



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN

**Reconocimiento de Rostros Humanos con PCNC Usando Distorsiones de Tipo
Rotación**

TESIS

QUE PARA OPTAR POR EL GRADO DE:

MAESTRA EN CIENCIAS (COMPUTACIÓN)

PRESENTA:

ZAMIRA CRUZ MONTERROSAS

TUTOR:

DRA. TETYANA BAYDYK

Centro de Ciencias Aplicadas y Desarrollo Tecnológico

MÉXICO, D.F. JULIO 2015



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

“Two things are infinite: the universe and human stupidity; and I'm not sure about the universe.”

Albert Einstein

“It would be possible to describe everything scientifically, but it would make no sense; it would be without meaning, as if you described a Beethoven symphony as a variation of wave pressure.”

Albert Einstein

Agradecimientos

Agradezco a mis tutores la Dra. Tetyana Baydyk y el Dr. Ernst Kussul por su incondicional apoyo y enorme paciencia. Sin lugar a dudas, unos de los investigadores más generosos, humildes y con vocación de enseñanza.

También agradezco a mis sinodales: Dra. Nidiyare Hevia Montiel, Dr. Francisco Javier García y Dr. Ángel Fernando Kuri Morales por tomarse el tiempo de leer mi trabajo y asesorarme. Al Mtro. José Luis Villarreal Benitez, por compartirme sus conocimientos y apoyarme constantemente. A la Mtra. Lizbeth Heras por su confianza.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) y los proyectos PAPIIT IN102014, IT 102814 por la beca y los incentivos económicos recibidos. A la UNAM y al Posgrado de Ciencia e Ingeniería en Computación.

Finalmente, mi más profundo agradecimiento a mi renovada y expandida familia consanguínea y no consanguínea. Sin ustedes definitivamente nada sería igual ya que probablemente hubiera optado por una actividad más lucrativa como la venta de comida. Tal vez ahora lo pueda hacer de forma científica.

A todos, muchas gracias.

Índice general

Agradecimientos	II
Contenido	III
Índice de Tablas	v
Índice de Figuras	VI
Abbreviations	VIII
1. Introducción	1
1.1. Objetivos	2
1.2. Contribución	3
1.3. Estructura de la Tesis	3
2. Fundamentos y Estado del Arte	5
2.1. Marco Teórico	6
2.2. Estado del Arte	8
3. Bases de Datos	15
3.1. Selección de las Bases de Imágenes	15
3.1.1. FRAV3D	23
3.1.2. FEI	25
4. Clasificador Neuronal de Codificación por Permutación (PCNC)	27
4.1. Introducción	27
4.2. Descriptores	29
4.2.1. RLD	30
4.3. Extractor de Propiedades	32
4.4. Codificación	34
4.4.1. Permutación	35
4.4.2. Análisis de Correlación	36
5. Algoritmo PCNC con distorsiones tipo rotación	39
5.1. Problemática	39
5.1.1. Hipótesis	40
5.2. Rotación	41

6. Experimentos y Resultados	45
6.1. Distorsiones	45
6.2. Desarrollo de experimentos	47
6.3. Resultados	49
6.3.1. Experimentos-Primera Fase	50
6.3.2. Experimentos- Segunda Fase	54
6.3.3. Experimentos-Tercera Fase	56
6.4. Comparativa	58
7. Conclusiones y Trabajo Futuro	59
7.1. Trabajo Futuro	61

Índice de tablas

3.1. Bases de imágenes de rostros 2D que consideran rotaciones en sus tomas	22
3.2. Orden de adquisición de imágenes (nomenclatura) y descripción	24
5.1. Resultados de prueba de PCNC y SVM con FRAV2D	41
5.2. Libertad de giro de la cabeza.	42
6.1. Tabla de rotaciones.	46
6.2. Descripción de desplazamientos	46
6.3. Descripción de las imágenes de la base de datos 2D de FRAV3D.	47
6.4. Descripción de las imágenes de la base de datos FEI.	48
6.5. Distribución de imágenes para entrenamiento y prueba en FRAV2D para la primera fase.	49
6.6. Distribución de imágenes para entrenamiento y prueba en FRAV2D para la segunda fase.	49
6.7. Distribución de imágenes para entrenamiento y prueba de la base de rostros FEI.	50
6.8. Resultado de los Experimentos con FRAV2D en la primera fase con la imagen de prueba 9.	51
6.9. Resultado de los Experimentos con FRAV2D en la primera fase con la imagen de prueba 10.	52
6.10. Resultado de los Experimentos con FRAV2D en la primera fase con la imagen de prueba 9 y 10.	53
6.11. Inclusión de distorsiones de rotación y desplazamiento para reconocimiento de la imagen 9.	54
6.12. Inclusión de distorsiones de rotación y desplazamiento para reconocimiento de la imagen 10.	55
6.13. Inclusión de distorsiones de rotación y desplazamiento para reconocimiento de las imágenes 9 y 10.	56
6.14. Resultados de los experimentos con FEI	56
6.15. Resultados de los experimentos con FEI sobre tomas de perfil	57
6.16. Comparativa entre resultados anteriores y nuevos resultados.	58

Índice de figuras

2.1. Proceso de Reconocimiento de rostros	7
3.1. Ejemplo de imágenes de FERET en escala de grises	16
3.2. Ejemplo de base de imágenes en la base FERET a color	16
3.3. Ejemplo de imágenes en Multi-PIE	17
3.4. Ejemplo de imágenes disponibles en ORL	17
3.5. Ejemplo de imágenes en la base CMU-Pittsburg	18
3.6. Ejemplo de imágenes en las base MIT-CBCL	18
3.7. Ejemplo de imágenes de la base CMU para el proyecto de reconocimiento	18
3.8. Ejemplo de imágenes de la base PIE	19
3.9. Ejemplo de imágenes de perfil y frontal que se estudian en NIST	19
3.10. Ejemplo de imágenes de la base M2VTS	19
3.11. Ejemplo de imágenes de la base UMIST	19
3.12. Ejemplo de imágenes de la base PUT	20
3.13. Ejemplo de imágenes de la base EURECOM KFD	20
3.14. Ejemplos de puntos de referencia etiquetados de la base EURECOM KFD	21
3.15. Ejemplo de imágenes de la base FEI	21
3.16. Ejemplo de imágenes de la base FRAV2D	21
3.17. Mallado 3D y mapa de profundidad	23
3.18. Mallado 3D y mapa de profundidad	24
3.19. Esquema de colocación del escáner y su colocación a la izquierda y derecha, respectivamente.	24
3.20. Esquema de colocación de la cámara	25
3.21. De izquierda a derecha se presentan ejemplos de toma frontal con expresión neutral, frontal sonriendo, giro a la derecha, giro a la izquierda.	26
4.1. Procesos de PCNC.	28
4.2. Esquema del sistema de reconocimiento de propósito general.	30
4.3. Estructura del RLD.	31
4.4. Resultado de la aplicación del filtro lineal.	32
4.5. Repetición de características en diferentes posiciones de la imagen. Los cuadros pequeños son regiones extraídas todas de diferentes zonas de los labios.	33
4.6. Permutación del vector U.	37
5.1. Flujo para la creación del vector de características de una imagen.	40
5.2. Rotaciones del rostro. La imagen a) muestra la posición inicial, b) las rotaciones sobre X (flexión y extensión), c) sobre el eje Z (inclinaciones) y d) sobre Y.	41

5.3. Rotación de una imagen sobre el origen.	42
5.4. Submuestreo en rotaciones.	43
5.5. Rotaciones con dos diferentes métodos.	44
6.1. Rotación de la imagen 1 de la base de datos FRAV3D	46
6.2. Imagen 9 (izquierda) e imagen 10 (derecha)	53

Abbreviations

PCNC	Permutation Coding Neural Classifier
LIRA	Limited Receptive Area
PCA	Principal Component Analysis
SVM	Super Vector Machine
K-NN	K-Nearest Neighbor
ORL	Olivetti Research Laboratory
2D	Two-Dimensional models
FEI	Faculdade de Engenharia Industrial (Facultad de Ingeniería Industrial)
FRVT	Face Recognition Vendor Test
FRGC	Face Recognition Grand Challenge
FERET	Face Recognition Technology
LDA	Linear Discriminant Analysis (Análisis Discriminante Lineal)
ICA	Independent Component Analysis (Análisis de Componentes Independientes)
LFA	Local Feature Analysis
TT	Trace Transform
RLBP	Robust Local Binary Pattern
RIFT	Rotation Invariant Feature Descriptor
SIFT	Scale-Invariant Feature Transform
SURF	Speed-Up Robust Feature
EP	Evolutionary Pursuit
EBGM	Elastic Bunch Graph Matching
AAM	Active Appearance Model
HMM	Hidden Markov Models
BRIEF	Binary Robust Independent Elementary Features
FRAV3D	Face Recognition and Artificial Vision 3D database

RLD	R andom L ocal D escriptor
PIE	P osition I llumination E xpression
ORL	O livetti R esearch L aboratory
CMU	C arnegie M elon U niversity
AU-Coded	A ction- U nit C oded
MIT-CBCL	M assachusetts I nstitute of T echnology - C enter for B iological & C omputational L earn
NIST	N ational I nstitute of S tandards and T echnology
M2VTS	M ulti M odal V erification for T eleservices and S ecurity
UMIST	U niversity of M anchester I nstitute of S cience and T echnology
PUT	P oznan U niversity of T echnology
KFD	K inect F ace D atabase
VRML	V irtual R eality M odeling L anguage
FACS	F acial A ction C oding S ystem
CBCL	C enter for B iological and C omputational L earning
AT&T	A merican T elephone and T elegraph
MTT	M asked T race T ransform

Capítulo 1

Introducción

El reconocimiento del entorno es una de las actividades que los humanos realizan con cierta facilidad pero que en realidad implica un alto grado de dificultad para una computadora. Ingenieros y neuro-científicos se han visto interesados en esta capacidad perceptiva humana y han desarrollado diversos estudios y experimentos con el fin de reproducirla e incluso automatizarla. Véase por ejemplo el reconocimiento y verificación de rostros.

Existen diversas razones que justifican el creciente interés por el reconocimiento facial, la principal radica en la forma no intrusiva y sencilla de recabar datos útiles para la autenticación de individuos. Si bien existen alternativas biométricas más certeras (reconocimiento de iris o huella dactilar), el reconocimiento de rostros es la herramienta más explotable en los sistemas de seguridad y vigilancia así como en los medios multimedia y de entretenimiento digital. Esto se debe a que hacen innecesaria la participación directa de los sujetos a reconocer al utilizar dispositivos de captura a distancia [1].

No obstante, y a pesar de que los sistemas de reconocimiento facial proporcionan un alto grado de libertad durante la adquisición de datos, también presentan significativas desventajas. En principio, los sistemas de reconocimiento degradan su desempeño cuando se enfrentan a situaciones reales donde las condiciones ambientales son inciertas o existe algún tipo de oclusión o distorsión sobre el objeto a reconocer. Algunos ejemplos de estas situaciones son: iluminación, variaciones de pose, expresiones faciales, envejecimiento, artefactos sobre el rostro, maquillaje, entre los más comunes. Luego se encuentran las limitantes de los propios dispositivos como la resolución, área de cobertura,

sensibilidad lumínica, calidad de la imagen y, finalmente pero no menos importante, están los factores intrínsecos al sistema como la galería de imágenes, el algoritmo de reconocimiento y los parámetros de ajuste.

Estos obstáculos hacen necesario definir la base del problema de reconocimiento que el sistema debe resolver para así realizar una selección efectiva del algoritmo, la galería de imágenes a trabajar y los dispositivos de entrada. En las subsiguientes secciones se definirán los fundamentos para el presente trabajo considerando las observaciones expuestas.

1.1. Objetivos

Actualmente existen diferentes algoritmos que permiten la clasificación de patrones para múltiples tareas como el análisis de componentes principales (PCA) [2], la máquina de vectores de soporte (SVM) [3], K-vecinos más cercanos (Knn) [4], redes neuronales, entre otros. Algunos clasificadores se apoyan en descriptores, estructuras que definen las características de una imagen [5–7]. Como alternativa a estos clasificadores y tomando como referencia el uso de descriptores, surge el clasificador neuronal de permutación de código o PCNC por sus siglas en inglés Permutation Coding Neural Classifier, algoritmo apto para trabajar con un amplio rango de imágenes y por tanto, multipropósito [8].

El clasificador PCNC se ha utilizado en tareas de reconocimiento de piezas micro-mecánicas [9], dígitos manuscritos [10] y rostros [11]. En cuando a esta última tarea, el sistema presentó un excelente desempeño con la base de imágenes ORL obteniendo tasas de error de 0.1 % para 200 imágenes de entrenamiento de un total de 400, 1.4 % para 160 imágenes, 2.15 % para 120 imágenes, 7.09 % para 80 imágenes y 16.1 % para 40 imágenes. No obstante, al probar el mismo clasificador con la base de imágenes 2D de FRAV3D [12] se presentó una tasa de error alta para imágenes donde aparecían distorsiones de tipo rotación sobre los ejes X, Y y Z. Estos resultados motivaron la investigación de las opciones que le proporcionarán al clasificador un mejor desempeño.

El presente trabajo retoma la investigación de reconocimientos de rostros con PCNC utilizando la base de imágenes 2D del set 3D creado por el grupo de Reconocimiento de Rostros y Visión Artificial (FRAV), y las imágenes generadas por grupo de la Facultad de

Ingeniería Eléctrica en Brasil (FEI). El objetivo es proveer una alternativa para mejorar los resultados del clasificador en la tarea de identificación de rostros que presenten distorsiones de tipo rotación. Esto conlleva a la adecuación de los parámetros y estructura general del actual clasificador manteniendo su funcionamiento básico.

1.2. Contribución

La tarea de reconocimiento de rostros con PCNC le permite al clasificador considerarse como una herramienta de clasificación de patrones apta para tareas de diferentes grados de complejidad y útil para ser utilizado en una amplia gama de imágenes. Resolver el problema de la sensibilidad a rotaciones en rostros proporciona robustez al clasificador y le permite consolidar su objetivo multipropósito.

1.3. Estructura de la Tesis

Capítulos – El contenido del presente trabajo se divide en siete secciones:

- Capítulo 1: Introducción
- Capítulo 2: Fundamentos y Estado del Arte. En este capítulo se expondrán los avances para el reconocimiento de rostros.
- Capítulo 3: Bases de Datos de Imágenes. La selección de los datos que entrenarán y probarán la red neuronal es de suma importancia, por esta razón se ha dedicado un capítulo para explicar las características de la base de imágenes así como las razones por la cual se eligió.
- Capítulo 4: Algoritmo PCNC. Aquí se explica la forma en que trabaja la red neuronal con el algoritmo PCNC, las ventajas respecto a otros algoritmos y los inconvenientes de su uso.
- Capítulo 5: Algoritmo PCNC con distorsiones tipo rotación. Una de las problemáticas que surgieron a partir del uso de PCNC fue el reconocimiento de imágenes que presentaban rotación. En la presente sección se explicará a detalle la forma en que se resolvió el problema.

- Capítulo 6: Experimentación y resultados. Se realizarán experimentos con el algoritmo PCNC en la base de datos FRAV y se comentarán los resultados obtenidos.
- Capítulo 7: Conclusiones y Trabajo Futuro. A partir de la experimentación y la investigación se dedujeron mejoras que pueden implementarse a futuro en colaboración con otras áreas.

Capítulo 2

Fundamentos y Estado del Arte

Hoy en día la tecnología biométrica ha expandido sus horizontes debido a la imperante necesidad de los gobiernos y empresas de llevar un control pasivo de la población para resguardar la seguridad física y virtual de personas e información. No obstante, aún quedan algunas incógnitas por resolver respecto a la efectividad y manejo de los sistemas biométricos, así como su impacto en las políticas de privacidad de los países.

Para describir las facilidades que ofrece la biometría, hemos de esclarecer su origen. El término biometría proviene del griego bios (vida) y metria (medición); esto es, el estudio de los métodos automáticos para el reconocimiento único de humanos basados en sus rasgos conductuales (comportamiento) o rasgos físicos [13].

Se tiene constancia que ya desde el siglo XV, en China, se utilizaba la biometría a través de huellas de la palma o dactilares como símbolos de distinción o firma. Inclusive se han encontrado tablas babilónicas de arcilla para realizar transacciones comerciales que datan del año 500 a. de C. en las que se registran las huellas dactilares. Pero no fue sino hasta el siglo XIX cuando William Herschel, Administrador británico del Distrito Indio, registra huellas dactilares y firmas en contratos civiles [14].

En 1870 Alphonse Bertillion, antropólogo francés, sugirió que ciertas medidas del cuerpo humano eran invariables con el tiempo. El sistema Bertillion, también conocido como Antropometría [15], incluía medidas del largo y ancho de la cabeza, y longitudes de pies y brazos para la identificación de sujetos, pero en 1903 el sistema falló al intentar identificar gemelos y sujetos de alto parecido físico.

Por el año de 1936 se sugiere utilizar como alternativa de identificación los patrones del iris y en la década de los 60's se inician investigaciones para el reconocimiento semi-automático de rostros. Fue hasta 1988 cuando se pudo implementar el primer sistema semi-automático de reconocimiento a la par que se publican los primeros estudios de la técnica llamada eigen-rostros. Kohonen [16] demostró que una simple red neuronal podía llevar a cabo el reconocimiento de rostros (previamente alineados y normalizados) aproximando los eigenvectores de la matriz de autocorrelación de la imagen del rostro y generando una descripción del mismo. Estos eigenvectores son los que se conocen como eigenrostros.

Ya por 1991 es posible realizar el reconocimiento facial en tiempo real y en 1993 se crea el programa FERET. En el año 2000 múltiples instancias del gobierno de los Estados Unidos de América, interesadas en la eficiencia del reconocimiento facial, patrocinan el FRTV, una tecnología abierta a gran escala que permite evaluar múltiples sistemas biométricos. Es en el año 2001 que se logra realizar la primera prueba real de reconocimiento facial durante el Super Bowl en Tampa, Florida. El reconocimiento facial se usó para identificar a más de cien mil fans a través de una cámara de seguridad con ayuda de la base de datos de la policía de Tampa; sin embargo, la tasa de error que se presentó fue muy alta [17].

Tras el análisis de fallas, en el 2004 el gobierno americano inicia el desafío FRGC con el fin de desarrollar algoritmos que mejorasen ciertas áreas de interés del reconocimiento facial [18]. Desde entonces, esta área ha tenido grandes avances en la técnica e implementación.

2.1. Marco Teórico

El reconocimiento de rostros es una tecnología biométrica ligada a diversos campos de estudio como el reconocimiento de patrones, visión computacional, gráficos, procesamiento de imágenes, computación estadística y aprendizaje automático. Los sistemas biométricos de reconocimiento de rostros operan bajo tres modalidades:

Verificación (autenticación). La verificación es un proceso de comparación uno a uno donde la imagen del rostro de un individuo se compara contra la imagen registrada del sujeto cuya identidad se reclama. El sistema determina si la imagen facial y la identidad declarada concuerdan.

Identificación (reconocimiento). La identificación implica una comparación uno a varios; esto es, la imagen de un rostro desconocido se compara contra los rostros que conforman la base de imágenes con el fin de asociar la identidad de entrada con la de alguno de los sujetos que pertenecen a la base.

Lista de seguimiento. El sistema de reconocimiento primero determina si el sujeto pertenece a la lista, si es así, identifica al individuo.

El reconocimiento de rostros, por lo tanto, confiere un problema de reconocimiento de patrones para la identificación o verificación de un rostro partir de una base de imágenes, generalmente bidimensionales, adquiridas previamente.

Desde el punto de vista humano, el reconocimiento de patrones puede entenderse como se muestra en la Fig.2.1.



FIGURA 2.1: PROCESO de Reconocimiento de rostros

Los sistemas automatizados de reconocimiento de rostros se basan en la abstracción del sistema de reconocimiento humano y por lo general dividen la tarea de reconocimientos en los módulos que a continuación se describen:

- Detección de rostros y puntos de referencia
- Normalización del rostro
- Extracción de características

- Comparación de características (identificación o verificación)

El presente trabajo considera que el rostro ya ha sido detectado y normalizado por lo que la base para la extracción de características la conforma el reconocimiento de patrones. Algunos métodos como el Análisis de Componentes Principales (PCA), Análisis Discriminante Lineal (LDA), Análisis de Componentes Independientes (ICA), Análisis de Componentes Locales (LFA), entre otros, utilizan un vector como forma de representación de un patrón. Las redes neuronales, por su parte, utilizan estos métodos más otras herramientas de modelación estadísticas dependiendo de la naturaleza del problema.

A continuación se explican algunas técnicas desarrolladas para la extracción de características y comparación durante las últimas décadas.

2.2. Estado del Arte

Actualmente existen diversos algoritmos empleados para el reconocimiento de rostros, entre los más importantes por su uso y baja tasa de error se encuentran los siguientes:

Análisis de Componentes Principales (PCA)

Es una técnica derivada de la transformación de Karhunen-Loeve que tiene como objetivo reducir la dimensión del conjunto de datos original. Cada rostro del grupo de entrenamiento es representado por un vector s -dimensional. PCA transforma linealmente este vector hacia un nuevo sistema coordenado o subespacio t -dimensional cuya base vectorial corresponde a la máxima variación de distancia respecto al espacio de la imagen original. La varianza de mayor tamaño del conjunto de datos es capturada en el primer eje (llamado el Primer Componente Principal), la segunda varianza más grande en el segundo eje, y así sucesivamente. Este nuevo subespacio es normalmente de mucho menor dimensión. La matriz de covarianza o de correlación construida es de naturaleza simétrica por lo que existe una base de vectores propios de la misma o también llamados eigenvectores de la matriz de covarianza [19].

Análisis de Componentes Independientes (ICA)

ICA surge como solución al problema de la separación ciega de fuentes. Utiliza una transformación lineal para expresar un grupo de variables aleatorias como

una combinación lineal de variables de entrada estadísticamente independientes. El criterio de selección de la función de transformación lineal involucra la minimización de información mutua expresada como una función de cúmulos de alto orden [20]. Básicamente PCA sólo considera los momentos de 2o orden y decorrelaciona datos, mientras que ICA considera estadísticas de mayor orden e identifica las variables independientes de entrada de entre las mezclas lineales (observables). Bartlett et al. [21] sugieren dos arquitecturas ICA para la tarea de reconocimiento: imágenes básicas estadísticamente independientes y representación de código factorial.

Análisis Discriminatorio Lineal(LDA)

LDA es una herramienta estándar de reconocimiento de patrones que clasifica a través de aproximaciones estadísticas de clases desconocidas con respecto a clases conocidas. Para todas las muestras de las clases se define la matriz de dispersión inter-clase S_B e intra-clase S_W . El objetivo es maximizar S_B y minimizar S_W . En otras palabras, maximizar el radio $\det|S_B|/\det|S_W|$. Esto se logra cuando los vectores columna de la matriz de proyección son los eigenvectores de 2.1.

$$S_W^{-1}S_B^2 \quad (2.1)$$

La propiedad uni-ejemplar de LDA (representación de cada clase con un solo ejemplar que es la muestra promedio de la clase) simplifica el proceso de clasificación; no obstante, en datos faciales de alta dimensión genera problemas con muestras de tamaño pequeño resultado de la insuficiencia de ejemplos de entrenamiento, comparados con la dimensionalidad del espacio muestra [22].

Seguimiento Evolutivo (EP)

Este es un método de representación con enfoque adaptativo para la codificación y clasificación de imágenes a través de una base óptica. Para que el sistema aprenda una base óptima necesaria para la compresión de datos y la clasificación de patrones, debe incrementar su habilidad de generalización. Esto se logra median-do entre la minimización del riesgo empírico en el entrenamiento y la reducción del intervalo de confianza que se presenta durante las pruebas con imágenes no registradas. EP implementa estrategias propias de los algoritmos genéticos para

la búsqueda del espacio de posibles soluciones con el fin de determinar una base óptima. El proceso inicia con la proyección de datos originales a un espacio PCA "blanqueado" de menor dimensión. Aquí la evolución queda definida por la función de aptitud y ésta, a su vez, depende de la precisión de ejecución (riesgo empírico) y separación de clases (intervalo de confianza). La precisión indica hasta qué punto el aprendizaje se considera exitoso mientras que la separación muestra una indicación de la adaptabilidad deseada en futuras pruebas [23].

EBGM (Correspondencia entre Agrupaciones de Grafos Elásticos)

La teoría considera que los rostros humanos comparten una estructura topológica similar y por lo tanto, pueden representarse de forma generalizada a través de grafos con nodos posicionados en puntos estratégicos (ejes, nariz, pómulos, etc.) y bordes etiquetados con vectores de distancia 2D. Cada nodo contiene un grupo de 40 coeficientes complejos de wavelets de Gabor en diferentes escalas y orientaciones (fase, amplitud) llamados "jets". El reconocimiento se basa en el etiquetado de estos grafos ya que con ello se mide la similitud de la respuesta del filtro de Gabor con respecto a cada nodo Gabor. Un grafo etiquetado es un conjunto de nodos conectados, donde cada nodo se etiqueta con jets y cada borde con distancias [24]. La dificultad del método radica en posicionar de forma precisa el punto de referencia. Algunas veces PCA y LDA ayudan a resolver esta tarea.

Métodos de Kernel

La técnica consiste en un mapeo de datos hacia un espacio de características de mayor dimensión donde cada coordenada corresponde a una característica del campo de datos. Este procedimiento resulta en un conjunto de puntos dentro de un espacio euclidiano. Se considera que el despliegue de datos en un subespacio no responde a un mapeo necesariamente lineal y por ende, la relación entre ellos es muy abierta. Las funciones MK permiten operar en el espacio de características sin tener que procesar las coordenadas de los datos sino el producto interno de las imágenes de todos sus pares, disminuyendo así la carga computacional. Los métodos de kernel pueden funcionar en conjunto con otros algoritmos tales como SVM, procesos gaussianos, LDA, PCA, análisis de correlación canónica, entre otros. Una de las desventajas que tienen estos algoritmos es su origen no paramétrico, lo que conlleva a una falta de transparencia en los resultados [25, 26].

Transformada de Traza (TT)

La Transformada de Traza por sus siglas en inglés TT es una generalización de la transformada Radon que consiste en trazar una imagen con líneas rectas a lo largo de las cuales se calculan ciertos funcionales de la función de la imagen. Dependiendo de la traza de funcionales seleccionada/usada/utilizada se producirán diferentes transformadas de traza de una misma imagen. El cálculo de características invariantes se puede utilizar para identificar objetos en una base de datos de tipo de inventario. La transformada de traza es global, por lo que no puede hacer frente a la oclusión o multiplicidad de objetos. Otra limitación es que las características que construye no son invariantes a los cambios de iluminación [27, 28].

Modelo de Apariencia Activa (AAM)

Es un modelo estadístico integrado que combina los modelos de variaciones de forma con el de apariencia en un marco normalizado de forma. El algoritmo utiliza la diferencia entre la estimación actual del modelo de apariencia y la imagen destino como referencia para la optimización del proceso. Valiéndose de la técnica de mínimos cuadrados se logran coincidencias con nuevas imágenes muy rápidamente [29].

Modelo Morfológico 3-D y Reconocimiento 3-D

El rostro humano es una superficie que yace en un espacio tridimensional, por lo tanto, la creación de este algoritmo consideró que un modelo 3D atendería mejor la necesidad de representar y reconocer rostros aún bajo variantes de iluminación, pose, etc. Blantz et al. [30] propusieron un modelo morfológico 3D del rostro que codifica su estructura geométrica y textura en términos de parámetros. El algoritmo debe considerar la textura y el rango del rostro, luego se procesa el rango de la imagen eliminando partes innecesarias como el cabello. Finalmente se calcula una forma canónica de la superficie. Este tipo de representación es insensible a cambios de orientación y expresiones faciales simplificando así la tarea de reconocimiento. El reconocimiento en sí se lleva a cabo sobre una superficie canónica. Uno de los inconvenientes que presenta este método es el costo en tiempo y hardware para generar los modelos 3D de los sujetos que conformarán la base de datos [31].

Enfoque Bayesiano

El enfoque Bayesiano contempla la similitud de características de forma probabilística. Parte de la creencia Bayesiana considera que las diferencias de intensidad de una imagen con otra son características de las variaciones normales de los individuos. Las variaciones en imágenes faciales son: intrapersonales y extrapersonales, su similaridad se mide a través de una regla Bayesiana. El modelo bayesiano, como todo modelo probabilístico, requiere de un importante grupo de datos para obtener resultados confiables. Los resultados son sensibles a la forma funcional si no se interpreta correctamente el término de error, lo que puede generar conclusiones variadas [32].

Modelo Oculto de Markov (HMM)

Es un modelo que caracteriza estadísticamente las propiedades desconocidas u ocultas de una señal a partir de las observables, asumiendo que el sistema a modelar es un proceso de Markov. HMM se compone de dos procesos inter-relacionados. El primero consta de una cadena de Markov con un número finito de estados, una matriz de probabilidad con transición de estados y una distribución de probabilidad del estado inicial; el segundo, un grupo de funciones de densidad de probabilidad para cada estado. Algunas desventajas que presenta este modelo es la cantidad de información necesaria para construir los modelos y las secuencias para generar la historia evolutiva [33].

Boosting y Soluciones de Ensamble

El método utiliza un aprendiz débil o "weak learner" en una versión probada de una muestra de entrenamiento generalizada perteneciente a un grupo de clasificadores de su tipo. Los algoritmos de boosting aprenden iterativamente clasificadores débiles con respecto a la distribución y los añaden a un clasificador final fuerte. Cada vez que se añade un clasificador débil, se actualiza el peso de los datos. Las muestras clasificadas correctamente pierden peso y las mal clasificadas lo ganan. Esto con el fin de que los clasificadores débiles se enfoquen en las muestras que no hayan podido ser clasificadas.

Existen algoritmos cuyo funcionamiento es muy semejante al de boosting conocidos como "algoritmos leveraging", pero para que un algoritmo pueda considerarse de boosting debe probar que utiliza el boosting en la aproximación de la formulación de aprendizaje correcta [34].

Viola y Jones [35] construyeron el primer sistema de detección de rostros en tiempo real utilizando AdaBoost [36], considerado un hito en la investigación de detección facial pero Guo et al. [37] realizaron los primeros acercamientos al reconocimiento de rostros utilizando los métodos de AdaBoost.

Máquina de Soporte Vectorial (SVM)

SVM es un modelo de aprendizaje supervisado que busca realizar una separación óptima de puntos de una clase a otra a través de un hiperplano usualmente de mayor dimensionalidad a la de los datos de entrada. SVM se ayuda de PCA para extraer las características a utilizar para realización de la clasificación y para obtener la función discriminante. SVM utiliza a PCA como herramienta para extraer las características del rostro en las imágenes y determinar las funciones discriminantes entre cada par de imágenes que deba procesar. Dependiendo de la naturaleza de los datos, el modelo del hiperplano puede resultar bastante complejo y único, además de incurrir siempre en un error de ajuste necesario para acercarlo más posible al modelo de la SVM al modelo original [38].

Algunas de las diferencias entre el reconocimiento de patrones en general y el reconocimiento de patrones en rostros surgen de la fotometría (direccionamiento de la luz, intensidad, modelo de iluminación, entre las principales) y geometría (pose, sistema coordinado, perspectiva, apariencia, entre otras) del sistema de adquisición imágenes. Es necesario comprender estas características tanto para desarrollar el sistema de reconocimiento como para probarlo, por esta razón diversos grupos de estudio generan bases de imágenes de acuerdo a las necesidades de la investigación a llevar a cabo. En el capítulo siguiente se habla de múltiples opciones de bases de imágenes disponibles para evaluar los sistemas de reconocimiento facial.

Una característica importante de un patrón visual, asociada al movimiento, es la ubicuidad. Esta se presenta principalmente en tomas de video a pesar de realizar el análisis cuadro por cuadro. En escenarios de vigilancia, no sólo lo anteriormente mencionado altera las condiciones de entrenamiento del sistema sino también las características propias de la imagen como su resolución.

El patrón facial especifica el patrón visual y agrega características propias del rostro tales como: la deformación, edad, superficie, similaridad intra-clase e inter-clase, y obstrucciones tales como maquillaje, lentes, cabello, por mencionar algunas. La mayoría de

los sistemas selecciona algunas de estas características con el fin de generar soluciones locales pero rápidas que permitan ampliar la eficiencia de los algoritmos de reconocimiento posteriormente.

Capítulo 3

Bases de Datos

Para la implementación de un clasificador es necesario contar con una fuente de objetos clasificados de forma certera; esto es, un conjunto de datos conocidos que permitan entrenarlo. Por esta razón, la selección de la base de datos es de gran importancia. En las siguientes secciones se detallan los criterios considerados para la elección de la base de datos y la descripción de algunas de las opciones disponibles.

3.1. Selección de las Bases de Imágenes

Los sistemas de reconocimiento de rostros requieren de una base de datos para realizar el entrenamiento y prueba del sistema. La elección de la base obedece a las necesidades del algoritmo a evaluar. En general, no existe una base de imágenes estándar que satisfaga los múltiples requerimientos de los problemas a estudiar por lo que se han desarrollado diferentes alternativas y protocolos.

Aunque ya existe un gran avance en la elaboración de bases de imágenes para el reconocimiento de rostros; aún no se logra tener un registro preciso y sistematizado de expresiones, razas, cambios de iluminación, rotaciones, y demás variaciones. El reto se debe a la naturaleza del gesto (natural o fingido), las propiedades del ambiente (iluminación, fondo, entre otras) y las características físicas propias de cada persona. Esto hace necesario contar con bases que contengan muestras de expresiones de diferentes sujetos bajo múltiples condiciones (iluminación, oclusiones, rotaciones, entre la más comunes), tarea bastante complicada de abarcar en su totalidad.

Para el trabajo presente se buscaron bases de datos cuyas imágenes presentaran rotaciones sobre Y y Z principalmente, debido a que la tarea principal era el estudio de distorsiones de tipo rotación más comunes en imágenes con rostros humanos. Otro criterio de selección fue la homogeneidad en el fondo de las imágenes, además de que no presentara más de un rostro a la vez. La centralización del rostro en la imagen era conveniente más no decisiva. Finalmente se consideró como factor importante que la base de imágenes tuviera licencia abierta para poder incluirla en los experimentos. A continuación se listan las bases de datos de imágenes faciales disponibles que trabajan con rotaciones o cambios de posiciones, algunas de ellas libres (con la correspondiente inclusión de agradecimientos) y otras con licencias de uso.

FERET. La base de datos fue dirigida por el Dr. Harry Wechsler de la Universidad de George Mason en colaboración con el Dr. Phillips. Se creó en 15 sesiones entre Agosto de 1993 a Julio de 1996 y está compuesta por un grupo de 14051 imágenes de 1199 individuos divididas en sets de estudio de pose e iluminación. También se incluyó un set tomado más de un año después entre la primera y última sesión fotográfica para estudiar el impacto de los cambios de apariencia por el paso del tiempo. Actualmente existen dos bases de datos, una en escala de grises (Figura 3.1) y otra a color (Figura 3.2).



FIGURA 3.1: Ejemplo de imágenes de FERET en escala de grises



FIGURA 3.2: Ejemplo de base de imágenes en la base FERET a color

Multi-PIE. En el año 2000 se creó esta base de datos a cargo de la Universidad Carnegie Mellon con el fin de estudiar pose, iluminación y cambios de expresión (Figura 3.3). En esta base se amplió el número de sujetos, la cantidad de sesiones y de expresiones capturadas, así como los ángulos de toma de la cámara. La base cuenta con 750000 imágenes de 337 sujetos diferentes.



FIGURA 3.3: Ejemplo de imágenes en Multi-PIE

AT&T ORL. Consta de diez diferentes imágenes de 40 distintos sujetos, algunas tomas hechas en tiempos diferentes con variaciones de iluminación, expresión facial (ojos abiertos/cerrados, sonrisa/sin sonrisa) y detalles faciales (lentes/ no lentes). Las tomas capturan poses frontales y rostros apuntando hacia arriba (con tolerancia para algunos movimientos laterales) como lo muestra la Figura 3.4.



FIGURA 3.4: Ejemplo de imágenes disponibles en ORL

CMU-Pittsburgh con Codificación AU. Base de datos para el estudio de la expresión y pose realizada con la participación de 182 sujetos de diferentes etnias y edades (18-50 años), cada uno cuenta con 1917 secuencias realizando múltiples muestras de las acciones FACS [39]. Las imágenes se encuentran en escala de grises y a color; fueron tomadas con dos cámaras Panasonic WV3230 conectadas a una video grabadora Panasonic AG-7500. Una de las cámaras se posicionó frente al sujeto y la otra 30° a su derecha. Un ejemplo de una imagen extraída de la base se muestra en la Figura 3.5.

MIT-CBCL Imágenes del rostro de 10 sujetos divididas en dos grupos: uno de alta resolución de perfil completo, medio perfil y frontal; y otro de imágenes sintéticas

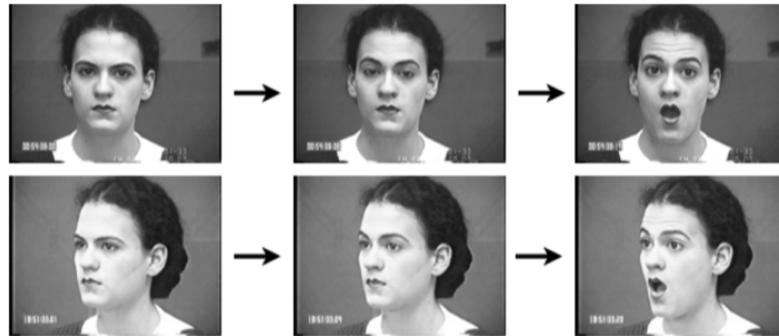


FIGURA 3.5: Ejemplo de imágenes en la base CMU-Pittsburg

renderizadas de modelos 3D (324 tomas por sujeto) de la cabeza de diez diferentes sujetos. El grupo de prueba está compuesto por 200 imágenes por sujeto con variaciones de iluminación, pose y fondo (Figura 3.6).



FIGURA 3.6: Ejemplo de imágenes en las base MIT-CBCL

Proyecto de Detección de Rostros CMU. La base de datos creada por Henry Schneiderman y Takeo Kanade en la Universidad de Carnegie Mellon, tiene como objetivo la evaluación de algoritmos para la detección frontal y vista de perfil de rostros humanos en ambientes naturales. La base está compuesta por 209 imágenes en formato .pgm de una o más personas por imagen con vistas de perfil en diferentes escenarios (Figura 3.7).



FIGURA 3.7: Ejemplo de imágenes de la base CMU para el proyecto de reconocimiento

Base de Datos PIE. Entre Octubre y Diciembre del año 2000, en el instituto de robótica de la universidad Carnegie Mellon, se obtuvieron 41368 imágenes de 68 sujetos; a cada sujeto se le hicieron tomas bajo 13 diferentes poses, 43 condiciones de iluminación y 4 expresiones, de ahí las iniciales PIE (Pose, Iluminación y Expresión) (Figura 3.8).



FIGURA 3.8: Ejemplo de imágenes de la base PIE

NIST. La base la componen 3 CD-ROMs con un total de 3248 imágenes de tamaño variable tomadas a 1573 individuos (1495 hombres y 78 mujeres) en pose frontal y de perfil (Figura 3.9). Las imágenes se encuentran en escala de grises de 8 bits.

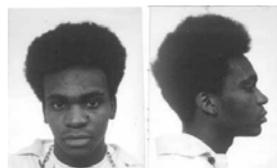


FIGURA 3.9: Ejemplo de imágenes de perfil y frontal que se estudian en NIST

M2VTS Extendida. M2VTS se conforma de 185 imágenes de 37 diferentes rostros, 5 tomas por cada persona tomadas en intervalos de una semana o cuando existían cambios drásticos en el rostro de los participantes (Figura 3.10). Las rotaciones abarcan poses graduales hacia ambos lados mientras los individuos contaban del uno al nueve en su lengua natal.



FIGURA 3.10: Ejemplo de imágenes de la base M2VTS

UMIST. La base la componen 564 imágenes de 20 sujetos de diferente raza/sexo/apariencia, cada uno cubriendo un rango de poses que van desde la vista frontal hasta el perfil completo. Las imágenes se encuentran en formato PGM y miden aproximadamente 220x220 píxeles (Figura 3.11).



FIGURA 3.11: Ejemplo de imágenes de la base UMIST

PUT. La base de datos contiene diez mil imágenes de alta resolución de 100 personas. Cada imagen fue capturada en un ambiente con condiciones controladas y se le añadieron datos adicionales tales como rectángulos que enmarcan el rostro, ojos, nariz y boca, puntos de interés de las posiciones y modelos de contorno editados manualmente (Fig. 3.12).



FIGURA 3.12: Ejemplo de imágenes de la base PUT

EURECOM KFD. El conjunto de datos consiste en imágenes faciales multimodales de 52 personas (14 mujeres, 38 hombres) adquiridas por medio del sensor del Kinect. Los datos se capturaron en dos sesiones con un intervalo de dos semanas de separación y se obtuvieron tomas bajo diferentes expresiones faciales (neutral, sonriendo, boca abierta), pose (perfil izquierdo y derecho), condiciones de iluminación y oclusiones (oclusión de ojos, boca y lateral con una hoja de papel) como se muestra en la Figura 3.13. Además de las imágenes RGB, también se obtuvo un mapa de profundidad (como bitmap y archivo de texto con los niveles de profundidad sensados por el Kinect), y los datos 3D asociados a cada imagen. Adicionalmente se incluyeron posiciones de referencia etiquetadas para: ojo izquierdo, ojo derecho, nariz, lado izquierdo y derecho de la boca, y mentón, así como referencia del sexo, año de nacimiento, origen étnico y fecha de la toma (Figura 3.14).

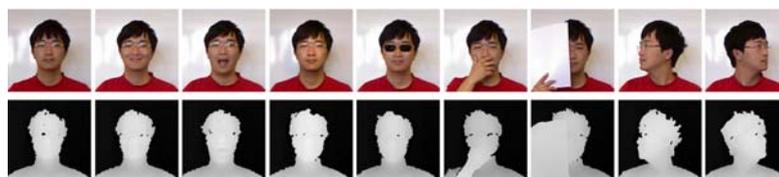


FIGURA 3.13: Ejemplo de imágenes de la base EURECOM KFD

FEI. Esta es una base de imágenes brasileña que contiene un set de 2800 fotografías a partir de 200 individuos (100 hombres y 100 mujeres). Las tomas se hicieron entre Junio de 2005 y Marzo de 2006 en el Laboratorio de Inteligencia Artificial de la Universidad de Sao Bernardo do Campo, Sao Paulo, Brasil. Las imágenes son a

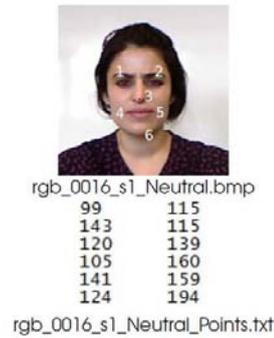


FIGURA 3.14: Ejemplos de puntos de referencia etiquetados de la base EURECOM KFD

color con fondo blanco y retratan cambios de pose gradual hasta llegar a los 180 grados, su tamaño es de 640x480 pixeles (Figura 3.15).



FIGURA 3.15: Ejemplo de imágenes de la base FEI

FRAV3D. El grupo de Reconocimiento Facial y Visión Artificial (Face Recognition & Artificial Visión en inglés) tomó como muestra a 106 sujetos y a cada uno le realizó 16 tomas a color (11236 imágenes en total) que consideraban cambios de pose, expresión e iluminación. Una muestra de las imágenes que contiene esta base se presenta en la Figura 3.16.



FIGURA 3.16: Ejemplo de imágenes de la base FRAV2D

La Tabla 3.1 muestra de forma sintetizada las bases de datos descritas anteriormente.

Tabla 3.1: Bases de imágenes de rostros 2D que consideran rotaciones en sus tomas

Nombre	Cantidad	Sujetos	Descripción	Pose	Iluminación	Expresión	Otros	Libre
FERET	14051	1199	Escala de grises, 8 bits	Si	Si	No	Retoque, tiempo	No
CMU-MultiPIE	750000	337	Imágenes a color de alta resolución	Si	Si	Si	Paso del tiempo	No
At&t (ORL)	400	40	92X112 pixeles en escala de gris	Si	Si	Si	Obstrucción por lentes	Si
CMU-Pittsburgh AU-Coded	348894	182	8 bit escala de grises	Si	No	Si	Razas	No
CBCL MIT	3240	10	Imágenes de alta resolución renderizadas de modelos 3D de sujetos	Si	Si	No	No	Si
Proyecto de Detección de Rostros CMU	209	-	Escala de grises con fondos realesy múltiples personas en una misma imagen	Si	No	Si	Situaciones reales	Si
CMU-PIE	41368	68	Imágenes a color con múltiples fondos	Si	Si	No	No	Si
NIST	3248	1573	Imágenes en escala de grises	Si	No	No	No	No
M2VTS Extendida	185	37	Imágenes a color	Si	No	Si	No	No
UMIST	564	20	PGM de 220X220	Si	No	No	No	Limitada
PUT	9971	100	Imágenes a color en alta definición	Si	No	No	No	Si
FEI	2800	200	Imágenes a color de 640X480	Si	Si	Si	No	Si
EURECOM-Kinect	468	52	Imágenes a color con profundidad	Si	Si	Si	Oclusión	No
FRAY	3488	109	imágenes a color de 320X240 pixeles	Si	Si	Si	Oclusión	No

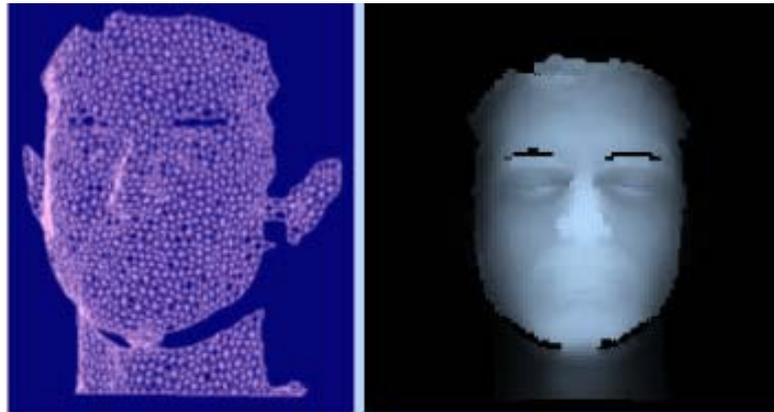


FIGURA 3.17: Mallado 3D y mapa de profundidad

En el presente trabajo se utilizaron las bases de imágenes FEI y 2D de FRAV3D, ya que ambas permiten estudiar el influjo de la pose en el reconocimiento de rostros, tema central del presente trabajo, y otras características importantes como la iluminación y expresión. De igual manera, estas dos bases de imágenes presentan fondos homogéneos, lo que permite trabajar directamente con el rostro y delegar a otro trabajo de investigación la detección facial en ambientes naturales.

A continuación se detallan las características de FEI y FRAV3D.

3.1.1. FRAV3D

FRAV proporciona una opción multimodal 2D (textura) y 3D (estructura), adquirida a través de un escáner láser de Minolta, modelo VIVID-700, con el software Polygon Editing Tool v 1.11 (PET). El escáner proporciona información de textura y 3D. Esta última se presenta a través de un mallado triangular en formato VRML (3D) y en una imagen de profundidad (2.5D). Ver Figura 3.17.

Las imágenes se adquirieron en condiciones controladas de iluminación (2 focos halógenos de luz difusa con la cara del individuo situada entre ambos haces de luz) y no controlada (2 focos halógenos de luz difusa más iluminación cenital de una lámpara fluorescente). En la Figura 3.18 se observa la posición de las luces y del sujeto así como el fondo que se eligió para facilitar la segmentación de la cara. La distancia entre el individuo y el escáner se fijó en 2 metros como lo muestra la Figura 3.19.

La base de datos consta de imágenes de 105 individuos (81 mujeres y 24 hombres), adultos de raza caucásica. El período de adquisición duró 10 meses (septiembre del 2004 -



FIGURA 3.18: Mallado 3D y mapa de profundidad

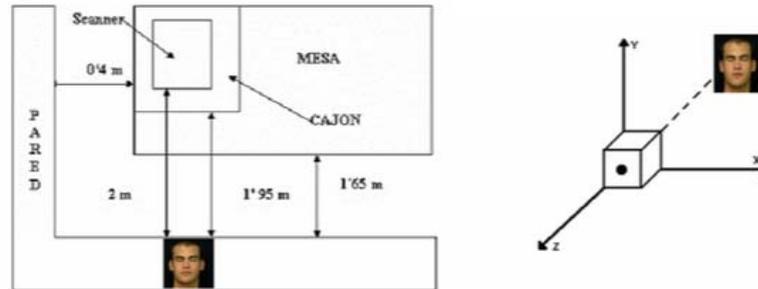


FIGURA 3.19: Esquema de colocación del escáner y su colocación a la izquierda y derecha, respectivamente.

junio del 2005) y cada individuo participó en una sesión completa de adquisición. En todas las imágenes los sujetos tienen los ojos cerrados y no se permitió el uso artefactos como sombreros, gafas u otros accesorios que alteraran el aspecto facial.

Existen 16 capturas por individuo (cada una contiene información de textura y tridimensionalidad) y se siguió un protocolo de adquisición de imágenes que abarca aspectos como nomenclatura de las imágenes, orden de adquisición, configuración del escáner, condiciones de adquisición, entre otros. El orden de adquisición de imágenes se muestra en el cuadro 3.2.

Nomenclatura	Tipo de imagen
01-04	Frontal
05-06	Giro derecho (25°)
07-08	Giro izquierdo(5°)
09	Giro severo a la derecha
10	Suave giro a la izquierda
11	Sonrisa
12	Boca abierta
13	Giro arriba en X
14	Giro abajo en X
15-16	Iluminación frontal cenital

TABLA 3.2: Orden de adquisición de imágenes (nomenclatura) y descripción

La resolución de las imágenes es de 400x400 píxeles en el caso de las imágenes de textura y de 200x200 píxeles en el de las imágenes de rango. La base de datos FRAV3D disponible para la comunidad científica (sólo para fines de investigación, no comerciales) por lo que se requiere solicitarla vía correo electrónico perteneciente a una entidad educativa.

3.1.2. FEI

Esta base se creó en el laboratorio de Inteligencia Artificial de la Facultad de Ingeniería Industrial a través de la participación de 200 individuos a los cuales se les hicieron 14 tomas diferentes. Los participantes pertenecen principalmente al estudiantado y staff de FEI por lo que sus edades oscilan entre los 19 y 40 años. Algunos de ellos presentan adornos como lentes. El número de hombres y mujeres es de 100 para cada uno.

La base de datos FEI contiene 14 variantes por cada individuo resultando en un total de 2800 imágenes en la base de datos [40–42]. Todas las imágenes son a color y se tomaron con un fondo homogéneo. Su tamaño es de 640X480 píxeles.

Esta base no sólo tiene como objetivo la alineación de las imágenes capturadas sino también generar una base de datos de rostros con imágenes pre-procesadas. La base contiene 800 diferentes imágenes faciales, 200 de frente, 200 sonriendo y 200 con giros de 10° a la derecha e izquierda (Ver Figura 3.21). El protocolo que se siguió en la adquisición de las tomas se presenta en la Figura 3.20.

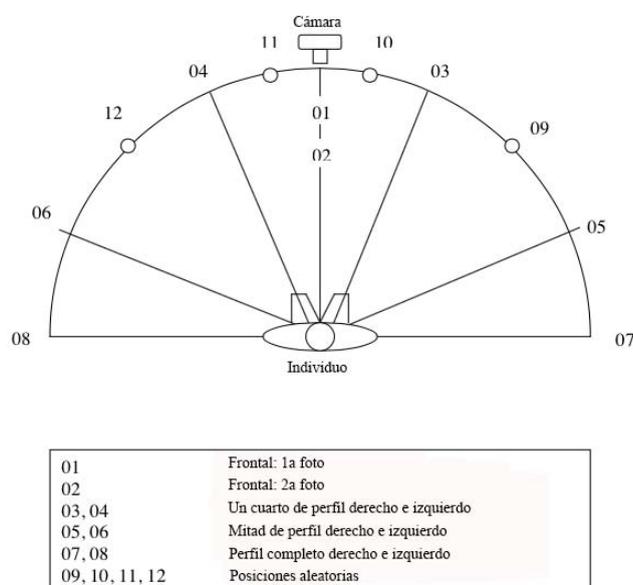


FIGURA 3.20: Esquema de colocación de la cámara



FIGURA 3.21: De izquierda a derecha se presentan ejemplos de toma frontal con expresión neutral, frontal sonriendo, giro a la derecha, giro a la izquierda.

Cabe destacar que esta base de datos presenta la oportunidad de trabajar con distintas razas ya que algunos de los participantes tienen rasgos orientales, piel oscura y piel clara.

Capítulo 4

Clasificador Neuronal de Codificación por Permutación (PCNC)

4.1. Introducción

La automatización del reconocimiento de rostros puede considerarse un problema multiclase donde el número de clases (rostros) equivale al número de sujetos registrados. Esto conlleva a la realización de dos tareas críticas, la representación de una clase a través de la extracción de las características que la definen y la construcción de un clasificador que las identifique [6].

El presente capítulo describe el funcionamiento del Clasificador Neuronal de Codificación por Permutación, PCNC. Este es un algoritmo heurístico que provee una forma de representar las características de una clase, en este caso el rostro de un sujeto, y lleva a cabo un proceso de clasificación para su reconocimiento o identificación. PCNC utiliza la teoría detrás de las redes neuronales de Rosenblatt [43], paradigma que permite crear una estructura generalizada de diferentes clasificadores, para crear un descriptor no global sino local.

En la Fig. 4.1 se presentan los procesos que ejecuta el clasificador PCNC, los cuales tienen como objetivo la identificación de un rostro a partir de su comparación con imágenes faciales registradas en una base de datos.

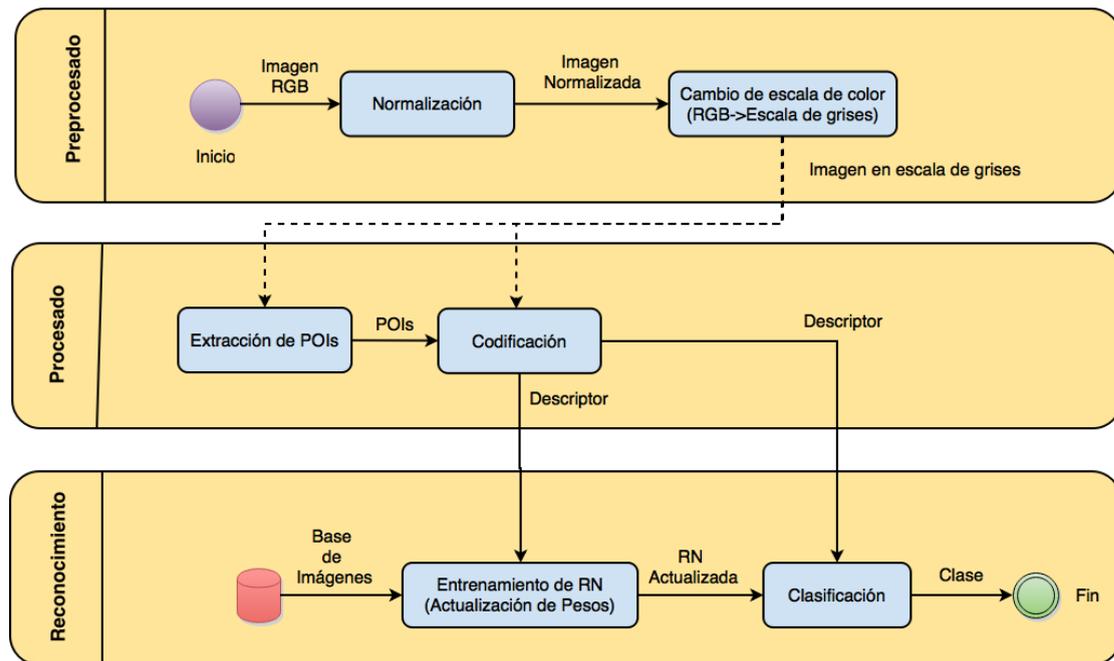


FIGURA 4.1: Procesos de PCNC.

Basándose en el artículo de Yip [44], la luminosidad tiene un papel de gran importancia en el reconocimiento de rostros ya que permite extraer información de formas; no obstante, cuando los archivos presentan una evidente disminución en la calidad de la imagen, las ventajas del color se hacen notorias al facilitar tareas de análisis de bajo nivel como la segmentación sobre el propio rostro. La distribución del color puede entonces complementar la información de luminosidad y permitir una mejor estimación de los límites, formas y tamaños de atributos faciales, como por ejemplo, la segmentación de ojos y cabello. Considerando que las imágenes de entrada están centradas, normalizadas, son nítidas y que el sistema trabaja con contornos, el procesamiento de color proporciona escasa información por lo que las imágenes se convierten directamente a escala de grises. Esto no sólo disminuye las operaciones a realizar sino también la cantidad de memoria a utilizar, lo que hace más eficiente la ejecución del programa. Cabe destacar que la normalización aquí hace referencia a la eliminación de fondos y centrado del rostro en la imagen independientemente de su orientación.

El cambio de escala de color de cada pixel, por su parte, calcula los valores de luminosidad a partir del promedio de los valores de cada canal como lo establece la ecuación

4.1.

$$f(x, y) = (R + G + B)/3. \quad (4.1)$$

Una vez pre-procesada la imagen, ésta sigue un conjunto de acciones que permiten la extracción de características y consiguiente construcción del descriptor. El descriptor funciona como la entrada necesaria para la actualización de los pesos en la red neuronal (entrenamiento) y posterior clasificación de una imagen a través de la red ya actualizada.

4.2. Descriptores

Los descriptores son herramientas que se encargan de extraer y procesar las características que identifican una imagen de manera independiente al patrón de referencia y región tratada, para posteriormente realizar una correspondencia con otras imágenes.

Existen una gran cantidad de descriptores y medidas de distancia asociadas que enfatizan diferentes propiedades de la imagen como intensidad del pixel, color, textura, contornos, por mencionar algunas. De acuerdo con la región que describe, un descriptor puede ser clasificado como global (se aplica a la imagen como un todo) o local (describe una región dentro de la imagen).

La idea de descriptores globales fue propuesta por Frank Rosenblatt [45]. En su perceptrón de tres capas, cada neurona de la capa asociativa desempeña un papel de descriptor aleatorio de la imagen. Aunque existen propuestas alternativas de descriptores, fue la idea de Rosenblatt la que se tomó como base y modificó para aplicarla localmente a un área específica de la imagen en vez de a toda ella.

Hoy en día el descriptor local ha evolucionado y, según el método utilizado, puede clasificarse como descriptor difuso (*sparse descriptor*) o denso (*dense descriptor*). Los descriptores difusos filtran puntos de interés en una imagen para después generar zonas de extracción alrededor de estos y proceder a la codificación. Los descriptores densos, por otra parte, extraen características pixel por pixel de la imagen de entrada.

PCNC utiliza descriptores locales difusos, y se basa en la teoría que describe los descriptores locales aleatorios o RLD (Random Local Descriptors) por sus siglas en inglés

para generar su vector característico. En la Figura 4.2 se muestra cómo intervienen los descriptores en la estructura del sistema de reconocimiento de propósito general.

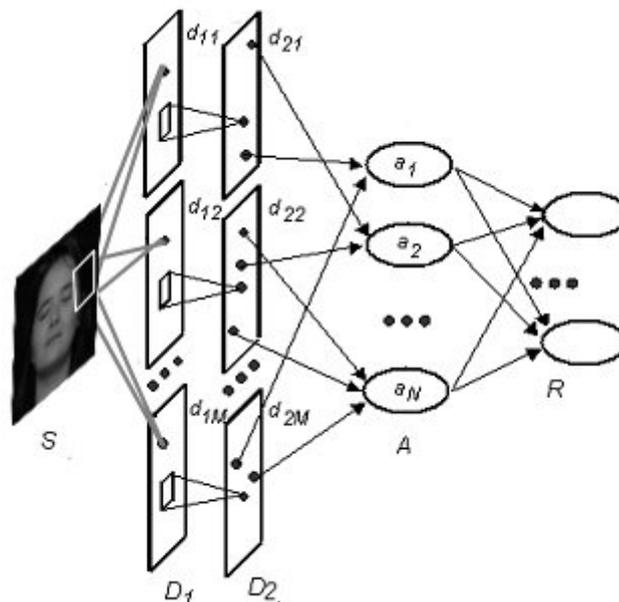


FIGURA 4.2: Esquema del sistema de reconocimiento de propósito general.

La primera capa del clasificador o capa de sensado corresponde a la imagen; la segunda capa $D1$ contiene neuronas RLD de bajo nivel o neuronas AND, la tercera capa $D2$ contiene neuronas RLD de más alto nivel o neuronas OR. La capa R contiene las neuronas de salida donde cada neurona corresponde a la clase a reconocer.

4.2.1. RLD

Como ya se mencionó, el descriptor RLD se conecta con las neuronas de la capa asociativa a un punto de la imagen de entrada y calcula una función de brillo en ese punto para generar puntos de interés que permitan evaluar la existencia o inexistencia de una característica. Una vez se haya establecido la presencia de características dentro de la imagen y generado su descriptor, se establecen conexiones entre capas, las cuales permanecerán estáticas durante todo el proceso. En la Figura 4.3, se muestra la estructura del descriptor RLD.

De forma general RLD funciona de la siguiente manera: Para cada punto de interés C de la imagen, representado por un pixel, se genera un rectángulo de dimensiones $w \times h$

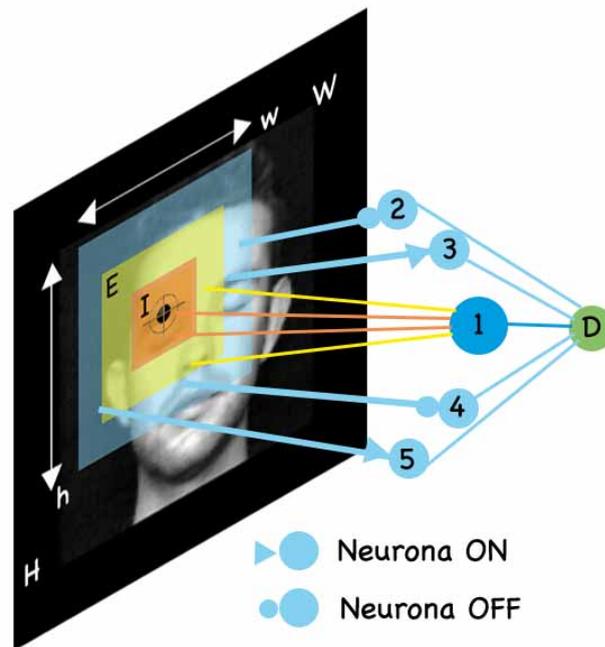


FIGURA 4.3: Estructura del RLD.

con centro en C . En el interior del rectángulo se distribuyen aleatoriamente un número fijo de puntos P_r cuyo umbral T_r se selecciona aleatoriamente del intervalo $T_{min} \leq T_{rs} \leq T_{max}$. Donde T_{min} y T_{max} representan el brillo máximo y mínimo de la imagen global respectivamente. Cada uno de esos puntos es analizado posteriormente por neuronas de tipo ON (positivas) y OFF (negativas).

Las neuronas ON se activan cuando el valor del brillo b_r del punto P_r es mayor que el umbral T_r de la neurona ($b_r \geq T_s$).

Las neuronas OFF se activan cuando el valor de brillo b_r del punto P_r es menor que el umbral T_r de la neurona ($b_r < T_r$).

Las neuronas ON y OFF son neuronas simples pero también existen neuronas complejas dentro de la estructura del RLD. En la Fig. 4.3 la neurona compleja lleva el número 1 y se caracteriza por tener conexiones excitatorias con todos los píxeles que pertenecen al rectángulo pequeño (Rectángulo E) e inhibitorias con los píxeles fuera del rectángulo E pero dentro del I . El filtrado que realiza esta neurona se determina a través de los pesos de las conexiones y sólo considera aquellos puntos cuyo gradiente de brillo sea mayor que el umbral preestablecido. Normalmente, las conexiones excitatorias tienen pesos +1 y las inhibitorias de -1. Para el reconocimiento de rostros, las conexiones excitatorias

deben ser inversamente proporcionales al área del rectángulo E y las inhibitorias deben ser inversamente proporcionales al área del rectángulo I .

Se considera que la propiedad o característica existe sólo si todas las neuronas de la capa $D1$ se activan por lo que la neurona D también recibe el nombre de neurona AND. Validada la característica, se envía en código binario al vector de características, el cual funcionará como valor de entrada de la red neuronal.

La topología de conexiones entre la capa de sensado y las neuronas 2 – 5 se mantiene fija para todos los planos D . Donde cada plano D representa una característica a detectar en toda la imagen y, por lo tanto, su número equivale al número de características presentes en la imagen.

4.3. Extractor de Propiedades

En el extractor de propiedades se generan los puntos de interés por medio de los cuales se evalúa la existencia de una característica. La extracción de los puntos de interés se realiza a través de la aplicación de un filtro lineal sobre la imagen en escala de grises. Lo que se busca con este filtro es obtener los puntos que mejor describan las características de un rostro. Para ello utiliza la diferencia del promedio de luminancia de un kernel A de tamaño m menos el promedio de luminancia de un kernel B de tamaño n centrado en A alrededor de un pixel Px , donde $n < m$. El valor de la diferencia es comparado con un umbral que permite evaluar si el pixel proporciona o no información y binarizar la imagen. El resultado se muestra en la Figura 4.4.



FIGURA 4.4: Resultado de la aplicación del filtro lineal.

Para explicar el proceso que le sigue a la extracción de puntos de interés es necesario comprender que una misma característica puede encontrarse en diferentes zonas o posiciones de la imagen, por ejemplo, las líneas en los labios aunque aparezcan en diferentes zonas de la cara pueden considerarse la misma característica y descartarse su evaluación en ambos perfiles. Véase la Figura 4.5.



FIGURA 4.5: Repetición de características en diferentes posiciones de la imagen. Los cuadros pequeños son regiones extraídas todas de diferentes zonas de los labios.

Con base en el funcionamiento de los descriptores RLD, la existencia local de una característica F_i dentro de una imagen se determina escaneándola a través una ventana rectangular de tamaño $w \times h$ con centro en un punto de interés C ; donde $w \ll W$ y $h \ll H$ siendo W y H el ancho y alto de la imagen en escala de grises. La posición del C es igual a la posición de los puntos de interés P_{ji} obtenidos por el filtrado de contornos.

Definimos entonces una característica F_i en la posición P_{ji} , como un conjunto de neuronas ON y OFF conectadas aleatoriamente a un pixel dentro de una ventana $w \times h$, de forma formal:

$$F_i = (Px_{ji}, Py_{ji}, \cap_1^C(ON_k, Th_{ON_k}), \cap_1^l(OFF_l, Th_{OFF_l})), \quad (4.2)$$

donde Px_{ji} y Py_{ji} son las coordenadas en X y Y del punto de interés o centro de la ventana, $\cap_1^C(ON_C, Th_{ON_C})$ es la función de conjunción de las k neuronas ON, ON_k es la posición de las neuronas ON generada aleatoriamente dentro de la ventana cuyos umbrales de luminancia están definidos aleatoriamente por Th_{ON_k} . De forma análoga, $\cap_1^l(OFF_l, Th_{OFF_l})$ es la función de conjunción de las l neuronas OFF, OFF_l es la posición de las neuronas OFF dentro de la ventana y Th_{OFF_l} su umbral.

Para evaluar la existencia de una característica se recorre la ventana sobre la imagen original (escala de grises) y posiciona su centro sólo donde exista un punto de interés, esto significa que toma como referencia la imagen filtrada. Si la conjunción de las neuronas ON y OFF es 1 para cada una en esa posición (ecuación 4.2), entonces la característica F_i existe, de lo contrario, se continúa con el escaneo de la imagen.

Para cada característica F_i extraída, el codificador crea un vector binario auxiliar o máscara de dimensión N con K unos cuya posición inicial se elige aleatoriamente y $K \ll N$. En

este trabajo se estableció a $k=16$ y $N=64000$. Así entonces el vector auxiliar queda definido como $U_i = (u_{i1}, u_{i2}, \dots, u_{iN})$ y corresponde a la posición inicial de la característica en la imagen. Si la característica se encuentra presente en otra posición, entonces se lleva a cabo una permutación de la máscara como se describirá posteriormente. Cada vez que una característica se presenta en una región de la imagen, esta se almacena dentro del vector de características aplicando una operación de disyunción.

Sea entonces Z diferentes posiciones en las que se encuentra una misma característica, su código queda definido por un vector binario U_i tal que:

$$U_i = \bigcap_{z=1}^Z U_z P_{zi} \quad (4.3)$$

Donde U_z es el vector auxiliar y P_{zi} su i -ésima permutación; por lo tanto $U_z P_{zi}$ es el resultado de la permutación del vector auxiliar.

Ya que U_i es de dimensión N y N es un valor muy grande, sólo se almacenan la posición de los unos del vector en una lista y se descartan los ceros.

4.4. Codificación

El problema de generar códigos binarios de las propiedades es lograr que la correlación entre dos características sea inversamente proporcional a su distancia. Las permutaciones no sólo permiten generar un código único para cada característica en la posición donde se localice sino que también proporcionan una forma de analizar su correlación.

Para codificar la localización de una característica F_i es necesario definir la distancia de correlación Dc ; esto es, la distancia entre propiedades como medida de similitud (el valor de este parámetro en el trabajo se estableció experimentalmente y es de 6 píxeles).

Sean los puntos $P_1(x_1, y_1)$ y $P_2(x_2, y_2)$ cuya característica es F_i , U'_{p1} , U'_{p2} los vectores binarios que codifican la característica para cada punto, dx la distancia euclidiana de P_1 y P_2 en X , y dy la distancia euclidiana de P_1 y P_2 en Y .

$$dx = \text{abs}(x_1 - x_2); dy = \text{abs}(y_1 - y_2) \quad (4.4)$$

Los vectores U'_{p1}, U'_{p2} están correlacionados en X si $dx < D_c$ y en Y si $dy < D_c$; en caso contrario, no hay correlación.

4.4.1. Permutación

La codificación se lleva a cabo a través de permutaciones en las direcciones X y Y . Esto es, la aplicación de P_m permutaciones en el vector U resulta en un nuevo vector U' cuyas componentes son las mismas que U pero en una posición permutada. En términos formales:

$$Pm(U) = U' \mid U' = \{u'_i\}, u'_i = u_{pi}, P, U, U' \in \mathbf{R}^m \quad (4.5)$$

El proceso de permutación retoma la constante definida como distancia de correlación D_c para calcular los valores de las ecuaciones 4.6,4.7. Este paso permite generar las coordenadas del punto de interés en función de D_c .

$$X = \frac{j}{D_c}, \quad (4.6)$$

$$Y = \frac{i}{D_c}, \quad (4.7)$$

Calculados los nuevos valores de X y Y a partir de las coordenada en i, j y la distancia de correlación D_c , se procede a encontrar el número de permutaciones completas respecto a cada eje con 4.8, 4.9.

$$E(X) = (int)X, \quad (4.8)$$

$$E(Y) = (int)Y. \quad (4.9)$$

$E(X)$ y $E(Y)$ indican la cantidad de veces que se permutará un vector completo. Para estimar el número de permutaciones parciales, primero se procesa la parte fraccionaria de X y Y como se muestra en 4.10, 4.11,

$$R(X) = j - E(X) \times D_c, \quad (4.10)$$

$$R(Y) = i - E(Y) \times D_c, \quad (4.11)$$

luego se obtiene el número de permutaciones parciales $P(X)$, $P(Y)$ con ayuda de los valores calculados y las ecuaciones 4.12, 4.13. Las permutaciones parciales no se aplican al vector completo sino sólo a los primeros $P(X)$ o $P(Y)$ elementos que lo componen.

$$P(X) = \frac{(int)(R(X)) \times N}{D_c}. \quad (4.12)$$

$$P(Y) = \frac{(int)(R(Y)) \times N}{D_c}. \quad (4.13)$$

Donde N indica el número de neuronas que componen el sistema. Ya que el modelo es generalizado, el valor de N puede cambiar de acuerdo a la complejidad y requerimientos del problema. En el caso del reconocimiento de rostros se emplearon 65,000 neuronas.

Las permutaciones son secuenciales y la técnica inicializa con el desplazamiento horizontal. Esto es, al vector U_k se le realizan $E(X)$ permutaciones y, en caso de que $P(x) \neq 0$, se concluye con una permutación más pero sólo a los primeros $P(X)$ elementos de U_k . Una vez finalizada la codificación en X , se procede con la de Y . La Fig.4.6 muestra un ejemplo gráfico de este proceso para x donde $E(X) = 2, P(X) = 2$.

4.4.2. Análisis de Correlación

Sean $P_1(x_1, y_1)$ y $P_2(x_2, y_2)$ dos puntos localizados en diferentes lugares de la imagen donde la misma característica F_i ha sido detectada y cuya máscara es U_{i1} y U_{i2} para P_1 y P_2 respectivamente, y por lo tanto $U_{i1} \neq U_{i2}$. Las distancias entre P_1 y P_2 se definen en 4.14 y 4.15.

$$dx = |x_2 - x_1| \quad (4.14)$$

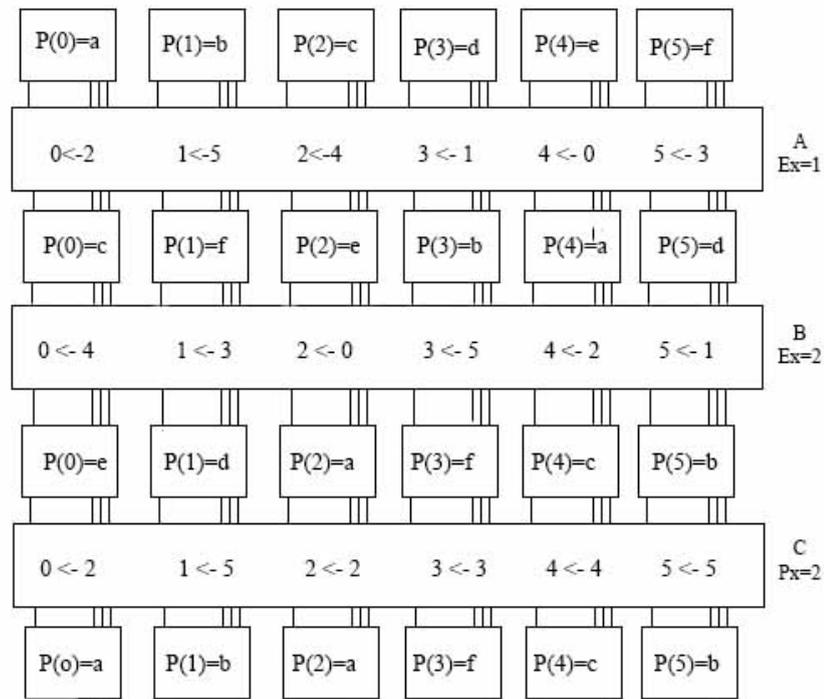


FIGURA 4.6: Permutación del vector U.

$$dy = |y_2 - y_1| \tag{4.15}$$

Sea Δn la diferencia en el número de 1's entre los vectores cuyo valor promedio en X puede calcularse de forma aproximada con 4.16:

$$\Delta n_x \approx \frac{K}{D_c} dx, \tag{4.16}$$

donde k es el número de 1's del vector auxiliar binario U de la característica F_i . Luego de las permutaciones verticales, las diferencias de los vectores U_{i1} y U_{i2} pueden estimarse por 4.17:

$$\Delta n_x = 1 - \frac{dx}{D_c} \quad \Delta n_y = 1 - \frac{dy}{D_c} \quad \Delta n \approx k(1 - \Delta n_x \Delta n_y), 1 > \Delta n > 0 \tag{4.17}$$

Lo cual señala que $\Delta n \rightarrow 0$ cuando $dx, dy \rightarrow 0$, lo cual obliga a mantener valores de $dx, dy < D_c$ para conservar la correlación. Por otra parte, la necesidad de que $k \ll N$ y de que existan permutaciones parciales previene la duplicidad de posiciones de un

mismo elemento en el vector al disminuir la probabilidad de que dicho evento suceda a menos del 0.01 %. Esto es porque la probabilidad de que tales casos indeseables sucedan está inversamente relacionada con las dimensiones de N y su relación con k (k/N).

Capítulo 5

Algoritmo PCNC con distorsiones tipo rotación

Esta sección expone primero la problemática a resolver; después la propuesta para resolverlo y, finalmente, cómo se aplicó el proceso elegido dejando para el siguiente capítulo los resultados experimentales. Esto con el fin de apegarse al método científico que dicta un orden cronológico de pasos a resolver, los cuales son:

1. Observación y descripción del fenómeno
2. Formulación e hipótesis que explican el fenómeno.
3. Uso de la hipótesis para predecir cuantitativamente los resultados de las nuevas observaciones
4. Realización de pruebas experimentales

5.1. Problemática

PCNC utiliza técnicas de permutación de código para el procesamiento de la imagen y una red neuronal para el reconocimiento. Para extraer los puntos de interés, PCNC aplica un filtro lineal a la imagen en escala de grises y la transforma en una imagen binarizada (Figura 5.1). Los puntos no nulos de la imagen binarizada (puntos de interés) sirven para evaluar la existencia de una característica en la imagen en escala de grises.



FIGURA 5.1: Flujo para la creación del vector de características de una imagen.

Esto es, cada punto de interés proporciona las coordenadas donde se centrará la ventana con neuronas ON y OFF, distribuidas aleatoriamente, en la imagen en escala de grises. A través de las neuronas de esta ventana se evalúa la existencia de una característica (Figura 5.1).

Hasta este punto el funcionamiento de PCNC es igual al del clasificador LIRA (Limited Image Receptive Area) [8]. La desventaja de este último radica en su sensibilidad a desplazamientos, escalas y rotaciones, características comunes en las imágenes de rostros. Como alternativa a LIRA, PCNC incluye como parte de la codificación a la permutación. La primera vez que PCNC fue utilizado en la tarea de reconocimiento de rostros, fue con la base de imágenes ORL como se muestra en [11, 46]. Las tasas de error que se obtuvieron de estas pruebas fueron significativamente buenas; no obstante, al trabajar con la base de imágenes 2D de FRAV3D en [12] se obtuvieron nuevos resultados que mostraban un desempeño pobre en el reconocimiento de rostros rotados como se muestra en la Tabla 5.1 .

5.1.1. Hipótesis

Basándose en [12] donde se agregaron distorsiones de desplazamiento para disminuir la tasa de error de reconocimiento, se planteó la adición de imágenes rotadas al set de entrenamiento como una posible solución del problema. Esto conllevó a la necesidad de conocer los ángulos clave para generar las distorsiones. Para ello se analizó la inclinación que puede alcanzar la cabeza al ser rotada según la Academia Americana de Cirujanos Ortopédicos (Figura 5.2). La Tabla 5.2 muestra los ángulos máximos alcanzados por un humano sobre el eje X , Y y Z , respectivamente.

La hipótesis planteada es alimentar el set de entrenamiento con imágenes sintetizadas de rotaciones clave dentro de un intervalo establecido por la capacidad humana para girar la cabeza. La generación de estas imágenes sintetizadas se detalla a continuación.

No.	Tipo de imagen		% de Error		
	Set de Prueba	Set de Entrenamiento	PCNC 1 dist	PCNC 12 dist	SVM
1	frontal	frontal	3.92	0.78	1.94
2	frontal	gestos (sonrisa)	11.76	1.94	5.13
3	frontal	gestos (boca abierta)	22.55	1.94	10.68
4	frontal	iluminación	19.61	2.45	8.70
5	frontal	rotación Y 5°	35.78	3.5	14.60
6	frontal	rotación pequeña Z	66.67	33.7	33.93
7	frontal	rotación X	42.65	16.4	12.62
8	frontal	rotación Y 25°	55.88	19.0	27.02
9	frontal	rotación grande Z	79.41	54.3	41.14
10	frontal, iluminación	frontal	2.94	0.59	1.94
11	frontal, rotación Y 5°	frontal, rotación Y 5°	5.88	0.39	4.85
12	frontal, rotación Y 5°, iluminación	frontal, rotación Y 5°,1 iluminación	3.92	0.33	4.17
13	frontal, escala, rotación	frontal, rotación Y, gestos	5.64	0.44	4.09
14	frontal	gestos	16.67	1.4	8.90
15	frontal	frontal	10.78	0.98	1.46

TABLA 5.1: Resultados de prueba de PCNC y SVM con FRAV2D

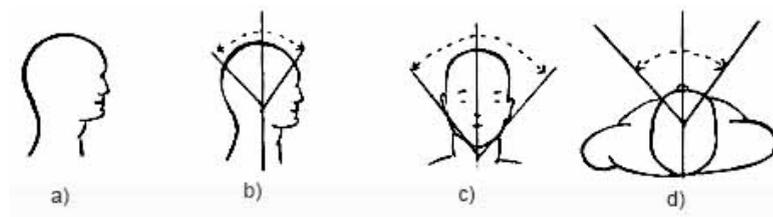


FIGURA 5.2: Rotaciones del rostro. La imagen a) muestra la posición inicial, b) las rotaciones sobre X (flexión y extensión), c) sobre el eje Z (inclinaciones) y d) sobre Y.

5.2. Rotación

La sintetización de las imágenes rotadas se creó a partir de la primera imagen frontal de cada clase. Para establecer las ecuaciones de rotación partimos de considerar un pixel A de coordenadas x, y sobre la imagen original. Ver Figura 5.3.

El cálculo de la nueva posición x', y' del pixel, dado un ángulo θ resulta de aplicar las expresiones geométricas en 5.1,

Eje	20-29 años	30-40 años	>50 años	Promedio
Z	$50.5^\circ \pm 5.5^\circ$	$46.5^\circ \pm 6.5^\circ$	$40^\circ \pm 8.5^\circ$	45°
X(adelante)	$75.5^\circ \pm 8.5^\circ$	$70.5^\circ \pm 17.5^\circ$	$64.5^\circ \pm 7^\circ$	60°
X(atrás)	$75.5^\circ \pm 8.5^\circ$	$70.5^\circ \pm 17.5^\circ$	$64.5^\circ \pm 7^\circ$	75°
Y	$91.5^\circ \pm 5.5^\circ$	$81.5^\circ \pm 6.5^\circ$	$77.5^\circ \pm 7.5^\circ$	80°

TABLA 5.2: Libertad de giro de la cabeza.

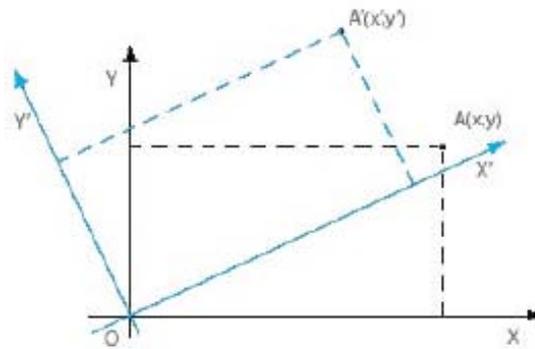


FIGURA 5.3: Rotación de una imagen sobre el origen.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\text{sen}\theta \\ \text{sen}\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (5.1)$$

Resolviendo el sistema en 5.1 obtenemos las ecuaciones 5.2.

$$\begin{aligned} x' &= x\cos\theta - y\text{sen}\theta \\ y' &= x\text{sen}\theta + y\cos\theta \end{aligned} \quad (5.2)$$

El resultado de aplicar estas ecuaciones en las imágenes frontales se muestra en la Figura 5.4. Como ahí se observa hay pérdida de información conocida como aliasing o sub-muestreo. Esto se debe a que el cálculo de senos y cosenos da como resultado números reales mientras que el manejo de píxeles se da a través de localidades cuantizadas con enteros, por lo tanto, los resultados deben redondearse para poder ser dibujados.

La forma de solucionar el sub-muestreo dependía de cómo se ingresaran las distorsiones al entrenamiento. Existían dos alternativas: la primera era rotar imágenes frontales y agregarlas a la base de rostros, y la segunda era utilizar un módulo que durante el proceso de entrenamiento rotara cada imagen de entrada y la añadiera como una distorsión más.

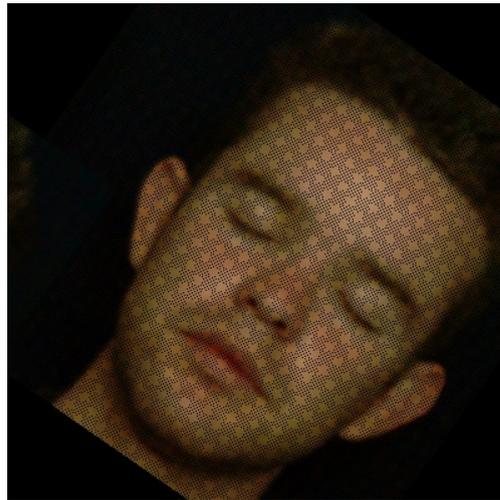


FIGURA 5.4: Submuestreo en rotaciones.

Con la primera alternativa, la cantidad de imágenes a procesar ascendía al producto de las S clases por la suma del total de muestras por clases a ingresar M más las imágenes rotadas por clase R . Además, a cada imagen rotada se le agregaban D distorsiones de desplazamiento, lo que resultaba en un total de $S \times (M + R) \times D$ imágenes a procesar. Con la segunda alternativa, se incrementaba el número de imágenes a un total de $S \times M \times (D + R)$ y las distorsiones de desplazamiento eran independientes de las de rotación.

La generación de imágenes rotadas durante el proceso de entrenamiento requería de un algoritmo ágil para no ralentizar el proceso, pero se perdía calidad en la imagen. El crear las imágenes anticipadamente permitía utilizar cualquier algoritmo de interpolación sin perder calidad pero incrementaba el número de imágenes a procesar. Por tal motivo, los experimentos se realizaron bajo las dos modalidades, antes y durante el entrenamiento, uno con una interpolación lineal (antes del entrenamiento) y otro con sesgo o "*shearing*".^{en inglés} [47] (durante el entrenamiento).

Un ejemplo de las imágenes generadas antes del entrenamiento se muestran en 5.5a, y de las generadas durante el proceso se muestran en 5.5b.

En el capítulo siguiente se describen los experimentos realizados y los resultados obtenidos de la aplicación de distorsiones tipo rotación antes y durante el entrenamiento con las imágenes 2D de la bases de datos FRAV3D. Adicionalmente se exponen los resultados de experimentos con la base de datos FEI siguiendo el mismo guión experimental que con las imágenes 2D de FRAV3D pero sin distorsiones tipo rotación sobre Z .



(A) Rotación con interpolación.

(B) Rotación con shearing.

FIGURA 5.5: Rotaciones con dos diferentes métodos.

Capítulo 6

Experimentos y Resultados

En este capítulo se detallan los experimentos realizados con PCNC sobre las imágenes 2D de la base de datos FRAV3D y FEI. Por la naturaleza de las base de datos, a FEI no se le generaron distorsiones de tipo rotación sobre Z , aunque sí se le aplicaron distorsiones de tipo desplazamiento. A continuación se detalla el orden de los experimentos llevados a cabo y las distorsiones trabajadas.

6.1. Distorsiones

Para aumentar el índice de reconocimiento se entrenó la red neuronal con imágenes que consideraran distorsiones de desplazamiento y rotación. Para el caso de las imágenes 2D de FRAV3D, se generaron imágenes con rotaciones sobre Z a partir de la primera imagen frontal de cada sujeto y se agregaron al set de entrenamiento, no así para FEI cuyas características hacían innecesaria la rotación sobre el eje Z ya que se enfoca en el estudio de rotaciones sobre el eje Y .

Es importante aclarar que la inclusión de imágenes rotadas en FRAV3D se hizo de dos formas diferentes, una como parte del set de entrenamiento y otra durante el entrenamiento. Las distorsiones de desplazamiento, por otro lado, siempre se incluyeron durante el entrenamiento tanto para FRAV3D como para FEI. La descripción de las rotaciones generadas sobre el eje Z se presenta en la Tabla 6.1 y la de desplazamiento en la Tabla 6.2.

Caso	Distorsión
R1	Rotación a la derecha 15°
R2	Rotación a la derecha 25°
R3	Rotación a la derecha 35°
R4	Rotación a la derecha 45°
R5	Rotación a la izquierda 15°
R6	Rotación a la izquierda 25°
R7	Rotación a la izquierda 35°
R8	Rotación a la izquierda 45°

TABLA 6.1: Tabla de rotaciones.

Caso	Distorsión
D0	Posición inicial
D1	Δ pixeles a la izquierda
D2	Δ pixeles a la derecha
D3	Δ pixeles hacia abajo
D4	Δ pixeles hacia arriba
D5	Δ pixeles hacia arriba y Δ pixeles hacia la derecha
D6	Δ pixeles hacia arriba y Δ pixeles hacia la izquierda
D7	Δ pixeles hacia abajo y Δ pixeles hacia la derecha
D8	Δ pixeles hacia abajo y Δ pixeles hacia la izquierda

TABLA 6.2: Descripción de desplazamientos

Las distorsiones por desplazamiento consideran una traslación de Δ número de pixeles de la imagen. Donde Δ , es un parámetro que puede variar según se requiera. En este trabajo, por ejemplo, el valor asignado a Δ para ambas bases de imágenes fue de 6 pixeles. El resultado de las rotaciones puede verse en la Figura 6.1.

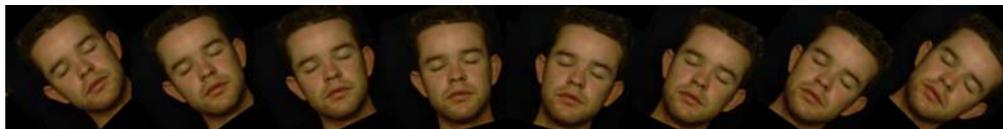


FIGURA 6.1: Rotación de la imagen 1 de la base de datos FRAV3D

En los experimentos realizados se consideraron tres grupos de desplazamiento, el primer grupo no contiene desplazamiento (D0), el segundo incluye la posición inicial y cuatro desplazamientos (D0-D4), las tercera incluye todos los desplazamientos. La siguiente sección describe cómo se aplicaron estas distorsiones en los experimentos.

6.2. Desarrollo de experimentos

Los experimentos se realizaron en dos ordenadores diferentes, uno con procesador Intel Core i7 870 a 2.93 Ghz con 4Gb de memoria RAM y sistema operativo Windows 8.1 (equipo A) y otro en una máquina virtual con 2GB de memoria y Windows XP (equipo B) instalada en una máquina con SO MacOS X Yosemite con 8GB en memoria RAM y procesador intel Core i7 a 2.8GHz. El IDE seleccionado fue Visual Studio 2013 para el equipo A y 2010 para el B. En el equipo A se trabajó la base de imágenes 2D de FRAV3D y en el equipo B, FEI.

Cada base de imágenes permite estudiar diferentes tipos de rotación. El set 2D de FRAV3D, por ejemplo, contiene imágenes donde la cabeza presenta inclinaciones a la izquierda y derecha (rotación sobre Z), mientras que en FEI existen imágenes de perfiles parciales y totales (rotación sobre Y). La descripción de las tomas obtenidas de cada sujeto para cada base de imágenes se encuentran en la Tabla 6.3 y 6.4.

Toma	Descripción
1	Frontal
2	Frontal
3	Frontal
4	Frontal
5	Giro 25° a la derecha
6	Giro 25° a la derecha
7	Giro 5° a la izquierda
8	Giro 5° a la izquierda
9	Inclinación a la derecha
10	Inclinación a la izquierda
11	Frontal sonriendo
12	Frontal boca abierta
13	Giro hacia arriba sobre X
14	Giro hacia abajo sobre X
15	Iluminación frontal cenital
16	Iluminación frontal cenital

TABLA 6.3: Descripción de las imágenes de la base de datos 2D de FRAV3D.

Para poder entrenar la red y probarla, ambas bases de datos fueron divididas en dos sets, uno de entrenamiento y otro de prueba. La división se hizo de acuerdo al tipo de distorsión o caso que se quería estudiar (FEI-rotaciones sobre Y, FRAV3D-rotaciones sobre Z). A su vez, los experimentos se dividieron en tres fases: la primera comprendía la

Toma	Descripción
1	Perfil derecho
2	Giro 40° a la derecha
3	Giro 30° a la derecha
4	Giro 20° a la derecha
5	Giro 10° a la derecha
6	Giro 10° a la izquierda
7	Giro 20° a la izquierda
8	Giro 30° a la izquierda
9	Giro 40° a la izquierda
10	Perfil izquierdo
11	Frontal
12	Frontal sonriendo
13	Frontal con variación de iluminación
14	Frontal con variación de iluminación

TABLA 6.4: Descripción de las imágenes de la base de datos FEI.

inclusión de distorsiones en el set de entrenamiento, la segunda consideraba las rotaciones como una distorsión aplicada durante la fase de entrenamiento, ambas fases con FRAV3D, y la última comprende el estudio de rotaciones con la base FEI.

El orden de los experimentos de la primera fase se muestra en la Tabla 6.5. Para esta etapa se incluyó un máximo de ocho distorsiones por rotación y se consideraron los casos donde el mismo ángulo se aplicaba en sentido de las manecillas del reloj y en contrasentido. La selección de las imágenes de prueba tomó en cuenta la necesidad de realizar comparativas entre la nueva tasa de reconocimiento y la obtenida en experimentos anteriores.

Los valores $R1 - R8$ de la Tabla 6.5 indican las distorsiones de rotación. A cada una de las pruebas también se le aplicaron los tres grupos de distorsiones de desplazamiento (cero, cuatro y ocho distorsiones).

En la fase dos, las rotaciones se agregaron durante el entrenamiento al igual que las distorsiones de desplazamiento. El orden de ejecución de las pruebas se muestra en la Tabla 6.6.

Para la base de datos FEI se organizaron los experimentos descritos en la Tabla 6.7. Como ya se mencionó anteriormente, con esta base de imágenes se estudió el comportamiento del algoritmo PCNC en rotaciones de tipo rotación sobre el eje Y .

No Prueba	Set de entrenamiento		Set de prueba	
	Descripción	Imagen	Descripción	Imagen
1	frontal, rotaciones	1,2,3,4,R1	rotación menor en z	9
2	frontal, rotaciones	1,2,3,4,R2	rotación menor en z	9
3	frontal, rotaciones	1,2,3,4,R3	rotación menor en z	9
4	frontal, rotaciones	1,2,3,4,R4	rotación menor en z	9
5	frontal, rotaciones	1,2,3,4,R5	rotación menor en z	9
6	frontal, rotaciones	1,2,3,4,R6	rotación menor en z	9
7	frontal, rotaciones	1,2,3,4,R7	rotación menor en z	9
8	frontal, rotaciones	1,2,3,4,R8	rotación menor en z	9
9	frontal, rotaciones	1,2,3,4,R1	rotación grande en z	10
10	frontal, rotaciones	1,2,3,4,R2	rotación grande en z	10
11	frontal, rotaciones	1,2,3,4,R3	rotación grande en z	10
12	frontal, rotaciones	1,2,3,4,R4	rotación grande en z	10
13	frontal, rotaciones	1,2,3,4,R5	rotación grande en z	10
14	frontal, rotaciones	1,2,3,4,R6	rotación grande en z	10
15	frontal, rotaciones	1,2,3,4,R7	rotación grande en z	10
16	frontal, rotaciones	1,2,3,4,R8	rotación grande en z	10
17	frontal, rotaciones	1,R1,R5	rotaciones en z	9,10
18	frontal, rotaciones	1,2,3,4,R1,R5	rotaciones en z	9,10
19	frontal, rotaciones	1,2,3,4,R2,R6	rotaciones en z	9,10
20	frontal, rotaciones	1,2,3,4,R3,R7	rotaciones en z	9,10
21	frontal, rotaciones	1,2,3,4,R4,R8	rotaciones en z	9,10

TABLA 6.5: Distribución de imágenes para entrenamiento y prueba en FRAV2D para la primera fase.

No Prueba	Set de entrenamiento		Set de prueba	
	Descripción	Imagen	Descripción	Imagen
1	frontal	1,2,3,4	1 rotación grande en z	9
2	frontal	1,2,3,4	1 rotación menor en z	10
3	frontal	1,2,3,4	2 rotación grande en z	9,10

TABLA 6.6: Distribución de imágenes para entrenamiento y prueba en FRAV2D para la segunda fase.

6.3. Resultados

Con el fin de disminuir el tiempo de entrenamiento, las imágenes se procesaron por grupos de semejanza; esto es, primero se procesaron las n imágenes del primer sujeto, posteriormente las n imágenes del segundo sujeto y así sucesivamente. Los resultados obtenidos se muestran en las siguientes subsecciones así como los tiempos registrados en cada etapa del proceso.

No Prueba	Set de entrenamiento		Set de prueba	
	Descripción	Muestras	Descripción	Muestras
1	perfil derecho, perfil izquierdo, frontal	1,10,11	restantes	2,3,4,5,6,7,8,9,12
2	giro derecha, giro izquierda, frontal	2,9,11	restantes	1,3,4,5,6,7,10,12
3	giro derecha, giro izquierda, frontal	3,8,11	restantes	1,2,4,5,6,7,9,10,12
4	giro derecha, giro izquierda, frontal	4,7,11	restantes	1,2,3,5,6,8,9,10,12
5	giro derecha, giro izquierda, frontal	5,6,11	restantes	1,2,3,4,7,8,9,10,12
6	perfil derecho	1	giros a la derecha	2,3,4,5
7	perfil derecho	2	giros a la derecha	1,3,4,5
8	perfil derecho	3	giros a la derecha	1,2,4,5
9	perfil derecho	4	giros a la derecha	1,2,3,5
10	perfil derecho	5	giros a la derecha	1,2,3,4
11	perfil izquierdo	6	giros a la izquierda	7,8,9,10
12	perfil izquierdo	7	giros a la izquierda	6,8,9,10
13	perfil izquierdo	8	giros a la izquierda	6,7,9,10
14	perfil izquierdo	9	giros a la izquierda	6,7,8,10
15	perfil izquierdo	10	giros a la izquierda	6,7,8,9

TABLA 6.7: Distribución de imágenes para entrenamiento y prueba de la base de rostros FEI.

6.3.1. Experimentos-Primera Fase

La Tabla 6.8 presenta los resultados obtenidos en la primera fase experimental. A cada experimento de la Tabla 6.5 se le aplicaron cero ($D0$), cuatro ($D0 - D4$) y hasta ocho distorsiones de desplazamiento ($D0 - D8$), tal como se muestra. Todos los experimentos

obtenidos en esta tabla tienen como set de prueba a la imagen número 9 y como set de entrenamiento, las primeras cuatro imágenes (1-4) y una rotación.

No. de Prueba	No. de Distorsiones	Codificación (s)	Entrenamiento (s)	Reconocimiento (s)	No. de Errores
1	0	31	11	6	81
	4	136	224	6	62
	8	243	503	6	56
2	0	31	60	6	81
	4	211	226	6	62
	8	242	581	7	56
3	0	32	19	5	84
	4	136	215	5	65
	8	243	643	5	56
4	0	32	18	5	87
	4	147	191	5	68
	8	248	553	5	56
5	0	32	23	5	85
	4	136	194	5	47
	8	240	500	5	48
6	0	32	24	5	82
	4	136	182	5	49
	8	244	461	5	45
7	0	30	21	5	83
	4	137	224	5	54
	8	239	433	6	46
8	0	30	17	5	89
	4	132	181	5	59
	8	237	489	6	44

TABLA 6.8: Resultado de los Experimentos con FRAV2D en la primera fase con la imagen de prueba 9.

En los experimentos se incluyeron pruebas con imágenes rotadas en sentido contrario a la del set de prueba (pruebas 1-4) ya que se quería obtener numéricamente el impacto que tenía la inclusión de datos poco útiles en el clasificador; aunque el objetivo principal era encontrar la rotación en el mismo sentido que mejorara la tasa de reconocimiento (pruebas 1-4). En los experimentos de la Tabla 6.8 se puede observar cómo la mayor influencia la tiene el desplazamiento seguido de la rotación y la poca mejora que refleja la introducción de ángulos opuestos al mantener el número de errores casi constantes.

En esta fase, por otra parte, la menor tasa de error se registró en la prueba No 8 con 44 errores de un total de 103 imágenes probadas; es decir, una tasa de error de 42.72 %. Cabe mencionar que en las pruebas 5-7 se obtuvieron tasas de error más o menos semejantes y

que en las pruebas 5 y 6, la diferencia entre aplicar 4 y 8 distorsiones de desplazamiento es muy baja. Este hecho es importante ya que se busca introducir el menor número de datos manteniendo un buen índice de reconocimiento.

El mismo procedimiento para probar el índice de reconocimiento en imágenes rotadas positivamente se hizo para imágenes rotadas negativamente. La Tabla 6.9 muestra los resultados obtenidos tomando como set de prueba la imagen No 10.

No. de Prueba	No. de Distorsiones	Codificación (s)	Entrenamiento (s)	Reconocimiento (s)	No. de Errores
9	0	44	48	6	63
	4	140	192	6	37
	8	251	477	5	21
10	0	39	36	6	63
	4	145	205	5	21
	8	252	494	5	22
11	0	32	23	5	76
	4	141	263	5	32
	8	241	470	4	22
12	0	32	19	5	67
	4	139	203	5	32
	8	255	524	6	22
13	0	32	23	5	76
	4	141	199	5	44
	8	248	508	5	39
14	0	38	52	5	82
	4	140	185	5	43
	8	248	461	5	37
15	0	36	43	5	83
	4	139	198	5	48
	8	249	441	5	45
16	0	32	17	6	78
	4	146	220	5	46
	8	252	481	5	45

TABLA 6.9: Resultado de los Experimentos con FRAV2D en la primera fase con la imagen de prueba 10.

En este caso los resultados mejoraron considerablemente teniendo una tasa de error de hasta 20.39% para todos los ángulos positivos (22 aciertos de 103). Los ángulos negativos que más error causaron fueron los más alejados a la rotación original; es decir; 35° y 45°. Si se comparan las imágenes 9 y 10 (Fig. 6.2), se puede observar que existen rotaciones sobre todos los ejes en la imagen 9, mientras que en la imagen 10 la rotación sobre Z es más simple.



FIGURA 6.2: Imagen 9 (izquierda) e imagen 10 (derecha) .

Finalmente se realizaron pruebas con las imágenes 9 y 10 en el set de pruebas y dos imágenes rotadas los mismos grados pero en sentido contrario. Lo que se buscaba era ver numéricamente el impacto que tenía la inclusión de este tipo de imágenes en la tasa de reconocimiento. Los resultados se muestran en la Tabla 6.10.

No. de Prueba	No. de Distorsiones	Codificación (s)	Entrenamiento (s)	Reconocimiento (s)	No. de Errores
17	0	19	14	5	167
	4	85	128	9	107
	8	282	782	14	72
18	0	38	26	5	156
	4	163	320	9	79
	8	287	639	9	73
19	0	39	28	5	75
	4	169	273	5	32
	8	302	796	5	24
20	0	38	26	5	74
	4	171	267	5	33
	8	300	707	5	29
21	0	39	21	5	70
	4	175	315	5	33
	8	301	713	5	29

TABLA 6.10: Resultado de los Experimentos con FRAV2D en la primera fase con la imagen de prueba 9 y 10.

El mejor resultado de la Tabla 6.10 se obtuvo en la prueba número 19, con rotaciones de 35° y -35° , y 8 desplazamientos. Cuando se estudiaron los ángulos por separado, se obtuvo que el mejor resultado para el reconocimiento de la imagen No 9 fue con -45° , seguido de -25° , -35° y -15° ; para el caso de la imagen No 10, el mejor resultado se reportó en la prueba 9 con la rotación de 15° seguida de 25° , 35° y 45° . La intersección

entre ambos grupos debía proveer una forma de disminuir las tasas de error del reconocimiento de las dos imágenes, lo cual se confirmó al obtener una tasa de error más baja para los ángulos de -25° y 25° en la prueba No 19.

La prueba No 17 se hizo con el fin de obtener un resultado numérico para comparar el desempeño del clasificador con una imagen frontal y dos rotaciones en vez de cuatro imágenes frontales y dos rotaciones; esto es, disminuir las entradas al clasificador. Lo que se observó es que las imágenes seleccionadas por sí mismas no proporcionan una mejora sino hasta aplicar ocho distorsiones de desplazamiento, entonces el porcentaje de error es casi el mismo pero con menores tiempos de codificación, entrenamiento y reconocimiento.

6.3.2. Experimentos- Segunda Fase

En los experimentos de la segunda fase se incluyen distorsiones de tipo rotación durante el entrenamiento; esto es, cada imagen de entrada es rotada y desplazada por lo que en realidad el número de imágenes procesadas se eleva al producto de $P \times C \times D$. Donde P es el número de imágenes de prueba, C el número de clases y D el número de distorsiones aplicadas. El número de distorsiones varía entre uno (una rotación), cinco (cuatro desplazamientos y una rotación), y nueve (ocho desplazamientos y una rotación). Los resultados que tomaron como set de prueba a la imagen No 9 se muestran en la Tabla 6.11.

Distorsión	Codificación (s)	Entrenamiento (s)	Reconocimiento (s)	No de errores
D0, R5	68	38	7	72
D0-D4, R5	142	188	5	60
D0-D8, R5	223	390	5	58
D0, R6	59	40	6	70
D0-D4, R6	142	177	5	54
D0-D8, R6	214	339	5	54
D0, R7	58	39	5	81
D0-D4, R7	141	182	6	53
D0-D8, R7	212	497	5	53
D0, R8	57	43	5	77
D0-D4, R8	140	207	5	64
D0-D8, R8	224	393	5	56

TABLA 6.11: Inclusión de distorsiones de rotación y desplazamiento para reconocimiento de la imagen 9.

De los resultados en la Tabla 6.11 el que menor tasa de error presentó fue al que se le aplicaron rotaciones de -35° (R7), mientras que el número de errores se mantuvo igual para 4 y 8 distorsiones de desplazamiento. La siguiente rotación en la que se registró un menor número de errores fue la de 25° (R6) e igualmente se mantuvo constante el número de errores tanto para 4 como para 8 distorsiones de desplazamiento.

El mismo procedimiento para estudiar la tasa de reconocimiento de una imagen rotada en sentido antihorario representada por la imagen no. 9, se realizó con la imagen rotada en sentido horario con la imagen No. 10. Los resultados se muestran en la Tabla 6.12.

Distorsión	Codificación (s)	Entrenamiento (s)	Reconocimiento (s)	No de errores
D0, R1	56	38	5	65
D0-D4, R1	139	200	5	41
D0-D8, R1	212	462	7	38
D0, R2	54	49	5	65
D0-D4, R2	140	225	5	47
D0-D8, R2	222	433	7	37
D0, R3	54	46	5	64
D0-D4, R3	138	224	5	31
D0-D8, R3	223	489	6	40
D0, R4	56	39	5	70
D0-D4, R4	137	202	5	44
D0-D8, R4	223	464	5	56

TABLA 6.12: Inclusión de distorsiones de rotación y desplazamiento para reconocimiento de la imagen 10.

Para la Tabla 6.12 el mejor resultado se presentó con el ángulo de 25° (R2) con una tasa de error de 35.92%; es decir, 37 errores de un total de 103 imágenes.

Una vez generados los resultados introduciendo un ángulo de rotación a la vez, se realizaron experimentos con dos ángulos de la misma magnitud y en sentido contrario en conjunto con ninguna, cuatro u ocho distorsiones de desplazamiento. Los resultados de esta etapa se muestran en la Tabla 6.13.

En esta fase, la mejor tasa de reconocimiento se obtuvo cuando se aplicaron rotaciones de -35° y 35° (R3 y R7) más 8 distorsiones de desplazamiento, la cual fue de 41.26%; es decir, 85 imágenes no reconocidas de un total de 206. La diferencia entre aplicar 4 y 8 distorsiones de desplazamiento fue más visible en esta fase que en la anterior pero se mantuvo casi contante para R3 y R7.

Distorsión	Codificación [s]	Entrenamiento [s]	Reconocimiento [s]	No de errores
D0, R1, R5	82	92	7	127
D0-D4, R1, R5	165	317	9	93
D0-D8, R1, R5	254	540	9	85
D0, R2, R6	90	95	8	133
D0-D4, R2, R6	81	98	9	95
D0-D8, R2, R6	254	560	9	92
D0, R3, R7	81	97	9	130
D0-D4, R3, R7	168	263	10	87
D0-D8, R3, R7	255	626	9	85
D0, R4, R8	82	100	9	138
D0-D4, R4, R8	169	294	9	95
D0-D8, R4, R8	263	550	9	89

TABLA 6.13: Inclusión de distorsiones de rotación y desplazamiento para reconocimiento de las imágenes 9 y 10.

6.3.3. Experimentos-Tercera Fase

La tercera fase comprendía experimentos sobre la base de imágenes de FEI. En esta fase se estudió la sensibilidad del clasificador ante rotaciones sobre el eje Y por lo que la selección de las imágenes en el set de entrenamientos incluían siempre ambos perfiles ya sea parciales o completos, más una imagen frontal con expresión neutra. Los resultados obtenidos se muestran en la Tabla 6.14.

No. de Prueba	No. de Distorsiones	Codificación (s)	Entrenamiento (s)	Reconocimiento (s)	No. de Errores
1	0	52	18	122	179
	4	410	158	119	10
	8	369	260	116	7
2	0	70	19	108	116
	4	366	142	121	11
	8	426	223	167	6
3	0	59	17	169	74
	4	243	105	109	30
	8	365	376	121	22
4	0	55	26	113	195
	4	224	146	155	41
	8	367	297	142	37
5	0	61	16	107	268
	4	279	200	134	79
	8	474	277	125	79

TABLA 6.14: Resultados de los experimentos con FEI

Los experimentos registrados en la Tabla 6.14 muestran la sensibilidad del clasificador ante rotaciones sobre el eje Y , teniendo como máxima tasa de error 14.89%; es decir, 268 imágenes mal clasificadas de un total de 1800 rostros con ninguna distorsión de desplazamiento y tres imágenes que conformaban el set de prueba: una frontal y dos perfiles parciales (-25° , 25°). El mejor resultado se obtuvo con una imagen frontal y dos rotaciones de $+/- 40^\circ$ (Prueba 2), seguida de perfiles completos (Prueba 1), rotación de $+/- 30^\circ$ (3), $+/- 20^\circ$ (4), y $+/- 10^\circ$ (Prueba 5). También se puede observar que la tasa de reconocimiento para 4 y 8 distorsiones es muy parecida, en la prueba 5, por ejemplo, son iguales y en la prueba 1 hay tres imágenes mal clasificadas de diferencia. En cuanto al tiempo para reconocer 1800 imágenes con 8 distorsiones fue en promedio de 134.2s (2.24 min), su codificación tomó en promedio 400.2s (6.67 min) y su entrenamiento fue de 286.6s (4.78 min).

Para estudiar la influencia de la inclusión de imágenes rotadas sobre Y en el reconocimiento de diferentes tomas de perfil se realizaron los experimentos número 6 al 15, sus resultados se muestran en la Tabla 6.15.

No. de Prueba	No. de Distorsiones	Codificación (s)	Entrenamiento (s)	Reconocimiento (s)	No. de Errores
6	8	116	40	49	174
7	8	120	55	54	36
8	8	114	42	30	37
9	8	55	42	29	35
10	8	58	54	33	66
11	8	171	86	66	77
12	8	191	75	69	50
13	8	176	86	81	40
14	8	199	64	61	57
15	8	145	89	63	164

TABLA 6.15: Resultados de los experimentos con FEI sobre tomas de perfil

En los experimentos de la Tabla 6.15, se aplicaron siempre 8 distorsiones de desplazamiento. Los resultados obtenidos indican que las imágenes de perfiles completos no ayudan a reconocer perfiles parciales, de hecho, son las que mayor número de errores registraron. El ángulo que permitió decrementar el número de errores fue el de 20° para rotaciones hacia la derecha (Prueba 9) y de 30° para rotaciones hacia la izquierda (Prueba 13).

6.4. Comparativa

Como se estableció en un principio, el presente trabajo buscaba mejorar los resultados obtenidos en experimentos pasados donde la tasa de error había sido de 66.67% con 1 distorsión de desplazamiento y de 33.7% con 12. Los nuevos resultados obtenidos se observan en la Tabla 6.16.

Versión	Distorsión	%Error	
		Rec. imagen 9	Rec. imagen 10
Original	D0	79.41	66.67
	D0-D12	54.3	33.7
Ampliación del set de entrenamiento	D0	79.61	61.16
	D0-D4	47.57	20.39
	D0-D8	43.69	21.36
Rotación como distorsión	D0, R2	63.11	67.96
	D0-D4, R2	45.63	52.43
	D0-D8, R2	35.92	52.43

TABLA 6.16: Comparativa entre resultados anteriores y nuevos resultados.

Capítulo 7

Conclusiones y Trabajo Futuro

El reconocimiento de rostros es una tarea de gran importancia para las áreas de seguridad y acceso a información que en el presente trabajo se ha trabajado a través de redes neuronales artificiales con propagación hacia adelante y un modelo de codificación de patrones basado en descriptores locales.

Ya que en los anteriores experimentos realizados con la base de datos FRAV3D se registraron bajas tasas de reconocimiento durante la clasificación de imágenes con distorsión tipo rotación sobre el eje Z y se obtuvo mejoría cuando se agregaron distorsiones de desplazamiento, se propuso la inclusión de distorsiones de rotación al set de entrenamiento o al proceso de entrenamiento. La propuesta se basa en lo establecido por Yu Lean en [48] donde se señala la importancia de los datos que se ingresan a una red neuronal. Una selección adecuada de estos puede afectar significativamente su desempeño ya sea positiva o negativamente.

La inclusión de rotaciones en el set de entrenamiento y durante el proceso de entrenamiento requirió la modificación de la estructura del programa original de PCNC así como la modificación de parámetros (umbrales, máscaras,). Los parámetros que se mantuvieron constantes para ambas bases de datos fueron el tamaño de la ventana de (13×13 píxeles), el valor de Δ (6 píxeles), el número de décadas que podían suceder hasta converger (199), el tamaño de la red (64000 neuronas), los puntos positivos y negativos por ventana (3 positivos, 3 negativos) y el número de características (400).

Debido a la diferencia de iluminación y contraste, los filtros cambiaron un poco entre la base de imágenes FEI y la de FRAV3D. Para FEI, por ejemplo, se estableció un umbral de brillo de 220, mientras que para FRAV3D fue de 50.

El registro del número de errores producido por diferentes ángulos en las pruebas no sólo permitía hacer una selección adecuada de datos al encontrar de forma experimental los que mejor representaban los patrones faciales sino también ver las deficiencias y bondades del clasificador. Se observó por ejemplo que las rotaciones sobre Y podían clasificarse considerablemente bien hasta el punto de registrar un porcentaje de error que tendía a 0% con 6 imágenes mal clasificadas de un total de 1800. Por otra parte, las rotaciones sobre Z sí producían errores mayores y aunque se logró disminuir la tasa de error, se requirió de la inclusión de varias distorsiones de desplazamiento.

Haciendo énfasis en que el uso de redes neuronales con grandes bases de datos requiere del tratamiento de los datos para que sean significativos para la red neuronal y una selección adecuada de los mismos, se estudiaron ángulos por separado y en pares, además de aplicar un pre-procesamiento a las imágenes de entrada. PCNC trabaja con valores de brillo y contornos por lo que el clasificador es insensible a tonos de piel y en general a razas como se pudo observar en los experimentos con FEI. Los datos que considera útiles provienen de un filtro de contornos de imágenes en escala de grises, lo que lo hace útil no sólo para rostros sino también para la clasificación de texturas en general (en este caso la textura equivale al rostro).

Una de las desventajas del clasificador es la sensibilidad ante grandes desplazamientos, razón por la cual el número de errores se incrementaba considerablemente cuando no se aplican distorsiones. Aunque el descriptor RLD y las permutaciones ayudan a mejorar las tasas de reconocimiento, se limitan sólo a desplazamientos horizontales y verticales. Este error puede corregirse agregando imágenes rotadas al set de entrenamiento o generando nuevas distorsiones de rotación pero esto implica un incremento en el tiempo de codificación y entrenamiento además de que el clasificador por sí mismo sigue siendo sensible a rotaciones. A pesar de ello, los tiempos registrados para cada etapa no sobrepasaron los 15 minutos para un máximo de 1800 imágenes, un tiempo asequible para la cantidad de imágenes procesadas.

7.1. Trabajo Futuro

Actualmente se han desarrollado descriptores que buscan ser insensibles a rotaciones y escala, problema muy común en el reconocimiento de patrones. Muchos de ellos se basan en la información proporcionada por histogramas, extracción de características o la proyección de la imagen en un plano diferente como la transformada de Gabor, transformada de Radon, transformada polar logarítmica, MTT [50], RLBP [49], RIFT [51], SURF [52], BRIEF [53].

Una de las opciones que más se parecen en funcionamiento al de PCNC es RLBP cuya estructura básica permite obtener un código invariante a rotaciones. La correcta fusión entre estas dos herramientas puede proporcionar una alternativa a los problemas de reconocimiento en presencia de rotaciones y traslaciones. PCNC es sensible a rotaciones porque no guarda una relación entre los píxeles vecinos; por lo tanto, una característica en un punto determinado al momento de presentarse una rotación, se considerará como una nueva característica y no como la misma en otra localidad.

Otra opción a desarrollar es la obtención de la proyección de la imagen a través de la aplicación de una transformada invariante a escala, traslación o rotación. El resultado obtenido puede entonces ser utilizarlo como la imagen de entrada. La desventaja de este método sería la diferenciación intraclase.

El desempeño de PCNC es considerablemente bueno, véase por ejemplo los resultados de los experimentos con FEI. No obstante lo anterior, aún pueden implementarse técnicas que le aporten robustez ante diferentes distorsiones sin que con ello incremente mucho el tiempo de procesamiento.

Bibliografía

- [1] Shaohua Kevin Zhou, Rama Chellappa, and Wenyi Zhao. *Unconstrained face recognition*, volume 5. Springer, 2006.
- [2] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1):37–52, 1987.
- [3] Edgar Osuna, Robert Freund, and Federico Girosi. Training support vector machines: an application to face detection. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 130–136. IEEE, 1997.
- [4] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 13(1):21–27, 1967.
- [5] Li Liu and Paul W Fieguth. Texture classification from random features. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(3):574–586, 2012.
- [6] Zhen Lei, Matti Pietikainen, and Stan Z Li. Learning discriminant face descriptor. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(2):289–302, 2014.
- [7] Xianbiao Qi, Rong Xiao, Jun Guo, and Lei Zhang. Pairwise rotation invariant co-occurrence local binary pattern. In *Computer Vision–ECCV 2012*, pages 158–171. Springer, 2012.
- [8] E Kussul, T Baidyk, and DC Wunsch. Image recognition systems with permutative coding. In *Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on*, volume 3, pages 1788–1793. IEEE, 2005.
- [9] Gengis K Toledo, Ernst Kussul, and Tatiana Baidyk. Neural classifier for micro work piece recognition. *Image and Vision Computing*, 24(8):827–836, 2006.

- [10] E Kussul, T Baidyk, L Kasatkina, and V Lukovich. Rosenblatt perceptrons for handwritten digit recognition. In *Neural Networks, 2001. Proceedings. IJCNN'01. International Joint Conference on*, volume 2, pages 1516–1520. IEEE, 2001.
- [11] E Kussul, Tatiana Baidyk, and Maksym Kussul. Neural network system for face recognition. In *Circuits and Systems, 2004. ISCAS'04. Proceedings of the 2004 International Symposium on*, volume 5, pages V–768. IEEE, 2004.
- [12] Ernst Kussul, Tatiana Baidyk, Cristina Conde, I Martin de Diego, and Enrique Cabello. Face recognition improvement with distortions of images in training set. In *Neural Networks (IJCNN), The 2013 International Joint Conference on*, pages 1–6. IEEE, 2013.
- [13] Anil K Jain and Stan Z Li. *Handbook of face recognition*. Springer, 2005.
- [14] Aleksandra Babich et al. Biometric authentication. types of biometric identifiers, 2012.
- [15] Alphonse Bertillon. *La photographie judiciaire: avec un appendice sur la classification et l'identification anthropométriques*. Gauthier-Villars, 1890.
- [16] Teuvo Kohonen. *Self-organizing maps*, volume 30. Springer Science & Business Media, 2001.
- [17] P Jonathon Phillips, Hyeonjoon Moon, Syed A Rizvi, and Patrick J Rauss. The feret evaluation methodology for face-recognition algorithms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(10):1090–1104, 2000.
- [18] P Jonathon Phillips, Patrick J Flynn, Todd Scruggs, Kevin W Bowyer, Jin Chang, Kevin Hoffman, Joe Marques, Jaesik Min, and William Worek. Overview of the face recognition grand challenge. In *Computer vision and pattern recognition, 2005. CVPR 2005. IEEE computer society conference on*, volume 1, pages 947–954. IEEE, 2005.
- [19] Jonathon Shlens. A tutorial on principal component analysis, april 2009. URL <http://www.sn1.salk.edu/~shlens/pca.pdf>.
- [20] Juha Karhunen, Erkki Oja, Liuyue Wang, Ricardo Vigarío, and Jyrki Joutsensalo. A class of neural networks for independent component analysis. *Neural Networks, IEEE Transactions on*, 8(3):486–504, 1997.

- [21] Marian Stewart Bartlett, Javier R Movellan, and Terrence J Sejnowski. Face recognition by independent component analysis. *Neural Networks, IEEE Transactions on*, 13(6):1450–1464, 2002.
- [22] Kamran Etemad and Rama Chellappa. Discriminant analysis for recognition of human face images. *JOSA A*, 14(8):1724–1733, 1997.
- [23] Chengjun Liu and Harry Wechsler. Face recognition using evolutionary pursuit. In *Computer Vision—ECCV’98*, pages 596–612. Springer, 1998.
- [24] Laurenz Wiskott, J-M Fellous, N Kuiger, and Christoph Von Der Malsburg. Face recognition by elastic bunch graph matching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):775–779, 1997.
- [25] Juwei Lu, Konstantinos N Plataniotis, and Anastasios N Venetsanopoulos. Face recognition using kernel direct discriminant analysis algorithms. *Neural Networks, IEEE Transactions on*, 14(1):117–126, 2003.
- [26] Ming-Hsuan Yang. Face recognition using kernel methods. *Advances in Neural Information Processing Systems 14*, pages 215–220, 2002.
- [27] Alexander Kadyrov and Maria Petrou. The trace transform and its applications. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(8):811–828, 2001.
- [28] Sanun Srisukh, Kongnat Ratanarangsankh, Werasak Kurutach, and Sahatsawat Waraklang. Face recognition using a new texture representation of face images. Technical report, Mahanakorn University of Technology, Bangkok, THAILAND, 2003.
- [29] Timothy F Cootes, Gavin V Wheeler, Kevin N Walker, and Christopher J Taylor. View-based active appearance models. *Image and vision computing*, 20(9):657–664, 2002.
- [30] Volker Blanz and Thomas Vetter. Face recognition based on fitting a 3d morphable model. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(9):1063–1074, 2003.
- [31] Alexander M Bronstein, Michael M Bronstein, and Ron Kimmel. Expression-invariant 3d face recognition. In *Audio-and Video-Based Biometric Person Authentication*, pages 62–70. Springer, 2003.

- [32] Baback Moghaddam, Tony Jebara, and Alex Pentland. Bayesian face recognition. *Pattern Recognition*, 33(11):1771–1782, 2000.
- [33] Ara V Nefian and Monson H Hayes III. Hidden markov models for face recognition. *choice*, 1:6, 1998.
- [34] Ron Meir and Gunnar Rätsch. An introduction to boosting and leveraging. In *Advanced lectures on machine learning*, pages 118–183. Springer, 2003.
- [35] Paul Viola and Michael J Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.
- [36] Yoav Freund, Robert Schapire, and N Abe. A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, 14(771-780):1612, 1999.
- [37] Guo-Dong Guo and Hong-Jiang Zhang. Boosting for fast face recognition. In *Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 2001. Proceedings. IEEE ICCV Workshop on*, pages 96–100. IEEE, 2001.
- [38] K Chan G Guo, S Z Li. Face recognition by support vector machines. *IEEE International Conference on Automatic Face and Gesture Recognition*, 3:196–201, 2000.
- [39] Paul Ekman and Wallace V Friesen. Facial action coding system. 1977.
- [40] Centro Universitario da FEI. Fei face database, 2014. URL <http://fei.edu.br/~cet/facedatabase.html/>.
- [41] Carlos Eduardo Thomaz and Gilson Antonio Giraldi. A new ranking method for principal components analysis and its application to face image analysis. *Image and Vision Computing*, 28(6):902–913, 2010.
- [42] LL OLIVEIRA JR and CE Thomaz. Captura y alineación de imágenes: Una base de rostros brasileña. *Relatório de iniciação científica, Depto. Eng. Elétrica da FEI, São Bernardo do Campo, SP*, 10, 2006.
- [43] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.
- [44] Andrew W Yip and Pawan Sinha. Contribution of color to face recognition. *Perception-London*, 31(8):995–1004, 2002.

- [45] Frank Rosenblatt. Principles of neurodynamics. perceptrons and the theory of brain mechanisms. Technical report, DTIC Document, 1961.
- [46] Erns M Kussul, Tatiana N Baydyk, Donald C Wunsch, Oleksandr Makeyev, and Anabel Martín. Permutation coding technique for image recognition systems. *Neural Networks, IEEE Transactions on*, 17(6):1566–1579, 2006.
- [47] Alan W Paeth. A fast algorithm for general raster rotation. In *Graphics Interface*, volume 86, pages 77–81, 1986.
- [48] Lean Yu, Shouyang Wang, and Kin Keung Lai. An integrated data preparation schelme for neural network data analysis. *Knowledge and Dta Engineering, IEEE Transactions on*, 18(2):217–230, 2006.
- [49] Timo Ahonen, Abdenour Hadid, and Matti Pietikäinen. Face recognition with local binary patterns. In *Computer vision-eccv 2004*, pages 469–481. Springer, 2004.
- [50] Sanun Srisuk, Kongnat Ratanarangsank, Werasak Kurutach, and Sahatsawat Waraklang. Face recognition using a new texture representation of face images. 2003.
- [51] Jie Chen, Shiguang Shan, Chu He, Guoying Zhao, Matti Pietikainen, Xilin Chen, and Wen Gao. Wld: A robust local image descriptor. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(9):1705–1720, 2010.
- [52] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, 2008.
- [53] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. Brief: Binary robust independent elementary features. In *Computer Vision–ECCV 2010*, pages 778–792. Springer, 2010.