



Universidad Autónoma de Querétaro
Facultad de Ingeniería
Maestría en Ciencias Instrumentación y Control Automático

Localización y verificación de rostros mediante de procesamiento de imágenes y herramientas de Morfología Matemática

TESIS

Que como parte de los requisitos para obtener el diploma/grado de
Maestro en Ciencias en Instrumentación y Control Automático

Presenta:

Ma. Isabel Morales Carrillo

Dirigido por:

Dr. Damián Vargas Vázquez

SINODALES

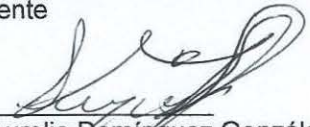
Dr. Damián Vargas Vázquez
Presidente


Dr. Jorge Domingo Mendiola Santibañez
Secretario

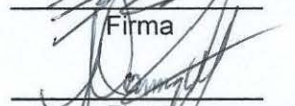
Dr. Alberto Pastrana Palma
Vocal


Dr. Edgar Alejandro Rivas Araiza
Suplente

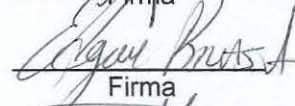
Dr. Jesús Carlos Pedraza Ortega
Suplente



Dr. Aurelio Domínguez González
Director de la Facultad



Firma


Firma


Firma


Firma


Firma


Firma

Dr. Irineo Torres Pacheco
Director de Investigación y
Posgrado

Centro Universitario
Querétaro, Qro.
Diciembre de 2014
México

RESUMEN

El reconocimiento facial ha sido un tema estudiado, especialmente en los últimos 20 años y se han desarrollado sistemas de reconocimiento donde identifican o autentifican a una persona mediante imágenes faciales que previamente se han analizado y almacenado. El objetivo de este trabajo de investigación es detectar y autentificar rostros de una base de datos desarrollada en la UAQ mediante filtros de Gabor y un perceptrón multicapa, con condiciones controladas como son la iluminación, expresión facial y pose. Para llevar a cabo la verificación(o autentificación) facial se precisa detectar el rostro dentro de una imagen dada, lo cual se realiza mediante el color de la piel, donde se utiliza el espacio de color YCbCr considerando umbrales en las componentes Cb y Cr. Después de la detección y localización del rostro, las características de la imagen facial son extraídas usando uno de los métodos más utilizados para el reconocimiento facial basado en las llamadas características Gabor. Mediante un proceso de entrenamiento el sistema aprende a reconocer a cada persona y rechaza a los impostores, donde es implementado un perceptrón multicapa con entrenamiento de retro-propagación para la clasificación de los vectores característicos de cada persona. Los resultados demuestran una eficacia de las técnicas implementadas con una tasa de clasificación del 94.41% considerando las imágenes de la base de datos obtenida.

(Palabras clave: Segmentación, Color de piel, Filtros de Gabor, Perceptrón Multicapa, retropropagación)

SUMMARY

Face recognition has been a topic studied, especially in the last 20 years and has developed recognition systems which identify or authenticate a person using facial images that have previously been analyzed and stored. The objective of this research is to detect and authenticate faces of a database developed in the UAQ using Gabor filters and multilayer perceptron with controlled conditions such as lighting, facial expression and pose. To perform the verification (or authentication) is necessary to detect the face in a given image, which the skin color is used in the YCbCr color space given thresholds in Cb and Cr components. After detection and face location, the features of the facial image are extracted using one of the most methods used for face recognition based on Gabor features called. Through a process of training the system learns to recognize each person and reject impostors, where it is implemented a multilayer perceptron with back-propagation training for classifying the feature vectors of each person. The results demonstrated the effectiveness of techniques implemented with a classification rate of 94.41% using the images of the database obtained.

(Key words: Segmentation, Skin color, Gabor filters, Multilayer Perceptron, backpropagation)

DEDICATORIA

Dedico esta tesis a mis padres Enriqueta Carrillo Trinidad y Pedro Morales Ramírez por darme educación a pesar de los obstáculos y por brindarme su apoyo en cada etapa de mi vida.

A mis hermanos Rosalino, Jesús y Juan Carlos por la motivación que siempre me han brindado.

AGRADECIMIENTOS

Primeramente quiero darle gracias a Dios por darme salud, oportunidad de cumplir cada objetivo. También quiero agradecer a mi familia por todo el apoyo incondicional en mis decisiones, por sus consejos, motivación, gracias.

Agradezco a mi asesor de tesis el Dr. Damián Vargas Vázquez por darme la oportunidad de trabajar con él en este proyecto, por su apoyo y dedicación.

A mis sinodales por el apoyo en el desarrollo de la tesis, por las sugerencias, observaciones, correcciones, en especial al Dr. Jorge Mendiola Santibáñez y al Dr. Alberto Pastrana Palma.

Al Jefe de la División de Investigación y Posgrado Dr. Manuel Toledano Ayala y al coordinador de la Maestría el Dr. Edgar Rivas Araiza por el apoyo y dedicación hacia los estudiantes.

A CONACyT por la beca otorgada para la realización de la maestría en la Universidad Autónoma de Querétaro.

A Fondo para el Fortalecimiento de la Investigación (FOFI) por el apoyo otorgado en el 2014.

ÍNDICE

	Página
Resumen	i
Summary	ii
Dedicatoria	iii
Agradecimientos	iv
Índice	v
Índice de cuadros	vii
Índice de figuras	viii
I. INTRODUCCIÓN	1
Hipótesis	3
Objetivo general	3
Objetivos específicos	3
II. REVISIÓN DE LITERATURA	5
ANTECEDENTES	5
Biometría facial	5
Sistema de reconocimiento facial	7
Modos de operación y análisis de errores del reconocimiento facial	8
Desafíos de un sistema de reconocimiento facial	10
Clasificación de métodos de reconocimiento facial General	11
Métodos de detección de rostros a través del color de piel	12
Métodos de reconocimiento facial basados en Gabor	14
FUNDAMENTACIÓN TEÓRICA	18
Detección del rostro	18
Espacio de color HSI	21
Transformación de RGB a HSI	22
Espacio de color YCbCr	22
Transformación de RGB a YCbCr	23
Segmentación	23
Morfología matemática	24
Filtros por reconstrucción	26

Apertura viscosa	27
Plantilla facial	29
Correlación normalizada	30
Filtros de Gabor	31
Perceptrón multicapa	33
Tipos de función de activación del perceptrón	37
Ecuaciones de entrenamiento del perceptrón multicapa con una sola salida	37
III. METODOLOGÍA	41
Adquisición de imágenes	41
Creación de la base de datos	42
Detección del rostro	43
Clasificación de píxeles de color	43
Pre-procesamiento de la imagen	45
Segmentación del rostro evitando el cuello por medio de apertura viscosa	48
Segmentación del rostro evitando el cuello por medio de proporciones	51
Normalización de las imágenes faciales	53
Comprobación del rostro detectado o candidato facial	53
Extracción de características basado en Gabor	54
Obtención del vector de características	57
Entrenamiento del perceptrón multicapa	61
IV. RESULTADOS Y DISCUSIÓN	62
Detección de color de piel	62
Recorte del rostro por medio de apertura viscosa	63
Detección del rostro por medio de proporciones más apertura viscosa	63
Evaluación de rostros candidatos mediante una plantilla facial	64
Verificación de rostros mediante perceptrón multicapa	66
Evaluación del algoritmo en condiciones no controladas	73
Conclusiones	79
LITERATURA CITADA	81

ÍNDICE DE CUADROS

Cuadro		Página
3.1	Parámetros del filtro de Gabor utilizados.	58
3.2	Valores de los parámetros del perceptrón multicapa.	61
4.1	Resultados de confirmación de rostros con plantilla facial.	65
4.2	Resultados de la verificación con pose frontal y condiciones de iluminación controlada.	71
4.3	Resultados de la verificación con pose frontal y diferentes condiciones de iluminación.	77
4.4	Resultados de la verificación bajo condiciones de iluminación controlada y con diferentes poses.	78

ÍNDICE DE FIGURAS

Figura		Página
2.1	Difusión de los rasgos biométricos más populares (international biometric group's, "biometric market and industry report 2007-2012").	6
2.2	Diagrama a bloques de un sistema de reconocimiento facial general.	7
2.3	Comparación de la respuesta de las funciones de Gabor con los campos receptivos (células simples v1).	16
2.4	Fuentes que causan variación facial.	19
2.5	Ejemplos de elementos de estructura.	25
2.6	Apertura viscosa. a) Imagen original, b),c),d),e) y f) Componentes erosionados conectados a viscosidad $\lambda = 20$, $\lambda = 22$, $\lambda = 27$, $\lambda = 36$ y $\lambda = 39$ respectivamente. Santillán et al.(2010).	28
2.7	Diagrama del filtrado Gabor. Ilonen et al. (2007)	33
2.8	Modelo de una neurona artificial.	34
2.9	Tipos de función de activación.	37
3.1	Ejemplos de imágenes de la base de la base de datos obtenida.	41
3.2	Ejemplos de imágenes de la base de datos GTAV.	43
3.3	Gráfica de densidad de color de piel en CbCr.	44
3.4	Clasificación de píxeles de color de piel. a) Imagen original, b) Resultado de color de piel en espacio de color hsi, c) Resultado de clasificación de color de piel en el espacio de color YCbCr.	45
3.5	Resultado de eliminación de objetos dentro del rostro.	46
3.6	Imagen después de una apertura morfológica.	46

3.7	Imagen de eliminación de objetos de áreas pequeñas.	47
3.8	Imagen resultante de la detección de color de piel.	47
3.9	Proceso de erosionado de la imagen. a) Imagen de entrada, b) Erosión1 de la imagen a y c) Erosión 2 de la imagen a.	48
3.10	Resultado de la apertura viscosa, a) Máscara obtenida, b) Rostro con colores de la imagen original.	49
3.11	Coordenadas x e y para el recorte del rostro.	49
3.12	Recorte del rostro mediante apertura viscosa.	50
3.13	Imagen resultante de la apertura viscosa agregando una apertura morfológica.	50
3.14	Imagen resultante del recorte del rostro de la apertura viscosa agregando una apertura morfológica.	51
3.15	Imágenes resultantes de la apertura viscosa junto con la apertura morfológica.	51
3.16	Resultado del recorte del rostro por medio de proporciones.	52
3.17	Plantilla facial.	53
3.18	Kernels de Gabor a diferentes escalas (frecuencias) y orientaciones parte real en el dominio espacial.	54
3.19	Parte real del filtro gabor en 3d, orientación: 45° , frecuencia: 0.1053, $\gamma = \sigma = \sqrt{2}$.	55
3.20	Parte imaginaria del filtro gabor en 3d, orientación: 45° , frecuencia: 0.1053, $\gamma = \sigma = \sqrt{2}$.	55
3.21	Imagen original en escala de grises.	56
3.22	Imágenes de magnitud resultantes de la convolución.	57
3.23	Imagen facial dividida en secciones.	58
3.24	Vector de características estimado de la imagen 3.23	60
3.25	Esquema del perceptrón multicapa con una sola salida.	60

4.1	Detección de color de piel. a) y c) Imágenes de entrada, b) y d) Imágenes resultantes de la detección de color de piel.	62
4.2	Uso de la apertura viscosa. a) y c) Imagen de entrada, b) y d) Imagen facial como resultado de la apertura viscosa.	63
4.3	Resultado de recorte por proporciones más viscosa. a) y c) Imagen original, b) y d) Imagen de rostro recortado.	64
4.4	Diagrama general de la etapa de verificación de rostros.	66
4.5	Imagen facial del usuario 15.	66
4.6	Vector de características de la figura 4.5.	67
4.7	Resultado de entrenamiento del perceptrón para la persona con identificación número 15.	68
4.8	Gráfica del error.	69
4.9	Resultado del usuario 15 con los vectores de prueba	69
4.10	Resultado de la salida del perceptrón respecto a los vectores intrusos.	70
4.11	Ejemplos de la base de datos GTAV bajo diferentes condiciones de iluminación, a) Fuente de luz natural, b) Fuente de luz fuerte y c) Fuente de luz medio fuerte.	73
4.12	Ejemplos de la base de datos FEI con diferentes poses.	74
4.13	Segmentación y recorte por medio de apertura viscosa.	74
4.14	Segmentación por medio de apertura viscosa, a) Imagen original y b) Imagen facial segmentado	75
4.15	Segmentación por medio de apertura viscosa, a) Imagen con iluminación compensada y b) Imagen facial segmentado	76
4.16	Segmentación y recorte por medio de apertura viscosa de imágenes de base de datos FEI.	76

I. INTRODUCCIÓN

El reconocimiento facial ha sido un campo de investigación activa en visión por computadora (Jafri y Arabnia, 2009), reconocimiento de patrones (Liu y Wechsler, 2002, Zhao et al., 2003), neurociencia (Liu et al., 2002, Tsao et al., 2006, Atkinson y Adolphs, 2011) y psicología (Pinsk et al., 2005, Barraclough y Perrett, 2011, Heron-Delaney et al., 2011) motivada no solo por los desafíos sino también por su amplio rango de aplicaciones prácticas en control de acceso, sistemas de identificación, video vigilancia, redes sociales, etc. y por la extraordinaria capacidad del ser humano de reconocer a las personas; aunado a lo anterior, la actividad humana es la principal preocupación tanto en la vida social como en el ciberespacio, pues el ser humano puede recordar cientos de rostros en toda su vida e identificarlos a pesar de grandes cambios en los estímulos visuales debido a condiciones ambientales, edades y distractores tales como: los lentes, cambios en el estilo de peinado, etc. Esta aparente simplicidad realizada por el ser humano no es una tarea fácil de trasladar a un sistema de reconocimiento automático. Por otro lado, un aspecto importante a considerar es que el cerebro humano tiene sus limitaciones en el número total de personas que puede recordar de manera precisa, una ventaja clave de un sistema de computadora es su capacidad de mantener un gran número de imágenes faciales.

El reconocimiento facial tiene varias ventajas sobre otras modalidades biométricas como huellas dactilares y el iris, además de ser natural y no intrusiva, la ventaja más importante es que la imagen facial se puede capturar a distancia y de manera encubierta.

Actualmente, los sistemas de reconocimiento facial tienen altos porcentajes de reconocimiento (entre 80 y 95), sin embargo no se ha logrado desarrollar un sistema robusto capaz de tener un 100% de efectividad, debido a las múltiples variables intrínsecos (tales como la edad, expresión facial,

parafernalia facial) y extrínsecos (tales como los cambios de iluminación, pose, escala) (Yang et al. 2002, Stan y Jain, 2011). Los sistemas de reconocimiento facial son un tipo de sistemas de reconocimiento biométrico, el cual se lleva a cabo gracias a un sistema automático basado en técnicas de inteligencia artificial y reconocimiento de patrones, que permiten el reconocimiento de una persona a partir de rasgos faciales.

Un sistema de reconocimiento facial tiene tres etapas fundamentales para realizar dicha función, estos son; la detección del rostro, la extracción de características y el reconocimiento (verificación o identificación).

Para la detección de rostros, el color ha sido una característica interesante y popular. A menudo se utiliza como un paso de pre-procesamiento para seleccionar las regiones de interés. Por otro lado, los filtros de Gabor han demostrado ser una poderosa herramienta para la extracción de características faciales, la principal razón que llamó la atención fue la similitud entre los filtros de Gabor y el campo receptivo de las células simples en la corteza visual (Daugman, 1985). Una razón más práctica es su éxito en muchas aplicaciones, por ejemplo, la detección y reconocimiento de rostros, reconocimiento del iris y la huella digital, en los que los métodos basados en función de Gabor se encuentran entre los de mejor desempeño (Messer et al, 2003). Finalmente, el reconocimiento depende del método de clasificación. La clasificación es una de las tareas más frecuentemente encontrados de toma de decisiones de la actividad humana. Un problema de clasificación se produce cuando un objeto necesita ser asignado a un grupo predefinido o clase sobre la base de un número de atributos observados relacionados con ese objeto. Las redes neuronales se han convertido en una herramienta importante para la clasificación. Las actividades de investigación recientes en la clasificación neuronal han establecido que las redes neuronales son una buena alternativa a los diversos métodos de clasificación convencionales.

En el presente trabajo de investigación, se trata de realizar la verificación de rostros con una base de datos desarrollada en la UAQ. Para llevar a cabo la verificación facial se precisa detectar el rostro dentro de una imagen dada, lo cual se realiza mediante el color de la piel, donde se utiliza el espacio de color YCbCr considerando umbrales en las componentes Cb y Cr. Después de la detección y localización del rostro, las características de la imagen facial son extraídas usando uno de los métodos más utilizados para el reconocimiento facial basado en las llamadas características Gabor. Mediante un proceso de entrenamiento el sistema aprende a reconocer a cada persona y rechaza a los impostores, donde es utilizado un perceptrón multicapa con entrenamiento de retro-propagación para la clasificación de los vectores característicos de cada persona.

Hipótesis

Por medio de técnicas de procesamiento de imágenes de detección del rostro, extracción de características y entrenamiento de aprendizaje es posible realizar la verificación de identidad de usuarios.

Objetivo general

Implementar un conjunto de herramientas para verificación de rostros a partir de una base de datos, usando el color de la piel en conjunción con técnicas de morfología matemática para detección del rostro, filtros de Gabor para extracción de rasgos distintivos, y redes neuronales para la validación del resultado.

Objetivos específicos

- Implementar un algoritmo para segmentar el rostro dentro de una imagen mediante el uso de color de la piel.

- Implementar una técnica para realizar el recorte dejando solo la imagen del rostro para su posterior análisis.

- Implementar un algoritmo para extraer características o rasgos distintivos del rostro para asignarle métricas que lo identifiquen en una base de datos de personas autorizadas.

- Implementar un algoritmo para entrenamiento y reconocimiento de los vectores característicos para la verificación de rostros.

- Obtener un prototipo de prueba para identificar los puntos de más interés para una futura investigación.

II. REVISIÓN DE LITERATURA

ANTECEDENTES

En este capítulo se realiza una revisión del estado del arte referente al tema del reconocimiento facial basado en Gabor. Como de partida, se explicarán las ideas básicas de la biometría facial, para posteriormente explicar las características de un sistema biométrico facial. También se resumen los métodos reconocimiento facial en general, poniendo especial énfasis en los basados filtros de Gabor.

Biometría facial

La biometría se refiere a todo conjunto de técnicas complejas para reconocer la identidad de una persona por medio de sus rasgos fisiológicos (patrón del iris, el rostro, huellas, geometría de la mano, etc.) o conductuales (por ejemplo, la manera de firmar o la voz). Estos rasgos biométricos son universales, distintivos, característicos, invariantes sobre un periodo de tiempo, además medibles de manera cuantitativa y objetiva (Jain, 2004).

Las tecnologías biométricas ofrecen altos niveles de seguridad, asegurándose que el individuo autorizado esté presente físicamente para permitir el acceso. Las consecuencias de la autenticación fraudulenta en entornos financieros, bancarios, minoristas y corporativos pueden ser desastrosos con la pérdida de información confidencial, el dinero y la integridad de los datos comprometidos.

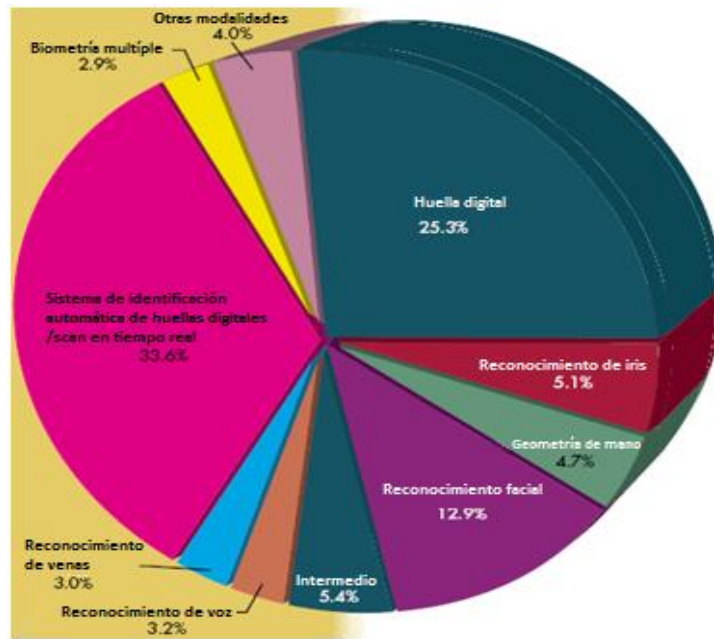


Figura 2.1. Difusión de los rasgos biométricos más populares (International Biometric Group's, "Biometric Market and Industry Report 2007-2012").

El reconocimiento facial ofrece varias ventajas sobre otros métodos biométricos, puesto que casi todas estas tecnologías requieren alguna acción voluntaria por parte del usuario; es decir, el usuario tiene que colocar su mano sobre un soporte para la toma de huellas dactilares o la detección de geometría de la mano, tiene que estar en una posición fija delante de una cámara por el iris o la identificación de retina. Además, múltiples individuos utilizan el mismo equipo para capturar sus características biológicas y se expone al usuario a la transmisión de gérmenes y las impurezas de otros usuarios. Sin embargo, el reconocimiento facial se puede hacer de forma pasiva sin ninguna acción explícita o participación por parte del usuario, las imágenes faciales se pueden adquirir desde una distancia por una cámara. Esto es particularmente beneficioso para los propósitos de seguridad y vigilancia. Por lo tanto, el reconocimiento facial es totalmente no invasivo y no lleva ningún tipo de riesgos para la salud.

Sistema de reconocimiento facial

Un sistema de reconocimiento facial general consiste en tres etapas principales como se muestra en la figura 2.2.

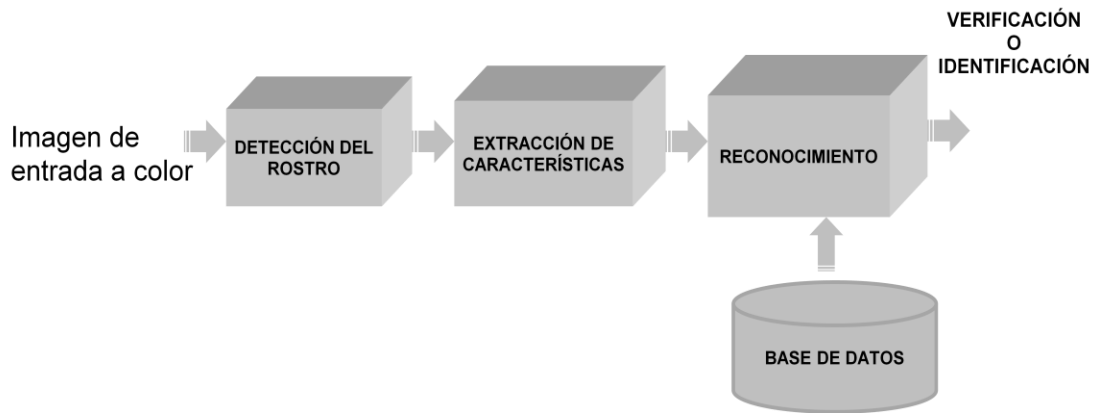


Figura 2.2. Diagrama a bloques de un sistema de reconocimiento facial general.

La detección del rostro en imagen consiste en detectar en una imagen el rostro (o los rostros en caso de ser varias personas), segmentar el área del rostro del fondo y localizarlo con un grado de precisión.

El objetivo de la extracción de características es encontrar una representación específica de los datos que pueden resaltar la información relevante. Esta representación se puede encontrar mediante la maximización de un criterio o puede ser una representación pre-definido. Por lo general, una imagen del rostro está representada por un vector de gran dimensión que contiene valores de píxeles (representación holística) o un conjunto de vectores, donde cada vector resume el contenido subyacente de una región local mediante el uso de una transformación de alto nivel (representación local) (Sarfranz, et al., 2010).

La extracción de características facial consiste en la localización de los componentes más característicos del rostro (ojos, nariz, boca, etc.) dentro de las imágenes que representan rostros humanos. Este paso es esencial para la inicialización de muchas técnicas de procesamiento; como el seguimiento del rostro, reconocimiento de la expresión facial o el reconocimiento de rostros.

La clasificación trata de asignar las diferentes partes del vector de características a grupos o clases, basándose en las características extraídas. En esta etapa se usa lo que se conoce como aprendizaje automático, cuyo objetivo es desarrollar técnicas que permitan a las computadoras aprender (Russ, 2007).

Modos de operación y análisis de errores del reconocimiento facial

El sistema de reconocimiento facial es una aplicación que identifica automáticamente a una persona en una imagen digital. Esto es posible mediante un análisis de las características faciales del sujeto extraídas de la imagen o de una secuencia de video, y las compara con una base de datos. Esto puede ser operado en dos modos, verificación de rostros(o autenticación) e identificación de rostros (reconocimiento) (Jain, 2004).

La verificación de rostros implica el emparejamiento (*match*) de datos uno a uno, en el cual se compara la imagen facial contra una plantilla del rostro cuya identidad está siendo reclamada (Problema 1:1). Para ello, el usuario proporciona una muestra de sus datos biométricos X_Q , además de una identidad reclamada I

El sistema compara estos datos con la plantilla X_I correspondiente a esa identidad. Si el parecido entre ambos datos según una similitud S excede un cierto umbral t , entonces la identidad del usuario es verificada y considerada como I . En caso contrario, el sistema rechaza la identidad reclamada. La regla de decisión para verificación es así (Serrano, 2009):

$$\begin{cases} X_Q \in I, si & S(X_Q, X_I) \geq t \\ X_Q \notin I, en otro caso \end{cases} \quad (2.01)$$

La identificación de rostros implica uno a muchos emparejamientos que compara una imagen facial contra todas las plantillas en la base de datos para determinar la identidad del rostro (problema 1: N). El usuario solo proporciona sus

datos biométricos X_Q , el sistema tiene que compararlos con todas las plantillas X_{IK} almacenadas en la base de datos correspondientes a N usuarios conocidos. Entonces se busca la identidad I_K del usuario cuya plantilla se ajuste a los datos con la máxima similitud S . Si esta similitud supera un cierto umbral t , entonces los datos de entrada se asignan a la identidad I_K , de otra manera, el sistema no puede asignar X_Q a ninguna identidad de la base de datos, la regla de decisión para la identificación es (Serrano, 2009):

$$\begin{cases} X_Q \in I_k, si \max_k [S(X_Q, X_{Ik})] \geq t & k = 1, 2, 3, \dots, N \\ X_Q \notin \text{cero}, en \text{ otro caso} \end{cases} \quad (2.02)$$

Ahora bien, teniendo claro los modos de reconocimiento facial, se procede a evaluar el sistema de verificación mediante la siguiente clasificación de usuarios. A los usuarios conocidos o clientes se les llama ‘verdaderos positivos’ (TP), sin embargo podría haber también algunos impostores que superen el umbral a los cuales se les llama o etiqueta como ‘falsos positivos’ (FP). Por otro lado, algunos de los usuarios conocidos o clientes podrían no conseguir exceder el umbral establecido y a estos usuarios se les etiqueta como ‘falsos negativos’ (FN). Finalmente se espera que los impostores produzcan un valor de similitud muy por debajo del umbral, y a estos se les etiqueta como ‘verdaderos negativos’ (TN).

Dos indicadores relacionados con los errores (FAR y FRR) se utilizan para medir el desempeño del sistema de verificación. La tasa de aceptación falsa, o FAR, es la medida de la probabilidad de que el sistema de seguridad biométrico aceptará incorrectamente un intento de acceso a un usuario no autorizado. La tasa de falso rechazo, o FRR, es la medida de la probabilidad de que el sistema de seguridad biométrico rechazará incorrectamente un intento de acceso de un usuario autorizado.

Analizando la cantidad de FP y FN, se definen las siguientes tasas de error (Messer et al., 2003, Wechsler, 2007):

Índice de falsa aceptación (FAR), proporción de impostores tomados por clientes:

$$FAR = \frac{FP}{TN + FP} \quad (2.03)$$

Índice de falso rechazo (FRR), proporción de clientes tomados por impostores:

$$FRR = \frac{FN}{TP + FN} \quad (2.04)$$

La precisión o tasa de reconocimiento es la proporción de verificaciones correctas considerando tanto los TP como TN con respecto al total de la población:

$$precisión = \frac{TP + TN}{TP + FN + TN + FP} \quad (2.05)$$

La tasa total de error (TER) es la suma de FAR y FRR.

Desafíos de un sistema de reconocimiento facial

Los desafíos asociados tanto para la detección del rostro como el reconocimiento facial (Gong et al., 2000; Yang et al. 2002; Abate et al., 2007), los cuales hacen que la apariencia del rostro varíe se muestran a continuación:

- Pose. Las imágenes de un rostro varía debido a la pose cámara-rostro (frontal, perfil, 45 grados, etc.) donde algunas características como un ojo o nariz podrían estar parcialmente ocluidos.

- Iluminación. Existen variaciones de iluminación debido a propiedades de reflectancia de la piel y al control de la cámara interno. Varios métodos en 2D realizan correctamente tareas de reconocimiento sólo bajo la variación de iluminación moderada, mientras el funcionamiento perceptiblemente cae cuando existen cambios de iluminación y postura.
- Tiempo. Otro factor importante es el retardo de tiempo, el rostro cambia con el tiempo, de una manera no lineal durante largos períodos.
- Presencia o ausencia de componentes estructurales. Las características faciales tales como, el bigote, barba y lentes podrían estar o no presentes con diferentes formas, color y tamaño.
- Expresión facial. La apariencia de los rostros son directamente afectados por la expresión facial.
- Oclusión. Los rostros podrían ser parcialmente ocluidos por otros objetos. En una imagen con un grupo de personas, algunos rostros podrían estar ocluidas por otros rostros.
- Orientación de la imagen. Las imágenes faciales varían directamente por las diferentes rotaciones alrededor del eje óptico de la cámara.

Clasificación de métodos de reconocimiento facial General

Existe gran variedad de métodos y técnicas empleadas para el reconocimiento facial, por lo tanto, dependiendo sobre la metodología de la adquisición de datos faciales (Abate et al., 2007; Jafri y Arabnia, 2009) y de cómo los seres humanos usan características holísticas y locales (Zhao et al., 2003) las técnicas se dividen en tres categorías en general:

- Reconocimiento basado en intensidad de imágenes
 - Métodos holísticos: estos métodos usan toda la región del rostro como entrada al sistema de reconocimiento (Turk y Pentland, 1991; Liu et al., 2004; Nhat y Lee, 2007; Sun et al., 2008).

- Métodos basados en características. Típicamente, en estos métodos, las características locales tales como, los ojos, nariz y boca son extraídas, su ubicación y estadísticas locales (apariciencia geométrica) se introducen a un clasificador estructural (Wiskott et al., 1999; Gottumukkal y Asari, 2004; Campadelli y Lantarotti, 2005; Shin et al., 2007; Albiol et al., 2008).
- Métodos híbridos. Como el sistema de percepción del ser humano usa ambas características tanto locales como globales para reconocer un rostro un sistema de reconocimiento automático debería usar ambos (Wan et al., 2004).
- Reconocimiento basado en secuencia de video (Zhou et al., 2002).
- Modelos basados en 3D (Bronstein et al., 2005; Lu y Jain, 2006; Zhou et al., 2008; Berreti et al., 2007).

Los métodos mencionados anteriormente se basan de reconocimiento facial en general, el sistema de reconocimiento facial tiene tres etapas claves como se ha mencionado: detección del rostro, extracción de características y reconocimiento (identificación o verificación). El presente trabajo de investigación se centra en detección de color de piel para la detección del rostro, los filtros de Gabor para la extracción de características y un perceptrón multicapa como clasificador para la verificación, a continuación se describirán más a fondo algunos antecedentes de los mismos.

Métodos de detección de rostros a través del color de piel

La clave en el reconocimiento de rostros es considerar el objeto de estudio; el rostro que a pesar de tener variaciones métricas mínimas es indispensable segmentarlo de otros objetos que puedan ser confundidos con el rostro y que conduzca al sistema a fallar. Un punto importante en el análisis de imágenes es la correcta segmentación del objeto de interés dentro de una imagen ya que determina la precisión de las etapas posteriores.

El color es considerado como un indicativo o descriptor importante y útil para detección de objetos, segmentación de imágenes, etc. La detección de color de piel juega un papel importante en un rango amplio de aplicaciones de procesamiento de imágenes; como el rastreo y detección del rostro, análisis de gestos y es usado como una etapa preliminar en el reconocimiento de rostros. Entre los espacios de color más utilizados para la detección de color de piel se encuentran, HSI, HSV, YCbCr, CIELab, entre otros.

a) Hsu et al. (2002) propusieron un algoritmo de detección de rostros, donde realizan una técnica de compensación de iluminación, la detección de color de piel se realiza mediante el espacio YCbCr, las características faciales como los ojos y la boca son localizadas utilizando los componentes de color, generando rostros candidatos, posteriormente checan la simetría de las localizaciones de las características y con un cierto umbral validan si hay un rostro o no con un porcentaje de 90.74%.

b) Singh et al. (2003) realizaron una técnica de detección de rostros utilizando una combinación de tres espacios de color RGB, YCbCr y HSI para generar un tercer modelo de color. Las características faciales son segmentadas en escala de grises considerando la simetría de los rasgos y zonas oscuras (mínimos del rostro), de donde concluyeron que los espacios de color YCbCr y HSI son más eficientes que el RGB y obtienen un porcentaje de detección del 95.18%.

c) Kim et al. (2004) desarrollaron un método de reconocimiento facial basado en imágenes de secuencia de video, utilizando el espacio de color YCbCr, para la detección del rostro y el PCA para el reconocimiento.

d) Kherchaoui y Houacine (2010) implementaron un algoritmo de detección de rostro que consiste en dos partes, la primera consiste en detectar zonas de color de piel utilizando el espacio de color YCbCr y en la segunda parte

utilizan una plantilla facial para confirmar si la región detectada es rostro o no, la evaluación fue de 141 imágenes faciales con un porcentaje de rostros detectados de 77.47%.

e) Chandrappa et al. (2011) realizaron una técnica de detección de color usando YCbCr para la detección de color y una plantilla facial para determinar o comprobar si hay un rostro o no en la imagen con un porcentaje del 96.8%.

f) Sagheer y Aly (2012) realizaron la detección del rostro, usando el espacio de color YCbCr, donde obtiene dos umbrales de color de piel, uno para condiciones de iluminación controlada y otro umbral para diferentes condiciones de iluminación; por último, utilizaron una red neuronal para determinar si la región detectada es rostro o no.

g) Amjad et al. (2012) desarrollan un algoritmo de detección del rostro en base a color de piel, donde realizan un modelado del Color de piel usando histogramas de Cb y Cr para obtener umbrales de color de piel, posteriormente utilizan herramientas morfológicas y la relación de aspecto para confirmar el rostro, obtienen un porcentaje de 90.87%.

Métodos de reconocimiento facial basados en Gabor

La amplia aplicación de las funciones de Gabor también ha dado lugar a diferentes terminologías, lo que puede ser muy confuso para los investigadores. Algunos ejemplos son wavelet Gabor, filtro de Gabor, la expansión Gabor, transformada de Gabor y la función Gabor. Las wavelets Gabor se llaman filtros de Gabor en el ámbito de las aplicaciones (Shen et al., 2007).

Las funciones de Gabor proporcionan la resolución óptima en ambos dominios de tiempo y frecuencia. Y la transformada wavelet Gabor parece ofrecer

las bases óptimas para extraer características locales por varias razones (Shen y Bai, 2006):

- Motivación biológica. Las células simples de la corteza visual de los cerebros mamíferos son modeladas como una familia de Gabor wavelets 2D. Además las formas de los Gabor wavelets son similares a los campos receptivos de estas células (Daugman, 1985).
- Motivación matemática. las wavelets Gabor son óptimas para la medición de las frecuencias espaciales locales (Kruger et al., 2002a, 2002b).
- Motivación empírica. De acuerdo a algunas pruebas de rendimiento (evaluación FERET, (Phillips et al., 2000) y la competencia FVC2004 (Messer et al., 2003)) se ha encontrado que el método empleando filtros de Gabor, para la extracción de características produce un mejor rendimiento comparado con otros métodos.

La función de Gabor en 1-D fue definido por Gabor (Gabor, 1945) en su artículo teoría de la comunicación, donde propuso una representación de señales como una combinación de funciones elementales y después extendido a 2-D por Daugman (Daugman, 1985) quien afirmó que las funciones de Gabor 2-D son buenos modelos de campos receptivos de una gran cantidad de células que se ubican en la corteza visual primaria de los mamíferos.

Jones y Palmer (1987) desarrollaron una comparativa para probar la hipótesis de que los campos receptivos simples en la corteza estriada de los gatos son filtros lineales que tienen la forma funcional de los filtros de Gabor 2D y llegan a la conclusión de que la función de Gabor proporciona una descripción útil y razonablemente exacta de la mayoría de los aspectos espaciales de campos receptivos simples.

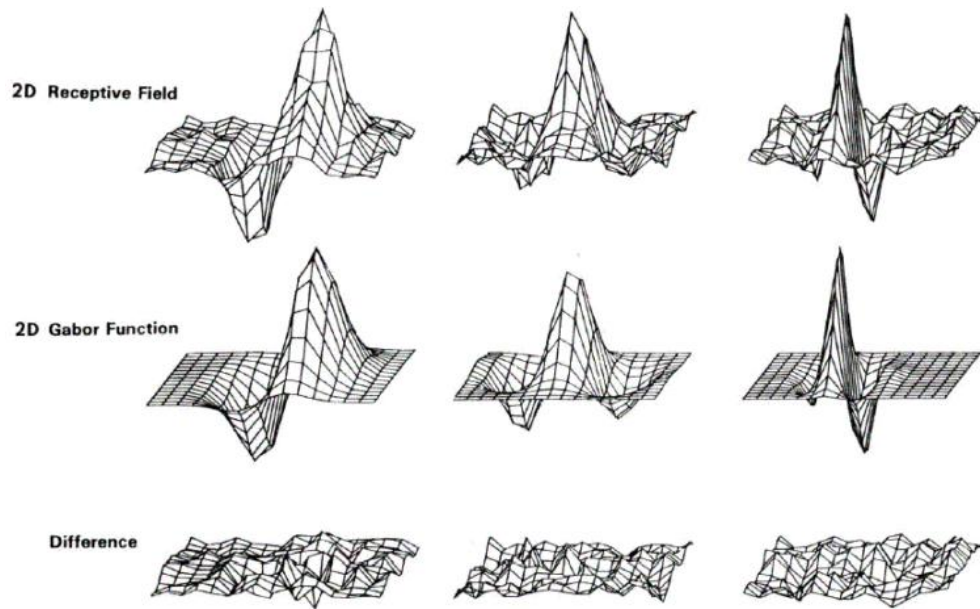


Figura 2.3. Comparación de la respuesta de las funciones de Gabor con los campos receptivos (células simples V1). Jonas y Palmer (1987).

Basado sobre estas ventajas de wavelet Gabor, ha sido usado en muchas aplicaciones de análisis de imágenes, tales como, reconocimiento facial, clasificación de textura, clasificación de expresión facial, y algunas otras excelentes investigaciones.

Las características Gabor, llamados también como Gabor jet, banco de Gabor o característica Gabor multi-resolución, son construidos de las respuestas de los filtros Gabor usando filtros sobre varias frecuencias y orientaciones. La frecuencia en este caso corresponde a la información de escala (Kamarainen, 2012).

a) Maajid et al. (2006) realizaron el reconocimiento facial utilizando 30 filtros de Gabor para extraer las características de la imagen facial, donde suman la respuesta de las magnitudes de las 30 imágenes resultantes de la convolución entre cada filtro con la imagen facial dada; para la clasificación emplean el método del vecino más cercano(nearest neighbor) obteniendo 93% de precisión.

b) Aguilar-Torres et al. (2007) implementaron el reconocimiento de rostros usando 6 frecuencias y 9 orientaciones, obteniendo 54 filtros de Gabor. Ellos dividieron la imagen original en 12x9 permitiendo reducir la cantidad de información, después realizaron una correlación cruzada entre la imagen dividida en secciones con cada uno de los filtros para generar el vector de características y utilizaron un perceptrón multicapa como clasificador con el algoritmo de entrenamiento back-propagation.

c) Shen et al. (2007) desarrollaron un método tanto para identificación como para la verificación de rostros, usando las wavelets Gabor junto con un análisis discriminante General (GDA), para ello emplearon toda la imagen del rostro para extraer las características, utilizando un banco Gabor clásico, es decir, 5 frecuencias y 8 orientaciones y las medidas de distancia en sus pruebas tales como, distancia euclidiana, Mahalanobis y correlación normalizada. Un punto importante es que observaron a partir de muestras de imágenes del rostro en el dominio de la frecuencia que la información útil está contenida principalmente en la banda de baja frecuencia. Por lo tanto, eligieron $f_{\max} = 0.25$ para el reconocimiento facial. Por otro lado, Bianconi y Fernandez (2007) en su investigación sobre los efectos de los parámetros de los filtros Gabor en la clasificación de textura mencionan que la frecuencia máxima depende del parámetro de γ donde especifican mediante una ecuación la relación de este parámetro con la frecuencia máxima.

d) Rahman y Bhuiyan (2008) en Bangladesh plantearon un método de identificación de rostros, donde todas las imágenes son procesadas con la misma iluminación y contraste RMS (Root Mean Square) equivalente a la desviación estándar de iluminación, luego utilizan los filtros de Gabor para la extracción de características con 3 frecuencias y 5 orientaciones donde cada uno se convoluciona con la imagen, y para la clasificación utilizan un perceptrón multicapa.

e) Por otro lado, los autores en Struc et al. (2009) desarrollaron nuevos filtros ortogonales construidos como una combinación lineal de los filtros originales de Gabor. Ellos emplearon 5 frecuencias y 8 orientaciones tomando la frecuencia máxima de 0.25. Dado que los nuevos filtros se derivan de las matrices de correlación de los filtros originales de Gabor por medio de análisis de componentes principales (PCA), fueron nombrados principales filtros de Gabor. Ellos emplearon estos filtros para realizar la verificación e identificación de rostros empleando como clasificador nearest neighbor, lo cual concluyen que estos filtros son capaces de alcanzar tasas de reconocimiento similares que las originales, pero usando una representación mucho más compacta.

f) Bui et al. (2011) desarrollaron un enfoque para el reconocimiento facial empleando filtros de Gabor junto con análisis de componentes principales (PCA) para obtener el vector de características y a diferencia de Struc *et al.* utilizaron máquina de soporte de vectores (SVM).

FUNDAMENTACIÓN TEÓRICA

Detección del rostro

El rostro contiene un conjunto de características que identifican de forma única a cada persona más que cualquier otra parte del cuerpo, también contiene los principales medios de comunicación, algunos de los cuales son evidentes, como los ojos como receptores de imagen y los labios como emisores de voz, y algunos de los cuales son menos obvios, como el movimiento de los ojos, el movimiento de los labios, el cambio de color en la piel y los gestos faciales. Las propiedades del rostro plantean un problema muy difícil para el análisis de imagen facial: una cara es un objeto dinámico y no rígido que es difícil de manejar. Como consecuencia, la mayoría de las tareas de análisis facial generalmente implican cálculos pesados debido a la complejidad de los patrones faciales. Por lo tanto,

uno puede necesitar algunas señales adicionales, tales como el color o el movimiento, a fin de ayudar y acelerar el análisis.

Hay numerosos factores que causan que la apariencia facial varíe. Las fuentes de variación en la apariencia facial pueden ser categorizadas en dos grupos: factores intrínsecos que son debido a la naturaleza física del rostro y los factores extrínsecos que causan que la apariencia del rostro se altere vía interacción del rostro con fuentes externas, como se presenta en la figura.

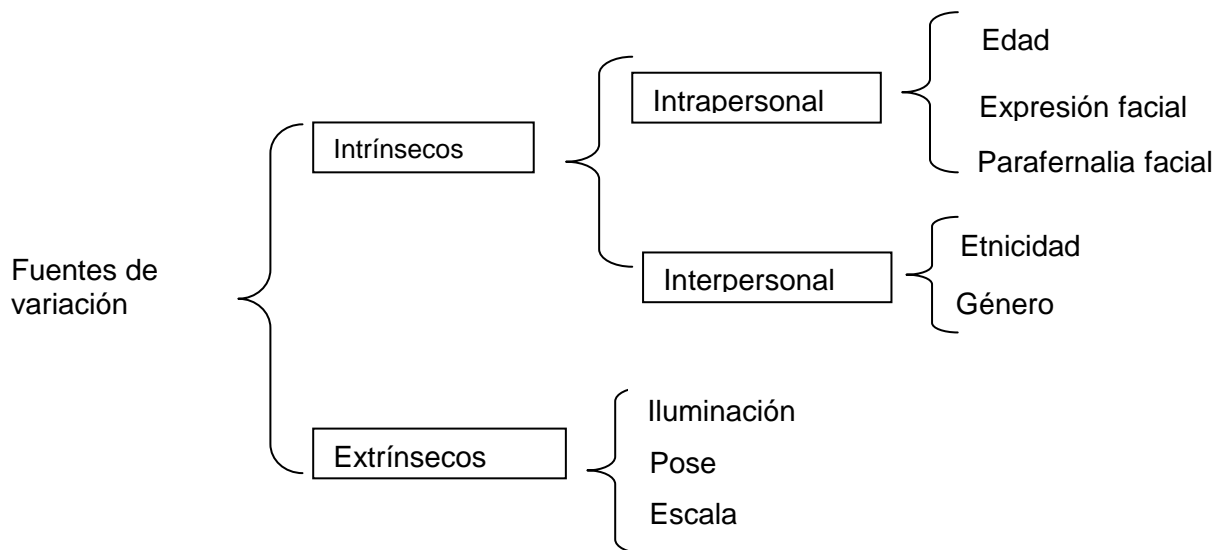


Figura 2.4. Fuentes que causan variación facial.

La detección del rostro es la primera etapa clave en aplicaciones tales como video vigilancia, interface humano-computadora, el reconocimiento facial, análisis de expresión facial, video conferencias, entre otros. El objetivo principal de la detección del rostro es segmentar la imagen en regiones que contienen rostros humanos o las partes de éste en imagen o video. Existen varios enfoques empleadas para la detección del rostro, tales como métodos basados en el conocimiento, enfoques de características invariantes, métodos de plantillas y métodos basados en la apariencia (Yang et al., 2002).

Dentro del enfoque de características invariantes se encuentra el uso de color. Este es una característica usada en aplicaciones de visión por computadora, como una señal que ofrece muchas ventajas: fácil de entender y utilizar. La implementación es computacionalmente rápida y eficiente. Bajo iluminación estable y uniforme, la señal de color se mantiene fuerte frente a los cambios geométricos, tales como la escala y la rotación. Su capacidad para separar los objetivos de fondo depende de la disimilitud de color entre los objetivos y el fondo. En algunas escenas, el color en sí mismo es suficiente para la detección de objetos. Por ejemplo, con la detección de rostros por la apariencia, se realiza una exploración exhaustiva (en diferentes lugares y escalas) de las imágenes en la búsqueda de rostros. Sin embargo, cuando la señal de color está disponible, se puede reducir las regiones de búsqueda por el pre-procesamiento de las imágenes y la selección de las áreas de la piel.

La principal dificultad en el uso del color en aplicaciones de visión artificial es que las cámaras no son capaces de distinguir los cambios de colores de la superficie a partir de los cambios de color producidos mediante la variación de los espectros de la iluminación. Por lo tanto, el color es sensible a cambios en la iluminación que son comunes en ambientes no controlados. Los cambios pueden ser debido a la variación de nivel de luz, por ejemplo, la presencia de sombras, la variación del color de la luz debido a los cambios en la distribución de potencia espectral (como la luz del día y la fuente de luz fluorescente), o ambos.

Varias estrategias se han empleado para reducir la sensibilidad de la iluminación. Una de las más comunes es que, la información de color se separa en dos componentes, la intensidad del color y la cromaticidad del color. El uso del componente de cromaticidad de color reduce el efecto de variar los niveles de luz. Para cancelar el efecto de color de la iluminación y por lo tanto diferentes distribuciones de potencia espectral, numerosos algoritmos de constancia de color se han sugerido.

Las características utilizadas generalmente para distinguir dos colores son el tono, el brillo y la saturación. El tono es la longitud de onda más importante que se percibe de un color, es decir, el color predominante. El brillo se entiende como la noción cromática de intensidad luminosa. La saturación se refiere a la pureza relativa o la cantidad de luz blanca mezclada con el tono de color, por ejemplo, el color rosa es un rojo no saturado.

Los colores de la piel humana de los diferentes individuos se agrupan en una pequeña región en un espacio de color a condición de que las imágenes sean tomadas en ambientes de iluminación controlada. El modelo de color de la piel es una descripción de los posibles tonos de piel. Para crear un modelo de este tipo, hay que seleccionar en primer lugar el espacio de color en el que se forma el modelo, y posteriormente el modelo matemático para describir los posibles colores de la piel. Los espacios de color más utilizados para este objetivo son: RGB, YCbCr, HSV, HSI, YUV, Cielab, etc. Entre ellos el mejor desempeño en términos de error de clasificación se logra mediante los modelos de HSI y YCbCr, y el mejor rendimiento en términos de tiempo de ejecución se consigue mediante el modelo YCbCr (Shin et al., 2002; Sánchez-Cuevas et al., 2013). Sin embargo, debido al hecho de que todas las imágenes son de captura en el modelo de color RGB, se debe realizar una transformación entre este modelo y el resto de modelos.

Espacio de color HSI

Por sus siglas H (Hue/Tonalidad), S (Saturation/Saturación), I (Intensity/Intensidad). El modelo HSI debe su utilidad a dos hechos básicos. Primero, la componente de intensidad, I, está desacoplada de la información cromática contenida en la imagen. Segundo, las componentes de tono y saturación están íntimamente relacionadas con la forma en que los seres humanos percibimos el color. Estas características hacen que el modelo HSI sea una herramienta ideal para desarrollar algoritmos de procesamiento de imágenes

basados en alguna de las propiedades de la percepción del color del sistema visual humano (González y Woods, 2007).

Transformación de RGB a HSI

$$H = \begin{cases} \cos^{-1}(z) & \text{if } (G \geq R) \\ 2\pi - \cos^{-1}(z) & \text{if } (G < R) \end{cases} \quad (2.06)$$

$$S = 1 - \frac{3 \min(R, G, B)}{R + G + B} \quad (2.07)$$

$$I = \frac{R + G + B}{3} \quad (2.08)$$

Donde:

$$z = \frac{(2B - G - R)/2}{\sqrt{(B - G)^2 + (B - R)(G - R)}} \quad (2.09)$$

Espacio de color YCbCr

El modelo YCbCr usado principalmente para la transmisión de señal de video, en el cual la información de luminancia está contenida en la componente Y, y la información de crominancia está en Cb y Cr. Cb refleja la diferencia entre la componente azul de la señal de entrada RGB y el brillo de la señal RGB, y Cr refleja la diferencia entre el componente rojo en señal de entrada RGB y el brillo de la señal RGB. En comparación con la transmisión de señal de vídeo RGB, la mayor ventaja del modelo YCbCr está en la separación de brillo y croma, donde la información de croma es muy eficaz en la aplicación de la segmentación de imágenes en color.

Transformación de RGB a YCbCr

$$Y = [0.299R + 0.587G + 0.114B] + 16 \quad (2.10)$$

$$Cb = [-0.169R - 0.332G + 0.5B] + 128 \quad (2.11)$$

$$Cr = [0.5R - 0.419G - 0.081B] + 128 \quad (2.12)$$

Segmentación

La segmentación es uno de los problemas clave en el análisis de imágenes y existen tantas técnicas como situaciones específicas. La segmentación de una imagen puede definirse como su partición en diferentes regiones, donde las cuales son conjuntos disjuntos conexos de píxeles.

El objetivo de la segmentación es simplificar y/o cambiar la representación de una imagen en otra más significativa y más fácil de analizar. La segmentación se usa tanto para localizar objetos como para encontrar los límites de estos dentro de una imagen. Más precisamente, la segmentación de la imagen es el proceso de asignación de una etiqueta a cada píxel de la imagen de forma que los píxeles que compartan la misma etiqueta también tendrán ciertas características visuales similares.

Una vez que el proceso de segmentación en la imagen haya terminado, es posible realizar mediciones en las regiones y analizar sus relaciones de adyacencia. Debido a lo anterior, el proceso de segmentación es un paso determinante para una interpretación cuantitativa de los datos de la imagen. El diseño de algoritmos para segmentar una imagen en regiones significativas requiere de un conocimiento previo acerca de aquellos componentes de la imagen que se pretenden investigar y diferenciar. El reconocer características como la forma, tamaño, orientación, distribución en niveles de gris, y textura puede permitir el discriminar entre diferentes estructuras de la imagen.

Patrón

Un patrón es un arreglo de descriptores. El nombre de característica es usado en el reconocimiento de patrones para denotar un descriptor. Una clase de patrón es una familia de patrones que comparten propiedades comunes (Gonzalez y Woods, 2007).

Morfología matemática

La morfología matemática es una de las ramas del procesamiento de imágenes con más fundamentos matemáticos (con vínculos a álgebra abstracta, topología, geometría discreta, geometría integral, probabilidad geométrica, ecuaciones diferenciales, etc.) y constituyen una alternativa al procesamiento de señales tradicional (Ronse, 2005; Russ, 2007).

El análisis morfológico permite extraer componentes de la imagen que son útiles en la representación y descripción de la forma de las regiones (bordes, esqueletos), permite obtener características relevantes de los objetos en la imagen.

El lenguaje de la morfología matemática es la teoría de conjuntos. Los conjuntos representarán la forma de los objetos en una imagen. Las primeras etapas de una operación de procesamiento de imágenes tienden a ser lineales, después se sigue con procesamiento morfológico y finalmente con cálculo estadístico o procedimientos de reconocimiento sintáctico.

Una parte esencial en las transformaciones en morfología matemática utilizan un conjunto geométrico conocido como elemento de estructura, el cual posee forma y un origen. Típicamente es de tamaño más pequeño que la imagen a analizar. El pixel central del elemento de estructura, llamado el origen, identifica el pixel de interés. El elemento de estructura es una imagen binaria, es decir, una matriz pequeña de pixeles, con valores de ceros o unos. Donde la dimensión de la

matriz especifica el tamaño y el patrón de ceros y unos especifica la forma del elemento de estructura.

El elemento de estructura se traslada sobre la imagen bajo estudio con la finalidad de determinar el conjunto de puntos que intervendrán en cada operación que se efectuó con las componentes de la imagen.

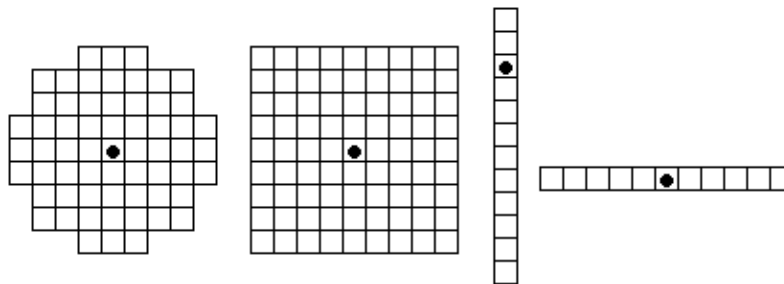


Figura 2.5. Ejemplos de elementos de estructura.

El tamaño del elemento de estructura se denota como μ , donde μ es un escalar que determina el tamaño de la estructura geométrica de B .

Los filtros morfológicos básicos son la apertura $\gamma_{\mu B}$ y cierre morfológico $\phi_{\mu B}$ con un elemento de estructura dado.

Así, la apertura morfológica se obtiene por la erosión seguida por la dilatación de la imagen resultante, dada por:

$$\gamma_{\mu B} = \delta_{\mu} \hat{B}(\varepsilon_{\mu} B(f)) \quad (2.13)$$

y el cierre morfológico se obtiene por la dilatación seguida por la erosión de la imagen resultante, y está dada por:

$$\varphi_{\mu B} = \varepsilon_{\mu} \hat{B}(\delta_{\mu} B(f)) \quad (2.14)$$

Donde $\hat{B} = \{-x : x \in B\}$ es el transpuesto del elemento de estructura B .

Donde la erosión morfológica $\varepsilon_{\mu B}$ esta expresado por:

$$\varepsilon_{\mu B}(f) = f \ominus \mu B: x \mapsto \inf_{y \in \mu B} f(x + y) \quad (2.15)$$

Y la dilatación morfológica:

$$\delta_{\mu B}(f) = f \oplus \mu B: x \mapsto \sup_{y \in \mu B} f(x - y) \quad (2.16)$$

Filtros por reconstrucción

El filtrado morfológico juega un papel fundamental no solo para simplificar la imagen de entrada sino también para la detección de marcadores. Particularmente, la clase de filtros conectados llamados por reconstrucción han sido aplicados no solo en el filtrado de la imagen sino también en segmentación de la imagen.

Los filtros por reconstrucción se forman por medio de transformaciones geodésicas, las transformaciones geodésicas básicas, la erosión y la dilatación geodésicas de tamaño 1 son iterados hasta que la idempotencia es alcanzada. La dilatación geodésica $\delta_f^1(g)$ Y la erosión geodésica $\varepsilon_f^1(g)$ de tamaño uno están dadas por:

$$\delta_f^1(g) = f \wedge \delta(g) \text{ Con } g \leq f \quad (2.17)$$

$$\varepsilon_f^1(g) = f \vee \varepsilon(g) \text{ Con } g \geq f \quad (2.18)$$

Cuando la función g es igual a la erosión o dilatación morfológicas de la función original, se obtiene la apertura y cerradura por reconstrucción.

Apertura por reconstrucción está dada por:

$$\tilde{\gamma}_{\mu}(f) = \lim_{n \rightarrow \infty} \delta_f^n(\varepsilon_{\mu}(f)) \quad (2.19)$$

Cerradura por reconstrucción está dada por:

$$\tilde{\phi}_{\mu}(f) = \lim_{n \rightarrow \infty} \varepsilon_f^n(\delta_{\mu}(f)) \quad (2.20)$$

Apertura viscosa

Una de las características de los filtros por reconstrucción es que permiten la extracción completa de los objetos marcados mediante la preservación de los bordes. Sin embargo la principal dificultad con estos filtros, es que se conecta demasiado, por lo tanto uno requiere una conexión pequeña en lugar de la habitual. De hecho, esta conectividad frecuentemente dificulta la identificación de las diferentes regiones en la imagen. Para atenuar esta inconveniencia, Serra extendió estos conceptos por medio de un enfoque de marcador y la noción de mapeo (Serra, 2002, 2005) en el cual caracteriza el concepto de propagación viscosa por medio de reconstrucción geodésica. Esta caracterización fue sugerida por el algoritmo de viscosidad propuesto por Vachier y Meyer (Vachier y Meyer, 2007).

Así una conexión sobre ventanas viscosas, las cuales no conectan demasiado, se define para permitir la separación de componentes conectados en un set de componentes conectados en un sentido de ventana viscosa. El principal objetivo de los operadores viscosos propuesto por Vachier y Meyer se centra en segmentación de objetos con contornos irregulares y discontinuos.

La idea de los operadores viscosos en (Maragos y Vachier, 2008) y (Vachier y Meyer, 2007) es combinar los efectos de una familia de cerraduras o

aperturas de actividad decreciente (tamaño del elemento de estructura) de tal manera que las áreas de baja luminancia son severamente suavizadas mientras que las aéreas con alta luminancia no son cambiados. Desde un punto de vista práctico, los filtros viscosos conectados tienen un comportamiento similar a los filtros con criterio de reconstrucción.

El número de componentes conectados depende del parámetro de viscosidad λ .

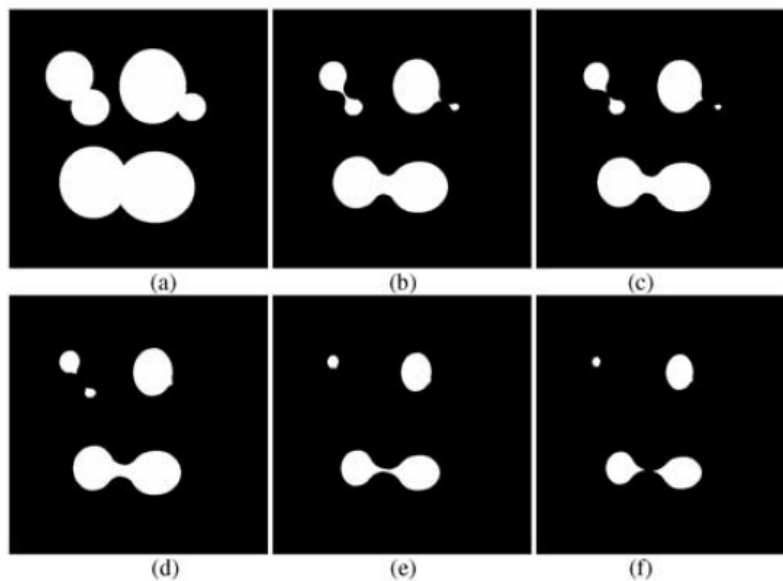


Figura 2.6. Apertura viscosa. a) Imagen original, b),c),d),e) y f) Componentes erosionados conectados a viscosidad $\lambda =20$, $\lambda =22$, $\lambda =27$, $\lambda =36$ y $\lambda =39$ respectivamente. Santillán et al. (2010).

Los filtros conectados son frecuentemente implementados por medio de algoritmos de reconstrucción, similarmente los filtros viscosos conectados sobre ventanas viscosas pueden ser implementados usando estos algoritmos. Así la apertura y cerradura viscosa a viscosidad λ (con $\lambda \leq \mu$) son dados por:

$$\mathcal{Y}_{\lambda,\mu}(f) = \delta_{\lambda} R(\varepsilon_{\lambda}(f), \varepsilon_{\mu-\lambda} \varepsilon_{\lambda}(f)) \quad (2.21)$$

$$\mathcal{Y}_{\lambda,\mu}(f) = \delta_{\lambda} R(\varepsilon_{\lambda}(f), \varepsilon_{\mu}(f)) \quad (2.22)$$

$$\tilde{\Phi}_{\lambda,\mu}(f) = \tilde{\varepsilon}_{\lambda} R^*(\delta_{\lambda}(f), \delta_{\mu-\lambda} \delta_{\lambda}(f)) \quad (2.23)$$

$$\tilde{\Phi}_{\lambda,\mu}(f) = \varepsilon_{\lambda} R^*(\delta_{\lambda}(f), \delta_{\mu}(f)) \quad (2.24)$$

Ambas transformaciones $\mathcal{Y}_{\lambda,\mu}(f)$ y $\tilde{\Phi}_{\lambda,\mu}(f)$ son expresados en términos de transformaciones por reconstrucción. El parámetro λ determina la viscosidad de la transformación y μ el tamaño del criterio.

Plantilla facial

La segmentación por medio de color de piel es útil, sin embargo las regiones que son detectadas no siempre será el rostro; existe la posibilidad de que se segmenten otras partes del cuerpo, tales como los brazos, el cuello, los hombros, etc. Por esta razón, es necesario confirmar si la región detectada es un rostro, para ello se utiliza una plantilla facial, la cual se puede obtener de varias formas; una de las opciones más fácil y simple de implementar es haciendo un promediado de varias imágenes en escala de grises.

La plantilla facial se crea mediante la realización de un cálculo de promedio de múltiples imágenes faciales en escala de grises.

$$T(x, y) = \frac{1}{N} \sum_{I=1}^N f_i(x, y) \quad (2.25)$$

Donde:

T : es la plantilla resultante.

N : es el número de imágenes.

f_i : es la imagen en escala de grises.

El grado de similitud entre la imagen facial dada y la plantilla facial se evalúa por medio de correlación normalizada. El resultado es un escalar, si el valor resultante es mayor a un cierto umbral establecido, la región detectada es considerada como rostro, en caso contrario es rechazado.

Correlación normalizada

La correlación es utilizada como una medida de similitud en el contexto de correspondencia de patrones o de plantillas. Los valores altos de correlación indican una buena correspondencia entre la imagen y el patrón o plantilla. Para compensar las variaciones de valores de intensidad, se normaliza utilizando la variación promedio de intensidades tanto de la imagen como de la plantilla.

$$C_n = \frac{\sum (I(x, y) - m_I)(T(x, y) - m_T) / N}{\sigma_I \sigma_T} \quad (2.26)$$

Donde:

Media de la imagen o patrón: valor promedio de los pixeles en el soporte.

$$m_I = \frac{\sum I(x, y)}{N} \quad (2.27)$$

Varianza de la imagen: variación promedio de los pixeles respecto a la media bajo el soporte.

$$\sigma^2 = \frac{\sum (I(x, y) - m)^2}{N} \quad (2.28)$$

Extracción de características

Filtros de Gabor

Básicamente los filtros de Gabor son un grupo de wavelets, cada wavelet captura la energía a una frecuencia y una dirección específica. La expansión de una señal usando estas bases proporciona una descripción de frecuencia localizada, por lo tanto, la captura de las características locales y/o energía de la señal (Huang et al., 2008).

Un filtro de Gabor bidimensional es un kernel gaussiano modulado por una onda plana sinusoidal. Formalmente un filtro de Gabor en el dominio espacial es definido como (Kamarainen et al., 2006; Ilonen et al., 2007; Shen et al., 2007):

$$\psi_{u,v}(x, y) = \frac{f^2}{\pi\gamma\eta} e^{-\left(\frac{f^2}{\gamma^2}x'^2 + \frac{f^2}{\eta^2}y'^2\right)} e^{j2\pi fx'} \quad (2.29)$$

Donde:

$$x' = x \cos \theta + y \sin \theta \quad (2.30)$$

$$y' = -x \sin \theta + y \cos \theta \quad (2.31)$$

f : es la frecuencia central del filtro.

θ : es el ángulo de rotación en sentido anti horario o de la envolvente gaussiana.

γ : es la nitidez de la gaussiana a lo largo del eje mayor paralelo a la onda.

η : es la nitidez a lo largo del eje menor, perpendicular a la onda.

Las características Gabor conocidos como Gabor jet, banco Gabor o característica Gabor multi-resolución son construidos de las respuestas de los

filtros de Gabor usando múltiples filtros de diferentes frecuencias f_m y orientaciones θ_n . Definidas como:

$$f_m = \frac{f_{\max}}{K^{(m/2)}} \quad (2.32)$$

$$\theta_n = \frac{n\pi}{N} \quad (2.33)$$

El enfoque más utilizado al usar filtros de Gabor para el reconocimiento facial es la construcción de un banco de filtros con filtros de diferentes escalas y orientaciones y filtrar la imagen del rostro dada con todos los filtros del banco. Es común aplicar varios filtros de Gabor con diferentes orientaciones y frecuencias con el fin de extraer las características más significativas de una imagen (Zhao y Chellappa, 2005; Struc et al., 2009; Serrano *et al.*, 2011).

Un banco de filtros se construye manteniendo γ , η constantes y dejando que θ y f tomen valores discretos dentro de un cierto rango. Las frecuencias están completamente especificadas haciendo referencia a: la frecuencia máxima (f_{\max}), limitado por Nyquist-Shannon (0.5pixel^{-1}); las orientaciones válidas generalmente se propagan entre 0 y π radianes en forma gradual. El valor exacto de θ_n es irrelevante, ya que la operación de magnitud anula su influencia, por lo que este factor no se tendrá en cuenta.

Un banco de Gabor se caracteriza por tener un total de $f_n \times \theta_n$ filtros, todos del mismo tamaño, y con valores constantes de los parámetros γ y η . El ángulo de orientación θ se suele hacer variar entre (0, π) radianes a intervalos regulares de π/θ_n radianes. Por otro lado, los filtros se crean primero con la frecuencia igual a f_{\max} , para ir posteriormente dividiendo este valor por el factor de proporción K hasta abarcar un total de f_n frecuencias.

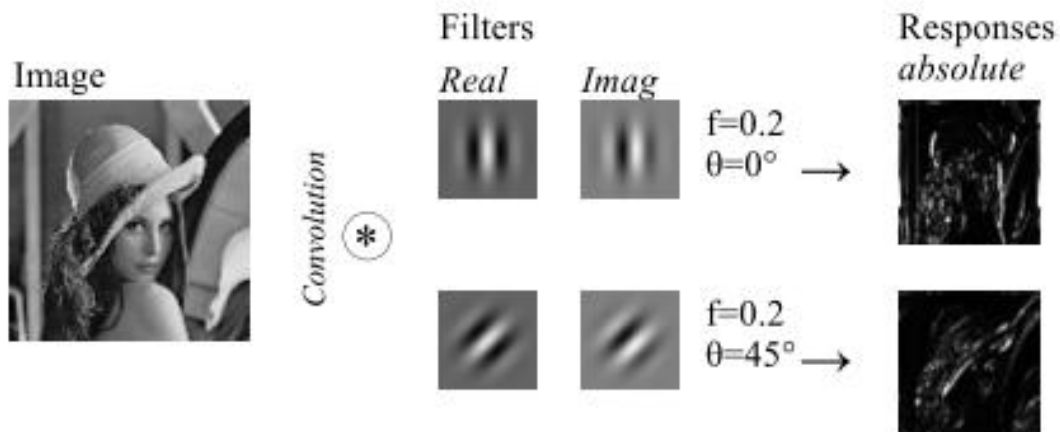


Figura 2.7. Diagrama del filtrado Gabor. Ilonen et al. (2007).

De acuerdo con la teoría de muestreo de Nyquist, una señal que contiene las frecuencias más altas de la mitad de la frecuencia de muestreo no puede ser reconstruido completamente. Por lo tanto, la frecuencia límite superior para una imagen 2D es 0.5 ciclos / píxel, mientras que el límite inferior es 0. Sin embargo, la banda de frecuencia útil de imágenes del rostro es mucho más estrecha (Shen et al., 2007). Basado en Bianconi y Fernandez (2007) la frecuencia máxima debería ser calculada de acuerdo al valor del parámetro γ , con la siguiente ecuación:

$$F_{\max} = \frac{\gamma}{2(\gamma + \sqrt{\log 2 / \pi})} \quad (2.34)$$

Clasificación

Perceptrón multicapa

Las Redes Neuronales Artificiales (RNAs) son sistemas de mapeo no lineal cuya estructura se basa vagamente en los principios observados en el sistema nervioso de los seres humanos.

Una red neuronal artificial (RNA) consta de un gran número de procesadores simples unidos por conexiones ponderadas (Figura 2.8). Por

analogía, las unidades de procesamiento pueden ser llamadas neuronas. Cada unidad recibe entradas de muchos otros nodos y genera una sola salida escalar que depende únicamente en la información disponible a nivel local, ya sea almacenada internamente o de llegar a través de las conexiones ponderadas. La salida se distribuye y actúa como una entrada a otros nodos de procesamiento (Reed y Marks, 1999).

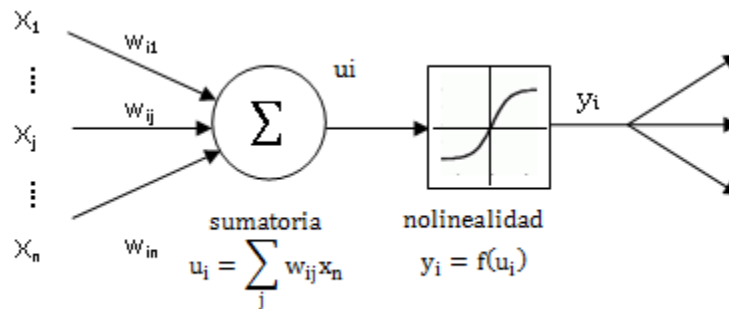


Figura 2.8. Modelo de una neurona artificial.

Una red neuronal entrenada puede ser considerada como un "experto" en la categoría de información que se ha dado a analizar.

Una aplicación importante de las redes neuronales es el reconocimiento de patrones. El reconocimiento de patrones se puede implementar mediante el uso de un feed-forward de la red neuronal que ha sido entrenado. Durante el entrenamiento, la red está capacitada para asociar salidas con patrones de entrada.

Una de las arquitecturas más utilizadas para el reconocimiento de rostros es la red de perceptrones multicapas (MLP), ya que a diferencia de una red mono capa, una red multicapa es capaz de resolver problemas más complejos.

El tipo de redes multicapa constan de tres componentes principales:

1. Capa de entrada. Recibe directamente la información de entrada, la información que será procesada por la red.

2. Capa(s) oculta(s). Se encuentra(n) en la parte interna de la red y no tiene(n) contacto directo con la información de entrada. El número de capas ocultas se determina manualmente y depende del problema a resolver.
3. Capa de salida. Transfiere hacia el exterior el resultado de la información de entrada procesada por la red.

Para entrenar una red, es necesario contar con un conjunto de patrones de entrada y sus correspondientes salidas deseadas, además de una función de error (función de coste) que mide el "costo" de las diferencias entre los productos de red y los valores deseados. Los pasos básicos son:

1. Presentar un patrón de entrenamiento x_i y propagar a través de la red para obtener las salidas.
2. Comparar los resultados con los valores deseados y calcular el error.
3. Calcular la $\frac{\partial E}{\partial w_{ij}}, \frac{\partial E}{\partial v_j}$ derivados del error con respecto a los pesos.
4. Ajustar los pesos para minimizar el error.
5. Repetir hasta que el error sea aceptablemente pequeña o el tiempo se agote.

En 1986 Rumelhart, Williams y Hinton desarrollaron el algoritmo conocido como propagación del error hacia atrás o backpropagation. Se basa en la generalización de la regla delta con conexiones hacia adelante y funciones de activación continuas no decrecientes y derivables. Este algoritmo permite que una RNA de varias capas aprenda la asociación existente entre patrones de entrada y sus clases correspondientes. La principal ventaja del Backpropagation es su capacidad genérica de mapeo de patrones. La red es capaz de aprender una gran variedad de relaciones de mapeo de patrones. No requiere un conocimiento matemático de la función que relaciona los patrones de la entrada y los patrones de salida. El Backpropagation sólo necesita ejemplos de mapeo para aprender.

El algoritmo de propagación hacia atrás es una regla de aprendizaje que organiza la capa intermedia para conseguir la correspondencia entre estímulos de entrada y patrones de salida.

El funcionamiento de una red entrenada por medio del algoritmo de propagación hacia atrás radica en el aprendizaje de un conjunto predefinido de pares de entrada-salida que se proporcionan como estímulos durante un ciclo conocido como ciclo de propagación-adaptación compuesto de dos fases.

Durante la primera fase se aplica un patrón de entrada a la primera capa de la red y éste se propaga a través de las capas superiores hasta generar un resultado a la salida. Este resultado es comparado con la salida deseada, obteniendo así un error para cada neurona de salida. El error es transmitido hacia atrás a todas las neuronas de la capa intermedia que contribuya directamente a la capa de salida, asimilando el porcentaje de error aproximado que corresponda a su participación en la salida original para realizar el ajuste de pesos correspondiente. Este proceso se realiza para cada una de las capas hasta que todas las neuronas de la red tengan un error describiendo su aportación relativa al error total. Con base en los errores obtenidos, los pesos entre neuronas son ajustados; de tal manera que cuando el mismo patrón se presente a la entrada, el error disminuirá. Conforme el error disminuye la salida está más cerca a la salida deseada.

Un aspecto común en muchas de las redes neuronales es la entrada especial llamada "bias". Esta entrada siempre presenta un valor fijo, +1 y funciona como una masa en un circuito eléctrico donde no varía de valor (se puede utilizar como un valor constante de referencia). La utilización de esta unidad tiene un doble objetivo; mejorar las propiedades de convergencia de la red y ofrecer un nuevo efecto umbral sobre la unidad que opera. El bias es agregado o embebido en el algoritmo de aprendizaje. Esto es deseable en algunos casos, ya que

proporciona cierta seguridad de que la red se comportará correctamente en situaciones críticas.

Tipos de función de activación del perceptrón

El desempeño del entrenamiento de una red neuronal depende del algoritmo de aprendizaje, del número de capas ocultas, del número de neuronas en cada capa, la arquitectura y la función de activación. La selección de la función de activación se realiza de acuerdo al problema a tratar.

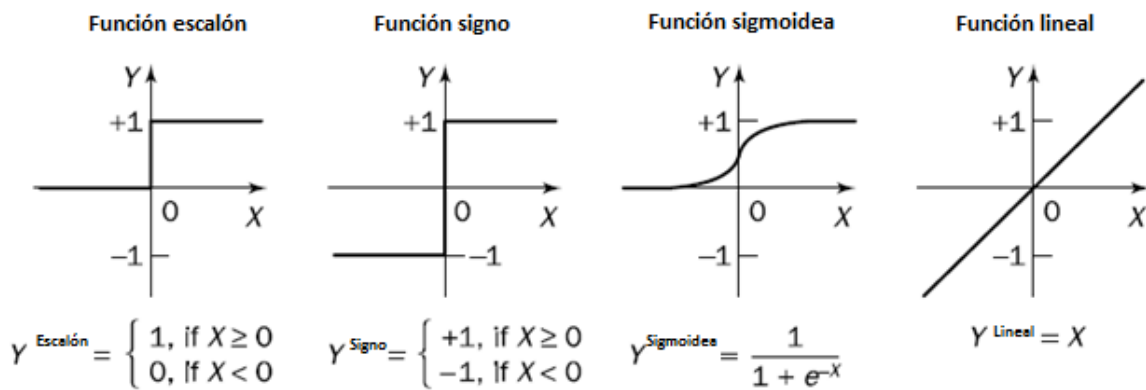


Figura 2.9. Tipos de función de activación.

Ecuaciones de entrenamiento del perceptrón multicapa con una sola salida

Primeramente se obtiene la multiplicación entre las entradas x_i y los pesos de entrada w_{ij} que son inicialmente aleatorios:

$$s_j = \sum_{i=1}^n w_{ij} x_i \quad (2.35)$$

El resultado pasa por una función de activación, el tipo sigmoidea es uno de las más utilizadas en problemas de reconocimiento de patrones.

$$h_j = f(s_j) = \frac{1}{1 + e^{-s_j}} \quad (2.36)$$

Para la siguiente capa se vuelve a calcular la multiplicación entre los pesos de la capa oculta v_j con las salidas de la capa anterior

$$rk = \sum_{j=1}^m v_j h_j \quad (2.37)$$

Donde:

w_{ij} : son los pesos de entrada que pueden ser aleatorias

x_i : son las entradas del perceptrón multicapa.

v_j : Pesos de la capa oculta.

La salida obtenida por el perceptrón pasa por una función de activación, que esta activa o desactiva la salida de la neurona, en este caso se usa la función sigmoidea:

$$\begin{aligned} out &= O \\ O &= f(rk) = \frac{1}{1 + e^{-rk}} \end{aligned} \quad (2.38)$$

yd : es la salida deseada

$$\begin{aligned} error &= e_y \\ e_y &= yd - O \end{aligned} \quad (2.39)$$

Medida del error:

$$E = \frac{1}{2} (e_y)^2 \quad (2.40)$$

$$\Delta E = \begin{bmatrix} \frac{\partial E}{\partial v_j} \\ \frac{\partial E}{\partial w_{ij}} \end{bmatrix} \quad (2.41)$$

Actualización de los pesos de la capa oculta

$$v_j(t+1) = v_j(t) - \eta \frac{\partial E}{\partial v_j} \quad (2.42)$$

$\Delta v_j = \eta \frac{\partial E}{\partial v_j}$ Es el gradiente descendiente

Donde η es un parámetro del tamaño del paso.

La velocidad de aprendizaje η es un parámetro que debe ser seleccionado. Cuando es demasiado pequeño, la convergencia será muy lenta. Cuando es demasiado grande, el proceso podría divergir.

Para obtener la actualización de los pesos se sigue la regla de las derivadas parciales de la función de error con respecto a v_j :

$$\frac{\partial E}{\partial v_j} = \frac{\partial E}{\partial e_y} \frac{\partial e_y}{\partial O} \frac{\partial O}{\partial rk} \frac{\partial rk}{\partial v_j} \quad (2.43)$$

$$\frac{\partial E}{\partial v_j} = e_y (-1) O (1 - O) h_j \quad (2.44)$$

Por lo tanto, el cálculo de los nuevos pesos de la capa oculta se obtienen mediante:

$$v_j(t+1) = v_j(t) + \eta \delta h_j \quad (2.45)$$

Donde:

$$\delta = (y_d - O)O(1 - O) \quad (2.46)$$

Calculando los nuevos pesos de la capa de entrada:

$$w_{ij}(t+1) = w_{ij}(t) - \eta \frac{\partial E}{\partial w_{ij}} \quad (2.47)$$

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial e_y} \frac{\partial e_y}{\partial O} \frac{\partial O}{\partial rk} \frac{\partial rk}{\partial h_j} \frac{\partial h_j}{\partial s_j} \frac{\partial s_j}{\partial w_{ij}} \quad (2.48)$$

$$\frac{\partial E}{\partial w_{ij}} = e_y(-1)O(1-O)v_j h_j(1-h_j)x_i \quad (2.49)$$

Por lo tanto, para los nuevos pesos de entrada se calcula mediante:

$$w_{ij}(t+1) = w_{ij}(t) + \eta \Delta_j x_i \quad (2.50)$$

Donde:

$$\Delta_j = \delta v_j h_j(1-h_j) \quad (2.51)$$

III. METODOLOGÍA

En esta sección, se describen los materiales y técnicas empleados para el desarrollo del presente trabajo.

Adquisición de imágenes

La adquisición de las imágenes se obtiene empleando una web-cam Logitech C920, con condiciones controladas de iluminación, pose, expresión facial, y una distancia fija entre cámara y usuario. Esto se realiza mediante una interfaz, donde la captura de las imágenes es de manera automática; obteniendo 25 imágenes por persona de un total de 37 voluntarios. Las imágenes capturadas son de 320 x 240 píxeles; se elige este tamaño para acelerar el proceso de detección de color de piel.



Figura 3.1. Ejemplos de imágenes de la base de datos obtenida.

Creación de la base de datos

La base de datos con la que se trabajará cuenta con imágenes faciales previamente pre-procesadas, es decir, la detección de color de piel para segmentar el área del rostro del fondo; se emplea una clasificación de píxeles de color de piel; en la cual se utiliza unos umbrales en los componentes Cb y Cr, los píxeles que superen un cierto valor de umbral toman un valor de 255, mientras que los demás píxeles toman un valor de cero, y el recorte para obtener solamente el área del rostro existen dos técnicas, las cuales se profundizará en las siguientes secciones. Por otro lado, es importante considerar una normalización del tamaño de las imágenes faciales, que consiste en llevar a un tamaño estándar las dimensiones de la imagen, sin provocar en ella alguna distorsión de importancia, para que en la etapa de extracción de características tengan la misma dimensión del vector de características y facilite el entrenamiento y clasificación. Debido a que en el recorte automático de los rostros, los tamaños varían, las imágenes faciales se normalizaron a un tamaño de 147 x 117.

Como se mencionó en la sección anterior se contó con el apoyo de 37 personas, con 25 imágenes por persona, teniendo un total de 925 imágenes faciales en la base de datos.

Por otro lado, se cuenta con una base de datos de imágenes faciales llamado GTAV, el cual se encuentra en la red con acceso gratuito, estas imágenes son de 320x240, a color y con diferentes cambios de iluminación, con oclusión, a diferentes poses y diferentes cambios de expresión facial. Debido a que en este trabajo se considera solo pose frontal, expresión neutra y sin oclusión, se tomaron solamente imágenes de este tipo (6 imágenes por 14 personas).



Figura 3.2. Ejemplos de imágenes de la base de datos GTAV.

Detección del rostro

Dada una imagen de entrada, la primera etapa consiste en detectar la región facial en la imagen para segmentarla del fondo, para su posterior análisis.

Las imágenes capturadas se encuentran en el espacio RGB, debido a que los componentes en este espacio están ligados a condiciones de iluminación, es necesaria la conversión del espacio RGB a otro espacio de color. En este trabajo se realizan pruebas con el espacio HSI y YCbCr donde solamente se trabajan con los componentes de H y S, (Cb y Cr para el segundo espacio excluyendo la componente de iluminación I (Y para el segundo espacio)).

Clasificación de píxeles de color

El color de piel humana tiene diferentes apariencias, con diferentes razas; incluso la misma persona bajo diferentes fuentes de luz o el uso de diferente tipo de ropa, sin embargo, este color es relativamente estrecha en un espacio de color. El color de la piel humana tiene una gama restringida de tonos y no está profundamente saturado, ya que la apariencia de la piel está formada por una combinación de la sangre (roja) y la melanina (marrón, amarillo). Por lo tanto, el

color de la piel humana no cae al azar en un espacio de color determinado, sino agrupados en una pequeña área en el espacio de color. Sin embargo, esa área no es la misma para todos los espacios de color; en la figura 3.3 se puede observar la densidad de distribución de color de piel en los componentes Cb y Cr utilizando algunos tonos de piel de la base de datos obtenida, en la cual se puede observar que los umbrales son similares a los que se presentan en la literatura; sin embargo es importante mencionar que los valores no son iguales debido a que en la literatura utilizan más muestras de tonos de piel.

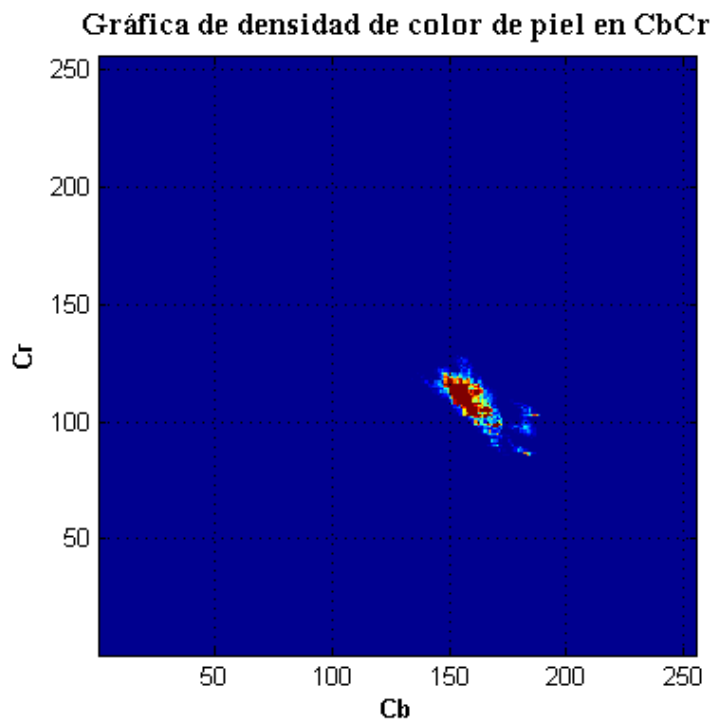


Figura 3.3. Gráfica de densidad de color de piel en CbCr.

El proceso de detección de color de piel consiste en clasificar cada pixel como piel tomando el valor de 255 en escala de grises (blanco) o no piel con el valor de cero (negro). Existen varios métodos de detección de color de piel empleando diferentes rango de umbrales (Kim *et al.*, 2004, Yin *et al.*, 2005, Phung, 2005, Horton *et al.*, 2010, Amjad *et al.*, 2012). Para este trabajo se tomaron algunos de los umbrales propuestos en la literatura, a continuación se muestran los umbrales de color con los cuales dieron mejores resultados

Para el espacio HSI:

$$I(x, y) = \begin{cases} 1 & \text{si } (0 < H < 50) \text{ and } (5 < S < 200) \\ 0 & \text{de otra manera} \end{cases} \quad (3.01)$$

Para el espacio YCbCr:

$$I(x, y) = \begin{cases} 1 & \text{si } (77 < Cb < 127) \text{ and } (133 < Cr < 173) \\ 0 & \text{de otra manera} \end{cases} \quad (3.02)$$

En la figura 3.4, se puede observar el resultado de cada espacio de color sin un pre-procesamiento, es decir, solamente el proceso de clasificación de píxeles, donde se detectan regiones de color de piel.



a)

b)

c)

Figura 3.4. Clasificación de píxeles de color de piel. a) Imagen original, b) Resultado de color de piel en espacio de color HSI, c) Resultado de clasificación de color de piel en el espacio de color YCbCr.

Pre-procesamiento de la imagen

El objetivo de la detección de color de piel es delimitar el área o contorno del rostro sin ningún tipo de ruido, por lo tanto, para eliminar objetos pequeños que se encuentran dentro del rostro, se emplea un etiquetado de objetos conectados donde las áreas pequeñas se eliminan mientras que el área mayor permanece, el

cual pertenece el área candidato a rostro. En la figura 3.5 se puede observar el resultado obtenido.



Figura 3.5. Resultado de eliminación de objetos dentro del rostro.

Posteriormente, a la figura 3.5 se le aplica una apertura morfológica (Schmitt y Mattioli, 1993; Heijmans, 1994; Ronse, 2005,) con el fin de suavizar las regiones de piel detectadas, ya que es posible que las zonas de color de piel detectadas produzca una imagen con ruido, la apertura morfológica está dada por la ecuación 2.13. Para este caso, se utilizó un elemento de estructura de tipo disco y el tamaño de 5x5 para mantener los bordes en su forma original y eliminar la mayor cantidad de ruido pero sin que se alteren demasiado los contornos de la imagen (Figura 3.6).



Figura 3.6. Imagen después de una apertura morfológica.

Después de una apertura morfológica a la imagen 3.5, se puede observar la figura 3.6 que un objeto de menor área comparado con el otro objeto permanece, por lo tanto, se aplica nuevamente una eliminación de objetos con

áreas pequeñas dejando solo el objeto de área mayor, que es un área candidata a rostro(Figura 3.7).



Figura 3.7. Imagen de eliminación de objetos de áreas pequeñas.

Se puede apreciar en la figura 3.7 que solo un objeto permanece, el cual se considera como una máscara facial, la cual sirve como base para obtener los colores de la imagen original; esto se realiza mediante una multiplicación punto a punto entre la máscara con cada canal del espacio RGB, el resultado se puede apreciar en la figura 3.8, donde se aprecia que el área resultante contiene el rostro.



Figura 3.8. Imagen resultante de la detección de color de piel.

Una vez que la parte de detección de color de piel se ha realizado, se puede apreciar como una gran parte del cuello también ha sido detectado y recuperado, pero el objetivo en esta primera etapa es segmentar solo el área del rostro evitando la región del cuello. Para ello, se cuentan con dos técnicas diferentes, una de ellas es por medio de la apertura viscosa y la otra es por medio de proporciones del rostro.

Segmentación del rostro evitando el cuello por medio de apertura viscosa

La apertura viscosa se define mediante la siguiente ecuación:

$$\mathcal{V}_{\lambda,\mu}(f) = \delta_{\lambda} R(\varepsilon_{\lambda}(f), \varepsilon_{\mu}(f)) \quad (3.03)$$

La apertura viscosa se realiza de manera automática; como primer paso, a la imagen de entrada (Figura 3.7) se le aplica un erosionado con un elemento estructural de tipo disco de tamaño variable sin que se pierda el objeto; con este proceso se obtienen dos marcadores para la reconstrucción posterior de la imagen.

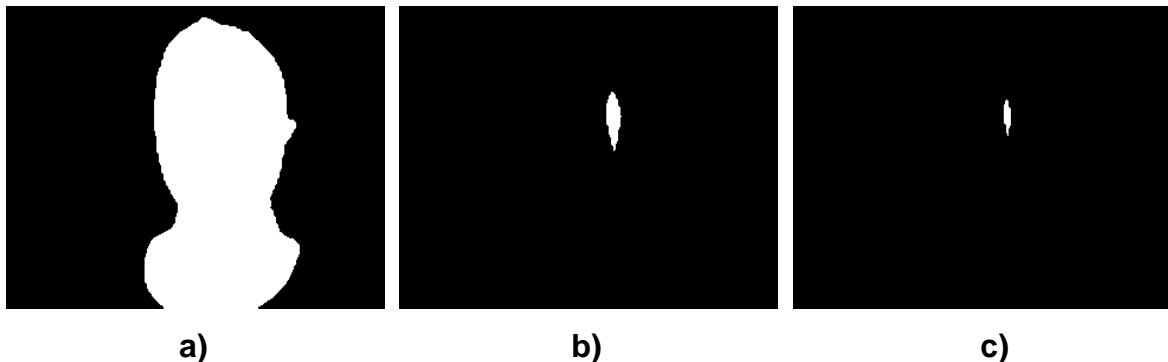


Figura 3.9. Proceso de erosionado de la imagen. a) Imagen de entrada, b) Erosión1 de la imagen a y c) Erosión 2 de la imagen a.

El siguiente paso es la reconstrucción de la imagen, se emplean la imagen de la erosión 2 como marcador y la erosión 1 como la máscara o la imagen que se va a reconstruir.

El último paso de la apertura viscosa es un dilatado con el último tamaño del elemento estructural del erosionado. En la figura 3.10 se muestra la imagen resultante.



Figura 3.10. Resultado de la apertura viscosa, a) Máscara obtenida, b) Rostro con colores de la imagen original.

Se puede observar que se tiene buen resultado y para realizar el recorte del rostro excluyendo el fondo, se encuentran las coordenadas del rostro, es decir, se encuentra dos coordenadas en el eje vertical (y_{min} y y_{max}) y dos en el eje horizontal (x_{min} y x_{max}), con el fin de tener los límites para el recorte, el resultado se muestra en la siguiente figura.

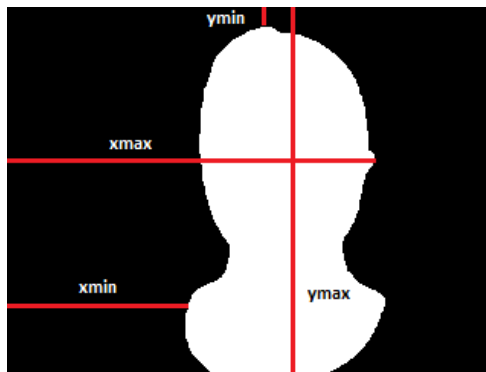


Figura 3.11. Coordenadas x e y para el recorte del rostro.

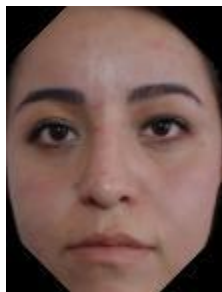


Figura 3.12. Recorte del rostro mediante apertura viscosa.

Se puede apreciar la figura 3.12 que el resultado ofrece un buen recorte del rostro, sin embargo algunas imágenes no se obtiene el mismo resultado. Con el fin de tener un algoritmo de segmentación que haga uso de la apertura viscosa y produzca imágenes con un nivel de segmentación aceptable, a las imágenes resultantes de la apertura viscosa se les aplica una apertura morfológica con un elemento de estructura de tipo cuadrado de tamaño variable para así obtener un resultado deseable a todas las imágenes.

El resultado de la apertura viscosa agregando la apertura morfológica se muestra en la figura 3.13, en la cual se puede observar que la separación del cuello y rostro se lleva a cabo, dejando una región deseable, además se conserva el contorno original del rostro, a diferencia de la figura 3.12.



Figura 3.13. Imagen resultante de la apertura viscosa agregando una apertura morfológica.

Como último paso, se encuentran las coordenadas limítrofes del rostro para efectuar una tarea adicional de recorte considerando los límites izquierdo y derecho del rostro, para dejar solo una pequeña imagen donde se elimine la región de fondo que no ofrece información de interés y contenga solo el rostro, como se muestra en la figura 3.14.



Figura 3.14. Imagen resultante del recorte del rostro de la apertura viscosa agregando una apertura morfológica.



Figura 3.15. Imágenes resultantes de la apertura viscosa junto con la apertura morfológica.

Segmentación del rostro evitando el cuello por medio de proporciones

Esta clase de segmentación del rostro sin cuello consiste en un análisis basado en las proporciones ancho versus largo del rostro. Primero, del conjunto de rostros en la base de datos propuesto se realiza manualmente una búsqueda

de la proporción entre la distancia del largo del rostro en pixeles, y el ancho del rostro en pixeles, con lo que se obtienen diferentes valores sin embargo se mantienen dentro de un rango limitado, estos valores no superan la proporción de 1.5 por lo que se toma un valor promedio considerando todos los valores obtenido; el valor utilizado para este trabajo fue de 1.3.

$$P = \frac{L}{A} \quad (3.04)$$

Donde:

P = Proporción del rostro

L = Largo del rostro

A = Ancho del rostro

Por lo tanto, la constante que se va a multiplicar para obtener el límite del rostro en el eje y es a igual a 1.3. Encontrando las coordenadas de y_{\min} , y_{\max} , x_{\min} y x_{\max} , el valor de y_{\max} se modifica debido a que en algunas imágenes se tiene parte del cuello por lo tanto se tomará el cuello y no es el resultado esperado, por lo tanto se realiza la diferencia entre x_{\max} y x_{\min} para encontrar el ancho del rostro y posteriormente multiplicarlo por la constante de proporción, es decir, por 1.3 y así se obtiene el límite del rostro en el eje y .

$$\text{ancho del rostro} = x_{\max} - x_{\min} \quad (3.05)$$

$$y_{\max} = \text{ancho del rostro} \cdot a \quad (3.06)$$



Figura 3.16. Resultado del recorte del rostro por medio de proporciones.

Normalización de las imágenes faciales

Una vez que se tengan las imágenes faciales recortadas adecuadamente, se procede a normalizarlos en tamaño, debido a que en las técnicas del recorte mencionadas, las imágenes faciales resultantes no tienen el mismo tamaño y es importante que todas las imágenes estén normalizadas en tamaño, es decir, un tamaño estándar para todas las imágenes para su posterior análisis.

Comprobación del rostro detectado o candidato facial

Una vez que se tenga las regiones de color de piel detectada, se sabe si un pixel es un pixel de piel o no, pero no se puede decir si ese pixel pertenece a una región que contiene un rostro o no. Por esta razón, es necesaria una confirmación de que la región segmentada sea exclusiva del rostro, ya que en la detección de color de piel es posible que se segmenten otras partes del cuerpo como los hombros, manos, cuello, etc. Por lo tanto, para este trabajo de investigación se emplea una manera sencilla y fácil de implementar para comprobar si la región segmentada y recortada es rostro o no.

Esto se realiza mediante una correlación normalizada entre el rostro candidato y una plantilla facial. La plantilla facial fue construida por el promedio de 20 imágenes en escala de grises, y el resultado se puede observar en la figura 3.17. Para que el rostro detectado sea confirmado es necesario que el resultado de la correlación normalizada supere un cierto umbral (0.6) en caso contrario será clasificado como no rostro.



Figura 3.17. Plantilla facial.

Extracción de características basado en Gabor

Los filtros de Gabor son filtros complejos que se obtienen variando diferentes frecuencias y orientaciones. En la figura 3.18 se pueden observar los filtros de la parte real, con un tamaño de 21 x 13.

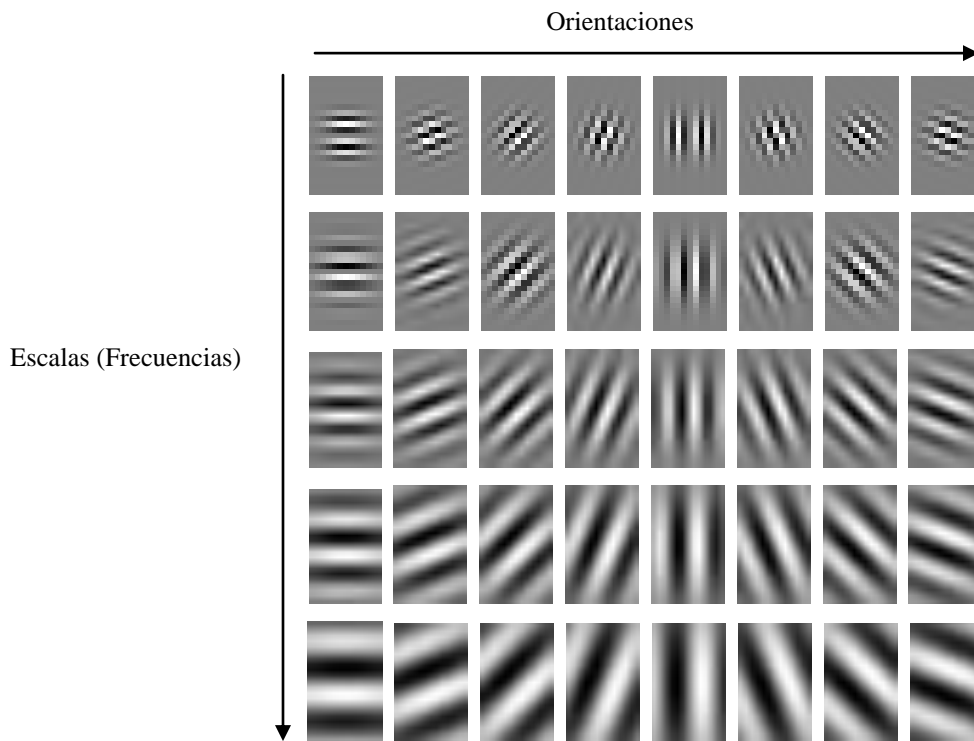


Figura 3.18. Kernels de Gabor a diferentes escalas (frecuencias) y orientaciones parte real en el dominio espacial.

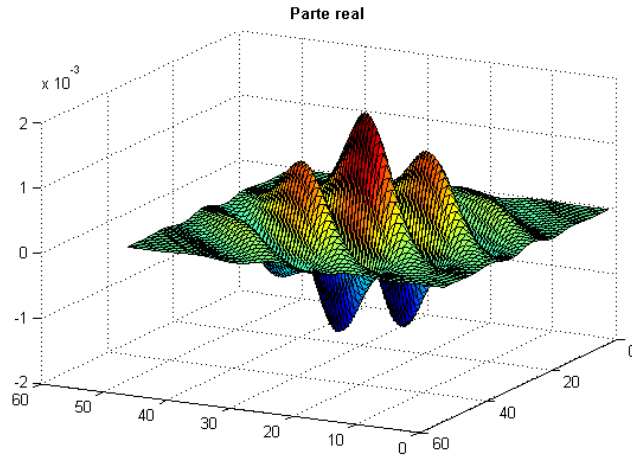


Figura 3.19. Parte real del filtro Gabor en 3D, orientación: 45° , frecuencia: 0.1053,

$$\gamma = \sigma = \sqrt{2}.$$

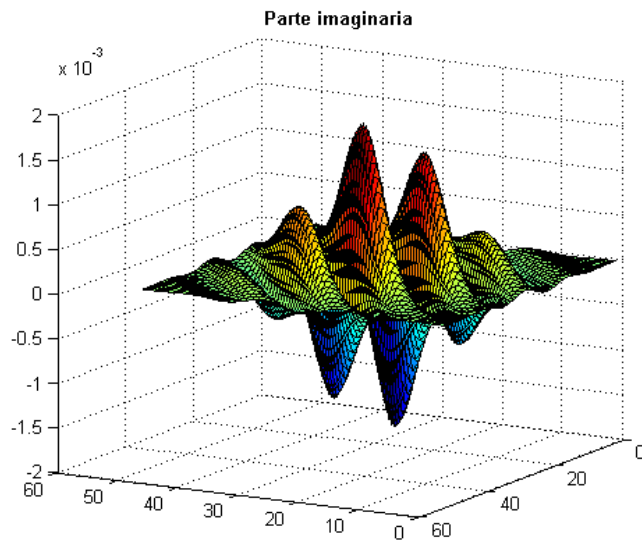


Figura 3.20. Parte imaginaria del filtro Gabor en 3D, orientación: 45° , frecuencia:

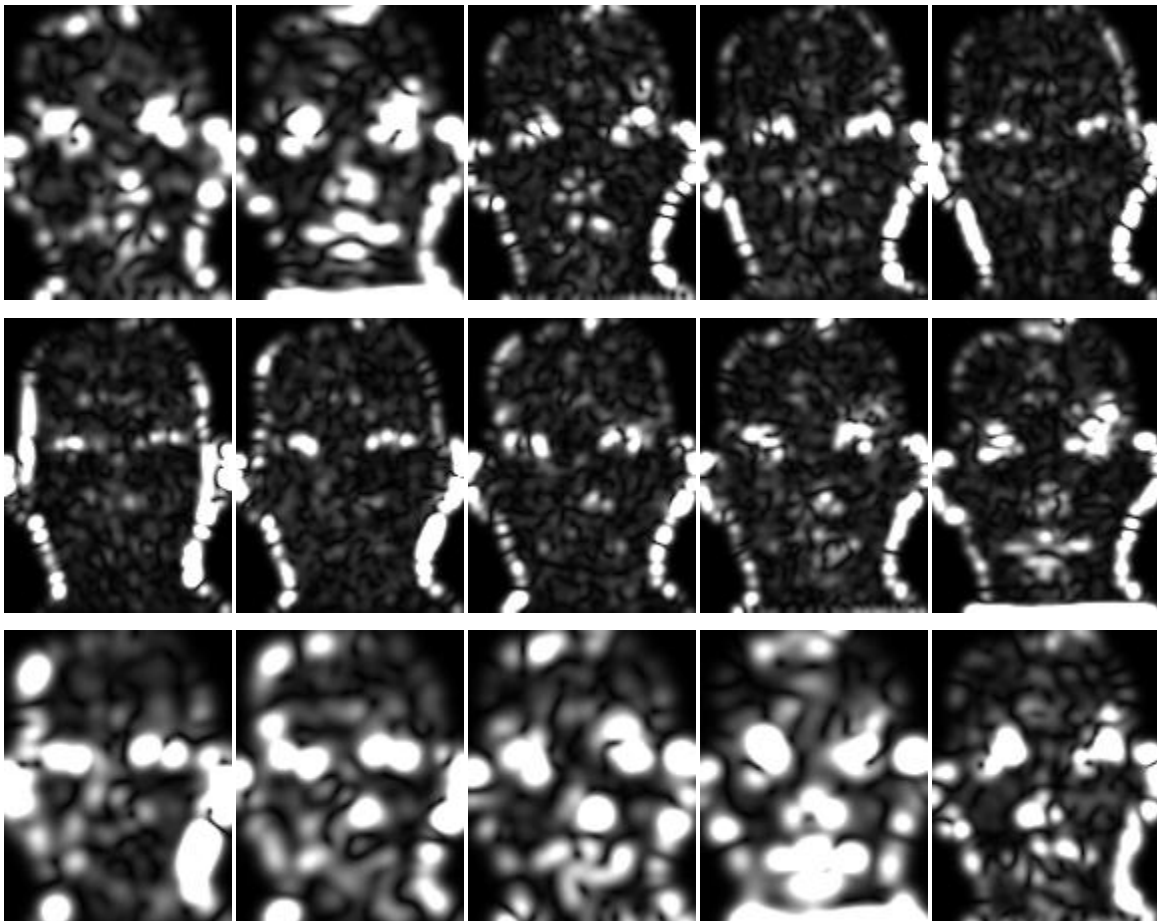
$$0.1053, \gamma = \sigma = \sqrt{2}.$$

Con el objetivo de observar la respuesta de cada filtro complejo (a diferente frecuencia y orientación) con la imagen facial, se realiza una convolución entre el filtro y la imagen facial, en este caso se emplearon 40 filtros de Gabor,

por lo que, se obtienen 40 imágenes resultantes. A continuación se muestran las imágenes resultantes de la convolución.



Figura 3.21. Imagen original en escala de grises.



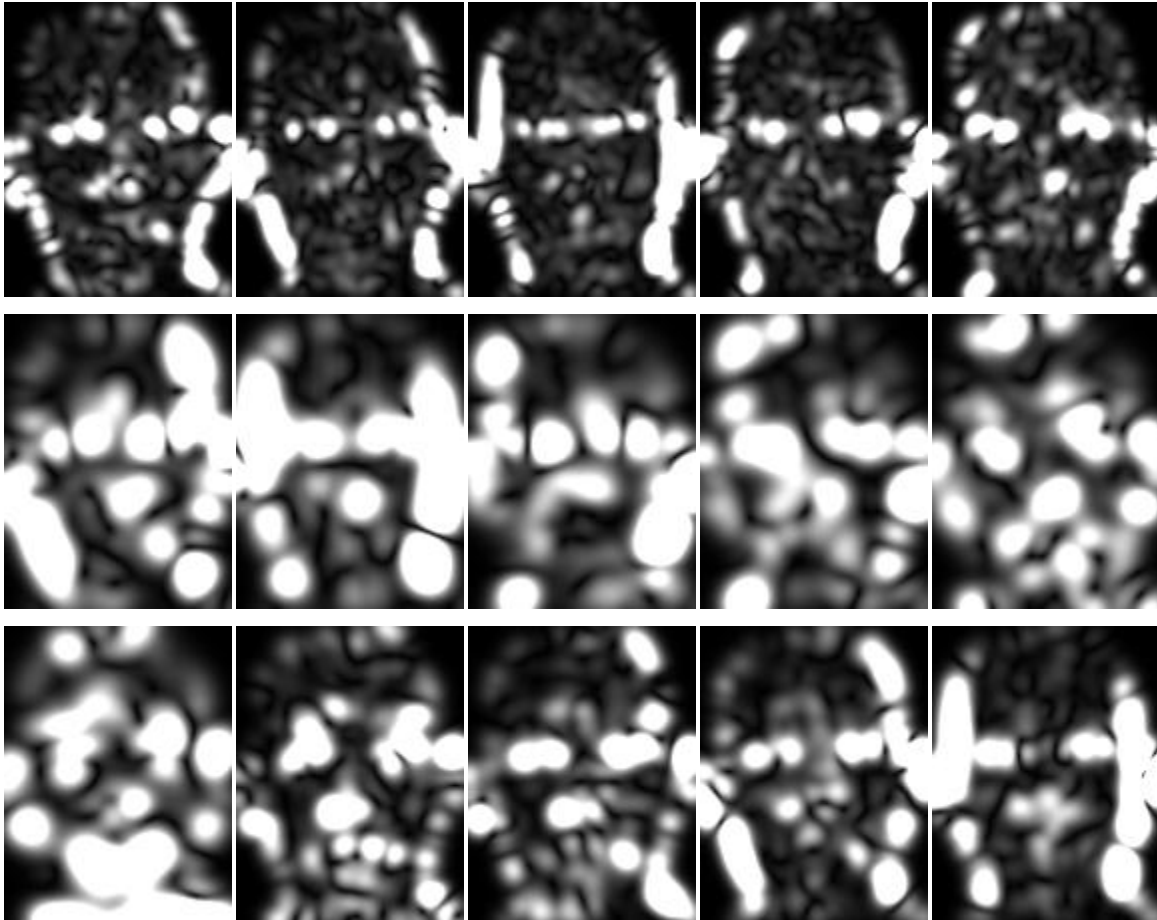


Figura 3.22. Imágenes de magnitud resultantes de la convolución.

Obtención del vector de características

Para este trabajo, los parámetros utilizados para la construcción de los filtros son los que se muestra en el cuadro 3.1. Estos parámetros fueron elegidos de acuerdo a los valores más utilizados en la literatura para el reconocimiento facial y los que mostraron mejores resultados (Bianconi y Fernandez, 2007; Serrano et al., 2011). El parámetro de f_{max} se calcula con base a la ecuación 2.34, el tamaño del kernel depende de la división de la imagen facial.

Cuadro 3.1. Parámetros del filtro de Gabor utilizados.

Parámetros del filtro	Valor
Número de frecuencias	5
Número de orientaciones	8
Factor de proporción K	$\sqrt{2}$
η	$\sqrt{2}$
γ	$\sqrt{2}$
f_{\max}	0.4211
Tamaño	21x13

La imagen facial de MN en escala de grises es dividida en $M_x M_y$ campos receptivos cada uno de tamaño $(2N_x + 1)(2N_y + 1)$ donde $N_x = (N - M_x) / 2M_x$, $N_y = (N - M_y) / 2M_y$. Este hecho permite que el tamaño del vector de características sea independiente de la imagen original (Aguilar-Torres *et al.*, 2007).

Por lo tanto, la imagen del rostro recortada se divide en 9 x 7 celdas, a cada celda se le aplica un filtro a diferente frecuencia y orientación, es decir, 40 diferentes filtros de Gabor.

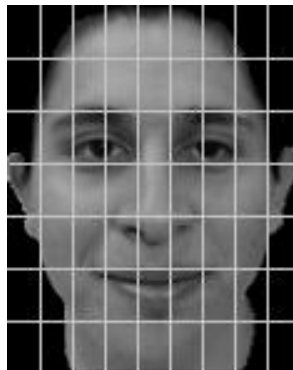


Figura 3.23. Imagen facial dividida en secciones.

Una vez que se tenga los kernel o máscaras del filtro de Gabor se realiza la Correlación cruzada entre la imagen facial en escala de grises y los filtros creados:

$$h(u, v) = \sum_{x=-N_x}^{N_x} \sum_{y=-N_y}^{N_y} I(x - ci, y - dk) \psi(x, y, f_m, \theta_n) \quad (3.07)$$

El contraste de energía es llamado también amplitud de la señal, al representarlo en niveles de gris, muestra la respuesta de la imagen al par en cuadratura en función de la posición espacial, que es independiente de la fase. Al promediar cada uno de estos contrastes de energía resultantes de cada banco de Filtros de Gabor, obtenemos los vectores característicos de la respuesta de la imagen en función de la posición y frecuencia espacial.

$$M(u) = \frac{1}{N_v} \sum_{v=1}^{N_v} |h(u, v)| \quad (3.08)$$

Donde:

$$N_v = N_\phi N_\omega$$

En la figura 3.24 se puede apreciar el vector de características de la figura 3.23 estimado mediante el procedimiento mencionado, el cual está normalizado entre 0-1 y se considera la magnitud de la respuesta entre el filtro y la imagen debido a que la respuesta toma valores complejos.

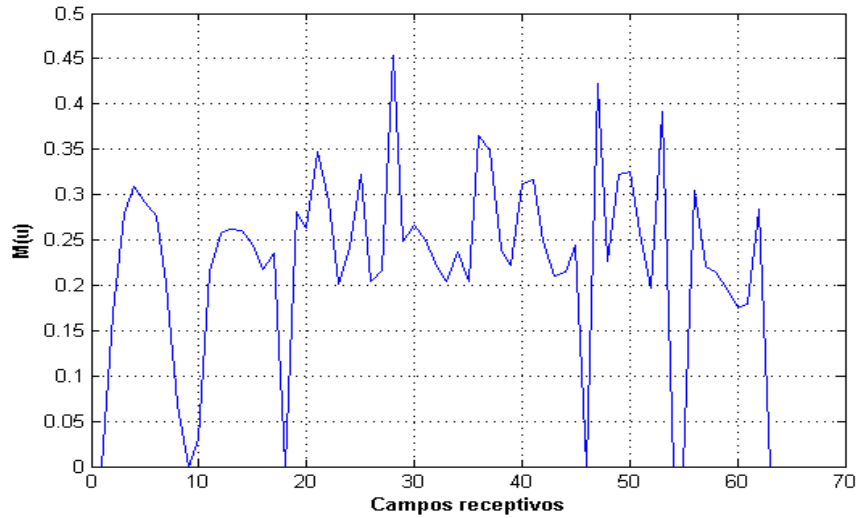


Figura 3.24. Vector de características estimado de la imagen 3.23

Diseño y parámetros del perceptrón multicapa

El diseño del perceptrón utilizado se muestra en la figura 3.25 donde el número de neuronas en la capa de entrada es igual al tamaño del vector característico, (n); el número de neuronas en la capa oculta es la mitad mas uno del tamaño total de entradas ($\frac{n}{2} + 1$).

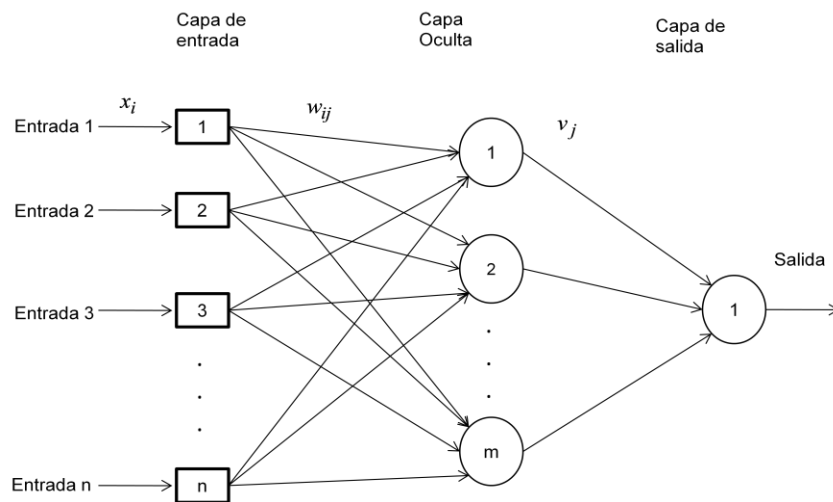


Figura 3.25. Esquema del perceptrón multicapa con una sola salida.

Entrenamiento del perceptrón multicapa

Una vez que se obtienen los vectores característicos de cada imagen para cada persona, se procede a entrenar el perceptrón multicapa. De la base de datos, con un total de 37 personas, (con 25 imágenes por persona), solo se seleccionan 27 personas para el entrenamiento 10 imágenes por persona; las imágenes restantes de estas 27 personas se utilizan en la verificación (test)). Por otro lado, queda un total de 10 personas en la base de datos que no fueron incluidos en el entrenamiento, los que servirán para la etapa de verificación donde la meta es que dichas imágenes no superen el umbral establecido debido a que son personas externas(o impostores).

Los valores de los parámetros del perceptrón fueron obtenidos a prueba y error hasta lograr tener mejores resultados de entrenamiento y verificación de vectores.

Cuadro 3.2. Valores de los parámetros del perceptrón multicapa.

Parámetros	Valores
Tamaño de paso η	0.05
Bias	1
Error de paro E	0.1
Número de neuronas de entrada n	64
Número de neuronas en la capa oculta	33
Pesos iniciales de entrada	Valores aleatorios entre 0-1
Pesos iniciales en la capa oculta	Valores aleatorios entre 0-1

IV. RESULTADOS Y DISCUSIÓN

En esta sección se exponen los resultados obtenidos de cada etapa mencionados anteriormente.

Detección de color de piel

La búsqueda o detección del rostro se realiza mediante el uso de color, se probaron dos espacios de color (HSI, YCbCr) pero al final se utilizó el espacio de color YCbCr donde se descarta la componente de iluminación (Y) y se trabaja solamente con los componentes de color Cb y Cr, debido a que tiene mejores resultados ante cambios de iluminación moderada, además de mejores tiempos de ejecución.

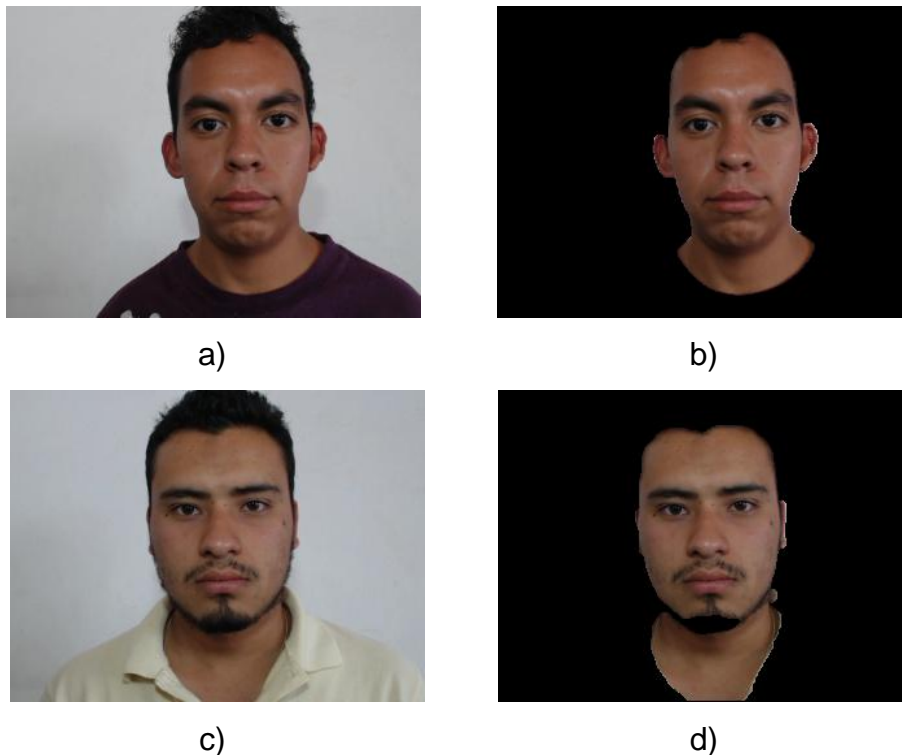


Figura 4.1. Detección de color de piel. a) y c) Imágenes de entrada, b) y d) Imágenes resultantes de la detección de color de piel.

Recorte del rostro por medio de apertura viscosa

La aplicación de la apertura viscosa en este trabajo es segmentar el rostro separándolo del cuello, la cual por si sola funciona correctamente en la mayoría de las imágenes, sin embargo para tener mejor rendimiento o reducir los casos no deseables se opta por adicionarle una apertura morfológica con los que se mejoran los resultados.

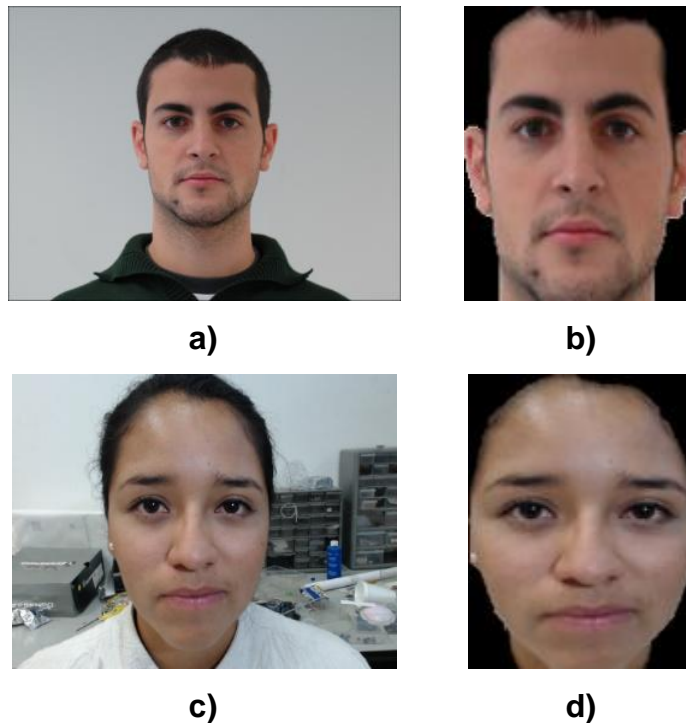


Figura 4.2. Uso de la apertura viscosa. a) y c) Imagen de entrada, b) y d) Imagen facial como resultado de la apertura viscosa.

Detección del rostro por medio de proporciones más apertura viscosa

El recorte del rostro por medio de proporciones ofrece buenos resultados, sin embargo, el rostro tiene algunas variaciones en proporciones, por ejemplo, la medida del largo del rostro es más grande que el ancho. Las medidas del largo y

ancho del rostro fueron utilizadas para calcular una constante que sirve para determinar la medida de largo del rostro y tener la coordenada del eje vertical del rostro, en algunas imágenes no fueron recortadas correctamente. Por ello fue, agregada al algoritmo la apertura viscosa, con lo que esos casos fueron mejorados. A continuación, se muestran imágenes resultantes de la técnica implementada.

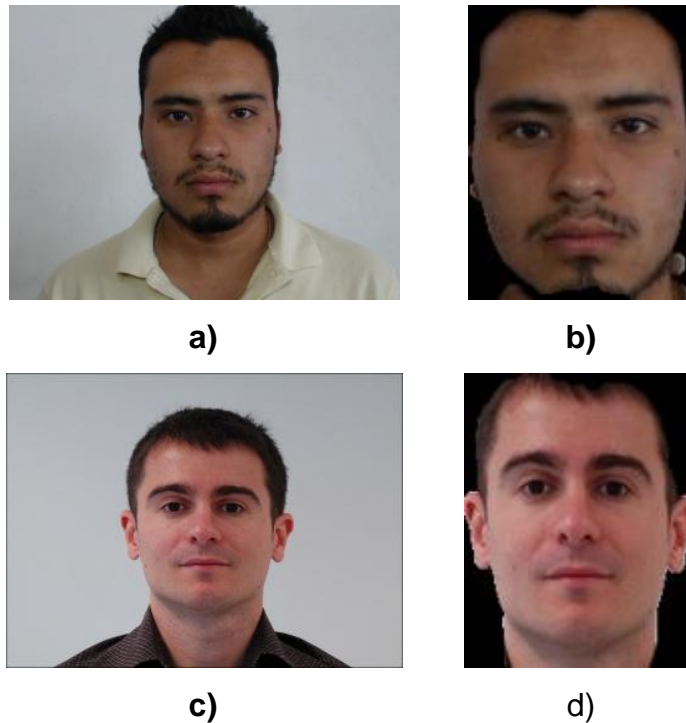


Figura 4.3. Resultado de recorte por proporciones más viscosa. a) y c) Imagen original, b) y d) Imagen de rostro recortado.

Evaluación de rostros candidatos mediante una plantilla facial

El objetivo de la detección del rostro es segmentar, localizar el rostro dentro de una imagen para su posterior análisis. Para confirmar si la región detectada es efectivamente un rostro se empleó la correlación normalizada entre el rostro candidato detectado y la plantilla facial, ambos en escala de grises. Se utilizó la base de datos propia con condiciones de iluminación y pose controladas,

y con una expresión facial neutra en la mayoría de las imágenes. Además, con el fin de observar el desempeño de la plantilla facial a la de datos anterior fueron agregadas 200 imágenes faciales o candidatos con cambios de iluminación moderada, así como imágenes faciales que incluían cuello e imágenes donde el rostro no se alcanzó a segmentar de manera correcta, es decir, el rostro estaba rostro incompleto.

Total de imágenes faciales: 1145

Total de falsos negativos (FN): 38

Total de falsos positivos (FP): 0

Total de verdaderos positivos (TP): 1107

Total de verdaderos negativos (TN): 5

Cuadro 4.1. Resultados de confirmación de rostros con plantilla facial.

Resultados	Porcentaje
índice de Falso Aceptación (FAR)	0
índice de Falso Rechazo (FRR)	3.3188
Precisión	96.6957

Verificación de rostros mediante perceptrón multicapa

La etapa de verificación de rostros se realiza empleando al perceptrón multicapa como clasificador utilizando los vectores característicos generados a partir de los 40 filtros de Gabor y la imagen facial dividida en secciones, en este caso se utilizaron 63 secciones. Durante el entrenamiento se obtuvieron los pesos óptimos para cada persona, los cuales se utilizan para esta etapa, la figura 4.4 muestra un diagrama a bloques de la verificación.

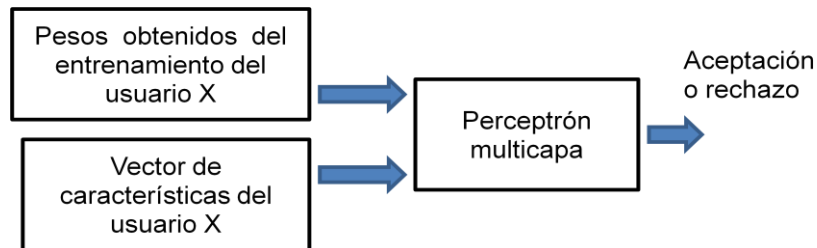


Figura 4.4. Diagrama general de la etapa de verificación de rostros.

Como ya se mencionó anteriormente, para la verificación del usuario se ocupa el dato biométrico en este caso la imagen facial con la cual se genera un vector de características; por otro lado se ocupa la identidad reclamada, y en este caso se toma un número de 1 al 27, debido a que se cuenta con 27 personas entrenadas.

A continuación se muestra resultados de entrenamiento del usuario 15.



Figura 4.5. Imagen facial del usuario 15.

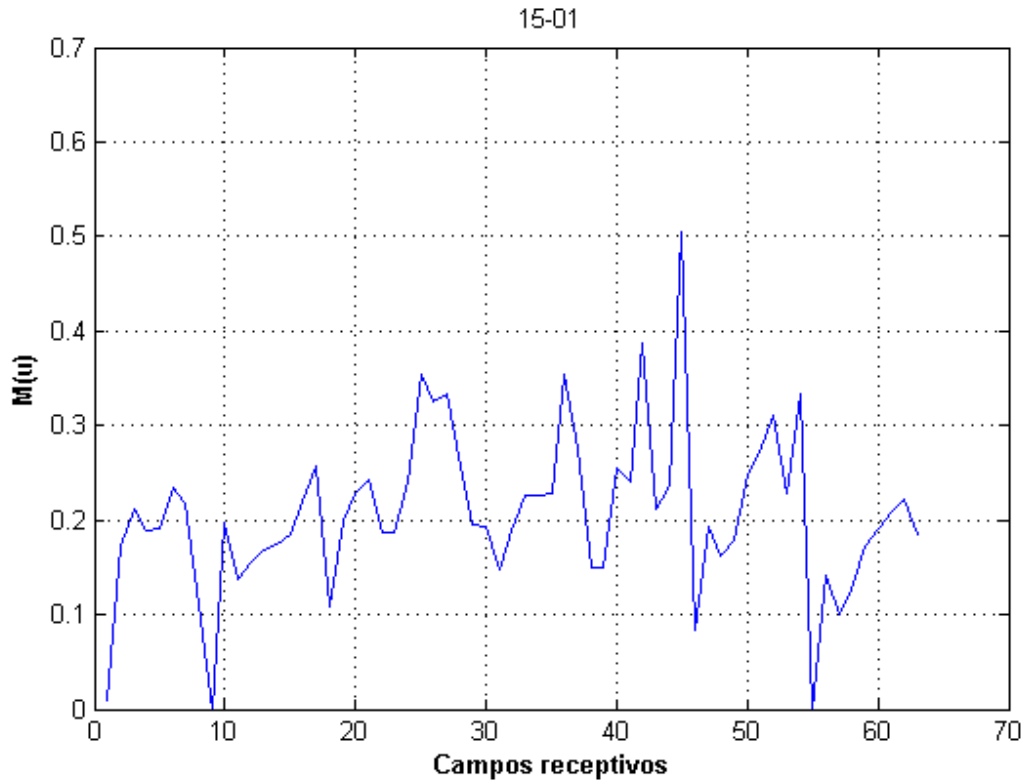


Figura 4.6. Vector de características de la figura 4.5.

Con los vectores característicos generados de todas las imágenes de los usuarios se procede con el entrenamiento del perceptrón multicapa, donde se busca que el usuario a aprender adopte valores cercanos a uno mientras que el resto a cero.

Como ejemplo de entrenamiento del perceptrón multicapa, para el usuario con identificación número 15. Se tomaron 10 vectores por usuario para entrenamiento y los vectores restantes se utilizaron para la etapa de prueba. En la figura 4.7 se presentan los resultados obtenidos de la verificación mediante el perceptrón multicapa.

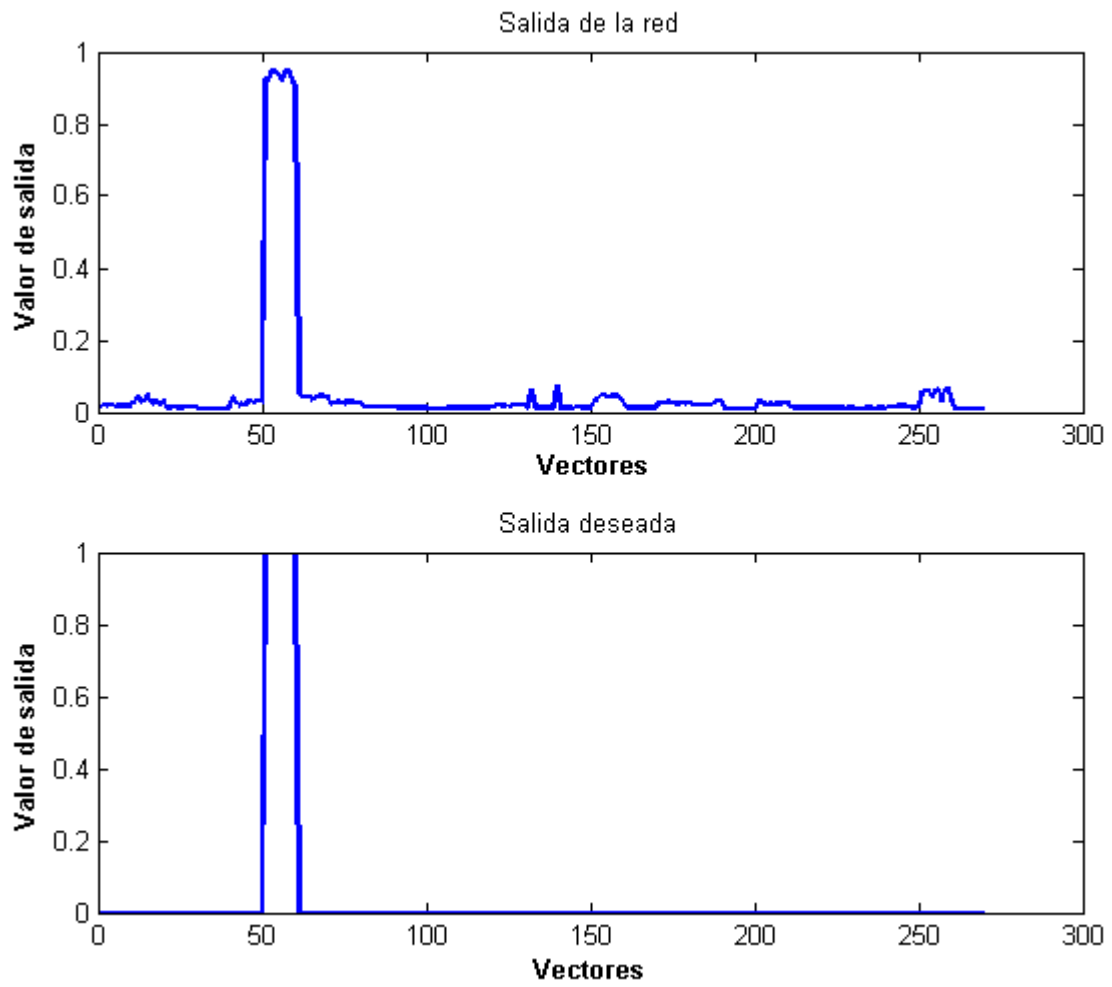


Figura 4.7. Resultado de entrenamiento del perceptrón para la persona con identificación número 15.

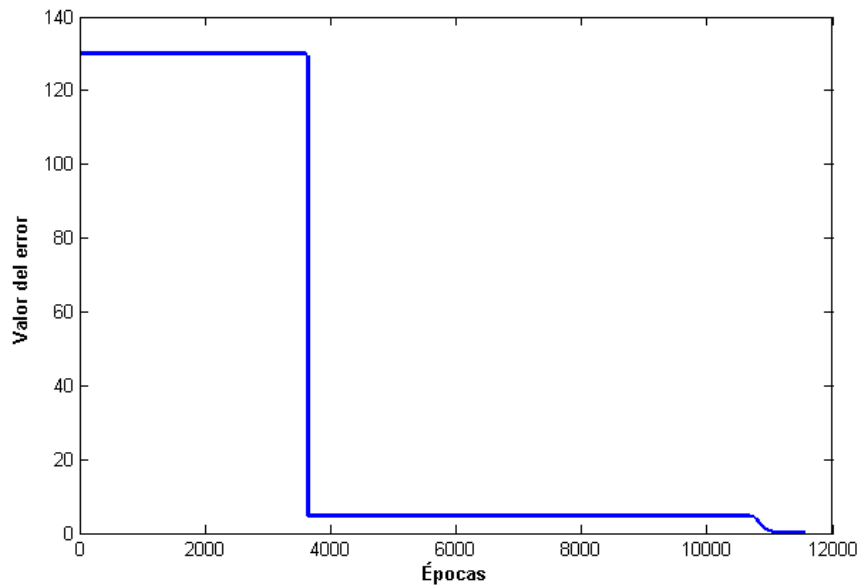


Figura 4.8. Gráfica del error.

Después de cada entrenamiento de los usuarios, los pesos obtenidos eran guardados para utilizarlos en la etapa de verificación de otros vectores de prueba de los mismos. En la siguiente figura se muestra el resultado de salida con respecto a los vectores de prueba, en este caso la salida deseada pertenece a los 10 vectores del usuario 15 y 10 vectores de los demás usuarios.

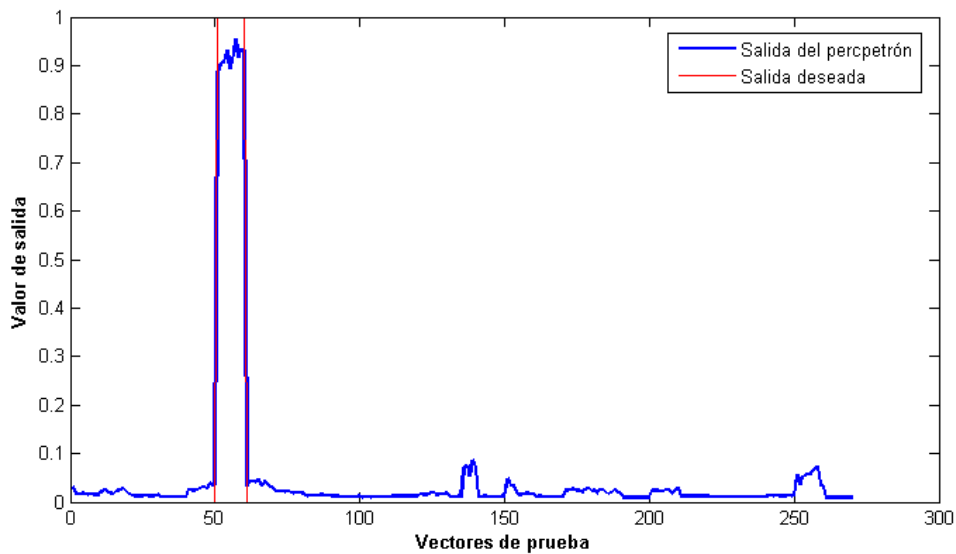


Figura 4.9. Resultado del usuario 15 con los vectores de prueba

Como se mencionó anteriormente se cuenta con 10 usuarios de la base de datos creada, además se toman en cuenta 14 usuarios de la base de datos GTAV (6 vectores por persona) y otras imágenes obtenidas en otra sesión de otros usuarios (6 usuarios) todos ellos no formaron parte del entrenamiento de la red. Cuyo fin es el de apreciar la respuesta de la red ante vectores desconocidos. Se puede observar en la figura 4.10 que el valor de salida de la red es bajo con respecto al umbral establecido que es de 0.8, por lo tanto, para estos vectores la respuesta del perceptrón multicapa es correcta.

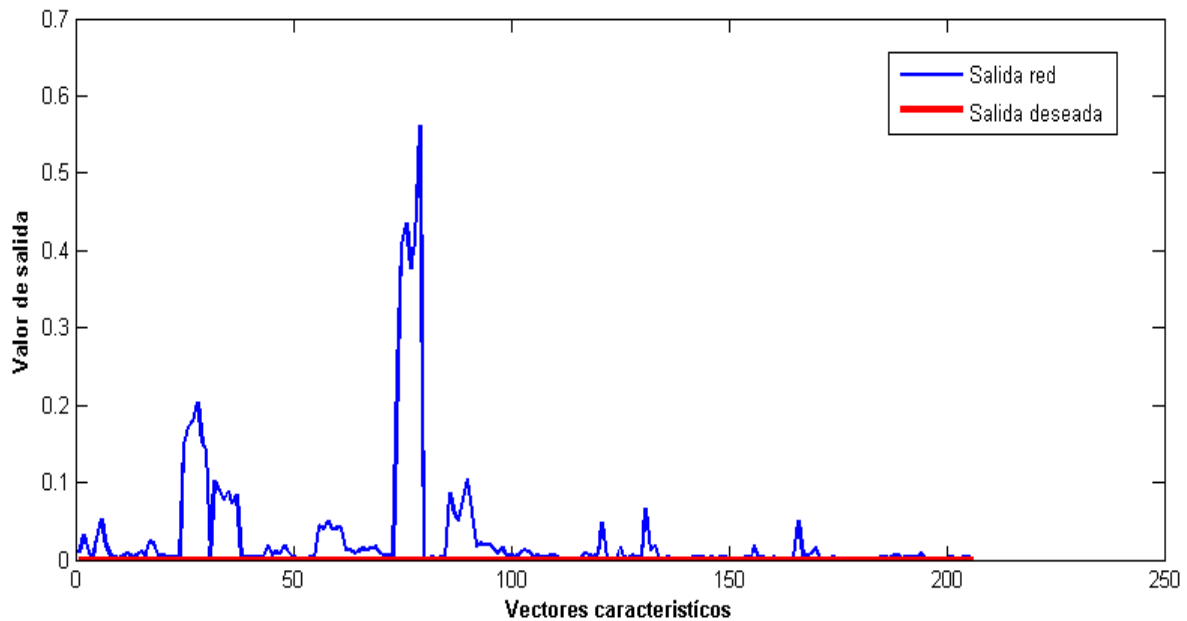


Figura 4.10. Resultado de la salida del perceptrón respecto a los vectores intrusos.

Cuadro 4.2. Resultados de la verificación con pose frontal y condiciones de iluminación controlada.

Usuario	Número de pruebas		Positivos	Negativos	FRR (%)	FAR (%)	Precisión (%)
	# TP	# TN	10	206			
1	8	206	2	0	20.00	0.00	90
2	10	206	0	0	0.00	0.00	100
3	7	206	3	0	30.00	0.00	85
4	10	206	0	0	0.00	0.00	100
5	9	206	1	0	10.00	0.00	95
6	7	206	3	0	30.00	0.00	85
7	10	204	0	2	0.00	0.97	99.77
8	10	206	0	0	0.00	0.00	100
9	10	206	0	0	0.00	0.00	100
10	2	203	8	3	80.00	1.46	59.65
11	10	206	0	0	0.00	0.00	100
12	10	206	0	0	0.00	0.00	100
13	10	206	0	0	0.00	0.00	100
14	10	206	0	0	0.00	0.00	100
15	10	206	0	0	0.00	0.00	100
16	10	206	0	0	0.00	0.00	100
17	9	206	1	0	10.00	0.00	95
18	10	206	0	0	0.00	0.00	100
19	10	206	0	0	0.00	0.00	100
20	10	206	0	0	0.00	0.00	100
21	3	206	7	0	70.00	0.00	65
22	9	204	1	2	10.00	0.97	94.77
23	6	206	4	0	40.00	0.00	100
24	10	206	0	0	0.00	0.00	100
25	10	206	0	0	0.00	0.00	100
26	10	206	0	0	0.00	0.00	100
27	6	206	4	0	40.00	0.00	80
TOTAL:	236	5555	34	7	12.5926	0.13	94.415

El número de pruebas negativas está conformada por 30 personas desconocidas, las cuales no forman parte del entrenamiento. 10 personas son de

la base de datos obtenida, 6 personas más donde sus imágenes son capturadas en otra sesión y 14 personas de la base de datos de GTAV; donde se obtuvieron 6 vectores característicos por persona (total de 180), más 26 vectores de las personas que fueron parte del entrenamiento (1 vector por persona usando otra imagen que no fue considerada en el entrenamiento ni en la prueba).

Se puede observar que los resultados del cuadro 4.1 tienen un porcentaje aceptable, esto sucede debido a que las imágenes de cada usuario fueron capturadas bajo condiciones controladas (iluminación, pose, expresión facial moderada y la distancia entre cámara-usuario) y en una sola sesión. Sin embargo, el porcentaje de precisión podría disminuir si se consideran cambios de iluminación, diferentes poses y expresiones faciales, con lo cual sería necesario un entrenamiento del perceptrón multicapa con todas las variaciones posibles del rostro para cada usuario.

Evaluación del algoritmo en condiciones no controladas

El reconocimiento (verificación o identificación) facial bajo diferentes condiciones de iluminación y pose resulta ser más complejo debido a que se presenta mayor variabilidad de la imagen facial a pesar de ser el mismo usuario.

Para observar el desempeño del algoritmo ante condiciones de iluminación y pose variables se utilizaron las bases de datos FEI y GTAV. Las imágenes de la base de datos GTAV fueron capturadas bajo tres iluminaciones diferentes (medio ambiente o la luz natural, fuente de luz fuerte desde un ángulo de 45° , y, finalmente, una medio-fuerte fuente de luz casi frontal. Las imágenes de la base de datos FEI cuenta con fondo blanco homogéneo en posición vertical con rotación de perfil de hasta aproximadamente 180 grados.

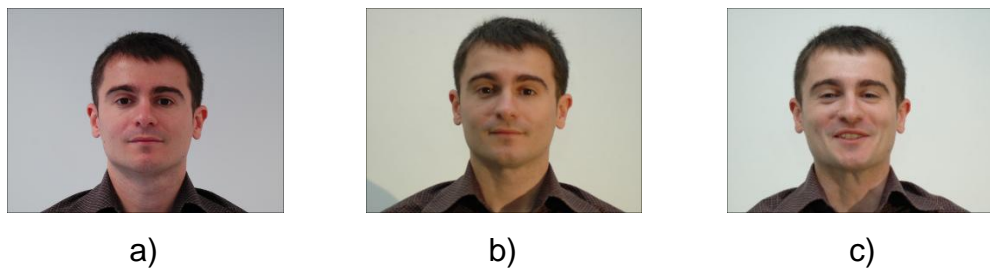


Figura 4.11. Ejemplos de la base de datos GTAV bajo diferentes condiciones de iluminación, a) Fuente de luz natural, b) Fuente de luz fuerte y c) Fuente de luz medio fuerte.

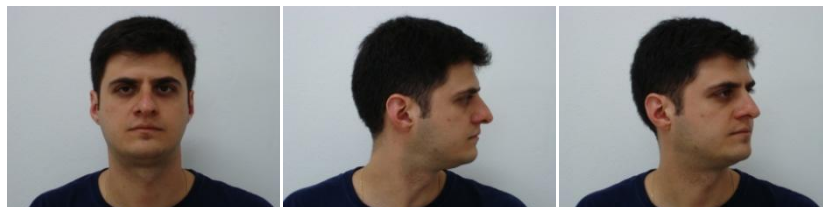




Figura 4.12. Ejemplos de la base de datos FEI con diferentes poses.

Segmentación y recorte facial ante diferentes condiciones de iluminación



Figura 4.13. Segmentación y recorte por medio de apertura viscosa.

Algunas imágenes con una iluminación fuerte se tuvo problemas en la clasificación de píxeles de color de piel y por lo tanto no se logra una segmentación adecuada de la imagen facial (figura 4.14).



Figura 4.14. Segmentación por medio de apertura viscosa, a) Imagen original y b) Imagen facial segmentado.

Se puede observa la figura 4.14 donde la segmentación no tuvo un resultado favorable, por lo tanto se tiene que compensar la iluminación para obtener un mejor resultado. Una técnica sencilla de compensar la iluminación es como se presenta en Wang et al. (2009). Siguiendo los siguientes pasos:

1. Calcular el promedio en cada componente RGB ($avgR, avgG, avgB$), el promedio en gris: $avgGray = (avgR + avgG + avgB)/3$.
2. Calcular los factores de ajuste de R, G, B. $aR = avgGray/avgR$, $aG = avgGray/avgG$ y $aB = avgGray/avgB$.
3. Ajustar los valores de R, G y B. $R = R * aR$, $G = G * aG$, $B = B * aB$
4. Si los valores de R, G y B son mayores a 255, se tienen que ajustar a 255.

Si se toma la figura 4.14 y se realiza una compensación de iluminación mejora el resultado de la segmentación del rostro como se puede observar en la figura 4.15



Figura 4.15. Segmentación por medio de apertura viscosa, a) Imagen con iluminación compensada y b) Imagen facial segmentado.

Segmentación y recorte facial ante variación de pose



Figura 4.16. Segmentación y recorte por medio de apertura viscosa de imágenes de base de datos FEI.

Resultados de la verificación de rostros bajo condiciones no controladas

Cuadro 4.3. Resultados de la verificación con pose frontal y diferentes condiciones de iluminación.

Pruebas negativas: 76								
Pruebas positivas	Usuario	# TP	# FN	# TN	# FP	FRR (%)	FAR (%)	Precisión (%)
5	1	2	3	75	1	60.0	1.3	69.33
6	2	3	3	76	0	50.0	0.0	75.00
3	3	3	0	76	0	0.0	0.0	100.00
4	4	2	2	76	0	50.0	0.0	75.00
3	5	1	2	76	0	66.7	0.0	66.66
5	6	2	3	75	1	60.0	1.3	69.33
3	7	1	2	74	2	66.7	2.6	65.33
4	8	4	0	76	0	0.0	0.0	100.00
4	9	3	1	76	0	25.0	0.0	87.50
4	10	1	3	76	0	75.0	0.0	62.50
7	11	5	2	76	0	28.6	0.0	85.71
3	12	3	0	64	12	0.0	15.8	92.00
TOTAL:		30	21	896	16	41.2	1.8	79.03

Cuadro 4.4. Resultados de la verificación bajo condiciones de iluminación controlada y con diferentes poses.

Pruebas negativas: 501								
Pruebas positivas	Usuario	# TP	# FN	# TN	# FP	FRR (%)	FAR (%)	Precisión (%)
6	1	0	6	493	8	100.0	1.6	49.20
6	2	1	5	499	2	83.3	0.4	58.13
6	3	3	3	499	2	50.0	0.4	74.80
4	4	0	4	501	0	100.0	0.0	50.00
6	5	3	3	501	0	50.0	0.0	75.00
6	6	2	4	495	6	66.7	1.2	66.06
6	7	0	6	490	11	100.0	2.2	48.90
6	8	1	5	498	3	83.3	0.6	58.03
4	9	2	2	500	1	50.0	0.2	55.63
7	10	1	6	489	12	85.7	2.4	69.90
7	11	3	4	490	11	57.1	2.2	55.94
7	12	0	7	500	1	100.0	0.2	64.29
7	13	4	3	499	2	42.9	0.4	70.32
6	14	3	3	501	0	50.0	0.0	62.50
7	15	4	3	498	3	42.9	0.6	49.90
7	16	3	4	497	4	57.1	0.8	78.37
6	17	1	5	499	2	83.3	0.4	75.00
5	18	0	5	496	5	100.0	1.0	78.27
6	19	1	5	495	6	83.3	1.2	64.29
5	20	3	2	490	11	40.0	2.2	71.13
5	21	2	3	501	0	60.0	0.0	58.13
5	22	2	3	498	3	60.0	0.6	49.50
5	23	2	3	501	0	60.0	0.0	73.90
6	24	2	4	501	0	66.7	0.0	57.73
6	25	0	6	501	0	100.0	0.0	78.90
11	26	8	3	501	0	27.3	0.0	70.00
7	27	1	6	486	15	85.7	3.0	69.70
7	28	2	5	501	0	71.4	0.0	70.00
7	29	2	5	501	0	71.4	0.0	66.67
6	30	3	3	490	11	50.0	2.2	86.36
	TOTAL:	59	126	14911	119	68.1	0.8	65.21

Conclusiones

En el presente trabajo, se desarrolló la verificación de rostros de pose frontal, poniendo especial énfasis a la detección y recorte del rostro. El algoritmo propuesto realiza una segmentación del color de piel dada una imagen de entrada, clasificando los píxeles como piel o no con base a unos umbrales en Cb y Cr del espacio de color YCbCr. Otro punto importante en esta primera etapa es el recorte del rostro, para lo que se presentaron dos técnicas para realizar esta función. La primera técnica consistió en considerar las proporciones en cuanto al largo y ancho del rostro; debido a que estos valores varían de persona en persona en algunos casos se presentó una desventaja al momento de realizar el recorte, puesto que se recupera el rostro incompleto en la parte del mentón. La segunda técnica consistió en el uso de una apertura viscosa junto con una apertura morfológica; en esta técnica se tiene como desventaja de que cuando en el área detectada, el área del rostro resulte menor en comparación al área del cuello, esto genera una imagen del rostro que incluye una parte del cuello; este caso se presenta cuando algunas regiones no deseadas en la imagen entran dentro de los umbrales de la piel. Para este trabajo se utilizaron de ambas técnicas ofreciendo mejores resultados para el recorte, primero se aplica una apertura viscosa y por último se utiliza las proporciones.

Además, se utilizó una plantilla facial para confirmar si la región detectada y recortada pertenece al rostro o no mediante una correlación normalizada, con lo que obtiene un buen resultado de confirmación con una precisión de más del 90% considerando las imágenes de la base de datos propia.

Para la etapa de la extracción de características se emplearon 40 filtros Gabor a diferentes frecuencias y orientaciones, la división de la imagen fue de 7x9 secciones o celdas, con lo cual se obtiene un vector característico de 63 valores normalizados entre 0 y 1. Para la clasificación de los vectores se implementó un perceptrón multicapa, con una tasa de clasificación bastante aceptable del

94.41%, considerando las imágenes de la base de datos obtenida de 37 personas y con un umbral de aceptación igual a 0.8. Sin embargo, como se pudo apreciar los cuadros 4.3 y 4.4 cuando se consideran diferentes condiciones de iluminación y pose el porcentaje de la tasa de clasificación baja de manera significativa, por lo tanto, el algoritmo es poco robusto bajo esas condiciones.

Por otro lado, para reducir el número de falsos negativos y positivos, es conveniente que la captura de las imágenes sea tomada en varias sesiones y con variación de pose y expresión facial con el fin de generar vectores con mayor variabilidad de cada usuario y realizar el entrenamiento con las distintas imágenes.

Como trabajo futuro, se puede aplicar filtros de Gabor solo a puntos importantes del rostro como en los ojos, boca, nariz; para extraer características relevantes, y observar el desempeño de la verificación. También, sería ideal probar con otro clasificador como máquina de soporte de vectores para observar la respuesta de la clasificación.

LITERATURA CITADA

Abate, A.F. ; Ricciardi, S. and Sabatino, G. 2007. 3D Face Recognition in a Ambient Intelligence Environment Scenario, In Face Recognition, (1-14), K. Delac and M. Grgic, Eds., In-Tech Education and Publishing, Vienna, Austria.

Aguilar-torres, G., G. Sanchez-perez, M. Nakano-miyatake, and H. Pérez-meana. 2007. Face Recognition Algorithm Using the Discrete Gabor Transform. 17th International Conference on Electronics, Communications and Computers.

Albiol, A., D. Monzo, A. Martin, J. Sastre. 2008. Face recognition using HOG-EBGM. Pattern Recognition Letters, 29: pp.1537-1543.

Amjad, A., A. Griffiths, and M. N. Patwary. 2012. Multiple face detection algorithm using colour skin modelling, IET Image Process., 6(8):1093–1101.

Atkinson, Anthony P. and Ralph Adolphs. 2011. The neuropsychology of face perception: beyond simple dissociations and functional selectivity. Philosophical transactions of the Royal society.

Barraclough N. E., Perrett D. I. 2011 From single cells to social perception. Phil. Trans. R. Soc. B 366, 1739–1752.

Berreti, S. A. Del Bimbo, P. Pala, and F. J. S. Mata. 2007. Geodesic distances for 3D-3D and 2D-3D face recognition, in Proc. IEEE Int. Conf. Mul- timed. Expo, pp. 1515–1518.

Bianconi, F. and A. Fernández. 2007. Evaluation of the effects of Gabor filter parameters on texture classification. Pattern Recognit., 40(12): 3325–3335.

Bronstein, A. M., M. M. Bronstein, R. Kimmel. 2005. Three-dimensional face recognition. In International Journal of Computer Vision 64(1): 5-30.

Bui, Len; Tran Dat, Huang Xu y Chetty Girija. 2011. Face recognition based on Gabor features.3rd European Workshop on Visual Information Processing. M. C.

Campadelli, P. and R. Lanzarotti. 2005. A Face Recognition System Based on Local Feature Characterization. In *Advanced Studies in Biometrics*, Vol.3161, *Lecture Notes in Computer Science*, M. Tistarelli, J. Bigun, and E. Grosso, Eds. Berlin: Springer.

Chandrappa, D. N., M. Ravishankar, and D. R. RameshBabu. 2011. Face detection in color images using skin color model algorithm based on skin color information. *3rd Int. Conf. Electron. Comput. Technol.*, pp. 254–258.

Daugman, J.G. 1985. Uncertainty Relation for Resolution in Space, Spatial-Frequency, and Orientation Optimized by TWO- dimensional Visual Cortical Filters, *J. OSA*, 2(7): 1 160- 11 69.

FEI Face Database. Centro Universitário da FEI. 2012. Disponible en: <http://fei.edu.br/~cet/facedatabase.html>

Gabor, D. 1945. *Theory of communication*.

Gong, S., S. J. McKenna, and A. Psarrou. 2000. *Dynamic Vision: From Images to Face Recognition*. London, U. K.: Imperial College Press.

Gonzalez, R. C. and R. E. Woods. 2007. *Digital Image Processing (3ra. Ed)*. Prentice Hall, New Jersey.

Gottumukkal, R. and Asari, V. K. 2004. An improved face recognition technique based on modular PCA approach, *Pattern Recognition Letter*, 25(4): 429–436.

Heijmans, H. 1994. *Morphological Image Operators (Advances in Electronics and Electron Physics; Series Editor: P. Hawkes)*. Boston: Academic Press.

Heron-Delaney, M., Wirth S., Pascalis O. 2011. Infants' knowledge of their own species. *Phil. Trans. R. Soc. B* 366, 1753–1763.

Horton, D., R. Road, and W. Yorkshire. 2010. Facial image processing for facial analysis. IEEE International Carnahan Conference on Security Technology(ICCST).

Hsu, R. and S. Member. 2002. Face Detection in Color Images. 24(5):696–706.

Huang, Z.-K. , W.-Z. Zhang, H.-M. Huang, and L.-Y. Hou. 2008. Using Gabor Filters Features for Multi-Pose Face Recognition in Color Images. Second Int. Symp. Intell. Inf. Technol. Appl., pp. 399–402.

Ilonen, J. , J.-K. Kamarainen, and H. Kalviainen. 2007. Fast extraction of multi-resolution Gabor features. 14th Int. Conf. Image Anal. Process. (ICIAP 2007), no. Iciap, pp. 481–486.

Jafri, R and H. R. Arabnia,H.R. 2009. A Survey of Face Recognition Techniques. J. Inf. Process. Syst., 5(2):41- 68.

Jain, A. K. A. Ross and S. Prabhakar. 2004. An Introduction to Biometric Recognition. IEEE Trans. on Circuits y Systems for Video Technology, 14(1):4–20.

Jones, J. P. and Palmer, L. A. 1987. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. J. of Neurophysiology, 58:1233–1258.

Kamarainen, J. 2012. Gabor Features in Image Analysis. Image Processing Theory, Tools and Applications. pp. 4–5.

Kamarainen, J.-K. , V. Kyrki, and H. Kälviäinen. 2006. Invariance properties of gabor filter-based features--overview and applications. IEEE Trans. Image Process., 15(5):1088–99.

Kherchaoui, S., A. Houacine, B. P. El Alia, and B. Ezzouar. 2010. Face Detection Based On A Model Of The Skin Color With Constraints And Template Matching. (1): 469–472.

Kim, J. O., S. J. Seo, C. H. Chung, and J. Hwang. 2004. Face Detection by Facial Features with Color Images and Face Recognition Using PCA. Pp 1-8. In Computational science and its applications. A. Laganà et al. (Eds.). Springer-Verlag Berlin Heidelberg.

Kruger, V. & Sommer, G. 2002a. Gabor wavelet networks for efficient head pose estimation. *Image and Vision Computing*, 20: 665-672.

Kruger, V. & Sommer, G. 2002b. Wavelet networks for face processing. *Journal of the Optical Society of America A-Optics Image Science and Vision*, 19:1112-1119.

Liu, Q., H. Lu, and S. Ma. 2004. Improving Kernel Fisher Discriminant Analysis for Face Recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14:42-49.

Liu, H. and H. Wechsler. 2002. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Trans. Image Process.*, 11(4):467–76.

Lu, X. and A.K. Jain. 2006. Deformation Modeling for Robust 3D Face Matching. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1377-1383.

Maajid, S., M. Muhammad, Y. Javed, and A. Anjum. 2006. Face Recognition using Bank of Gabor Filters. pp. 144–150.

Maragos, P., Vachier, C. 2008. A PDE formulation for viscous morphological operators with extensions to intensity-adaptative operators. In: *Proc. 15th IEEE-International Conference in Image Processing*, San Diego, California, pp. 2200–2203.

Messer, K., J. Kittler, M. Sadeghi, S. Marcel, C. Marcel, S. Bengio, et al., 2003. Face Verification Competition on the XM2VTS Database. *Proc. of the Int. Conf. on AVBPA*, pp. 964–974.

Nhat, V. D. M. and S.Y. Lee. 2007. Kernel-Based 2DPCA for Face Recognition, Proc. of the Int. Symp. on Signal Processing y Information Technology, pp 35–39.

Phillips, P. J., H. Moon, S. A. Rizvi, and P. J. Rauss. 2000. The FERET evaluation methodology for face-recognition algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(10):1090-1104.

Phung, S. L., A. Bouzerdoun, and D. Chai. 2005. Skin segmentation using color pixel classification: analysis and comparison. IEEE Trans. Pattern Anal. Mach. Intell., 27(1): 148–54.

Pinsk, MA., Desimone K, Moore T, Gross CG, Kastner S. 2005. Representations of faces and body parts in macaque temporal cortex: A functional MRI study. Proc Natl Acad Sci U S A. 102:6996–7001.

Rahman, T. and A. Bhuiyan. 2008. Face Recognition using Gabor Filters. Proceedings of 11th International Conference on Computer and Information Technology. 2: 25–27.

Reed, Russell and Robert Marks. 1999. Neural Smithing – Supervised Learning in Feedforward Artificial Neural Networks. The MIT Press.

Ronse, C. 2005. Mathematical Imaging and Vision, Special Issue on Mathematical Morphology after 40 years, 22:103–106.

Russ, J. 2007. The image processing handbook. 6th ed. CRC Press. London, NY.

Sagheer, A. and S. Aly. 2012. An Effective Face Detection Algorithm Based on Skin Color Information. Eighth Int. Conf. Signal Image Technol. Internet Based Syst., pp. 90–96.

Sanchez-Cuevas, M. C., R. M. Aguilar-Ponce, and J. L. Tecpanecatl-Xihuitl. 2013. A Comparison of Color Models for Color Face Segmentation, Procedia Technol., 7(444):134–141.

Santillán, Israel, Ana M. Herrera-Navarro, Jorge D. Mendiola-Santibáñez, Iván R. Terol-Villalobos. 2010. Morphological Connected Filtering on Viscous Lattices. *J Math Imaging Vis.*, 36: 254–269.

Sarfraz, M., Olaf Hellwich and Zahid Riaz. 2010. Feature Extraction and Representation for Face Recognition. Pp. 9-20. In *Face Recognition*. M. Oravec. In-Tech. Croacia.

Schmitt, M. y Mattioli, J. 1993. *Morphologie Mathématique*. Paris: Masson.

Serra, J. 2002. Viscous lattices. In: Talbot, H., Beare, R. (eds.) *Mathematical Morphology*, pp. 79–89. CSIRO Publishing, Melbourne.

Serra, J. 2005. Viscous lattices. *J.Math. Imaging Vis.* 22(2–3), 269–282.

Serrano, Á. , I. Martín de Diego, C. Conde, and E. Cabello. 2011. Analysis of variance of Gabor filter banks parameters for optimal face recognition. *Pattern Recognit. Lett.*, 32(15): 1998–2008.

Serrano, Á. 2009. Parametrización óptima de un banco de filtros de Gabor para su aplicación a un problema de reconocimiento facial. Universidad Rey Juan Carlos, Madrid España.

Shen, L., L. Bai, and M. Fairhurst. 2007. Gabor wavelets and General Discriminant Analysis for face identification and verification. *Image Vis. Comput.*, 25(5): 553–563.

Shen, L., L. Bai. 2006. A review on Gabor wavelet for face recognition. *Pattern Analysis Appl.* 9:273-292.

Shin, H., S. D. Kim, and H. C. Choi. 2007. Generalized elastic graph matching for face recognition. *Pattern Recognition Letters*, 28:1077–1082.

Shin, K. I. Chang, L. V Tsap, and N. Dame. 2002. Does Colorspace Transformation Make Any Difference on Skin Detection ?. Proceedings Sixth IEEE Workshop on Applications of Computer Vision.

Singh, S. K., D. S. Chauhan, M. Vatsa, and R. Singh. 2003. A Robust Skin Color Based Face Detection Algorithm. 6(4): 227–234.

Stan, Z and K. Jain. 2011. Handbook of face recognition. Springer. USA.

Struc, Vitomir, Rok Gajsek and Nikola Pavesic. 2009. Principal Gabor Filters for Face Recognition. IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems.

Sun, N., H.-x. Wang, Z.-h. Ji, C.-r. Zou, and L. Zhao. 2008. An efficient algorithm for Kernel two-dimensional principal component analysis, Neural Computing & Applications, 17: 59-64.

Tarrés, F, A. Rama. “GTAV Face Database”. Disponible en: <https://gtav.upc.edu/research-areas/face-database>.

Tsao, DY., Freiwald WA, Tootell RB, Livingstone MS. 2006. A cortical region consisting entirely of face-selective cells. Science. 311:670–4.

Turk, M. and A. Pentland. 1991. Eigenfaces for Recognition. J. of Cognitive Neuroscience, 3(1):71-86.

Vachier, C.,Meyer, F. 2007. News from viscous land. In: Banon, G.J.F., Barrera, J., Braga-Neto, U.D.M., Hirata, N.S.T. (eds.) Proceedings of the 8th International Symposium onMathematical Morphology, San José dos Campos, Brazil, pp. 189–200.

Wan, Y. H., S. M. Ji, Y. Xie, X. Zhang, and P. J. Xie. 2004. Video program clustering indexing based on face recognition hybrid model of hidden Markov model and support vector machine. In Combinatorial Image Analysis, Proceedings, Vol.3322, Lecture Notes In Computer Science, pp.739-749.

Wang, B., X. Chang, and C. Liu. 2009. A Robust Method for Skin Detection and Segmentation of Human Face. *Second Int. Conf. Intell. Networks Intell. Syst.*, pp. 290–293.

Wechsler, H. 2007. *Reliable Face Recognition Methods*. Springer Verlag.

Wiskott, L. N. Kr, and C. Von Der Malsburg. 1999. Face Recognition by Elastic Bunch Graph Matching. pp. 1–23.

Yang, M.-H. D. J. Kriegman, and N. Ahuja. 2002. Detecting faces in images: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(1): 34–58.

Yin, H., P. Fu, and S. Meng. 2005. An Efficient Face Detection Method in Color Images 2 Face Detection Algorithm. pp. 875–880.

Zhao, W, R. Chellappa, P. Phillips and A. Rosenfeld. 2003. Face recognition: A literature survey. *ACM Computing Survey*, 35(4): 399–458.

Zhao, W., R. Chellapa. 2005. *Face processing advanced modeling and methods*. Academic Press.USA.

Zhou, X., H. Seibert, C. Busch, and W. Funk. 2008. A 3D Face Recognition Algorithm Using Histogram-based Features. *EurographicsWorkshop on 3D Object Retrieval*.

Zhou,S. V. Krueger, and R. Chellappa. 2002. Face Recognition from Video: A condensation Approach. in *Proc. of Fifth IEEE International Conference on Automatic Face and Gesture Recognition*. Washington D.C., USA, pp.221-228.