



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN

**MODELOS BIOLÓGICOS DE VISIÓN APLICADOS AL
RECONOCIMIENTO DE ROSTROS**

T E S I S

QUE PARA OPTAR POR EL GRADO DE:
MAESTRO EN INGENIERÍA
(COMPUTACIÓN)

P R E S E N T A

JORGE PANTALÉON MARTÍNEZ

TUTOR

DR. BORIS ESCALANTE RAMÍREZ
FACULTAD DE INGENIERÍA

CODIRECTOR

DR. URIEL RODRIGO NAVA VELAZCO
FACULTAD DE INGENIERÍA

MÉXICO D.F. AGOSTO 2014

© agosto 2014 – Jorge Pantalón Martínez

– All rights reserved.

– Todos los derechos reservados.

A mis padres y hermanos a quienes les agradezco el apoyo que siempre me han brindado.

Agradecimientos

Dr. Boris Escalante Ramírez por haberme otorgado la oportunidad de integrarme al grupo del laboratorio de procesamiento de imágenes, además de su paciencia, confianza y apoyo durante estos años.

Dr. Uriel Nava por su amistad, apoyo, paciencia, orientación, y dirección de este trabajo, sus comentarios y revisiones del mismo.

Dr. Pablo Perez Alcázar por su ayuda en la revisión de este escrito y por las sugerencias del mismo.

Dra. Lucía Medina Gómez por la revisión de este documento.

Dr. Fernando Arámbula Cosío por su valiosa ayuda y aclaraciones en este trabajo.

Dra. Nidiyare Helvia Montiel por participar en las revisiones y ser parte del grupo de sinodales.

Al Posgrado en Ciencia e Ingeniería de la Computación de la UNAM, a los docentes y personal (Lulú, Diana y Amalia) por los servicios que me prestaron durante mis estudios, que además fueron de excelente calidad.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por haber otorgado la beca para la realización de los estudios de maestría con el número de registro 419835.

A la Universidad Nacional Autónoma de México (UNAM) por todo su apoyo a lo largo de mis estudios desde la preparatoria, licenciatura y posgrado, también por permitirme ser parte de una institución de excelencia y aportarme una formación humana y académica de la máxima calidad.

Modelos biológicos de visión aplicados al reconocimiento de rostros

Resumen

La tecnología de reconocimiento de rostros es un área de investigación muy activa en los últimos años, retoma muchas disciplinas, y tienen una gran aceptación por los investigadores en diferentes campos: ciencias de la computación, ingeniería, matemáticas, física, neurociencia y ciencia cognitiva. Una de las razones que ha llevado a la aceptación de esta tecnología es la necesidad cada vez mayor de aplicaciones de seguridad y vigilancia utilizadas en diferentes ámbitos de la vida real. Este trabajo propone un método para reconocimiento de rostros de bajo costo computacional e invariante a los cambios de iluminación, pose y expresión facial. Nuestro enfoque se divide en tres etapas fundamentales: 1) Extracción de características faciales globales utilizando la transformada Hermite discreta (DHT); 2) Extracción de características faciales locales utilizando patrones binarios locales (LBP) ; 3) Además el uso del análisis discriminante lineal (LDA) utilizado para construir un espacio característico de pocas dimensiones. La representación DHT y LBP describe rasgos faciales más robustos y sencillos para su reconocimiento que la representación de la imagen original. Para la etapa de clasificación de patrones utilizamos el método de la mínima distancia, usando 3 métricas: distancia euclideana, Canberra y cityblock. La combinación DHT, LBP y LDA nos permite eliminar o al menos reducir los inconvenientes de muchos métodos de reconocimiento de rostros. Las pruebas realizadas con diferentes combinaciones de DHT y LBP así como una comparación con algunos de los métodos comunes en el área del reconocimiento de rostros se presentan en este trabajo.

Índice general

| | Página |
|---|---------------|
| 1 Introducción | 1 |
| 1.1 Motivación | 2 |
| 1.2 Retos y desafíos | 2 |
| 1.3 Objetivos y contribución | 4 |
| 1.4 Metas | 4 |
| 2 Antecedentes | 5 |
| 2.1 Sistema de visión humano | 5 |
| 2.1.1 Anatomía | 6 |
| 2.1.2 La retina como parte del sistema nervioso central | 7 |
| 2.1.3 Proyección de los campos receptivos a la corteza visual | 11 |
| 2.2 Tecnología de reconocimiento de rostros | 13 |
| 2.3 Técnicas de preprocesamiento | 13 |
| 2.4 Extracción de características | 14 |
| 2.4.1 Basados en características invariantes | 15 |
| 2.4.2 Métodos holísticos | 17 |
| 2.4.3 Métodos híbridos | 22 |
| 2.5 Bases de datos y sus evaluaciones | 22 |
| 2.5.1 Base de datos FERET | 22 |
| 2.5.2 Base de datos YALE | 23 |
| 2.5.3 Base de datos YALE Extendida B | 23 |
| 2.5.4 Base de datos ORL | 24 |
| 2.6 Aplicaciones comerciales | 25 |
| 3 Sistema propuesto | 27 |
| 3.1 Metodología | 27 |
| 3.2 Transformada Hermite | 30 |
| 3.2.1 Transformada Polinomial | 30 |
| 3.2.2 Transformada Hermite en una dimensión | 31 |
| 3.2.3 Transformada Hermite discreta | 34 |
| 3.2.4 Transformada Hermite discreta 2D | 35 |
| 3.2.5 Vectores característicos | 36 |
| 3.3 Patrones Locales Binarios (LBP) | 38 |
| 3.3.1 Funcionamiento | 38 |
| 3.3.2 Modificación del método original de Patrones Locales binarios | 39 |
| 3.3.3 Descripción del rostro usando LBP | 40 |
| 3.4 Análisis discriminante lineal (LDA) | 42 |
| 3.4.1 Discriminante lineal de Fisher | 43 |

| | |
|---|-----------|
| 3.5 Reconocimiento de rostros (Clasificador) | 44 |
| 3.5.1 Clasificación por mínima distancia | 44 |
| 3.5.2 Métricas utilizadas en el reconocimiento de características biométricas | 45 |
| 4 Resultados experimentales | 47 |
| 4.1 Experimentos | 47 |
| 4.2 Evaluación de los métodos propuestos sin el uso del LDA | 49 |
| 4.3 Evaluación de los métodos propuestos usando LDA | 53 |
| 4.4 Comparación con otros métodos | 56 |
| 5 Conclusiones y trabajo futuro | 57 |
| A Anexo A | 59 |
| A.1 Base de Datos ORL | 59 |
| Bibliografía | 65 |

Índice de figuras

| | Página |
|--|---------------|
| 2.1 Esquema del proceso visual humano. | 6 |
| 2.2 Estructura del ojo humano. Obtenido de teleformacion.edu.aytolacoruna.es/FISICA | 7 |
| 2.3 Distribución de bastones y conos en la retina humana. Imagen tomada de la página www.hgculiacan.com | 8 |
| 2.4 Diagrama simple de la organización de la retina. [1]. | 10 |
| 2.5 Esquema de campo receptivo ON-OFF. | 12 |
| 2.6 Vía visual principal vista desde la base del cerebro [1]. | 13 |
| 2.7 Diagrama general de los sistemas de reconocimiento de rostros. | 14 |
| 2.8 Plantillas geométricas(blancas)[2]. | 15 |
| 2.9 Características faciales identificadas manualmente[3]. | 16 |
| 2.10 <i>Grids</i> para el reconocimiento de rostros[4]. | 16 |
| 2.11 Ejemplo de un conjunto de entrenamiento[5]. | 18 |
| 2.12 Cara promedio[5]. | 18 |
| 2.13 7 <i>eigenfaces</i> de la Fig 2.11[5]. | 19 |
| 2.14 La misma persona vista bajo diferentes condiciones de iluminación puede cambiar la apariencia dramáticamente así como su identidad [6]. | 20 |
| 2.15 A) <i>Eigenfaces</i> y sus variaciones de iluminación, b) <i>Fisherfaces</i> mostrando lo poco que le afecta los cambios de iluminación [7]. | 20 |
| 2.16 Ejemplo de imágenes FERET de [8]. | 23 |
| 2.17 Ejemplo de imágenes Yale, se muestra las diferentes variaciones de imágenes para una sola persona [6]. | 23 |
| 2.18 Ejemplo de imágenes Yale Extendida B [9]. | 24 |
| 2.19 Ejemplo de 3 sujetos y sus grupos de imágenes de la base ORL [10] | 24 |
| 2.20 Aplicaciones comerciales en reconocimiento de rostros [11]. | 25 |
| | |
| 3.1 Función de sensibilidad al contraste como envolvente de los canales selectivos a frecuencias espaciales .Obtenida de www.contrastsensitivity.net/es/cm.html | 29 |
| 3.2 Diagrama simplificado del sistema propuesto. | 30 |
| 3.3 Diagrama de bloques de la transformada polinomial directa e inversa. | 31 |
| 3.4 Diagrama de bloques de la transformada polinomial directa e inversa. | 33 |
| 3.5 Descomposición de una imagen con la transformada Hermite 2do Orden. | 36 |
| 3.6 Variación vectores característicos intraclase usando la DHT. | 37 |
| 3.7 Variación vectores característicos interclase usando la DHT. | 38 |
| 3.8 Máscara rectangular de 3 x 3, el algoritmo LBP calcula comparaciones entre un píxel central y sus vecinos. En este ejemplo, el valor central es PC = 10 y la etiqueta final es 27 | 38 |
| 3.9 Los patrones circulares (8;1),(16;2) y (8;2) | 39 |
| 3.10 Histogramas con LBP Uniforme. | 40 |

| | | |
|------|---|----|
| 3.11 | Imagen del rostro dividida en $7 \times 7, 5 \times 5$ y 3×3 | 41 |
| 3.12 | Histograma mejorado espacialmente 4 regiones LBP (24;3) | 41 |
| 3.13 | Clases bien separadas en un espacio 2D. | 42 |
| 4.1 | Imágenes ejemplo de la base de datos <i>ORL</i> | 48 |

Índice de tablas

| | Página |
|--|---------------|
| 4.1 Hermite global | 49 |
| 4.2 LBP global | 50 |
| 4.3 LBP 4 y 16 regiones | 50 |
| 4.4 Hermite combinación LBP globales | 51 |
| 4.5 Hermite combinación LBP 4 regiones | 51 |
| 4.6 Hermite combinación LBP 16 regiones | 52 |
| 4.7 Hermite global combinación LDA | 53 |
| 4.8 LBP 4 y 16 regiones combinación LDA | 54 |
| 4.9 Hermite combinación LBP globales y LDA | 54 |
| 4.10 Hermite combinación LBP 4 regiones y LDA | 55 |
| 4.11 Hermite combinación LBP 16 regiones y LDA | 55 |
| 4.12 Comparación de rendimiento con algunos métodos todos de extracción de características faciales. | 56 |

1

Introducción

Identificar una persona es una actividad que para el ser humano es obvia y relativamente fácil. La sociedad moderna requiere sistemas de identificación personal confiables y prácticos, lo cual ha generado un enorme interés en el desarrollo de métodos de identificación biométricos. Estos métodos utilizan características fisiológicas intrínsecas de un individuo (p.ej. huellas dactilares, rostro, iris, voz, etc), un objetivo es la identificación automática de una persona en diferentes escenarios del mundo real, lo que ha permitido aumentar la eficiencia y seguridad en la identificación y autenticación de individuos.

Entre los métodos biométricos más usados en los últimos años tenemos al reconocimiento de rostro. Algunas propuestas en la tecnología de reconocimiento de rostros utilizan modelos inspirados en sistemas biológicos de visión [12]. En esta tesis utilizamos un modelo que incorpora algunos fenómenos simples de la visión que se describen a detalle en el capítulo 2.

En este capítulo se presenta la motivación y objetivos del trabajo. También se reseñan los retos y desafíos que enfrentan en la actualidad los métodos para el reconocimiento de rostros.

1.1 Motivación

Los avances tecnológicos aplicados al reconocimiento de personas se han diversificado a diversos ambientes de la vida común, por ejemplo: ingresos seguros a computadoras, salas de acceso, entradas a edificios, aeropuertos, estadios, eventos donde se necesita un acceso o asegurarse que la persona correcta tenga acceso.

Los primeros trabajos en el área de reconocimiento de personas datan de finales del siglo XIX, con el uso de la huella dactilar; el uso que se le dio fue identificar personas utilizando bases de datos de esas huellas dactilares, el objetivo era realizar comparaciones para identificar personas en escenas de crímenes.

El uso de la huella dactilar es el antecedente inmediato de un sistema biométrico, esta palabra deriva de los vocablos griegos "*bios*" (vida) y "*metrón*" (medición) [13]. Estos sistemas están basados en características biológicas, intrínsecas de la persona y por tanto difícilmente transferibles o sustraíbles.

El proceso general de autenticación con métodos biométricos siguen algunos pasos en general: captura de los datos del usuario a validar; extracción de características específicas de la muestra; comparación de esas características con los registros de una base de datos; y por último validación de los datos para aceptar o rechazar al usuario.

Algunas de las medidas biométricas que existen son relativamente confiables [14]; por ejemplo, el uso de la huella digital, el iris o la voz; sin embargo el reconocimiento de rostros es una alternativa que nos ofrece ventajas sobre el resto de las otras métricas. El reconocimiento de rostros destaca por ser una técnica con una alta capacidad de respuesta frente a múltiples características y se encuentra entre los más aceptados por el usuario, [14][15], lo que hace que los sistemas basados en rostros sean útiles para el desarrollo de aplicaciones **no intrusivas**.

Debido a ciertos problemas en los sistemas biométricos mencionados en [14], han surgido opciones que se han enfocado al campo de reconocimiento de rostros; es por eso que hoy en día muchos sistemas se desarrollan utilizando el rostro como una manera mucho más práctica para el reconocimiento; lo cual indica, que se está trabajando mucho en esta área, además de prometer grandes expectativas en el campo del reconocimiento de rostros, se espera en un futuro excelentes resultados en este tipo de sistema biométrico.

1.2 Retos y desafíos

El reconocimiento de rostros es una tarea que, sin esfuerzo, realizamos los humanos diariamente. Constituye un reto científico, que ha ocupado a diferentes grupos de investigación de todo el mundo en los últimos años.

Aunque se han hecho grandes avances en el campo del reconocimiento de rostros, aún quedan abiertos demasiados desafíos (p.ej. iluminación, identidad, pose, expresiones faciales, etc), a pesar de la capacidad computacional que tenemos hoy en día están lejos de resolverse.

Existen en la literatura algoritmos que tratan de resolver esos problemas [16], pero lejos están de tener éxito, debido a que las imágenes tienen una gran variabilidad en distintos factores que impiden un buen rendimiento. Los principales obstáculos para el reconocimiento de rostros son:

- **Orientación.** La imagen de un rostro varía dependiendo de su pose. Un rostro puede aparecer en la imagen de frente, de perfil, rotado o inclinado. Esto puede provocar que la apariencia varíe significativamente con el simple cambio de pose.
- **Condiciones de la imagen.** Durante el proceso de adquisición de las imágenes se pueden encontrar diferentes condiciones de iluminación y ruido, debido a la forma de adquisición o del sensor con las que son adquiridas, estos elementos cambian significativamente la apariencia del rostro.
- **Expresión facial.** La apariencia puede ser modificada con la expresión facial de la persona.
- **Oclusión.** El rostro puede cubrirse con objetos en elementos clave (p.ej. ojos, nariz y boca), dificultando su identificación (presencia o ausencia de anteojos, barba, bigotes, gorras).
- **Envejecimiento.** La apariencia de la persona cambia significativamente con los años, ya que se generan arrugas y cambios en la piel, esto también complica la identificación adecuada de la persona.

Algunos de los algoritmos más comunes tienden a codificar los rostros como vectores de características, el tamaño de los vectores de características debe ser lo más compacto posible para evitar el problema de la dimensionalidad. Los vectores característicos usados en algunos de los métodos del estado del arte del reconocimiento de rostros, tienen una dimensión muy alta; un ejemplo sencillo de estos casos: dada una imagen de solo 48x48 píxeles se obtiene un vector de 2304 elementos [17][18]. Debe considerarse que el tamaño de un conjunto de entrenamiento debe aumentar exponencialmente con la dimensión del espacio de entrada. Uno de los problemas es que se tiene un conjunto de datos limitado. Por lo tanto la importancia de disminuir la dimensión de los vectores, existen numerosas técnicas de reducción de dimensionalidad, que buscan representaciones más compactas de los datos, generando algoritmos de clasificación eficientes que conduzcan un ahorro en espacios físicos y de memoria en la computadora.

Entonces, un desafío importante para estos sistemas de reconocimiento de rostros es obtener una representación de baja dimensionalidad que permita el manejo eficiente y de bajo costo computacional de los datos.

1.3 Objetivos y contribución

El objetivo de esta tesis es desarrollar un sistema clasificador de rostros de bajo costo computacional y que sea robusto a cambios de iluminación, postura y expresión facial, utilizando modelos que combinen descriptores globales inspirados en el sistema de visión, junto con descriptores locales de textura.

El sistema propuesto consiste en las siguientes etapas principales: 1) Extracción de características globales del rostro por medio de la transformada Hermite y características locales mediante descriptores de regiones locales. 2) Combinación de características globales y locales en un solo vector. 3) Reducción de dimensión utilizando análisis de Fisher. 4) Clasificación mediante distancias.

Las principales contribuciones de esta tesis son:

1. Se emplea de manera novedosa la transformada Hermite en la caracterización de rostros.
2. Combinación de métodos globales y locales, haciendo más robusta la representación de los rostros.
3. Espacios característicos de bajas dimensiones mediante la técnica del discriminante de Fisher.

En los siguientes capítulos se presentan en detalle las propiedades teóricas y tecnológicas de esta metodología desarrollada, así como su desempeño.

Este trabajo se concentra únicamente en la tarea de reconocimiento, por lo tanto, se supone que la detección y normalización han sido realizadas con anterioridad.

Se comprueba la eficiencia del sistema comparando los resultados obtenidos con otros métodos propuestos en la literatura.

1.4 Metas

1. Analizar y revisar el estado del arte en reconocimiento de rostros.
2. Diseñar un modelo de extracción de características usando las herramientas propuestas anteriormente que sea robusto a cambios de iluminación, expresión facial y pose.
3. Realizar la etapa de extracción de características.
4. Evaluar el método propuesto, su rendimiento y sus posibles alcances.
5. Diseñar el clasificador que nos de los mejores resultados.
6. Evaluar el rendimiento del método propuesto.

2

Antecedentes

EN este capítulo se realiza una revisión del estado del arte en el área del reconocimiento de rostros, con un apartado sobre conceptos de algunos de los procesos que el sistema de visión de algunos mamíferos, lleva a cabo durante el procesamiento de la información visual, además de algunos esquemas generales de los métodos de reconocimiento de rostros. Se continuará con un estudio de las principales técnicas de procesamiento y métodos de extracción de características en rostros propuestos hasta nuestros días. Se finaliza con la muestra de algunas bases de datos de rostros más usadas en la literatura, además de mencionar aplicaciones comerciales de reconocimiento de rostros usados hoy día.

2.1 Sistema de visión humano

Muchos algoritmos de procesamiento de imágenes han sido ideados y propuestos como modelos de visión biológicos de primates y seres humanos. A continuación se dará una revisión general que Estudillo planteó en [19], describiendo órganos y procesos involucrados en el procesamiento de la visión.

El proceso de la visión se divide en tres etapas: la primera corresponde al procesado óptico, el cual comienza en el ojo cuando la luz viaja a través de él hasta la retina; la segunda está determinada por la fotorrecepción, aquí la luz que entra al ojo hace contacto primeramente con la cornea, donde es refractada la luz provocando y logra que la imagen converja en su camino hacia el iris y la pupila. En situaciones de baja intensidad luminosa la pupila será más grande, permitiendo el paso de la luz suficiente para formar una imagen discernible y haciendo posible el registro de dicha imagen

mediante los procesos fisiológicos que convierten la energía luminosa en impulsos nerviosos, donde la retina captura la luz a través de sus fotorreceptores y células epiteliales pigmentarias. Las moléculas de fotorpigmento del fotorreceptor absorben la luz provocando un cambio de potencial en la membrana del fotorreceptor. Con esto se inicia una serie de señales que viajan a través de las neuronas de la retina hacia el nervio óptico para llegar al cerebro. La señal es entonces recibida y procesada por el cerebro como una imagen interpretable y finalmente, el procesado neural de la imagen que permite la percepción de la misma, Fig.2.1.

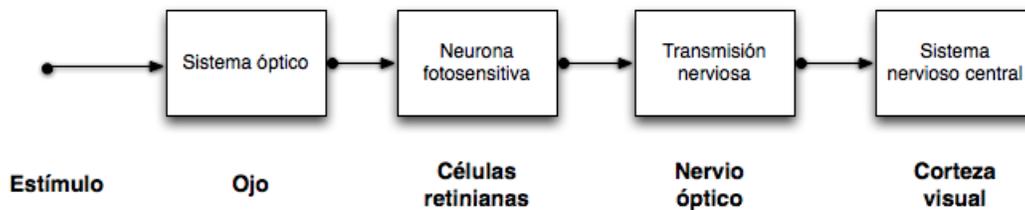


Figura 2.1: Esquema del proceso visual humano.

2.1.1 ANATOMÍA

A continuación se describirán brevemente las partes que interactúan en el proceso visual en el orden en el que la luz va encontrando a su paso por el ojo, hasta darle una interpretación como imagen.

Cornea

En primer lugar, como se mencionó anteriormente, la luz atraviesa la córnea, Fig. 2.2, que es un tejido transparente avascular pero con gran número de terminaciones nerviosas. Tiene unos 12 mm de diámetro, unos 0.6 mm de espesor central y un índice de refracción de 1.366. Siguiendo las leyes principales de la refracción, cuando la luz incide sobre un medio que posee un índice de refracción mayor que el del aire, la velocidad de propagación disminuye y la luz cambia de dirección. La cornea podría ser un ejemplo de ese medio, ya que cuando la luz alcanza la superficie de la cornea, la constante de refracción disminuye.

Pupila e iris

Inmediatamente después de la cornea se halla el iris, Fig. 2.2, que es músculo coloreado que actúa como diafragma, controlando la cantidad de luz que entra en el ojo. La apertura del diafragma y por donde penetra la luz es lo que se conoce como pupila, cuyo diámetro puede variar desde 2 mm hasta 8 mm en un adulto joven. Esta variación producida en función del nivel de iluminación, controla la cantidad de luz que llega a la retina y juega un papel fundamental en la calidad de la imagen retiniana. En caso de que exista demasiada luz, el iris disminuye la apertura de la pupila con la ayuda del músculo esfínter, ya que demasiada luz podría provocar opacidad al momento de la formación de la imagen. Lo opuesto ocurre cuando la luz es pobre; la pupila es dilatada para

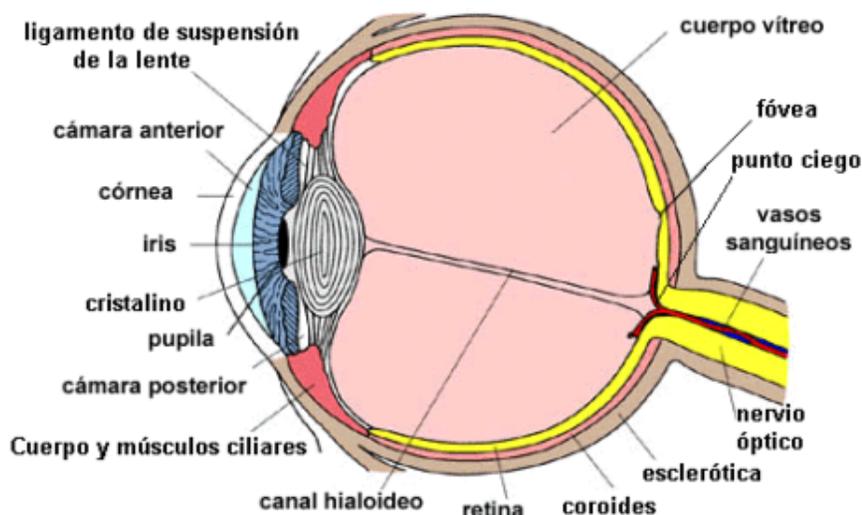


Figura 2.2: Estructura del ojo humano. Obtenido de teleformacion.edu.aytolacoruna.es/FISICA

tratar de capturar tantos fotones como sea posible para la formación adecuada de la imagen.

La lente

Después de atravesar la pupila, la luz encuentra la lente del cristalino, una estructura transparente que, controlada por el músculo ciliar, es capaz de modificar su propia forma para conseguir una imagen nítida en la retina.

La lente se sostiene y se mantiene en posición gracias a las fibras zonulares o ligamentos de suspensión. Estos ligamentos o fibras están unidos al área de los músculos ciliares que rodea al ojo llamado cuerpo ciliar. El cuerpo ciliar y las fibras zonulares trabajan en conjunto para alterar el punto focal del ojo. Se le llama acomodación al proceso por el cual la contracción del músculo ciliar hace que la lente adquiera una forma más redondeada. Como resultado de la acomodación, las imágenes son enfocadas y enviadas a la retina.

2.1.2 LA RETINA COMO PARTE DEL SISTEMA NERVIOSO CENTRAL

La luz procedente de la lente es enviada a la retina, que es la membrana que contiene a los fotorreceptores, llamados conos y bastones. Son células que cuando reciben un estímulo luminoso adecuado se excitan y transducen energía electromagnética en señales electroquímicas, que se transmiten a través de sucesivas neuronas en la propia retina y luego hacia el cerebro. La parte funcional de la retina cubre toda la porción posterior del ojo, excepto la mancha ciega, que es una salida del nervio óptico llamada papila o disco óptico. La agudeza visual es máxima en la parte central de la retina llamada la mancha amarilla o mácula lútea, sobre todo en la fóvea, una depresión con forma de fosa en el centro de la mácula donde se enfocan los objetos de mayor

interés visual; es el centro de fijación o de enfoque cuando los rayos de luz se dirigen hacia el objetivo visual correspondiente. La fovea contiene la densidad más alta de fotorreceptores [20].

Organización de la retina

La retina puede ser dividida en varias capas externas e internas. La capa que se encuentra hasta el fondo de la retina es la capa de epitelio pigmentario (CEP), la cual ayuda a evitar que la luz se disperse por todo el globo ocular permitiendo una visión nítida, Fig. 2.3. Esta capa, además, otorga el soporte a las células fotorreceptoras que se encuentran arriba de esta capa. La capa de los fotorreceptores (CF) consiste de los segmentos interno y externo de los conos y bastones, los cuales reciben y transforman los fotones de la luz. El núcleo de estos fotorreceptores se encuentra en la capa nuclear externa (CNE) mientras que sus axones y terminaciones celulares se encuentran en la capa plexiforme externa (CPE). La capa nuclear interna (CNI) contiene el núcleo de las células bipolares, a las células horizontales y a la mayor parte de las células amacrinas. La capa plexiforme interna (CPI) es el lugar donde se lleva a cabo la comunicación vertical entre las células bipolares y ganglionares, haciendo un segundo contacto sináptico. En la última capa, la capa de células ganglionares (CCG), se encuentran los cuerpos de las células ganglionares. Las dendritas de estas células se extienden hacia la (CPI), mientras que sus axones lo hacen en la dirección opuesta, o sea a la capa de fibras del nervio óptico (CFNO).

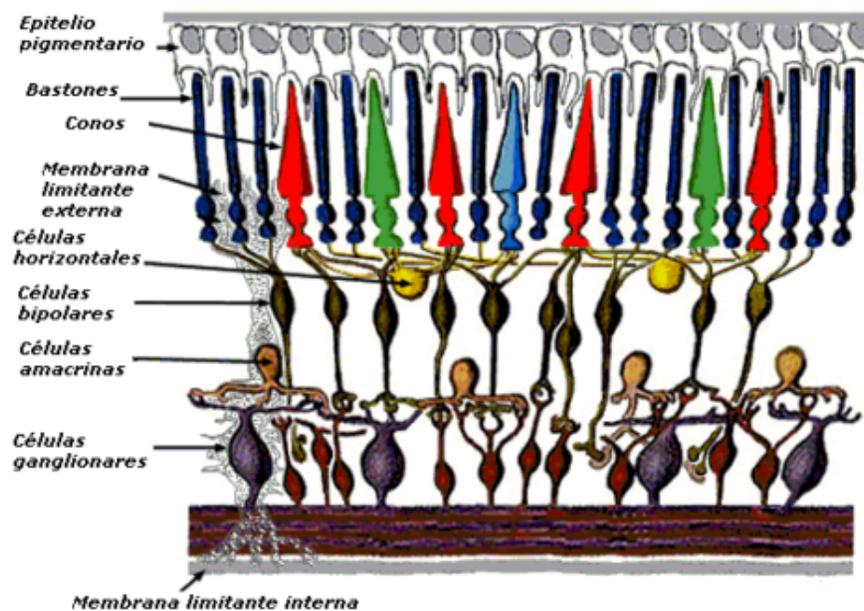


Figura 2.3: Distribución de bastones y conos en la retina humana. Imagen tomada de la página www.hgcultiacan.com

Conos y bastones

Las células fotorreceptoras son las responsables del inicio de una serie de eventos en cascada para convertir una imagen proyectada sobre el tejido de la retina, de fotones a señales electroquímicas que pueden ser procesadas por el cerebro. A esta conversión se le denomina fototransducción. Los dos tipos de células involucradas en este proceso son las células fotorreceptoras llamadas bastones y conos. De los 130 millones de fotorreceptores que poseen las retinas de los seres humanos, aproximadamente cerca de 120 millones son estructuras largas y cilíndricas conocidas como bastones.

Los bastones son altamente sensibles a la luz y por ello están especializados en la detección de luz de baja intensidad (visión escotópica), siendo homogéneos en su sensibilidad de longitud de onda. Los conos son usualmente más cortos que los bastones y están especializados en la detección de intensidades altas que registran el color y detalles finos; individualmente son más sensibles a longitudes de onda largas (conos L), medias (conos M) y cortas (conos C) [20]. El elemento más crítico en este proceso de fototransducción es el fopigmento que contienen los conos y bastones; una proteína sensible a la luz llamada opsina. En los bastones, cuando un solo fotón es absorbido por la opsina, se genera un cambio de potencial que genera a su vez un impulso eléctrico. Los conos poseen tres distintos tipos de opsina que son los responsables de los tres colores primarios (rojo, azul, amarillo).

Mencionado anteriormente, los conos son menos sensibles por lo que necesitan una longitud de onda muy específica para iniciar un impulso eléctrico. La fovea, la región más sensible y de mayor agudeza de la retina, contiene exclusivamente conos. La densidad de los conos decrece al aumentar la distancia a la fovea. Los bastones por su parte, no están presentes en la fovea y se encuentran mayormente concentrados en la parafovea [1]. En la zona del disco óptico, que es el lugar donde se juntan los axones de las células ganglionares para abandonar el ojo y formar el nervio óptico, no existen fotorreceptores y constituye un punto ciego. La distribución y proporción de los conos y bastones varía entre las especies de animales. En la Fig. 2.4 se muestra un gráfico de la densidad de conos y bastones en diferentes partes de la retina.

Células bipolares

El siguiente conjunto de neuronas son las llamadas células bipolares, cuya función es propagar en la dirección vertical la comunicación. El cuerpo de estas células reside en la CNI de la retina, Fig. 2.3 mientras que sus dendritas reciben las señales de los fotorreceptores en la primera unión sináptica. En la terminación opuesta del cuerpo de la célula, la señal viaja a través del axón de la célula bipolar para realizar una segunda sinapsis con la siguiente neurona vertical, la célula ganglionar.

Neuronas laterales: células horizontales y amacrinas

Las señales eléctricas que fluyen a través de las neuronas verticales no son completamente independientes entre sí, ya que muchas están ligadas por neuronas laterales. Un tipo de estas neuronas

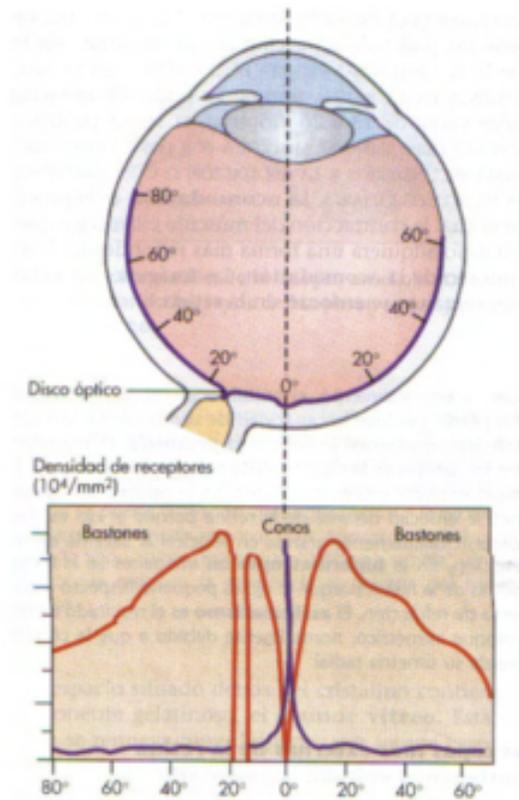


Figura 2.4: Diagrama simple de la organización de la retina. [1].

laterales son las células horizontales que se encuentran en la CNI de la retina Fig. 2.3.

Comúnmente estas células están ligadas a más de un fotorreceptor, lo que significa que las células bipolares subsecuentes reciben señales de más de un fotorreceptor. Estas células hacen conexiones inhibitorias con las células bipolares cuando hay luz en su alrededor; en otras palabras, estas células tienen como función la variación del contraste. Por su parte, las células amacrinas forman conexiones entre neuronas en la dirección vertical, en la capa interna, algunas veces con las células ganglionares. La función de estas células no es clara aún pero se cree que contribuyen al efecto de contraste.

Células ganglionares

Las últimas neuronas de la red que reciben las señales son las células ganglionares de la retina. Cuando se activan debido a la llegada de una señal, la célula ganglionar produce un potencial de acción que comienza su travesía hacia los axones de la célula. Los axones de las células ganglionares de la retina convergen, formando el nervio óptico. El nervio óptico, compuesto de un manojó de fibras ópticas, representa el camino de las señales que se dirigen hacia el cerebro.

2.1.3 PROYECCIÓN DE LOS CAMPOS RECEPTIVOS A LA CORTEZA VISUAL

Campos receptivos

En cada una de las estructuras descritas anteriormente, una célula individual responde a imágenes en una pequeña parte del campo de visión y sólo responde fuertemente a patrones particulares de la imagen. La parte del campo de visión a la cual una célula responde se le llama el campo receptivo de la célula. A la relación entre los patrones de la imagen en el campo receptivo y la actividad de la célula es referida como las propiedades del campo receptivo de la célula [20].

El campo receptivo de un fotorreceptor suele ser una pequeña área circular que coincide con el área de la retina ocupada por éste. Las células bipolares y ganglionares de muchos mamíferos tienen una organización de campo receptivo antagonista concéntrico formado por dos áreas llamadas centro y periferia respectivamente. Las células bipolares y ganglionares pueden ser de dos tipos: de encendido central (también conocidas como células ON) o de apagado central (o células OFF) Fig. 2.5.

Estas células son excitadas por un estímulo en el centro de dicho campo receptivo e inhibido por otro estímulo en el área que rodea al centro, llamada periferia.

En las células bipolares, las respuestas al estímulo del área en torno al centro del campo receptivo quedan determinadas por las vías de interneuronas en las que participan las células horizontales [1]. Las salidas de estas células horizontales siempre son inhibitorias, lo que provoca que se mejore el contraste.

Existen tres tipos de células ganglionares, las cuales se clasifican de acuerdo a las funciones y extensión de sus campos receptivos. Las células ganglionares de tipo W constituyen el 40 % del total de las células ganglionares. La excitación que estas células reciben proviene mayormente de las células bipolares conectadas sinápticamente con los bastones. Las células ganglionares de tipo W tienen un campo receptivo amplio, ya que sus dendritas se extienden por toda la capa plexiforme interna y parecen estar especializadas en la detección de dirección de movimiento en el campo de visión.

El grupo de células ganglionares de tipo X es el más numeroso y representa un 55 % del total de las células ganglionares. Estas células tienen un campo receptivo pequeño debido a que sus dendritas no se extienden ampliamente en la retina, sino que solamente representan áreas discretas de la retina, transmitiendo los detalles finos de la imagen. Debido a que cada célula X recibe una señal de por lo menos un cono, la transmisión de una célula X es probablemente la responsable de toda la visión a color. El último tipo de células ganglionares y el menos numeroso, representando el 5 % del total de células ganglionares, es el tipo Y. Estas células poseen campos receptivos extensos y responden a cambios rápidos en la imagen visual y movimientos rápidos o cambios rápidos en la intensidad de la luz.

Corteza visual

Los nervios ópticos procedentes de ambos ojos convergen en el quiasma óptico. Algunas de las fibras de este nervio se cruzan en el quiasma y se unen al tracto óptico mientras que otras continúan

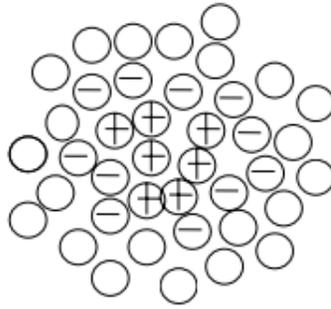


Figura 2.5: Esquema de campo receptivo ON-OFF.

su camino en sentido posterior en el tracto homolateral al ojo de origen. Como resultado de esta organización, cada tracto óptico contiene fibras cruzadas y no cruzadas. Las fibras de los tractos ópticos establecen sinapsis en el núcleo geniculado lateral (NGL) [1].

La mayoría de las neuronas del NGL proyectan la información visual hacia la corteza visual; sin embargo, algunas son interneuronas. En los primates, casi en un 90% de estas proyecciones se llevan a cabo sobre el NGL, donde los axones de las células ganglionares son agrupados en capas dependiendo del ojo de origen y otras propiedades Fig. 2.6. El NGL sirve de relevo y envía la información hacia la corteza visual primaria (V1). Las neuronas que hacen la sinapsis en el NGL, llamadas neuronas de relevo, poseen las mismas propiedades que las ganglionares W, X y Y respectivamente [20]. Los axones de las neuronas de relevo, que representan cada ojo son agrupados en una forma alternada, dando lugar a una banda llamada radiación óptica que viaja hacia la V1. Antes de llegar a la V1, los axones de las neuronas de relevo hacen contactos sinápticos con diferentes capas córticas.

En la corteza visual, existen neuronas en algunas capas que tienen propiedades de campo receptivo muy diferentes a las que se encuentran en el NGL. Estas células responden mejor a bordes o barras a una orientación en particular; estos bordes orientados son una característica importante para el sistema nervioso ya que frecuentemente los bordes definen los contornos de los objetos. V1 tiene dos tipos de neuronas sensibles a la orientación: células simples y células complejas. Los campos receptivos de células simples tienen regiones separadas que responden a variaciones de luz, respondiendo entonces a barras o bordes en una posición en particular en el espacio. Las células complejas en contraste, responden a la presencia de una barra localizada en cualquier lugar dentro de su campo receptivo y no tienen regiones específicas que puedan ser estimuladas por cambios de luz. Estas evidencias de Hubel y Wiesel provienen de estudios anatómicos y fisiológicos realizados en la corteza visual primaria de gato.

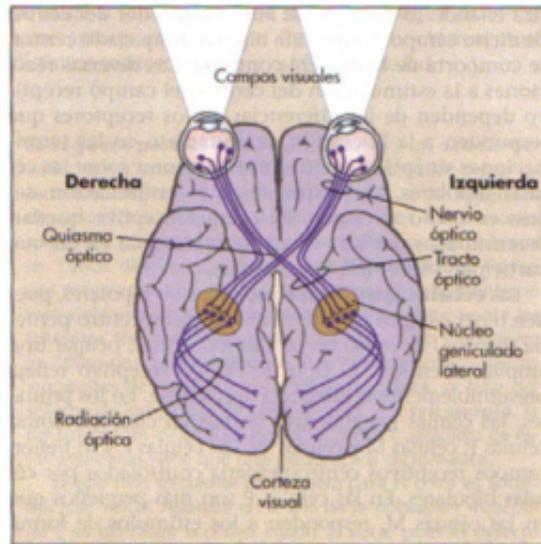


Figura 2.6: Vía visual principal vista desde la base del cerebro [1].

2.2 Tecnología de reconocimiento de rostros

A pesar de que el ser humano es capaz de reconocer los rostros de personas con relativa facilidad y más sorprendente poder memorizar gran cantidad de rostros en el transcurso de nuestras vidas, conseguir que las computadoras realicen ese mismo proceso no es nada sencillo y considerándolo aún muy complejo. Normalmente, para que una computadora pueda entender o realizar la difícil tarea de reconocer un rostro, se realizan descripciones de las imágenes para cuantizarlas, permitiéndonos usar técnicas estadísticas y de probabilidad; con los datos que se obtienen (extracción de características), se usan para compararlos con otras descripciones y finalmente se obtiene un valor numérico con el que se puede visualizar las similitudes que existen entre los diferentes rostros.

En el reconocimiento de rostros tenemos varios subprocesos, entre ellos la detección del rostro, el preprocesamiento de las imágenes, la extracción de características y la clasificación, Fig. 2.7. Estos procesos realizan lo siguiente: la segmentación de la región de la imagen en la que se encuentra localizado el rostro, la normalización de las imágenes durante el preprocesamiento y muchas veces, para acelerar procesos, la extracción de características que permiten una representación de baja dimensionalidad de los rostros mediante un mapeo del espacio original de la imagen a un subespacio característico más simple. La clasificación es la última etapa de un sistema común de reconocimiento de rostros; en esta etapa el subespacio característico es particionado y los vectores característicos extraídos son clasificados de acuerdo a su identidad.

2.3 Técnicas de preprocesamiento

Las imágenes de los rostros regularmente son tomadas en momentos diferentes o en diversas circunstancias que no pueden ser controladas, lo cual provoca que, aunque la imagen sea de la

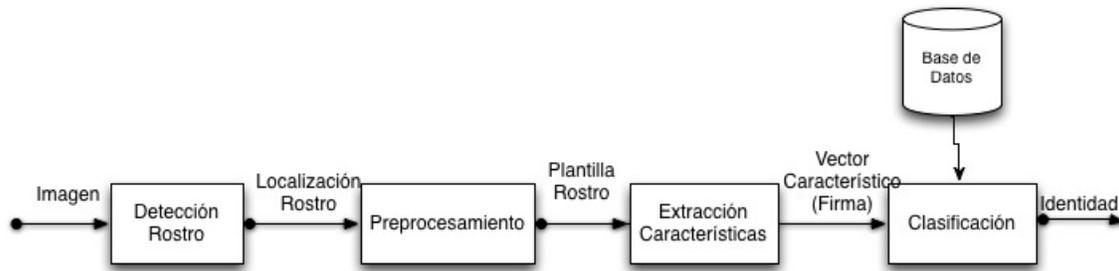


Figura 2.7: Diagrama general de los sistemas de reconocimiento de rostros.

misma persona, se tengan variaciones en cuanto a la iluminación, la orientación y el tamaño del rostro. Por tal motivo es que se necesita un preproceso de la imagen y así poder ser utilizada. Entre las tareas de preprocesamiento más comunes podemos encontrar las siguientes:

Normalización geométrica. Con esto se asegura que en el caso que los rostros no estén en la misma posición o en la misma orientación se ajustan las imágenes de tal manera que todas se encuentren en la misma posición.

Ecuilibración del histograma. Esta técnica de preprocesamiento intenta normalizar el histograma, es decir la transformación pretende obtener para una imagen un histograma con una distribución uniforme. Esto con el fin de que en nuestras imágenes se pueda reducir variaciones causadas por los cambios de iluminación.

Extracción de rostro (enmascaramiento). Consiste en eliminar información irrelevante en imágenes que contengan al rostro.

2.4 Extracción de características

El desempeño de un sistema de reconocimiento en general depende en buena medida de una buena extracción de características. En esta parte del sistema se detecta el rostro, se normaliza y se parametriza la imagen; después se obtiene el vector de características que represente a un rostro. La importancia en todo el proceso es que la información más discriminante de un rostro se extraiga y se elimine aquella que resulte irrelevante para el reconocimiento. En esta etapa existen técnicas que obtienen las características de manera manual, hasta técnicas más sofisticadas que extraen las características de forma automática.

Se espera que los métodos que extraen características tengan 3 objetivos principales:

- Reducir dimensiones del espacio original de la imagen.
- Maximizar la diferencia entre rostros de distintas personas.

- Minimizar la diferencias entre rostros de la misma persona.

La búsqueda de los objetivos mencionados, ha hecho que se creen un número grande de métodos, cuyo objetivo es la extracción de características del rostro. Según Zhao en [16] la extracción de características en los rostros se pueden clasificar en 3 categorías, las cuales son:

1. Basados en características invariantes (locales)
2. Métodos holísticos
3. Métodos híbridos

2.4.1 BASADOS EN CARACTERÍSTICAS INVARIANTES

Esta se basa en características de la imagen de entrada para identificar y extraer (o medir) un distintivo facial, por ejemplo, características tales como los ojos, la boca, la nariz, etc, así como otras marcas que nos sirvan como puntos de referencia y luego calcular las relaciones geométricas entre esos puntos faciales, reduciendo así la imagen de entrada a un vector de características geométricas. Aplicando estándares estadísticos de técnicas de reconocimiento de patrones se encuentran correspondencias con alguna otra imagen de rostro con la que se está midiendo.

Los primeros trabajos realizados en el reconocimiento de rostros automatizado se basaron principalmente en técnicas geométricas. Uno de los primeros en incursionar en esto fue Kanade [21], que empleó un método de procesamiento de imágenes sencillo para extraer un vector de 16 parámetros faciales, que eran las proporciones de distancias, áreas y ángulos (para compensar la variación del tamaño de las imágenes) y se utiliza una medida de distancia euclidiana simple de adaptación para conseguir un rendimiento máximo de 75 % en un base de datos de 20 personas diferentes con 2 imágenes por persona (una para referencia y una para la prueba).

Brunelli y Poggio [2], aprovechando el enfoque de Kanade, calculan un vector de 35 elementos geométricos, Fig. 2.8, de una base de datos de 47 personas (4 imágenes por persona) e informó un 90 % en la tasa de reconocimiento.

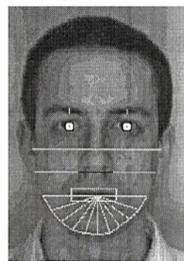


Figura 2.8: Plantillas geométricas(blancas)[2].

Técnicas más sofisticadas de extracción involucran plantillas deformables [22][23][24], transformada de Hough [25], el operador Reisfeld de simetría [26] y de Grafiltrado y operaciones

morfológicas [27]. Sin embargo, todas de estas técnicas se basan en gran medida en métodos heurísticos, tales como restringir el subespacio de búsqueda con las restricciones geométricas [28]. Lo más destacado en este tipo de métodos fue el trabajo de Cox [3], reportaron un rendimiento en el reconocimiento del 95 % en una base de datos de 685 imágenes (una sola para cada individuo) usando un vector de características de 30 elementos, derivados de 35 características faciales, Fig. 2.9.

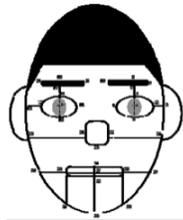


Figura 2.9: Características faciales identificadas manualmente[3].

Se aprecia que las características faciales se extrajeron manualmente, lo que es pensable asumir que el rendimiento del reconocimiento habría sido mucho menor que un sistema automatizado, y por lo tanto menos preciso.

Otro método muy conocido, basado en funciones, es el método del grafo elástico de correspondencia propuesto por Wiskott en, [29]. El reconocimiento del rostro se realiza mediante la comparación de su grafo con todas las imágenes de rostros conocidos y recogiendo el de mayor valor de similitud. Con esta arquitectura, la tasa de reconocimiento puede llegar a 98 % para el primer rango y el 99 % durante las primeras 10 filas con una galería de 250 individuos.

El sistema se ha mejorado para permitir manipular con diferentes poses, Fig. 2.10 [4], pero el rendimiento del reconocimiento en caras en la misma orientación sigue siendo el mismo. Aunque este método fue uno de los mejores en la evaluación FERET, en su época [30] [31], sufre el serio inconveniente de requerir la colocación gráfica de las primeras 70 caras que deben hacerse manualmente antes de usar el grafo elástico de correspondencia, esto hace que no se vuelve suficientemente confiable [32].

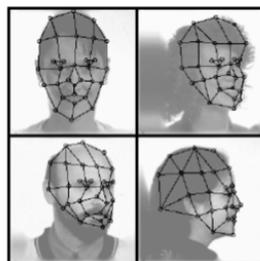


Figura 2.10: Grids para el reconocimiento de rostros[4].

Campadelli y Lanzarotti [33], experimentando con esta técnica, han eliminado la necesidad de hacer la colocación grafo manualmente utilizando modelos paramétricos, basados en las plantillas deformables propuestos en [22].

Ventajas y desventajas

La principal ventaja que ofrecen estos métodos destacamos que, dado que la extracción de los puntos característicos precede el análisis realizado para hacer coincidir la imagen para un individuo conocido, por ejemplo usando los grafos [4]. En principio, los basados en características con esquemas pueden ser invariante a la medida, orientación y/o iluminación [3].

Otros beneficios de estos regímenes incluyen una buena representación de la imágenes de la cara y la coincidencia es muy rápida. La principal desventaja de estos enfoques es la dificultad de la detección automática de características (quien implemente estas técnicas tiene que tomar decisiones arbitrarias sobre las características que son más importantes), por ejemplo en los siguientes métodos mencionados en [34]. Así mismo, estos tipos de métodos generalmente requieren imágenes de mayor resolución que los métodos holísticos.

2.4.2 MÉTODOS HOLÍSTICOS

Los métodos holísticos tratan de reconocer rostros utilizando representaciones globales de la imagen, es decir, toman las descripciones basadas como un todo en la imagen (cada pixel de la imagen es una característica para estos métodos) más que en las características locales del rostro. En la versión más simple de los métodos holísticos, la imagen se representa como una matriz 2D de valores de intensidad y el reconocimiento se realiza por comparaciones de correlación directa entre la cara de entrada y todas las otras caras de la base de datos. Aunque este enfoque ha sido demostrado que funciona [35] bajo circunstancias limitadas (p.ej. igualdad en iluminación, escala, postura, etc), otro punto que se menciona es que computacionalmente es costoso y adolece de las deficiencias habituales de enfoques sencillos basados en la correlación, tales como la sensibilidad a la orientación, el tamaño, las condiciones de iluminación, fondo y el ruido [36].

La principal traba para el *directmatching* es que intentan llevar a cabo la clasificación en el reconocimiento de las imágenes, en un espacio de muy alta dimensionalidad [36]. Para contrarrestar esta maldición de la dimensionalidad¹ [37], otros sistemas han propuesto el uso métodos estadísticos de reducción de la dimensionalidad, estos propuestas obtienen y conservan las dimensiones de características más significativas antes de realizar el reconocimiento. Algunas de ellas se mencionan a continuación.

Sirovich y Kirby [38] fueron los primeros en utilizar *Análisis de componentes principales* PCA [39]

¹Bellman llamó al fenómeno de la alta dimensionalidad como "la maldición de la dimensionalidad"

para reducir la dimensionalidad de representar imágenes de la cara.

Ellos demostraron que cualquier rostro se puede representar eficientemente a lo largo del espacio de coordenadas de lo que llamaron las *eigenpictures*, y que cualquier cara puede ser reconstruida utilizando sólo una pequeña colección de *eigenpictures*.

Turk y Pentland [40],[5] se dieron cuenta, con base en Sirovich y resultados de Kirby, que las proyecciones en las *eigenpicture* se podrían utilizar como características de clasificación para reconocer rostros. Emplearon este razonamiento para desarrollar un reconocimiento de rostros, en un sistema que se basa en los *eigenfaces*, estos corresponden a los vectores propios asociados con los valores dominantes en una cara conocida (patrones) de la matriz de covarianza, se reconocen los rostros mediante la comparación de sus proyecciones a lo largo de los *eigenfaces* (caras propias) y con las imágenes de la cara de los individuos conocidos como *eigenfaces*. Los *eigenfaces* definen un espacio de características que reduce drásticamente la dimensionalidad del espacio.

Se pueden observar un conjunto de entrenamiento, la cara y el promedio de siete principales *eigenfaces* derivados de las imágenes de entrenamiento en las Figs. 2.11, 2.12 y 2.13, respectivamente. El método se prueba utilizando una base de datos de 2.500 imágenes de 16 personas, bajo todas las combinaciones, 3 orientaciones de la cabeza, 3 escalas y 3 condiciones de iluminación y varias resoluciones. Las tasas de reconocimiento que fueron reportadas 96%, 85% y 64% , corresponden para distinta iluminación, orientación y variación de la escala. Aunque el método parece ser bastante robusto frente a variaciones de iluminación, el rendimiento se degrada con los cambios de escala.



Figura 2.11: Ejemplo de un conjunto de entrenamiento[5].



Figura 2.12: Cara promedio[5].

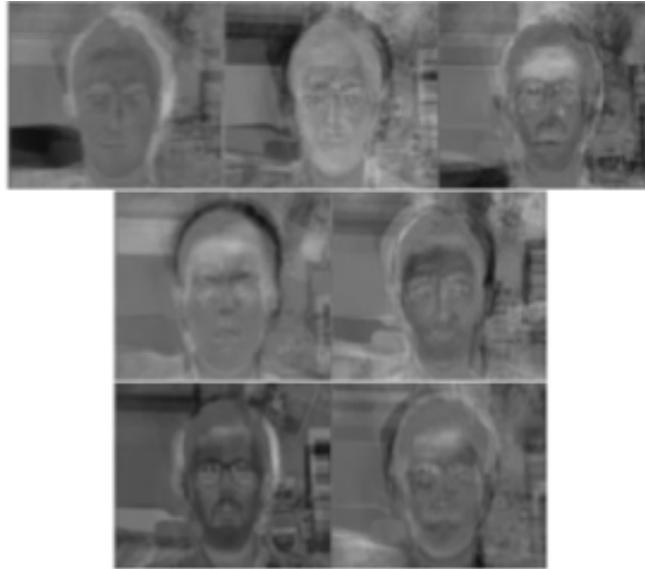


Figura 2.13: 7 eigenfaces de la Fig 2.11[5].

Las capacidades del sistema de Turk y Pentland han sido extendidas de varias maneras en [41] y probado en un base de datos de 7.562 imágenes, de aproximadamente 3.000 personas.

PCA parece funcionar bien cuando existe en la base de datos la imagen de una misma persona, pero cuando la persona tiene varias imágenes de ella misma, el rendimiento del algoritmo disminuye; entonces Belhumeur en [6] sostuvo la idea de cambiar el algoritmo y encontrar un subespacio que discrimine mejor entre diferentes clases (intraclase e interclase). Particularmente se utilizan todas las muestras de todas las clases y se calcula la matriz de dispersión entre interclases y la matriz de dispersión entre intraclases.

PCA mantiene variaciones no deseadas debido a la iluminación y la expresión facial. Como indica Moses [42], "las diferencias entre las imágenes de la misma persona por la iluminación y la dirección de la iluminación son casi siempre muy grandes, debido a estas variaciones provocan en las imágenes un cambio de identidad en la misma persona", Fig. 2.14.

Por lo tanto, para contrarrestar los problemas de PCA se propone la utilización del método llamado análisis discriminante lineal de Fisher [43], el cual maximiza la relación entre la dispersión de la clase y la dispersión dentro de la clase y es por lo tanto es mejor clasificando que el PCA. La realización de varias pruebas en 330 imágenes, 5 personas (66 imágenes para cada uno), informan que el método, llamado *Fisherfaces*, que utiliza proyección del subespacio a priori de la proyección de *Linear Discriminant Analysis* LDA, (para evitar la dispersión dentro de las clases), es mejor manejando simultáneamente las variaciones en la iluminación y las expresiones faciales. Swets y Weng [7] informan resultados similares cuando se emplea el mismo procedimiento no sólo para los rostros sino también para objetos en general (90 % de precisión en una base de datos de 1316 + 298 imágenes de 504 clases). La, (Fig. 2.15 muestra algunos ejemplos de *eigenfaces* y *Fisherfaces* y como *Fisherfaces* capturan información discriminatoria mejor que *Eigenfaces*).



Figura 2.14: La misma persona vista bajo diferentes condiciones de iluminación puede cambiar la apariencia dramáticamente así como su identidad [6].

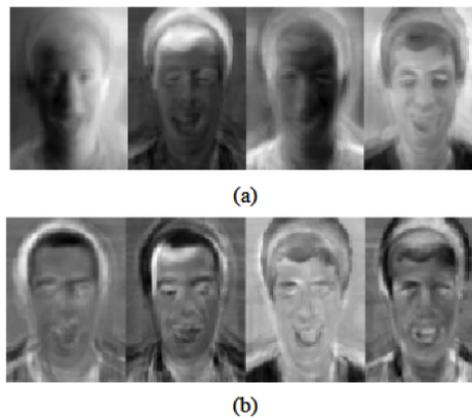


Figura 2.15: A) Eigenfaces y sus variaciones de iluminación, b) Fisherfaces mostrando lo poco que le afecta los cambios de iluminación [7].

Cabe señalar que algunos trabajos, por ejemplo Martínez en [44] muestra que cuando el conjunto de datos de entrenamiento es pequeño en PCA puede superar a LDA y también que PCA es menos sensible a los conjuntos de entrenamiento diferentes.

Los *eigenfaces* estándar y los enfoques *Fisherfaces* suponen la existencia de una proyección óptima, la cuál proyecta las imágenes del rostro para que no se solapen a distintas regiones del subespacio reducido, donde cada una de estas regiones corresponde a un sujeto único. Esto, en realidad, puede no ser necesariamente cierto ya que las imágenes de diferentes personas pueden ser mapeadas a la misma región de la cara espacio. Moghaddam en [45] propone un enfoque alternativo que utiliza imágenes de diferencia, donde una imagen de diferencia se define como la media aritmética de los valores de la intensidad de los correspondientes píxeles entre imágenes. Propone 2 clases de imágenes de diferencia: intrapersonal, dos imágenes de la misma persona y interpersonal, que consiste en imágenes de diferencia, derivada de dos imágenes de diferentes personas.

Entonces, dada la diferencia de entre 2 imágenes $I_1 e I_2$, la probabilidad de que la imagen de diferencia pertenezca a la clase intrapersonal es dada por regla de Bayes: 2.1

$$P(\Omega_I | d(I_1, I_2)) = \frac{P(d(I_1, I_2) | \Omega_I) P(\Omega_I)}{P(d(I_1, I_2) | \Omega_I) P(\Omega_I) + P(d(I_1, I_2) | \Omega_E) P(\Omega_E)} \quad (2.1)$$

Donde:

$d(I_1, I_2)$ = Diferencia entre las 2 imágenes $I_1 e I_2$.

Ω_I = Clase intrapersonal.

Ω_E = Clase interpersonal

Esta clasificación convierte en un M-ario (M es el número de individuos) problema a un simple espacio binario donde la clase interpersonal e intrapersonal se pueden meter en un clasificador MAP donde dada las dos imágenes se declaran pertenecer al mismo individuo sí:

$$P(\Omega_I | d(I_1, I_2)) > P(\Omega_E | d(I_1, I_2)) \text{ o, equivalentemente, sí } P(\Omega_E | d(I_1, I_2)) > \frac{1}{2}$$

Para un enfoque computacionalmente menos costoso y más conveniente, Moghaddam y Pentland [46], también sugieren ignorar la información de clase interpersonal y el cálculo de la similitud basada sólo en la información de clase intrapersonal. En el clasificador de máxima verosimilitud (ML), la puntuación de similitud se da sólo por $P(d(I_1, I_2) | \Omega_I)$.

Algunos avances recientes PCA y LDA basado en algoritmos incluyen análisis multi-lineal PCA dos dimensiones [47], Kernel PCA [48], entre otros. En algoritmos basados en LDA podemos encontrar LDA directo [49], Análisis Función Discriminante Local [50], Fourier-LDA [51], Gabor-LDA [52], Todos estos métodos supuestamente obtienen mejores resultados en el reconocimiento que las técnicas de base, debido a que usan combinaciones de diferentes métodos.

Ventajas y desventajas

La principal ventaja de los enfoques holísticos es que destruyen la información en las imágenes para concentrarse en regiones limitadas o puntos de interés. Esta misma propiedad es su mayor inconveniente, ya que la mayoría de estos enfoques comienzan con la premisa básica de que todos los píxeles de la imagen son igualmente importantes [34].

Por consiguiente, estas técnicas no sólo son computacionalmente caras, sino que requieren un alto grado de correlación entre las imágenes de prueba y entrenamiento, y no se desempeñan tan eficazmente bajo grandes variaciones de cambios de pose, escala y en la iluminación, etc. [53]. Mencionado anteriormente, varios de estos algoritmos han sido modificados y/o han realizado mejoramientos para compensar esas variaciones, junto con las técnicas de reducción de dimensionalidad han sido bastante utilizadas (que a pesar de que tales técnicas aumentan la generalización de capacidades, el único inconveniente es que potencialmente pueden causar la pérdida de información discriminativa [54]); estos enfoques parecen producir mejores resultados que en el reconocimiento basado en características locales.

2.4.3 MÉTODOS HÍBRIDOS

Realizan el reconocimiento combinando las características locales de los rostros y las características globales, su idea es tratar a la imagen del rostro como lo hace la percepción humana, así como se sugiere por Thompson [55]; Bartlett y Searcy (Cognitive Psychology) [56]; Chellappa [57] y Wechsler [58], lo realizan los seres humanos.

2.5 Bases de datos y sus evaluaciones

Hoy en día debido al incremento en técnicas, métodos y teoría en el ámbito del reconocimiento de rostros, se han creado una buena cantidad de bases de datos estándares y con características específicas, todo esto con el propósito de realizar comparaciones de los diferentes métodos en la literatura.

A lo largo del tiempo distintas bases de datos son adaptadas a las necesidades de cada método o con una necesidad específica. A continuación muestro algunas bases más comunes para la realización de pruebas:

2.5.1 BASE DE DATOS FERET

Base de datos FERET es la prueba de-facto [8] dentro de evaluaciones en los sistemas de reconocimiento de rostros. FERET es administrado por la Agencia (DARPA) (Defense Advanced Research Projects Agency) y (NIST) (National Institute of Standards and Technology). La base de datos contiene imágenes de rostro, obtenidas a lo largo de diciembre de 1993 y agosto de 1996. En 2003 publicó una versión de alta resolución, 24 bits de color, de estas imágenes.

Esta base contiene una galería de imágenes que incluyen 1196 imágenes de 1196 personas y cuatro clases de conjuntos de pruebas: **fafb** (1195 imágenes con variación en expresión), **fafc** (194 imágenes con variación en iluminación), **dupI** (722 imágenes tomadas en menos de 18 meses) y **dupII** (234 imágenes tomadas en los 18 meses siguientes), en la Fig. 2.16 se muestra un ejemplo de las imágenes que contiene esta base.

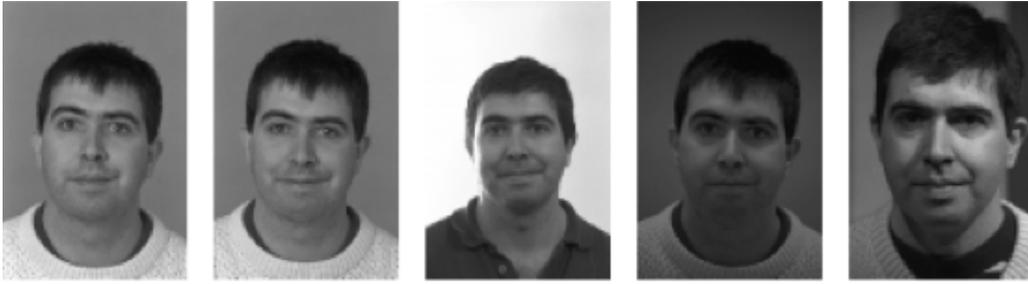


Figura 2.16: Ejemplo de imágenes FERET de [8].

2.5.2 BASE DE DATOS YALE

Esta base fue construida por Belhumeur en [6] y está formada por 165 imágenes en escala de grises, con 15 personas diferentes. Lo interesante en esta base de datos es que cada 11 imágenes de cada persona contienen variaciones en expresiones faciales, cambios en iluminación y usos de lentes para cada individuo. En la Fig. 2.17 se muestra un ejemplo de esta base de datos.



Figura 2.17: Ejemplo de imágenes Yale, se muestra las diferentes variaciones de imágenes para una sola persona [6].

2.5.3 BASE DE DATOS YALE EXTENDIDA B

La base de datos de rostros de Yale Extendida B, construida por Georghiades en [9], contiene 16128 imágenes de 640x480 en escala de grises, consta de 28 individuos, bajo 9 poses y 64 condiciones de iluminación; además se proporciona un conjunto de imágenes recortadas mostrando sólo el rostro de cada individuo. En la Fig. 2.18 se muestra un ejemplo de esta base.



Figura 2.18: Ejemplo de imágenes Yale Extendida B [9].

2.5.4 BASE DE DATOS ORL

La base de datos ORL [10], contiene un conjunto de imágenes de rostros con 256 niveles de gris y una resolución de 92x112 píxeles. Estas imágenes fueron tomadas entre abril de 1992 y abril de 1994. En total, la base de datos contiene 10 imágenes diferentes de 40 sujetos distintos. Para algunos de los sujetos, las imágenes fueron tomadas en diferentes momentos. Todas las imágenes fueron tomadas con los sujetos en una posición frontal, con una tolerancia para alguna inclinación y rotación de hasta 20 grados. También existen variaciones en la escala de hasta 10%. En la Fig. 2.19 se muestra un ejemplo de 3 grupos de imágenes de diferentes sujetos.



Figura 2.19: Ejemplo de 3 sujetos y sus grupos de imágenes de la base ORL [10]

2.6 Aplicaciones comerciales

Como se había mencionado en 1.1; el reconocimiento de rostros hoy en día ya tiene alcances comerciales y existen compañías que se dedican a ofrecer soluciones con la tecnología del reconocimiento de rostros. En nuestra vida cotidiana ya contamos con cámaras digitales o bien en nuestros celulares inteligentes, que tienen la capacidad de detectar rostros en las fotos; incluso empieza a surgir el reconocimiento de rostro para autenticación ya sea en una computadora o en el celular inteligente. En la Figura. 2.20 se muestran algunas páginas que ofrecen servicios usando este tipo de tecnología.

| Producto | Dirección web |
|---|---|
| L-1 Identity Solutions | http://www.l1id.com/ |
| smarti de TAB Systems | http://www.tab-systems.com/ |
| FastAccess de Sensible Vision | http://www.sensiblevision.com/ |
| Vision Access de Bioscrypt | http://www.bioscrypt.com/ |
| Animetrics90 de Animetrics | http://www.animetrics.com/ |
| MyFaceID of Betaface | http://www.betaface.com/ |
| FaceVACS de Cognitec Systems | http://www.cognitec-systems.de/ |
| Morpheus de Kee Square S.r.l. | http://www.keesquare.com/ |
| Tecnologías FRS de IITS | http://www.iits.es/ |
| Facial Recognition Solutions de Cross Match | http://www.crossmatch.com/ |
| Ex-Sight Products | http://www.ex-sight.com/ |
| FaceEnforce de Cybula | http://www.cybula.com/ |
| Affinity de OmniPerception | http://www.omniperception.com/ |
| idfend de The Covenant Consortium | http://www.tcc.us.com/ |
| Productos biométricos de ImageWare | http://www.iwsinc.com/ |
| BioID SDK de BioID | http://www.bioid.com/ |
| IMC Security Products | http://www.imcsecurity.com/ |
| Sistemas EntryGuard de FaceKey | http://www.facekey.com/ |

Figura 2.20: Aplicaciones comerciales en reconocimiento de rostros [11].

3

Sistema propuesto

EN este capítulo se plantea un nuevo método de reconocimiento de rostros basado en la transformada Hermite discreta DHT ¹ y el uso del método de Patrones Binarios Locales LBP ² que fue diseñado para la descripción texturas. La DHT es empleada como un método holístico de extracción de características faciales y la usamos como método basado en el sistema visual humano, más adelante abordaré a detalle este punto; y contamos con un método que actúa localmente que son los LBP, combinando ambos métodos se genera un método híbrido, que como en la literatura se menciona, es una tendencia de desarrollo en la actualidad; recordando que se mezclan las ventajas de los métodos holísticos y los basados en características locales, la combinación de estos métodos fueron sugeridos por Wechsler [58], O’Toole en [59] y Johnston en [60] en la cual los seres humanos realizamos el proceso de reconocimiento. Se verifica también que los vectores característicos extraídos son invariantes a iluminación, pose y expresión facial, además de reducir la dimensionalidad de estos vectores mediante el método Fisher-LDA ³.

3.1 Metodología

El uso de los métodos híbridos como establece Chellappa en la sección III de su trabajo [57], dice que en un sistema de reconocimiento de rostros se debería usar en general los 2 tipos de extracción

¹Discrete Hermite Transform

²Local Binary Pattern

³Lineal Discriminant Analysis

de características, tanto holísticas como locales. En su resumen en la sección III de su trabajo propuso que algunas de las características del comportamiento humano son aplicables al diseño de sistemas de reconocimiento de rostros, considerando un sistema que manipule solamente imágenes de rostros frontales. Cabe destacar que algunas características del rostro son más significativas y útiles que otras. A manera de ejemplo, se puede hacer referencia al hecho de que los ojos y la boca juegan un papel quizá más importante en el reconocimiento de personas, dejando a la nariz con poca importancia en el proceso [57].

Algunos estudios prueban desde un punto de vista visual que un rostro lo podríamos ver como un patrón de distintas intensidades de luz, de este modo el rostro se analizaría por medio de su orientación, frecuencia y amplitud de sus componentes. La capacidad de distinguir los rasgos de las personas a distinta distancia y con distintos grados de iluminación estará en función de la frecuencia y el contraste de los rasgos, y la sensibilidad que el testigo tenga para la combinación de ambas dimensiones en lo que se denomina como la *Función de Sensibilidad al Contraste (FSC)*. Cuanto más grande sea la distancia a la que se presenta la persona a identificar mayor será la frecuencia espacial, cuanto menos iluminación menor será el contraste; por lo tanto, el incremento de la distancia y el decremento de la iluminación disminuyen la posibilidad de percibir algunos rasgos faciales.

Por otro lado, el sistema visual humano está preparado para procesar frecuencias bajas, medias y altas por diferentes canales, como en algunos estudios citados en [61],[62],[63], señalan. Se puede observar en la Fig. 3.1, las frecuencias bajas permitirán percibir la configuración global de una cara, pero los detalles permanecen borrosos, y en las frecuencias altas se distinguen los detalles, que en el caso de los rostros serían los rasgos.

Otros estudios también demuestran lo mencionado anteriormente, por ejemplo en [64] dice que las bandas de baja frecuencia juegan un papel dominante en el reconocimiento de rostros; Sergent [65] dice que las componentes de bajas y altas frecuencia en la imagen de un rostro, se usaban dependiendo del tipo de tarea de reconocimiento que se pidiera en el uso de la imagen (género, identidad, etc.). Lai en [66] menciona que la banda de baja frecuencia de la imagen de un rostro es menos sensible a las variaciones de expresiones faciales y que el espectro del rostro es invariante a cambios de escala, traslación y rotación en el plano [66]. En [16] también se comprobó que los componentes de baja frecuencia realizan la descripción global en una imagen, mientras que los de alta frecuencia describen los detalles.

Motivado por la explicación anterior, se propone el método que se muestra en la Fig. 3.2 dividido en 4 etapas:

1. Extracción de características mediante un método holístico; con la transformada Hermite discreta DHT, utilizando el 3er, 4to y 5to orden de la transformada, y por cada coeficiente de la transformada obtener 4 estadísticos para generar la firma o el vector de características representando a cada persona.

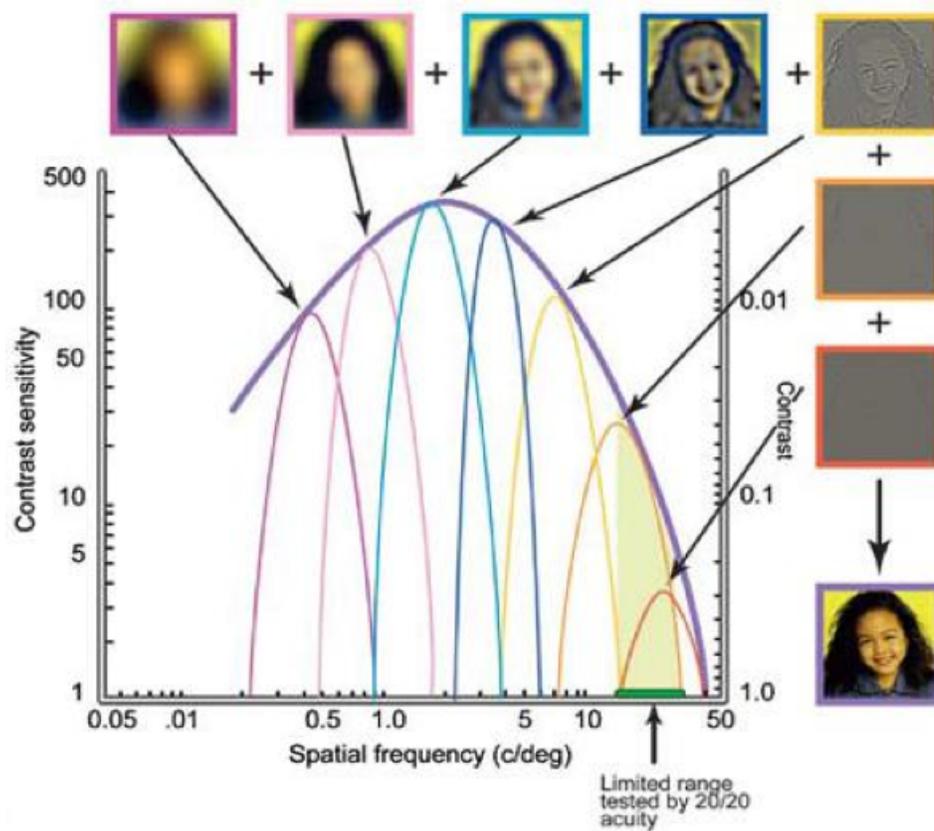


Figura 3.1: Función de sensibilidad al contraste como envoltura de los canales selectivos a frecuencias espaciales. Obtenida de www.contrastsensitivity.net/es/cm.html

2. Extracción de características mediante los LBP (como método local), aquí se evaluarán los rostros en forma global como por regiones; se dividirá cada rostro en 4 y 16 regiones respectivamente. Se obtendrán los histogramas globales y por cada región (4 y 16 parches), obteniendo los histogramas generales o concatenados.
3. Una vez obteniendo los vectores de los 2 pasos anteriores, se usan firmas ya sea de manera individual o se procede a concatenar la firma de ambos métodos y reducir la dimensión con el método LDA.
4. Después del LDA, el siguiente paso es el reconocimiento mediante distancias, podemos aquí visualizar como se comporta y las tasas de reconocimiento al momento de clasificar usando la *matriz de confusión*; en este proceso el tipo de aprendizaje es supervisado.
5. Se evaluarán todos los experimentos. Nuestra hipótesis es que 2 métodos como son DHT y LBP, combinados con un reductor de dimensionalidad que en este caso es LDA, se obtendrán las mejores tasas de clasificación en comparación con los métodos evaluados individualmente, teniendo tasas de reconocimiento muy cercanas a los algoritmos más populares en la literatura.

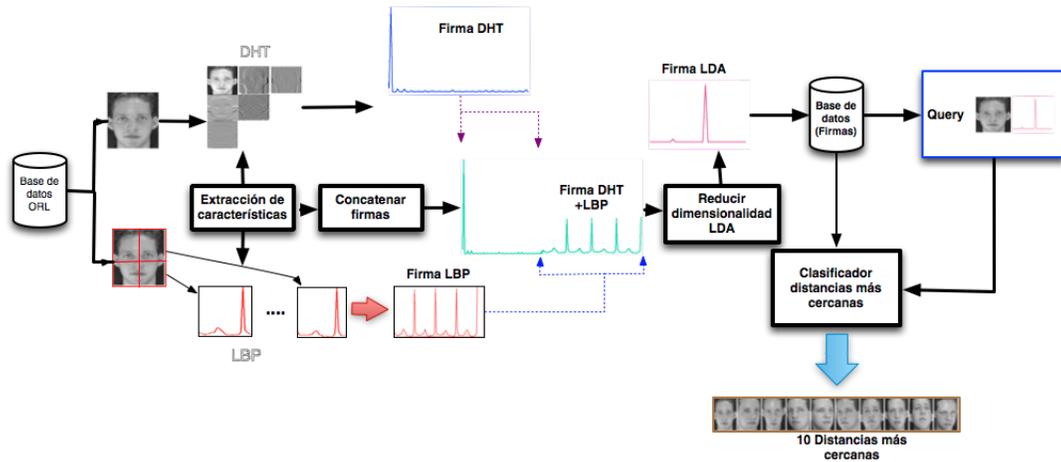


Figura 3.2: Diagrama simplificado del sistema propuesto.

3.2 Transformada Hermite

La transformada Hermite, introducida al área de procesamiento digital de imágenes por Martens[67], es un modelo de representación basado en el sistema de visión humano. La transformada Hermite modela los campos receptivos presentes en el ojo humano para las primeras etapas de visión (similares a las derivadas de Gaussiana excepto por un factor de escala); en su etapa de análisis, separa la información visual contenida en la imagen en información de baja frecuencia (promedio) y alta frecuencia que hace a la transformada Hermite un buen descriptor de patrones característicos como bordes y texturas. La transformada Hermite es un caso particular de una transformada más general denominada transformada polinomial.

3.2.1 TRANSFORMADA POLINOMIAL

La tranformada polinomial es una técnica de descomposición local de señales por medios de polinomios.[67] La transformada polinomial consta de dos pasos. El primero, la señal original $L(x)$ es localizada al multiplicarla por una función ventana $V(x)$ en varias posiciones equidistantes. A partir de la función ventana $V(x)$ es posible definir una función de ponderación periódica T :

$$W(x) = \sum_k V(x - kT) \quad (3.1)$$

El periodo de la función de ponderación es T . Haciendo $W(x)$ diferente de cero para cualquier posición x , se obtiene:

$$L(x) = \frac{1}{W(x)} \sum_k L(x)V(x - kT) \quad (3.2)$$

El segundo paso consiste en aproximar la información local dentro de cada ventana de análisis en términos de una familia de polinomios ortogonales. Para llevar a cabo la expansión, se toman

como funciones base los polinomios $Q_n(x)$, donde n es el grado del polinomio. Estos polinomios están determinados totalmente por la función ventana de tal forma que deben ser ortonormales respecto a $V^2(x)$ es decir:

$$\int_{-\infty}^{+\infty} V^2(x)Q_m(x)Q_n dx = \delta_{mn} \tag{3.3}$$

En una transformada polinomial, los coeficientes de una señal $L(x)$ (proyecciones) pueden ser obtenidos como:

$$L_n(kT) = \int_{-\infty}^{+\infty} L(x)Q_n(x - kT)V^2(x - kT)dx \tag{3.4}$$

seguido de un submuestreo en múltiplos de T. Es posible notar que la señal original $L(x)$ se convoluciona con las funciones de análisis (funciones filtro) denotadas por:

$$D_n(x) = Q_n(-x)V^2(-x) \tag{3.5}$$

es decir:

$$L_n(kT) = \int_{-\infty}^{+\infty} L(x)D_n(x - kT)dx \tag{3.6}$$

El mapeo de la señal original $L(x)$ a los coeficientes $L_n(kT)$ es conocido como transformada polinomial directa. El proceso de reconstrucción de la señal a partir de los coeficientes es conocido como transformada polinomial inversa 3.7 y básicamente consiste en interpolar los coeficientes $L_n(kT); k - entero$ con las funciones de síntesis (funciones patrón) $P_n(x)$ 3.8 y sumar sobre todos los órdenes n .

$$L(x) = \sum_{n=0}^{\infty} \sum_k L_n(kT)P_n(x - kT) \tag{3.7}$$

$$P_n(x) = \frac{Q_n(x)V(x)}{W(x)} \tag{3.8}$$

Las transformadas polinomiales directa e inversa se muestran en la Fig. 3.3

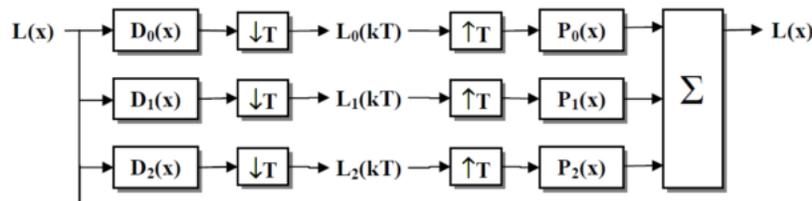


Figura 3.3: Diagrama de bloques de la transformada polinomial directa e inversa.

3.2.2 TRANSFORMADA HERMITE EN UNA DIMENSIÓN

Cuando en la transformada polinomial las ventanas de análisis empleadas son funciones gaussianas se habla de la transformada Hermite. Dicha ventana Gaussiana tiene la propiedad de ser isotrópica

Capítulo 3. Sistema propuesto

(invariable con la rotación), separable en coordenadas cartesianas y sus derivadas simulan algunos procesos a nivel de corteza visual del sistema de visión humano; se define de la siguiente manera:

$$V(x) = \frac{1}{\sqrt{\sqrt{\pi}\sigma}} e^{-\frac{x^2}{2\sigma^2}} \quad (3.9)$$

donde el factor de normalización $\frac{1}{\sqrt{\sqrt{\pi}\sigma}}$ es tal que $V^2(x)$ tiene energía unitaria. Los polinomios ortogonales que están asociados a $V^2(x)$ son conocidos como polinomios de Hermite y son las soluciones a la ecuación diferencial [68]:

$$y'' - 2xy' + 2ny = 0 \quad (3.10)$$

donde x es una variable espacial y y es una función de x . La ecuación 3.10 se conoce como la ecuación de onda del oscilador armónico (resultado de la ecuación de Schrödinger). La forma matemática de los polinomios de Hermite dada por la fórmula de Rodrigues [69] es:

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n e^{-x^2}}{dx^n}, n = 0, 1, 2, \dots, \quad (3.11)$$

Donde el operador $\frac{d^n}{dx^n}$ denota la n -ésima derivada de la función. La definición de la transformada Hermite involucra tres funciones cuyas propiedades se describen a continuación.

FUNCIÓN DE PONDERACIÓN. Como la función de ponderación $W(x)$ es periódica con periodo T , puede ser expandida a través de una serie de Fourier:

$$W(x) = \frac{\sqrt{2\sqrt{\pi}\sigma}}{T} w(x) \quad (3.12)$$

donde:

$$w(x) = 1 + 2 \sum_{k=1}^{\infty} e[-\frac{1}{2}(k\frac{2\pi\sigma}{T})^2] \cos k\frac{2\pi x}{T} \quad (3.13)$$

El contraste de esta función de ponderación se determina por el parámetro de muestreo $\tau = T/\sigma$. Como usualmente se quiere limitar el número de descomposiciones locales, es mejor hacer que τ sea tan grande como sea posible. Por otro lado, considerando la ecuación 3.2, se observa que $W(X)$ debe ser aproximadamente constante ya que la división entre $W(x)$ podría introducir una sensibilidad a la variación de la escala.

FUNCIÓN DE ANÁLISIS. Las funciones de análisis (funciones filtro) determinan que información se hace explícita en los coeficientes de la transformada Hermite, por lo que las principales propiedades de esta transformada están determinadas por estas funciones. A partir de la ecuación

3.5 se pueden derivar las funciones filtro específicas:

$$D_n(x) = \frac{(-1)^n}{\sqrt{2^n n!}} \cdot \frac{1}{\sigma \sqrt{\pi}} H_n\left(\frac{x}{\sigma}\right) e^{-\frac{x^2}{\sigma^2}} \quad (3.14)$$

que podemos reescribirla en términos de la definición de derivadas de Gaussiana evaluada en $x = (x/\sigma)$ como:

$$D_n(x) = \frac{1}{\sqrt{2^n n!}} \cdot \frac{1}{\sigma \sqrt{\pi}} \frac{d^n}{dx^n} e^{-\frac{x^2}{\sigma^2}} \quad (3.15)$$

Es aquí donde se produce la conexión entre los perfiles de los campos receptivos del Sistema de Visión Humano (HVS Human Vision System) modelados por las derivadas de Gaussiana, y los polinomios de Hermite ortonormales respecto a la ventana de análisis.

$$G_n(x) = (-1)^n H_n(x) e^{-x^2} \quad (3.16)$$

La transformada de Fourier de $D_n(x)$ es:

$$d_n(\omega) = \frac{1}{\sqrt{2^n n!}} (j\omega\sigma)^n e^{-\frac{(\omega\sigma)^2}{4}} \quad (3.17)$$

teniendo un valor extremo cuando $(\omega\sigma)^2 = 2n$, indica que los filtros de órdenes mayores analizan frecuencias sucesivamente más altas de la señal. Sin embargo, para filtros de órdenes muy grandes, los picos en frecuencia se aproximan demasiado, por lo que tales filtros proporcionan poca información adicional. Debido a esto, en la práctica, la transformada de Hermite se limita a pocos términos. Las funciones filtro para $n = 0, \dots, 4$ se muestran en la Figura 3.4.

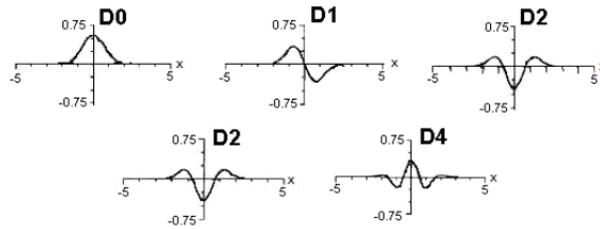


Figura 3.4: Diagrama de bloques de la transformada polinomial directa e inversa.

FUNCIONES DE SÍNTESIS. Las funciones de síntesis (funciones patrón) se usan durante la reconstrucción de la señal original que se hace a partir de los coeficientes de la transformada Hermite. Estas funciones están dadas por la ecuación:

$$P_n(x) = H_n\left(\frac{x}{\sigma}\right) \frac{V(x)}{W(x)} = \frac{T}{\sqrt{2^n n!}} \frac{1}{\sigma \sqrt{2\pi}} H_n\left(\frac{x}{\sigma}\right) \frac{e^{-\frac{x^2}{2\sigma^2}}}{w(x)} \quad (3.18)$$

donde $w(x)$ es la función de ponderación de la ecuación 3.13. Si $w(x) = 1$ (para valores de muestreo del parámetro $\tau < 2$), la función patrón es igual a la función Hermite de grado n . La función Hermite tiene la propiedad de ser isomorfa a su transformada de Fourier:

$$P_n(\omega) = \frac{T}{\sqrt{2^n n!}} (-j)^n H_n(\omega\sigma) e^{-\frac{(\omega\sigma)^2}{2}} \quad (3.19)$$

3.2.3 TRANSFORMADA HERMITE DISCRETA

La transformada Hermite cuenta con una aproximación discreta (DHT: Discrete Hermite Transform) basada en los polinomios de Krawtchouk [19]. Dado que los filtros de análisis de la transformada Hermite son similares a las derivadas Gaussianas excepto por un factor de escala, es posible utilizar la forma discreta de las derivadas Gaussianas dadas por los coeficientes binomiales:

$$C_N^x = \frac{N!}{x!(N-x)!} \quad (3.20)$$

donde N es la longitud de la ventana binomial. Los polinomios ortonormales de Krawtchouk son obtenidos a partir del producto de los polinomios de Krawtchouk por una ventana binomial definida por:

$$v^2(x) = C_N^x / 2^N \quad (3.21)$$

de tal manera que los polinomios ortonormales de Krawtchouk están definidos como:

$$K_n(x) = \frac{1}{\sqrt{C_N^n}} \sum_{\tau=0}^n (-1)^{n-\tau} C_{N-x}^{n-\tau} C_x^\tau \quad (3.22)$$

para $x = 0, \dots, N$ y $n = 0, \dots, D_{max}$ donde D_{max} es el máximo orden de la derivada y $D_{max} \leq N$. La transformada discreta de Hermite de longitud N aproxima a la transformada Hermite continua por la siguiente relación $\sigma = \sqrt{M/2}$. Ejemplos de los polinomios de Krawtchouk se presentan a continuación para $D = 1, D = 2$ y $D = 4$ con $D = N$:

$$K_1 = \frac{1}{\sqrt{2^1}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$K_2 = \frac{1}{\sqrt{2^2}} \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & -2 \\ 1 & -1 & 1 \end{bmatrix}$$

$$K_4 = \frac{1}{\sqrt{2^4}} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 4 & 2 & 0 & -2 & -4 \\ 6 & 0 & -2 & 0 & 6 \\ 4 & -2 & 0 & 2 & -4 \\ 1 & -1 & 1 & -1 & 1 \end{bmatrix}$$

3.2.4 TRANSFORMADA HERMITE DISCRETA 2D

Un caso especial se presenta en las transformadas polinomiales bidimensional resulta que cuando la función ventana es separable entonces:

$$V(x, y) = V(x)V(y) \quad (3.23)$$

y la rejilla de muestreo es cuadrada. Las funciones de análisis y síntesis son por ende separables y pueden ser implementadas de manera muy eficiente. En el caso de la transformada Hermite, se trabaja con ventanas Gaussianas, las cuales al ser isotrópicas podemos escribir $\sigma = \sigma_x = \sigma_y$

$$V(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2 + y^2)}{2\sigma^2}} \quad (3.24)$$

Los polinomios de Hermite ortonormales con respecto a $V^2(x, y)$ pueden escribirse como:

$$Q_{n-m,m}(x, y) = \frac{1}{\sqrt{2^n(n-m)!m!}} H_{n-m}\left(\frac{x}{\sigma}\right) H_m\left(\frac{y}{\sigma}\right) \quad (3.25)$$

Los coeficientes de la señal localizada $f_{n-m,m}(x, y)$ se obtienen al filtrar la imagen con las funciones filtro separables dada por:

$$D_{n-m,m}(x, y) = Q_{n-m,m}(-x, -y)V^2(-x, -y) \quad (3.26)$$

y haciendo enseguida un submuestro a la señal de salida en las direcciones horizontal y vertical en múltiplos de T; es decir, que para todo $p, q \in S$,

$$L_{n-m,m}(p, q) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} L(x, y) D_{n-m,m}(x - p, y - q) dx dy \quad (3.27)$$

La imagen sintetizada se obtiene por un proceso de interpolar los coeficientes polinomiales con funciones de síntesis

$$P_{n-m,m}(x, y) = \frac{Q_{n-m,m}(x, y)V(x, y)}{W(x, y)}$$

y sumar para todo orden n y m de la siguiente manera:

$$L(x, y) = \sum_{n=0}^{\infty} \sum_{m=0}^n \sum_{(p,q) \in S} L_{n-m,m}(p, q) P_{n-m,m}(x - p, y - q) \quad (3.28)$$

En la siguiente imagen se muestra la descomposición de Hermite sobre un rostro de la base de datos Fig. 3.5.

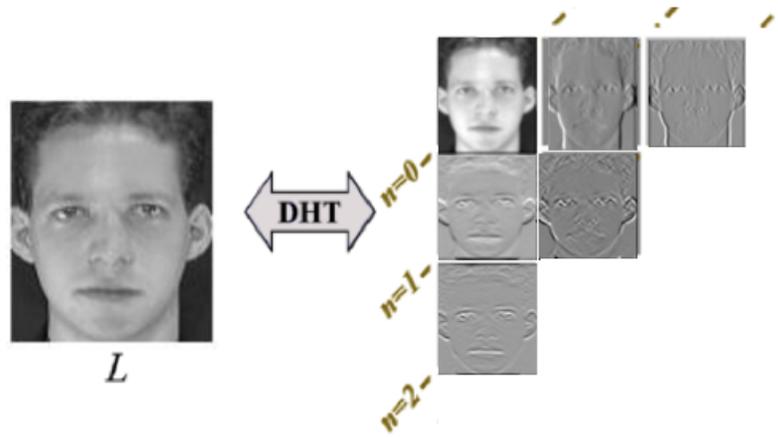


Figura 3.5: Descomposición de una imagen con la transformada Hermite 2do Orden.

3.2.5 VECTORES CARACTERÍSTICOS

Los componentes de baja de frecuencia como se mencionó en 3.1 nos describe de manera global un rostro, con esta idea se crea la etapa para extraer los vectores característicos de cada rostro usando la DHT.

Lo que se propone es obtener 4 estadísticos; ya que los coeficientes de Hermite pueden ser considerados como funciones de densidad de probabilidad, la media, la varianza, el skewness y la kurtosis son suficientes como para proporcionar una buena aproximación a ellos [70]. Hemos investigado el siguiente conjunto de datos estadísticos, en donde M y N son el tamaño de los coeficientes:

- Media

$$\mu_{(s,\Theta)} = \frac{1}{NM} \sum_{x=1}^N \sum_{y=1}^M C_{(s,\Theta)}(x, y) \quad (3.29)$$

- Varianza

$$\sigma_{(s,\Theta)}^2 = \frac{1}{NM} \sum_{x=1}^N \sum_{y=1}^M (C_{(s,\Theta)}(x, y) - \mu_{(s,\Theta)})^2 \quad (3.30)$$

- skewness(Υ): Es una medida de asimetría; es positiva si la distribución tiende a la derecha, negativo cuando la distribución tiende a la izquierda, o incluso nulo, que normalmente implica una distribución simétrica.

$$\Upsilon_{(s,\Theta)} = \frac{\mu_{(s,\Theta)}^3}{\sigma_{(s,\Theta)}^3} \quad (3.31)$$

- kurtosis(κ)

$$\kappa(s, \Theta) = \frac{\mu_{(s,\Theta)}^4}{\sigma_{(s,\Theta)}^4} \quad (3.32)$$

Después de definir los estadísticos, la idea es: por cada coeficiente de la DHT se obtendrán los 4 estadísticos mencionados; en la Fig. 3.5, el análisis realizado en el ejemplo, fue hasta el 2do orden y se pueden ver 6 coeficientes de la transformada; en nuestro caso se trabajará con hasta el 5to orden de la DHT, evaluando desde el 3er orden⁴. En total obtendremos vectores para cada rostro de 84, 60 y 40 elementos, según el orden de la transformada (4 estadísticos por coeficiente). Para demostrar que la DHT puede extraer características faciales relativamente estables y adecuadas que más adelante se observará que no es suficiente sólo el uso de la DHT, pero en combinación con el método local LBP, los pequeños rasgos locales se podrán utilizar para aumentar el porcentaje de clasificación combinando los 2 métodos. A continuación se muestran los vectores característicos a nivel de la DHT y como se están comportando a nivel intraclase Fig. 3.6 e intertraclase Fig. 3.7.

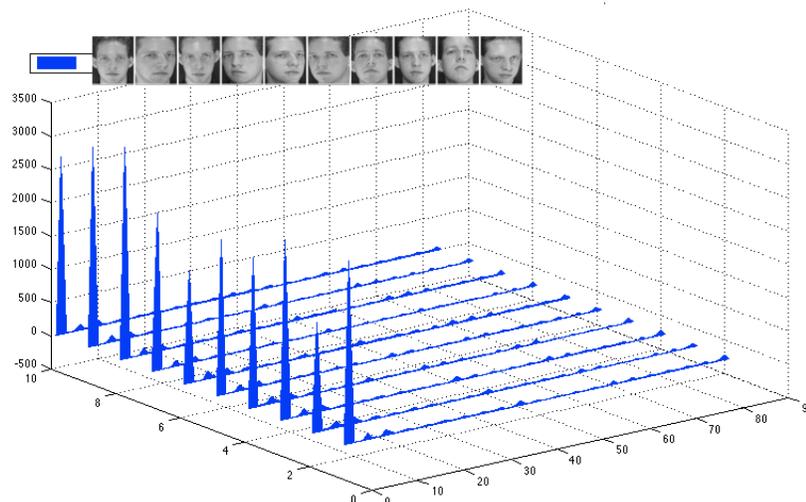


Figura 3.6: Variación vectores característicos intraclase usando la DHT.

⁴Se usan el 3er, 4to y 5to orden de la DHT porque los resultados fueron los mejores alcanzados con base a pruebas.

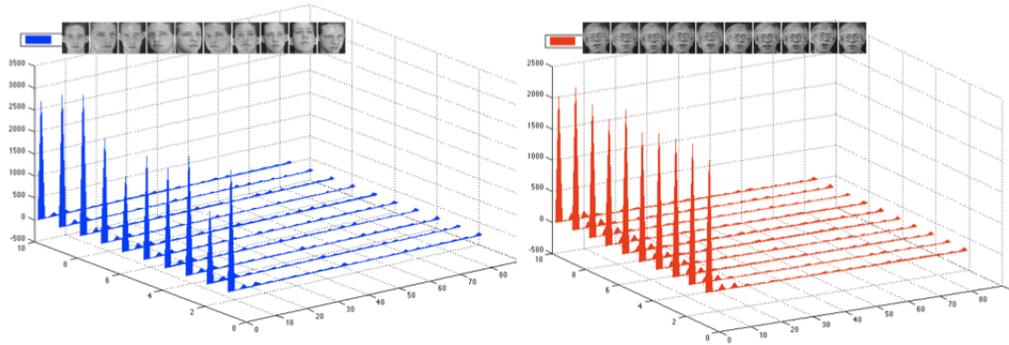


Figura 3.7: Variación vectores característicos interclase usando la DHT.

3.3 Patrones Locales Binarios (LBP)

El método patrones locales binarios es un descriptor que fue propuesto por Ojala [71], para la descripción de texturas. Se basa en la idea de que las propiedades de textura dentro de una región homogénea pueden representarse como micro-patrones, Ahonen [72] puntualizó que un rostro se podría considerar como una composición de micropatrones, y que por lo tanto puede ser descrito efectivamente usando este método [72]. En el trabajo de Ahonen se demuestra que usando LBP para el reconocimiento de rostros, el método da muy buenos resultados, representados en altas tasas de reconocimiento y eficiencia computacional.

3.3.1 FUNCIONAMIENTO

El método original utiliza una máscara fija rectangular de 3×3 , esta asigna etiquetas a cada uno de los píxeles de la imagen tomando en cuenta la distribución de sus píxeles vecinos. Es decir, cada píxel es comparado con los píxeles alrededor para establecer que etiqueta debe ser asignada a dicho píxel, lo que representa un barrido cuadrado alrededor de un píxel central. Los valores dentro de la máscara rectangular se comparan con su píxel central; aquellos menores que el valor central se marcan con "0", en caso contrario, con "1". Los píxeles marcados se multiplican por una función de ponderación de acuerdo con sus posiciones para formar una cadena de patrón. Después, la suma de los ocho píxeles sustituye el valor del píxel central, ver Fig. 3.8.

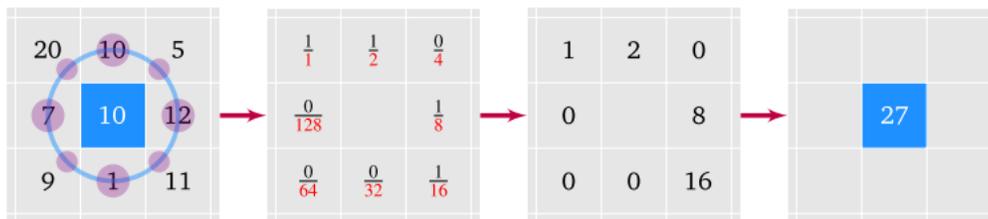


Figura 3.8: Máscara rectangular de 3×3 , el algoritmo LBP calcula comparaciones entre un píxel central y sus vecinos. En este ejemplo, el valor central es $PC = 10$ y la etiqueta final es 27 [73]

Sin embargo diferentes tamaños de máscara son utilizados para lograr la descripción invariante a escala de texturas de diferentes tamaños [71]. La forma de realizar el proceso pasa primero por definir la vecindad de píxeles como una serie de puntos igualmente espaciados en un patrón circular centrado en el píxel evaluado. En caso de que las coordenadas de alguno de los píxeles vecinos no correspondan exactamente a un píxel sobre la imagen, se realiza una interpolación bilineal para hallar la coordenada correspondiente.

La notación general para definir el tipo de patrón utilizado es $(P;R)$, donde P es el número de puntos sobre un patrón circular de radio R . La Fig. 3.9 muestra algunos ejemplos de este tipo de distribuciones circulares de píxeles.

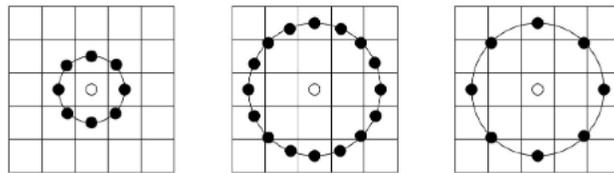


Figura 3.9: Los patrones circulares $(8;1)$, $(16;2)$ y $(8;2)$ [72].

Ojala en su artículo [71] definió los llamados patrones uniformes. Un patrón binario local es considerado uniforme si contiene como máximo 2 transiciones de 0 a 1 o viceversa cuando es leído de forma circular. Por ejemplo, los patrones 11111111 ó 00000000 (Ninguna transición "0"), y 11100111 (2 transiciones) son uniformes, mientras que los patrones 01001100 (4 transiciones) no lo son.

Ojala a través de su investigación, verificaron que aproximadamente el 90% de los patrones obtenidos con un vecindario $(8;1)$ son uniformes, y que cerca del 70% de los patrones son también uniformes al usar un vecindario de $(16;2)$ [71]. Por otra parte, Ahonen [72] obtuvo resultados con cifras del 90.6% y 85.2% con vecindades de $(8;1)$ y $(8;2)$ respectivamente, sobre las fotografías de rostros preprocesadas de la base de datos FERET [72]. Estos altos porcentajes permiten disminuir el tamaño del histograma LBP, ya que todos los patrones no uniformes son asignados a una misma etiqueta común.

3.3.2 MODIFICACIÓN DEL MÉTODO ORIGINAL DE PATRONES LOCALES BINARIOS

En esta subsección, se analiza una modificación del método de LBP. Se resume la principal características del algoritmo seleccionado en esta tesis. En particular, el estudio se centra sobre el sistema LBP uniforme que proporciona invariancia rotacional, señalando que en este trabajo no representa importancia esta característica pero es una ventaja que se podría aprovechar en un futuro.

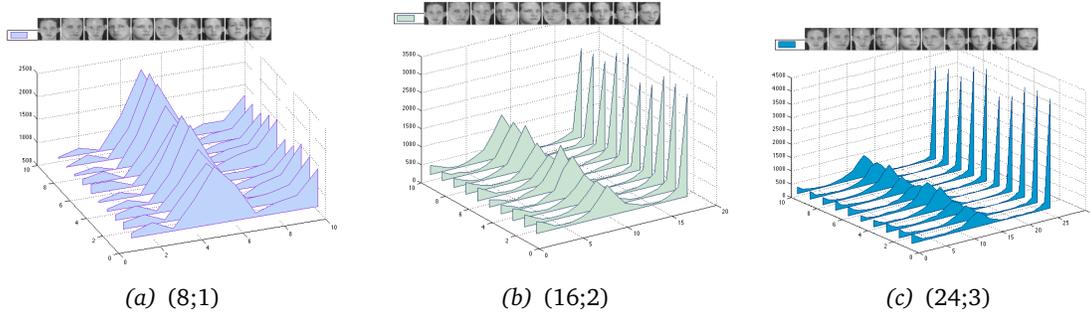


Figura 3.10: Histogramas con LBP Uniforme.

LBP-UNIFORME Mencionado anteriormente en la investigación de Ojala [71], se observó que en el LBP original, más del 90% de los patrones de textura se puede describir con pocas transiciones espaciales, que son los cambios (0/1) [71]. Por lo tanto, introdujeron una medida de uniformidad, $U(LBP_{PR}(g_c))$, que describe las características fundamentales como puntos brillantes, áreas planas y bordes. Por ejemplo, cuando $P = 8$, hay 9 patrones uniformes de los 36 invariantes rotacionales únicos.

$$U(LBP_{PR}(g_c)) = |g_{P-1} - g_c| + \sum_{P=0}^{P-1} |s(g_P - g_c) - s(g_{P-1-g_c})| \quad (3.33)$$

Que se corresponde con el número de transiciones espaciales. Así que, el **LBP-uniforme** ($LBP_{P,R}^{uni}$), se puede obtener como:

$$(LBP_{P,R}^{uni}) = \begin{cases} \sum_{P=0}^{P-1} & \text{si } U(LBP_{P,R}(g_c)) \leq 2 \\ P + 1 & \text{otro caso} \end{cases} \quad (3.34)$$

3.3.3 DESCRIPCIÓN DEL ROSTRO USANDO LBP

El operador LBP uniforme descrito en la sección 3.3.2 es utilizado para describir regiones locales, que después son combinadas para dar una descripción global del rostro. Según Chellapa en [57] se opta por este tipo de método debido a que algunas regiones del rostro aportan más información que otras, además de que los descriptores de texturas promedian la descripción que describen, lo cual no es conveniente al describir rostros puesto que mantener la información de las relaciones espaciales es importante.

Para formar la descripción global, la imagen del rostro es dividida en diferentes números de regiones, a las que se les aplica el operador LBP consiguiendo histogramas diferentes por cada región Fig. 3.12 . Estos histogramas después son concatenados para construir una descripción global del rostro. La Fig. 3.11 muestra algunas de las posibles configuraciones rectangulares de división de la imagen de rostros.



Figura 3.11: Imagen del rostro dividida en 7×7 , 5×5 y 3×3 [72].

Con la combinación de los histogramas correspondientes a las descripciones locales, puede construirse lo que Ahonen llama *histograma mejorado espacialmente*. Este nuevo histograma codifica tanto la apariencia del rostro, como las relaciones espaciales entre las regiones individuales. Luego de que las m regiones R_0, R_1, \dots, R_{m-1} del rostro son determinadas, se procede a aplicar el operador ($LBP_{P,R}^{uni}$) a cada región m . Las m regiones resultantes son concatenadas para generar al llamado *histograma mejorado espacialmente*, el tamaño de histograma es de $m \times n$, donde n es longitud de cada histograma.

A través del *histograma mejorado espacialmente*, se describe a los rostros con tres niveles de localización: los patrones binarios locales representan información a nivel de los píxeles, cada uno de los histogramas representa la información de cada una de las regiones del rostro y el *histograma mejorado espacialmente* comprende la descripción global del rostro.

En la Fig. 3.11 muestra como ejemplo regiones rectangulares iguales, Ahonen menciona que no importa que sea el mismo caso siempre. Argumenta que el utilizar histogramas como descripción, permite que las regiones R_0, R_1, \dots, R_{m-1} sean de cualquier tamaño y forma, y que además estas regiones pueden estar sobrepuestas unas sobre otras, sin cubrir necesariamente toda o no toda la imagen.

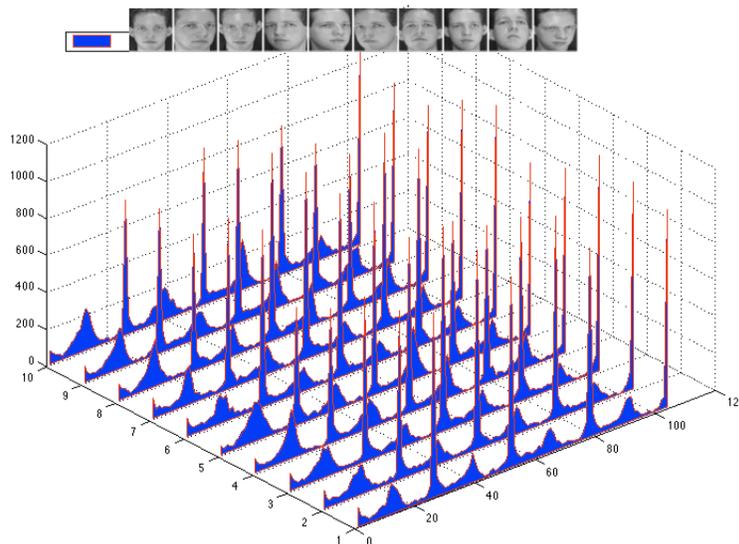


Figura 3.12: Histograma mejorado espacialmente 4 regiones LBP (24;3)

3.4 Análisis discriminante lineal (LDA)

La técnica LDA fue desarrollada por R.A. Fisher, C.Rao y J.G.Bryan, [74], como método de clasificación, a través del cual una nueva observación x podía ser asignada a una de entre dos o más clases C_i .

R.A. Fisher en 1936 el artículo titulado: “The use of multiple measurements in taxonomic problems” [43], en este se genera una familia de métodos de clasificación conocidos como *Discriminantes Lineales de Fisher*. LDA tiene como objetivo convertir un problema de alta dimensionalidad en uno de baja. Para ello LDA proyecta los datos(imágenes) en un espacio vectorial de baja dimensionalidad de manera que la razón entre la distancia entre clases y la distancia dentro de la clase se maximiza. De este modo se garantiza una máxima discriminación entre las clases Fig. 3.13.

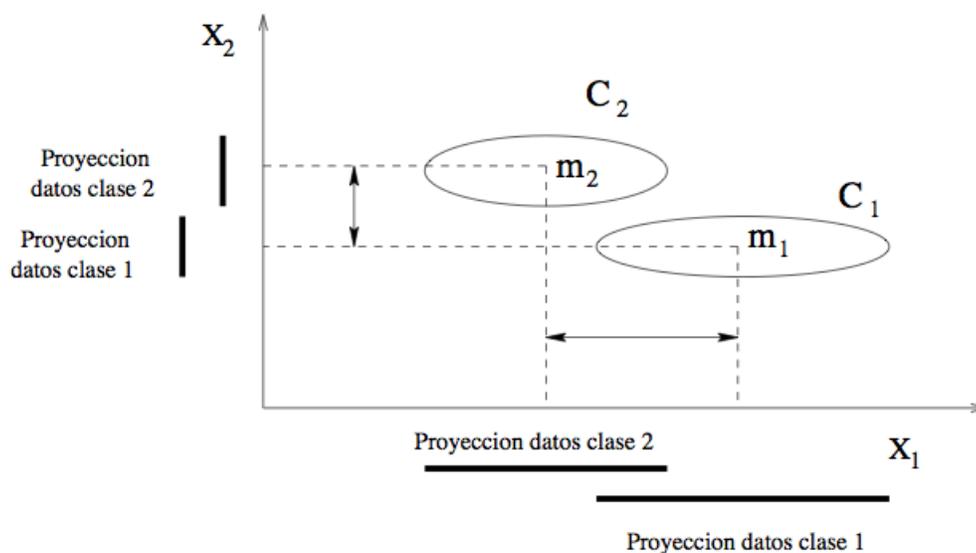


Figura 3.13: Clases bien separadas en un espacio 2D.

A diferencia de PCA, que es un método de reducción dimensional sin supervisión, LDA es supervisado y utiliza información de los datos.

Otra motivación para reducir la función de la dimensión del vector es que los resultados psicofísicos indican que las tareas perceptivas como el juicio similitud tienden a ser realizadas en una representación de pocas dimensiones, como lo dice Liu en [75], que si bien no existe evidencia al respecto con la que el cerebro humano es capaz de reconocer rostros, se sugiere que en él se representan en bajas dimensiones; si se analiza con rapidez justo como procesa la información el cerebro, pensaríamos que durante la asimilación de datos se aplica un reducción de la dimensión; por ejemplo si miramos una foto de alguien conocido, no nos fijamos en puntos o características de la imagen, basta con mirar la foto como un todo logramos asociar quien es en la imagen, igual nos indica que la dimensionalidad intrínseca es mucho menor que el número de píxeles que contiene esa imagen.

3.4.1 DISCRIMINANTE LINEAL DE FISHER

Existe una familia de métodos de clasificación entre ellas un grupo llamado Fisher-LDA [76].

Para darnos una idea del método se explicará la versión más simple del problema: Encontrar el vector W que proyecte los datos a un espacio uni-dimensional de manera de obtener la mayor separabilidad posible entre sus clases. Consideramos la siguiente ecuación:

$$J(W) = \frac{W^T S_B W}{W^T S_W W} \quad (3.35)$$

En donde S_B es la matriz de dispersión entre clases y S_W es la matriz de dispersión intrínseca a cada clase, la definición de ambas matrices es:

$$S_B = \sum_c N_c (\mu_c - \bar{x})(\mu_c - \bar{x})^T \quad (3.36)$$

$$S_W = \sum_c \sum_{i \in c} (x_i - \mu_c)(x_i - \mu_c)^T \quad (3.37)$$

$$\mu_c = \frac{1}{N_c} \sum_{i \in c} x_i \quad (3.38)$$

$$\bar{x} = \frac{1}{N} \sum_i x_i = \frac{1}{N} \sum_c N_c \mu_c \quad (3.39)$$

Siendo N_c número de muestras para la clase c .

Debido a que J es invariante al escalado los vectores $W \rightarrow \alpha W$, es posible elegir una W tal que el denominador sea $W^T S_W W = 1$. Por lo que el problema de optimización queda de la siguiente manera (considerando restricciones):

$$\min_W \quad -\frac{1}{2} W^T S_B W \quad \text{s.t.} \quad W^T S_W W = 1 \quad (3.40)$$

el cual corresponde con la lagrangiana:

$$L_p = -\frac{1}{2} W^T S_B W + \frac{1}{2} \lambda (W^T S_W W - 1) \quad (3.41)$$

Con la siguiente solución:

$$S_B W = \lambda S_W W \Rightarrow S_W^{-1} S_B W = \lambda W \quad (3.42)$$

Donde se puede observar que se trata de un problema de resolución de autosistemas generalizados, utilizando el hecho de que la matriz S_B es simétrica positiva, puede ser escrita como $S_B^{\frac{1}{2}} S_B^{\frac{1}{2}}$, en donde $S_B^{\frac{1}{2}}$ se construye a partir de su descomposición por autovalores como:

$S_B = U\Lambda U^T \rightarrow S_B^{\frac{1}{2}} = U\Lambda^{\frac{1}{2}}U^T$. De donde definiendo $V = S_B^{\frac{1}{2}}W$ se obtiene:

$$S_B^{\frac{1}{2}}S_W^{-1}S_B^{\frac{1}{2}}V = \lambda V \quad (3.43)$$

que es un problema de un autosistema con una matriz simétrica positiva $S_B^{\frac{1}{2}}S_W^{-1}S_B^{\frac{1}{2}}$, con soluciones λ_k como autovalor y V_k como autovector, dando lugar a la solución:

$$W = S_B^{\frac{1}{2}}V \quad (3.44)$$

Finalizando con introducir esta solución en la función objetivo $J(W)$ se encuentra que la solución buscada que maximiza dicha función objetivo es la que tiene los autovalores mayores. Vale la pena mencionar que cuando se aplica el discriminante lineal de Fisher a cualquier número de clases, siempre hay que tener en cuenta que la dimensionalidad del espacio muestral en este caso d nunca debe ser menor que el número de clases C ($d \geq C$). En el discriminante lineal de varias clases habrá que tener en cuenta $C-1$ funciones discriminantes similares a las deducidas para 2 clases. Por todo esto, la proyección se realizará en un espacio $C-1$ dimensional.

3.5 Reconocimiento de rostros (Clasificador)

Teniendo las descripciones de los rostros a través de sus *histogramas mejorados espacialmente* o bien los histogramas concatenados, se necesita una medida para poder establecer la similitud entre los histogramas, de entre el rostro que se desea reconocer y los que se encuentran en la base de datos.

3.5.1 CLASIFICACIÓN POR MÍNIMA DISTANCIA

Uno de los métodos en la teoría de la clasificación de patrones, encontramos el método de la clasificación por mínima distancia, donde cada clase se representa por un vector patrón, luego se calcula la distancia entre el patrón desconocido y cada uno de los vectores patrón, eligiendo la menor distancia para asignar el patrón a una clase.

Supongamos que se tienen M clases y que cada una está representada por un vector prototipo:

$$m_j = \frac{1}{N_j} \sum_{x \in w_j} x \quad j = 1 \dots M \quad (3.45)$$

n_j = es el número de vectores patrones para la clase w_j

Para determinar de qué clase es miembro un vector patrón desconocido x , **se asignan a la clase más cercana a su prototipo**; por lo tanto, nos basamos en su similitud, midiendo las distancias de clase, usando la distancia del vector prototipo de la clase.

Se puede usar una distancia para determinar la proximidad, entonces tenemos que realizar las

medidas de distancias:

$$D_j(x) = \|x - m_j\|_2 \quad j = 1 \dots M \quad (3.46)$$

$$\|a\| = (a^T a)^{\frac{1}{2}} \quad \text{Norma Euclidiana} \quad (3.47)$$

Por esto, se asigna x a la clase w_i , si $d_i(x)$ es la menor distancia encontrada. Con esto implicamos que la mejor concordancia o "matching" se basa en esta formulación.

O bien, es equivalente a demostrar que es equivalente a evaluar las funciones:

$$D_j(x) = x^T m_j - \frac{1}{2} m_j^T m_j \quad j = 1 \dots M \quad (3.48)$$

y se asignan x a la clase w_i si $d_i(x)$ lleva el valor numérico más grande.

El límite de decisión entre la clase w_i y w_j para un clasificador de mínima distancia es:

$$d_{ij} = d_i(x) - d_j(x) = 0 \quad (3.49)$$

3.5.2 MÉTRICAS UTILIZADAS EN EL RECONOCIMIENTO DE CARACTERÍSTICAS BIOMÉTRICAS

Un elemento importante en el reconocimiento de patrones son las medidas de similitud que se emplean para medir la distancia entre los vectores característicos que se obtienen mediante diferentes algoritmos de extracción de características. Se muestran 3 métricas las cuales se utilizaron en este proceso de la distancia mínima basándose en el trabajo de Montoya [77] donde se proponen estas medidas para utilizarlas en el reconocimiento de algunos métodos biométricos. Se definirán las distancias entre vectores de la siguiente forma. Dados $\vec{x} = (x_1, x_2, \dots, x_n)$ y $\vec{y} = (y_1, y_2, \dots, y_n)$ entonces las distancias se definen como:

Distancia euclidiana

$$d_E^2 = \sum_{i=1}^n (x_i - y_i)^2 \quad (3.50)$$

Distancia Canberra

$$d_C = \sum_{i=1}^n \frac{|x_i - y_i|}{|x_i| + |y_i|} \quad (3.51)$$

Distancia cityblock

$$d_{Cb} = \sum_{i=1}^n |x_i - y_i| \quad (3.52)$$

Se cuentan con estas 3 medidas de similitud, son de suma importancia ya que no sólo se debe tener en cuenta como obtener los vectores de características, sino también las medidas de similitud pertenecientes a espacios métricos que se emplean pues las mismas tienen influencia en la precisión con que se obtienen los resultados; se notará más adelante que algunas medidas no

Capítulo 3. Sistema propuesto

son las más adecuadas para clasificar o por el contrario con algunas de ellas se obtienen tasas de reconocimiento excelentes al momento de la clasificación.

4

Resultados experimentales

EN este capítulo se presentan los resultados obtenidos en diferentes experimentos realizados. Primero se muestra el rendimiento con la DHT y LBP en diferentes configuraciones, los experimentos fueron realizados de manera individual o bien usando la combinación de ambos métodos sin el uso LDA. Después se expone el rendimiento con el uso LDA, donde se evalúan la DHT y LBP en todas sus configuraciones.

Las distancias que se usaron en los experimentos para la clasificación son solo 3 (euclidiana, Canberra y cityblock).

Revisando los resultados se observa que la metodología propuesta en la Fig. 3.2, como método híbrido, resulta ser una herramienta poderosa para la extracción de características faciales frente a métodos clásicos empleados en el área de reconocimiento de rostros.

4.1 Experimentos

Los experimentos fueron realizados empleando la base de datos de rostros de Olivetti Research Laboratory [10]. Esta base de datos consiste de 10 imágenes de 40 individuos, las imágenes se encuentran en formato *pgm* con una resolución de 96x112 píxeles en 256 niveles de gris. La base de datos cuenta con imágenes de rostros con variaciones de luminosidad, diferentes expresiones faciales, orientación del rostro, tomas con lentes y sin lentes, lo que permite probar la robustez del sistema contra esas posibles variaciones. Tomando en cuenta que la base de datos de rostros ORL es una base de datos estándar utilizada por muchos investigadores, lo que permite realizar comparaciones con diferentes algoritmos propuestos.

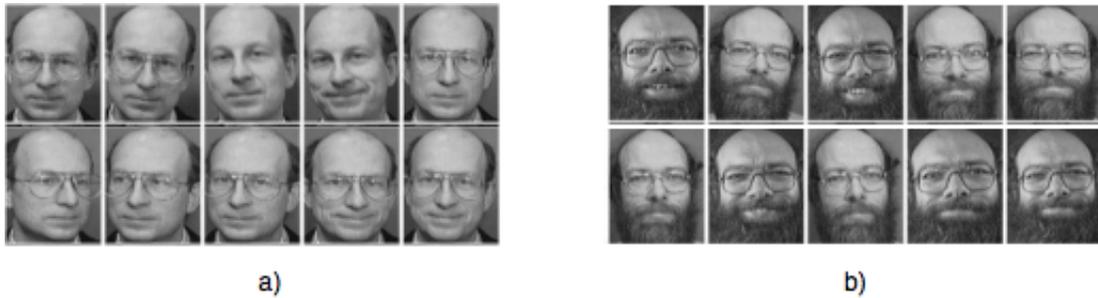


Figura 4.1: Imágenes ejemplo de la base de datos ORL

Se muestran algunas imágenes de ejemplo tomadas de la base de datos de rostros ORL *ORL*, en la Fig. 4.1. Los ejemplos consisten de 10 tomas diferentes de 2 individuos, las tomas incluyen cambios de expresión facial, iluminación accesorios (p. ej, lentes o barba) y postura. Todos experimentos aquí reportados fueron realizados bajo las siguientes condiciones:

- La extracción de características fueron tanto globales como locales, pero para comprobar la hipótesis propuesta en la tesis se experimentó usando la transformada Hermite en las siguientes configuraciones: $DHT(5,5)$, $DHT(4,4)$, $DHT(3,3)$ ¹, de los cuales por cada coeficiente de la DHT se requirieron 4 estadísticos (media, varianza, skewness y kurtosis); los LBP se usaron en sus formas globales y también en forma de parches dividiendo la imagen en 4 y 16 regiones respectivamente. Se comprobaron por separado y en conjunto, además se prueba que realmente un sistema híbrido funciona adecuadamente como lo sugiere la literatura; se muestra que ambos métodos por separado carecen de eficiencia en el momento de la clasificación.

Los métodos DHT y LBP en todas sus configuraciones generarán el vector de características de cada rostro en la base de datos.

- Obteniendo las firmas de cada rostro (vector de características) el siguiente paso es aplicar método Fisher-LDA, el objetivo es reducir la dimensión de los vectores y una mejor clasificación; es importante mencionar que dependiendo de la configuración usada en la extracción de características las formas tendrán un número de elementos distintos en cada vector; cabe destacar que en LDA las formas son siempre de 39 elementos en cada vector, debido a que una propiedad de LDA 3.4.1 nos dice que se obtendrán $C-1$ clases, y nosotros contamos con 40 clases que son los sujetos de cada carpeta de la base de datos mostrada en el anexo A.
- En el reconocimiento se usaron las mismas imágenes de la base de datos propuesta, y se basó en distancias, las usadas fueron euclidiana, Canberra y cityblock, mientras que para obtener

¹Los números 5,5 se refieren al orden del filtro y al orden de la transformada en la configuración del toolbox de Hermite usado en la tesis

el rendimiento del algoritmo o de los algoritmos se usó la técnica *matriz de confusión*, con la cual se observa el comportamiento de las clases al momento de la clasificación.

- Por último los resultados a comparar son de la literatura y se pueden consultar en los respectivos artículos citados

4.2 Evaluación de los métodos propuestos sin el uso del LDA

Los experimentos fueron realizados con todas las configuraciones posibles de la DHT y LBP, se usaron las 400 imágenes de la base datos ORL [10]; los métodos fueron combinados y la clasificación de los vectores fue realizada por medio de las distancias *euclidea*, *Canberra* y *cityblock*².

Las tablas que se muestran en esta sección, 4.1, 4.2, 4.3, 4.4, 4.5, son los resultados obtenidos y las tasas de clasificación sin el uso de LDA. Se puede observar que en la tabla 4.6 la distancia *Canberra* tiene la mejor tasa de clasificación, de igual forma muestro algunos otros experimentos, que realmente no fueron muy satisfactorios.

La combinación *Hermite*(5, 5)+*LBP*_{16,2}^{uni} 16*parches*, es la mejor descripción tanto global como local del rostro, junto con la métrica que mejor clasificó. El número de elementos del vector de características es de 372 elementos, es bastante alto y computacionalmente no muy adecuado. La mayoría de los resultados no parecen ser competitivos, carecen de robustez y no cumplen con requerimientos que se están buscando obtener en este trabajo.

Tabla 4.1: *Hermite global*

| Método | # Longitud vector | Distancias de prueba | | |
|-----------------------|-------------------|-----------------------|---------------------|----------------------|
| | | <i>Euclidiana</i> (%) | <i>Canberra</i> (%) | <i>Cityblock</i> (%) |
| <i>Hermite</i> (5, 5) | 84 | 54.80 | 67.40 | 50.46 |
| <i>Hermite</i> (4, 4) | 60 | 34.05 | 64.95 | 48 |
| <i>Hermite</i> (3, 3) | 40 | 34.78 | 65.93 | 46.88 |

Todos los resultados en las distancias de prueba están en porcentaje %,

Hermite(n,m) n=orden filtro y m=orden transformada,

LBP (Patrón Binario Local) Uniforme (P,R), donde P número de vecinos y R radio de la vecindad.

²Las métricas usadas en estos experimentos fueron las de mejores resultados, se usaron también algunas distancias como son Chis-q, sqEuclidean, coseno, correlación, Emd y Chebychev

Tabla 4.2: *LBP global*

| Método | # Longitud vector | Distancias de prueba | | |
|--------------------|-------------------|----------------------|--------------|---------------|
| | | Euclidiana(%) | Canberra(%) | Cityblock(%) |
| $LBP_{1,8}^{Uni}$ | 10 | 22.08 | 25.80 | 24.40 |
| $LBP_{16,2}^{Uni}$ | 18 | 30.53 | 25.05 | 30.53 |
| $LBP_{24,3}^{Uni}$ | 26 | 12.65 | 38.27 | 31.90 |

Todos los resultados en las distancias de prueba están en porcentaje %, Hermite(n,m) n=orden filtro y m=orden transformada, LBP (Patrón Binario Local) Uniforme (P,R), donde P número de vecinos y R radio de la vecindad.

Tabla 4.3: *LBP 4 y 16 regiones*

| Método | # Longitud vector | Distancias de prueba | | |
|--------------------|-------------------|----------------------|--------------|---------------|
| | | Euclidiana(%) | Canberra(%) | Cityblock(%) |
| $LBP_{1,8}^{Uni}$ | 40 | 38.82 | 39.35 | 38.82 |
| $LBP_{16,2}^{Uni}$ | 72 | 42.83 | 49.35 | 46.77 |
| $LBP_{24,3}^{Uni}$ | 104 | 43.95 | 54.73 | 52.02 |
| $LBP_{1,8}^{Uni}$ | 160 | 55.65 | 55.50 | 56.20 |
| $LBP_{16,2}^{Uni}$ | 288 | 57.55 | 61.62 | 60.60 |
| $LBP_{24,3}^{Uni}$ | 416 | 56.90 | 62.05 | 62.10 |

Todos los resultados en las distancias de prueba están en porcentaje %, Hermite(n,m) n=orden filtro y m=orden transformada, LBP (Patrón Binario Local) Uniforme (P,R), donde P número de vecinos y R radio de la vecindad.

4.2 Evaluación de los métodos propuestos sin el uso del LDA

Tabla 4.4: Hermite combinación LBP globales

| Método | # Longitud vector | Distancias de prueba | | |
|--|-------------------|----------------------|-------------|--------------|
| | | Euclidiana(%) | Canberra(%) | Cityblock(%) |
| <i>Hermite(5,5) + LBP_{8,1}^{uni}</i> | 94 | 44.97 | 68.03 | 51.42 |
| <i>Hermite(4,4) + LBP_{8,1}^{uni}</i> | 70 | 44.87 | 65.65 | 49.70 |
| <i>Hermite(3,3) + LBP_{8,1}^{uni}</i> | 50 | 45.15 | 67.27 | 49.55 |
| <i>Hermite(5,5) + LBP_{16,2}^{uni}</i> | 102 | 45.62 | 68.40 | 53 |
| <i>Hermite(4,4) + LBP_{16,2}^{uni}</i> | 78 | 45.25 | 66.55 | 51.58 |
| <i>Hermite(3,3) + LBP_{16,2}^{uni}</i> | 58 | 45.50 | 68.28 | 51.78 |
| <i>Hermite(5,5) + LBP_{24,3}^{uni}</i> | 110 | 45.62 | 68.97 | 54.70 |
| <i>Hermite(4,4) + LBP_{24,3}^{uni}</i> | 86 | 45.10 | 67.15 | 53.50 |
| <i>Hermite(3,3) + LBP_{24,3}^{uni}</i> | 66 | 45.60 | 69.17 | 53.30 |

Todos los resultados en las distancias de prueba están en porcentaje %,

Hermite(n,m) n=orden filtro y m=orden transformada,

LBP (Patrón Binario Local) Uniforme (P,R), donde P número de vecinos y R radio de la vecindad.

Tabla 4.5: Hermite combinación LBP 4 regiones

| Método | # Longitud vector | Distancias de prueba | | |
|--|-------------------|----------------------|-------------|--------------|
| | | Euclidiana(%) | Canberra(%) | Cityblock(%) |
| <i>Hermite(5,5) + LBP_{8,1}^{uni}</i> | 124 | 46.65 | 68.50 | 55.30 |
| <i>Hermite(4,4) + LBP_{8,1}^{uni}</i> | 100 | 46.20 | 67.20 | 53.85 |
| <i>Hermite(3,3) + LBP_{8,1}^{uni}</i> | 80 | 46.47 | 69.15 | 53.67 |
| <i>Hermite(5,5) + LBP_{16,2}^{uni}</i> | 156 | 48.70 | 70.35 | 59.42 |
| <i>Hermite(4,4) + LBP_{16,2}^{uni}</i> | 132 | 48.85 | 69.55 | 58.23 |
| <i>Hermite(3,3) + LBP_{16,2}^{uni}</i> | 112 | 48.95 | 71.55 | 57.93 |
| <i>Hermite(5,5) + LBP_{24,3}^{uni}</i> | 188 | 49.58 | 71.77 | 62.18 |
| <i>Hermite(4,4) + LBP_{24,3}^{uni}</i> | 164 | 48.60 | 71.15 | 61.55 |
| <i>Hermite(3,3) + LBP_{24,3}^{uni}</i> | 144 | 49.78 | 71.50 | 61 |

Todos los resultados en las distancias de prueba están en porcentaje %,

Hermite(n,m) n=orden filtro y m=orden transformada,

LBP (Patrón Binario Local) Uniforme (P,R), donde P número de vecinos y R radio de la vecindad.

Tabla 4.6: Hermite combinación LBP 16 regiones

| Método | # Longitud vector | Distancias de prueba | | |
|--|-------------------|----------------------|--------------|--------------|
| | | Euclidiana(%) | Canberra(%) | Cityblock(%) |
| <i>Hermite</i> (5, 5) + $LBP_{8,1}^{uni}$ | 244 | 52.35 | 72.02 | 65.93 |
| <i>Hermite</i> (4, 4) + $LBP_{8,1}^{uni}$ | 220 | 52.58 | 71.75 | 65.30 |
| <i>Hermite</i> (3, 3) + $LBP_{8,1}^{uni}$ | 200 | 52.53 | 71.07 | 65.02 |
| <i>Hermite</i> (5, 5) + $LBP_{16,2}^{uni}$ | 372 | 54.25 | 73.48 | 67.50 |
| <i>Hermite</i> (4, 4) + $LBP_{16,2}^{uni}$ | 348 | 54 | 72.40 | 66.97 |
| <i>Hermite</i> (3, 3) + $LBP_{16,2}^{uni}$ | 328 | 54.43 | 70.97 | 68.86 |
| <i>Hermite</i> (5, 5) + $LBP_{24,3}^{uni}$ | 500 | 56.63 | 72.55 | 67.93 |
| <i>Hermite</i> (4, 4) + $LBP_{24,3}^{uni}$ | 476 | 54.63 | 71.20 | 67.63 |
| <i>Hermite</i> (3, 3) + $LBP_{24,3}^{uni}$ | 456 | 54.90 | 69.50 | 67.40 |

Todos los resultados en las distancias de prueba están en porcentaje %,

Hermite(n,m) n=orden filtro y m=orden transformada,

LBP (Patrón Binario Local) Uniforme (P,R), donde P número de vecinos y R radio de la vecindad.

4.3 Evaluación de los métodos propuestos usando LDA

Siguiendo el método mostrado en la Fig. 3.10, una vez evaluados los métodos individuales y en conjunto; el siguiente paso es aplicar el método LDA a los experimentos de la sección anterior. En estos experimentos al igual que los anteriores se usaron las 400 imágenes de la base de datos ORL [10], pero aquí se implementó el método de validación cruzada.

Esta es una técnica sencilla y ampliamente usada para validación; consiste en dividir los datos en K conjuntos de forma generalmente aleatoria. En cada iteración se extrae un conjunto y con los k-1 conjuntos restantes se entrena el modelo, esta operación se repite de tal forma que en cada entrenamiento un conjunto se separa para disminuir el sesgo, para dar una estimación del comportamiento del modelo se extrae la media de todos los experimentos.

Los resultados que arroja el método cruzado en las clasificaciones LDA son competitivos con los de la literatura; los mejores resultados están resaltados en cada tabla, 4.7, 4.8, 4.9, 4.10, 4.11. En la tabla 4.10 encontramos un resultado del **98.37 %**, lo que nos indica que es la mejor forma global y local de representación de los rostros para disminuir los cambios de iluminación, poses y expresiones faciales, además los vectores de características obtenidos pueden ser representados con sólo 39 elementos.

Computacionalmente el sistema es bastante eficiente, al momento de la clasificación, algunos experimentos consumen menos de 5 segundos en realizar el proceso.

Los resultados obtenidos son bastante adecuados y equiparables con cualquiera de los métodos y resultados más trabajados en el área del reconocimiento de rostros.

Tabla 4.7: Hermite global combinación LDA

| Método | # Longitud vector | Distancias de prueba | | |
|-----------------------|-------------------|----------------------|-------------|--------------|
| | | Euclidiana(%) | Canberra(%) | Cityblock(%) |
| <i>Hermite</i> (5, 5) | 39 | 94.67 | 81.16 | 95.09 |
| <i>Hermite</i> (4, 4) | 39 | 85.26 | 66.98 | 86.80 |
| <i>Hermite</i> (3, 3) | 39 | 71.69 | 44.28 | 74.04 |

Todos los resultados en las Distancias de prueba están en porcentaje %,

Hermite(n,m) n=orden filtro y m=orden transformada,

LBP (Patrón Binario Local) Uniforme (P,R), donde P número de vecinos y R radio de la vecindad.

Tabla 4.8: LBP 4 y 16 regiones combinación LDA

| Método | # Longitud vector | Distancias de prueba | | |
|-------------------------------|-------------------|----------------------|-------------|--------------|
| | | Euclidiana(%) | Canberra(%) | Cityblock(%) |
| $LBP_{1,8}^{Uni} 4regiones$ | 39 | 30.10 | 19.37 | 29.54 |
| $LBP_{16,2}^{Uni} 4regiones$ | 39 | 59.77 | 29.89 | 58.35 |
| $LBP_{24,3}^{Uni} 4regiones$ | 39 | 74.59 | 38.59 | 74.96 |
| $LBP_{1,8}^{Uni} 16regiones$ | 39 | 77.80 | 56.43 | 82.56 |
| $LBP_{16,2}^{Uni} 16regiones$ | 39 | 82.61 | 73.92 | 83.33 |
| $LBP_{24,3}^{Uni} 16regiones$ | 39 | 71.63 | 79.14 | 79.47 |

Todos los resultados en las Distancias de prueba están en porcentaje %, Hermite(n,m) n=orden filtro y m=orden transformada, LBP (Patrón Binario Local) Uniforme (P,R), donde P número de vecinos y R radio de la vecindad.

Tabla 4.9: Hermite combinación LBP globales y LDA

| Método | # Longitud vector | Distancias de prueba | | |
|------------------------------------|-------------------|----------------------|-------------|--------------|
| | | Euclidiana(%) | Canberra(%) | CityBlock(%) |
| $Hermite(5, 5) + LBP_{8,1}^{uni}$ | 39 | 95.91 | 72.48 | 96.65 |
| $Hermite(4, 4) + LBP_{8,1}^{uni}$ | 39 | 87.36 | 70.25 | 91 |
| $Hermite(3, 3) + LBP_{8,1}^{uni}$ | 39 | 66.73 | 55.26 | 79.97 |
| $Hermite(5, 5) + LBP_{16,2}^{uni}$ | 39 | 96.62 | 70.79 | 97.36 |
| $Hermite(4, 4) + LBP_{16,2}^{uni}$ | 39 | 89.32 | 61.58 | 83.75 |
| $Hermite(3, 3) + LBP_{16,2}^{uni}$ | 39 | 67.14 | 58.86 | 76.73 |
| $Hermite(5, 5) + LBP_{24,3}^{uni}$ | 39 | 96.80 | 79.01 | 97.40 |
| $Hermite(4, 4) + LBP_{24,3}^{uni}$ | 39 | 90.26 | 72.93 | 93.44 |
| $Hermite(3, 3) + LBP_{24,3}^{uni}$ | 39 | 74.44 | 74.31 | 81.00 |

Todos los resultados en las Distancias de prueba están en porcentaje %, Hermite(n,m) n=orden filtro y m=orden transformada, LBP (Patrón Binario Local) Uniforme (P,R), donde P número de vecinos y R radio de la vecindad.

4.3 Evaluación de los métodos propuestos usando LDA

Tabla 4.10: Hermite combinación LBP 4 regiones y LDA

| Método | # Longitud vector | Distancias de prueba | | |
|--|-------------------|----------------------|-------------|--------------|
| | | Euclidiana(%) | Canberra(%) | Cityblock(%) |
| <i>Hermite(5,5) + LBP_{8,1}^{uni}</i> | 39 | 97.96 | 69.66 | 98.37 |
| <i>Hermite(4,4) + LBP_{8,1}^{uni}</i> | 39 | 94.22 | 74.51 | 95.91 |
| <i>Hermite(3,3) + LBP_{8,1}^{uni}</i> | 39 | 71.80 | 57.93 | 84.16 |
| <i>Hermite(5,5) + LBP_{16,2}^{uni}</i> | 39 | 97.13 | 77 | 98.06 |
| <i>Hermite(4,4) + LBP_{16,2}^{uni}</i> | 39 | 95.29 | 82.90 | 96.67 |
| <i>Hermite(3,3) + LBP_{16,2}^{uni}</i> | 39 | 77.06 | 75.86 | 91.78 |
| <i>Hermite(5,5) + LBP_{24,3}^{uni}</i> | 39 | 97.54 | 82.64 | 98.18 |
| <i>Hermite(4,4) + LBP_{24,3}^{uni}</i> | 39 | 95.90 | 83.79 | 97.40 |
| <i>Hermite(3,3) + LBP_{24,3}^{uni}</i> | 39 | 82.05 | 73.04 | 93.18 |

Todos los resultados en las Distancias de prueba están en porcentaje %,

Hermite(n,m) n=orden filtro y m=orden transformada,

LBP (Patrón Binario Local) Uniforme (P,R), donde P número de vecinos y R radio de la vecindad.

Tabla 4.11: Hermite combinación LBP 16 regiones y LDA

| Método | # Longitud vector | Distancias de prueba | | |
|--|-------------------|----------------------|-------------|--------------|
| | | Euclidiana(%) | Canberra(%) | Cityblock(%) |
| <i>Hermite(5,5) + LBP_{8,1}^{uni}</i> | 39 | 96.63 | 89.70 | 97.28 |
| <i>Hermite(4,4) + LBP_{8,1}^{uni}</i> | 39 | 96.46 | 77.65 | 97.13 |
| <i>Hermite(3,3) + LBP_{8,1}^{uni}</i> | 39 | 91.68 | 80.57 | 96.06 |
| <i>Hermite(5,5) + LBP_{16,2}^{uni}</i> | 39 | 80.57 | 80.17 | 80.77 |
| <i>Hermite(4,4) + LBP_{16,2}^{uni}</i> | 39 | 80.48 | 80.06 | 80.25 |
| <i>Hermite(3,3) + LBP_{16,2}^{uni}</i> | 39 | 31.17 | 56.61 | 38.25 |
| <i>Hermite(5,5) + LBP_{24,3}^{uni}</i> | 39 | 80.60 | 80.20 | 80.80 |
| <i>Hermite(4,4) + LBP_{24,3}^{uni}</i> | 39 | 80.42 | 80.38 | 80.43 |
| <i>Hermite(3,3) + LBP_{24,3}^{uni}</i> | 39 | 72.12 | 79.91 | 80.43 |

Todos los resultados en las Distancias de prueba están en porcentaje %,

Hermite(n,m) n=orden filtro y m=orden transformada,

LBP (Patrón Binario Local) Uniforme (P,R), donde P número de vecinos y R radio de la vecindad.

4.4 Comparación con otros métodos

Se realizó una comparación con los métodos mencionados usualmente en literatura de reconocimiento de rostros, mostrando resultados de algoritmos clásicos y algunos que combinan propiedades similares a las que se usaron en esta tesis, ver tabla 4.12; cabe destacar que los resultados mostrados son equiparables, ya que son tomados de artículos que usaron la misma base de datos ORL [10] y con métodos de clasificación similares.

Tabla 4.12: Comparación de rendimiento con algunos métodos todos de extracción de características faciales.

| Método | RR % |
|------------------------|-------|
| Eigenfaces[78] | 97.50 |
| ICA[78] | 93.80 |
| 2D PCA[79] | 98.30 |
| Fisher-Faces[78] | 98.50 |
| Kernel Eigen-faces[78] | 98 |
| LBP[80] | 94.25 |
| Gabor[80] | 94.5 |
| Gabor+LBP[80] | 95.93 |
| Gabor+LDA[12] | 95.38 |

RR (Tasa de reconocimiento.)

5

Conclusiones y trabajo futuro

LA transformada de Hermite junto con los patrones locales binarios constituyen una metodología adecuada para caracterizar rostros en la tarea del reconocimiento. Además, al incluir una etapa de análisis discriminante con el método de Fisher, se logra obtener una representación compacta que permite discriminar rostros de forma robusta.

La transformada de Hermite analiza rostros de forma global conservando las características visualmente importantes, dividiendo la imagen en coeficientes de baja y alta frecuencia.

Sin embargo, al ser un método global, no considera cambios locales lo que conduce a tasas de clasificación bajas. Por lo que es necesario introducir una medida local que permita aumentar la discriminación de las imágenes. El método por sí sólo no es muy robusto pero en combinación con LDA genera tasas de clasificación de más del 95 %.

Por otro lado, los patrones binarios permiten una descripción local de la imagen al considerar las variaciones alrededor de un píxel. Esta operación se puede realizar sobre toda la imagen para generar un descriptor global, esta representación describe las características faciales de forma más robusta que el espacio original LBP. En los resultados obtenidos se puede verificar que, al igual que el método de Hermite, no es suficiente su uso individual, pero en combinación con LDA, se obtienen clasificaciones cercanas al 83 % dependiendo de la configuración usada.

Los vectores característicos extraídos por la combinación de ambos métodos son adecuados para el reconocimiento de rostros. El sistema propuesto posee propiedades que lo convierten en una alternativa viable a algunos de los problemas que enfrentan muchos sistemas de reconocimiento de rostros (p.ej. iluminación, pose, etc).

La última parte de la metodología incluye un clasificador mediante distancias. En este caso uti-

lizamos las 3 distancias mencionadas en 3.5.2, que fueron con las que mejores resultados de clasificación se obtuvieron.

Los resultados muestran que el método logra tasas de clasificación superiores que los obtenidos por los algoritmos en el estado del arte, en condiciones controladas, representan una opción atractiva para muchas aplicaciones; además junto con análisis el Fisher ofrece un modelo robusto y eficiente que permite eliminar o al menos reducir algunos inconvenientes de varios métodos de reconocimiento de rostros, algunos de ellos mencionados en capítulos anteriores.

De acuerdo a los resultados, la mejor tasa de clasificación fue de 98.38 %. Existe un patrón recurrente con respecto a las imágenes que no se clasificaron correctamente, que es la que provoca que bajen tasas de clasificación; el rostro que contienen esa imagen presenta características únicas con respecto a su color de piel. Debido a que el algoritmo no considera esta característica, esto hace pensar que el color de piel influye en la identificación de rostros de manera importante al igual que el resto de los rasgos biométricos (ojos, boca, expresiones, etc).

Un aspecto que limita el rendimiento del algoritmo son las matrices singulares. En las tablas de la sección 4.3 algunos resultados se ven afectados debido a que los vectores de características son similares, lo que conduce a que no exista la matriz inversa que requiere el método de Fisher, produciendo regiones incorrectas.

Con respecto a los LBP, se dividieron las imágenes en regiones para aumentar la discriminación local. Dos esquemas fueron considerados: 4 y 16 regiones. Aunque se puede aumentar el número de regiones, no se consideró por dos razones: el costo computacional y la generación de vectores dispersos.

Como trabajo futuro quedaría pendiente evaluar bases de datos más complejas, comparando los resultados con los que existen en el estado del arte. Otro punto a probar es la transformada de Hermite a nivel multirresolución, además de utilizar otra variante llamada *Hermite rotada* y usando las mismas metodologías que en este trabajo se presentan.

El algoritmo LBP usado en la tesis posee la propiedad de invarianza a rotación, de la misma forma que se aplicaron los métodos en esta tesis, quedaría verificar si realmente esa propiedad puede aportar más robustez a los sistemas actuales de reconocimiento de rostros.



Anexo A

A.1 Base de Datos ORL

La base de datos ORL [10] cuenta con 40 imágenes de 10 individuos, estas imágenes de cada individuo fueron tomadas en diferentes sesiones; las variaciones incluyen distinta iluminación, expresión facial, uso de lentes, así como sujetos con barba, hombres, mujeres, un chico de raza negra y personas de edad avanzada. Todas las imágenes fueron tomadas en fondo homogéneo oscuro, con las personas en posición frontal. Se tomaron entre abril de 1992 y abril de 1994. Esta fue elaborada por AT&T Laboratories Cambridge en colaboración con el departamento de ingeniería de la Universidad de Cambridge.

Los archivos se encuentran en formato PGM, siendo su resolución de 92x112 píxeles, con 256 niveles de gris. La base está ordenada en 40 directorios, los cuales se identifican por una *s* y el número de persona, ejemplo (*s1,s2,s3...,s40*) y dentro de esos directorios las 10 imágenes comienzan con *1.pgm,2.pgm..10.pgm*, así cada directorio tiene sus 10 imágenes en ese formato. La base de datos *ORL* [10], se puede descargar gratuitamente en la página de internet de AT&T Laboratories Cambridge, formato comprimido *zip* o *tar*.









Bibliografía

- [1] Levy, M. N., Stanton, B. A., and Koeppe, B. M., [*Fisiología Elsevier*], Principles of physiology, 4th ed (2006).
- [2] Brunelli, R. and Poggio, T., “Face recognition: features versus templates,” in [*IEEE Transactions on Pattern Analysis and Machine Intelligence*], 1042–1052 (1993).
- [3] Cox, I. J., Ghosn, J., and Yianilos, P. N., “Featurebased face recognition using mixture distance,” in [*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*], 209–216 (1996).
- [4] Wiskott, L., Fellous, J.-M., Kuiger, N., and Von der Malsburg, C., “Face recognition by elastic bunch graph matching,” in [*Intelligent Biometric Techniques in Fingerprint and Face Recognition*], 355–396 (1999).
- [5] Turk, M. and Pentland, A., “Eigenfaces for recognition,” *Journal Of Cognitive Neuroscience*, (1991).
- [6] Belhumeur, P. N. and Kriegman, D. J., “Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* , 711–720 (1997).
- [7] Swets, D. L. and Weng, J. J., “Using discriminant eigenfeatures for image retrieval,” *IEEE Transactions On Pattern Analysis And Machine Intelligence* , Vol. 18 , 831–836 (1996).
- [8] Phillips, P., Moon, H., Rizvi, S. A., and Rauss, P., “The FERET evaluation methodology for facerecognition algorithms,” in [*Tecnical Report NISTIR 6264*], 137–143 (1999).
- [9] Georghiades, A., Belhumeur, P., and Kriegman, D., “From few to many: Illumination cone models for face recognition under variable lighting and pose,” in [*IEEE Trans. Pattern Anal Mach. Intelligence*], 643–660 (2001).
- [10] ATT, “<http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.htm>, the orl database of faces,” (1992-1994).
- [11] Pineda, G. F., *Reconocimiento de rostros mediante wavelets y redes neuronales*, Master’s thesis, ESIME Culhuacan IPN (2008).
- [12] Shen, L., Bai, L., and Fairhurst, M., “Gabor wavelets and general discriminant analysis for face identification and verification,” *Pattern Recognition Letters* (2007).
- [13] (Madrid), R. A. E., [*Diccionario de la lengua española*], 22ed. Real Academia Española (2002).
- [14] Bolle, R. M., Conell, J., Pankanti, S., and Ratha, N. K., [*Guide to Biometrics*], Springer (2003).

BIBLIOGRAFÍA

- [15] Kung, S. Y., Mak, M. W., and Lin, S. H., [*Biometric Authentication: A Machine Learning Approach*], Prentice Hall PTR (2005).
- [16] Zhao, W. and Rosenfeld, A., “Face recognition: A literature survey,” in [*ACM Computing Survey*], 399–458 (2003).
- [17] Maaten, L., Van der, J., Postma, E., and van den Herik, H. J., “Dimensionality reduction: A comparative review,” in [*Technical report, MICC, Maastricht University*], (2008).
- [18] Lu, X., “Image analysis for face recognition: a brief survey,” in [*Personal Notes*], (2003).
- [19] Alfonso, E., *Reconocimiento de iris usando la transformada Hermite multiescala*, Master’s thesis, UNAM (2008).
- [20] Hooser, S. D. and Nelson, S. B., [*Visual System*], In Encyclopedia of Life Sciences. John Wiley and Son (2005).
- [21] Kanade, T., “Picture processing system by computer complex and recognition of human faces,” in [*Kyoto University, Japan, PhD. Thesis*], (1973).
- [22] Yuille, A., Cohen, D., and Hallinanv, P., “Feature extraction from faces using deformable templates,” in [*IEEE Computer Society Conference on Computer Vision and Templates. San Diego, CA, USA*], 104–109 (1989).
- [23] Roeder, N. and Li, X., “Experiments in analyzing the accuracy of facial feature detection,” in [*Vision Interface ’95*], 8–16 (1995).
- [24] Colombo, C., Bimbo, A. D., and Magistris, H. J., “Human-computer interaction based on eye movement tracking,” in [*Computer Architectures for Machine Perception*], 258–263 (1995).
- [25] Nixon, M., “Eye spacing measurement for facial recognition,” in [*SPIE Proceedings*], 279–285 (1985).
- [26] Reisfeld, D., *Generalized symmetry transforms: attentional mechanisms and face recognition*, PhD thesis, Tel-Aviv University, PhD. Thesis, technical report (1994).
- [27] Graf, H. P., Chen, T., Petajan, E., and Cosatto, E., “Locating faces and facial parts,” in [*International Workshop on Automatic Face and Gesture Recognition*], 41–46 (1995).
- [28] Craw, I., Tock, D., and Bennett, A., “Finding face features,” in [*Second European Conference on Computer Vision*], 92–96 (1992).
- [29] Wiskott, L., Fellous, J.-M., Kuiger, N., and Von der Malsburg, C., “Face recognition by elastic bunch graph matching,” in [*IEEE Transactions on Pattern Analysis and Machine Intelligence*], 775–779 (1997).
- [30] Phillips, P., Rauss, P., and Der, S., “FERET (face recognition technology) recognition algorithm development and test report,” in [*U.S. Army Research Laboratory ARL-TR-995*], (1996).
- [31] Phillips, P., Moon, H., Rizvi, S. A., and Rauss, P. J., “The FERET evaluation methodology for facerecognition algorithms,” in [*IEEE Conference on Computer Vision and Pattern Recognition*], 137–143 (1997).

- [32] Sukthankar, G., “Face recognition: a critical look at biologically inspired approaches,” in [*Carnegie Mellon University, Pittsburgh, PA*], (1997).
- [33] Campadelli, P. and Lanzarotti, R., “A face recognition system based on local feature characterization,” in [*n Advanced Studies in Biometrics Vol.3161 LectureNotes in Computer Science*], 147–152 (2005).
- [34] Cendrillon, R. and Lowell, B. C., “Real-time face recognition using eigenfaces,” in [*Proceedings of the SPIE International Conference on Visual Communications and Image Processing*], 269–276 (2000).
- [35] Baron, J., “Mechanisms of human facial recognition,” in [*International Journal of Man-Machine Studies ,Vol.15*], 137–178 (1981).
- [36] Huang, J., “Detection strategies for face recognition using learning and evolution,” in [*Mason University, Fairfax, Virginia, Ph. D.*], (1998).
- [37] Bellman, R. E., [*Adaptive control processes*], Princeton University Press (1961).
- [38] Sirovich, L. and Kirby, M., “Low-dimensional procedure for the characterization of human faces,” *Journal of the Optical Society of America A: Optics, Image Science, and Vision* (1987).
- [39] Jain, A. K. and Dubes, R. C., “Algorithms for clustering data,” in [*New Jersey: PrenticeHall*], (1988).
- [40] Turk, M. and Pentland, A., “Face recognition using eigenfaces,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (1991).
- [41] Pentland, A., Moghaddam, B., and Starner, T., “Viewbased and modular eigenspaces for face recognition,” *IEEE Conference on Computer Vision and Pattern Recognition* , 84–90 (1994).
- [42] Moses, Y., Adini, Y., and Ullman, S., “Face recognition: the problem of compensating for changes in illumination direction,” *European Conf. Computer Vision* , 286–296 (1994).
- [43] Fisher, A., “The use of multiple measures in taxonomic problems,” *Annals of Eugenics* , Vol.7 , 179–188 (1936).
- [44] Martínez, A. M. and Kak, A. C., “PCA versus LDA,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* , 228–233 (2001).
- [45] Moghaddam, B. and Pentland, A., “A bayesian similarity measure for direct image matching,” *Proceedings 13th International Conference on Pattern Recognition* , 350–358 (1996).
- [46] Moghaddam, B. and Pentland, A., “Probabilistic visual learning for object representation,” in [*Transactions On Pattern Analysis And Machine Intelligence*], 666–710 (1997).
- [47] Meng, J. and Zhang, W., “Volume measure in 2d PCA based face recognition,” in [*Pattern Recognition Letters*], 1203–1208 (2007).
- [48] Sun, N., Wang, H., Ji, Z. H., Zou, C., and Zhao, L., “An efficient algorithm for kernel two-dimensional principal component analysis,” in [*Neural Computing and Applications*], 59–64 (2008).

BIBLIOGRAFÍA

- [49] F. Song, F., Zhang, D., Liu, H., and Tao, Q., “A parameterized direct lda and its application to face recognition,” in [*Neurocomputing*], 191–196 (2007).
- [50] Yang, Q. and Ding, X. Q., “Discriminant local feature analysis of facial images,” in [*IEEE International Conference on Image Processing*], 863–866 (2003).
- [51] Jing, X. Y. and Zhang, D., “A fourier LDA approach for image recognition,” in [*Pattern Recognition*], 453–457 (2005).
- [52] Pang, Y., Zhang, L., Li, M., Liu, Z., and Ma, W., “A novel gabor LDA based face recognition method,” in [*Advances In Multimedia Information Processing*], 352–358 (2004).
- [53] Beumier, C. and Acheroy, M., “Automatic face recognition,” in [*Proceedings symposium IMAGING, Eindhoven, The Netherland*], 77–89 (2000).
- [54] Yang, Q. and Tang, X., “Recent advances in subspace analysis for face recognition,” in [*SINOBIOMETRICS*], 275–287 (2004).
- [55] Thompson, P., [*Margaret Thatcher: A new illusion*], Perception (1980).
- [56] Bartlett, J. and Searcy, J., [*Inversion and configuration of faces*], Cognitive Psychology.
- [57] Chellappa, R., Wilson, C., and Sirohey, S., “Human and machine recognition of faces: a survey,” *Proceedings of the IEEE*, vol. 83(5) , 705–741 (1995).
- [58] Wechsler, H. and Huang, T., [*Face recognition: From theory to applications*], Proceeding of NATO-ASI (1998).
- [59] O’Toole, A., Roark, D., and Abdi, H., [*Recognitizing moving faces a psychological and neural synthesis.*], Trends in Cognitive Science, pag. 261–266 (2002).
- [60] Knight, B. and Johnston, A., [*The role of movement in face recognition.*], Visual Cognition, pag 265 – 273 (1997).
- [61] Campbell, F. W. and Robson, J. G., “Application of Fourier analysis to the modulation response of the eye,” *Journal of the Optical Society of America* , 581 (1964).
- [62] Graham, N. and Nachmias, J., “Detection of grating patterns containing two spatial frequencies: a comparison of single channel and multichannel models,” *Vision Research* , 251–259 (1971).
- [63] Wilson, H. R. and Bergen, J. R., “A four mechanism model for threshold spatial vision,” *Vision Research* , 19–32 (1979).
- [64] Ginsburg, A. G., “Visual information processing based on spatial filters constrained by biological data,” tech. rep., Aerospace Medical Research Laboratory (1978).
- [65] Sergent, J., “Microgenesis of face perception,” *Aspects of Face Processing*, NATO ASI (Holanda, 1986).
- [66] Lai, J. H., Yuen, P. C., and Feng, G. C., “Face recognition using holistic Fourier invariant features,” *Pattern Recognition* 34 (2001).
- [67] Martens, J. B., “The Hermite transform-theory,” *Speech and Signal Processing* 38 , 1595–1606 (1990).

- [68] Riley, K., Hopson, M., and Bence, S., [*Mathematical methods for physics and engineering*], Cambridge University Press (2002).
- [69] Abramowitz and Stegun, [*A handbook of mathematical functions*], Dover (1965).
- [70] Schwartz, W. R., Roberti de Siqueira, F., and Pedrini, H., “Evaluation of feature descriptors for texture classification,” *Journal of Electronic Imaging* 21 (2012).
- [71] Ojala, T., Pietikainen, T., and Maenpaa, T., “Multiresolution grayscale and rotation invariant texture classification with local binary patterns,” *Pattern Analysis and Machine Intelligence* , 971–987 (2002).
- [72] Ahonen, T., Hadid, A., and Pietikainen, T., “Face description with local binary patterns: application to face recognition,” *Pattern Analysis and Machine Intelligence* , 2037–2041 (2006).
- [73] Nava, U., *Overcomplete image representations for texture analysis*, PhD thesis, UNAM (2013).
- [74] Ripley, B., [*Pattern recognition and neural networks*], Cambridge University Press, United Kingdom (2000).
- [75] Liu, C. and Wechsler, H., “Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition,” *IEEE Transactions on Image Processing* 11 , 467–476 (2002).
- [76] Welling, M., “Fisher linear discriminant analysis,” tech. rep., Department of Computer Science, University of Toronto.
- [77] Montoya, J. A. and Leite, N., “Wavelet based feature extraction for fingerprint image retrieval,” tech. rep., Instituto de Computacao, Universidade Estadual de Campinas. Brasil (2006).
- [78] Yang, M. H., “Kernel eigenfaces vs. kernel fisherfaces : face recognition using kernel methods,” *Fifth IEEE International Conference on Automatic Face and Gesture Recognition* , 215–220 (2002).
- [79] Yang, J., Zhang, D., and Yang, J. Y., “Two-dimensional PCA: A new approach to appearance-based face representation and recognition,” *IEEE Trans on Pattern Analysis and Machine Intelligence* (2004).
- [80] Yang, J., Luo, M., and Jiao, Y., “Face recognition based on image latent semantic analysis model and SVM,” *Image Processing and Pattern Recognition* (2013).