen), mientras que la negatividad se manifiesta en el hecho de que no detecta rostro en la imagen.

3. Verdadero–Positivo. La palabra “verdadero” se refiere al hecho de que es verdad que “hay un rostro en la imagen” y la palabra “positivo” se refiere a que se detecta un rostro en la imagen, suponiendo que en realidad existe un rostro en la imagen. En otras palabras, “habiendo rostro en la imagen”, se detecta rostro (Verdadero–Positivo):  $(\text{sí hay rostro en la imagen}) = \text{detecta rostro en la imagen}$ , donde el símbolo “ $\sim$ ” no aparece más porque no se está falseando nada. La verdad radica en la proposición “sí hay rostro en la imagen” (se afirma que sí hay rostro en la imagen), mientras que la positividad se manifiesta en el hecho de que sí se detecta rostro en la imagen.

4. Verdadero–Negativo. La palabra “verdadero” se refiere al hecho de que es cierto que “no hay rostro en la imagen” y la palabra “negativo” se refiere a que no se detecta rostro en la imagen, suponiendo que realmente en la imagen no hay rostro. En otras palabras, “no habiendo rostro en la imagen”, no se detecta rostro (Verdadero–Negativo):  $(\text{no hay rostro en la imagen}) = \text{no detecta rostro en la imagen}$ . La verdad radica en la proposición “no hay rostro en la imagen” (se afirma que no hay rostro en la imagen), mientras que la negatividad se manifiesta en el hecho de que efectivamente no hay rostro en la imagen.

## 2.4. Viola-Jones

De acuerdo a la referencia (Skelly y Latifi, 2019), para comprender mejor el funcionamiento del algoritmo original de Viola–Jones, es importante profundizar en las cuatro partes principales del mismo, las cuales son: las características rectangulares de Haar, la imagen integral, el algoritmo de aprendizaje y el clasificador en cascada.

### 2.4.1. Características de Haar

Las características rectangulares tipo Haar se utilizan para clasificar una subventana como “potencialmente conteniendo una cara” o “ciertamente no conteniendo una cara”. Estas entidades se clasifican en tres grupos: entidades de dos rectángulos, entidades de tres rectángulos y entidades de cuatro rectángulos.

De acuerdo con la Figura 2.1, las funciones (a) y (b) se clasifican como características de dos rectángulos. La función (a) probablemente se usaría para detectar la diferencia de color en el área de los ojos y las cejas de una persona y en su frente. La función (b) se puede usar para detectar un borde vertical entre la cara de alguien y el fondo detrás de la cara.

La función (c) se clasificaría como una característica de tres rectángulos y se puede usar para detectar la diferencia de color entre los ojos de una persona y el puente de la nariz. La función (d) define una característica de cuatro rectángulos y podría usarse para detectar diferencias diagonales en varias partes de la cara.

Las diferentes combinaciones, orientaciones y tamaños de características rectangulares como estas, le dan al algoritmo una guía diferente sobre si una subventana contiene una cara o no. Para todas estas funciones, la suma de píxeles en los rectángulos blancos se resta de la suma de los píxeles en los rectángulos grises para calcular las diferencias en la oscuridad o la luminosidad en una subventana determinada.

Con una gran cantidad de píxeles, la cantidad de operaciones de suma sería muy grande y el algoritmo se ejecutaría muy lentamente. Para evitar esto, Viola y Jones introdujeron el concepto de la imagen integral.

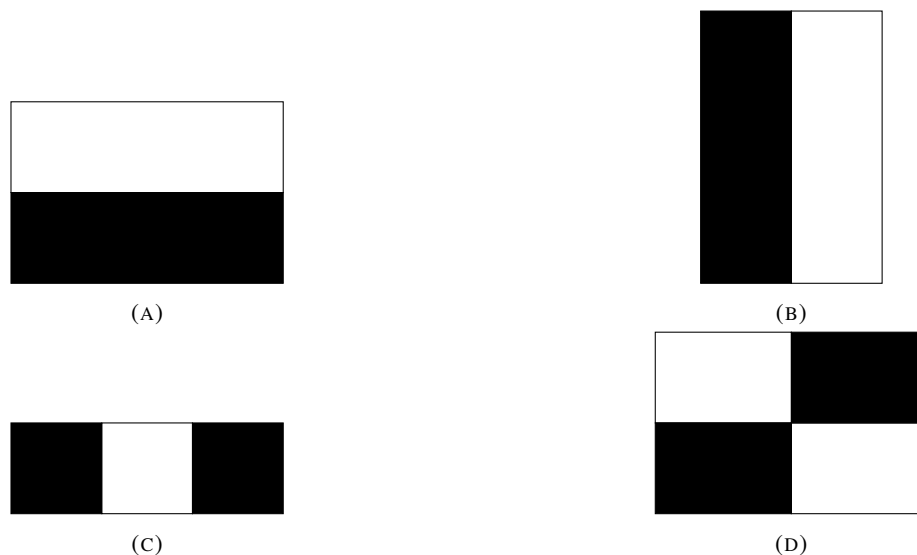


FIGURA 2.1: Ejemplo de las características similares a Haar (Skelly y Latifi, 2019).

### 2.4.2. Imagen integral

La imagen integral es una forma inteligente de calcular previamente todas las sumas de valores de píxeles que podrían ser necesarios durante la detección. En la imagen original, cada píxel contiene un valor que representa el valor de intensidad de ese píxel.

En una imagen en escala de grises, este valor sería un número entero entre 0 (negro) y 255 (blanco). La imagen original de ejemplo que se muestra en la Figura 2.2 solo contiene valores de píxeles que son múltiplos de 10, por lo que el cálculo mental no requiere tareas para facilitar la comprensión del ejemplo.

En la imagen integral, cada píxel contiene el valor de la suma de todos los píxeles de arriba y de la izquierda, incluido él mismo. Entonces, para este ejemplo, en la imagen integral, el píxel (1,1) tiene un valor de 240.

	0	1	2	3
0	30	50	70	80
1	70	90	50	60
2	90	10	30	60
3	60	40	40	20

(A) Imagen de entrada

	0	1	2	3
0	30	80	150	230
1	100	240	360	500
2	190	340	490	690
3	250	440	630	850

(B) Imagen integral

FIGURA 2.2: Ejemplo de una imagen integral (Skelly y Latifi, 2019).

Ese valor se obtiene sumando los píxeles (0,0), (0,1), (1,0) y (1,1) de la imagen original. Esta suma se calcula como  $30 + 50 + 70 + 90$ , lo que da como resultado un valor de 240. Siguiendo el mismo proceso, el píxel (3,3) es la suma de todos los valores de píxel en la imagen original.

Esta imagen integral se calcula antes de realizar cualquier detección, de modo que para cada región de las características del rectángulo, el número de cálculos es pequeño y el proceso es rápido.

### 2.4.3. El algoritmo de impulso (AdaBoost)

Al entrenar el algoritmo de Viola–Jones, es importante explorar todas las características rectangulares posibles que pueden caber en una subventana de imagen dada para determinar qué características son las más críticas para detectar rostros.

Viola y Jones afirman: “Dentro de cualquier subventana de imagen, el número total de características Haar–like es muy grande, mucho mayor que el número de píxeles. Para garantizar una clasificación rápida, el proceso de aprendizaje debe excluir una gran mayoría de las funciones disponibles y centrarse en un pequeño conjunto de funciones críticas” (Paul Viola y Jones, 2004).

Esta es la motivación para impulsar el proceso de convertir a los algoritmos que aprenden de manera ineficiente en algoritmos más eficientes en el aprendizaje automático. Por sí solos, cada una de las características rectangulares Haar–like definen algoritmos poco eficientes y harían un mal trabajo al detectar rostros en subventanas de imágenes.

Viola–Jones utiliza un algoritmo AdaBoost adaptado para seleccionar qué características son las más críticas. AdaBoost comienza con un algoritmo ineficiente y clasifica cada muestra como muestra positiva o negativa.



Específicamente, AdaBoost comienza con el algoritmo ineficiente que clasifica mejor cada una de las subventanas independientemente de cualquier otra característica. En el caso de la detección de rostros, un ejemplo de un algoritmo ineficiente sería la característica (a) de la Figura 2.1.

Supóngase que esta característica es la que mejor clasifica todas las subventanas inicialmente. AdaBoost ejecutará esta característica particular de un tamaño específico sobre cada subventana en el conjunto de datos de entrenamiento y clasificará cada uno como un ejemplo positivo (puede contener una cara) o un ejemplo negativo (no contiene una cara).

Está claro que esta característica por sí sola hará un mal trabajo al separar las muestras, pero supóngase que hace un mejor trabajo que cualquiera de las otras características por sí solas. Se calcula el error y, en la siguiente ronda, se hace hincapié en clasificar correctamente todas aquellas subventanas que se clasificaron incorrectamente en la ronda anterior.

Se selecciona un nuevo algoritmo ineficiente para clasificar todos los datos de entrenamiento nuevamente, pero esta vez se hace énfasis adicional en clasificar correctamente las subventanas previamente clasificadas incorrectamente.

Cuando se completa el entrenamiento, hay una gran cantidad de algoritmos ineficientes, cada uno con una cantidad diferente de información sobre si una subventana en particular contiene una cara o no. La cantidad de información se basa en qué tan bien esa característica clasificó el conjunto de datos ponderado.

Las características que tienen la menor cantidad de información se pueden descartar o dejar de lado, y sólo aquellas con la mayor cantidad de información se utilizan para realizar la clasificación final de nuevas imágenes. Descartar las funciones con la menor cantidad de información da como resultado un aumento insignificante del error.

Esto se debe a que ciertas características hacen un trabajo muy pobre al clasificar el conjunto de datos, por lo que su opinión en la decisión es insignificante. Con este proceso, Viola y Jones pudieron reducir la cantidad de funciones de 180,000 a sólo unos pocos cientos con una precisión notable.

De acuerdo con (Paul Viola y Jones, 2004), “los experimentos iniciales demostraron que un clasificador de cara frontal construido a partir de 200 características produce una tasa de detección del 95 % con una tasa de falsos positivos de 1 en 14,084”.

Esta precisión, sin embargo, fue considerada “no suficiente para muchas tareas del mundo real” por Viola y Jones. El detector final tiene más de 6,000 funciones, pero sigue siendo 15 veces más rápido que los algoritmos de la competencia de su época.

#### 2.4.4. Arquitectura en cascada

El detector final toma la forma de una cascada, también llamada “árbol de decisión degenerado”, en el que un resultado positivo del primer clasificador da como resultado una solicitud de evaluación a un segundo clasificador, y un resultado positivo del segundo clasificador da como resultado una solicitud de evaluación a un tercero, y así sucesivamente hasta completar la cascada.

Las características presentes en los clasificadores están determinadas por el proceso AdaBoost descrito anteriormente, y cada etapa de la cascada contiene un clasificador más complejo que la etapa anterior. Si en alguna etapa el resultado es negativo, se rechaza toda la subventana.

Si una subventana atraviesa toda la cascada sin rechazo, entonces se determina que la subventana contiene una cara. Esta estructura en cascada también se conoce como “cascada atencional”, porque rechaza rápidamente las subventanas que probablemente no contengan un rostro, y así el detector pueda prestar más atención a las subventanas que pueden contenerlo. El rápido rechazo de las subventanas sin rostro da como resultado la impresionante velocidad del detector.

## 2.5. Redes Neuronales Artificiales

Las redes neuronales artificiales, o ANNs o RNAs, son modelos basados en el comportamiento de redes de neuronas biológicas, aunque simplifican muchas de las funciones de estas últimas (Bravo y Sotelo, 2017). Las propiedades que preservan son la capacidad de integrar información y la capacidad de activarse en función de esta información recabada. Existen muchas diferentes arquitectura de RNAs (Bravo y Sotelo, 2017), siendo una de las más conocidas la arquitectura con alimentación hacia adelante (feedforward) como la que muestra Figura 2.3.

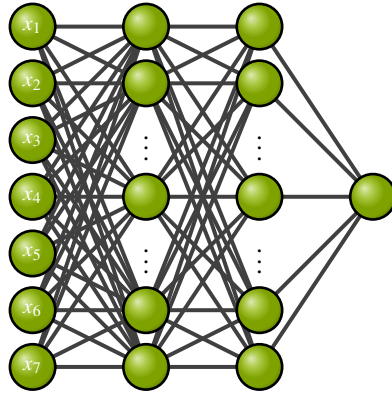


FIGURA 2.3: Estructura específica de una red neuronal artificial con alimentación hacia adelante que tiene una capa de entrada con siete componentes (sin contar la variable muda), dos capas ocultas con más de diez unidades de procesamiento y una capa de salida con una neurona artificial.

Esta arquitectura es también conocida como Perceptrón Multicapas y juega en la actualidad un papel sumamente importante cuando se combina con arquitecturas mucho más sofisticadas, que van desde la presencia de varios filtros adaptables en secuencia para el procesamiento de imágenes, hasta el uso de redes recurrentes tipo LSTM para el procesamiento de lenguaje natural.

En este trabajo, se analizan diferentes estructuras de perceptrón multicapa para analizar el impacto que ellas tienen en el proceso de detección de rostro. Además, algunas de las arquitecturas propuestas se combinan con otros métodos para analizar el impacto que tienen también en la detección de rostros.

## 2.6. Invariantes de Hu

De acuerdo a (Flusser, 1999), los invariantes de momento se han convertido en una herramienta clásica para el reconocimiento de objetos durante los últimos 30 años. Fueron presentados por primera vez a la comunidad de reconocimiento de patrones por Ming-Kuei Hu, quien empleó los resultados de la teoría de las invariantes algebraicas y derivó sus siete famosas invariantes a la rotación de objetos 2-D:

$$\phi_1 = \mu_{20} + \mu_{02},$$

$$\phi_2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2,$$

$$\phi_3 = (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2,$$

$$\phi_4 = (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2,$$

$$\phi_5 = (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12})((\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2) + (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})(3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2),$$

$$\phi_6 = (\mu_{20} - \mu_{02})((\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2) + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03}),$$

$$\phi_7 = (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})((\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2) - (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03})(3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2).$$

donde:

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - x_c)^p (y - y_c)^q f(x, y) dx dy \quad (2.1)$$

es el momento central del objeto  $f(x, y)$  y  $(x_c, y_c)$  son las coordenadas del centroide del objeto. Hu también mostró cómo lograr la invariancia a la escala. Desde entonces, se han dedicado numerosos trabajos a las diversas mejoras y generalizaciones de los invariantes de Hu y también a su uso en muchas áreas de aplicación.

## 2.7. Componentes Conexas

De acuerdo con (Sehgal y Venkatesh, 2019), el etiquetado de componentes conectados es uno de los aspectos más importantes en el campo de la visión artificial y el procesamiento de imágenes.

Técnicamente, todos los objetos de imagen están formados por componentes que a su vez están formados por pixeles conectados. Por lo tanto, el etiquetado de componentes conectados es uno de los aspectos más importantes del procesamiento de imágenes y la visión artificial.

La entrada a este procedimiento es una imagen binaria y la salida es una imagen simbólica, es decir, una imagen etiquetada donde todos los componentes están etiquetados.

A todos los pixeles conectados con el mismo valor se les asigna una etiqueta de identificación. Los componentes conectados se refieren al conjunto de pixeles que tienen el mismo valor y están conectados entre sí de manera que existe una ruta entre cada dos pixeles del conjunto de componentes conectados.

Los componentes conectados se refieren a dos pixeles  $p(x, y)$ ,  $q(s, t)$  que están conectados si, y solamente si, existe un camino entre ellos. Aquí la ruta, o camino, se refiere a la secuencia de pixeles distintos con coordenadas  $(x_1, y_1) \dots (x_n, y_n)$  donde  $(x_1, y_1) = (x, y)$  y  $(x_n, y_n) = (s, t)$  y  $n =$  longitud del camino.

Cada imagen está compuesta de pixeles. Para una imagen binaria, el valor de los pixeles puede ser 1 o 0. Si la imagen es una en escala de grises, o una imagen RGB, primero se convierte en una imagen binaria utilizando uno de los varios métodos disponibles.

El más utilizado es el método de umbralización. En el método de umbral, se establece un nivel de umbral, digamos 100, luego todos los valores de pixeles superiores a 100 se reemplazan con 1 y los pixeles con un valor inferior a 100 se reemplazan con 0 o viceversa.

El enfoque básico para el análisis/etiquetado de componentes conectados para cada imagen de tamaño  $m \times n$  es el siguiente:

1. Escanea la imagen pixel por pixel y actualiza la matriz de etiquetas asignando una etiqueta al pixel actual que se está escaneando. El escaneo verifica todos y cada uno de los pixeles.
2. Para cada pixel que se escanea, primero verifica si tiene asignada alguna etiqueta o no. Si no, entonces continúa con el siguiente paso.
3. Si el pixel no tiene una etiqueta asignada, busca los 4 vecinos inmediatos de manera muy similar a como se hace con la regla de conectividad de 8.
  - a) Regla de conectividad de 4 vecinos. Pocos algoritmos también usan esta regla.
  - b) Si el pixel no tiene vecinos etiquetados conectados, al pixel actual se le asigna una nueva etiqueta.

- c) De lo contrario, si tiene vecinos conectados con etiquetas asignadas, el mínimo de todas las etiquetas se asigna al pixel actual.

Después de etiquetar todos los pixeles, la matriz de etiquetas pasa por un segundo escaneo para garantizar que todos los pixeles tengan exactamente una etiqueta asignada y que se resuelven todas las irregularidades entre las etiquetas.

Esto proporcionará una cantidad precisa de componentes conectados mediante la asignación de una etiqueta diferente para cada componente conectado (Sehgal y Venkatesh, 2019).

## 2.8. Filtro de Gabor

De acuerdo a (Montero, Márquez y Rodríguez, 2011), el filtro de Gabor es un filtro lineal cuya respuesta de impulso es una función sinusoidal multiplicada por una función gaussiana.

El filtrado de una imagen con funciones de Gabor está relacionado con los procesos en la corteza visual. Concretamente, son un buen modelo para los campos receptivos de las células simples de la corteza cerebral si se supone que éstas poseen un comportamiento lineal.

Además, los filtros de Gabor se han empleado en el procesamiento digital de imágenes, donde se han mostrado eficientes a la hora de realizar diferentes tareas, tales como segmentación de texturas, compresión, etc.

El primer hecho sirvió de inspiración para el desarrollo de aplicaciones en el campo del tratamiento de imágenes. Las funciones de Gabor son funciones que operan en el conjunto de los números complejos, y que la parte real es la función de Gabor simétrica (simetría par). La parte imaginaria es la función de Gabor asimétrica (simetría impar).

La parte real del filtro de Gabor está dada por:

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi\frac{x'}{\lambda} + \psi\right) \quad (2.2)$$

La parte imaginaria del filtro de Gabor está dada por:

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \sin\left(2\pi\frac{x'}{\lambda} + \psi\right) \quad (2.3)$$

donde:

$$x' = x \cos \theta + y \sin \theta$$

$$y' = -x \sin \theta + y \cos \theta,$$

$\sigma$ : varianza de la función gaussiana.

$\lambda$ : ancho de banda de la función sinusoidal.

$\theta$ : orientación de la normal.

$\gamma$ : elipticidad de la función de Gabor.



## Capítulo 3

# Estado del Arte

La detección de rostros por medio de cámaras o videos es un tema complejo, aunque suene muy sencillo, y existan gran variedad de dispositivos que cuenten con esto. Algunas personas han desarrollado algoritmos basándose en redes neuronales artificiales, algunas de las cuales se implementaron a través de prueba y error, obteniéndose cada vez una mejor eficiencia.

Existen varios documentos de investigación acerca del tema desde los años 70, uno de ellos da a entender que el problema consiste en reconocer rostros (C. J. Hawthorn y Scholten, 2001), pero previamente utiliza un paso de identificación, en donde las características más importantes consideran rostros rotados y la aplicación de una red neuronal recurrente que es calificada como universal.

Adicionalmente, justifica el estudio de la detección de rostros haciendo referencia al hecho de que las imágenes con que se trabaja finalmente carecen de muchas características deseables que podrían encontrarse en la imagen original, pero que finalmente, debido a perturbaciones, obligan a reducir el tamaño en píxeles original.

Sin embargo, los autores asumen que previamente se ha logrado detectar el rostro, procediendo entonces a su identificación o reconocimiento. La parte que resulta interesante de este trabajo, es la que se refiere a la operación de rotación, lo cual también puede ser un problema a resolver solamente en el caso de detección de rostros.

Otro trabajo realizado tuvo como propósito crear un sistema que sea lo suficientemente rápido para la detección de rostros en tiempo real (Devrari y Kumar, 2011), para ello se emplean GPUs, por medio de las cuales primero se verifica si se trata de un rostro o no.

La arquitectura propuesta se basa en el algoritmo de Viola–Jones, permitiendo buscar las características del rostro humano, se hace una comparación del tiempo que tarda en detectar el rostro entre el CPU y GPU, contando con una gran velocidad de procesamiento en la arquitectura propuesta con el GPU, teniendo ésta una gran efectividad ya que, a pesar de tener oclusiones, aún así detecta el rostro de la persona en la imagen.

El inconveniente de esta aproximación radica en el hecho de que no se utilizan redes neuronales artificiales, sino tecnologías basadas en tarjetas gráficas, lo cual acelera el tiempo de procesamiento, tema que no es de interés en este trabajo. El trabajo se hará teniendo en cuenta procesamiento secuencial solamente, lo que permite simplificar en cierta forma la implementación del algoritmo necesario.

Utilizando la transformada de Fourier y una red neuronal artificial con arquitectura feed-forward (El-Bakry y Zhao, 2007), algunas personas intentaron verificar si en la imagen de muestra ésta un rostro o no. Para ello, utilizaron una plantilla estándar de la cara siendo ésta comparada con la imagen en la que se desea saber si hay un rostro o no, teniendo resultados muy favorables en la detección del rostro; no solamente el rostro humano, sino también el rostro de animales, sin embargo la resolución de las imágenes es muy baja y todas, durante el proceso de la detección, se convierten en escala de grises.

Las imágenes utilizadas sólo están enfocadas al rostro, siendo esto un inconveniente debido a que influyen ciertos factores como la posición, la iluminación, además de que la resolución de las imágenes no se toma en cuenta, dejando con incertidumbre si la red neuronal funciona apropiadamente en la detección de los rostros con estos factores.

Se ha diseñado un sistema que es capaz de detectar y reconocer el rostro (Gonzalez y Ramirez, 2011), basándose en el tono de piel. Primero detectan la tonalidad de la piel y determinan si existen similitudes con la tonalidad de la piel humana, llevando la imagen a una RGB normalizada con el fin de suavizarla.

Posteriormente, se detecta el rostro utilizando la red neuronal, teniendo éxito con la mayoría de las muestras; sin embargo, el inconveniente que se tuvo radicó en la iluminación, además de no tomar en cuenta si hay rotación en la persona, oclusión u algunos otros obstáculos.

Igualmente, se realizó una comparación del tiempo que tardan dos algoritmos en detectar (Shivesh Bajpai y Karthik, 2013), en tiempo real, el rostro de las personas. Uno de ellos es el algoritmo basado en el gradiente, el cual se determinó que tiene desventajas en comparación con el algoritmo Silhouettes. El algoritmo basado en el gradiente tarda treinta y un segundos en detectar el rostro, y cuando la persona está muy cerca o lejos, no funciona apropiadamente (no detecta algo).

En cambio, Silhouettes no presenta esos problemas y lo hace en menos de un segundo, lo impresionante es el tiempo de detección en tiempo real, además del equipo utilizado consistente de un procesador Pentium con un core 2 duo y 1.83 GHz, actualmente hay equipos con mejores características.

El inconveniente de esta aproximación radica en el hecho de que no utiliza redes neuronales artificiales, sino sólo compara el tiempo en que se tardan los algoritmos no siendo de interés para este trabajo.

Por otro lado, otros investigadores en el área utilizaron diferentes enfoques de redes neuronales artificiales, los cuales incluyen el algoritmo de propagación hacia atrás en el entrenamiento de un perceptrón multicapas, análisis de componentes principales y redes neuronales artificiales con funciones de base radial, además de mapeos autoorganizados, etc.

El trabajo menciona de manera general el proceso para la obtención del reconocimiento del rostro (Nisha y Dahiya, 2015), primero mediante una cámara web se capturan las imágenes siendo éstas almacenadas para su posterior clasificación.

La imagen pasa por un preprocesamiento que consta de cuatro módulos: la detección de rostros, la igualización de histograma, un proceso de detección de bordes y la eliminación del ruido no deseado, para posteriormente pasar al entrenamiento orientado al aprendizaje de la clasificación, la cual determina si es rostro o no. Finalmente, se proporciona la información sobre el reconocimiento del rostro.

No se menciona la estructura de la Red Neuronal Artificial que utilizan; sin embargo, resultan interesantes los diferentes enfoques utilizados, uno de los cuales, el Perceptron Multicapas, inspira la realización de este documento.

Existe otra forma de identificar el rostro según otros investigadores (Szlavik y Szirany, 2003), en la cual se utilizó CNN-UM, y que se enfoca en buscar la nariz del rostro para así generar un eje simétrico y posteriormente rotar la imagen, teniendo ellos una base de 493 imágenes en escala de grises de 20 personas en diferente rotación de sus respectivos rostros.

No se menciona qué software utilizan, ni entran en muchos detalles, y solamente hacen referencia a que obtuvieron una tasa de éxito de identificación facial del 92.242%. El que se tengan imágenes de cada persona con diferente rotación es de mucho interés en este artículo, por que de esta forma podría mejorarse la eficiencia de los algoritmos empleados en este trabajo.

En otro trabajo, para la detección de rostros se ha manejado una red neuronal artificial con arquitectura feedforward que cuenta con dos capas (Tarun Kumar y Malik, 2011), junto

con la utilización del algoritmo Fast Fourier Transform. En este caso, se tiene una base de imágenes de rostros y no rostros con un tamaño de  $32 \times 36$ , con lo cual se va haciendo una comparación con la imagen de entrada para así determinar si hay un rostro o no en ella.

Se menciona que la imagen se convierte a escala de grises, no estipula qué método se utiliza para el procesamiento de la imagen. También se hicieron pruebas con imágenes de animales como el koala y el chimpancé, con las cuales se obtuvieron muy buenos resultados ya que se detectó su rostro. Solamente se utilizaron seis imágenes para la detección, donde se obtuvo el 92.006% de Verdadero-Positivo y el 9.89% de la tasa de Falso-Negativo.

Es sorprendente el resultado obtenido; sin embargo, la calidad de la imagen no tiene una muy buena resolución ya que solamente funciona para imágenes de baja resolución. Sería interesante saber si funciona para resoluciones más grandes, o si habría que modificar algo durante el entrenamiento de la red neuronal y/o en el proceso de detección de rostros.

En un estudio realizado para la detección ocular y estimación de la mirada (Zeynep Orman y Kemer, 2011), se determinó que para ello se tiene que realizar la detección del rostro, mencionando las técnicas y métodos que existen para poder llevar a cabo esta detección, describiendo ventajas y desventajas.

Su principal propósito está enfocado a detectar cuándo un conductor se encuentra cansado por medio del estado en que se encuentre su mirada para así poder evitar accidentes, lo que lo hace un estudio muy ambicioso y que ayudaría demasiado a la sociedad el resolverlo.

En las técnicas mencionadas, se encuentran las que se emplean para este proyecto y al igual que ellos, se encuentran las mismas dificultades para la detección del rostro, como la iluminación, las oclusiones, la rotación.

Se realizó un proyecto para la detección de rostros (Shahrin Azuan Nazeer y Khalid, 2007), describiendo que el proceso inicia con la entrada de la imagen en escala de grises. Posteriormente, se realiza el procesamiento de la imagen integral y se van extrayendo las características a considerar que permiten discernir si se trata de un rostro o no, utilizando el algoritmo de aprendizaje de AdaBoost y la RNA, pasando así a la clasificación utilizando la técnica en cascada de Viola-Jones con la RNA.

Ha sido muy impactante el resultado obtenido al combinar estas técnicas pero, debido a que en este documento se trabaja comparándolas por separado, se podría considerar el estudio de la combinación RNA con Viola-Jones como un trabajo a futuro.

Otro de los trabajos realizados para la detección de rostros está basado en momentos invariantes de Hu (Milan Tuba y Vukovic, 2014), los cuales son clasificados por una red neuronal artificial. El primer paso es detectar la piel por medio del método YCbCr, obteniendo una imagen binaria, posteriormente se utilizan dos operaciones morfológicas; a saber, erosión y dilatación, de lo cual sólo quedan las partes del cuerpo que es piel para así posteriormente aplicar la técnica de los invariantes de Hu.

Posteriormente a lo escrito en el párrafo inmediato anterior, con la RNA se define si se detecta o no un rostro. Los resultados son muy favorables, ya que se obtuvo el 94.74% de efectividad en la detección correcta de rostros, cabe destacar que el procesamiento de la imagen es interesante.

En 2004, se diseñó un sistema que se utiliza en tiempo real para detectar y reconocer rostros mediante un escáner (Jain, 2004). En dicho sistema, se utilizan imágenes en 2D para luego aplicar un modelo en imágenes 3D, empleando técnicas y algoritmos distintos a los utilizados en este trabajo.

Sin duda, existe una infinidad de modelos para la detección de rostros (Zhu y Ramanan, 2011). Tal es el caso de un modelo unificado para este fin, basándose en la estimación de poses y la localización de puntos de referencia utilizando una combinación de árboles con un conjunto compartido de partes, obteniendo muy excelentes resultados; sin embargo, una investigación de este tipo no es importante en el desarrollo de este trabajo, ya que no se utilizan RNA.



Cabe destacar que, para la detección de rostros (Aamer .S.S.Mohamed y Jiang, 2007), algunos métodos se basan en el color de piel, tal como un algoritmo desarrollado empleando la transformada discreta del coseno, así como las cromaticidades Cb y Cr de una imagen RGB. Posteriormente, una RNA clasifica si es un rostro o no. Se obtuvieron resultados favorables para imágenes verticales.

Sin embargo, en la imagen el rostro se encuentra en el escenario perfecto considerándolo como un ambiente en el que no existe oclusión alguna y el rostro está de frente.

En otro sistema realizado para la detección de rostros (Gurjar y Tiwari, 2014), se empleó el diseño de una red neuronal feedforward en conjunto con el filtro de Gabor. Para ello, se hace un preprocesamiento de la imagen donde pasan de RGB a escala de grises, entrando así por el filtro de Gabor, y finalmente determinar si hay un rostro o no. Para el entrenamiento de la red, se crea una base de datos de rostros y no rostros de  $27 \times 18$ .

Se menciona que se obtuvieron resultados favorables, aunque no se menciona cuántas neuronas y capas utilizaron, ni tampoco el total de imágenes que emplearon para el entrenamiento, y si hubo problemas con la iluminación, oclusión, la posición de la cara, etc., sin duda sería muy interesante saberlo ya que en este documento uno de los algoritmos que se emplearon es la combinación de RNA con el filtro de Gabor.

También se desarrolló un sistema para la detección de rostros (Hay Mar Yu Maung y Naing, 2016), el cual se implementó en MATLAB, cuya interfaz cuenta con un menú que contiene: creación de una base de datos, la cual contiene dos carpetas, una con rostros y otra donde no los hay, con 69 imágenes en cada una de estas carpetas. Dentro del menú de opciones que presenta la interfaz, se encuentra la inicialización de la RNA.

La misma interfaz gráfica presenta la opción de entrenamiento de la red neuronal artificial, así como las pruebas a realizar con las imágenes con el objeto de detectar si hay un rostro o no, mencionándose que se utilizó una capa de entrada, dos capas ocultas con cien neuronas y una capa de salida, con la función de transferencia "tansig". Para el procesamiento de la imagen se utilizó un histograma para así manejar la iluminación.

Los resultados que los investigadores obtuvieron dejan en claro que mientras más neuronas se tengan mejor es la detección, se aclara que este sistema es solamente para imágenes en donde no hay oclusiones, sin duda se enfrentan a las mismas problemáticas como: la oclusión, la pose y la resolución de imagen.

Se presentó un algoritmo de detección de rostros basado en una cascada de conjuntos de árboles de decisión (V. V. Khryashchev y Priorov, 2017). Este algoritmo modifica el enfoque estándar de Viola-Jones al utilizar una cascada de clasificadores binarios para explorar imágenes. Cada clasificador se entrena para un rango específico de ángulos de rotación de la cabeza, lo que permite obtener mejores resultados en comparación con el algoritmo estándar de Viola-Jones.

Sin duda, esta es una forma innovadora de detectar rostros y resulta extremadamente interesante para su posible uso en conjunto con una Red Neuronal Artificial (RNA) como trabajo futuro en esta tesis. .

Se ha diseñado un algoritmo robusto para la detección de rostros en imágenes a color (Tin, 2012). Este algoritmo comienza detectando la piel y segmentando el color. Posteriormente, se extraen características faciales, identificando así la boca, la nariz y los ojos. Una vez hecho esto, se realiza un análisis utilizando el Análisis de Componentes Principales (PCA).

El conjunto de entrenamiento consta de 200 imágenes, que incluyen 100 imágenes frontales verticales, 50 imágenes frontales con gafas y 50 imágenes de rostros ligeramente girados. Estas imágenes fueron seleccionadas de manera selectiva de Internet.

Los resultados obtenidos son favorables, sin embargo, la iluminación presenta un desafío, ya que los rostros con gafas pueden afectar la detección. Esta técnica permite la detección de rostros y sería interesante explorar la eficiencia que se podría lograr al utilizarla en conjunto con una Red Neuronal Artificial (RNA).

Como se puede observar, hasta este punto se ha realizado una recopilación de diversas técnicas que existen para la detección de rostros y resulta interesante comparar cuáles son las ventajas y desventajas de algunas de ellas (Ming-Hsuan Yang y Fellow, 2002), así como cuáles son las más efectivas, dando la posibilidad de mejorar los algoritmos empleados.

Esta recopilación es de gran ayuda porque así se puede partir de las características de los algoritmos y retroalimentar los que se emplean en este trabajo.

Hasta ahora se concluye que las problemáticas a las que actualmente se enfrenta el problema de detección de rostros son las definidas por la iluminación, las oclusiones, las poses y la rotación del rostro, así como el rendimiento de la red neuronal artificial usada, el cual es afectado por los ajustes en el número de neuronas en cada capa o el número de éstas.

Para resaltar aún más la importancia de lo anterior, algunos investigadores llevaron a cabo una investigación de trabajos existentes acerca de las metodologías (Manisha M. Kasar y Kim, 2016), técnicas y algoritmos empleados para la detección de rostros utilizando Redes Neuronales Artificiales, mencionando sus ventajas y desventajas. En algunos casos se menciona la cantidad de imágenes que utilizan, así como la descripción general del proceso para la detección de rostros.

Los autores que llevaron a cabo lo descrito en el párrafo inmediato anterior, intentan desarrollar un sistema híbrido, con una alta tasa de éxito.

Sin duda, el investigar y hacer la comparativa de los trabajos existentes para crear un sistema con mayor eficiencia es muy favorable, pero aún así, como los autores mencionan, se enfrentan a las mismas problemáticas ya descritas líneas arriba, siendo ejemplo de ellas la iluminación de la imagen, la rotación del rostro, y las oclusiones.

De manera similar, los autores (Nisha Soni y Mathu, 2013) desarrollaron un sistema para la detección de rostros utilizando Redes Neuronales Artificiales SOM en Matlab. Este sistema emplea tres técnicas diferentes para la extracción de características de la imagen: la Transformada Discreta Wavelet (DWT), la Transformada Discreta del Coseno (DCT) y la detección de bordes Sobel. Lograron una tasa de detección del 100 % durante 5 pruebas consecutivas.

Sin embargo, las imágenes utilizadas solo muestran la cara sin oclusiones y con buena iluminación, lo que podría explicar la alta tasa de detección. Sería interesante emplear la RNA SOM en conjunto con Viola-Jones, Invariantes de Hu o Filtro de Gabor, especialmente en condiciones de iluminación y oclusión más desafiantes.

Similarmente, los autores (Marta Lucía Guevara y Urueña, 2008) implementaron un método de detección de rostros y características faciales sobre imágenes digitales, basándose en la extracción de características con filtros de base Haar y clasificadores en cascada, obteniendo una eficiencia en la detección de rostro y los ojos con un valor entre el 92 % y el 100 %, todo en un tiempo de procesamiento rápido, y sin utilizar redes neuronales artificiales.

Por otro lado, algunos investigadores han desarrollado un modelo para el reconocimiento de rostros (Kubota, 2011), en el que primero detectan en la imagen digital si hay un rostro o no. El proceso de detección se hace por medio de AdaBoost en combinación con una red neuronal artificial del tipo feedforward con tres capas de neuronas.

Se menciona que en la primera capa el número de neuronas es equivalente a la longitud del vector de características extraído y que sirve como entrada a la red neuronal artificial, para la segunda capa sólo se menciona que el número de neuronas se selecciona en función del experimento, y en la última capa sólo hay una neurona, utilizando la función de transferencia "tansig".

Al implementar el modelo en forma híbrida, obtuvieron un tasa de éxito considerable, tenían muy pocos Falsos-Positivos; Sin embargo, sus muestras revelaron que el rostro no se detecta si no está adecuadamente iluminado.

En ese trabajo, sólo consideran que su modelo presente la menor tasa de Falsos – Positivos y sea más rápido el proceso. Pese a ello, al pasar al reconocimiento de rostro utilizan

el método de Análisis de Componentes Independientes en conjunto con redes neuronales artificiales.

Cabe resaltar que deberían mejorar la detección de rostros considerando todos los factores que influyen, intentando hacer que la eficiencia mejore para dar paso así al reconocimiento de rostros como se propone en este trabajo.

En otro artículo (Ma. Christina D. Fernandez y Dadios, 2014), también se desarrolló un sistema para la detección y verificación de rostros en tiempo real utilizando el algoritmo de Viola–Jones y redes neuronales artificiales. El sistema generado puede reconocer rostros humanos con una precisión del 87.05 %, se menciona que funciona mejor si la persona se encuentra a 150 cm de la cámara, o si se encuentra directamente enfrente de la cámara.

Como en algunos de los casos anteriores, los autores de ese artículo se enfrentan a la problemática de la iluminación en el rostro. Estos resultados sugieren que primero se debería atacar completamente la resolución de estos tipos de problemas, como lo es la iluminación, las oclusiones que tiene el rostro, así como la distancia en la que se encuentran las personas .

Los autores (Tikoo y Malik, 2014), implementan en Matlab una interfaz capaz de detectar el rostro utilizando el algoritmo de Viola-Jones, para posteriormente reconocer el rostro. Para ello, una vez que se detecta el rostro la imagen, ésta se pasa a un formato de escala de grises, para posteriormente determinar, por medio de una red neuronal feedforward, a quién pertenece el rostro.

En dicho trabajo no se menciona que tuvieran problemas como las oclusiones o de rotación del rostro; sin embargo, por la imagen que muestran de ejemplo se observa un escenario perfecto: el rostro frente a la cámara, con buena iluminación y sin ninguna oclusión, lo que permite tener una tasa de éxito muy buena .

La revisión que se hace en este trabajo (Al-blushi, 2021), de varias arquitecturas de redes neuronales artificiales con técnicas de análisis para el reconocimiento de rostros, se hizo una comparación con la efectividad que tiene cada una de ellas, anotando las ventajas y desventajas.

Como se menciona en estos trabajos revisados, los autores de ellos se enfrentan con las mismas dificultades, que son nuevamente la iluminación, la pose del rostro y las oclusiones. Sin duda, estos problemas siguen siendo un gran desafío .

Más recientemente, se ha desarrollado una interfaz en la que se determina el género de la persona a la que pertenece el rostro (Berbar, 2022), previamente haciendo un reconocimiento facial utilizando una red convolucional, el modelo clasificador de género binario obtuvo una alta precisión y el modelo del reconocimiento obtuvo un 98 %.

En ese trabajo, se menciona que las imágenes utilizadas de la base de datos son con el rostro de perfil, de frente y algunos tienen lentes, de acuerdo a las imágenes mostradas como ejemplo que se utilizan para la base solo se encuentran de frente.

Su modelo funciona mejor con las imágenes a color comparado con las imágenes a escala de grises, sin duda al utilizar redes convolucionales arroja muy buenos resultados; sin embargo, en ese trabajo se utiliza otro tipo de red neuronal y técnicas de extracción de características.

Un reporte de investigación adicional que optimiza el diseño de una red neuronal artificial feedforward para el reconocimiento de rostros (Maryam Mahmood Hussein y Shareef, 2020), mediante el algoritmo de búsqueda Harmony, implantado en MatLab, la tasa de éxito subió considerablemente de un 82 % a un 94 %, debido a que se optimiza la red buscando cuantas neuronas se necesitan en la capa de entrada, oculta y de salida obteniendo así la mejor eficiencia de la red neuronal artificial que definen.

En ese trabajo, no se menciona si se tuvieron problemas tales como la iluminación, oclusiones y poses del rostro; sin embargo, sólo se basan en la optimización de la red híbrida. El

trabajo resulta interesante ya que podría optimizar la red neuronal que se utiliza en este documento, para así poder tener una mejor eficiencia una vez resueltos los problemas ya antes mencionados que se tienen al detectar el rostro.

Finalmente, otros investigadores desarrollaron (Dr. Vinayak Bharadi, 2022), un sistema en tiempo real para el reconocimiento de rostros. Utilizaron imágenes del rostro de 20 personas y emplearon la arquitectura VGG16 de Redes Neuronales Convolucionales (CNN) y Transfer Learning, así como la biblioteca de código abierto OpenCV para el preprocesamiento de imágenes.

Los resultados obtenidos oscilaron entre el 97 % y el 99 %. Sin embargo, las imágenes que utilizaron solo mostraban a una persona de frente, con buena iluminación y sin ninguna oclusión. Esto podría explicar la alta tasa de éxito. Sería interesante ver qué sucede con imágenes que presenten a más personas, con diferentes posiciones del rostro, diferentes condiciones de iluminación y oclusiones

Una vez hecha la revisión que se ha expuesto en esta sección, se concluye que existe una infinidad de investigación que se ha realizado sobre el tema, lo cual resalta la importancia del mismo. En esta revisión se han encontrado bastantes inconvenientes y características que se toman en cuenta para realizar la detección de rostros.



## Capítulo 4

# Desarrollo Experimental

El desarrollo experimental se lleva a cabo definiendo un arreglo experimental que está soportado principalmente por una interfaz gráfica implementada en MatLab. La interfaz se divide en varias secciones: una relacionada con las características de imagen/video, una segunda de visualización, y una tercera de acciones.

Las características de imagen/video consideran el hardware y ciertas propiedades de imagen/video. Las primeras se refieren al tipo de cámara usada, el tipo de procesador con que se trabaja, el tipo de tarjeta gráfica y el tipo de computadora. Las segundas toman en cuenta la resolución, el color de la imagen, el número de persona en ésta, el tipo de oclusión, si es que existe, la cobertura (si es medio cuerpo, solamente el rostro, etc.) y el entorno dentro del cual se toma la imagen.

La sección de visualización en la interfaz solamente muestra la imagen siendo analizada. Finalmente, la sección de acciones da la posibilidad de seleccionar la imagen a mostrar, identificando ésta con una etiqueta. Asimismo, presenta las opciones referentes al algoritmo con que se desea trabajar y deja, para un desarrollo posterior, la posibilidad de trabajar con videos en tiempo real, así como de presentar los resultados y las comparaciones.

Sin embargo, el trabajo manual fuera de la interfaz también juega un papel importante y en él los algoritmos son la parte medular. Este es el tema a tratar en la parte de Algoritmos.

### 4.1. Herramientas utilizadas

Para poder llevar a cabo este proyecto fue necesario la utilización de ciertas herramientas tanto de Software como de Hardware. Dichas herramientas se enlistan como se muestran en las siguientes subsecciones.

#### 4.1.1. Software

- MATLAB  
R2015a(8.5.0.197613)  
64-bit (win64)
  
- Access 2021
  
- Excel 2021

#### 4.1.2. Hardware

- Laptop marca HP  
Procesador: Intel(R) Core(TM) i3-5005U CPU @ 2.00GHz 2.00GHz

Memoria instalada (RAM): 8.00 GB

Tipo de sistema: Sistema operativo Windows 10 Home de 64 bits, procesador x64

- Cámara réflex Nikon  
Modelo D3300  
Lente: 18-55mm

## 4.2. Algoritmos

Se seleccionan tres algoritmos para llevar a cabo el análisis de la eficiencia de los mismos en la detección de rostros. El esquema de cada uno de ellos se muestra en las Figuras 4.1, 4.2 y 4.3. La forma en que estos algoritmos se utilizan en la interfaz gráfica es ilustrada por la Figura 4.4.

La Figura 4.1 corresponde a la secuencia que sigue el algoritmo de Viola–Jones para detectar un rostro. Como ya se comentó en la Sección 2.4, y las Subsubsecciones 2.4.1 y 2.4.3, el bloque de imagen integral tiene la función de acelerar el proceso de detección de características en el rostro. Posteriormente, el bloque Adaboost se encarga de trabajar con las subventanas y pasar finalmente al clasificador en cascada.

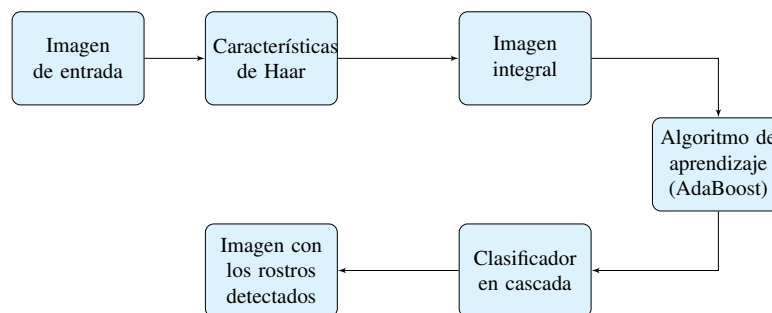


FIGURA 4.1: Diagrama de bloques del algoritmo Viola–Jones

Por otro lado, la Figura 4.2 muestra la secuencia que se sigue cuando se combinan Redes Neuronales Artificiales y un Filtro de Gabor. La imagen original es transformada a una en escala de grises, y posteriormente esto sirve como entrada a un bloque dentro del cual se combina, en forma similar a como se hace con filtros adaptables combinados con Redes Neuronales Artificiales, una Red Neuronal con un Filtro de Gabor, produciendo como salida los rostros detectados.

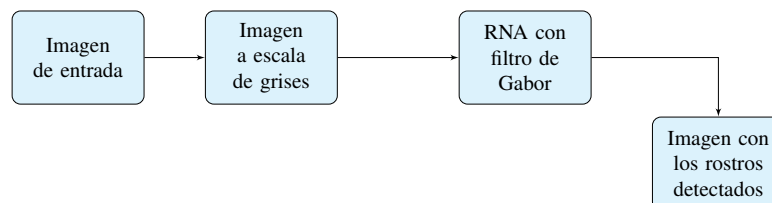


FIGURA 4.2: Diagrama de bloques de RNA con filtro de Gabor

Finalmente, la Figura 4.3 presenta la secuencia a seguir cuando los invariantes de Hu en la imagen original sirven como entrada a una Red Neuronal Artificial. La imagen original recibe un preprocesamiento ya descrito líneas arriba, y el resultado del mismo sirve como insumo al bloque dedicado a determinar los Invariantes de Hu.

Los invariantes detectados sirven como alimentación a una Red Neuronal Artificial previamente entrenada, la cual finalmente indica qué rostros aparecen en la imagen original.

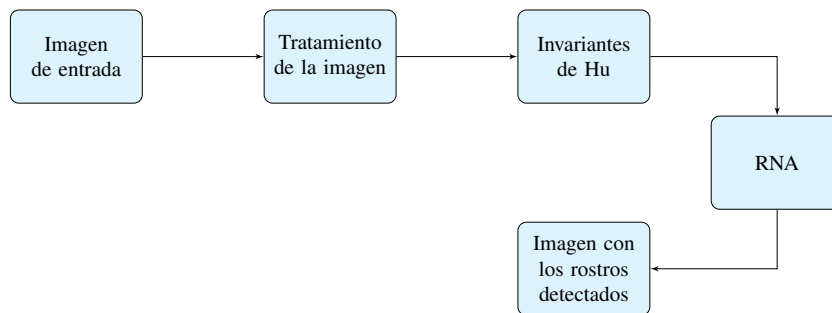


FIGURA 4.3: Diagrama de bloques de RNA con Invariantes de Hu.

Con este escenario experimental, se llevaron a cabo tres experimentos. En el primero de ellos, se compara Viola–Jones contra RNA combinadas con Filtro de Gabor. En el segundo experimento, se lleva a cabo una evaluación del comportamiento de la combinación Invariantes de Hu con Redes Neuronales Artificiales con el fin de detectar rostros en imágenes. En el tercero se compara Viola–Jones versus RNA–Filtro de Gabor versus Invariantes de Hu con tratamiento de imagen previo usando Igualización de histogramas adaptable (adaphisteq), igualización de histograma (histeq) y ajuste de contraste (imajust).

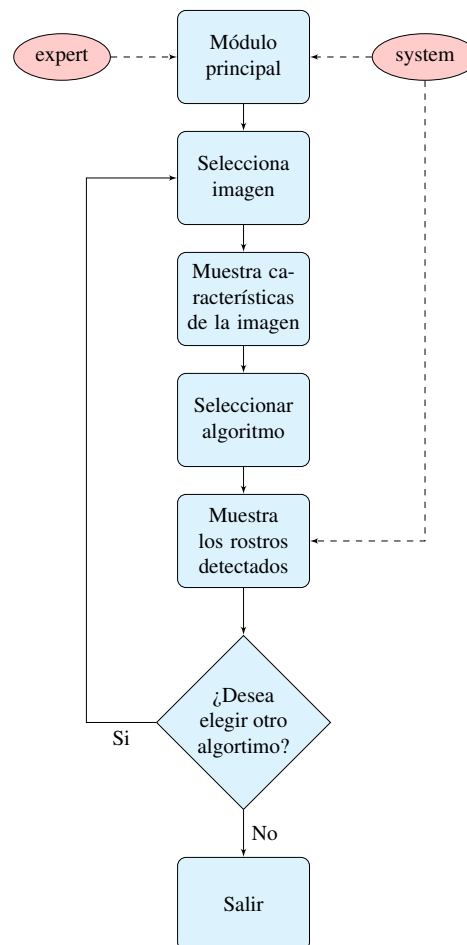


FIGURA 4.4: Diagrama de flujo de la interfaz de la detección de rostros



### 4.3. Experimento 1: Viola–Jones versus RNA–Filtro de Gabor

Para realizar esta prueba se ocupó un concentrado de imágenes que se tomaron a un equipo de fútbol. Por medio de la interfaz realizada para la detección de rostros, se introdujeron las imágenes del equipo de fútbol, las cuales tienen diferentes resoluciones, coberturas del cuerpo, número de personas y oclusiones, para así saber cuántos rostros se detectan con los algoritmos empleados (Viola–Jones versus RNA) y determinar cuál es el más eficiente. Algunos de los resultados experimentales se pueden observar en la Tabla 4.1, así como en la Figura 4.5.

En este experimento, se tienen 451 imágenes, en las cuales las primeras 19 tienen resolución diferente dadas por la Tabla 4.1 y la Figura 4.5. Las restantes 432 imágenes están clasificadas dentro de una de tres resoluciones posibles:  $600 \times 400$ ,  $300 \times 200$  y  $150 \times 100$ .

Se consideran estas resoluciones debido a que, con resoluciones menores o mayores a ellas, se encuentra que los algoritmos tienen problemas en la detección de rostros. Las imágenes en estas clases contienen una sola persona y además, de una u otra forma, la escena contiene oclusiones, algunas no son de cuerpo completo, algunas están giradas, etc.

La Figura 4.1 y la Figura 4.2 ilustran el flujo de ejecución de los algoritmos para este experimento, mostrando los bloques principales en los mismos y sin entrar en detalles.

Como un ejemplo del resultado que se obtiene al aplicar el diagrama de flujo dado por la Figura 4.1, la Figura 4.6 muestra la interfaz gráfica que se empleó a lo largo del trabajo para desarrollar los experimentos en el mismo, e indica los rostros detectados por Viola–Jones en la imagen específica seleccionada.

Como se puede observar, no se detectan todos los rostros en la imagen, aunque los cuadros de detección corresponden efectivamente a rostros.

Como comparación con este ejemplo, la Figura 4.7 muestra cómo reacciona la interfaz experimental cuando se emplean redes neuronales artificiales para detectar rostros en la misma imagen. Obsérvese que nuevamente no se detectan todos los rostros en la imagen, dos de ellos están detectados correctamente, y además, a diferencia de Viola–Jones, se detectan áreas que, erróneamente, se identifican como rostros.

Experimentalmente se encuentra que la alta resolución tiene un efecto favorable en la detección de rostros, como se puede observar en las imágenes con números 15, 16, 17 y 18, sin importar aparentemente que tengan o no oclusiones, sean o no mixtas o con cuerpo parcial o no; mostrándose el efecto contrario en imágenes con resolución más baja, como pueden ser las imágenes complementarias a las ya comentadas.

Sin embargo, comparativamente, el algoritmo de Viola–Jones muestra mejores resultados que el algoritmo basado en Redes Neuronales Artificiales.

TABLA 4.1: Resultados de la prueba número 1.

No. de Imagen	Resolución	Oclusión	Cobertura	No. de Personas	No.de rostros detectados por Viola-Jones	No.de rostros detectados por RNA
1	262 × 175	Gorra	Mixto	9	2	3
2	230 × 153	Gorra	Cuerpo completo	13	1	1
3	253 × 169	Gorro	Mixto	10	3	4
4	156 × 234	Ninguna	Medio cuerpo	3	1	1
5	164 × 245	Ninguna	Medio cuerpo	3	1	1
6	258 × 172	Lentes oscuros	Medio cuerpo	6	1	1
7	200 × 300	Ninguna	Cuerpo completo	8	1	1
8	230 × 153	Gorro	Medio cuerpo	7	1	1
9	300 × 200	Gorro	Medio cuerpo	5	2	2
10	301 × 200	Gorro, gorra	Medio cuerpo	6	3	2
11	253 × 169	Gorra	Medio cuerpo	5	2	1
12	305 × 457	Ninguna	Mixto	4	2	3
13	600 × 400	Gorra	Medio cuerpo	9	2	3
14	600 × 400	Gorra	Mixto	9	5	3
15	400 × 600	Ninguna	Mixto	5	5	3
16	433 × 289	Ninguna	Cuerpo completo	5	3	2
17	559 × 373	Ninguna	Medio cuerpo	6	5	2
18	502 × 302	Gorra, lentes oscuros	Mixto	16	14	12
19	494 × 333	Gorra, lentes oscuros	Medio cuerpo	3	1	1

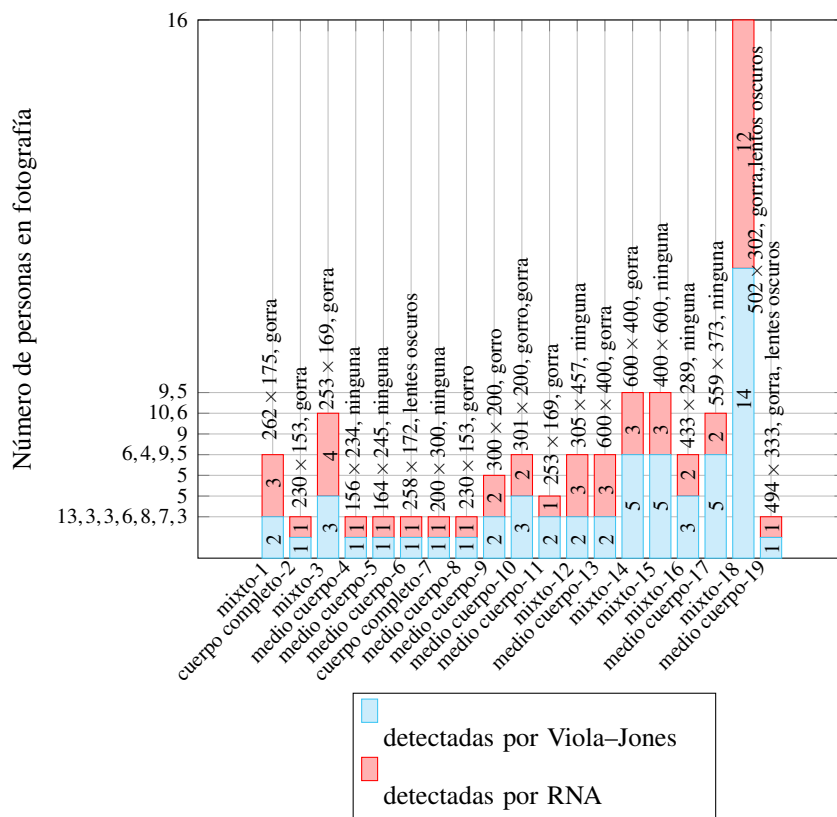


FIGURA 4.5: Número de rostros detectados por Viola-Jones y RNA.

Esto no debería sorprender ya que Viola-Jones es un algoritmo muy conocido que se emplea en diferentes plataformas, no solamente en MatLab, como pueden ser OpenCV y,

posiblemente, Octave.

Por esto mismo, cuenta ya con una base de datos de rostros pre-entrenada y lista para ser usada en cuestiones de aplicación real. Por otro lado, la arquitectura de red neuronal artificial se presta a realizar modificaciones posibles en el número de capas y en el número de neuronas por capa, dando también la posibilidad de trabajar con aprendizaje profundo, cosa que no se hizo en este trabajo.

Se llevó cabo este experimento primero, porque solamente se contaba con 19 imágenes y los algoritmos de Viola-Jones y RNAs, y se desconocían las otras técnicas.

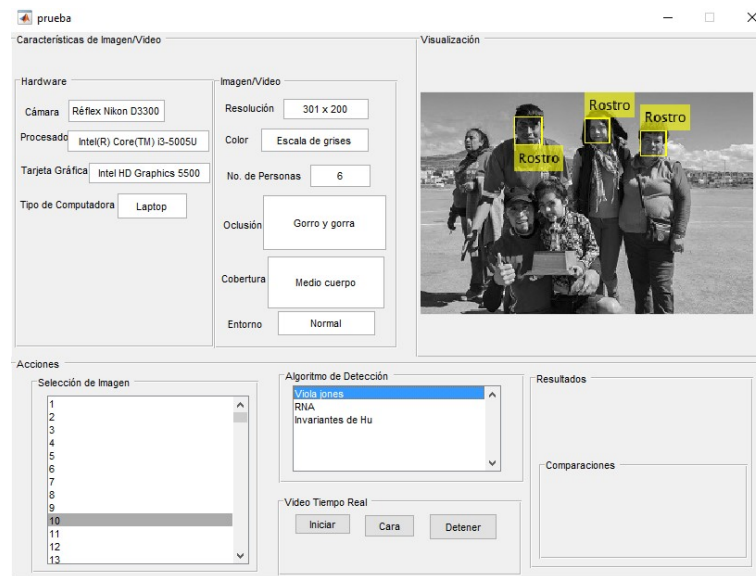


FIGURA 4.6: Interfaz gráfica utilizando el algoritmo de Viola-Jones

FIGURA 4.7: Interfaz gráfica utilizando la Red Neuronal Artificial

#### 4.4. Experimento 2: Invariantes de Hu antecediendo a redes neuronales artificiales para detectar rostros según los invariantes

En esta sección, con el fin de ilustrar la manera en que se combinan las técnicas de Invariantes de Hu con las técnicas de Redes Neuronales Artificiales, para detectar rostros, se lleva a cabo la construcción del algoritmo que se muestra en el Diagrama 4.3 y éste se aplica sobre las 451 imágenes con la intención de analizar la forma en que el algoritmo de entrenamiento de la red neuronal artificial, precedida por la construcción de los invariantes de Hu de la imagen bajo estudio, afecta la eficiencia de la detección de rostros.

Detrás de la construcción del diagrama mostrado por la Figura 4.3, y por lo tanto del respectivo algoritmo, se encuentra el análisis, y su efectividad, de dos arquitecturas de redes neuronales artificiales con alimentación hacia adelante, las cuales están definidas por una capa de entrada con siete componentes, que son los invariantes de Hu, una primera capa oculta con diez neuronas artificiales, una segunda capa oculta con 90 neuronas artificiales y una capa de salida con una neurona artificial, mientras que la segunda arquitectura tiene una capa de entrada con siete componentes, definidas por los invariantes de Hu, una primera capa oculta con diez neuronas artificiales, una segunda capa oculta con 50 neuronas artificiales, y una capa de salida con una neurona artificial.

En ambas arquitecturas, las funciones de transferencia de las capas ocultas y de salida son de tipo tangente hiperbólica, lo que quiere decir que la capa de salida señala que se ha detectado un rostro cuando la neurona artificial se encuentra en el estado  $+1$ , y en el estado  $-1$  para indicar que la región de interés en la imagen de entrada no es un rostro.

Las regiones de interés se detectan utilizando el concepto de invariantes de Hu, ya comentados en la Subsección 2.6. Conocidos los invariantes de Hu en la región, se pasan éstos como entradas a la arquitectura de red neuronal artificial, la cual previamente ha sido entrenada en forma supervisada.

El entrenamiento se probó con diversos algoritmos desde el más simple que es el Gradient Descent, hasta uno de los más complejos que es el de Gradientes Conjugados.

Cabe mencionar que la técnica de invariantes de Hu requiere de un preprocesamiento de la imagen, por medio de algoritmos como AdaptHisteq, Histeq o Imajust. Posteriormente, se presentaron imágenes que no se usaron para entrenamiento, y con las cuales se probó el funcionamiento de las arquitecturas de redes neuronales artificiales ya comentadas.

Para el entrenamiento de las redes neuronales artificiales se emplearon un total de 432 imágenes, algunas de las cuales contienen solamente un rostro y el resto contienen dos o más personas en el cuadro.

El entrenamiento solamente utilizó cuadros fotográficos con una sola persona. Para la prueba de efectividad del algoritmo de detección de rostros, una vez entrenadas las arquitecturas, se emplearon imágenes con dos o más personas en la escena.

Recuérdese que la idea de Falso-Positivo en el caso de detección de rostros se aplica al presentar en secuencia imágenes que contienen rostros, y se cuentan aquellas imágenes en las que el algoritmo detecta regiones que no son rostros y, sin embargo, afirma que sí lo son.

Por otro lado, un resultado de “Falso-Negativo” quiere decir que regiones que sí son rostros se identifican como que no son rostros. En oposición a un resultado “Falso-Positivo”, para determinar el valor de “Falso-Negativo” en la detección de rostros, se le presenta al algoritmo una secuencia de imágenes que contienen rostros, y se cuentan aquellas imágenes en las que el algoritmo afirma que no contiene rostros.

Por ejemplo, la Figura 4.7 proporciona cuatro Falso-Positivo (los cuatro rostros no enmarcados), cuatro Falso-Negativo (las cuatro regiones enmarcadas). Por otro lado, la Figura 4.6 da como resultado tres Falso-Positivo (los tres rostros no enmarcados) y cero Falso-Negativo.

Finalmente, la Figura 4.8 proporciona seis Falso - Negativo (los rostros no encuadrados de las personas en la imagen) y dos Falso-Positivo (las áreas enmarcadas que no contienen un solo rostro alguno).

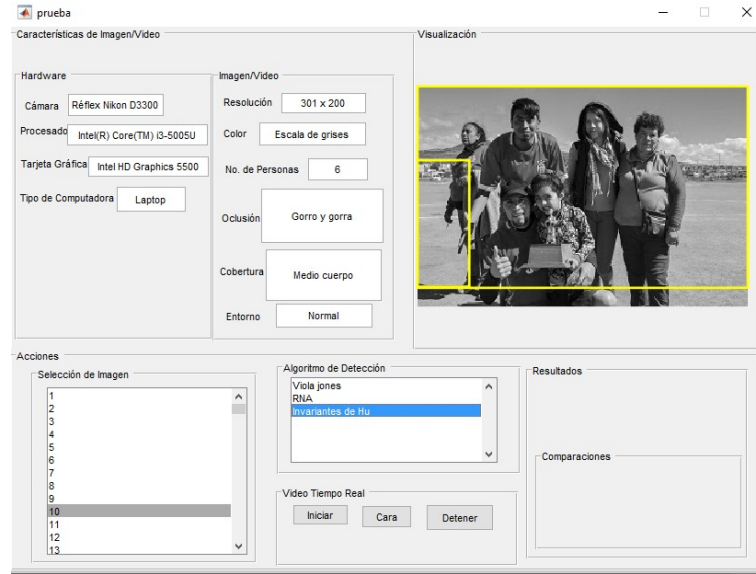


FIGURA 4.8: Interfaz gráfica utilizando el algoritmo de Invariantes de Hu.

Las Tablas 4.2 – 4.7 muestran el total de rostros detectados con invariantes de Hu (con preprocesamiento Adaphthisteq, Histeq o Imajust) para las dos arquitecturas ya señaladas, mostrándose los métodos de entrenamiento de las redes neuronales artificiales (Gradient Descent (Gd), Gradient Descent with Momentum (Gdm), Variable Learning Rate Backpropagation (Gdx), Rescilent Propagation (Rp), Conjugate Gradient Fletcher(Cgf), Conjugate Gradient Powell (Cgp), Conjugate Gradient with Powell–Beale Restarts (Cgb), Scaled Conjugate Gradient (Scg), BFGS quasi–Newton (Bfg), One–Step Secant (Oss), Levenberg–Marquardt (Lm) y Bayesian Regularization (Br)).

Las Tablas 4.2 – 4.7 muestran el número de rostros detectados de un total de 726, considerando que se emplearon las imágenes de entrenamiento y otro conjunto de imágenes que no se usaron para entrenar, pero que contenían uno o más rostros.

La segunda fila de la tabla indica el número de rostros detectados, habiendo entrenado previamente la red neuronal artificial con el algoritmo que se indica en la parte superior de la primera fila, en la columna correspondiente. También la tercera fila señala el porcentaje con respecto al total de rostros considerados (726).

Como se puede observar, la red neuronal artificial entrenada con el algoritmo de tasa de aprendizaje variable produce un mayor número de rostros detectados, resaltando el hecho de que los otros algoritmos no son satisfactorios.

TABLA 4.2: Resultados de la Técnica Adaphthisteq con [7 10 90 1] neuronas en cada capa de la RNA, empezando con la capa de entrada y terminando con la capa de salida. La red neuronal artificial se entrenó con 432 imágenes, todas ellas teniendo un solo rostro.

Gd	Gdm	Gdx	Rp	Cgf	Cgp	Cgb	Scg	Bfg	Oss	Lm	Br
13	29	<b>34</b>	12	21	13	15	16	6	12	3	12
1.79 %	3.99 %	<b>4.68 %</b>	1.65 %	2.89 %	1.79 %	2.07 %	2.20 %	0.83 %	1.65 %	0.41 %	1.65 %

TABLA 4.3: Resultados de la Técnica Adaphisteq con [7 10 50 1] neuronas en cada capa de la RNA, empezando con la capa de entrada y terminando con la capa de salida. La red neuronal artificial se entrenó con 432 imágenes, todas ellas teniendo un solo rostro.

Gd	Gdm	Gdx	Rp	Cgf	Cgp	Cgb	Scg	Bfg	Oss	Lm	Br
20	14	17	22	11	10	6	26	18	6	7	11
2.75 %	1.93 %	2.34 %	3.03 %	1.52 %	1.38 %	0.83 %	3.58 %	2.48 %	0.83 %	0.96 %	1.52 %

TABLA 4.4: Resultados de la Técnica Histeq con [7 10 90 1] neuronas en cada capa de la RNA, empezando con la capa de entrada y terminando con la capa de salida. La red neuronal artificial se entrenó con 432 imágenes, todas ellas teniendo un solo rostro.

Gd	Gdm	Gdx	Rp	Cgf	Cgp	Cgb	Scg	Bfg	Oss	Lm	Br
42	31	76	278	128	239	73	75	264	36	67	260
5.79 %	4.27 %	10.47 %	38.29 %	17.63 %	32.92 %	10.06 %	10.33 %	36.36 %	4.96 %	9.23 %	35.81 %

TABLA 4.5: Resultados de la Técnica Histeq con [7 10 50 1] neuronas en cada capa de la RNA, empezando con la capa de entrada y terminando con la capa de salida. La red neuronal artificial se entrenó con 432 imágenes, todas ellas teniendo un solo rostro

Gd	Gdm	Gdx	Rp	Cgf	Cgp	Cgb	Scg	Bfg	Oss	Lm	Br
36	248	60	88	264	255	232	45	70	60	257	78
4.96 %	34.16 %	8.26 %	12.12 %	36.36 %	35.12 %	31.96 %	6.20 %	9.64 %	8.26 %	35.40 %	10.74 %

TABLA 4.6: Resultados de la Técnica Imagejust con [7 10 90 1] neuronas en cada capa de la RNA, empezando con la capa de entrada y terminando con la capa de salida. La red neuronal artificial se entrenó con 432 imágenes, todas ellas teniendo un solo rostro.

Gd	Gdm	Gdx	Rp	Cgf	Cgp	Cgb	Scg	Bfg	Oss	Lm	Br
184	13	229	8	239	25	241	234	241	232	246	238
25.34 %	1.79 %	31.54 %	1.10 %	32.92 %	3.44 %	33.20 %	32.23 %	33.20 %	31.96 %	33.88 %	32.78 %

TABLA 4.7: Resultados de la Técnica Imagejust con [7 10 50 1] neuronas en cada capa de la RNA, empezando con la capa de entrada y terminando con la capa de salida. La red neuronal artificial se entrenó con 432 imágenes, todas ellas teniendo un solo rostro.

Gd	Gdm	Gdx	Rp	Cgf	Cgp	Cgb	Scg	Bfg	Oss	Lm	Br
16	236	153	11	34	249	247	229	18	246	192	31
2.20 %	32.51 %	21.07 %	1.52 %	4.68 %	34.30 %	34.02 %	31.54 %	2.48 %	33.88 %	26.45 %	4.27 %

Los porcentajes en cada una de las Tablas 4.2 hasta 4.7 se calculan considerando el total de rostros en las imágenes utilizadas, que son 451, y el total de rostros en ellas que son 726, determinando así cuál es más eficiente en la detección de rostros con el método de invariantes de Hu, como se muestra en la gráfica 4.9.

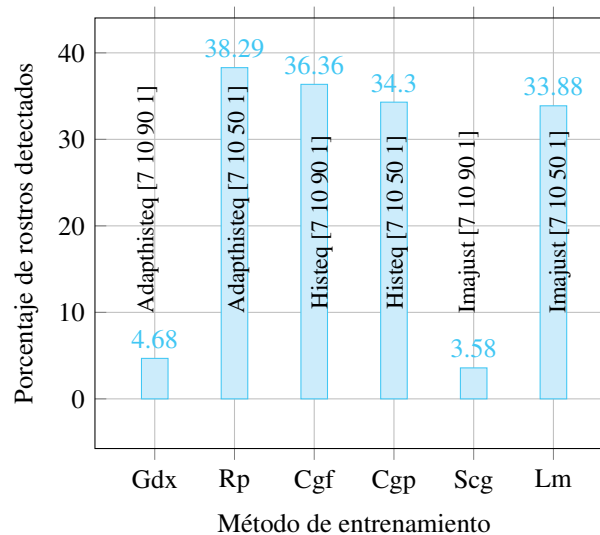


FIGURA 4.9: Mejores métodos de entrenamiento de una red neuronal artificial cuando se combinan con Invariantes de Hu.

#### 4.5. Experimento 3: Viola–Jones versus RNA–Filtro de Gabor versus Invariantes de Hu con tratamiento de imagen previo usando Igualización de histogramas adaptable (adapthisteq), igualización de histograma (histeq) y ajuste de contraste (imajust)

Una vez que se obtiene la mejor arquitectura de la Red Neuronal Artificial en combinación con Invariantes de Hu, tal como se especifica en el Experimento 2, se procede ahora a comparar sus resultados con aquellos obtenidos a través de los algoritmos de Viola–Jones y la combinación de Redes Neuronales Artificiales con Filtro de Gabor. Se toman en cuenta la cantidad de rostros detectados, y los resultados se muestran en Tabla 4.8 y en el diagrama de barras mostrado por Figura 4.10.

Como se puede observar, el algoritmo de Viola–Jones produce los mejores resultados, ocupando las Redes Neuronales Artificiales una mejor segunda posición y quedando al final la combinación que se hace con ellas con Invariantes de Hu.

Sin duda, una mejoría de los resultados requiere profundizar en la forma en que se definen las arquitecturas de Redes Neuronales Artificiales, posiblemente a través de la automatización del proceso mismo de construcción de la arquitectura, tanto para la combinación de Redes Neuronales Artificiales con Filtro de Gabor y la combinación de Redes Neuronales con Invariantes de Hu.

TABLA 4.8: Resultados del número de rostros detectados por Viola-Jones, RNA e Invariantes de Hu.

Algoritmos	Rostros detectados	Porcentaje de los rostros detectados
Viola-Jones	322	44.35 %
RNA	201	27.69 %
Invariantes de Hu	278	38.29 %

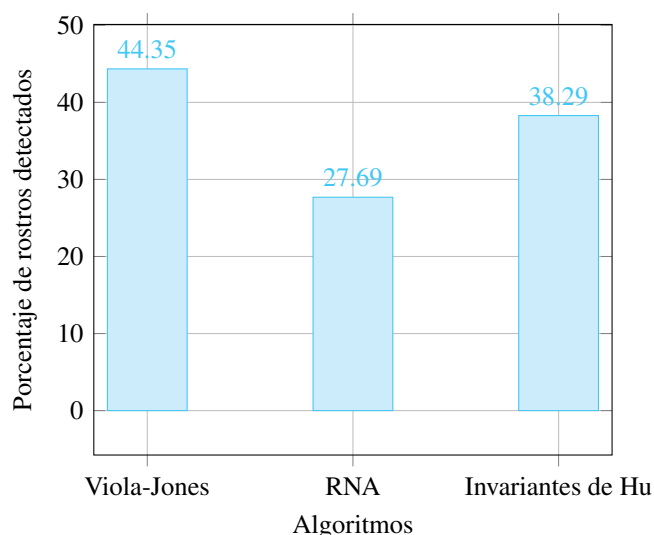


FIGURA 4.10: Comparación de algoritmos empleados para la detección de rostros.

No solamente se analizan en este trabajo el número de Verdadero–Positivo, sino también el número de Falso–Positivo, en los que se enmarcan partes de la imagen dada, pero en realidad no existen rostros. Se registro el número de recuadros Falso–Positivo para determinar cuál de los algoritmos empleados tiene este valor más grande, los resultados se pueden visualizar en la Tabla 4.9, así como en la grafica mostrada por Figura 4.11.

TABLA 4.9: Resultados de los Falso–Positivo para cada uno de los tres algoritmos analizados.

Algoritmos	Total de Falso–Positivo
Viola-Jones	3
RNA	951
Invariantes de Hu	1173

Como era de esperar, los Invariantes de Hu combinados con Redes Neuronales Artificiales arrojan el peor resultado, mientras que Viola–Jones logra una mejor puntuación, coincidiendo este hecho con los resultados encontrados a través de los Verdadero–Positivo.



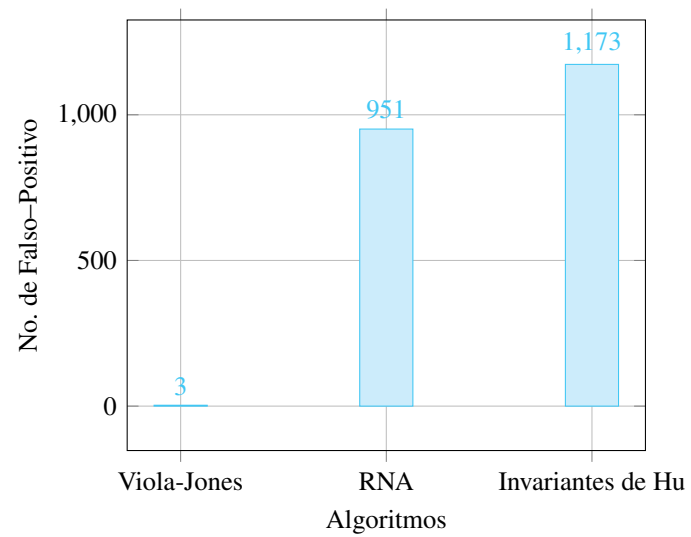


FIGURA 4.11: Número de Falso-Positivo en relación con de Viola-Jones, RNA e Invariantes de Hu.

## Capítulo 5

# Conclusiones y Trabajo Futuro

Viola – Jones resulta ser el mejor de los otros algoritmos analizados, debido a que utiliza funciones de Haar y métodos en cascada, además de que los comandos que lo representan, al menos en MatLab, han sido construidos a través de entrenamientos que consideran bases de imágenes bastante ricas en contenido.

Sin embargo, al algoritmo enfrenta dificultades si la resolución son mayores o menores a las aquí trabajadas, por ello se debería buscar la manera de hacer la detección con resoluciones mayores,  $600 \times 400$ ,  $150 \times 100$ . El problema de detección de rostro aparece porque Viola–Jones sólo funciona para resoluciones de  $384 \times 288$  píxeles, y a 15 fotogramas por segundo en el caso de imagen/video.

Adicionalmente, se necesitan más imágenes de entrenamiento para el caso de los algoritmos donde intervienen Redes Neuronales Artificiales, a resoluciones menores o mayores estos algoritmos tienen dificultades para hacer la detección, incluido Viola – Jones.

Existen muchas otras deficiencias, sobre todo cuando no solamente se trata de detectar el rostro, sino también reconocer cuándo se trata del rostro de una persona real, o de una fotografía de ella, o cuando se trata de reconocer rostros de gemelos.

Para resolver este tipo de problemas, suele acudir a obligar que la persona mueva la cabeza, o bien parpadee, o bien se instala un sensor térmico, o bien introduciendo adicionalmente un identificador de huellas digitales.

En el caso de los algoritmos que emplean Redes Neuronales Artificiales, debería buscarse la forma de automatizar la búsqueda de la mejor arquitectura, con la cual se tenga un mejor desempeño.

Para una buena detección se debería contar con una cámara fotográfica o de video con buena resolución. Asimismo, para mejorar el rendimiento del algoritmo, por ejemplo para aplicaciones en videovigilancia, el ambiente debe ser perfecto; es decir, que no existan oclusiones en el rostro de la persona a detectar, entre otros aspectos.

Hay trabajos que se enfocan a determinar la nacionalidad de la persona identificando el color de la piel, los rasgos oculares, etc. Esto conduce al área de análisis de rostros, en donde también se puede determinar la edad, los estados anímicos de la persona, etc.

Otro aspecto que se propone trabajar es el que se refiere a la forma de medir la eficiencia de un algoritmo de detección de rostros. En este sentido, se propone trabajar con los cuatro posibles resultados: Falso–Positivo, Falso–Negativo, Verdadero–Positivo y Verdadero – Negativo, cada una de las cuales se muestra en Figura 5.1.

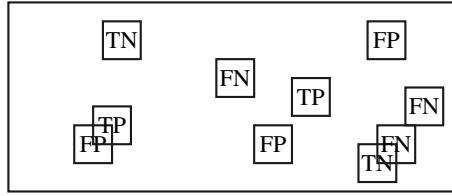


FIGURA 5.1: Distribución de resultados posibles en la detección de rostro y su posible impacto en el cálculo de la eficiencia de un algoritmo de detección de acuerdo a lo que se comenta en el documento.

De acuerdo con la Figura 5.1, en la situación real, el número total de rostros es igual a  $NRD = \text{número de } FN + \text{número de } TP$ . Recuérdese que el primer término en la parte derecha de esta ecuación indica que realmente existe rostro en la imagen pero el algoritmo indica que no.

Por otro lado, la misma figura indica que el número total de cuadros o marcos colocados por el algoritmo, pero que en realidad no contienen rostros, es dada por la ecuación  $NRF = \text{número de } FP + \text{número de } TN$ .

Sin embargo, en general, las áreas que definen estas regiones enmarcadas pueden intersectarse, de tal forma que si  $A$  y  $B$  son cualquier par de estas regiones enmarcadas, entonces, la contribución real al área de la imagen es dada por la siguiente deducción,

$$\text{área de } A = \text{área de } (B^C \cap A) + \text{área de } (B \cap A)$$

$$\text{área de } B = \text{área de } (A^C \cap B) + \text{área de } (B \cap A)$$

$$\begin{aligned} \text{área de } A + \text{área de } B &= \text{área de } (B^C \cap A) + \text{área de } (A^C \cap B) + \\ &2\text{área de } (B \cap A) \end{aligned}$$

lo que significa que

$$\begin{aligned} \text{área de } A + \text{área de } B - \text{área de } (B \cap A) &= \text{área de } (B^C \cap A) + \text{área de } (A^C \cap B) + \\ &\text{área de } (B \cap A) \\ &= \text{área de } (A \cup B) \end{aligned}$$

En una primera aproximación, se tiene la siguiente relación entre las áreas de imágenes,

$$\begin{aligned} \text{área total en la que no se enmarcan rostros} &= \text{área total de la imagen} - \\ &- NRD - NRF + \\ &FP \cap TN \cap FN \cap TP \end{aligned}$$

En otras palabras,

$$\begin{aligned} NRF &= \text{área total de la imagen} - NRD - \\ &- \text{área total en la que no se enmarcan rostros} + \\ &+ FP \cap TN \cap FN \cap TP \end{aligned}$$

cuando el valor de  $NRF$  sea suficientemente pequeño, entonces se puede decir que el algoritmo es bastante eficiente. Sobre este tema se está trabajando actualmente por parte de algunos de los autores de este trabajo.



# Bibliografía

- Aamer .S.S.Mohamed Ying Weng, Stan S Ipson y Jianmin Jiang (2007). «Face Detection based on Skin Color in Image by Neural Networks». En: págs. 1-5. URL: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1084.1715&rep=rep1&type=pdf>.
- Al-blushi, Ameera A. (abr. de 2021). «Face Recognition Based on Artificial Neural Network: A review». En: *Artificial Intelligence and Robotics Development Journal* 1. URL: [https://pdfs.semanticscholar.org/9205/111662fba5243109a0eeacb3b8d5b2cc6579.pdf?\\_ga=2.208880919.484360636.1658266660-1957770660.1652034753](https://pdfs.semanticscholar.org/9205/111662fba5243109a0eeacb3b8d5b2cc6579.pdf?_ga=2.208880919.484360636.1658266660-1957770660.1652034753), págs. 117-131.
- Alcalá, José Ramón y Guillermo Navarro (2008). *Una introducción a la imagen digital y su tratamiento*. MEDECIANT.
- Berbar, Mohamed A. (jul. de 2022). «Faces Recognition and Facial Gender Classification using Convolutional Neural Network». En: *Menoufia J. of Electronic Engineering Research (MJEER)* 31.2, págs. 1-10. URL: [https://mjeer.journals.ekb.eg/article\\_244525\\_0a5a9d18714686e23f966317858a65ac.pdf](https://mjeer.journals.ekb.eg/article_244525_0a5a9d18714686e23f966317858a65ac.pdf).
- Bouzalmat, Anissa et al. (sep. de 2011). «FACE DETECTION AND RECOGNITION USING BACK PROPAGATION NEURAL NETWORK AND FOURIER GABOR FILTERS». En: *International Journal (SIPIJ)* 2.3, págs. 15-21. URL: [https://www.researchgate.net/publication/266603162\\_Face\\_Detection\\_and\\_Recognition\\_Using\\_Back\\_Propagation\\_Neural\\_Network\\_and\\_Fourier\\_Gabor\\_Filters](https://www.researchgate.net/publication/266603162_Face_Detection_and_Recognition_Using_Back_Propagation_Neural_Network_and_Fourier_Gabor_Filters).
- Bravo, Eduardo Francisco Caicedo y Jesús Alfonso López Sotelo (2017). *UNA APROXIMACIÓN PRÁCTICA A LAS REDES NEURONALES ARTIFICIALES*. Universidad del Valle Programa Editorial.
- C. J. Hawthorn, K. P. Weber y R. E. Scholten (dic. de 2001). «Littrow Configuration Tunable External Cavity Diode Laser with Fixed Direction Output Beam». En: *Review of Scientific Instruments* 72.12, págs. 4477-4479. URL: <http://link.aip.org/link/?RSI/72/4477/1>.
- Devrari, Kailash y K.Vinay Kumar (dic. de 2011). «Fast Face Detection Using Graphics Processor». En: *International Journal of Computer Science and Information Technologies* 2.12, págs. 1082-1086. URL: <https://hal.archives-ouvertes.fr/hal-00782740/document>.
- Dr.Vinayak Bharadi Mr.Rutik Sansare, Mr.Tushar Padelkar andMr.Vishant Shinde (abr. de 2022). «Real Time Face Recognition System Using Convolutional Neural Network». En: *International Journal of creative research thoughts(IJCRT)* 10, págs. 75-79. URL: <https://ijcrt.org/papers/IJCRT0020015.pdf>.
- El-Bakry, Hazem M. y Qiangfu Zhao (2007). «Fast Object/Face Detection Using Neural Networks and Fast Fourier Transform». En: *International Journal of Computer, Electrical, Automation, Control and Information Engineering* 1, págs. 3752-3757. URL: <file:///C:/Users/arime/Downloads/Fast-ObjectFace-Detection-Using-Neural-Networks-and-Fast-Fourier-Transform1.pdf>.
- Fernández, Sergio Domínguez (jun. de 2014). «GENDER AND AGE CLASSIFICATION BASED ON FACIAL FEATURES». Tesis doct. UNIVERSIDAD DE MÁLAGA.

- Flusser, Jan (mayo de 1999). «On the independence of rotation moment invariants». En: *Pattern Recognition Society*, págs. 1405-1410. URL: <http://library.utia.cas.cz/prace/20000033.pdf>.
- Gonzalez, Angel Noe Martinez y Victor Ayala Ramirez (2011). «Real Time Face Detection Using Neural Networks». En: *Department of Electronics Engineering, Universidad de Guanajuato DICIS*, págs. 88-89. URL: <https://www.uv.mx/rossum/CD/papers/088.pdf>.
- Gurjar, Lalita y Nidhish Tiwari (mar. de 2014). «FACE DETECTION USING FEED FORWARD NEURAL NETWORK IN MATLAB». En: *International Journal of Application or Innovation in Engineering and Management* 3, págs. 1-3. URL: <https://www.ijaiem.org/volume3issue4/IJAIEM-2014-04-28-065.pdf>.
- Hay Mar Yu Maung, Hla Myo Tun y Zaw Min Naing (jul. de 2016). «Implementation of Neural Network Algorithm for Face Detection Using MATLAB». En: *International Journal of Scientific and Research Publications* 6, págs. 239-243. URL: <http://www.ijserp.org/research-paper-0716/ijserp-p5539.pdf>.
- Jain, Anil K. (ago. de 2004). «Face Detection and Recognition». En: págs. 1-8. URL: <https://apps.dtic.mil/dtic/tr/fulltext/u2/a430264.pdf> AIM-2001.
- Kubota, Naoyuki (jul. de 2011). «Applying Artificial Neural Networks for Face Recognition». En: págs. 1-16. URL: <https://www.hindawi.com/journals/aans/2011/673016/>.
- Kumar, Ashu, Amandeep Kaur y Munish Kumar (jul. de 2018). «Face Detection Techniques: A Review». En: págs. 1-21. URL: [https://www.researchgate.net/publication/326667118\\_Face\\_Detection\\_Techniques\\_A\\_Review](https://www.researchgate.net/publication/326667118_Face_Detection_Techniques_A_Review).
- Ma. Christina D. Fernandez Kristina Joyce E. Gob, Aubrey Rose M. Leonidas Ron Jason J. Ravara Argel A. Bandala y Elmer P. Dadios (2014). «Simultaneous Face Detection and Recognition using Viola-Jones Algorithm and Artificial Neural Networks for Identity Verification». En: págs. 672-676. URL: [https://www.researchgate.net/publication/269987856\\_Simultaneous\\_Face\\_Detection\\_and\\_Recognition\\_Using\\_Viola-Jones\\_Algorithm\\_and\\_Artificial\\_Neural\\_Networks\\_for\\_Identity\\_Verification](https://www.researchgate.net/publication/269987856_Simultaneous_Face_Detection_and_Recognition_Using_Viola-Jones_Algorithm_and_Artificial_Neural_Networks_for_Identity_Verification).
- Manisha M. Kasar, Debnath Bhattacharyya y Tai hoon Kim (mar. de 2016). «Face Recognition Using Neural Network: A Review». En: *International Journal of Security and Its Applications* 10.3, págs. 81-100. URL: [https://www.researchgate.net/profile/Manisha-Kasar/publication/301727666\\_Face\\_Recognition\\_Using\\_Neural\\_Network\\_A\\_Review/links/5ef18af5a6fdcc73be96ccc2/Face-Recognition-Using-Neural-Network-A-Review.pdf](https://www.researchgate.net/profile/Manisha-Kasar/publication/301727666_Face_Recognition_Using_Neural_Network_A_Review/links/5ef18af5a6fdcc73be96ccc2/Face-Recognition-Using-Neural-Network-A-Review.pdf).
- Marta Lucía Guevara, Julian David Echeverry y Willian Ardila Urueña (jun. de 2008). «DETECCIÓN DE ROSTROS EN IMÁGENES DIGITALES USANDO CLASIFICADORES EN CASCADA». En: 14.38, págs. 1-5. URL: <https://www.redalyc.org/articulo.oa?id=84903801>.
- Maryam Mahmood Hussein, Ammar Hussein Mutlag y Hussain Shareef (2020). «An Improved Artificial Neural Network Design for Face Recognition utilizing Harmony Search Algorithm». En: págs. 1-11. URL: <https://iopscience.iop.org/article/10.1088/1757-899X/745/1/012038/pdf>.
- Milan Tuba, Romana Capor-Hrosik y Mirjana Vukovic (2014). «Face detection based on invariant moments classified by neural network». En: *INTERNATIONAL JOURNAL OF CIRCUITS, SYSTEMS AND SIGNAL PROCESSING* 8. [https://www.researchgate.net/publication/280096634\\_Face\\_detection\\_based\\_on\\_invariant\\_moments\\_classified\\_by\\_neural\\_network](https://www.researchgate.net/publication/280096634_Face_detection_based_on_invariant_moments_classified_by_neural_network) INTERNATIONAL\_JOURNAL\_OF\_CIRCUITS\_SYSTEMS\_AND\_SIGNAL\_PROCESSINGVol\_8\_2014\_pp\_203-210, págs. 203-210.

- Ming-Hsuan Yang, David J. Kriegman y Narendra Ahuja Fellow (ene. de 2002). «Detecting Faces in Images: A Survey». En: *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE* 24.1, págs. 34-58. URL: [https://www.researchgate.net/publication/3193340\\_Detecting\\_Faces\\_in\\_Images\\_A\\_Survey](https://www.researchgate.net/publication/3193340_Detecting_Faces_in_Images_A_Survey).
- Montero, R. Santiago, J. M. López Márquez y M. Ornelas Rodríguez (nov. de 2011). «Aplicando filtros de Gabor a segmentación de caracteres alfanuméricos en placas vehiculares». En: págs. 1-5. URL: [https://www.researchgate.net/publication/291822075\\_Aplicando\\_filtros\\_de\\_Gabor\\_a\\_segmentacion\\_de\\_caracteres\\_alfanumericos\\_en\\_placas\\_vehiculares](https://www.researchgate.net/publication/291822075_Aplicando_filtros_de_Gabor_a_segmentacion_de_caracteres_alfanumericos_en_placas_vehiculares).
- Nisha y Sandeep Dahiya (feb. de 2015). «Face Detection and Expression Recognition using Neural Network Approaches». En: *International Journal of Computer Theory and Engineering* 15.1, págs. 47-51. URL: [https://globaljournals.org/GJCST\\_Volume15/3-Face-Detection-and-Expression.pdf](https://globaljournals.org/GJCST_Volume15/3-Face-Detection-and-Expression.pdf).
- Nisha Soni, Mahendra Kumar y Garima Mathu (ago. de 2013). «Face Recognition using SOM Neural Network with Different Facial Feature Extraction Techniques». En: *International Journal of Computer Applications (0975 – 8887)* 76.3, págs. 7-11. URL: [https://www.researchgate.net/profile/Mahendra-Kumar-8/publication/260833478\\_Face\\_Recognition\\_using\\_SOM\\_Neural\\_Network\\_with\\_Different\\_Facial\\_Feature\\_Extraction\\_Techniques/links/00b7d52198bbf6f818000000/Face-Recognition-using-SOM-Neural-Network-with-Different-Facial-Feature-Extraction-Techniques.pdf](https://www.researchgate.net/profile/Mahendra-Kumar-8/publication/260833478_Face_Recognition_using_SOM_Neural_Network_with_Different_Facial_Feature_Extraction_Techniques/links/00b7d52198bbf6f818000000/Face-Recognition-using-SOM-Neural-Network-with-Different-Facial-Feature-Extraction-Techniques.pdf).
- Paul Viola y Michael J. Jones (jul. de 2004). «Robust Real-Time Face Detection». En: *International Journal of Computer Vision*, págs. 137-154. URL: <https://web.archive.org/web/20190202042433/http://www.vision.caltech.edu/html-files/EE148-2005-Spring/pprs/viola04ijcv.pdf>.
- Sehgal, Isha y Prof. K. S. Venkatesh (ago. de 2019). «CONNECTED COMPONENT LABELING FOR BINARY IMAGES.» En: *International Journal of Advanced Research*, págs. 916-917. URL: [https://www.researchgate.net/publication/336044568\\_CONNECTED\\_COMPONENT\\_LABELING\\_FOR\\_BINARY\\_IMAGES](https://www.researchgate.net/publication/336044568_CONNECTED_COMPONENT_LABELING_FOR_BINARY_IMAGES).
- Shahrin Azuan Nazeer Nazaruddin Omar, Khairol Faisal Jumari y Marzuki Khalid (2007). «FACE DETECTION USING ARTIFICIAL NEURAL NETWORK APPROACH». En: págs. 1-7. URL: [https://www.researchgate.net/publication/221296288\\_Face\\_Detecting\\_Using\\_Artificial\\_Neural\\_Network\\_Approach/link/575b355408aed884620d9bf2/download](https://www.researchgate.net/publication/221296288_Face_Detecting_Using_Artificial_Neural_Network_Approach/link/575b355408aed884620d9bf2/download).
- Shetty, Anirudha B et al. (ago. de 2021). «Facial recognition using Haar cascade and LBP classifiers». En: *Global Transitions Proceedings* 2, págs. 330-335. URL: [https://www.sciencedirect.com/science/article/pii/S2666285X21000728?ref=pdf\\_download&fr=RR-2&rr=82e0b070296e1549](https://www.sciencedirect.com/science/article/pii/S2666285X21000728?ref=pdf_download&fr=RR-2&rr=82e0b070296e1549).
- Shivesh Bajpai, Amarjot Singh y K. V. Karthik (feb. de 2013). «An Experimental Comparison of Face Detection Algorithms». En: *International Journal of Computer Theory and Engineering* 5.1, págs. 47-51. URL: <http://www.ijcte.org/papers/644-W00073.pdf>.
- Skelly, J. y S. Latifi (2019). «Efficiency and Accuracy Improvement to the Viola-Jones Object Detection Algorithm Using Machine Learning Methods». En: págs. 1-18. URL: [https://cmosedu.com/jbaker/students/james\\_s/ECG703\\_SKELLY\\_Final\\_Report.pdf](https://cmosedu.com/jbaker/students/james_s/ECG703_SKELLY_Final_Report.pdf).
- Szlavik, Zoltan y Tamas Szirany (2003). «Face Identification With CNN-UM». En: 1. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.107.6837&rep=rep1&type=pdf>.



- Tarun Kumar, Kushal Veer Singh y Shekhar Malik (ene. de 2011). «Artificial Neural Network in Face Detection». En: *International Journal of Computer Applications* 14.3. URL: <https://pdfs.semanticscholar.org/0a27/1e0c14ce95256c8dd98b9bb1534c36a1c9f3.pdf>, págs. 1-3.
- Tikoo, Smriti y Nitin Malik (2014). «Detection of Face using Viola Jones and Recognition Using Back Propagation Neural Network». En: págs. 1-10. URL: <https://arxiv.org/ftp/arxiv/papers/1701/1701.08257.pdf>.
- Tin, Hlaing Htake Khaung (feb. de 2012). «Robust Algorithm for Face Detection in Color Images». En: *I.J.Modern Education and Computer Science*, págs. 31-37. URL: [https://www.researchgate.net/publication/263547528\\_Robust\\_Algorithm\\_for\\_Face\\_Detection\\_in\\_Color\\_Images](https://www.researchgate.net/publication/263547528_Robust_Algorithm_for_Face_Detection_in_Color_Images).
- V. V. Khryashchev, A. A. Lebedev y A. L. Priorov (mayo de 2017). «ENHANCEMENT OF FAST FACE DETECTION ALGORITHM BASED ON A CASCADE OF DECISION TREES». En: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 6, págs. 237-241. URL: [https://www.researchgate.net/publication/316852638\\_ENHANCEMENT\\_OF\\_FAST\\_FACE\\_DETECTION\\_ALGORITHM\\_BASED\\_ON\\_A\\_CASCADE\\_OF\\_DECISION\\_TREES](https://www.researchgate.net/publication/316852638_ENHANCEMENT_OF_FAST_FACE_DETECTION_ALGORITHM_BASED_ON_A_CASCADE_OF_DECISION_TREES).
- Zeynep Orman, Abdulkadir Battal y Erdem Kemer (ago. de 2011). «A STUDY ON FACE, EYE DETECTION AND GAZE ESTIMATION». En: *International Journal of Computer Science an Engineering Survey* 2.3, págs. 29-46. URL: [https://www.researchgate.net/publication/264878015\\_A\\_Study\\_On\\_Face\\_Eye\\_Detection\\_And\\_Gaze\\_Estimation](https://www.researchgate.net/publication/264878015_A_Study_On_Face_Eye_Detection_And_Gaze_Estimation).
- Zhu, Xiangxin y Deva Ramanan (2011). «Face Detection, Pose Estimation, and Landmark Localization in the Wild». En: págs. 1-8. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.227.5872&rep=rep1&type=pdf>.