

UNIVERSIDAD IBEROAMERICANA



“RECONOCIMIENTO DE ROSTROS UTILIZANDO ANÁLISIS DE COMPONENTES PRINCIPALES: LIMITACIONES DEL ALGORITMO”

TESIS

Que para obtener el grado de

MAESTRO EN SISTEMAS Y PLANEACION

P r e s e n t a:

CARLOS VILLEGAS QUEZADA

Director

MTRO. JORGE RIVERA ALBARRAN

Asesores:

MTRO. PEDRO FERNANDO SOLARES SOTO

MTRO. FELIPE ANTONIO TRUJILLO FERNANDEZ

Resumen

Una de las tecnologías emergentes que muestra grandes posibilidades de desarrollo en los próximos años, es la Biometría, la cual tiene como propósito la identificación de personas por medio de sus diversas características físicas (voz, huellas digitales, características faciales, etc.). A partir de los ataques terroristas a las dos torres del World Trade Center en New York, el 11 de septiembre de 2001, los sistemas biométricos cobraron un gran auge. Sobre todo, los sistemas de reconocimiento de rostros. Dichos sistemas tendrán el objetivo de detectar posibles sospechosos, a partir de diversas bases de datos con fotografías de personas consideradas terroristas o delincuentes. También, tienen otros múltiples usos de identificación en la vida cotidiana: acceso a cajeros automáticos, a estacionamientos, acceso de personal a oficinas, reconocimiento de personas desaparecidas y fallecidas que no presenten identificación, acceso a computadoras (en lugar de utilizar password), etc.

En la actualidad, los dos paquetes de software comercial más utilizados en el reconocimiento de rostros, utilizan la técnica de Análisis por Componentes Principales o alguna derivación de ella. Asimismo, la gran mayoría de los paquetes computacionales realizados en proyectos de investigación para reconocimiento e identificación de rostros, que utilizan otros métodos estadísticos (regresión, redes neuronales, análisis cluster, etc.), comparten algo en común con el análisis de componentes principales: todos ellos utilizan cálculos y métricas que se llevan a cabo en un espacio denominado Euclidiano o L_2 .

El software basado en técnicas que utilizan la métrica Euclidiana, no ha presentado resultados completamente satisfactorios en el proceso de reconocimiento de rostros. Se considera que dichas limitaciones se presentan por dos motivos principales: el primero, debido al propio algoritmo, en segundo lugar, al tipo de características estadísticas que presentan los datos de entrada al sistema (los rostros). Para poder utilizar adecuadamente métodos estadísticos en espacios L_2 , los datos de entrada deben cumplir los supuestos de linealidad, homoscedasticidad y sobre todo, el de normalidad multivariada. Se postula en el presente trabajo, que los datos numéricos que caracterizan la fotografía digital de un rostro, no cumplen los presupuestos mencionados. Por tanto, dicha violación, contribuye a la obtención de resultados erróneos en el proceso de reconocimiento a pesar de la “robustez” de los métodos de análisis de componentes principales y técnicas similares.

En la tesis se describen los elementos teóricos que subyacen a la técnica de componentes principales y con ellos, se desarrolló un sistema computacional en MATLAB basado en dicha técnica. Asimismo, se realizó un análisis estadístico de las características de las imágenes digitales de rostros (base AT&T y fotografías adicionales del portal del FBI), para comprobar si cumplían los supuestos de normalidad, homoscedasticidad y linealidad.

Con el sistema computacional elaborado, se realizaron diversos experimentos de reconocimiento de rostros utilizando diversas fotografías por sujeto en la base de entrenamiento, con el propósito de observar el comportamiento del sistema. Los resultados muestran que no se cumplen con los supuestos estadísticos mencionados, los cuales se requieren para utilizar adecuadamente técnicas basadas en espacios Euclidianos. Por otra

parte, el método de componentes principales presentó resultados adecuados, pero también, muestra serias dificultades y limitaciones que impiden un reconocimiento de rostros totalmente confiable.

Finalmente, se postuló un nuevo método aplicable al reconocimiento de rostros, que se basa en la utilización de una *familia de funciones polinomiales multivariadas de aproximación manejadas en un Espacio L_∞* . Este método, al trabajar el problema de aproximación para caracterizar a los rostros en un espacio *n-dimensional* bajo la norma L_∞ (norma mínimax o Tchebyshev), no requiere el cumplimiento de los supuestos de linealidad, homoscedasticidad y normalidad. Por tanto, se piensa que eliminará o reducirá las limitaciones que presentan los sistemas actuales basados en técnicas de componentes principales o similares al utilizar datos que no satisfacen los supuestos estadísticos mencionados. La técnica que se propone, utilizará el denominado Algoritmo Genético Ecléctico, para encontrar el aproximante que caracterizará a los rostros bajo un enfoque de optimización combinatoria y aproximación multivariada.

Tabla de Contenido

Introducción	13
1. Planteamiento del estudio	28
1.1. El problema de investigación	28
1.2. Delimitación del problema	30
1.3. Preguntas de investigación	31
1.4. Hipótesis	31
1.5. Objetivos	32
1.5.1. Objetivo general	32
1.5.2. Objetivos particulares	32
1.6. Limitaciones del problema	33
1.7. Resultados que se pretende lograr con el proyecto	34
2. El “Estado del conocimiento” en el reconocimiento de rostros humanos por métodos computacionales	36
2.1. Enfoque psicológico y de neurociencias en el reconocimiento de rostros en el ser humano	37
2.2. Enfoque computacional	41
2.2.1. Segmentación de rostros	42
2.2.2. Extracción de características	45
2.2.3. Reconocimiento de rostros	52
2.2.4. Reconocimiento de rostros a partir de perfiles	70
2.2.5. Métodos evolutivos	72
3. Análisis de componentes principales	74
3.1. Características generales del análisis de componentes principales	75
3.2. Planteamiento matemático del análisis de componentes principales	77
3.3. Representación de rostros utilizando el análisis de componentes principales	85
3.4. Algoritmos para reconocimiento de rostros utilizando análisis de componentes principales	87

4.	Análisis estadístico de las imágenes digitales de rostros	92
4.1.	Supuestos estadísticos del análisis multivariante tradicional	93
4.2.	Análisis estadístico de la imagen de un rostro	95
4.3.	Prueba de normalidad univariada y multivariada para fotografías de rostros	100
4.3.1.	Prueba de normalidad univariada	100
4.3.2.	Prueba de normalidad multivariada	106
4.3.3.	Prueba de homoscedasticidad y linealidad	109
5.	Análisis del reconocimiento de rostros utilizando técnicas de componentes principales	113
5.1.	Características de la base de datos de rostros utilizada	114
5.2.	Elaboración de la Base de Entrenamiento	118
5.3.	Configuración de los experimentos	119
5.4.	Características, Algoritmo y Funcionamiento del sistema de reconocimiento de rostros	127
5.4.1.	Características generales del sistema	127
5.4.2.	Algoritmo	127
5.4.3.	El sistema computacional para reconocimiento de rostros	132
5.4.4.	Operación y tipo de resultados que proporciona el sistema	135
5.5.	Experimentación	144
5.5.1.	Resultados con la base AT&T	149
5.5.2.	Experimentos con la base combinada AT&T y la del FBI	167
6.	Propuesta de un método para reconocimiento holístico de rostros utilizando aproximación multivariada y algoritmos genéticos eclécticos en un espacio L_∞	178
6.1.	Reconocimiento de rostros como un problema de aproximación multivariada	180
6.2.	Reconocimiento de patrones y Aprendizaje	182
6.3.	Algoritmos Genéticos	186
6.4.	Aproximación multivariada en un Espacio L_∞ como un problema de optimización combinatoria	197
6.5.	Método general para el reconocimiento de rostros utilizando algoritmos genéticos eclécticos y aproximantes polinomiales en un espacio L_∞	202
6.5.1.	Muestreo aleatorio de píxeles y configuración de variables	203

6.5.2. Reconocimiento de rostros como un problema de aprendizaje y obtención del aproximante multivariado mediante un algoritmo genético ecléctico	204
6.5.3. El rostro y sus espectros de señal	208
Conclusiones	214
Referencias bibliográficas.....	219
Anexos	237
Anexo A. Rostros de la Base AT&T	238
Anexo B. Resultados del análisis univariado de normalidad para los 400 rostros de la base AT&T	259

Lista de Tablas

TABLA

1.	Estadística descriptiva de los valores a nivel píxel que presenta la Fotografía s0802.pgm	96
2.	Estadística descriptiva del “rostro promedio” de las 400 fotografías de la base AT&T	99
3.	Resultados de la prueba de Normalidad Univariada (Kolmogorov Smirnov) para algunos de los rostros de la base AT&T	102
4.	Prueba de normalidad univariada para el “rostro promedio” de la base AT&T	106
5.	Resultados de la prueba de normalidad multivariada utilizando el software PRELIS, a una muestra de la base AT&T	108
6.	Resultados posibles considerando la decisión del sistema de reconocimiento en relación a la verdad sobre la fotografía a identificar	145
7.	Porcentajes y numero de rostros identificados en cada rubro de acuerdo a diversos Intervalos de confianza (con 40 sujetos y 9 fotografías por cada uno)	151
8.	Número de rostros identificados en cada rubro de acuerdo a diversos intervalos de confianza (con 40 sujetos en b.d. y 8 fotografías por cada uno, 53 rostros para identificar)	152
9.	Numero de rostros identificados en cada rubro de acuerdo a diversos intervalos de confianza (con 40 sujetos y 5 fotografías por cada uno)	152
10.	Numero de rostros y porcentajes de identificación considerando los umbrales $(\bar{X} + 2\sigma)$ y $(\bar{X} + 3\sigma)$ con la menor tasa posible de “falsos positivos” / “falsos negativos” y el mayor porcentaje posible de “positivos verdaderos” para los experimentos realizados con 40 sujetos (con diverso numero de fotografías por sujeto) y 53 fotografías a reconocer	154
11.	Número de rostros identificados en cada rubro de acuerdo a diversos intervalos de confianza (con 33 sujetos y 9 fotografías por cada uno)	156
12.	Número de rostros identificados en cada rubro de acuerdo a diversos intervalos de confianza (con 33 sujetos y 7 fotografías por cada uno)	156

13.	Número de rostros identificados en cada rubro de acuerdo a diversos intervalos de confianza (con 33 sujetos y 5 fotografías por cada uno)	157
14.	Número de rostros identificados en cada rubro de acuerdo a diversos intervalos de confianza (con 33 sujetos y 3 fotografías por cada uno)	157
15.	Número de rostros identificados en cada rubro de acuerdo a diversos intervalos de confianza (con 33 sujetos y 2 fotografías por cada uno)	158
16.	Número de rostros identificados en cada rubro de acuerdo a diversos intervalos de confianza (con 33 sujetos y 1 fotografía por cada uno)	158
17.	Número de rostros y porcentajes de identificación considerando un umbral de $(\bar{X} + 2\sigma)$ para los seis experimentos realizados con 33 sujetos (con diverso numero de fotografías por sujeto) y 53 fotografías a reconocer	160
18.	Número de rostros y porcentajes de identificación considerando una tasa de “falsos positivos” del 0.0% y máximo porcentaje de “positivos verdaderos” para los seis experimentos realizados con 33 sujetos (con diverso numero de fotografías por sujeto) y 53 fotografías a reconocer	162
19.	Número de rostros y porcentajes de identificación considerando una tasa de “falsos positivos” del 0.0% y clasificadas por máximo porcentaje de “positivos verdaderos” para los seis experimentos realizados con 33 sujetos (con diverso numero de fotografías por sujeto) y 53 fotografías a reconocer	163
20.	Número de rostros y porcentajes de identificación considerando el mejor umbral $(\bar{X} + 2\sigma)$ con la menor tasa posible de “falsos positivos” y el mayor porcentaje posible de “positivos verdaderos” para los seis experimentos realizados con 33 sujetos (con diverso numero de fotografías por sujeto) y 53 fotografías a reconocer	165
21.	Número de rostros y porcentajes de identificación considerando el mejor umbral $(\bar{X} + 2\sigma)$ clasificados por mayor porcentaje general de reconocimiento, menor porcentaje de “falsos positivos” y menor numero de fotografías por sujeto para los seis experimentos realizados con 33 sujetos (con diverso número de fotografías por sujeto) y 53 fotografías a reconocer	165
22.	Resultados de reconocimiento utilizando la base combinada del FBI con AT&T	173
23.	Porcentajes de reconocimiento verdadero para las fotografías de sospechosos del FBI (con respecto a 11 fotografías)	174

24	Comparación de resultados considerando un umbral de $\bar{X} + 2\sigma$ entre el experimento 9 y el experimento con la base AT&T / FBI	175
25.	Población inicial en un algoritmo genético simple	191
26.	Población inicial y selección de cromosomas	193
27.	Cruzamiento y nueva población	194
28.	Población final	195
29.	Matriz de características de píxeles (“m” características x “tm” píxeles) para obtener el aproximante que pueda caracterizar al rostro	203

Lista de Figuras

FIGURA

1.	Sistema de reconocimiento antropométrico de Alphonse Bertillion para el reconocimiento de criminales en Europa, 1877-1884	14
2.	Portada del libro “Finger Prints” publicado por Francis Galton en 1892, primer texto sobre identificación de huellas dactilares	15
3.	Reconocimiento de rostros utilizando (a) características geométricas y (b) plantillas	55
4.	Reconocimiento utilizando puntos seleccionados del rostro para ser procesados por la técnica de componentes principales	56
5.	Reconocimiento utilizando un “clasificador por muestreo” a partir de “n-tuplas” que se obtienen de los valores de píxeles	57
6.	Reconocimiento de rostros utilizando el metodo “Eigenfaces” propuesto por Turk y Pentland	58
7.	Ejemplo de fotografía de la base AT&T (segunda imagen de la octava persona de la base – s0802.pgm)	96
8.	Histograma de los niveles de gris de los pixeles correspondientes a la fotografía s0802.pgm	98
9.	Histograma de los niveles de gris correspondientes a los pixeles del “rostro promedio” de la base at&t	99
10.	Grafico Q-Q de normalidad univariada para la fotografía s0101.pgm	103
11.	Grafico Q-Q de normalidad univariada para la fotografía s0108.pgm	103
12.	Nube de puntos en “x” para los pixeles del sujeto 8/foto 8	110
13.	Nube de puntos en “y” para los pixeles del sujeto 8/foto 8	111
14.	Nube de puntos 3-D para los pixeles del sujeto 8/foto 8	112
15.	Configuración de la base de rostros AT&T conteniendo fotografías de 40 sujetos con 10 imágenes por cada persona	117
16.	Los primeros 25 rostros de la base AT&T utilizados como “fotografías a identificar”	121
17.	Los siguientes 15 rostros de la base AT&T utilizados como “fotografías a identificar”, para dar un total inicial de 40 fotografías a reconocer	122

18.	Trece fotografías adicionales para ser utilizadas como imágenes a identificar	123
19.	Modelo simplificado del “espacio de rostros” que se origina al realizar un análisis de componentes principales, ilustrando los cuatro resultados posibles que se obtienen al proyectar la imagen de un rostro en el espacio	131
20.	Ejecución del sistema de reconocimiento y Menú principal	132
21.	Selección de una fotografía a reconocer	135
22.	Reconocimiento de un rostro (verdadero positivo)	136
23.	Reconocimiento de un rostro (verdadero positivo). El sujeto a reconocer presenta anteojos y mayor edad	138
24.	Reconocimiento de un rostro (verdadero positivo). Las fotografías utilizadas no se encuentran controladas (luz, fondo de imagen, resolución, etc.)	138
25.	Reconocimiento “verdadero positivo” de una fotografía controlada de la base AT&T	139
26.	Se intenta reconocer una imagen que no corresponde a un rostro real. El sistema reporta adecuadamente “rostro no reconocido”	140
27.	Ejemplo de una identificación de tipo “falso negativo”	141
28.	Ejemplo de una identificación de tipo “falso positivo”	142
29.	Ejemplo de una identificación combinada de tipo “falso negativo” y “falso positivo”	143
30.	Reporte de periódico a partir del cual se obtuvieron dos fotografías digitalizadas para experimentación	168
31.	Ejemplo de un reporte de delincuentes publicado en internet por el FBI	169
32.	Ejemplo de un reporte de terroristas publicado en internet por el FBI	170
33.	Rostros que se obtuvieron del portal del FBI y se consideran como parte de la base de entrenamiento	171
34.	Rostros de delincuentes y terroristas utilizados como entrada al sistema (rostro a reconocer)	172
35.	Cruzamiento de cromosomas en un solo punto	188
36.	Óptimos locales y globales en un espacio multidimensional	200
37.	Clasificador Mahalanobis de distancia mínima	207
38.	Representación vectorial de los valores de gris correspondientes a la imagen de un rostro	209
39.	Señales correspondientes a 10 rostros de la base de entrenamiento AT&T. (a) – gradiente, (b) – frecuencia baja, (c) – máxima entropía	211

Introducción

En la vida cotidiana rara vez nos maravillamos de nuestra capacidad para reconocer un rostro. Es sorprendente que podamos memorizar un número tan considerable de rostros en el transcurso de nuestra vida. En general lo hacemos de golpe, de manera holística, sin necesidad de una descripción verbal ni de un análisis consciente de las diversas características faciales. Podemos dejar de ver a una persona por algunos años y cuando la encontramos nuevamente, generalmente la reconocemos, a pesar de que haya cambiado de peinado o tenga barba. Sin embargo, el lograr que una computadora pueda reconocer un rostro, es un problema muy complejo.

El reconocimiento de rostros por medios computacionales, forma parte del área denominada **Biometría**. La palabra biometría, se deriva del griego (bio y metria): *βιο* (vida) y *μετρία* (medida)

El propósito de la Biometría, consiste en la elaboración de métodos automatizados para la identificación o verificación de personas mediante el uso de características físicas o de comportamiento. Esta tecnología se basa en la premisa de que cada persona es única y

posee rasgos distintivos que pueden ser utilizados para identificarla. Algunos ejemplos de características físicas son la cara, las huellas digitales, el iris de los ojos y en los últimos años se ha considerado el ADN. Por su parte, algunos ejemplos de características de comportamiento son la voz, la manera de firmar, la forma de caminar.

Los antecedentes del uso de medidas físicas del ser humano para propósitos de identificación, se remonta al siglo XIV, en donde los chinos, ya utilizaban ciertas impresiones de los dedos para identificación. En el año de 1877, Alphonse Bertillion (antropólogo y policía de París) está considerado como la primera persona que desarrolló y utilizó de manera sistemática, un método para clasificar e identificar criminales a partir de diversas medidas del cuerpo y la cabeza (figura 1).



FIGURA 1. SISTEMA DE RECONOCIMIENTO ANTROPOMÉTRICO DE ALPHONSE BERTILLION PARA EL RECONOCIMIENTO DE CRIMINALES EN EUROPA, 1877-1884

FUENTE: "A brief history of Biometrics", en:

http://www.galwayeducationcentre.ie/athenry/a_brief_history_of_biometrics.htm

En 1884, con el sistema de Bertillion se tuvo la posibilidad de identificar a 241 criminales. Los cuerpos de policía, tanto de Inglaterra, Francia y los Estados Unidos, adoptaron el sistema. Sin embargo, años después se comprobó que presentaba fallas importantes en el proceso de identificación.

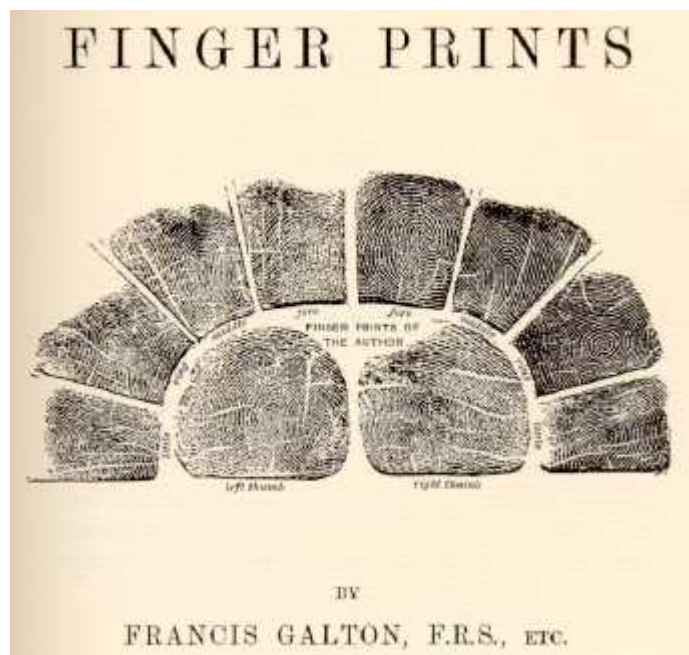


FIGURA 2. PORTADA DEL LIBRO “FINGER PRINTS” PUBLICADO POR FRANCIS GALTON EN 1892, PRIMER TEXTO SOBRE IDENTIFICACIÓN DE HUELLAS DACTILARES.
FUENTE: “from Finger Prints”, en: <http://etext.lib.virginia.edu/railton/wilson/galtonfp.html>

Por su parte, Sir William Herschel realizaba desde principios de 1800, investigaciones sobre el uso de huellas dactilares para tratar de identificar personas. Logró obtener aproximadamente 8000 conjuntos de huellas. Sin embargo fue Sir Francis Galton, antropólogo británico y primo de Charles Darwin, quien logró obtener el primer sistema de clasificación de huellas dactilares completamente operacional. En el año 1892, publicó el primer libro sobre huellas dactilares con el nombre de *Finger Prints* (Galton, 1892). En la

figura 2, se puede apreciar la portada de dicha publicación. Utilizando la investigación de Galton, Edward Richard Henry desarrolló entre 1896 y 1897 el que está considerado como el primer sistema de la policía para identificación por medio de huellas dactilares. Dicho sistema, fue utilizado por Scotland Yard en 1901 y utilizado posteriormente, por diversos cuerpos de policía en el mundo.

Por lo que respecta al desarrollo de sistemas aplicados al reconocimiento de rostros, las primeras aplicaciones se remontan a la década de los años sesentas con una compañía denominada Panoramic Research, Inc. en Palo Alto, California y financiada por el Departamento de Defensa de los Estados Unidos y otras agencias de Inteligencia. Uno de los fundadores de dicha empresa, fue Woodrow Wilson Bledsoe (pionero en el campo del razonamiento automatizado). Bledsoe desarrolló un sistema “semiautomático” para extraer características de la fotografía de un rostro, por medio de señalar en una tableta digitalizadora elementos faciales del rostro (diversos puntos de los ojos, boca, cabeza, etc.). A partir de dichos puntos, se obtenían coordenadas y se comparaban contra una Base de Datos que contenía nombres de personas y un registro de las coordenadas de su correspondiente fotografía. La computadora calculaba las distancias entre los registros de la base de datos y los puntos de la fotografía a identificar. En 1966, el sistema de Bledsoe continuó en desarrollo en el Stanford Research Institute (SRI).

El siguiente trabajo pionero en el área, se debe a Sakai y Fujibayashi (1969), quienes presentan un programa computacional (que no requiere intervención de algún operador humano) para confirmar la existencia o ausencia de un rostro en una imagen.

Por su parte, Kelly (1970) elabora una disertación doctoral sobre reconocimiento de rostros en Stanford. Su técnica permite que una computadora extraiga de manera automática el contorno de la cabeza y el cuerpo de una determinada persona a través de una fotografía; y pueda entonces localizar los ojos, nariz y boca. Sin embargo, el procedimiento requiere de tres imágenes de cada individuo: una imagen del cuerpo, una imagen del “fondo” de la fotografía (sin el cuerpo) y un acercamiento de la cabeza.

El siguiente desarrollo, se realiza en la disertación doctoral de Takeo Kanade (1973) en la Universidad de Kyoto, Japón. Quien reporta los mismos resultados que obtuvo Kelly, pero con la ventaja de utilizar únicamente una sola fotografía del rostro. Asimismo, propone un nuevo esquema más flexible para el análisis de la imagen, logrando reconocer adecuadamente, quince de veinte personas.

Los anteriores, son los trabajos pioneros en el área de reconocimiento de rostros, como se puede apreciar, el tiempo transcurrido entre dichos desarrollos (años 60 y 70) y la actualidad, es de apenas 45 años como máximo. Si lo comparamos contra el trabajo de huellas dactilares en 1892 (a partir del sistema de Galton), a la fecha, se tienen 113 años de desarrollo, por lo cual, los sistemas computacionales aplicados a dicha área presentan éxitos notables y sumamente confiables en la época actual.

De manera general, la mayoría de los sistemas biométricos funcionan de forma muy similar y se puede resumir en dos pasos. El primero, consiste en que la persona debe *registrarse* en el sistema. Durante el proceso de registro, el sistema captura el rasgo característico de la persona, como por ejemplo la huella digital, y lo procesa para crear una representación electrónica denominada “modelo de referencia” o “modelo de entrenamiento”. El modelo de entrenamiento debe ser guardado en una base de datos, una tarjeta inteligente, o en algún

otro lugar del cual será extraído en cualquier ocasión futura para dar origen al segundo paso.

El segundo paso depende si la función del sistema biométrico consiste en *verificar* la identidad de la persona o *reconocer* a la persona. En el caso de verificación, la persona le informa al sistema cual es su identidad ya sea presentando una tarjeta de identificación, dando una clave o proporcionando su huella digital o imagen del rostro. El sistema captura el rasgo característico de la persona y lo procesa para crear una representación también electrónica llamada “modelo en vivo”. Por último, el sistema compara el modelo en vivo con el modelo de referencia de la persona. Si ambos modelos son idénticos o la diferencia es menor que un determinado umbral, se considera una verificación adecuada. A este proceso se le conoce también como comparación uno-a-uno (one-to-one).

Cuando la función del sistema es de reconocimiento, la persona no le informa al sistema cual es su identidad. El sistema tan solo captura el rasgo característico de la persona y lo procesa para crear el modelo en vivo. Luego el sistema procede a comparar los modelos respectivos para determinar la identidad de la persona en la base de datos o en diversas bases de datos. Es un proceso denominado uno-a-muchos (one-to-many).

Los sistemas biométricos se han considerado elementos clave para cuestiones de seguridad, desde hace varias décadas. En enero del año 2000, la revista Technology Review publicada por el MIT, propuso al área de la biometría como una de las “diez tecnologías emergentes que cambiarán al mundo”. Y de acuerdo a la empresa Internacional Biometric Group (una de las empresas líderes en el desarrollo de sistemas biométricos), se consideró que las

ventas de sistemas biométricos crecerían de \$399 millones de dólares en el año 2000, a \$1.9 billones de dólares para el año 2005.

A partir de los ataques terroristas a las dos torres del World Trade Center en New York, el 11 de septiembre de 2001, los sistemas biométricos cobraron un gran auge. Sobre todo, los sistemas de reconocimiento de rostros. La propuesta de las instituciones de seguridad de los Estados Unidos, consistía en proporcionar a los aeropuertos y lugares de concentración masiva (estadios, auditorios, centros comerciales, etc.), sistemas de reconocimiento de rostros. Dichos sistemas tendrían el objetivo de detectar posibles sospechosos (delincuentes y sobre todo terroristas), a partir de diversas bases de datos con fotografías de personas consideradas terroristas o ligados de alguna forma a ellos.

Existen diversos procedimientos para elaborar sistemas computacionales de reconocimiento de rostros, los cuales se detallan en el siguiente capítulo. Sin embargo, los dos algoritmos que más eficiencia has mostrado y con los cuales se desarrollaron los dos sistemas comerciales aplicados al reconocimiento de rostros y que se encuentran en uso por diversas agencias de seguridad, tanto en Estados Unidos como en otros países, son:

- **Algoritmo Eigenfaces**, utilizado en el sistema **FaceNet/FaceFinder**¹ de la compañía Viisaje.
- **Algoritmo de Análisis de Características Locales (LFA)** utilizado en el sistema **FaceIt**² de la compañía Visionics.

¹ FaceNet/FaceFinder son marcas registradas de Viisaje. <http://www.viisage.com>

² FaceIt es marca registrada de Visionics Corp/Identix. <http://www.identix.com>

Ambos algoritmos se basan en la técnica denominada **Componentes Principales**, desarrollada hace décadas y utilizada originalmente en las áreas de Psicología, Ciencias Sociales y Educación.

El primer sistema comercial exitoso, fue **FaceNet**, el cual se desarrolló a partir de la técnica de análisis de componentes principales. Originalmente fueron Sirovich y Kirby (1987), quienes plantearon la posibilidad de caracterizar un rostro humano por medio de componentes principales y denominaron al modelo con el nombre de “eigenpictures”. Posteriormente, Turk y Pentland (1991) investigadores del MIT, modificaron y hicieron operativa la propuesta de los eigenpictures para desarrollar el algoritmo denominado *Eigenfaces*. Dicho algoritmo fue patentado por el MIT y posteriormente lo adquirió la empresa Viisaje para desarrollar comercialmente el sistema FaceNet/FaceFinder.

Por lo que se refiere al algoritmo LFA, fue desarrollado por Joseph Atick, Paul Griffin y Norman Redlich (1996), basándose también en la técnica de componentes principales. El algoritmo LFA, intenta reducir la información de la imagen digital de un rostro, y obtener únicamente de 12 a 40 elementos que caracterizan al rostro. Posteriormente, el Dr. Atick fundó la empresa Visionics y comercializó su sistema de reconocimiento de rostros con el nombre de **FaceIt**.

Ambos sistemas presentan cierto éxito, sus tasas de eficiencia reportadas por las compañías respectivas, indican un porcentaje de reconocimiento entre el 90% y el 95%. Sin embargo, en un estudio específico para evaluar al sistema FaceIt (Heo, Abidi, Paik y Abidi, 2003), reportan porcentajes de reconocimiento eficiente de rostros únicamente, entre el 79.8% y un

máximo de 95.2%, dependiendo de las características de las fotografías de los rostros: expresión, edad, iluminación, pose, etc. Incluso, en algunos experimentos con fotografías del sujeto tomadas después de varios meses y con variaciones en la resolución y detalles de las fotografías, la eficiencia únicamente fue del 32%.

Se considera que las deficiencias que presentan estos sistemas, entre otras cosas, se deben a la utilización de algoritmos que se basan en la técnica de componentes principales. Esta técnica presupone el cumplimiento de ciertas características estadísticas por parte de los datos que utiliza. Específicamente, uno de los principales supuestos que deben cumplir los datos, es el de ajustarse a una distribución de probabilidad normal multivariada. Asimismo, las técnicas de análisis de componentes principales, LFA y otras técnicas similares que utilizan la gran mayoría de las aplicaciones de reconocimiento de rostros, se manejan en un espacio Euclideo (L_2).

Dichos métodos son considerados robustos y permiten obtener resultados con cierta confiabilidad, a pesar de que los datos no cumplan estrictamente con los supuestos estadísticos requeridos. Sin embargo, es indudable, que entre más se alejen los datos manipulados de los supuestos estadísticos que se requieren para utilizar análisis de componentes principales y otras técnicas similares, los resultados cada vez serán menos confiables. Por otra parte, las técnicas mencionadas se basan también, en la utilización de métricas en espacios lineales normados o Euclideos, lo cual está directamente relacionado con el requerimiento de los supuestos estadísticos y el manejo de métricas en espacios L_1 , L_2 y L_p .

Los métodos tradicionales generalmente estadísticos (regresión, análisis factorial, análisis de componentes principales, etc.), utilizados en la mayoría de los sistemas actuales de reconocimiento de rostros presentan ciertas limitaciones al tener que trabajar bajo la norma L_2 (Dean, 1988; Jonson, 1991) o bajo algún subconjunto de ella. Entre las limitaciones, se requiere que las distribuciones de los datos deben satisfacer las características de normalidad, homoscedasticidad e independencia.

El autor del presente trabajo ha postulado un potencial método aplicable al reconocimiento de rostros, que se basa en la utilización de una **familia de funciones polinomiales multivariadas de aproximación manejadas en un Espacio L_∞** (Villegas, 1999). La propuesta anterior, se basa en la idea intuitiva, acerca de que los datos digitales que configuran la fotografía de un rostro a nivel de píxel (elemento básico para cualquier sistema de reconocimiento de rostros), no presentan una distribución normal multivariada. Por tanto, como se ha mencionado, la utilización de técnicas basadas en análisis de componentes principales y en general cualquier técnica tradicional de análisis multivariado, tiene grandes posibilidades de encontrar resultados erróneos y/o presentar deficiencias en el análisis, lo cual provocará disminuciones en la eficiencia de los algoritmos como sucede con las aplicaciones comerciales actuales.

El método propuesto considera, que al trabajar el problema de aproximación para caracterizar a los rostros como un sistema que encuentre el mejor aproximante en un espacio n -dimensional bajo la norma L_∞ (norma mínimax o Tchebyshev), se eliminarán (o

por lo menos se reducirán) las limitaciones que presentan los sistemas actuales al utilizar técnicas de componentes principales o similares y utilizar datos que no satisfacen presumiblemente, el supuesto de normalidad.

Por tanto, un elemento indispensable para conocer si tiene caso utilizar técnicas en un espacio L_∞ , es verificar si efectivamente los datos que representan a un determinado rostro y a un conjunto de rostros (base de datos), presentan respectivamente una distribución normal univariada y multivariada. Asimismo, se desearía conocer el comportamiento de un sistema de reconocimiento de rostros basado en análisis de componentes principales, para experimentar diversas alternativas con respecto a las características de las fotografías, el número adecuado de fotografías por sujeto que se deben tener en la base de entrenamiento y algunos otros elementos que incidan en la eficiencia del sistema.

Con el propósito de verificar lo anterior, se planteó el presente trabajo de tesis, que pretende realizar un análisis sobre las limitaciones de la técnica denominada componentes principales en su aplicación al reconocimiento de rostros. Así como proponer un nuevo método que se considera, eliminará algunas de las limitaciones que se presentan en las técnicas basadas en componentes principales y en general, en los sistemas tradicionales que utilizan la métrica L_2 .

Como se ha mencionado, la tecnología en reconocimiento de rostros, se espera que tenga grandes incrementos en ventas en las próximas décadas. De los 399 millones de dólares por ventas de sistemas biométricos en el año 2000 (como se mencionó en párrafos anteriores),

\$34.4 millones en ventas correspondió a sistemas de reconocimiento de rostros. Sin embargo, lo anterior se ha incrementado sustancialmente en los últimos años. Para el año 2007, se pronostican ventas por \$429.1 millones de dólares.

Por tanto, los sistemas de reconocimiento de rostros es un nicho de mercado que no se debe perder de vista y es importante continuar con investigaciones que contribuyan a su desarrollo. A continuación, se presenta una panorámica de las temáticas que se presentan en la tesis.

Descripción general de la tesis

En el capítulo 1 se realizó el planteamiento general de la investigación, presentando las principales preguntas de investigación, la hipótesis, los objetivos de la tesis y los resultados que se pretenden obtener con el proyecto.

En el capítulo 2, se presenta el “Estado del Conocimiento” en el área de reconocimiento de rostros. Se investigaron las principales aplicaciones, técnicas y algoritmos que se han considerado relevantes desde el año 1968 al año 2004.

Como se ha mencionado, dos de los sistemas comerciales que se consideran más eficientes en la actualidad para reconocimiento de rostros, basan sus algoritmos en el uso de la técnica de componentes principales o alguna derivación de ella. En el capítulo 3, se mencionan las características básicas de los elementos matemáticos que subyacen a la técnica de

componentes principales. Asimismo, se presenta en este capítulo el algoritmo basado en componentes principales para realizar reconocimiento de rostros. A partir del cual, se desarrollará como parte del trabajo de tesis, un programa computacional para reconocimiento de rostros basado en dicha técnica y que se reportará en un capítulo posterior.

Uno de los objetivos principales de la tesis, consiste en realizar un análisis estadístico de las características que presentan los valores numéricos de los píxeles que configuran cualquier fotografía digital de un rostro. Lo anterior se presenta en el capítulo 4, en el cual se utilizó el paquete SPSS³ para llevar a cabo el análisis respectivo.

Otro elemento esencial del proyecto, consistió en verificar si los píxeles que configuran los datos de un rostro, se ajustan a una distribución normal. Lo anterior se realizó en el capítulo 5, utilizando los paquetes computacionales SPSS y LISREL/PRELIS⁴.

La verificación del supuesto de normalidad multivariada, es un elemento indispensable, los métodos estadísticos tradicionales basan sus algoritmos en el supuesto de normalidad de los datos de entrada, entre otras cosas. El autor del presente trabajo, ha postulado *a priori*, que las fotografías digitales de rostros humanos no se ajustan a una distribución normal multivariada. Si efectivamente no cumplen con una distribución normal, puede ser una de las causas de las fallas que presentan los sistemas comerciales actuales en el eficiente reconocimiento de rostros.

³ SPSS es una marca registrada de SPSS, Inc. (<http://www.spss.com>)

⁴ LISREL y PRELIS son marcas registradas de SSI Scientific Software Internacional. (<http://www.ssicentral.com>)

En el capítulo 6, se mencionan las características principales del sistema de reconocimiento de rostros basado en el análisis de componentes principales que se desarrolló utilizando MATLAB⁵ y las rutinas proporcionadas por Image Processing Toolbox⁶.

La mayor parte del capítulo, se dedica a los diversos experimentos de reconocimiento de rostros que se llevaron a cabo, utilizando principalmente las fotografías de la base AT&T⁷. Se efectuaron diversos experimentos utilizando 40 sujetos en la base de entrenamiento, en donde cada sujeto puede tener 9, 8 y 5 fotografías por cada persona en la base de datos. Y se utilizaron 52 fotografías, como los posibles sujetos a reconocer. Asimismo, se contemplaron algunas fotografías de objetos, para probar el adecuado reconocimiento del sistema.

Otro conjunto de experimentos se llevó a cabo con 33 sujetos en la base de entrenamiento, utilizando 9, 7, 5, 3, 2 y 1 fotografía por cada persona de la base de entrenamiento. Al igual que en el grupo de experimentos anteriores, se utilizaron 52 fotografías para reconocimiento.

El último conjunto de experimentos, se realizó con una base de entrenamiento que combina fotografías controladas de la base AT&T (como en los casos anteriores) e imágenes de fugitivos y terroristas que se obtuvieron del portal internet del FBI (dichas fotografías no se encuentran controladas en cuanto a luz, fondo de imagen, resolución, etc.). El propósito consiste en verificar la eficiencia del sistema de reconocimiento cuando se trabaja con imágenes que presentan diversas características. Al igual que en los casos anteriores, se

⁵ MATLAB es marca registrada de MathWorks, Inc. (<http://www.mathworks.com>)

⁶ Image processing Toolbox es marca registrada de MathWorks, Inc.

⁷ La base de fotografías AT&T fue desarrollada por los Laboratorios de Investigación AT&T – Cambridge junto con Computer Laboratories of Cambridge University, UK. Se ha utilizado en diversas investigaciones de reconocimiento de rostros a nivel internacional.

utilizan 52 fotografías para reconocer. A partir de los experimentos anteriores, se realizan diversas comparaciones y se obtienen algunas limitaciones que presenta el sistema.

Por último, en el capítulo 7 se presenta de manera general, la propuesta de un método alternativo para el reconocimiento de rostros. Dicho método, se basa esencialmente en utilizar un espacio de trabajo en L_∞ en el cual, a partir de diversas características que presentan los píxeles de la imagen digital de un rostro, se obtendrá un aproximante mediante una familia de polinomios de grado “ n ”.

Dichos polinomios caracterizarán al rostro respectivo. La obtención del aproximante en un espacio L_∞ se conceptualizará como un problema de optimización combinatoria, el cual se obtendrá utilizando Algoritmos Genéticos y la teoría de aproximación multivariada. Al utilizar un paradigma de cálculo bajo la norma L_∞ , no se requiere que los datos cumplan los presupuestos estadísticos de normalidad, homoscedasticidad y linealidad, entre otros, que se requieren en el uso de técnicas estadísticas tradicionales. Se considera que la elaboración de un sistema de reconocimiento de rostros utilizando el método propuesto, ayudará a la eficacia y eficiencia de dichos sistemas. La elaboración y programación de tal sistema, se empezará en el corto plazo y es motivo de otra investigación.

Capítulo 1

Planteamiento del Estudio

A raíz de los ataques terroristas a los Estados Unidos el 11 de septiembre de 2001, la utilización de software aplicado al reconocimiento de rostros y en general en el área de Biométrica, ha presentado gran relevancia e impulso en su investigación a partir de esa fecha.

1.1. El problema de investigación

Para el reconocimiento de rostros se han utilizado diversos métodos, entre éstos: elementos geométricos del rostro, análisis estadístico, redes neuronales, componentes principales, etc.

La técnica más utilizada en los últimos años, ha sido el de componentes principales. Uno de los paquetes de software aplicado al reconocimiento de rostros y que ha presentado resultados exitosos, utiliza el algoritmo denominado *Eigenfaces* (Sirovich y Kirby,1987; Kirby y Sirovich,1990; Turk y Pentland,1991), basado en la técnica de análisis de componentes principales. Sin embargo, aunque dicho software se considera uno de los

mejores, presenta ciertos porcentajes de error. Otros sistemas de reconocimiento de rostros, que utilizan otras técnicas derivadas del análisis de componentes principales (LFA, propuesto por Penev y Atick, 1996), también han presentando resultados inciertos.

La gran mayoría de los sistemas computacionales para reconocer rostros, comparten algo en común, sus respectivos algoritmos matemáticos trabajan en un espacio en L_1 o L_2 . Lo anterior, presupone que las características de los datos deben cumplir con ciertos supuestos estadísticos y matemáticos al aplicar las diversas técnicas de análisis multivariado que se utilizan.

El software basado en la técnica de análisis de componentes principales y algunos otros desarrollos basados en técnicas similares, presentan diversas limitaciones y por tanto, deficiencias en los resultados. Tales limitaciones se considera que se presentan por dos motivos principales: el primero, debido al propio algoritmo; en segundo lugar, al tipo de características estadísticas que presentan los valores que se obtienen de los pixeles de una imagen digital de un rostro. Se considera que los datos numéricos de los rostros utilizados en el sistema de *Eigenfaces*, viola el supuesto de normalidad multivariada, elemento importante para utilizar las técnicas de análisis con componentes principales.

Por tanto, en este trabajo se pretende conocer los elementos teóricos que subyacen a la técnica de componentes principales, para proponer y desarrollar un sistema computacional basado en dicha técnica, así como realizar un análisis estadístico de las características de las imágenes digitales de rostros y realizar diversos experimentos de reconocimiento de rostros

utilizando el sistema propuesto; con el propósito de obtener las principales ventajas y desventajas que brinda la técnica de componentes principales.

A partir de las desventajas encontradas, se propondrá de forma teórica, otro método (diferente también a las otras técnicas utilizadas actualmente) aplicable al reconocimiento de rostros.

1.2. Delimitación del problema

La base de fotografías que se utilizará para el análisis, estará limitada a 400 fotografías, correspondientes a 40 sujetos (10 fotografías por cada sujeto). Dicha base de datos, fue realizada por los Laboratorios AT&T de la Universidad de Cambridge, UK.¹

De las 400 imágenes, un determinado porcentaje se utilizará como la base de entrenamiento del sistema, y las imágenes adicionales a cada sujeto, se considerarán fotografías que se tomaron en el momento a la persona y se trata de reconocer si es un probable delincuente.

La base AT&T se ha utilizado en gran cantidad de investigaciones a nivel mundial, para probar sistemas de reconocimiento de rostros. Adicionalmente, se utilizarán otras imágenes de objetos y personas que se obtendrán de Internet y del portal del FBI (fugitivos más buscados)². Se considera que el sistema computacional a desarrollar en la presente tesis, sería una aplicación para el reconocimiento de sospechosos (delincuentes y terroristas).

¹ La base de datos de los Laboratorios AT&T (denominada “ORL Database of Faces”) se puede recuperar como archivo ZIP en: http://www.cl.cam.ac.uk/Research/DTG/attarchive.pub/data/att_faces.zip

² Las fotografías del FBI, se obtuvieron en: <http://www.fbi.gov>

1.3. Preguntas de investigación

¿Se tiene posibilidad de realizar un reconocimiento de rostros utilizando la técnica de análisis de componentes principales?

¿Las características estadísticas que presentan los píxeles de imágenes digitales de rostros cumplen con los requisitos para realizar un análisis por medio de componentes principales?

¿Cuántas fotografías por cada sujeto en la “base de entrenamiento” de rostros son requeridas para llevar a cabo un reconocimiento adecuado utilizando la técnica de componentes principales?

¿Cuáles serían las principales desventajas que presenta el análisis de componentes principales para llevar a cabo un adecuado reconocimiento de rostros?

¿Qué otra técnica sería susceptible de emplearse para llevar a cabo un adecuado reconocimiento de rostros?

1.4. Hipótesis

Utilizando una base de datos conteniendo una sola fotografía digital por persona, frontal y con elementos controlados de luz será posible llevar a cabo un adecuado reconocimiento del rostro de una persona utilizando otra fotografía digital del sujeto.

1.5. Objetivos

1.5.1. Objetivo general

Desarrollar un sistema computacional para el reconocimiento de rostros mediante aprendizaje supervisado basado en el análisis de componentes principales, con el propósito de realizar diversos experimentos para obtener las ventajas y limitaciones del procedimiento y proponer de manera conceptual una nueva alternativa de solución.

1.5.2. Objetivos particulares

- Identificar los elementos teórico-matemáticos del análisis de componentes principales y su aplicación en el reconocimiento de rostros.
- Desarrollar un sistema computacional en MATLAB basado en análisis de componentes principales para reconocimiento de rostros.
- Realizar análisis estadístico de las características principales que presentan imágenes digitales de rostros humanos.
- Elaborar una “base de entrenamiento” a partir de diversas fotografías de personas, para ser utilizada en el proceso de reconocimiento de rostros.
- Experimentar el reconocimiento de rostros por medio del sistema desarrollado, variando el número de fotografías por sujetos (de 1 a 9 imágenes), utilizadas para configurar la “base de entrenamiento”.
- Experimentar con diversos “umbrales” de la métrica de identificación, para obtener la mejor alternativa de reconocimiento.

- Obtener a partir de los resultados de experimentación, las posibles ventajas y desventajas al utilizar la técnica de análisis por componentes principales.
- Proponer de forma teórica-general, un método alternativo al de componentes principales, para reconocimiento de rostros.

1.6. Limitaciones del problema

En el reconocimiento de rostros se consideran las siguientes limitaciones:

- Se utilizará como base principal de datos, la configurada por las 400 fotografías recopiladas por los Laboratorios AT&T de la Universidad de Cambridge.

Las fotografías utilizadas para entrenar al sistema cumplen los siguientes requisitos:

- Imagen frontal
- La imagen no tendrá ningún otro elemento además del rostro, se utilizarán fotografías del tipo “pasaporte”, “licencia”.
- Las fotografías tendrán un cierto control de luz.
- Los rostros pueden presentar anteojos, barba, bigote
- Se permitirán variaciones en la inclinación del rostro
- Se permitirán ciertos cambios en la expresión facial.
- Todas las fotografías son en tonos de grises (grayscale)
- Las imágenes a reconocer, tendrán características similares a las anteriores.
- Un porcentaje reducido de fotografías para configurar alguno de los experimentos, se obtuvieron de Internet y no cumplen los requisitos de control de la base AT&T.

1.7. Resultados que se pretenden lograr con el proyecto.

- Sistema computacional de reconocimiento de rostros basado en la técnica de análisis de componentes principales y programado en MATLAB.
- Análisis estadístico descriptivo de las características de un rostro utilizando el paquete computacional SPSS.
- Probar estadísticamente que los datos digitales de un simple rostro y considerando un conjunto de rostros, no presentan una distribución normal univariada y multivariada.
- Verificar si un sistema de reconocimiento de rostros que se basa en análisis de componentes principales, funciona adecuadamente cuando únicamente se tiene una sola fotografía por persona en la base de datos de sospechosos (base de entrenamiento).
- Obtener cual sería el número idóneo de fotografías por sujeto en la base de entrenamiento, para un adecuado funcionamiento del sistema de reconocimiento de rostros basado en componentes principales.
- Verificar la eficiencia del sistema de reconocimiento de rostros, cuando se mezclan fotografías no controladas (en cuanto al fondo de la imagen, cantidad de luz, resolución, etc) y fotografías controladas.
- Ventajas y desventajas del uso de análisis de componentes principales en el reconocimiento de rostros a partir de diversos experimentos con una base de datos de rostros utilizada en diversas investigaciones.

- Propuesta teórica-general de un método para reconocimiento de rostros, como alternativa al uso de análisis de componentes principales y métodos similares, el cual no requiere los presupuestos estadísticos que se requieren al trabajar en un espacio L_2

Cápítulo 2

El “Estado del Conocimiento“ en el reconocimiento de rostros humanos por métodos computacionales

El propósito de este capítulo, consiste en presentar una revisión de la literatura acerca de los trabajos de investigación sobre reconocimiento automatizado de rostros por medio de computadoras. Se seleccionaron las investigaciones que se consideraron más relevantes y que presentan una panorámica de los métodos más famosos que se han utilizado en el reconocimiento de rostros.

Es importante señalar, que se tienen dos grandes áreas o enfoques acerca del estudio del reconocimiento de rostros:

- El psicológico y de neurociencias,
- El reconocimiento por medio de computadoras.

Algunas de las características que subyacen al reconocimiento de rostros desde el punto de vista biológico y psicológico, se han utilizado para proponer diversos métodos algorítmicos que se utilizan en el reconocimiento automatizado.

Sin embargo, el enfoque computacional, es el que más interesa en el presente trabajo. Por tanto, a lo largo del presente capítulo, se mencionan algunas cuestiones sobre el enfoque psicológico y de neurociencias; realizando un énfasis en los métodos computacionales.

2.1. Enfoque psicológico y de neurociencias en el reconocimiento de rostros en el ser humano

Esta primera área, escapa del ámbito del presente trabajo, pero es necesario remarcar, que diversos resultados de sus investigaciones han sido y son, la base para el planteamiento de diversos algoritmos computacionales que se utilizan en el procesamiento de rostros en computadora. Cabe mencionar que los enfoques psicológicos y de neurociencias han estudiado principalmente los siguientes aspectos:

- Unicidad de caras.
- Investigación acerca de que si el reconocimiento es holístico o por análisis de características locales.
- Análisis y uso de expresiones faciales para reconocimiento e investigación acerca de :
 - como los niños perciben las caras.
 - organización de la memoria para el reconocimiento de rostros.
 - desordenes en la habilidad para reconocer en forma precisa rostros.

- investigación acerca de la existencia de una gran neurona “madre” para el reconocimiento de rostros.
- papel del hemisferio derecho del cerebro en la percepción de rostros.

Entre las principales preguntas que se intentan responder en las investigaciones psicológicas y de neurociencias se encuentran:

¿Es el reconocimiento de caras un proceso en el que se tengan sistemas neuronales dedicados específicamente?

La evidencia para la existencia de un sistema en el ser humano dedicado para el procesamiento de caras, proviene de tres temáticas principales que han sido motivo de investigación por diversas personas e instituciones:

a) Las caras son más fáciles de recordar por los humanos que cualquier otro objeto cuando se les presenta en orientación invertida.

b) Los pacientes con *prosopagnosia* no pueden reconocer caras familiares, pero no presentan otra agnosia profunda. Reconocen a las personas por sus voces, color de cabello, vestimenta, etc. Aunque perciben a los ojos, nariz, boca, pelo, etc., no pueden conjuntar todos los elementos para propósitos de identificación. Reconocen si el objeto es una cara o no, pero tienen dificultad para identificar a la cara en cuestión.

c) Se dice que los bebés vienen al mundo “prelambrados” para ser atraídos por las caras. Los neonatos prefieren mirar algún objeto que tenga patrones similares a caras, que a otros objetos.

¿La percepción de la cara es resultado de un análisis holístico o de características?

Se ha encontrado que tanto la información holística (global) y las características particulares son cruciales para la percepción y reconocimiento de caras. Algunos estudios sugieren la posibilidad de que las descripciones globales sirvan como un “front end” para posteriormente, afinar la percepción basada en características.

¿Cuál es el rango de significado de las características faciales ?

La forma de la cara, el pelo, los ojos y la boca, se ha encontrado que son importantes para percibir y recordar caras. Algunos estudios muestran que la nariz tiene un papel insignificante. Aunque, existen pocos estudios acerca de “perfiles”, generalmente, el reconocimiento se realiza en imágenes frontales. Asimismo, se ha encontrado que la parte superior de la cara es más útil para el reconocimiento de caras, que la parte inferior.

Los atributos estéticos, juegan un papel importante (tales como la belleza, el atractivo y la simpatía), se ha encontrado que entre más atractiva sean las caras, es mejor su reconocimiento; las caras mucho menos atractivas le siguen en lograr su reconocimiento, seguidas por las caras comunes, en términos de la facilidad de ser reconocidas.

¿Cuál es el papel del análisis de la frecuencia espacial ?

Algunos estudios (Buhmann, Lades, Malsburg, 1990) han encontrado que la información en bandas bajas de la frecuencia espacial juegan un papel dominante en el reconocimiento de caras. Otras investigaciones (Sergent, 1986) muestran que, la tarea de juzgar el sexo se facilita con la utilización de componentes de baja frecuencia, la tarea de identificación requiere del uso de componentes de frecuencias superiores. Los componentes en baja

frecuencia contribuyen a la descripción global, los componentes de frecuencia superior contribuyen a los detalles finos en la tarea de identificación.

¿ Qué papel desempeña el Cerebro en el procesamiento de rostros ?

El papel del hemisferio derecho en la percepción de caras ha sido estudiado y soportado por diversos estudios. Con respecto a la prosopagnosia y el hemisferio derecho, los estudios han mostrado que el 73% de las víctimas tienen lesiones unilaterales del hemisferio derecho, 17% lesiones bilaterales y el 10% lesiones unilaterales del hemisferio izquierdo. Lo anterior indica considerablemente, que el hemisferio derecho se involucra grandemente con el reconocimiento de caras. Otros estudios han mostrado que el hemisferio derecho tiene ventaja en la recepción y almacenamiento de caras.

¿ Cómo se realiza el reconocimiento de caras en los niños ?

Parece ser que los niños por debajo de los diez años de edad, codifican caras no familiares utilizando características aisladas. El reconocimiento de estas caras se realiza utilizando señales derivadas de la vestimenta, tales como lentes, tipo de cabello, collares, etc.

Niños mayores de 10 años, cambian sus mecanismos de reconocimiento de características aisladas y “señales” a un análisis holístico.

¿Cuál es el papel del Género y la Raza en el reconocimiento ?

Los humanos reconocen en forma más eficiente a personas de su propia raza. Lo anterior se cree, que se debe al hecho de que los humanos pueden “codificar” una *cara promedio* con atributos *promedio*, dichos atributos pueden ser diferentes para diversas razas, haciendo más difícil el reconocimiento en personas de raza diferente.

Se han realizado estudios diversos, en los cuales se considera que dicha dificultad se debe a razones psicosociales, en las cuales las razones de una identificación deficiente se deben a efectos de prejuicio, falta de familiaridad con las clases de estímulos con otras razas, entre otras.

En cuanto a la identificación de género, estudios realizados en Japón han encontrado que el 65% de las características faciales de las mujeres (en población japonesa), son más heterogéneas que las características de los hombres. Por otra parte, las caras de mujeres blancas son levemente más variables que las de los hombres.

2.2. Enfoque Computacional

Bastantes de las teorías e hipótesis acerca de los anteriores temas se han investigado, mediante el uso de pequeños conjuntos de imágenes. Sin embargo, los diversos resultados obtenidos han servido como base para el diseño de algoritmos y sistemas para el reconocimiento de rostros a través de computadoras.

Actualmente (en los últimos cinco años aproximadamente), se han incrementado los sistemas para reconocimiento de caras, pero con pocas excepciones, la mayoría de los enfoques se han probado con conjuntos de datos relativamente pequeños (menores a 100 imágenes), únicamente se tienen algunas investigaciones que se han procesado con bases de datos de 3,000 rostros humanos. La excepción, son los sistemas FaceIt y FaceNet utilizados en aeropuertos, se creó que accede a diversas bases de datos con un gran número de fotografías (se desconoce el tamaño de la base).

A continuación se describen los principales métodos y técnicas que se han utilizado para realizar procesamiento computacional tendientes a la identificación de rostros.

2.2.1. Segmentación de rostros

Craw y otros (1987), describen un método para extraer el área de la cabeza a partir de una imagen. Utilizan una escala de imagen jerárquica y una plantilla de escala. Utilizan resoluciones de 8x8, 16x16, 32x32, 64x64, 128x128 píxeles y consideran restricciones para la localización de la cabeza.

Al nivel más bajo de resolución, una plantilla se construye del contorno de la cabeza; la dirección del borde se calcula a partir del nivel de grises de la imagen utilizando una máscara de Sobel. Una línea continua se utiliza para conectar el contorno de la cabeza.

La plantilla se utiliza para compensar posibles errores que se dan al confundir bordes de la cabeza con otros bordes de la imagen. Los resultados que se obtienen con este nivel bajo de resolución, se utiliza como guía en el siguiente nivel de resolución .

Una vez que se tiene el contorno de la cabeza, una búsqueda de características de bajo nivel tales como ojos, cejas y labios, se lleva a cabo. Los resultados que se obtienen, en general, son satisfactorios para detectar el perfil de la cabeza, la búsqueda de los ojos no es tan exitosa. Asimismo, se reporta que se obtuvieron resultados satisfactorios utilizando los métodos de Canny (1986) o el de Burr (1981), para encontrar bordes.

Craw, Tock y Bennet (1992), describen un sistema para reconocer y medir características faciales, trabajo motivado por un proyecto de indexado automatizado de fotos policíacas. Ellos tratan de localizar 40 puntos característicos a partir de una imagen de niveles de gris;

dichos puntos se seleccionaron con base a la propuesta de Shepherd (1986), el cual también fue utilizado como un criterio de juicio. El sistema utiliza una búsqueda jerárquica de grueso a fino. La plantilla se basa en el principio de transformación poligonal aleatoria propuesta por Grenander y otros (1991).

La localización aproximada, escala y orientación de la cabeza se obtiene por una deformación iterativa de la plantilla total por medio de escalamiento aleatorio, traslación y rotación.

La optimización se logra por medio de “recocido simulado” (simulated annealing) (Davis, 1987). Una vez que se logra una idea general de la localización de la cabeza, se realiza un refinamiento por transformaciones individuales de vectores del polígono (Crac, Tock, Bennett, 1992).

Los autores reportan una segmentación exitosa de la cabeza en las 50 imágenes utilizadas. En 43 de los casos se logra un perfil completamente distinguible de la cabeza, en las restantes, existen fallas para encontrar el mentón.

En lo referente a la segmentación detallado de la cara incluyendo ojos, nariz, boca, etc., se buscaron 1462 posibles puntos característicos por medio de plantillas. Los autores reportan la identificación de 1292 de dichos puntos. La única característica no encontrada fue la ceja, atribuyen el 6% de identificaciones incorrectas debido a la presencia de bigotes y barba que presentaban algunas fotografías de la base de datos. Por otra parte, el sistema requiere de gran tiempo de computación.

Govindaraju y otros (1990) consideran un modelo para localizar la cara en una imagen desordenada, su técnica utiliza una plantilla deformable que es ligeramente diferente a la propuesta por Yuille y otros (1989). Trabajan sobre los bordes de la imagen, basando en

esto su plantilla sobre el contorno de la cabeza. La plantilla se compone de tres segmentos que se obtienen a partir de las discontinuidades de curvatura del contorno de la cabeza. Estos tres segmentos forman la línea del lado derecho, la línea del lado izquierdo y la línea del cabello. Cada una de estas curvas se le asigna una quadri-tupla que consiste de la longitud de la curva, la cuerda en forma vectorial, el área encerrada entre la curva y la cuerda y el centroide de dicha área. El centro de dichos tres segmentos da la localización del centro de la cara y posteriormente, las plantillas permiten trasladar, escalar y rotar de acuerdo a ciertos modelos .

En una propuesta de Brunelli y Poggio (1992), la cara se segmenta a partir de un fondo moderadamente desordenado. El enfoque involucra trabajar con la intensidad de la imagen de la cara así, como con los bordes encontrados mediante el método de Canny. Las tareas de preprocesamiento incluyen localizar la intersección de puntos de bordes (oclusión de objetos), asignar etiquetas a segmentos contiguos de bordes y enlazar segmentos contiguos en bordes a puntos de intersección.

La cara se aproxima mediante la utilización de una elipse como herramienta analítica. Pares de segmentos de bordes etiquetados L_i, L_j se ajustan a una ecuación lineal de la elipse.

El conjunto de parámetros resultante se verifica contra la tasa de aspecto de la cara, y si es satisfactoria, se incluye en la clase del conjunto de parámetros para la selección final. Se reporta una precisión superior al 80% en el proceso cuando se aplicó a un conjunto de datos de 48 imágenes.

2.2.2. Extracción de Características

Los primeros trabajos en reconocimiento de caras fueron realizados por Sakai y otros en el año de 1969. Utilizaron una imagen digitalizada con 8 niveles de gris. El trabajo se realizó sobre un conjunto de datos consistente de imágenes frontales de caras, y se utilizó una cuadrícula de 3x3 para determinar los pixeles que presentan el mayor valor de gradiente, de tal forma que la cantidad de información se reduce a elementos esenciales. Estos pixeles se conectan a pixeles vecinos que exhiban características similares para formar líneas y segmentos de contorno. Un enfoque de “grueso a fino” se utiliza para determinar características individuales de la cara. El reconocimiento no diferencia entre caras diferentes, determina únicamente la existencia de una cara en la imagen. Los autores anotan que el procedimiento empleado presenta una dependencia de la dirección de la iluminación, cambios en ésta, causan problemas en el enfoque.

Reisfeld y Yeshurun (1992), utilizan un operador de simetría generalizado con el propósito de encontrar los ojos y boca en una cara. Su motivación proviene de la naturaleza simétrica de la cara sobre una línea vertical que pase a través de la nariz. Simetrías subsecuentes se permiten dentro de las características de ojos, nariz y boca. El operador de simetría localiza puntos en la imagen correspondientes a valores altos de una medida de simetría. No es aparente, como los autores determinan si una característica es ojo o nariz. Se indica que el procedimiento presenta una superioridad sobre otros esquemas basados en correlación, similares al de Baron (1981) en el sentido de que dicho esquema es independiente de la escala u orientación. El método es computacionalmente intenso, se mencionan tasas de éxito del 95% , con la restricción de que la cara ocupe entre el 15-60% de la imagen

Yuile, Cohen y Hallinan (1989), extraen características faciales utilizando plantillas deformables, las cuales permiten trasladar, rotar y deformar con el propósito de ajustarse a la mejor representación de la forma actual de la imagen. El preprocesamiento se realiza a la imagen de intensidad inicial para dar las representaciones de picos y valles. Filtros morfológicos se utilizan para determinar dichas representaciones. Las plantillas para el ojo tienen 11 parámetros consistentes de los arcos superiores e inferiores, el círculo para el iris, los puntos centrales y el ángulo de inclinación del ojo. Esta plantilla se ajusta a la imagen tratando de minimizar de la energía, por tanto, se determinan las funciones de energía para potencial de valle, potencial de borde, potencial de imagen, potencial de pico y potencial interno. Los coeficientes se seleccionan para cada potencial y una regla de actualización se emplea para determinar el mejor conjunto de parámetros. En los experimentos de los autores, se reporta que la localización inicial de la plantilla es crítica para determinar la localización exacta del ojo. Cuando la plantilla se inició arriba de la ceja, el algoritmo falló en la distinción entre el ojo y la ceja. Por otra parte, la complejidad computacional de este algoritmo, requiere de 5 a 10 minutos en un equipo SUN 4, una vez que se seleccionó un apropiado punto de inicio.

Nixon (1985) utilizó la Transformada de Hough para lograr reconocimiento facial. La transformada localiza formas descritas analíticamente al usar la magnitud del gradiente y la información direccional proporcionada por el operador de gradiente con el objeto de auxiliar en el proceso de reconocimiento.

Dos partes del ojo son atractivas para su reconocimiento, el perímetro redondeado del iris es atractivo, porque la detección de formas circulares es relativamente sencilla. El

perímetro de la esclera del ojo es una parte distinta y también se puede emplear en la detección. La esclera tiene la ventaja de que la forma es reflejada por la región que se encuentra debajo de las cejas.

La forma analítica que representa al iris es un círculo con direcciones de gradiente esperadas en cada cuadrante, dada la luz del fondo de la esclera. Una elipse aparece la forma más indicada para la representación aproximada de la forma del perímetro de la esclera, pero no es satisfactoria para aquellas partes alejadas del ojo a partir del centro de la cara.

La elipse se forma para ojo de la cara utilizando una función exponencial. Los gradientes de magnitud que se obtuvieron mediante un operador de Sobel, se “umbralizan” utilizando cuatro niveles de brillo para representar la dirección del gradiente en ese punto. La información de la dirección se incorpora dentro de la técnica de la transformada de Hough. El procedimiento para localizar cada ojo se restringe a la mitad de la imagen. La transformada de Hough se aplica para detectar la instancia de cada forma en un conjunto de datos de seis sujetos. La desviación de la posición del centro del iris a partir del valor estimado tiene un valor medio de 0.33 píxeles. La aplicación de la transformada de Hough para detectar el perímetro de la forma de la región debajo de las cejas aparece en promedio dar un espaciado 20% mayor que el espaciado entre los dos iris. Utilizando la Transformada de Hough para encontrar la esclera, se muestra que el espaciado difiere en promedio por menos de 1.33 píxeles. Los resultados muestran que es posible derivar una medida del espaciado por la detección de la posición de ambos iris, y la forma que describe el perímetro de la esclera y las cejas. La medida para detectar la posición del iris es más precisa. La detección del perímetro de la esclera es más sensitiva a los métodos. La detección de la posición de las cejas proporciona una medida del espaciado del ojo, que es

mayor que el proporcionado por las otras técnicas, pero que se puede utilizar cuando las otras no discriminan adecuadamente.

Hong, considera que las características de la imagen se dividen en cuatro grupos: características visuales, características estáticas de pixel, características de coeficientes de transformación y características algebraicas, éstas últimas representan los atributos intrínsecos de una imagen. La Descomposición en valores singulares (SVD – singular value decomposition) de una matriz, se utiliza para extraer las características del patrón. Los valores singulares (SV) que se extraen del SVD tienen buena ejecución como descriptores de forma. Los SV de una imagen son muy estables y representan los atributos algebraicos de la imagen, son intrínsecos pero no visibles.

Representando una imagen como un vector de características SV n-dimensional, el problema de reconocimiento se resuelve en un espacio de características n-dimensional. Una foto de una cara de un tamaño de 32 mm x 27 mm, típicamente requiere de un vector de características SV de 70-dimensiones para describirla. El vector original de características SV altamente dimensional, se puede comprimir a un espacio de características de una menor dimensión (2D o 1D) utilizando varias transformaciones.

La transformada de Foley-Sammon se utiliza para obtener el conjunto óptimo de vectores discriminantes. Con un pequeño conjunto de fotos, Hong utilizó únicamente dos de los vectores para reconocimiento. El conjunto consistió de nueve fotos de caras de tamaño 50mm x 35 mm. Cada foto se muestreo cinco veces, variando la posición relativa entre la foto y la cámara de TV, con un total de 45 imágenes, 5 imágenes en cada clase.

La operación de SVD se aplicó a cada matriz de imagen, extrayendo características SV y el correspondiente vector; el plano de discriminación óptimo y el clasificador cuadrático del patrón normal se construyó para las 45 muestras del vector de características.

El clasificador pudo reconocer las 45 muestras de entrenamiento de los nueve sujetos. La prueba se realizó utilizando 13 fotografías consistentes de nueve fotos muestreadas del conjunto original, con dos fotos de un sujeto y tres muestras del sujeto con diferente edad. Se obtuvo una tasa de error del 42.67%, debido a las limitaciones estadísticas del pequeño número de muestras de entrenamiento.

Los enfoques de clasificación basados en parámetros de la estructura, generalmente no son robustos para el reconocimiento de imágenes complejas de caras humanas. Son sensitivos a cambios en rotación, escala y expresión facial.

El desarrollo de lo que se denomina descriptores de forma (Goshtasby, 1985), auxilia en la descripción de la forma de un objeto, independientemente de la traslación o rotación. En una propuesta de Cheng y otros (1991), el vector SV se comprime en un espacio de baja dimensión por medio de diversas transformaciones, la más popular es una transformación discriminante óptima basada en el criterio de Fisher, el cual representa la proyección del conjunto de muestras sobre una dirección K , seleccionando los patrones que tienen un mínimo de dispersión dentro de cada clase y una máxima dispersión entre las clases en el espacio unidimensional. Tres vectores de características SV se extraen a partir del conjunto de entrenamiento. La transformación discriminante óptima comprime el espacio de características altamente dimensional SV a un nuevo espacio de características r -dimensional. Las nuevas características secundarias son algebraicamente independientes y

la información redundante se reduce. Este enfoque fue probado con 64 caras de ocho personas (las clases).

Las imágenes fueron representadas por *matrices de forma* de Goshtasby, que son invariantes a la traslación, rotación y escala de las imágenes faciales y se obtienen por cuantización polar de la forma (Goshtasby, 1985).

Tres fotos de cada clase fueron utilizadas para proporcionar un conjunto de entrenamiento de 24 vectores característicos SV. Dichos vectores fueron tratados con la transformación discriminante óptima para obtener nuevos vectores característicos para las 24 muestras de entrenamiento. La clase de vectores de centro se obtuvieron utilizando los segundos vectores característicos. El experimento utilizó seis vectores discriminantes óptimos. La separabilidad de las muestras de conjuntos de entrenamiento fue de un 100%. Las remanentes 40 caras, fueron utilizadas como el conjunto de prueba, cinco imágenes para cada persona.

Manjunath y otros (1992) presentan un método para la extracción de puntos característicos pertinentes a partir de la imagen de una cara. Emplea la descomposición de ondeleta (wavelet) de Gabor y una interacción de escala local para extraer características en puntos de curvatura máxima en la imagen, correspondiente a la orientación y vecinos locales. Dichos puntos se almacenan en una base de datos y subsecuentes imágenes objetivo se aparean utilizando una técnica de apareamiento de grafos.

Un enfoque estadístico para detectar y reconocer el ojo humano en una imagen de intensidad, con la restricción de que la cara se encuentre de forma frontal, se describe en una investigación de Hallinan (1991). Este método, elaborado por Hallinan utiliza un

enfoque basado en plantilla, la cual cuenta con dos regiones de intensidad uniforme. La primera es la región del iris y la otra, es la región blanca del ojo. El enfoque construye un ojo “arquetipo” y modela varias distribuciones como variaciones de esta. Para el ojo ideal, se selecciona una intensidad uniforme, tanto para el iris y lo blanco. Los resultados que se obtuvieron con 25 imágenes utilizadas como un conjunto prueba, dio 107 ojos positivos.

Conlin (1986), describe un sistema de visión basado en el conocimiento para detectar caras humanas a partir de croquis realizados “a mano”. El sistema emplea reglas del tipo IF-THEN para realizar los proceso: IF línea-parte-superior-boca No se encuentra, Pero línea-inferior se encuentra THEN mirar por la línea superior de la boca en el área de la imagen directamente arriba .

La plantilla para la cara consiste de ambos ojos, la nariz y la boca. El procesamiento se realiza sobre cuatro diferentes niveles de abstracción de la información de la imagen: segmento lineal, parte componente, componente y cara. Los segmentos de línea son seleccionados como candidatos de partes componentes con valores probabilísticos asociados. Un componente tratará de ver si un área particular en la imagen tiene las partes componentes necesarias (en orientación correcta relativa a la otra) y determina la existencia del componente. El nivel de cara tratará de determinar la geometría del componente que mejor se ajusta para describir una cara a partir de los datos de la imagen

La estructura del sistema se basa en una arquitectura de pizarrón, todas las tareas tienen acceso al pizarrón, y un controlador activa las tareas de acuerdo a la información disponible. El autor reporta buenos resultados en la detección de caras y comenta, que la modularidad permite expandir las fuentes de conocimiento referentes a cejas, orejas, mentón, etc.

2.2.3. Reconocimiento de Rostros

Kaya y Kobayashi (1972) presentan un estudio básico en la clasificación de caras utilizando vistas frontales, con boca cerrada, sin barba, sin anteojos. Se utilizan distancias euclidianas entre puntos similares sobre la cara como parámetros, para la caracterización respectiva. Dichos parámetros son resistentes a cambios en luz y grado de desarrollo en pequeños cambios de la expresión facial. Kaya y otros estiman que el número de parámetros debe ser mayor que $\log_2 N$ bits, en donde N es igual al número de caras a clasificar. Los parámetros se normalizan dividiéndolos entre la longitud de la nariz, para contabilizar cualquier diferencia debida al tamaño de la fotografía y la distancia del sujeto a la cámara.

Se utilizaron fotos de 62 japoneses adultos entre los 20 y los 30 años de edad y bajo las mismas condiciones de luz. Los parámetros característicos fueron medidos a mano, y se calculó la media y la desviación estándar. La correlación de los parámetros indica que la actual dimensión del vector de parámetros puede ser más pequeña de 9.

Una de las métricas utilizadas con gran efectividad de clasificación es el número promedio de parámetros utilizados para identificar la cara.

Uno de los métodos de caracterización de la cara es el uso de parámetros geométricos (distancias y ángulos entre puntos, tales como las corneas del ojo, extremidades de la boca, mentón, etc.). Los datos utilizados por Kanade (1977), consisten de 17 hombres y 3 mujeres. Dos fotos de cada uno se tomaron, la segunda, un mes después.

Los puntos característicos se localizan en dos fases, un estado de “grueso a fino” simplifica la operación diferencial y los algoritmos para encontrar características. Una vez que los

ojos, nariz y boca son aproximadamente localizados, información más precisa se extrae al confinar el procesamiento a cuatro regiones pequeñas.

Estas cuatro regiones son los ojos izquierdo y derecho, nariz y boca. La intensidad se basa en el histograma de área local que se obtuvo en el estado de “grueso a fino”.

Después de la adquisición fina de los datos de la imagen, información más precisa se extrae de cada región utilizando un proceso de “thresholding” (umbral), diferenciación y proyección integral. Kanade utiliza un conjunto de 16 parámetros faciales que son proporciones de distancias, áreas y ángulos, para compensar la variación de tamaño de las fotos. Para eliminar diferencias de escala y dimensión los componentes del vector resultante se normalizan.

El conjunto entero de 40 imágenes se procesa y una foto de cada individuo se utiliza en el conjunto de referencia, las remanentes 20 fotos se utilizan como conjunto de prueba. Se utilizó una simple medida de distancia para verificar la similitud entre una imagen del conjunto de prueba y una imagen en el conjunto de referencia; se obtuvo una precisión entre el 45% y el 75%, dependiendo de los parámetros utilizados.

Recientemente, el uso de la expansión denominada “Karhunen-Loeve” o KL (denominación que también se le ha dado a la técnica de análisis de componentes principales), se ha utilizado para la representación de rostros (Kirby y Sirovich, 1990; Sirovich y Kirby, 1987) y el reconocimiento de rostros (Turk y Pentland, 1991; Pentland y otros, 1994). Dicha técnica ha generado nuevos intereses de investigación y en la actualidad, los dos principales sistemas de reconocimiento de rostros se basan en la técnica de componentes principales y sus variaciones. La expansión KL se ha utilizado para compresión de imágenes desde hace más de treinta años, su utilización en el

reconocimiento de patrones se ha dado en pocas ocasiones (Jain, 1989). Una de las razones por las cuales la expansión KL no se había utilizado, es su complejidad computacional.

Sirovich y Kirby (1987) retoman el problema de representar imágenes mediante componentes principales, utilizando caras sin orejas. Se puede notar que el número de imágenes “M” para calcular la matriz de covariancia de los datos, es mucho menor que la dimensionalidad de renglones o columnas de dicha matriz, generando que la matriz sea singular. Los autores utilizan un método estándar del álgebra lineal que calcula únicamente los M eigenvectores que no pertenecen al espacio nulo de la matriz degenerada. Una vez que los eigenvectores (referidos como eigenfotos) se obtienen, cualquier imagen en el ensamble se puede reconstruir aproximadamente utilizando una combinación pesada de eigenfotos.

En trabajos posteriores, Kirby y Sirovich (1990) incluyen la simetría de las caras en la representación de eigenfotos de las caras, utilizando un ensamble extendido de imágenes, consistentes de las caras originales y sus imágenes espejo.

Entre los métodos que más relevancia tuvieron a principios de la década de los noventas y que sirvieron para desarrollos posteriores, es el propuesto por Brunelli y Poggio (1993). Dichos investigadores han realizado métodos para el reconocimiento de rostros a partir de la obtención de características geométricas (figura 3.a). A partir de 35 características se realiza el reconocimiento mediante un clasificador bayesiano.

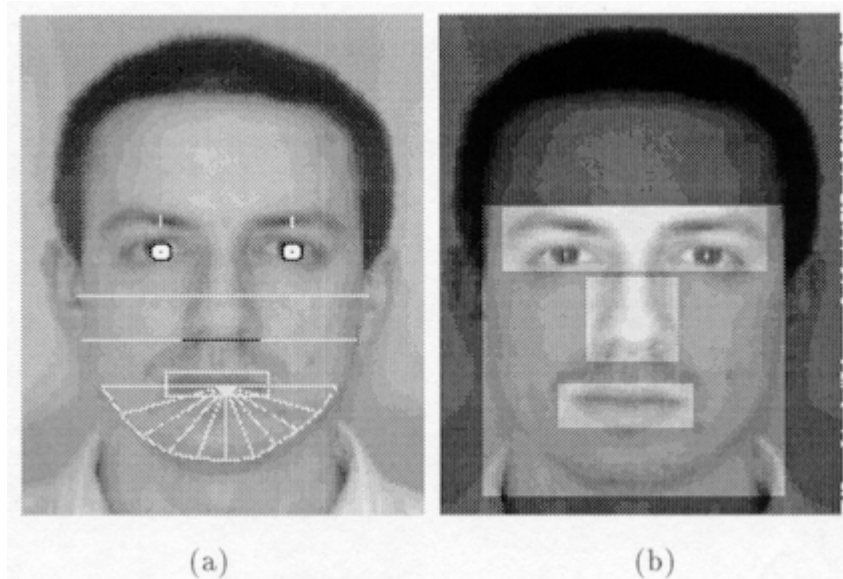


FIGURA 3. RECONOCIMIENTO DE ROSTROS UTILIZANDO
(a) Características Geométricas (b) Plantillas

Otro método muy utilizado hasta la fecha, consiste en la utilización de “plantillas” (máscaras), en el cual se configuran diversas plantillas que abarcan ciertas regiones del rostro: ojos, nariz, boca, etc., como se puede observar en la figura 3.b. La imagen a reconocer se compara a través de las plantillas con las imágenes de la base de datos, y se obtiene un vector con las puntuaciones de apareamiento mediante procesos de correlación (Brunelli y Poggio, 1994).

Un método también ampliamente utilizado en la actualidad, consiste en la utilización del denominado Análisis de Componentes Principales. El cual, es una técnica estadística que se utiliza para la reducción de la dimensionalidad de los datos, preservando la información relevante. Hancock, Bruce y Burton (1996, 1998), proponen la utilización de componentes principales a partir de 38 puntos geométricos del rostro (figura 4).

Por su parte, Lucas (1994, 1995) propone el método denominado Clasificador de *n-tuplas*, el cual consiste en la selección de píxeles por medio de un muestreo de “*n-tuplas*” de éstos (obteniendo únicamente valores del nivel de gris). Cada *tupla* presenta una longitud de “*m*”. Por ejemplo, en la figura 5, se representa el conjunto de *n-tuplas* ($n=3$), con una longitud por tupla de $m=3$. El reconocimiento se efectúa mediante la utilización de una métrica (distancia Manhattan), para comparar el conjunto de entrenamiento con la imagen a clasificar.

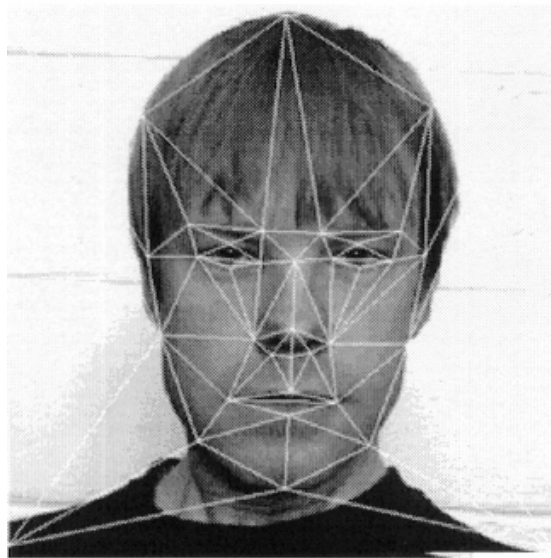


FIGURA 4. RECONOCIMIENTO UTILIZANDO PUNTOS SELECCIONADOS DEL ROSTRO PARA SER PROCESADOS POR LA TÉCNICA DE COMPONENTES PRINCIPALES

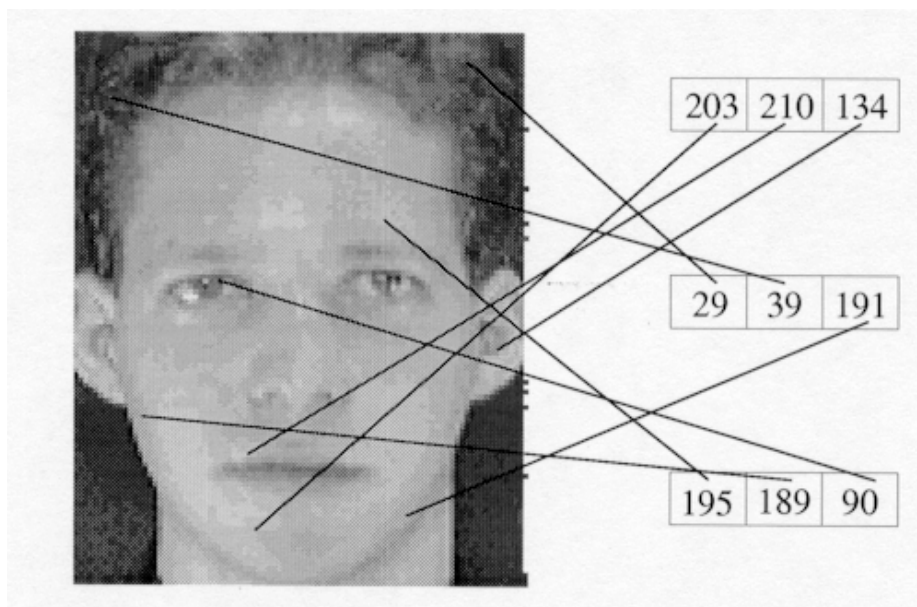
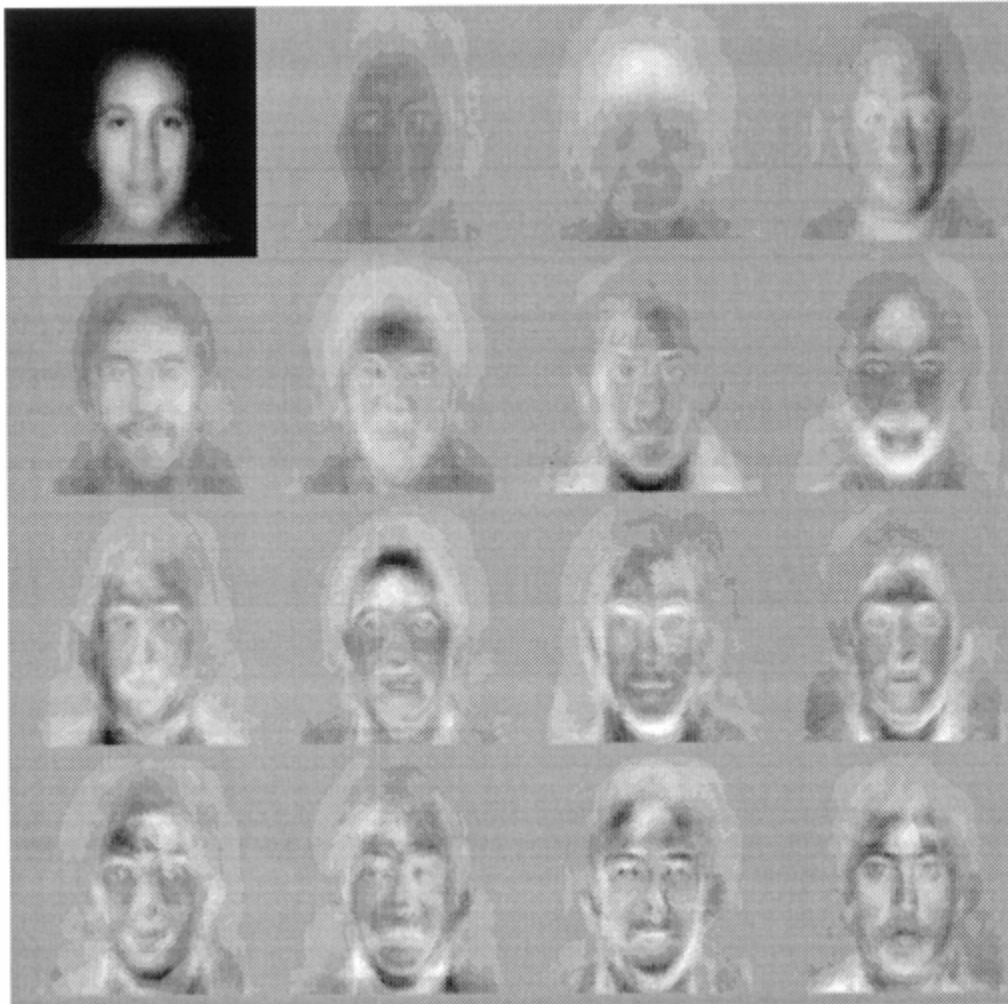


FIGURA 5. RECONOCIMIENTO UTILIZANDO UN “CLASIFICADOR POR MUESTREO” A PARTIR DE “n-tuplas” QUE SE OBTIENEN DE LOS VALORES DE PÍXELES

Turk y Pentland (1991) utilizan eigenfotos (conocidas como **eigenfaces**) para la detección de caras y su respectiva identificación (figura 6). Dadas las eigenfaces, toda cara en la base de datos se puede representar como un vector de pesos, los pesos se obtienen por la proyección de la imagen en los componentes de la eigenface por medio de un simple producto punto.

Cuando una nueva imagen prueba cuya identificación se requiere es dada, la nueva imagen también se representa por su vector de pesos. La identificación de la imagen prueba se realiza al localizar la imagen en la base de datos cuyos pesos son los más cercanos (en distancia euclídeana), a los pesos de la imagen prueba.



**FIGURA 6. RECONOCIMIENTO DE ROSTROS UTILIZANDO EL METODO “EIGENFACES”
PROPUESTO POR TURK Y PENTLAND**

Al utilizar la observación de que la proyección de la imagen de una cara y la imagen que no sea una cara, son muy diferentes, se tiene un método para detectar la presencia de caras en una determinada imagen.

Los autores utilizan una base de datos de 3000 caras, correspondientes a 16 sujetos, digitalizando todas las combinaciones de tres orientaciones de la cabeza, tres tamaños de la

cabeza y tres condiciones de luz. Se reporta que el enfoque es robusto para cambios en la condición de la luz, pero se degrada rápidamente ante cambios de la escala.

Akamatsu y otros (1991), utilizan el método KL para la extracción de características a partir de imágenes de caras, éste método, se combina con otras dos operaciones para mejorar la ejecución de la técnica de extracción en la clasificación de caras con vista frontal. La aplicación de la expansión KL directamente sobre una imagen facial sin estandarización, no logra robustez ante variaciones en la adquisición de la imagen. El método reportado por Akamatsu, utiliza estandarización de la posición y tamaño de la cara. Los puntos centrales son las regiones correspondientes a los ojos y boca.

Cada imagen objetivo se traslada, escala y rota a través de transformaciones afines, tal que, los puntos de referencia de los ojos y boca estén en un arreglo espacial específico con una distancia constante. La expansión KL aplicada a la estandarización de imágenes de caras se conoce como la transformada Karhunen-Loeve de patrón de intensidad en imagen objetivo de transformada afín (KL-IPAT).

El KL-IPAT se extrajo de 269 imágenes con 100 eigenfaces. El reconocimiento de la imagen mejoró, comparado con el enfoque de eigenface utilizando KL sobre la imagen. Como segundo paso se aplica la transformada de Fourier a la imagen estandarizada y utiliza el espectro de Fourier resultante en vez del dominio espacial de la imagen estandarizada. La expansión KL que se aplica al espectro de Fourier es denominado como la transformada Karhunen-Loeve del espectro de Fourier en la imagen objetivo de la transformada afín (KL-FSAT). La robustez del KL-IPAT y el KL-FSAT se verificó con variaciones geométricas utilizando las características estándar de 269 imágenes de caras. En el primer

experimento, las muestras de entrenamiento y prueba fueron adquiridas bajo condiciones tan similares como fue posible. El conjunto prueba consistió de cinco muestras a partir de 20 individuos. El KL-IPAT tuvo una precisión del 85% y el KL-FSAT una precisión del 91%. Ambos métodos identificaron erróneamente el ejemplo en donde existía diferencia en el uso de anteojos, entre el conjunto prueba y el conjunto de entrenamiento. El KL-FSAT mostró gran robustez sobre el KL-IPAT ante diferentes orientaciones de la cabeza. Se realizó un buen reconocimiento al restringir los parámetros de adquisición de imagen. Ambos métodos presentan dificultades cuando la orientación de la cabeza muestra variaciones.

Pentland y otros (1994), extendieron las capacidades de sistemas anteriores en varias direcciones. Reportan extensas pruebas basadas sobre 7562 correspondientes a 3000 personas, la mayor base de datos sobre la cual se ha reportado un estudio de reconocimiento de caras, hasta la fecha.

Veinte eigenvectores fueron calculados utilizando un subconjunto seleccionado aleatoriamente de 128 imágenes. Adicionalmente a la eigen-representación, se tiene información acerca de la raza, sexo, edad aproximada y expresión facial se incluyó. Similar a las aplicaciones de áreas de la policía, únicamente se guardó la vista frontal y de perfil.

Una de las aplicaciones que los autores consideran, es la búsqueda interactiva a través de la base de datos. Cuando se le pregunta al sistema por caras de ciertos tipos de personas (por ejemplo, mujeres menores a 30 años), imágenes que satisfacen dicha pregunta se presentan en grupos de 21 fotos. Cuando el usuario selecciona una de las imágenes, el sistema

presenta caras que son similares, para un nuevo reconocimiento por parte del usuario. En una prueba involucrando 200 imágenes, se logró un 95% de precisión en el reconocimiento. Para evaluar la precisión en el reconocimiento como función de la raza, imágenes de adultos masculinos de raza blanca, negra y asiática fueron probadas. Para blancos y negros se logró una precisión entre el 90% y el 95% respectivamente, y una precisión del 80% para los asiáticos.

En aplicaciones en que se cuenta con más de dos vistas de la cara, se pueden realizar principalmente dos enfoques para su manejo. El primero, consiste en manejar en conjunto todas las imágenes y construir un conjunto de eigenfaces que representen a todas, desde todas las vistas. El otro enfoque, es usar eigenspaces separados para diferentes vistas, de tal forma que la colección de imágenes tomadas para cada vista, tengan su propio eigenspace. Este segundo enfoque, conocido como el eigenspace basado en vistas, ha presentado mejores resultados.

El concepto de eigenfaces se puede extender a eigenfeatures, tales como eigenojos, eigenboca, etc. De tal forma, que así como las eigenfaces se utilizaron para detectar la presencia de la cara (Reisfeld y Yeshurun, 1992), los eigenfeatures se utilizan para detectar características (ojos, nariz, boca, etc.). Las tasas de detección reportadas son del 94%, 80% y 56% para los ojos, nariz y boca, respectivamente, sobre una base de datos de 7562 imágenes.

Utilizando un conjunto limitado de imágenes (45 personas, con dos vistas por persona correspondientes a diferentes expresiones faciales, tales como neutral y sonriendo), se

realizaron experimentos de reconocimiento como una función del número de eigenvectores para eigenfaces únicamente y para la representación combinada. Los eigenfeatures y eigenfaces se ejecutaron adecuadamente para espacios de orden bajo. Al combinarse ambos, únicamente se obtuvo una mejora marginal.

El uso de líneas de isodensidad (curvas de nivel constante de gris), para reconocimiento de caras se ha investigado en muy pocas ocasiones (Nakamura, Mathur, Minami, 1991). Tales líneas sin embargo, no están directamente relacionadas a la estructura tridimensional de la cara, pero proporcionan una imagen en relieve de ésta. Utilizando imágenes de caras que se tomaron con un fondo negro, un operador de Sobel y algunos pasos de postprocesamiento se utilizan para obtener el contorno de la región de la cara. Un histograma de nivel de gris (8 bits) se utiliza para trazar líneas de contorno sobre los niveles de isodensidad. Un procedimiento de plantilla de apareamiento se utiliza para el reconocimiento.

El método se utilizó con 10 pares de imágenes de caras, con tres pares de fotos de hombres con anteojos, dos pares de fotos de hombres con barba y dos pares de fotos de mujeres. Se reporta un adecuado reconocimiento.

La utilización de redes neuronales en el reconocimiento de caras se ha dirigido a diversos problemas: clasificación de género, reconocimiento de la cara, clasificación de expresión facial. Una de las más antiguas aplicaciones se reporta en el *mapa asociativo* de Kohonen (1988). Utilizando un pequeño conjunto de caras, se reporta una precisión cuando la imagen de entrada presenta bastante “ruido” o cuando porciones de la imagen se omiten.

Una simple capa adaptativa (una por cada persona en la base de datos) para reconocimiento de caras, análisis de expresión y verificación de cara, se reporta en un estudio de Stonham (1986). Este sistema se denominó WISARD (Wilkie, Aleksander and Stonham's Recognition Device), el sistema necesita de 200 a 400 presentaciones para entrenamiento de cada clasificador, los patrones de entrenamiento incluyen traslación y variación en expresión facial. 16 clasificadores se utilizaron para construir el conjunto de datos utilizando 16 personas. La clasificación se logró, determinando el clasificador que proporciona la más alta respuesta para la imagen de entrada. Se presentan extensiones de la técnica para verificación de caras y análisis de expresión. Sin embargo, el reducido tamaño de la muestra no permite realizar conclusiones adecuadas sobre la viabilidad de este enfoque aplicado a grandes conjuntos de personas en bases de datos.

Golomb y Sejnowski (1991), utilizan cascadas de dos redes neuronales para clasificación de genero. El primer estado es una red neuronal para comprimir imágenes, cuyos nodos ocultos sirven como entradas a la segunda red, la cual ejecuta la clasificación del genero.

Ambas redes se encuentran totalmente conectadas y redes de tres capas, se entrenan por un algoritmo estándar "back-propagation". Las imágenes utilizadas para la prueba y entrenamiento fueron adquiridas de tal forma que se evita el pelo facial, joyería, maquillaje, fueron preprocesadas de tal forma que los ojos están a nivel, así como la boca. Un bloque sin orejas de 30 x 30 pixeles se extrajo para entrenamiento y prueba. El conjunto de datos consistió de 45 hombres y 45 mujeres, 80 se utilizaron para entrenamiento, sirviendo 10 como ejemplos de prueba

La compresión de la red indirectamente sirve como un extractor de características, en el que la actividades de los 40 nodos ocultos (en una red de $900 \times 40 \times 900$), sirven como características para la segunda red, la cual ejecuta clasificación de género.

La red para clasificación de género es una red de $40 \times n \times 1$, en donde el número “n” de nodos ocultos es 2, 5, 10, 20 o 40. Experimentos con 80 imágenes de entrenamiento y 10 imágenes de prueba han mostrado la factibilidad del enfoque,

Utilizando un vector de 16 atributos numéricos, tales como ancho de la ceja, ancho de la nariz y boca, etc., Brunelli y Poggio (1992) desarrollaron una red neuronal para la clasificación de género. Los autores entrenaron dos redes HyperBF (Poggio y Girosi, 1990), una para cada tipo de género. Las imágenes de entrada se normalizaron respecto a la escala y rotación, utilizando las posiciones de los ojos, los cuales son detectados de manera automática. El vector de características de 16 dimensiones se extrae también de manera automática.

Las salidas de las dos redes HyperBF se comparan, la etiqueta de género para la imagen de prueba se decide por la red con mayor salida. En experimentos de clasificación, únicamente un subconjunto de vector con 16 características se utilizó. La base de datos consistió de 21 hombres y 21 mujeres. La estrategia “leave-one-out” (Fukunaga, 1989) se utilizó para la clasificación. Cuando el vector de características a partir del conjunto de entrenamiento se utilizó como vector prueba se logró una precisión del 92.5%, para caras que no se encontraban en el conjunto de entrenamiento, la precisión fue del 87.5%.

La clasificación de género se ha extendido al reconocimiento de caras, utilizando una red HyperBF por persona y un vector de características de 35 dimensiones. La motivación para

la estructura subyacente es el concepto de una gran neurona “abuela”, una simple neurona (función Gausiana en la red HyperBF) .

Se generó una base de datos “sintética” a falta de imágenes reales de personas, dicha generación se realizó mediante la perturbación del promedio de vectores característicos de personas disponibles y las personas disponibles fueron utilizadas como muestras de prueba. Para diferentes conjuntos de parámetros (coeficientes, centros y métricas de las HyperBF), los resultados de clasificación han sido reportados.

Otros usos de redes HyperBF aplicadas al reconocimiento de caras se reporta en Brunelli y Poggio (1992). En esta investigación se utilizan para remover variaciones debidas a cambios en el punto de vista. Las imágenes se transforman utilizando transformaciones afines 2-D, los parámetros de transformación se obtienen utilizando las posiciones detectadas de los ojos y boca en la imagen y las posiciones deseadas de estas características. La imagen transformada se somete a un operador direccional para reducir los efectos de iluminación y la imagen resultante, se multiplica por una función Gausiana y es integrada sobre un campo receptivo para lograr reducción de la dimensionalidad.

El laboratorio Media Lab del MIT utiliza una base de datos de 27 imágenes de cada una de 16 diferentes personas, utilizando imágenes de 17 personas para entrenamiento, y las restantes para muestras de prueba. Se reportaron razonables resultados.

Rahardja, Sowmya y Wilson (1991), presentan los resultados utilizando un modelo conexionista de expresión facial. El modelo utiliza la estructura de pirámide para representar datos de la imagen. Cada nivel de la pirámide se representa por una red consistente de una entrada, una capa oculta y una capa de salida. Las capas de entrada de

los niveles medios de la pirámide son las salidas de los niveles previos de las unidades ocultas cuando el entrenamiento se completa. El entrenamiento de la red a su más bajo nivel se lleva a cabo en forma convencional. Cada red se entrena utilizando una variación rápida del algoritmo backpropagation, el conjunto de patrones de entrenamiento para los subsecuentes niveles se obtiene al combinar y particionar las salidas de las unidades ocultas del nivel precedente.

Las imágenes originales del conjunto de entrenamiento se particionan en bloques de cuadrados traslapados, estos bloques simulan los campos receptivos locales del sistema visual humano. Cada bloque consiste del conjunto de bloques particionados en las mismas posiciones sobre la imagen del conjunto de patrones. El conjunto de entrenamiento consiste de seis caras dibujadas a mano con seis diferentes expresiones.

La red presenta cuatro niveles, niveles 1-3 consiste de 25 unidades de entrada, seis unidades ocultas y 25 unidades de salida. El cuarto nivel tiene 18 unidades de entrada, 8 unidades ocultas y 25 unidades de salida. El último nivel de la pirámide tiene la representación más abstracta.

La red únicamente reconoce el 50% de las expresiones faciales.

Los sistemas presentados por Buhmann, Lades y Malsburg (1990), así como la investigación de Lades y otros (1993) se basan en la denominada Arquitectura de liga dinámica (DLA). Dicha arquitectura intenta solucionar algunos de los problemas conceptuales de las redes neuronales convencionales, siendo el problema más prominente la expresión de interrelación sintáctica en redes neuronales. DLA utiliza plasticidad sináptica y esta en posibilidad de formar instantáneamente conjuntos de neuronas agrupadas en grafos estructurados y mantener las ventajas de sistemas neuronales. Una DLA permite

generalización sobre un grandes grupos de operaciones simétricas y la adquisición de nuevos objetos por aprendizaje “one-shot”, reduciendo los pasos y el tiempo de aprendizaje. Se utilizan filtros Gabor basado en wavelets como detectores de características, caracterizados por sus frecuencias, posición y orientación. Dos transformadas no lineales se utilizan para ayudar durante el proceso de apareamiento. Se requiere un mínimo de dos niveles, el dominio de la imagen y el dominio del modelo. Se supone que el dominio de la imagen corresponde a las áreas corticales visuales primarias y el modelo de dominio a la corteza intertemporal en el campo de la visión biológica.

El dominio de la imagen consiste de un arreglo 2-D de nodos y cada nodo en la posición x , consiste de F diferentes detectores de características (neuronas), que proporcionan descriptores locales de la imagen. Las imágenes se representan como grafos con atributos, en donde los atributos asignados a los nodos del grafo son vectores de actividad de detectores de características locales. Un objeto en la imagen se representa por un subgrafo del dominio de la imagen.

El modelo de dominio es un ensamblado de todos los grafos con atributo, copias idealizadas de subgrafos en el dominio de la imagen.

Durante el proceso de reconocimiento un objeto se selecciona a partir del dominio del modelo, una copia del grafo del modelo se posiciona en el centro del dominio de la imagen. La calidad del apareamiento se evalúa utilizando una función de costo. Si el costo total se reduce, el nuevo valor se acepta, lo anterior se repite hasta que el costo óptimo se encuentra. El reconocimiento tiene lugar después de que el costo total óptimo se determina para cada objeto.

El objeto con el mejor apareamiento a la imagen se determina, la identificación es un proceso de apareamiento de una grafo elástico. El sistema identifica la cara de una persona

al comparar un grafo que se extrajo con respecto a un conjunto de grafos almacenados. Un experimento consistió de un conjunto de más de 40 diferentes caras, con un pequeño esfuerzo de estandarizar las imágenes, el sistema reconoció en forma consistente (Buhmann, Lades, Malsburg, 1990).

El sistema utilizado por Lades y otros (1993) tiene un conjunto de caras mucho mayor y las reconoció bajo diferentes tipos de distorsión y rotación en profundidad, logrando menos del 5% de falsos

Manjunath y otros (1992) almacenaron puntos característicos que se detectaron mediante descomposición “wavelet Gabor” para cada imagen, lo cual reduce considerablemente los requerimientos de almacenamiento en la base de datos. Típicamente 35 a 45 puntos por imagen se generan y almacenan. El proceso de identificación utiliza la información presente en una representación gráfica topológica de los puntos característicos. Después de compensar por localizaciones de centroídes diferentes, dos funciones de costo se evalúan, una es el costo topológico y la otra, un costo de similitud. Se reporta una precisión del 94%, el método muestra dependencia de la dirección de iluminación y debe trabajar sobre imágenes con fondo controlado (pasaporte, licencias).

Cheng y otros (1991) desarrollaron un método algebraico para reconocimiento de caras utilizando SVD y “umbralizando” los eigenvalores así obtenidos a algún valor mayor que el del umbral. Utiliza un análisis proyectivo con el conjunto de entrenamiento de las imágenes sirviendo como el espacio de proyección. El conjunto de entrenamiento en sus experimentos consistió de tres instancias de caras de la misma persona.

Si $A \in \mathfrak{R}^{m \times n}$ representando la imagen, y $A_j^{(i)}$ representa la j -ésima cara de la persona i , entonces la imagen promedio para la persona i se encuentra dada por:

$$\frac{1}{N} \sum_{j=1}^N A_j^{(i)} \quad (1)$$

Los eigenvalores y eigenvectores se determinan por medio de esta imagen promedio utilizando SVD. Una imagen prueba se proyecta sobre el espacio determinado por los eigenvectores, la norma de Frobenius se utiliza como un criterio para determinar que persona pertenece a la imagen prueba. Los autores han reportado una precisión del 100% trabajando con una base de datos de 64 caras de 8 diferentes personas. Cada persona presentó 8 fotos, tres imágenes de cada persona se utilizaron para determinar el vector de características para la imagen respectiva.

Seibert y Waxman (1991) han propuesto un sistema para reconocimiento de caras a partir de sus partes utilizando una red neuronal. El sistema es similar a un sistema modular desarrollado para reconocer objetos en 3-D, a partir de la combinación de vistas 2-D desde diversos puntos, en el caso de las caras, los arreglos de características tales como ojos y nariz, tienen un papel importante en las vistas 2-D. Los pasos del procesamiento son: segmentación de la región de la cara utilizando técnicas de detección de cambio entre frames, extracción de características como ojos, boca, etc. utilizando detección de simetría, agrupamiento y mapeo log-polar de las características y sus atributos tales como centroides, codificación de arreglos de características, *clustering* de vectores característicos en categorías de visión utilizando ART 2, e integración de evidencia acumulada utilizando una red de aspecto.

Por su parte, Yang y otros (1993), realizaron un sistema para detectar y reconocer caras en imágenes monocromáticas, en primer lugar, un algoritmo basado en reglas se utiliza para localizar caras en la imagen. Posteriormente, cada cara se reconoce por medio de una estructura similar a una red neuronal denominada Cresceptron (Wenhg, Ahuja, Huang, 1993). El Cresceptron tiene una estructura piramidal de multiresolución. En experimentos pequeños, involucrando 50 personas, el método funcionó adecuadamente.

2.2.4. Reconocimiento de rostros a partir de perfiles

La investigación en esta área es generalmente aplicada a requerimientos de identificación policiaca. Las imágenes de perfil proporcionan información estructural detallada acerca de la cara, la cual no se puede obtener en imágenes frontales. Específicamente, el tamaño y orientación de la nariz, así como la obtención de puntos locales de interés. El reconocimiento involucra la determinación de interrelaciones entre dichos puntos.

Kaufman y Breeding (1976) desarrollaron un sistema de reconocimiento a partir de siluetas de perfil. La imagen adquirida por una cámara de TV blanco y negro, se genera una imagen binaria, con el negro correspondiente a la región de la cara. Un paso de preprocesamiento extrae la porción frontal de la silueta que limita la imagen de la cara. Un conjunto de autocorrelaciones normalizadas expresada en coordenadas polares, se utiliza como un vector de características, una regla del tipo “K-vecinos próximos” con distancia “pesada” se utiliza para la clasificación. Se realizaron experimentos con 120 perfiles correspondientes a 10 personas, la mitad fue utilizada para entrenamiento.

Un conjunto de 20 características de autocorrelación se utilizó como un vector de características. Se realizaron tres conjuntos de experimentos, en los primeros dos, 60 muestras de entrenamiento seleccionadas de forma aleatoria se utilizaron. En el tercer experimento, 90 muestras se utilizaron en el conjunto de entrenamiento. Se reportan precisiones del 90% logrado con las 90 muestras almacenadas en el conjunto de entrenamiento y con una dimensionalidad del vector de características de 4. Comparaciones con las características derivadas a partir de las invariantes de momento mostraron que las autocorrelaciones circulares se ejecutaron mejor.

Harmon y Hunt (1977) presentan un sistema de reconocimiento semiautomático para reconocimiento de perfiles, tratando el problema como un apareamiento de “waveform”. Las fotos de perfil de 256 hombres se procesaron manualmente reduciendo las curvas de contorno. A partir de estas curvas, se obtuvo un conjunto de 9 marcas, un conjunto de 6 marcas características se derivaron. Y un total de 11 características numéricas se extrajeron a partir de las características mencionadas anteriormente.

Por su parte, Gutta y otros (1996) utilizando los métodos de extracción de características desarrollados por Grötschel y Lobas (1993), los utilizan para crear 11 componentes del vector de características. La clasificación se realizó basándose en distancias euclidianas y particionamiento de conjuntos. El particionamiento se utilizó para reducir el número de candidatos para incluir en las medidas euclidianas y reducir el tiempo de ejecución. Se reporta un sistema robusto para identificar un perfil desconocido.

Harmon y otros (1981), presentan una continuación de su investigación de años anteriores (Harmon y Hunt, 1977; Harmon, 1973). El objetivo es desarrollar procedimientos robustos y económicos para sistemas de tiempo real de identificación de perfiles de caras. El trabajo define 17 puntos que parecen ser la mejor combinación para lograr el reconocimiento. Se utiliza distancia euclideana mínima entre el archivo conocido y el de referencia, asimismo, se usan ventanas para reducir la población durante la búsqueda, estas ventanas se basan en el vector promedio que se obtuvo de múltiples muestras de perfiles individuales. Se obtienen resultados del 96% .

Wu y Huang (1990) reportan también un sistema para reconocimiento de perfiles utilizando enfoques similares al de Harmon (1977). Primero, a partir de los perfiles se utilizan B-splines para extraer seis puntos de interés (pico de la nariz, fondo de la nariz, punto de la boca, mentón y punto del ojo). Un vector de características con dimensión de 24, se construye a partir del cálculo de distancias entre dos puntos vecinos, longitud, ángulo entre segmentos de curvatura juntando dos puntos adyacentes, etc. El reconocimiento se realiza al comparar el vector de características que se extrajo contra la imagen de prueba con vectores almacenados utilizando un método de búsqueda secuencial y una norma absoluta. Las características almacenadas se obtienen a partir de tres instancias de los perfiles de personas, en todas, se utilizaron 18 personas para la fase de entrenamiento.

2.2.5. Métodos Evolutivos

Liu y Wechsler (1998), proponen una combinación de Análisis de componentes principales con un enfoque evolutivo. Se utiliza un Algoritmo Genético con el propósito de buscar los

mejores ejes de rotación definidos por el Análisis de componentes principales, con el propósito de encontrar los ejes que mejor realicen la separación de clases. Reportan experimentos que se realizaron con 1107 imágenes de rostros de la base de datos FERET, correspondientes a 369 sujetos, utilizando en promedio, tres imágenes por sujeto.

Otras aplicaciones utilizando algoritmos genéticos, se han realizado combinándolos con algunas de la técnicas mencionadas en las secciones anteriores. Entre las investigaciones reportadas, se encuentran, las de Bhanu (1995) para segmentación de imágenes. Huang (1998) realiza el reconocimiento utilizando componentes principales y algoritmos genéticos. Pinto y Sossa (1998) proponen un método que utiliza las primeras cuatro invariantes de Hu y a partir de éstas realizar la identificación con la ayuda de un algoritmo genético.

Capítulo 3

Análisis de Componentes Principales

El análisis de componentes principales (ACP) tiene sus antecedentes en Psicología, a través de las técnicas de regresión lineal iniciadas por Galton. Específicamente, Pearson (1901) presentó la primera propuesta del método de componentes principales. Sin embargo, el nombre de “**componentes principales**” y su primer desarrollo teórico se debe a Hotteling (1933), quien desarrolló un método de extracción de factores.¹

Por su parte, Thurstone (1947), expresó la relación entre las correlaciones y las saturaciones de las variables en los factores. Asimismo, introdujo el concepto de estructura simple y desarrolló la teoría y método de las rotaciones factoriales para obtener la estructura factorial más sencilla. En un principio las rotaciones eran gráficas. Kaiser (1958) desarrolló el método denominado VARIMAX para realizar rotaciones ortogonales mediante

¹ La idea central del método de componentes principales, consiste en reducir n variables a p factores o componentes (principales), en donde $p \ll n$. De los p componentes que se obtienen, el primero, corresponde a un subconjunto de las n variables que presentan la mayor varianza y por tanto, dicho subconjunto de variables comparten características similares en los datos originales. El segundo factor, corresponderá al siguiente subconjunto de variables que comparten elementos similares, pero la varianza que presentan es menor a la del primer factor; y así sucesivamente.

procedimientos matemáticos. Dicho método, es el que actualmente utilizan los diversos paquetes computacionales, como SPSS y SAS.

El análisis de componentes principales se empleó inicialmente (y se sigue utilizando), en la psicología, las ciencias sociales y naturales. Sin embargo, desde hace algunos años se ha visto lo útil de su aplicación en las ciencias físicas, la ingeniería, la economía, la educación, el reconocimiento de patrones, etc.

Por otra parte, a Fukunaga (1972), se le considera como el primer investigador en aplicar el análisis de componentes principales al Reconocimiento de Patrones.

3.1. Características generales del análisis de componentes principales

El análisis con componentes principales es una técnica de análisis estadístico multivariante que se clasifica entre los métodos de simplificación o reducción de la dimensionalidad de variables, y que se aplica cuando se dispone de un conjunto elevado de variables con datos cuantitativos y con el fin de obtener un menor número de variables: Las nuevas variables, son una combinación lineal de las variables originales y se denominan componentes principales o factores.

El análisis con componentes permite la descripción, de forma sintética, de la estructura y las interrelaciones de las variables originales en el fenómeno que se estudia a partir de los factores que se obtuvieron.

Por tanto, el método de componentes principales tiene como propósito transformar un conjunto de variables, a las que se denominan variables originales interrelacionadas, en un nuevo conjunto de variables, combinación lineal de las originales denominadas factores o componentes principales. Estas últimas se caracterizan por estar incorrelacionadas entre sí.

En general, la extracción de los factores se efectúa sobre variables tipificadas para evitar problemas derivados de escala, aunque también se puede aplicar sobre variables expresadas en desviaciones respecto a la media.

Si p variables se encuentran tipificadas, la suma de las varianzas es igual a p , ya que la varianza de una variable tipificada es por definición igual a 1. El nuevo conjunto de variables que se obtienen utilizando ACP, es igual en número al de variables originales. Sin embargo, es importante destacar que la suma de sus varianzas es igual a la suma de las varianzas de las variables originales. Las diferencias entre ambos conjuntos de variables estriba en que, como ya se indicó, los factores se calculan de manera que estén incorrelacionados entre sí (a no ser, que se especifique lo contrario en el análisis).

Cuando las variables originales están muy correlacionadas entre sí, la mayor parte de su variabilidad se puede explicar con muy pocos componentes (factores). Si las variables originales estuvieran completamente incorrelacionadas entre sí, entonces el ACP carecería de aplicación, ya que en ese caso, los factores coincidirían con las variables originales.

También se puede decir, que mediante el ACP, se tiene la posibilidad de identificar “patrones” en un conjunto de datos y expresar los datos de una forma en que recopilemos sus similitudes y diferencias.

3.2. Planteamiento matemático general

Supongamos que cada individuo está descrito por k variables. Los datos de n sujetos se pueden representar mediante la matriz X de orden $n \times k$:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \text{M} & \text{M} & \text{M} & \text{M} \\ x_{n1} & x_{n2} & \Lambda & x_{nk} \end{bmatrix} \quad (2)$$

Los datos de la matriz forman una nube de n puntos en un espacio k -dimensional.

Por tanto, sea $\hat{x}_1, \dots, \hat{x}_k$ una base de \mathfrak{R}^k . Y considerando la nube k -dimensional de n puntos a la que se denominará M_1, M_2, \dots, M_n , en los cuales se tendrán las coordenadas $(x_{i1}, x_{i2}, \dots, x_{ik})$ correspondientes al punto M_i en la base $\hat{x}_1, \dots, \hat{x}_k$. Por tanto, cada línea de la matriz X definida en (2) corresponde a los componentes de un punto de la nube en el sistema de coordenadas inducido por esta base.

Considerando ahora otra base, $\hat{u}_1, \dots, \hat{u}_k$ de \mathfrak{R}^k . Y las coordenadas $(z_{i1}, z_{i2}, \dots, z_{ik})$ del punto M_i en esta nueva base, se puede construir la matriz Z de orden $n \times k$.

$$Z = \begin{bmatrix} z_{11} & z_{12} & \Lambda & z_{1k} \\ z_{21} & z_{22} & \Lambda & z_{2k} \\ \text{M} & \text{M} & \text{M} & \text{M} \\ z_{n1} & z_{n2} & \Lambda & z_{nk} \end{bmatrix} \quad (3)$$

Para relacionar Z con X , consideramos U , como la matriz de paso de la base x_1, \dots, x_k a la nueva base. La matriz U es una matriz de orden $k \times k$ cuyas columnas tienen las coordenadas de los vectores u_1, \dots, u_k en la base inicial, es decir si

$$\begin{aligned} u_1 &= u_{11}x_1 + u_{21}x_2 + \Lambda + u_{k1}x_k \\ u_2 &= u_{12}x_1 + u_{22}x_2 + \Lambda + u_{k2}x_k \\ &\text{M} \qquad \qquad \text{M} \\ u_k &= u_{1k}x_1 + u_{2k}x_2 + \Lambda + u_{kk}x_k \end{aligned} \quad (4)$$

entonces, la matriz de paso será

$$U = \begin{bmatrix} u_{11} & u_{12} & \Lambda & u_{1k} \\ u_{21} & u_{22} & \Lambda & u_{2k} \\ \text{M} & \text{M} & \text{M} & \text{M} \\ u_{k1} & u_{k2} & \Lambda & u_{kk} \end{bmatrix} \quad (5)$$

Por tanto, la relación entre Z , X y U es

$$X = ZU^T \quad (6)$$

Como U es invertible, se puede deducir que

$$X(U^T)^{-1} = Z \quad (7)$$

Si los vectores u_1^0, \dots, u_k^0 forman una base ortonormal, la matriz U satisface la relación

$$U^T U = I \text{ y por lo tanto } (U^T)^{-1} = U.$$

Matriz de covarianzas y cambio de bases

La matriz de covarianzas para la matriz X se encuentra representada por

$$S_X = \begin{bmatrix} s_{X_1}^2 & s_{X_1 X_2} & \Lambda & s_{X_1 X_k} \\ s_{X_1 X_2} & s_{X_2}^2 & \Lambda & s_{X_2 X_k} \\ \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} \\ s_{X_1 X_k} & s_{X_2 X_k} & \Lambda & s_{X_k}^2 \end{bmatrix} \quad (8)$$

donde $s_{X_i}^2$ representa la varianza de la variable X_i en el conjunto y, $s_{X_i X_j}$ es la covarianza de X_i y X_j .

Simplificando, si se asume que las variables en X_1, \dots, X_k están “centradas”, es decir que tienen una media de cero: $\bar{x}_1 = \dots = \bar{x}_k = 0$. Con lo anterior, se puede comprobar que la matriz de covarianzas de X se puede expresar como

$$S_X = \frac{1}{n-1} X' X \quad (9)$$

Utilizando la ecuación (9), se puede deducir que la matriz S_Z de covarianzas de las variables Z_1, \dots, Z_k , componentes de los puntos en la nueva base $\hat{u}_1, \dots, \hat{u}_k$ se relaciona con S_X por medio de

$$S_Z = \frac{1}{n-1} Z^T Z = \frac{1}{n-1} U^{-1} X^T X (U^T)^{-1} = U^{-1} S_X (U^T)^{-1} \quad (10)$$

En el caso en que U es ortogonal, la relación se puede escribir como

$$S_Z = U^{-1} S_X U \quad (11)$$

Por tanto, de forma general: si las variables Z_1, \dots, Z_k se consideran incorrelacionadas y son una combinación lineal de las variables iniciales y se asume que explican la mayor parte de su variabilidad, se puede escribir como:

$$\begin{aligned} Z_1 &= u_{11}X_1 + u_{12}X_2 + \dots + u_{1p}X_p \\ Z_2 &= u_{21}X_1 + u_{22}X_2 + \dots + u_{2p}X_p \\ &\cdot \quad \quad \quad \cdot \\ &\cdot \quad \quad \quad \cdot \\ &\cdot \quad \quad \quad \cdot \\ Z_p &= u_{p1}X_1 + u_{p2}X_2 + \dots + u_{pp}X_p \end{aligned} \quad (12)$$

El sistema anterior es reversible, por lo cual, se puede expresar a las variables X_j en función de los componentes principales Z_j , por tanto:

$$\begin{aligned}
X_1 &= u_{11}Z_1 + u_{12}Z_2 + \dots + u_{1p}Z_p \\
X_2 &= u_{21}Z_1 + u_{22}Z_2 + \dots + u_{2p}Z_p \\
&\cdot \quad \quad \quad \cdot \\
&\cdot \quad \quad \quad \cdot \\
&\cdot \quad \quad \quad \cdot \\
X_p &= u_{p1}Z_1 + u_{p2}Z_2 + \dots + u_{pp}Z_p
\end{aligned} \tag{13}$$

Utilizando los componentes Z_j como tipificados:

$$Y_j = \frac{Z_j}{\sqrt{\lambda_j}} \quad j = 1, 2, \dots, p \tag{14}$$

Entonces, en el segundo sistema (13), se puede sustituir los Z_j por $Y_j\sqrt{\lambda_j}$, resultando la ecuación j -ésima del sistema con la siguiente forma:

$$X_j = u_{ij}Y_1\sqrt{\lambda_1} + u_{2j}Y_2\sqrt{\lambda_2} + \dots + u_{pj}Y_p\sqrt{\lambda_p} \tag{15}$$

De la teoría de componentes principales (Pérez, 2004; Batista y Martínez, 1989), se conoce que $u_{hj}\sqrt{\lambda_h}$ es el coeficiente de correlación entre la variable j -ésima y la componente h -ésima, lo que permite escribir la ecuación como:

$$X_j = r_{1j}Y_1 + r_{2j}Y_2 + \dots + r_{pj}Y_p \tag{16}$$

De la ecuación anterior, se tiene la posibilidad de separar sus últimos $p-k$ términos, lo que permite escribirla como:

$$X_j = r_{1j}Y_1 + r_{2j}Y_2 + \dots + r_{kj}Y_k + (r_{k+1,j}Y_{k+1} + \dots + r_{pj}Y_p) \tag{17}$$

Considerando que una ecuación de modelo factorial se puede representar como:

$$X_j = l_{j1}F_1 + l_{j2}F_2 + \dots + l_{jk}F_k + e_j \quad (18)$$

se puede observar que los k factores F_h se estiman mediante las k primeras componentes principales tipificadas (Y_h) y la estimación de los coeficientes l_{jh} , se encuentra dada por:

$$\hat{l}_{j1} = r_{1j}, \hat{l}_{j2} = r_{2j}, \dots, \hat{l}_{jk} = r_{kj} \quad (19)$$

La comunalidad de la variable X_j se puede estimar como:

$$\hat{h}_j^2 = \hat{l}_{j1}^2 + \hat{l}_{j2}^2 + \dots + \hat{l}_{jk}^2 \quad (20)$$

y el factor único e_j se estimará como:

$$\hat{e}_j = r_{k+1,j}Y_{k+1} + r_{m+2,j}Y_{k+2} + \dots + r_{pj}Y_p \quad (21)$$

Por último, la especificidad o parte de la varianza que se debe al factor único se puede estimar como:

$$\hat{\omega}_j^2 = 1 - \hat{h}_j^2 \quad (22)$$

Utilizando los elementos matemáticos descritos anteriormente, se tiene la posibilidad de configurar diversos sistemas para análisis de componentes principales.

En general, de manera básica se puede considerar que cualquier problema de componentes principales, se representaría en notación matricial como:

$$\mathbf{y} = \mathbf{A}'\mathbf{x} \quad (23)$$

y para un componente cualquiera:

$$y_j = a_j\mathbf{x} \quad (24)$$

donde:

- \mathbf{A} es una matriz cuadrada de orden “p” cuyas columnas representan los pesos de las combinaciones lineales o Componentes Principales
- \mathbf{x} es el vector de las variables originales

Y para eliminar la indeterminación que se puede producir por el hecho de que la varianza de los componentes principales puede llegar a modificarse al multiplicar los pesos por constantes, se impone la restricción de que el vector de coeficientes de cada componente sea de módulo 1, por tanto:

$$a'_j a_j = 1 \quad (25)$$

De manera final, se pueden plantear que las principales definiciones que subyacen a las formulaciones matemáticas que se han presentado, son las siguientes:

- El primer componente principal F_1 es una combinación lineal de las variables originales \mathbf{X} , dada por $a'_1 \mathbf{x}$ y con varianza máxima. Con la restricción de que $a'_1 a_1 = 1$.

- El segundo componente principal F_2 , es la combinación de las variables originales dada por $a'_2 \mathbf{x}$, que maximiza la varianza. Y con las restricciones: $a'_2 a_2 = 1$ y $Cov(a'_1 \mathbf{x}, a'_2 \mathbf{x}) = 0$.
- En general, el j -ésimo componente principal será la combinación lineal de las variables observadas, dada por $a'_j \mathbf{x}$ que maximiza la varianza no explicada y con las restricciones: $a'_j a_j = 1$ y $Cov(a'_k \mathbf{x}, a'_j \mathbf{x}) = 0$ para $(k < j)$.
- Adicionalmente, $Var(F_1) > Var(F_2) > \dots > Var(F_p)$ y $Cov(Y_j, Y_k) = 0$ para $(j \neq k)$

Diversos paquetes de software disponibles actualmente (SPSS, SAS, Minitab, etc.), tienen la posibilidad de realizar análisis de componentes principales con diversas alternativas. Sin embargo, para problemáticas muy específicas, como es el caso del problema que nos ocupa sobre reconocimiento de rostros, se requiere de un sistema *ad hoc* para calcular los componentes y que incluya diversos elementos adicionales.

Por tanto, en la siguiente sección, se presenta los elementos principales para configurar el algoritmo de reconocimiento de rostros basado en análisis de componentes principales. Se utilizarán diversos elementos matemáticos que se plantearon en la presente sección. La programación del sistema, se realizará utilizando MATLAB y la Herramienta de Procesamiento de Imágenes de Matlab.

3.3. Representación de rostros utilizando el análisis de componentes principales

La imagen de un rostro, se puede representar utilizando un vector. Si el ancho y la altura de la imagen del rostro es de $w \times h$ pixeles respectivamente, el número de componentes del vector será $n = w \times h$.

Cada valor numérico del píxel, corresponderá a un elemento del vector. La construcción de dicho vector a partir de la imagen, se realiza por medio de una simple concatenación: los renglones de la imagen se localizan una tras otra.

Espacio de imagen y espacio de rostros

El vector que representa al rostro pertenece a un espacio, al que se denominará “espacio de imagen”. Y es en donde se encuentran todas las imágenes cuya dimensión es n .

Utilizando el método de componentes principales y obteniendo los denominados “factores” o componentes, se obtendrá el “espacio de rostros”.

La dimensión de un rostro en el espacio de imagen es n , sin embargo, no todos los pixeles del rostro pueden ser relevantes. Además, cada píxel depende de sus vecinos. Por tanto, se presupone, que la dimensión del espacio del rostro es menor que la dimensión del espacio de la imagen.

El objetivo de utilizar el método de Componentes Principales, consiste en reducir la dimensión de un conjunto o espacio de modo que la nueva base describa mejor los “modelos” o “patrones” típicos del conjunto total. En el caso presente, nuestros “patrones” corresponden al conjunto de rostros de entrenamiento.

En el espacio de imágenes, los rostros se encuentran no se encuentran aleatoriamente distribuidos, como se ha mencionado, utilizando el ACP, se reducirá la alta dimensionalidad del conjunto de datos.

La idea central del ACP consiste en encontrar un subespacio de baja dimensionalidad (denominado espacio de características o espacio de rostros), el cual capturará la mayor parte de la variación que se presenta en el conjunto de datos originales (espacio de imágenes).

Por tanto, dado un conjunto de vectores de entrenamiento $\{x\}$, que corresponden a los rostros ejemplo, y que presentan una matriz de covarianza S_x , los factores se pueden calcular al solucionar el problema de eigenvalues

$$D = P^T S_x P \quad (26)$$

en donde P es la matriz de eigenvectores de S_x y D corresponde a la matriz diagonal de los eigenvalores.

La matriz de proyección ortogonal P_M dentro del subespacio principal M-dimensional (para $M \ll N$), está dado por los M eigenvectores, correspondientes a los más grandes eigenvalores. Estos eigenvectores, forman las columnas de la matriz de proyección P_M .

El vector y de componentes principales, se obtiene al proyectar la imagen x en el espacio de rostros:

$$y = P_M^T (x - \bar{x}) \quad (27)$$

en donde \bar{x} , denota el “rostro promedio”.

3.4. Algoritmo para reconocimiento de rostros utilizando ACP

Como se ha mencionado en secciones anteriores, la imagen de un rostro en 2-D, se puede representar mediante un vector unidimensional al concatenar cada columna de la imagen, a lo largo del vector. Por tanto, el procedimiento algorítmico será:

1. Concatenar todas las imágenes de los rostros susceptibles a identificar en un vector, en donde cada imagen del rostro bidimensional presenta nr renglones y nc columnas en un vector \mathbf{x}_i , de tamaño $N(nr \times nc)$: Por tanto, habrá M vectores de tamaño N .

$$\mathbf{x}_i = [p_1, \dots, p_N]^T, i = 1, \dots, M \quad (28)$$

en donde, p_j representa el valor numérico del píxel.

2. Obtener la imagen promedio (m) de todos los rostros (rostro promedio)

$$m = \frac{1}{M} \sum_{i=1}^M x_i \quad (29)$$

3. Substraer a cada imagen de rostro, el rostro promedio

$$w_i = x_i - m \quad (30)$$

4. Obtener el conjunto de M vectores ortonormales (e_i), de tal forma que presenten la máxima proyección posible sobre las w_i . Por tanto, la cantidad

$$\lambda_i = \frac{1}{M} \sum_{n=1}^M (e_i^T w_n)^2 \quad (31)$$

se debe maximizar con la restricción de ortonormalidad: $e_i^T e_k = \delta_{ik}$.

5. Obtener la matriz de covarianza. Se ha señalado en secciones anteriores, que los valores de (e_i) y (λ_i) se obtienen por medio de los eige-vectores y eige-valores de la matriz de covarianza

$$C = WW^T \quad (32)$$

donde W es una matriz compuesta por los vectores columna w_i colocados de forma seguida. El tamaño de la matriz C (NxN) puede ser muy grande, por ejemplo, para el caso de la base de datos AT&T con un tamaño de cada rostro de 64 x 112 pixeles, se genera una matriz de covarianza de 4096 x 4096. Para tratar de disminuir el tamaño de dicha matriz, se puede utilizar un teorema de álgebra lineal. Dicho teorema, establece que los vectores (e_i) y los escalares λ_i se pueden obtener al solucionar la matriz de tamaño (M x M): $W^T W$. Dado que d_i y μ_i sean los eige-vectores y eingevalores de $W^T W$, respectivamente, se puede representar

$$W^T W d_i = \mu_i d_i, \text{ multiplicando ambos lados por W}$$

$$W W^T (W d_i) = \mu_i (W d_i) \quad (33)$$

Lo anterior, permite que los primeros M-1 eigevectores (e_i) y los eigevalores λ_i de $W^T W$ estén dados por $W d_i$ y por μ_i respectivamente. $W d_i$ debe ser normalizado,

con el propósito de que sea similar a (e_i) . Los eige-vectores que corresponden a los eigevalores diferentes de cero de la matriz de covarianza, producen una base ortonormal para el “subespacio” dentro del cual, la mayoría de los datos de las imágenes de los rostros se pueden representar.

6. Clasificar los eigevectores. Los eigevectores se clasifican de mayor a menor, de acuerdo a sus correspondientes eigevalores. El eigevector (o componente principal), asociado con el más grande eigevalor, reflejará el componente con la mayor varianza de las imágenes de rostros. Se considera que aproximadamente el 90% de la varianza total, se encuentra en el 5% al 10% de los factores que se obtuvieron.

7. Reconocer un rostro. Si se denomina a Ω como la descripción de la contribución de cada eigevector de rostro para representar la imagen facial, al tratar los eigevalores como un conjunto base del conjunto de entrenamiento de los rostros. Entonces, un método para determinar si un determinado rostro (Ω) se encuentra en la base de entrenamiento, consiste en encontrar la clase de rostro k que minimíze la distancia Euclideana

$$\varepsilon_k = \|\Omega - \Omega_k\| \quad (34)$$

$$\varepsilon_k < \textit{umbral}$$

en donde Ω_k es un vector que describe la k -ésima clase de rostro. Si ε_k es menor que un “umbral predefinido”, entonces, el rostro se clasifica como perteneciente a la clase k . Lo cual implica, haber identificado el rostro en la base de datos correspondiente.

Para efectuar la clasificación del rostro, se debe utilizar una métrica o norma. Anteriormente, se mencionó que se usará la distancia euclídeana, la cual corresponde a la norma L_2 . Sin embargo, existen otras posibilidades de normas, las cuales se mencionan a continuación.

Medidas de distancia (normas)

Un elemento importante en la fase del reconocimiento, es la norma para comparar la proyección del rostro a reconocer con la base de datos. Las principales normas que se pueden utilizar en el Paso 7 del algoritmo, son: la euclídeana (L_2), la norma L_1 y la distancia de Mahalanobis.

Norma L_1

Esta norma, suma la diferencia absoluta entre las componentes de los dos vectores

$$L1 = \sum_{i=1}^M |\Omega_i - \Omega_{k_i}| \quad (35)$$

donde M, es el tamaño de ambos vectores.

Norma L_2 o Euclídiana

Esta norma, suma la diferencia al cuadrado entre las componentes de los dos vectores

$$L2 = \sum_{i=1}^M (\Omega_i - \Omega_{k_i})^2 \quad (36)$$

Distancia de Mahalanobis

Calcula el producto de las componentes y el valor propio (eige-valor) asociado a cada componente, realizando la sumatoria de dichos productos.

$$Mah = -\sum_{i=1}^M \Omega_i \bullet \Omega_{k_i} \bullet h_i \quad \text{en donde : } h_i = \frac{1}{\sqrt{\lambda_i}} \quad (37)$$

A partir de los elementos matemáticos que subyacen al análisis de componentes principales y al algoritmo para reconocimiento de rostros planteado anteriormente, se desarrolló un programa en MATLAB para realizar diversos experimentos y probar las bondades del método de componentes principales. En capítulo 5, se presentan las características principales del programa y el desarrollo de los experimentos realizados con fotografías de la base de rostros AT&T, fotografías del portal del FBI y algunas fotografías adicionales.

Capítulo 4

Análisis estadístico de las imágenes digitales de rostros

Un rostro humano analizado por medio de una imagen digital, presentará determinadas características estadísticas a partir de la configuración de los píxeles que la constituyen. En el presente trabajo, únicamente se consideran fotografías digitales en tonos de grises (denominadas tipo “grayscale”).

Como se ha mencionado, las fotografías de la base de datos AT&T tienen una dimensión de 92 píxeles de ancho por 112 píxeles de alto. Por tanto, cada imagen tendrá 10304 píxeles en total. Y cada píxel presenta un valor entre 0.0 y 1.0, correspondiente a la tonalidad del nivel de gris que representa.

Se puede considerar una fotografía digital de una persona, como un vector de 10,304 posiciones. Si se tienen para un determinado experimento de reconocimiento de rostros, una base de entrenamiento con 40 personas y 9 fotografías por persona, se tendrán 360 vectores (o variables) con 10,304 valores cada uno (lo que implicará 3,709,440 valores para analizar).

Como se puede observar, al realizar un análisis para reconocimiento de rostros, se presenta un problema de análisis multivariado con un gran número de datos.

Algunos de los sistemas computacionales de reconocimiento de rostros que se aplican actualmente, utilizan técnicas que se basan en determinados supuestos estadísticos que deberían de presentar los datos de los píxeles de las imágenes.

En el caso específico que nos ocupa, la técnica de análisis de componentes principales, su aplicación presupone el cumplimiento de ciertos supuestos estadísticos.

4.1. Supuestos estadísticos del análisis multivariante tradicional

Un elemento importante para aplicar alguna de las técnicas de análisis multivariado, radica en el cumplimiento de algunas características estadísticas por parte de los datos de entrada, en nuestro caso, los píxeles de cada imagen digital del rostro. Si bien, los diversos métodos que se aplican en el análisis multivariado se consideran robustos y por ende, admiten algunas violaciones a los supuestos estadísticos, en la medida que los datos de entrada se alejen más de los supuestos ocasionarán fallas en los resultados del sistema.

Se piensa, que el análisis de componentes principales puede generar fallas en el proceso de reconocimiento de rostros, debido a violaciones de los supuestos estadísticos por parte de los datos numéricos que presentan los píxeles de imágenes de rostros. Por tanto, en la presente sección se realizará un análisis estadístico de una muestra de imágenes digitales de rostros humanos para conocer si cumplen con los supuestos estadísticos que requieren las técnicas de análisis multivariado, específicamente, el análisis de componentes principales.

El análisis multivariado requiere que los supuestos subyacentes a las técnicas estadísticas sean contrastados para las variables aisladas que se consideran y en segundo lugar, para el

valor teórico del modelo multidimensional. Sin embargo, una prueba multivariante, en muchos casos, es difícil de aplicar. Por tanto, en la mayoría de las ocasiones, se realizan pruebas para las variables aisladas y si la mayoría de ellas cumple, se extiende al modelo multivariable en su totalidad.

A continuación se presentan los supuestos estadísticos que deberían cubrir los datos al aplicarse el análisis multivariado (Hair, Anderson, Tatham y Black, 1998):

Normalidad

El supuesto fundamental del análisis multivariante es la prueba de normalidad de los datos. Los datos, deben configurar una distribución normal multivariable.

La prueba de normalidad univariante para una única variable, es relativamente fácil de contrastar. En el caso de de dos o más variables (multivariada), si la distribución es normal multivariada, implica que todas las variables individuales que intervienen deben ser normal univariante y sus combinaciones son también normales. Sin embargo, lo contrario no es necesariamente cierto: dos o más variables normales univariantes, no son necesariamente normal multivariante.

Una situación en que todas las variables presentan normalidad univariante, indica indicios de que se tiene “normalidad multivariada”, aunque no lo garantiza. En la práctica, si la mayoría de las variables presenta normalidad univariante, se considera una “normalidad multivariada”. En el caso, de que alguna o algunas de las variables aisladas, presentan una “no normalidad”, implicará que no se tiene normalidad multivariada.

Supuesto de Homoscedasticidad

Es un supuesto relativo a las relaciones de dependencia entre variables. Se refiere al supuesto de que las variables dependientes exhiban iguales niveles de varianza a lo largo del rango del predictor de la variable. Se puede aplicar el test de Levene para probar si se tiene homoscedasticidad.

Linealidad

La linealidad es un supuesto implícito de todas las técnicas multivariantes basadas en medidas de correlación incluyendo la regresión múltiple, análisis factorial, componentes principales y modelos de ecuaciones estructurales.

La forma más común de evaluar la linealidad es examinar los gráficos de dispersión de las variables e identificar cualquier pauta no lineal en los datos. Otra alternativa de análisis consiste en aplicar regresión múltiple y examinar los residuos.

4.2. Análisis estadístico de la imagen de un rostro

Se realizaron diversas pruebas estadísticas a las fotografías digitales de la base de datos AT&T para evaluar si cumplen con los supuestos estadísticos que se han mencionado en la sección anterior. Para el análisis de las imágenes digitales se utilizó el paquete de estadísticas SPSS y el paquete LISREL/PRELIS.

Análisis descriptivo de la imagen digital de un rostro.

Como se ha mencionado, las imágenes utilizadas en el presente trabajo presentan una configuración de 92 píxeles (columnas) por 112 píxeles (renglones). Por ejemplo, para la segunda fotografía de la octava persona de la base AT&T (S0802.PGM, figura 7), la estadística descriptiva correspondiente a dicha imagen se presenta en la tabla 1.



FIGURA 7. EJEMPLO DE FOTOGRAFIA DE LA BASE AT&T (SEGUNDA IMAGEN DE LA OCTAVA PERSONA DE LA BASE – S0802.PGM)

TABLA 1. ESTADÍSTICA DESCRIPTIVA DE LOS VALORES A NIVEL PÍXEL QUE PRESENTA LA FOTOGRAFIA S0802.PGM

Descriptives			Statistic	Std. Error
FOTO0802	Mean		.52081214	2.90E-03
	95% Confidence Interval for Mean	Lower Bound	.51512073	
		Upper Bound	.52650355	
	5% Trimmed Mean		.51869707	
	Median		.52709360	
	Variance		8.687E-02	
	Std. Deviation		.29472950	
	Minimum		.000000	
	Maximum		1.000000	
	Range		1.000000	
	Interquartile Range		.54679800	
	Skewness		.153	.024
	Kurtosis		-1.354	.048

De los resultados descriptivos que se presentan en la tabla 1, se puede observar que el valor promedio de los tonos de los pixeles es de 0.5208, con una desviación estándar de +/- 0.2947. Lo anterior, indica una gran dispersión de los datos. Asimismo, el valor mínimo (0.0) y el máximo (1.0), nos indican que se tienen pixeles que cubren toda la gama de tonalidades de grises: desde el negro al blanco. Por tal motivo, análisis de rostros humanos tomando únicamente en cuenta análisis de estadística descriptiva no se consideran adecuados, debido a que cualquier fotografía en tonalidades de grises, puede tener características estadísticas equivalentes.

Sin embargo, la interrelación entre los diversos valores de los pixeles, la estructura de la imagen y las diversas configuraciones latentes a nivel local, son las que proporcionan información sobre la imagen y están en condiciones de marcar la diferencia entre las imágenes de rostros humanos y cualquier otra imagen.

En la figura 8, se muestra el histograma para la fotografía S0802.PGM bajo análisis. Se observa a simple vista, una configuración multimodal. Lo anterior, en combinación con los valores de la simetría y curtosis que se obtuvieron (tabla 1), dan la pauta para considerar que el comportamiento de la distribución de los datos no se asemeja a una distribución normal. Sin embargo, se realizará posteriormente, un prueba de hipótesis para probar normalidad de las imágenes de rostros humanos.

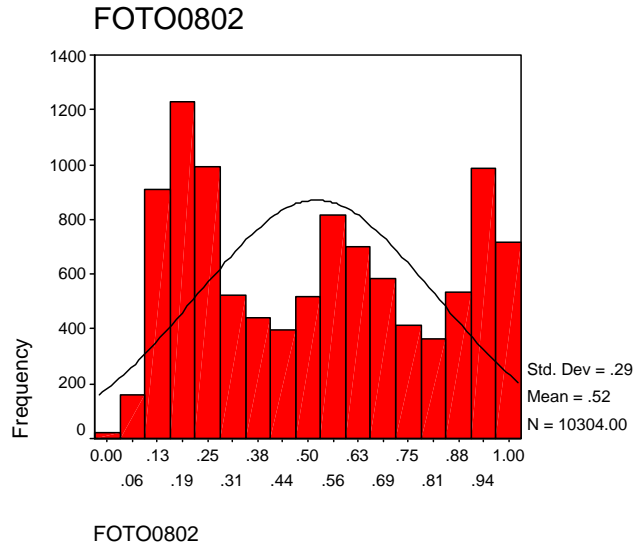


FIGURA 8. HISTOGRAMA DE LOS NIVELES DE GRIS DE LOS PÍXELES CORRESPONDIENTES A LA FOTOGRAFIA S0802.PGM

El análisis estadístico descriptivo que se realizó para las diversas imágenes de la base de datos AT&T, presentan características similares a los resultados del rostro con identificación S0802, presentados anteriormente.

Se obtuvo el promedio numérico de los 400 rostros de la base de datos, equivalente a lo que se ha denominado por diversos autores “rostro promedio”. El resultado del análisis estadístico correspondiente, se presenta a continuación (tabla 2 y figura 9).

TABLA 2. ESTADÍSTICA DESCRIPTIVA DEL “ROSTRO PROMEDIO” DE LAS 400 FOTOGRAFÍAS DE LA BASE AT&T

Statistics

PROFOTOS

N	Valid	10304
	Missing	0
Mean		.49638077
Median		.51238174
Mode		.248812 ^a
Std. Deviation		.14386572
Variance		2.07E-02
Skewness		.014
Std. Error of Skewness		.024
Kurtosis		-1.240
Std. Error of Kurtosis		.048
Minimum		.248812
Maximum		.776146

a. Multiple modes exist. The smallest value is shown

En la figura 9, se observa el histograma correspondiente al “rostro promedio”. Se aprecia una distribución de tipo bimodal, lo cual sugiere que el rostro promedio, tampoco se ajusta a una distribución de tipo normal.

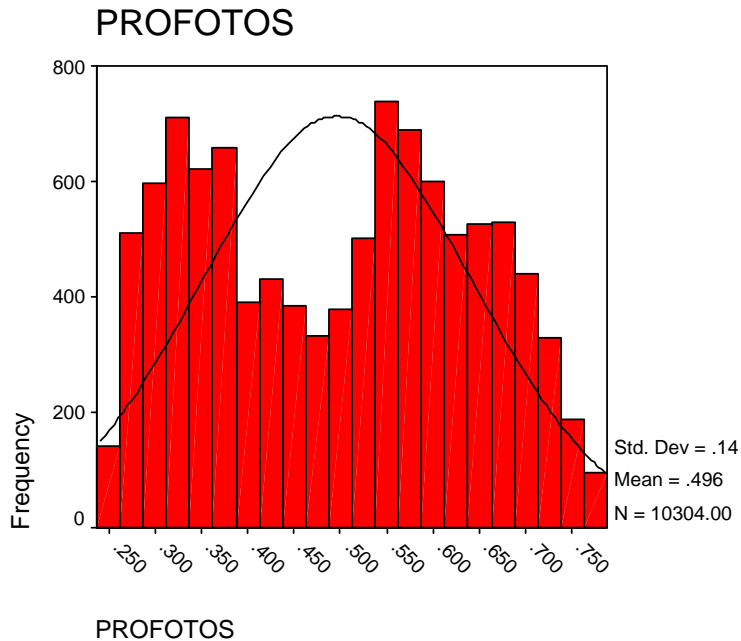


FIGURA 9. HISTOGRAMA DE LOS NIVELES DE GRIS CORRESPONDIENTES A LOS PÍXELES DEL “ROSTRO PROMEDIO” DE LA BASE AT&T

En la sección siguiente se comprobará si efectivamente, la distribución que presentan los datos de la fotografía de un rostro, cumple con las características de una distribución normal o nó.

4.3. Prueba de normalidad univariada y multivariada para fotografías de rostros.

Para comprobar si los datos de los píxeles que configuran un rostro humano, cumple con las características de una distribución normal, se realizaron pruebas de normalidad univariada con el paquete SPSS y pruebas de normalidad multivariada a una muestra representativa de la base de datos AT&T, utilizando el paquete PRELIS/LISREL. Cabe recordar, que el cumplimiento del supuesto de normalidad, es importante para tener resultados adecuados en sistemas de reconocimiento de rostros que utilizan técnicas bajo métricas en L_2 .

4.3.1. Prueba de normalidad univariada

Para comprobar la normalidad univariada de rostros de la base AT&T, se utilizó la prueba de Kolmogorov-Smirnov que proporciona el paquete SPSS.

La hipótesis que se utiliza en SPSS es:

H_0 : No existe diferencia entre la distribución del conjunto de datos y una distribución normal.

H_1 : Existe una diferencia significativa entre la distribución de datos y una distribución normal.

La prueba de significancia se realiza a un valor de confiabilidad del 95%. Por tanto, si el valor de “p” es menor a 0.05, se rechazará H_0 y se considerará que la distribución no es normal.

En la tabla 3 se presentan algunos de los resultados de la prueba de normalidad univariada para la muestra seleccionada. Se puede observar, que ninguna de las fotografías cumple con el supuesto de normalidad.

En el anexo B, se presentan los resultados de la prueba de normalidad para los 400 rostros de la base de datos AT&T.

TABLA 3. RESULTADOS DE LA PRUEBA DE NORMALIDAD UNIVARIADA (KOLMOROV-SMIRNOV) PARA ALGUNOS DE LOS ROSTROS DE LA BASE AT&T.

Tests of Normality

	Kolmogorov-Smirnov ^a		
	Statistic	df	Sig.
FOTO0103	.208	10304	.000
FOTO0201	.138	10304	.000
FOTO0308	.058	10304	.000
FOTO0409	.059	10304	.000
FOTO0510	.074	10304	.000
FOTO0604	.065	10304	.000
FOTO0709	.126	10304	.000
FOTO0801	.108	10304	.000
FOTO0907	.051	10304	.000
FOTO1006	.128	10304	.000
FOTO1105	.136	10304	.000
FOTO1202	.057	10304	.000
FOTO1310	.046	10304	.000
FOTO1406	.113	10304	.000
FOTO1504	.088	10304	.000
FOTO1607	.166	10304	.000
FOTO1706	.155	10304	.000
FOTO1801	.076	10304	.000
FOTO1907	.097	10304	.000
FOTO2002	.087	10304	.000
FOTO2109	.069	10304	.000
FOTO2208	.077	10304	.000
FOTO2309	.110	10304	.000
FOTO2410	.101	10304	.000
FOTO2510	.063	10304	.000
FOTO2610	.067	10304	.000
FOTO2706	.143	10304	.000
FOTO2804	.098	10304	.000
FOTO2909	.070	10304	.000
FOTO3001	.068	10304	.000
FOTO3104	.054	10304	.000
FOTO3209	.076	10304	.000
FOTO3310	.068	10304	.000
FOTO3401	.065	10304	.000
FOTO3502	.051	10304	.000
FOTO3603	.109	10304	.000
FOTO3706	.118	10304	.000
FOTO3805	.050	10304	.000
FOTO3907	.048	10304	.000
FOTO4001	.041	10304	.000

a. Lilliefors Significance Correction

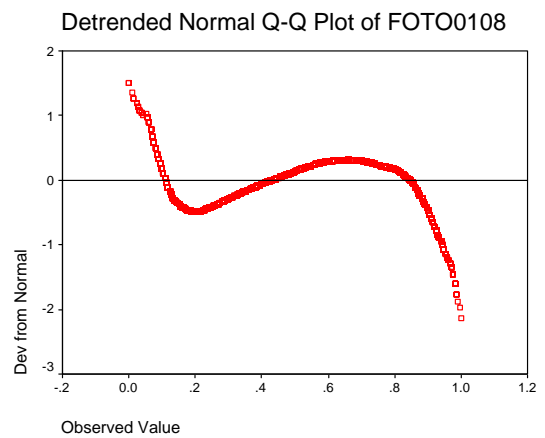
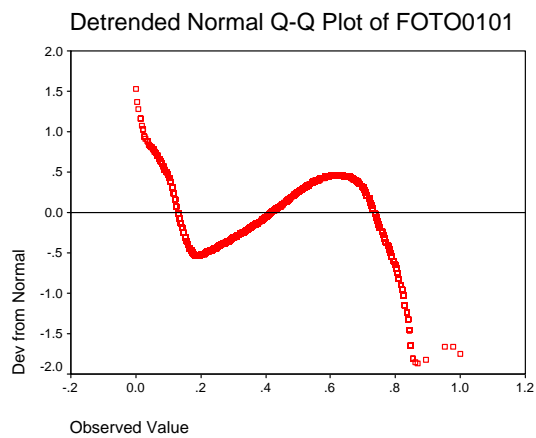
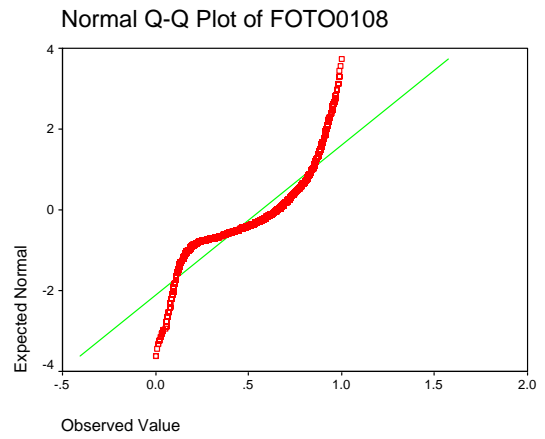
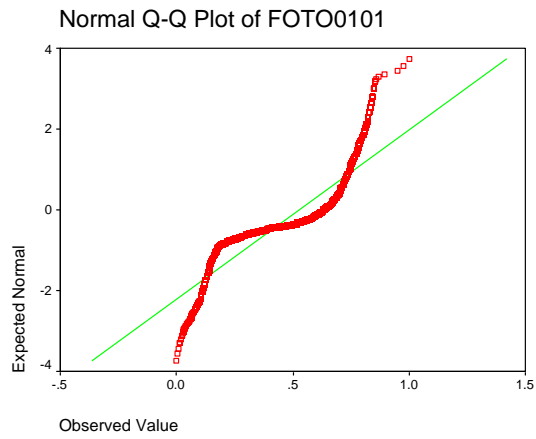


FIGURA 10. GRAFICO Q-Q DE NORMALIDAD UNIVARIADA PARA LA FOTOGRAFIA S0101.PGM

FIGURA 11. GRAFICO Q-Q DE NORMALIDAD UNIVARIADA PARA LA FOTOGRAFIA S0108.PGM

Asimismo, en las figuras 10 y 11, se muestran a manera de ejemplo, dos de los gráficos Q-Q de la prueba de distribución normal. En ellos, se aprecia que los puntos se alejan de la recta, lo cual indica un inadecuado ajuste a una distribución normal.

Para el caso del “rostro promedio”, también se realizó la prueba de normalidad. En la tabla 4 se presenta el resultado que se obtuvo con el paquete SPSS. Se aprecia, que tampoco se ajusta a una distribución normal.

TABLA 4. PRUEBA DE NORMALIDAD UNIVARIADA PARA EL “ROSTRO PROMEDIO” DE LA BASE AT&T

Tests of Normality

	Kolmogorov-Smirnov ^a		
	Statistic	df	Sig.
PROFOTOS	.095	10304	.000

a. Lilliefors Significance Correction

4.3.2. Prueba de normalidad multivariada

La prueba de normalidad multivariada, es una prueba compleja de realizar. Pocos paquetes de software tienen la capacidad de llevarla a cabo. Entre los paquetes de software que permiten la prueba de normalidad multivariada, se encuentra PRELIS, que forma parte del software de análisis de ecuaciones estructurales denominado LISREL.

El análisis de normalidad multivariada para el problema que nos ocupa, requiere del uso de una matriz de 10304 renglones (píxeles de la fotografía digital de una

persona), por 400 columnas (número total de imágenes de la base AT&T), lo que implicaría el análisis de 4,121, 600 valores en la matriz.

Dicha matriz es demasiado grande para realizar una prueba de normalidad multivariada con los paquetes computacionales de que se disponen en la actualidad.

Por tanto, una forma de atacar el problema sin menoscabo de la calidad del resultado, consiste en realizar el análisis de normalidad multivariada a una muestra estadísticamente significativa de la base de datos AT&T.

Se calcula el tamaño de la muestra mediante:

$$n = \frac{N Z^2 p q}{(N - 1) e^2 + Z^2 p q} \quad (38)$$

en donde: se considera un error del 5%, una $Z = 1.96$ (95% de fiabilidad) y $p=q=0.5$.

Por tanto

$$n = \frac{400(1.96)^2 (0.5)(0.5)}{399(0.05)^2 + (1.96)^2 (0.5)(0.5)} = 35.12 \approx 8.78\%$$

El tamaño de la muestra correspondería a 35 fotografías, un 8.78% del total de imágenes. Para efectos del presente trabajo, se decide utilizar como tamaño de la muestra, un 10% de la base de datos (40 fotografías).

Seleccionando de forma aleatoria, una imagen (de las 10 posibles) de cada persona.

Utilizando el paquete PRELIS, se realiza el análisis respectivo, los resultados se pueden observar en la tabla 5.

TABLA 5. RESULTADOS DE LA PRUEBA DE NORMALIDAD MULTIVARIADA UTILIZANDO EL SOFTWARE PRELIS, A UNA MUESTRA DE LA BASE AT&T

DATE: 08/16/2005						
Test of Univariate Normality for Continuous Variables						
Variable	Skewness		Kurtosis		Skewness and Kurtosis	
	Z-Score	P-Value	Z-Score	P-Value	Chi-Square	P-Value
VAR 1	-69.035	0.000	40.364	0.000	6395.139	0.000
VAR 2	-20.957	0.000	-22.077	0.000	926.581	0.000
VAR 3	4.211	0.000	-14.038	0.000	214.785	0.000
VAR 4	0.126	0.900	-19.803	0.000	392.185	0.000
VAR 5	-2.100	0.036	-14.083	0.000	202.738	0.000
VAR 6	2.710	0.007	-16.669	0.000	285.205	0.000
VAR 7	15.173	0.000	-25.171	0.000	863.814	0.000
VAR 8	6.329	0.000	-28.065	0.000	827.728	0.000
VAR 9	2.107	0.035	-21.351	0.000	460.290	0.000
VAR 10	6.306	0.000	-28.793	0.000	868.821	0.000
VAR 11	11.936	0.000	-28.708	0.000	966.634	0.000
VAR 12	-1.438	0.151	-17.224	0.000	298.722	0.000
VAR 13	-5.472	0.000	-14.423	0.000	237.979	0.000
VAR 14	2.892	0.004	-26.919	0.000	732.999	0.000
VAR 15	13.641	0.000	-9.342	0.000	273.334	0.000
VAR 16	-15.779	0.000	-29.063	0.000	1093.667	0.000
VAR 17	6.689	0.000	-24.234	0.000	632.028	0.000
VAR 18	-9.455	0.000	-14.356	0.000	295.484	0.000
VAR 19	-5.603	0.000	-24.922	0.000	652.483	0.000
VAR 20	3.116	0.002	-20.490	0.000	429.535	0.000
VAR 21	8.641	0.000	-13.505	0.000	257.054	0.000
VAR 22	9.868	0.000	-3.492	0.000	109.576	0.000
VAR 23	8.795	0.000	-24.833	0.000	694.047	0.000
VAR 24	-9.383	0.000	-28.834	0.000	919.435	0.000
VAR 25	-1.551	0.121	-16.896	0.000	287.888	0.000
VAR 26	-2.147	0.032	-20.409	0.000	421.142	0.000
VAR 27	-20.250	0.000	-22.059	0.000	896.679	0.000
VAR 28	4.041	0.000	-27.476	0.000	771.275	0.000
VAR 29	8.438	0.000	-7.196	0.000	122.970	0.000
VAR 30	-2.654	0.008	-16.145	0.000	267.699	0.000
VAR 31	-3.641	0.000	-12.995	0.000	182.118	0.000
VAR 32	-24.811	0.000	-5.722	0.000	648.311	0.000
VAR 33	15.019	0.000	-8.230	0.000	293.293	0.000
VAR 34	-12.042	0.000	-3.224	0.001	155.407	0.000
VAR 35	1.580	0.114	-17.501	0.000	308.768	0.000
VAR 36	0.927	0.354	-28.069	0.000	788.745	0.000
VAR 37	17.617	0.000	-19.763	0.000	700.947	0.000
VAR 38	3.677	0.000	-21.022	0.000	455.451	0.000
VAR 39	20.049	0.000	-2.879	0.004	410.265	0.000
VAR 40	1.860	0.063	-19.313	0.000	376.451	0.000

Relative Multivariate Kurtosis = 1.140

Test of Multivariate Normality for Continuous Variables							
Skewness			Kurtosis			Skewness and Kurtosis	
Value	Z-Score	P-Value	Value	Z-Score	P-Value	Chi-Square	P-Value
227.265	509.023	0.000	1914.627	102.952	0.000	269703.839	0.000

De los resultados, al obtener un valor de la Curtosis Multivariada mayor a 1.0 (y con una significancia $p < 0.05$), se puede concluir: que la muestra de fotografías no se ajusta a una distribución normal multivariada. Y por tanto, de manera individual, la imagen de una persona tampoco presenta una distribución normal. Lo anterior, ya se había probado en la sección 4.3.1 y los resultados completos de SPSS se presentan en el anexo B.

Si bien, el estudio de normalidad sería conveniente realizarlo a una muestra más extensa de rostros humanos que incluya diversas razas. Estadísticamente, los resultados que se obtuvieron, indican en principio, que ningún rostro cumple con una distribución normal. Además, sería de esperarse que la mayor parte de los rostros de la población deberían ser diferentes entre ellos. Únicamente, en algunos casos, habría conjuntos significativos de personas que serían muy similares en sus rostros.

4.3.3. Prueba de homoscedasticidad y linealidad

Por lo que respecta a la prueba para comprobar la homogeneidad de las varianzas (homoscedasticidad), se puede llevar a cabo con el test de Levene. En el cual, se contrasta la hipótesis nula de la homogeneidad de varianzas. Se rechaza la hipótesis nula, si la significancia del estadístico F es menor que 0.05. Generalmente, se aplica esta prueba cuando se tiene una variable dependiente y una o más, variables independientes (o factores).

En el caso que nos ocupa, no se tienen variables dependientes e independientes, únicamente se está considerando al conjunto de valores de píxeles de cada fotografía. Por tanto, no se realizará dicha prueba.

Sin embargo, sin realizar procedimientos matemáticos, una forma de visualizar el cumplimiento de la homoscedasticidad y la linealidad es por medio del denominado gráfico denominado “nube de puntos” o scatter. La nube de puntos para los píxeles se presentan en las figuras 12 y 13.

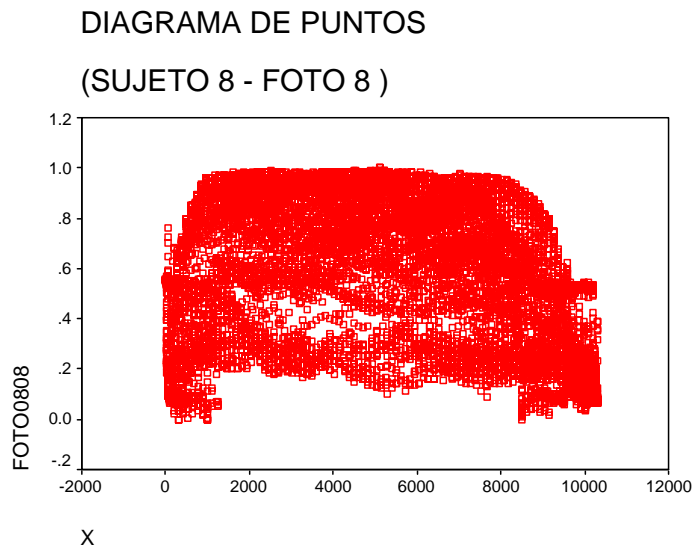


FIGURA 12. NUBE DE PUNTOS EN “X” PARA LOS PÍXELES DEL SUJETO 8/FOTO 8

DIAGRAMA DE PUNTOS
(SUJETO 8 - FOTO 8)

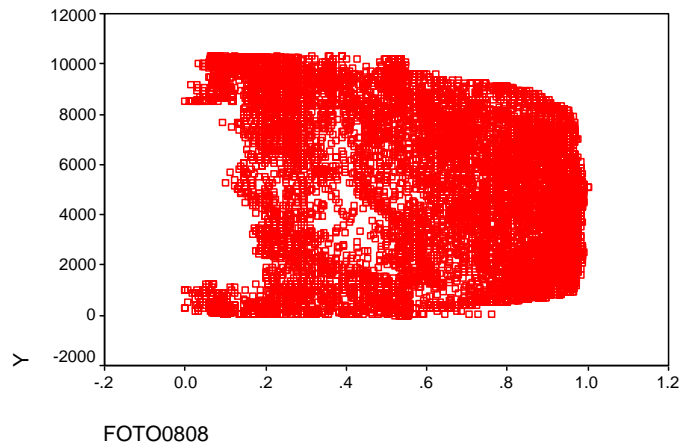


FIGURA 13. NUBE DE PUNTOS EN “Y” PARA LOS PÍXELES DEL SUJETO 8/FOTO 8

En las figuras, se puede observar, la inmensa cantidad de puntos (píxeles) que se tienen en una fotografía. Lo anterior, genera un verdadero conglomerado que no permitiría obtener un adecuado planteamiento de “linealidad”. Por ejemplo: el tratar de aplicar procedimientos de regresión, en donde (como se sabe) se intenta ajustar una recta o curva al conjunto de puntos, sería imposible. Se observa, que debido a la gran cantidad de información, no se vislumbra la posible curva que aplicaría. Por tanto, se considera que los píxeles no cumplen adecuadamente el supuesto de linealidad.

DIAGRAMA DE PUNTOS 3-D

(SUJETO 8 - FOTO 8)

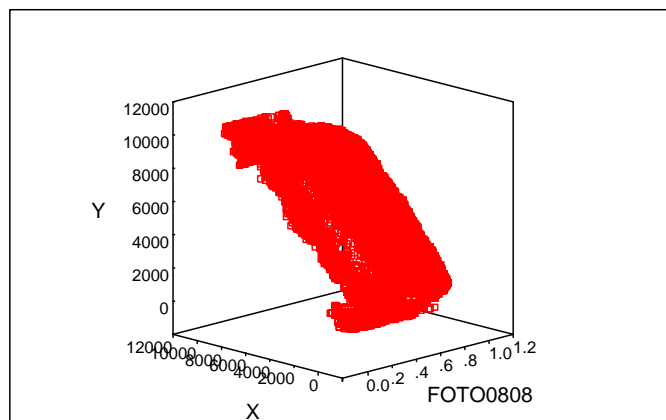


FIGURA 14. NUBE DE PUNTOS 3-D PARA LOS PÍXELES DEL SUJETO 8/FOTO 8

En la figura 14, se presenta el diagrama 3-D, de los píxeles correspondientes a la octava fotografía del sujeto 8 de la base AT&T. Se observa, la configuración de un plano muy saturado. Lo cual, también proporciona indicios, de una falta de linealidad a nivel d-dimensional. Se requeriría reducir la cantidad de información, para tener la posibilidad de aplicar técnicas que requieren los supuestos de linealidad.

Capítulo 5

Análisis del reconocimiento de rostros utilizando técnicas de componentes principales

Parte de éste capítulo es producto de los resultados y las ideas planteadas en dos reportes de investigación presentados por el autor del presente trabajo de tesis, y que se mencionan a continuación: “Reconocimiento de rostros utilizando componentes principales: Algunas limitaciones”¹ y “Reconocimiento de rostros utilizando Wavelets y Análisis de Componentes Principales”²

Este capítulo tiene como propósito, plasmar los resultados que se obtuvieron al realizar experimentos de reconocimiento de rostros utilizando la técnica denominada Componentes Principales (ya mencionada en el capítulo 3), así, como el análisis que se llevó a cabo para obtener algunas ventajas y desventajas que puede presentar la técnica.

Si bien, diversos autores han utilizado con anterioridad la técnica de componentes principales para el reconocimiento de rostros y de hecho, algunos de los mejores sistemas con que se cuenta actualmente están basados en dicha técnica o alguna extensión de ella, se

¹ Villegas, Carlos (2004). “Reconocimiento de rostros utilizando componentes principales: Algunas limitaciones”, Seminario de Visión por Ordenador, Department d’ Enginyeria de Sistemes, Automàtica i Informàtica Industrial, Universitat Politècnica de Catalunya, España.

² Villegas, Carlos (2005). “Reconocimiento de rostros utilizando Wavelets y Análisis de componentes principales”, Seminario de Reconocimiento de Patrones, Department d’ Enginyeria de Sistemes, Automàtica i Informàtica Industrial, Universitat Politècnica de Catalunya, España.

considera que presenta diversas limitaciones, reportadas únicamente de manera somera en algunas investigaciones.

El análisis de algunas de las limitaciones que presenta la técnica de componentes principales en el reconocimiento de rostros, permitirá contestar a las preguntas de investigación planteadas en la tesis, así como a la hipótesis planteada.

Las ventajas y desventajas que se pretenden obtener con el análisis de los experimentos, se utilizarán como elementos fundamentales a tomar en cuenta para el desarrollo de un proyecto de investigación que se encuentra realizando el autor de la presente tesis. Dicho proyecto, pretende contribuir a subsanar alguna o algunas de las limitaciones que tenga la técnica de componentes principales (y posiblemente, alguno de los otros métodos que se utilizan actualmente para el reconocimiento de rostros).

5.1. Características de la base de datos de rostros utilizada para la experimentación.

Se utilizaron imágenes de rostros de la base de datos desarrollada por **AT&T Laboratories, Cambridge**.³ Dicha base está constituida por imágenes de 40 personas, con 10 fotografías (en diferentes poses) para cada sujeto. Cada imagen tiene una resolución de 92 x 112 píxeles. Una fotografía de la base AT&T, requiere de 10,304 píxeles para almacenarse. La base de datos de AT&T ha sido utilizada por diversos autores para probar diversos programas computacionales sobre reconocimiento de rostros.

³ Información sobre la base de datos de AT&T y la base en sí, se puede recuperar en la siguiente dirección de Internet: <http://www.uk.research.att.com/facedatabase.html>

Entre las características de la base de rostros de AT&T, se encuentran:

- Se tienen 36 personas masculinas y 4 femeninas
- Una de las personas es de raza negra.
- Las personas presentan diversas edades.
- Las fotografías se tomaron en diferentes días.
- La cantidad de luz que recibe la persona se encuentra controlada, presentando algunas variaciones ligeras.
- Todas las fotografías presentan el mismo fondo por detrás del sujeto.
- Se presentan diversas expresiones faciales: ojos abiertos/cerrados, sonrisa/no sonrisa
- 15 personas tienen anteojos.
- 7 personas presentan barba, de las cuales, 4 tienen barba y anteojos
- 1 persona tiene bigote
- De las personas que tienen anteojos, cinco de ellos, presentan algunas fotografías sin anteojos.
- Todas las imágenes son frontales (tipo fotografía para credencial o pasaporte), algunas de las fotografías presentan ligeras variaciones angulares del rostro.

Esta base de datos ha sido utilizada en diversas investigaciones sobre reconocimiento de rostros, utilizando diversos métodos (análisis estadístico, elementos geométricos, redes

neuronales, regresión multivariada, etc.). Por tal motivo, se tiene la posibilidad de realizar algunas comparaciones utilizando los resultados que se obtengan.

La galería completa de los 400 rostros que componen la base de datos de AT&T se puede observar en la figura 15. Una visualización más detallada de cada fotografía, se puede encontrar en el anexo A, en donde se muestran las 360 fotografías utilizadas como base de entrenamiento. Las imágenes del anexo A, junto con las 40 fotografías que se presentan en las figuras 16 y 17, configuran el total de la base AT&T (400 imágenes).

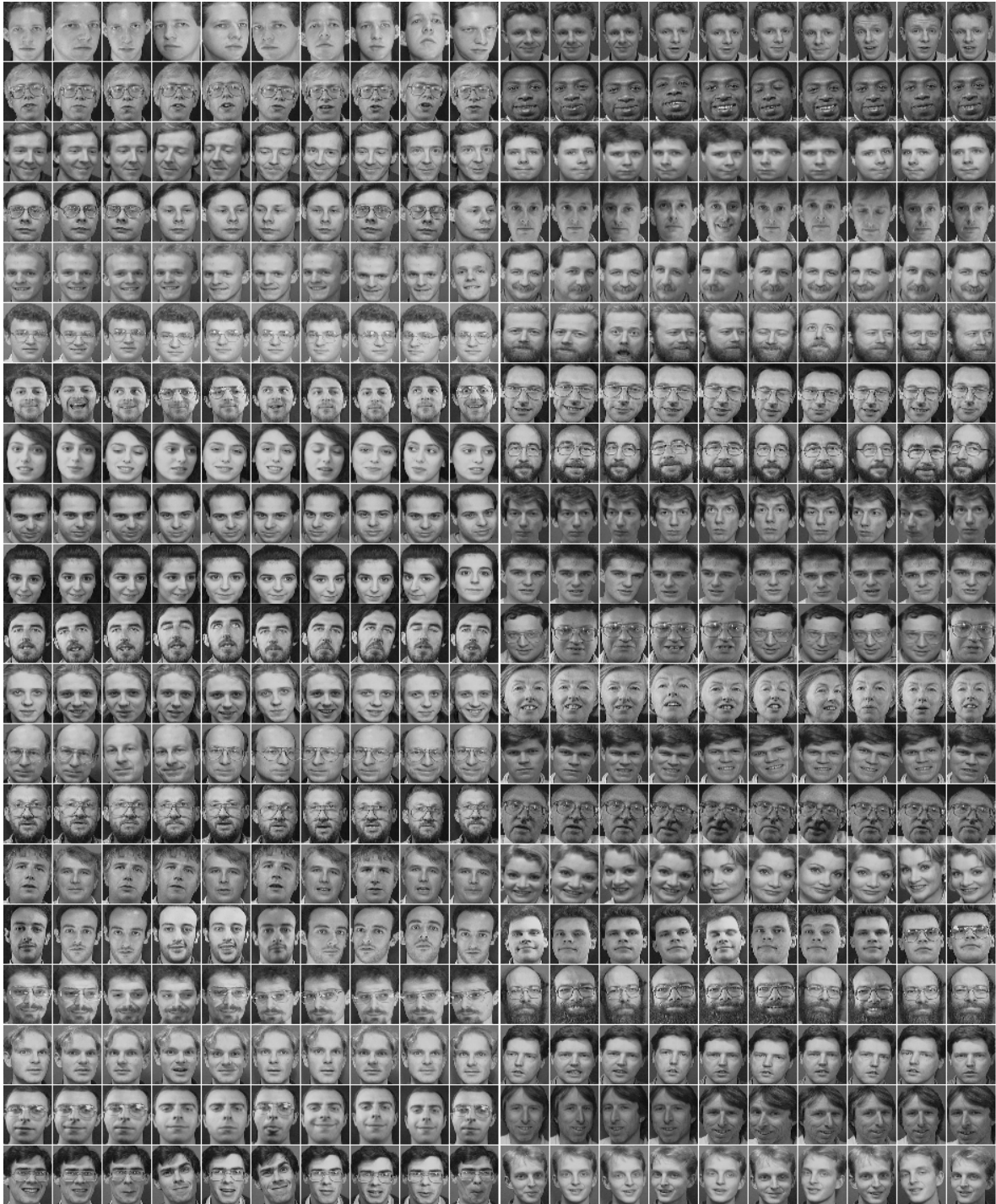


FIGURA 15. CONFIGURACIÓN DE LA BASE DE ROSTROS AT&T CONTENIENDO FOTOGRAFÍAS DE 40 SUJETOS CON 10 IMÁGENES POR CADA PERSONA

5.2. Elaboración de la base de entrenamiento

Se realizaron diferentes experimentos de reconocimiento utilizando diversos números de fotografías de la base AT&T para configurar la base de entrenamiento. Asimismo, se incluyeron en algunos de los experimentos, fotografías de delincuentes y terroristas buscados por el FBI. Por lo que respecta a los rostros por reconocer, como ya se indicó anteriormente, la primera fotografía de cada persona de la base de AT&T se utilizó para el procedimiento. Adicionalmente, se incluyeron fotografías de personas que no se encuentran en la base AT&T y en algunos de los experimentos (utilizando únicamente 33 de los sujetos como la base de entrenamiento), se utilizaron 7 fotografías de AT&T que no se consideraron como parte de la base de entrenamiento

Las fotografías de la base de rostros AT&T se encuentran en formato PGM (Portable Gray Map), se decidió conservar el formato original para el desarrollo de la base de entrenamiento.

Las 10 fotografías de cada sujeto se encuentran en subdirectorios separados. Y la simbología que se utilizó para el desarrollo del proyecto es la siguiente:

Cada nombre asignado a las fotografías presenta el nombre : **SXXTT.pgm**

En donde:

XX = indica el número de la persona (de 01 a 40)

TT = número de fotografía (01 a 10)

Por ejemplo: S0108.pgm, se refiere a la octava fotografía de la primera persona.

Se asignaron de forma arbitraria, nombres a cada uno de los 40 rostros de la base de datos de sospechosos. El propósito es únicamente ilustrar la posibilidad de encontrar el Nombre de la persona a reconocer y diversos atributos que se le pueden asociar (dirección, Teléfono, edad, nacionalidad, etc.).

Se calcularon 10 bases de entrenamiento, correspondiendo a las diferentes configuraciones de experimentos, los cuales se detallan en la siguiente sección.

5.3. Configuración de los experimentos realizados:

Con las diversas bases de entrenamiento utilizadas en la experimentación, siempre se tratará de reconocer 53 imágenes: 40 imágenes pertenecientes a la base AT&T (primera fotografía de cada persona), y 13 imágenes adicionales, que el sistema debería de indicar que no se encuentran (negativo verdadero). En las figuras 16 y 17 se pueden observar las fotografías de la base de rostros AT&T utilizadas como imágenes a reconocer. Se asume que dichas fotografías serían el equivalente a las imágenes que se le tomarían a la persona a la entrada de algún lugar restringido para poder ingresar o en un aeropuerto, entre otros.

Las 13 imágenes adicionales para identificar se pueden observar en la figura 18. Estas fotografías presentan imágenes de algunos rostros, una caricatura de un rostro, el dibujo a lápiz de un rostro, una máscara, la imagen de rostros en piedra (la montaña Rushmore), una

pareja bailando, una persona en una motocicleta, un sujeto disfrazado de mago, la imagen de un supuesto rostro en el planeta Marte y una ventana enclavada en un muro.

Si bien, se asumiría que un sistema para reconocer rostros, únicamente se le presentarían fotografías de personas; en los experimentos se incluyen algunas imágenes que no corresponden a rostros humanos (como se mencionó en el párrafo anterior), con el propósito de validar con más rigor el sistema computacional de reconocimiento.

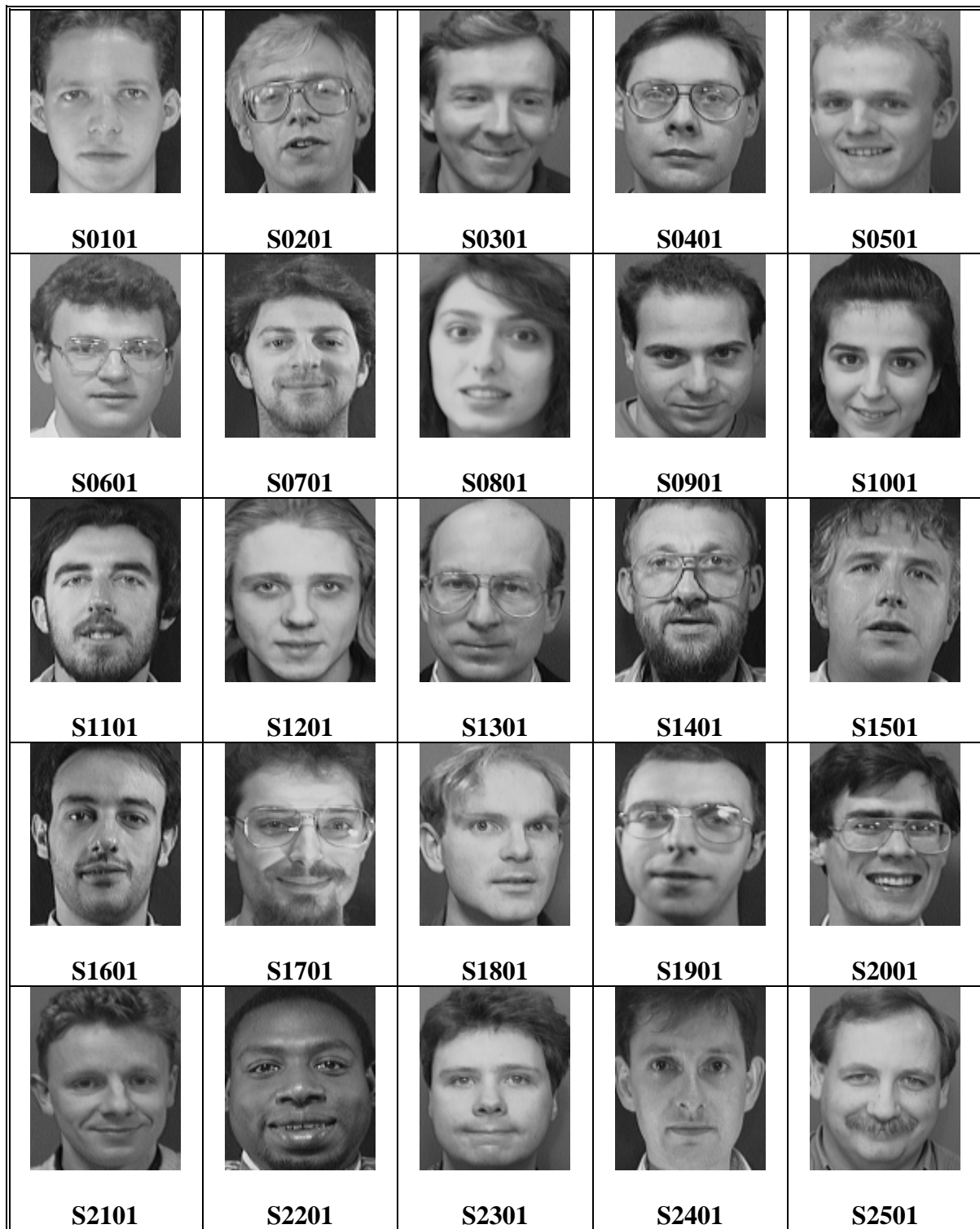


FIGURA 16. LOS PRIMEROS 25 ROSTROS DE LA BASE AT&T UTILIZADOS COMO “FOTOGRAFÍAS A IDENTIFICAR”

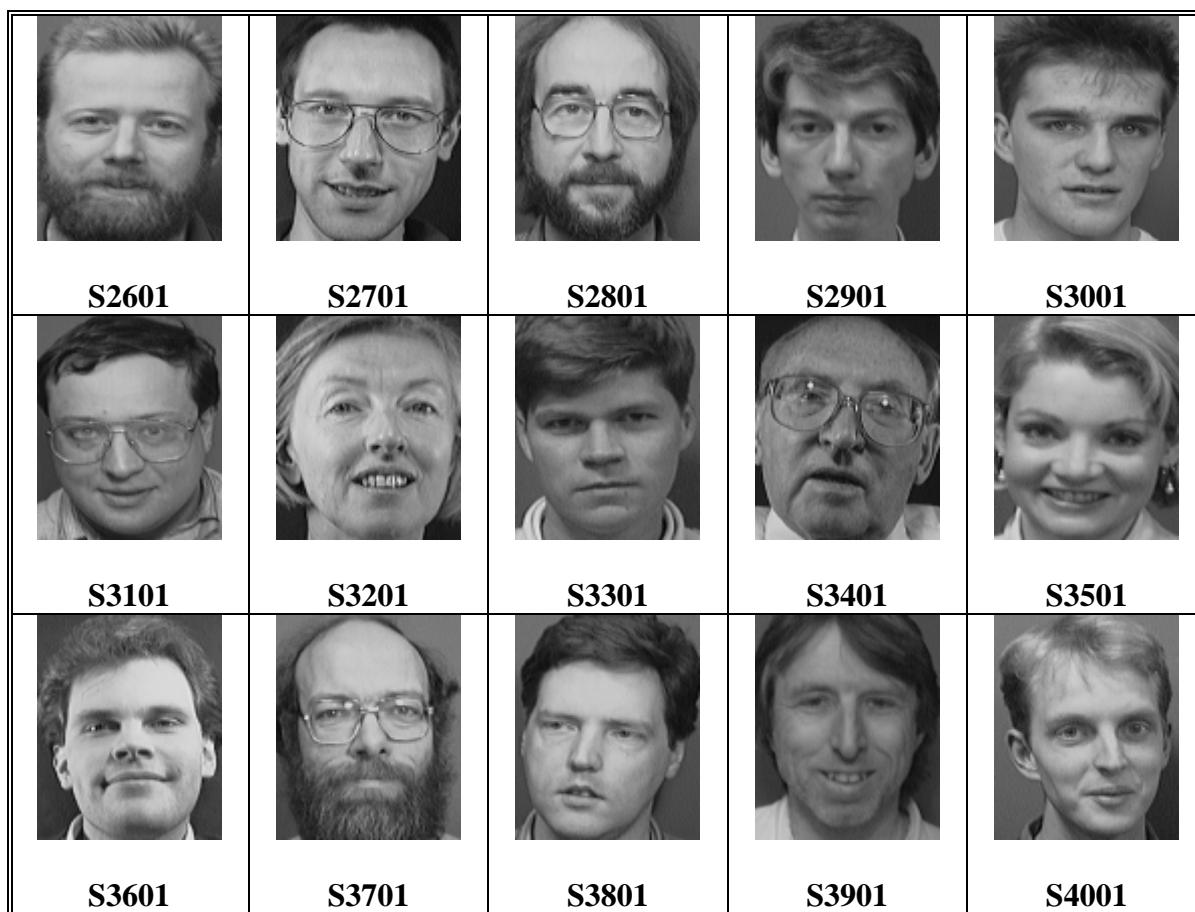


FIGURA 17. LOS SIGUIENTES 15 ROSTROS DE LA BASE AT&T UTILIZADOS COMO “FOTOGRAFIAS A IDENTIFICAR”, PARA DAR UN TOTAL INICIAL DE 40 FOTOGRAFIAS A RECONOCER

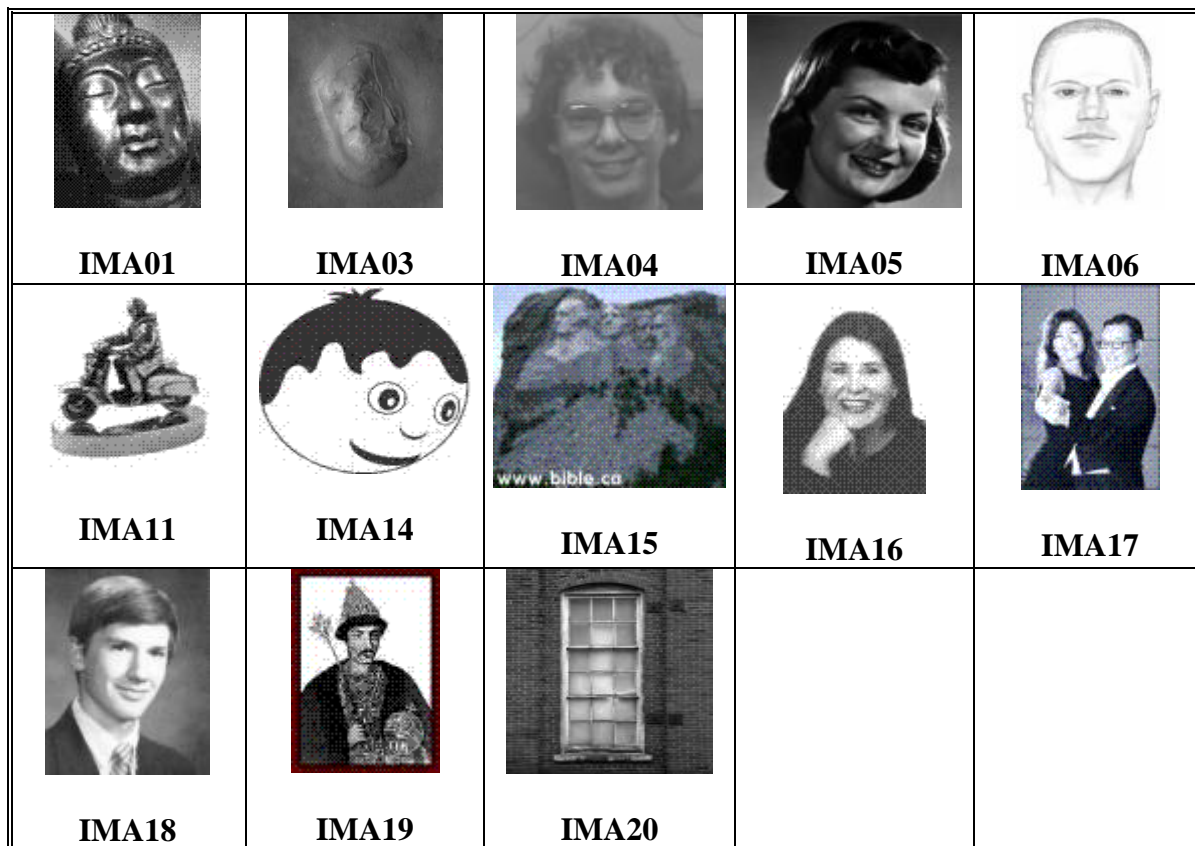


FIGURA 18. TRECE FOTOGRAFÍAS ADICIONALES PARA SER UTILIZADAS COMO IMÁGENES A IDENTIFICAR

Los experimentos se realizaron disminuyendo el número de fotografías similares por sujeto que se utilizaron para “entrenar” la base de datos para reconocimiento. Se utilizó un máximo de 9 fotografías por sujeto y un mínimo de 1 fotografía. En aplicaciones reales, generalmente se cuenta únicamente con una fotografía por persona para elaborar la base de datos utilizada como entrenamiento del sistema. Por ejemplo, la base de datos del IFE, la base de datos de licencias para conducir o la base de datos de las fotografías de pasaportes, entre otras.

Sin embargo, a partir de los actos terroristas del 11 de septiembre de 2001 contra los Estados Unidos, dicho país ha implementado fotografiar a la mayoría de las personas que llegan a territorio estadounidense. Por tanto, se considera que tendrán un mínimo de 3 fotografías por sujeto: la imagen de la “visa”, fotografía del pasaporte y la fotografía que se tomará en las aduanas americanas.

A continuación se describen los experimentos y el número de imágenes consideradas en cada uno de ellos.

Experimentos con 40 personas de la base AT&T

EXP01. Experimento 1.

Base de entrenamiento:	9 fotografías por sujeto (base AT&T)
Total de imágenes de la B.D.	360
Imágenes a reconocer:	53 fotografías: 40 de la base AT&T (1era. Fotografía) 13 fotografías adicionales

EXP02. Experimento 2.

Base de entrenamiento:	8 fotografías por sujeto (base AT&T)
Total de imágenes de la B.D.	320
Imágenes a reconocer:	53 fotografías: 40 de la base AT&T (1era. Fotografía) 13 fotografías adicionales

EXP03. Experimento 3.

Base de entrenamiento:	5 fotografías por sujeto (base AT&T)
Total de imágenes de la B.D.	200
Imágenes a reconocer:	53 fotografías: 40 de la base AT&T (1era. Fotografía) 13 fotografías adicionales

Experimentos con 33 personas de la base AT&T

En estos experimentos se utilizaron como parte de la base de datos a reconocer, las fotografías de las 7 personas que no se utilizaron de la base AT&T. Dichas fotografías corresponden a las claves: S09, S18, S22, S33, S38, S39, S40.

Por tanto, la base de fotografías a reconocer quedó constituida por 53 fotografías: 33 de la base AT&T, 7 sujetos de la base AT&T (no utilizados) y las 13 fotografías adicionales

EXP04. Experimento 4.

Base de entrenamiento: 9 fotografías por sujeto (base AT&T)
Total de imágenes de la B.D. 297

Imágenes a reconocer: 53 fotografías:
33 de la base AT&T (1era. Fotografía)
7 de las personas no utilizadas AT&T
13 fotografías adicionales

EXP05. Experimento 5.

Base de entrenamiento: 7 fotografías por sujeto (base AT&T)
Total de imágenes de la B.D. 231

Imágenes a reconocer: 53 fotografías:
33 de la base AT&T (1era. Fotografía)
7 de las personas no utilizadas AT&T
13 fotografías adicionales

EXP06. Experimento 6.

Base de entrenamiento: 5 fotografías por sujeto (base AT&T)
Total de imágenes de la B.D. 165

Imágenes a reconocer: 53 fotografías:
33 de la base AT&T (1era. Fotografía)
7 de las personas no utilizadas AT&T
13 fotografías adicionales

EXP07. Experimento 7.

Base de entrenamiento: 3 fotografías por sujeto (base AT&T)
Total de imágenes de la B.D. 99

Imágenes a reconocer: 53 fotografías:
33 de la base AT&T (1era. Fotografía)
7 de las personas no utilizadas AT&T
13 fotografías adicionales

EXP08. Experimento 8.

Base de entrenamiento: 2 fotografías por sujeto (base AT&T)
Total de imágenes de la B.D. 66

Imágenes a reconocer: 53 fotografías:
33 de la base AT&T (1era. Fotografía)
7 de las personas no utilizadas AT&T
13 fotografías adicionales

EXP09. Experimento 9.

Base de entrenamiento: 1 fotografías por sujeto (base AT&T)
Total de imágenes de la B.D. 33

Imágenes a reconocer: 53 fotografías:
33 de la base AT&T (1era. Fotografía)
7 de las personas no utilizadas AT&T
13 fotografías adicionales

Experimento utilizando fotografías adicionales de algunos delincuentes y terroristas reportados por el FBI

EXP10. Experimento 10.

Base de entrenamiento: 1 fotografía para 22 sujetos (base AT&T)
1 fotografía de 10 delincuentes (FBI)
1 fotografía que se obtuvo con escáner.

Total de imágenes de la B.D. 33

Imágenes a reconocer: 53 fotografías:
22 de la base AT&T (1era. Fotografía)
15 de fotografías del FBI
16 fotografías adicionales de AT&T

5.4. Características, algoritmos y funcionamiento del sistema

5.4.1. Características generales del sistema

A partir de las “n” imágenes utilizadas para cada uno de los diversos experimentos, se configura la base de datos para entrenamiento. Por ejemplo, para el experimento EXP04 (con 33 sujetos y 9 fotos por cada uno), se utilizó un arreglo de $112 \times 92 \times 297$, lo cual permite acceder fácilmente a cada una de las fotografías. Asimismo, se asignaron nombres ficticios a cada una de las personas, con el propósito de recuperar dicha información al realizar el reconocimiento.

Se desarrolló un programa en MATLAB, que utiliza la ToolBox de Procesamiento de Imágenes para llevar a cabo la experimentación de reconocimiento de rostros utilizando la técnica de Componentes Principales. El programa utilizó el método planteado en la sección 3.4 de la presente tesis.

5.4.2. Algoritmos

El algoritmo empleado es la técnica de Componentes Principales (técnica descrita en el capítulo 3), método bien definido y utilizado en diversas áreas de aplicación.

Con el propósito de manejar dicha técnica, se generó un vector de 10304 (92×112) posiciones para cada fotografía. Por ejemplo, para el experimento EXP04, se generó una matriz de 10304×297 (matriz D).

Enseguida, se obtienen valores “promedio” de imagen en la matriz M . Dichos valores promedio, se restan a cada columna de D y se almacenan en la matriz S . A continuación se obtiene la matriz de covariancia $C = S \bullet S^T$. Sin embargo, para reducir el tiempo de proceso, se transforma el cálculo de la covariancia a la forma $A = S^T \bullet S$. A partir de ella, se utilizan funciones Matlab para obtener los eigenvalues y se normalizan los resultados.

Utilización de componentes principales para clasificar un rostro

Una vez que se obtienen los componentes principales a partir de la base de entrenamiento, la identificación de un determinado rostro, será un proceso de reconocimiento de patrones. Así, los componentes principales más significativos configuran un subespacio dimensional de tamaño M' . Los M' componentes seleccionados, serán aquellos que presentan los valores más altos de varianza. En el caso presente, se han seleccionado 30 componentes (o eigenvalues).

Con lo anterior, el rostro a identificar (F), se transforma en sus componentes de tipo eigenvalue, al proyectar el rostro en el denominado “espacio de rostros” (base de entrenamiento transformada en eigenvalues), utilizando la operación

$$w_k = \mathbf{u}_k^T (F - \psi) \quad \text{para } k = 1, \dots, M'$$

Los “pesos” w_k forman un vector $\Omega^T = [w_1 \ w_2 \ w_3 \ \dots \ w_{M'}]$, el cual describe las contribuciones de cada componente (eigenvalue) para representar la imagen de entrada de un

rostro, utilizando los eigenvalues como un conjunto base para el total de las imágenes de los rostros.

El vector se utiliza como una forma (reconocimiento de patrón) de encontrar una determinada “clase” a la que pertenecen diversos rostros de la base de entrenamiento. Por tanto, el objetivo consiste en encontrar por una parte, si el rostro a identificar pertenece a alguna de las “clases” (la clase que mejor describe al rostro).

Un método para determinar cual es la clase que proporciona la mejor descripción para el rostro de entrada a identificar, consiste en encontrar la clase k que minimice la distancia Euclideana $\varepsilon_k = \|(\Omega - \Omega_k)\|$, en donde: Ω_k es un vector que describe la k -ésima clase de rostros y, Ω describe el rostro a identificar. Así, un rostro se clasifica como perteneciente a la clase k , cuando la distancia mínima ε es menor que un determinado valor de umbral θ (propuesto por el usuario del sistema, dentro de ciertos límites establecidos).

<i>Si</i>	$\varepsilon \leq \theta$	
<i>Entonces</i>		ROSTRO_IDENTIFICADO
<i>Encasocontrario</i>		ROSTRO_NO_IDENTIFICADO

Espacio de rostros

De manera adicional al procedimiento planteado anteriormente para reconocer un rostro, se pueden presentar a su vez, varias alternativas según se localice el rostro a identificar en el espacio M -dimensional de componentes principales. Dicho espacio se le denomina generalmente, “espacio de rostros” (figura 19).

Así, la imagen de un rostro y en particular los rostros que pertenecen a la base de entrenamiento, deberían de estar cercanos al espacio de rostros. Y por tanto, los rostros de sospechosos (fotografías de sujetos en la base de entrenamiento), deben de proyectarse dentro del espacio de rostros y muy cercanos a la correspondiente clase de rostro. Se pueden presentar 4 posibilidades (los números corresponden a los puntos indicados en la figura 19):

1 Rostro reconocido como tal e identificado como sospechoso. Cuando el rostro de entrada se proyecta de manera cercana al espacio de rostros y cercano a la *clase* respectiva. Sería el equivalente a lo que se denomina **Verdadero Positivo**.

2 Rostro reconocido como tal, pero no identificado. Cuando el rostro de entrada se proyecta de manera cercana al espacio de rostros, pero no está cercano a una determinada *clase*. Es el equivalente a un rostro no reconocido en la base –**Verdadero Negativo**–.

3 Rostro distante del espacio de rostros, pero cercano a una clase. Se identifica un rostro de entrada como similar a un rostro de la base, pero la identificación es realmente falsa. Corresponde a un **Falso positivo**.

4 Rostro distante del espacio de rostros y alejado de alguna clase. La imagen correspondería a un objeto (no es un rostro). O bien, puede ser un rostro, pero por sus características de imagen (luz, fondo, etc.), es muy diferente a las características que presenta la mayor parte de los rostros de la base de entrenamiento y por tanto, se

identifica como muy lejano. En general se identifica como *Rostro no reconocido*, lo cual puede corresponder en realidad a un **Falso Negativo** o bien, a un **Objeto**.

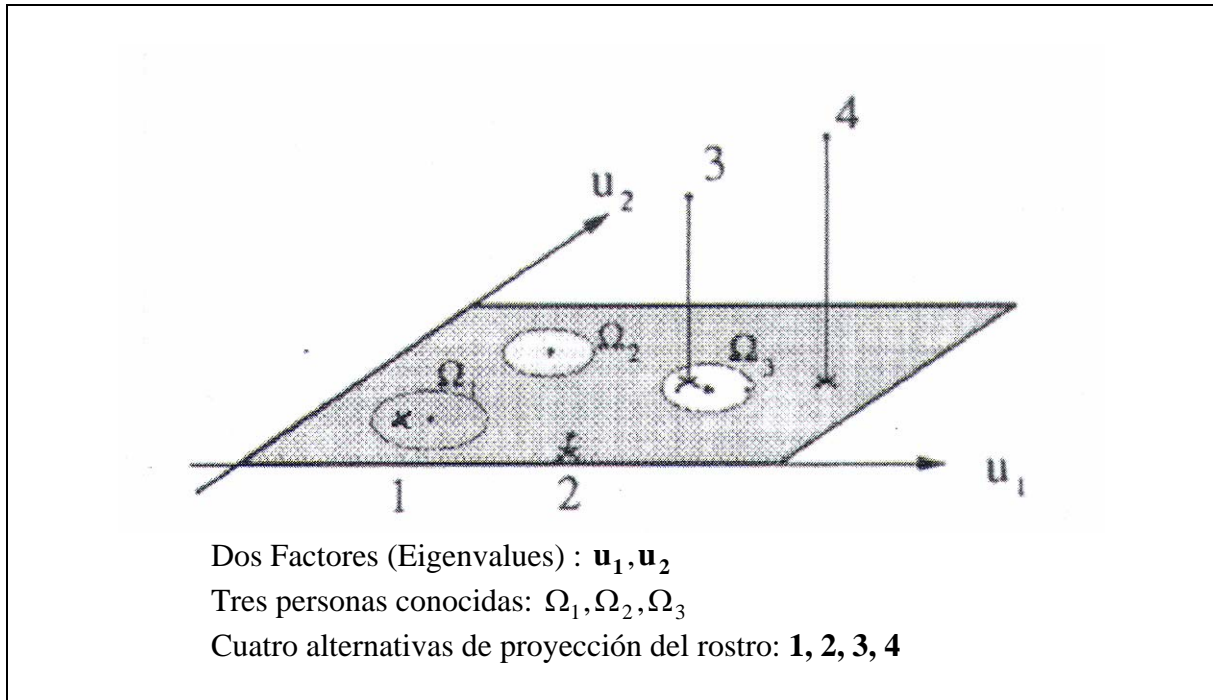


FIGURA 19. MODELO SIMPLIFICADO DEL “ESPACIO DE ROSTROS” QUE SE OBTIENE AL REALIZAR UN ANÁLISIS DE COMPONENTES PRINCIPALES, ILUSTRANDO LOS CUATRO RESULTADOS POSIBLES QUE SE OBTIENEN AL PROYECTAR LA IMAGEN DE UN ROSTRO EN EL ESPACIO

5.4.3. El sistema computacional para reconocimiento de rostros

Con base a los algoritmos planteados anteriormente y la técnica matemática de componentes principales, se elaboró un sistema computacional para el reconocimiento de rostros utilizando MATLAB y su ToolBox de Procesamiento de Imágenes.

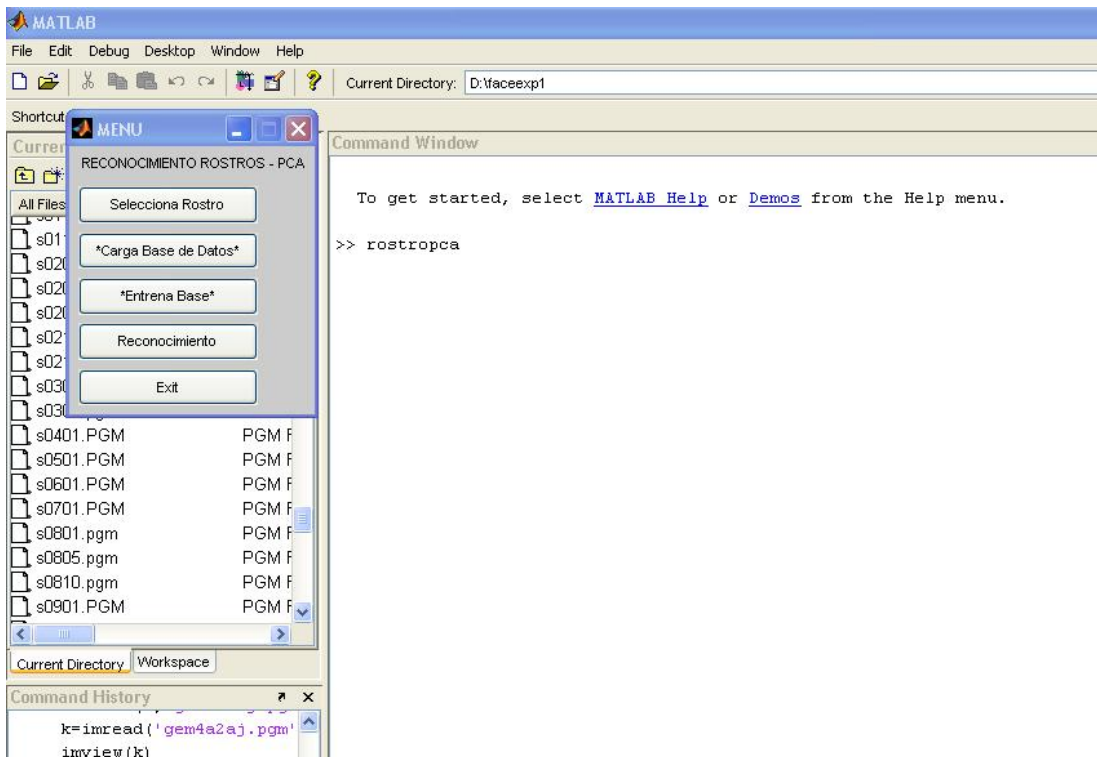


FIGURA 20. EJECUCION DEL SISTEMA DE RECONOCIMIENTO Y MENU PRINCIPAL

La función principal en Matlab que inicia el programa, se denominó *rostropca* (figura 20).

Al invocar el procedimiento, se obtiene el menú que se observa en la misma figura y por medio del cual se pueden realizar los siguientes procesos:

- **Carga la base de Datos** con el conjunto de fotografías que se utilizará para el entrenamiento, es el primer paso del proceso.
- **Entrena base**, procedimiento que utiliza el análisis de componentes principales para “entrenar la base”. El tiempo de procesamiento promedio para entrenar una base con 360 imágenes (92 x 112 pixeles), fue entre 5 y 6 minutos.⁴

El procedimiento anterior, únicamente se lleva a cabo una sola vez para el conjunto de fotografías iniciales. Cada vez que se tuvieran nuevas fotografías de sospechosos para adicionar a la base, se tendría que ejecutar nuevamente el procedimiento de “entrenamiento”. Conforme el número de fotografías aumenta, el tiempo de procesamiento se incrementará. Si consideramos un tiempo de 5 minutos para procesar las 360 fotografías y un tiempo de procesamiento lineal (no necesariamente cierto...), requerimos de 13 minutos para entrenar una base de 1,000 fotografías y de 2.31 horas para 10,000 imágenes. Obviamente, dicho tiempo dependerá también, de la velocidad de la computadora utilizada. Sin embargo, para efectos de un software comercial, se tendría que investigar algún método para reducir el tiempo de entrenamiento (independientemente de las características del hardware). Lo anterior, queda fuera de los objetivos del presente trabajo, pero es un elemento importante a tomar en cuenta.

Continuando con la descripción del sistema en Matlab, una vez “entrenada la base”, las otras opciones que presenta el sistema son (figura 20).

⁴ Se utilizó una computadora SONY Vaio con un procesador Intel Pentium a 1.8 GHz y 512 MB de RAM.

- **Selecciona Rostro** a reconocer, indicando el nombre de alguna imagen que se utilizará como dato al sistema (por ejemplo: S2301.PGM). En un caso real, correspondería a la fotografía que se le está tomando a la persona en algún acceso restringido (aeropuerto, oficina, cajero automático, etc.).
- **Reconocimiento**, corresponde al procesamiento de la imagen seleccionada , para verificar si es reconocida como alguna de las imágenes que se encuentran en la base de entrenamiento.
- **Exit**, la opción para abandonar el sistema de reconocimiento.

A continuación, se describen brevemente, las principales rutinas que configuran el programa de reconocimiento de rostros.

Rutinas del sistema de reconocimiento de rostros

Las rutinas del programa de reconocimiento de rostros se han denominado:

rostropca.m	función principal
generaDBrostro.m	genera la base de datos de rostros
entrena.m	realiza el entrenamiento utilizando PCA
reconoce.m	realiza el reconocimiento del rostro
sorteigen.m	función auxiliar para clasificación de eigenvalues

5.4.4. Operación y tipo de resultados que proporciona el sistema

Una vez entrenada la base, se procede a las opciones de selección de rostro y reconocimiento del mismo. Como ilustración de los posibles resultados que proporciona el sistema, se presentan los siguientes ejemplos:

En la figura 21, se ilustra el proceso cuando se ha seleccionado la fotografía de una persona a reconocer. Específicamente, el ejemplo corresponde a la imagen del delincuente William Junior Jordan, delincuente buscado por el FBI (figura 31).

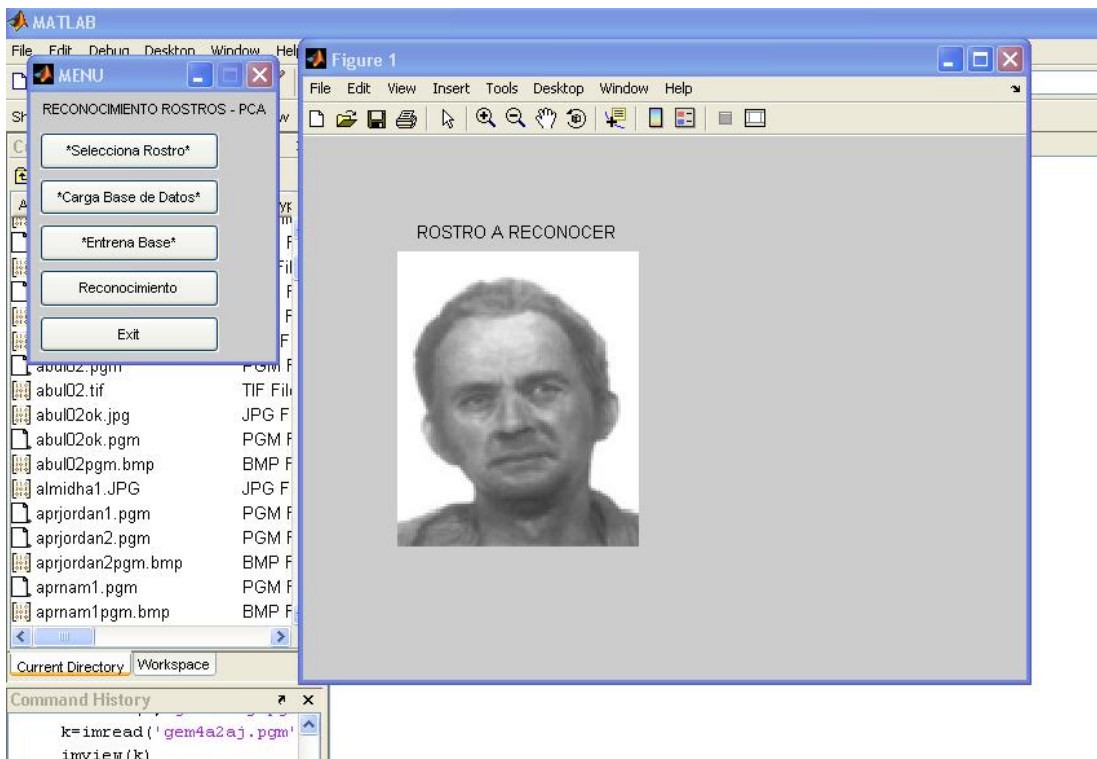


FIGURA 21. SELECCIÓN DE UNA FOTOGRAFIA A RECONOCER

Enseguida, se selecciona la opción de reconocimiento en el programa. El resultado que muestra el proceso se presenta en la figura 22.

Se puede observar, que el sistema identificó a la persona como **ApJordan**, nombre que se le dio al sujeto en la base de entrenamiento. El procedimiento accede a la fotografía que para el sistema, se encuentra “más cercana” a la imagen de entrada, de acuerdo al análisis de componentes principales y al umbral seleccionado.

Se puede constatar (figura 31), que la presente identificación corresponde a lo que se ha denominado **Verdadero Positivo**. La persona identificada, realmente corresponde a un sospechoso que se encuentra en la base de entrenamiento.

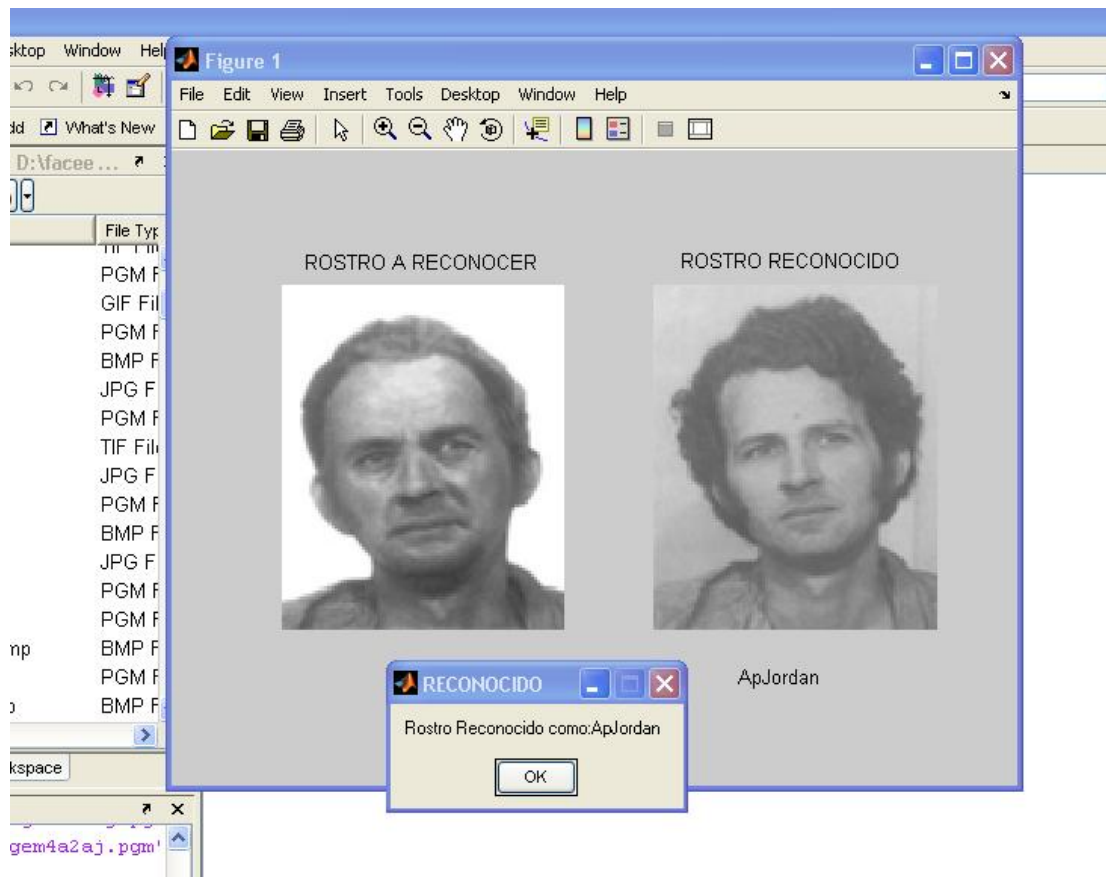


FIGURA 22. RECONOCIMIENTO DE UN ROSTRO (VERDADERO POSITIVO)

En las figuras 23 y 24, se presentan otros ejemplos de reconocimiento verdadero a partir de fotografías que se obtuvieron del portal internet del FBI.

La figura 23 corresponde al delincuente Decharmon, se puede observar que la fotografía que se utiliza como dato, correspondería a una imagen tomada algunos años posteriores a la imagen que se encuentra en la base de entrenamiento. La persona presenta anteojos, un peinado más corto y algo diferente, el cabello se nota más entrecano. El sistema reconoció adecuadamente al sujeto.

En la figura 24, se realiza el reconocimiento del terrorista Fahid Mohammed Ally Msalam, buscado por estar involucrado en la colocación de bombas en las embajadas americanas de dar es Salam y Nairobi. En este caso, las fotografías utilizadas no están controladas en cuanto al nivel de luz y fondo de la imagen. La identificación es efectivamente, un Verdadero Positivo, como se puede constatar con el reporte del FBI que se muestra en la figura 32.



FIGURA 23. RECONOCIMIENTO DE UN ROSTRO (VERDADERO POSITIVO). EL SUJETO A RECONOCER PRESENTA ANTEOJOS Y MAYOR EDAD.



FIGURA 24. RECONOCIMIENTO DE UN ROSTRO (VERDADERO POSITIVO). LAS FOTOGRAFÍAS UTILIZADAS NO SE ENCUENTRAN CONTROLADAS (LUZ, FONDO DE IMAGEN, RESOLUCION, ETC.)

Por lo que respecta a las imágenes de la base AT&T, en la figura 25, se muestra también un Verdadero Positivo para la octava persona de la base.

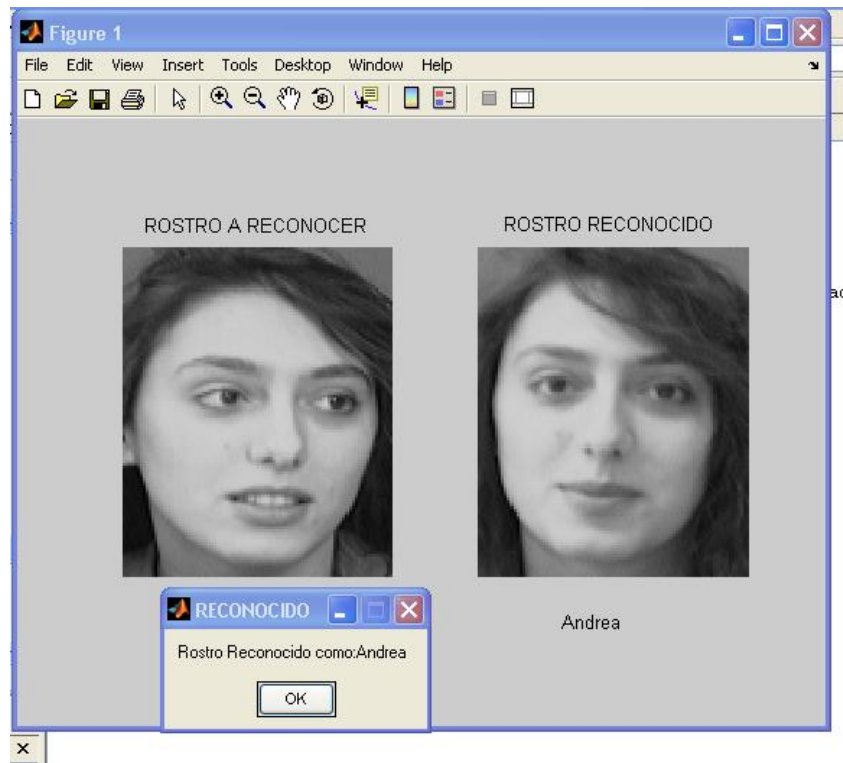


FIGURA 25. RECONOCIMIENTO “VERDADERO POSITIVO” DE UNA FOTOGRAFIA CONTROLADA DE LA BASE AT&T

Si se intenta reconocer un objeto o “algo” (figura 26) que no corresponde a un rostro (revisar la sección del “espacio de rostros”), el sistema indica el resultado como **Rostro no reconocido**. Lo anterior corresponde a un Verdadero Negativo. Se tendría la posibilidad de que el sistema indicara un mensaje de “fotografía no corresponde a un rostro”. Sin embargo, no se contempló la programación de tal opción, se considera que un sistema real de reconocimiento, no se le presentarían imágenes que no fueran de personas.

Sin embargo, el sistema que se desarrolló clasifica adecuadamente imágenes que no corresponden a rostros humanos, como se presentará en secciones posteriores.

En la figura 27, se presenta el caso de un **Falso Negativo**. La fotografía que se desea reconocer, al procesarla, se indica como resultado: Rostro no reconocido. Sin embargo, la imagen corresponde al delincuente Decbello, como se puede verificar en las figuras 33 y 34. El sistema debería de haberlo reconocido, pero no fue así.

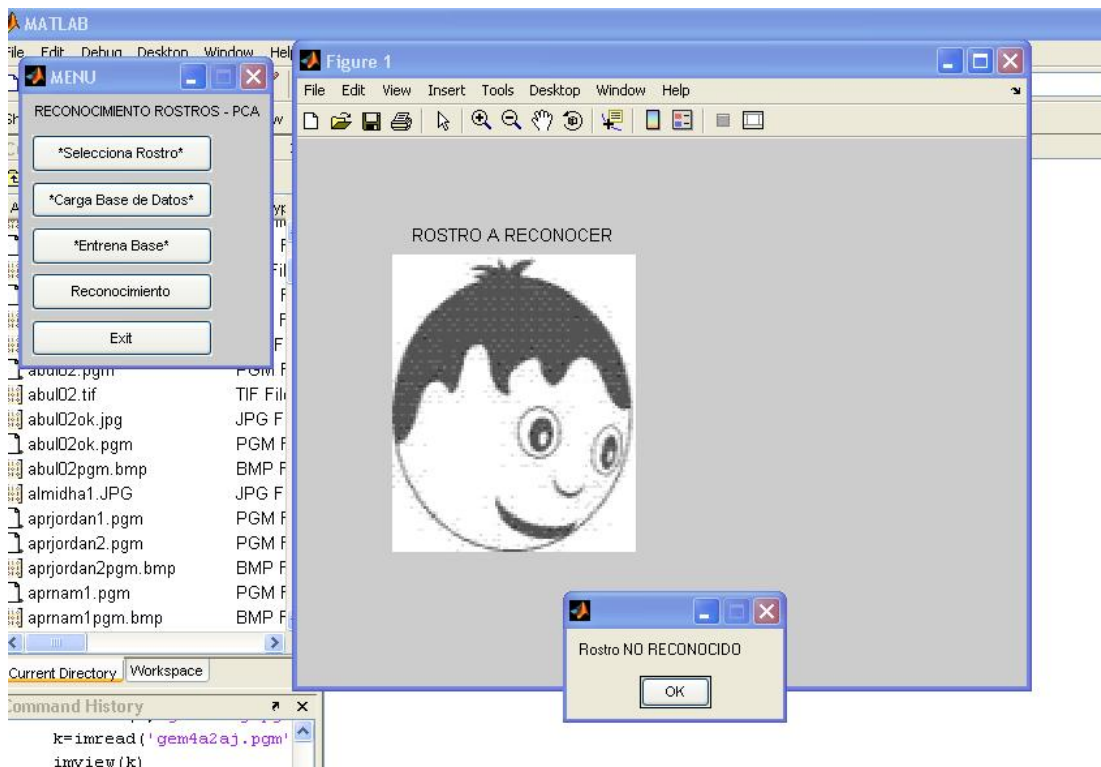


FIGURA 26. SE INTENTA RECONOCER UNA IMAGEN QUE NO CORRESPONDE A UN ROSTRO REAL. EL SISTEMA REPORTA ADECUADAMENTE “ROSTRO NO RECONOCIDO”

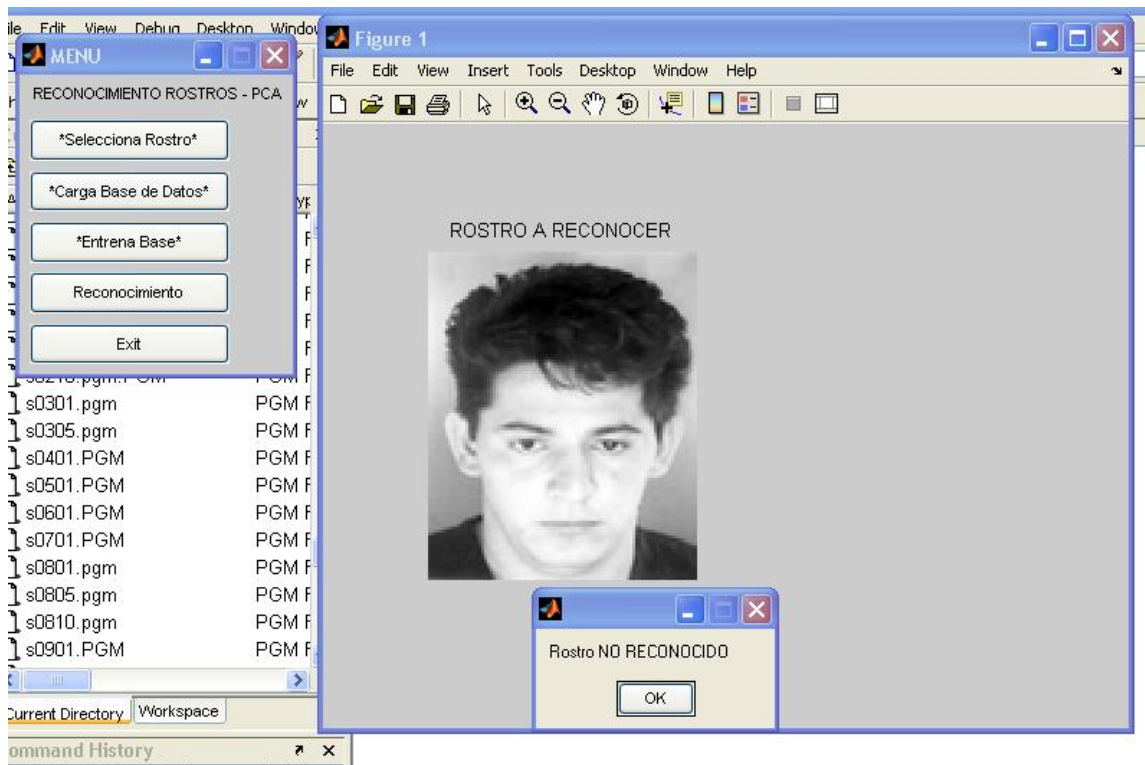


FIGURA 27. EJEMPLO DE UNA IDENTIFICACION DE TIPO “FALSO NEGATIVO”

En la figura 28, se presenta el caso del denominado **Falso Positivo**. Que corresponde a la identificación de una persona inocente, que se le identifica como sospechoso. Por tanto, el sistema identifica a la persona como alguien similar que encontró en la base de entrenamiento. En realidad, es una persona que no corresponde a la fotografía que se encontró.

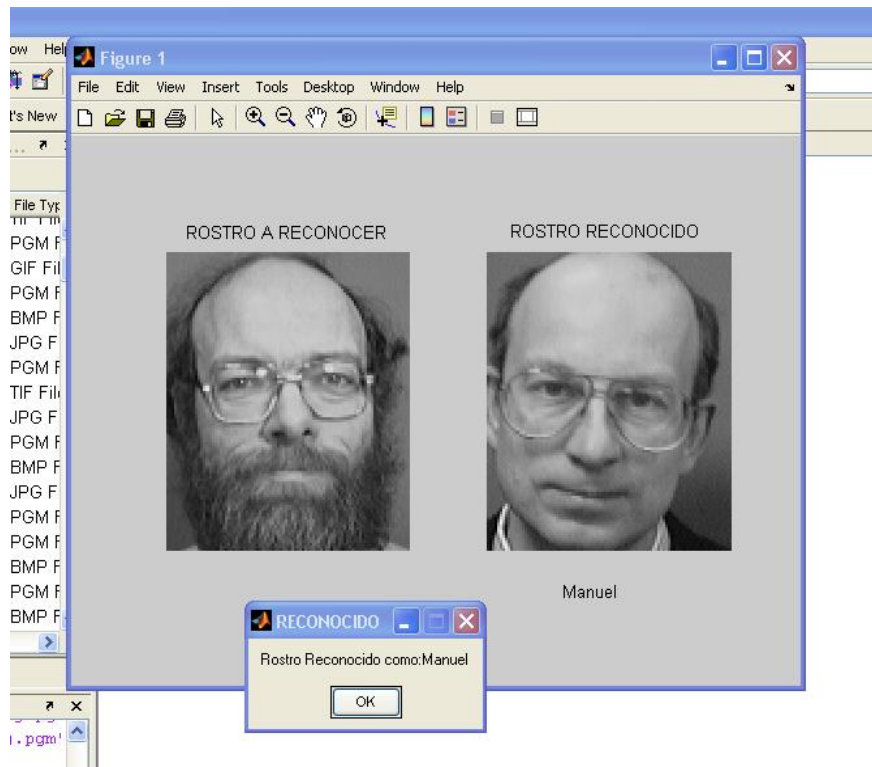


FIGURA 28. EJEMPLO DE UNA IDENTIFICACION DE TIPO “FALSO POSITIVO”

Por último, en la figura 29 se presenta un caso interesante. La fotografía a reconocer, el sistema lo presenta como una identificación positiva, supuestamente correspondiendo al terrorista Abul Faraj Al Libi (figura 30).

Sin embargo, la persona cuya fotografía se ingresa como dato no es dicho terrorista, pero sí se encuentra en la base de entrenamiento reportado como el delincuente Novrios (figura 33 y 34). El sistema no lo identificó como quien debería de ser (correspondiendo a un Falso Negativo) y lo identifica como a otro delincuente (equivalente a un Falso Positivo, pues es inocente de ser un terrorista).

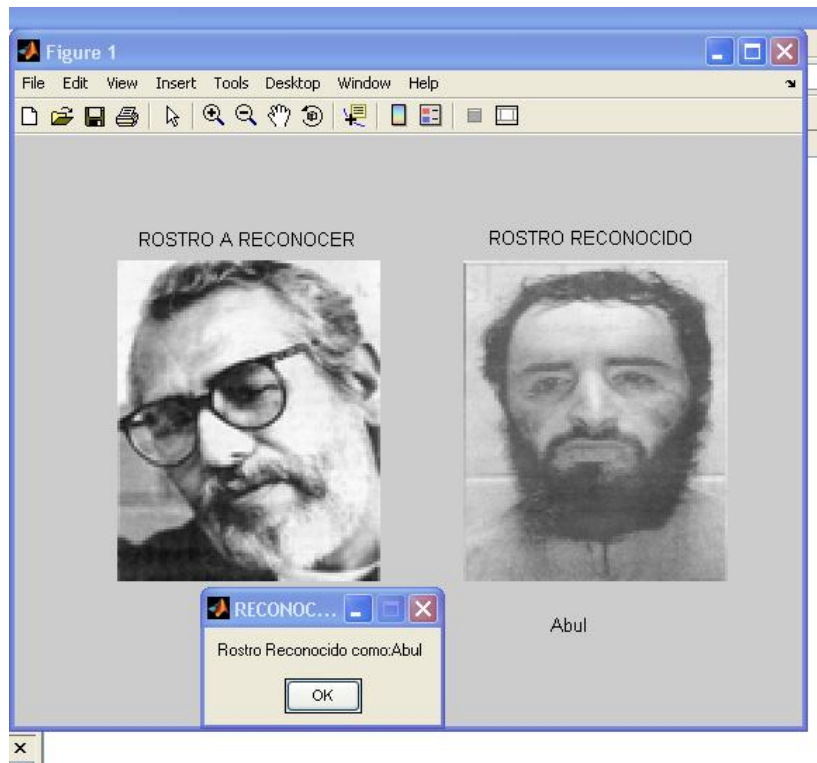


FIGURA 29. EJEMPLO DE UNA IDENTIFICACION COMBINADA DE TIPO “FALSO NEGATIVO” Y “FALSO POSITIVO”

Se han presentado las principales características de operación del sistema de reconocimiento de rostros que se elaboró para propósitos de la presente tesis.

El sistema tiene grandes posibilidades para efectuar reconocimientos adecuados, incluso sorprendentes en algunos casos. Aunque también presenta fallas y errores. Cabe mencionar, que en la actualidad, no se cuenta con ningún sistema comercial de reconocimiento biométrico (y en especial, de reconocimiento de rostros), que tenga por lo menos un 99% de confiabilidad.

En la siguiente sección, se presentará el análisis detallado de los experimentos que se realizaron y los resultados respectivos, para tratar de responder a las preguntas de investigación que se han planteado en el presente trabajo.

5.5. Experimentación

Se realizaron 10 experimentos de reconocimiento de rostros, de acuerdo a las configuraciones experimentales descritas en la sección 5.3.

Los resultados se evalúan de acuerdo a los estándares internacionales para éste tipo de sistemas, en los que se considera la interrelación que se tiene sobre la “verdad de la imagen a reconocer” y la “decisión que tomó el sistema” (tabla 6). Por tanto, los porcentajes de identificación se clasifican en alguna de las cuatro opciones siguientes:

- **Verdadero Negativo** La persona es inocente y el sistema lo reconoce como tal.
- **Verdadero Positivo:** La persona es un “sospechoso” reportado en la base de datos, y el sistema lo logra identificar.
- **Falso Positivo:** La persona es inocente, pero el sistema lo identifica como Sospechoso. Su rostro se clasifica como similar a alguno de los rostros de la base de datos.
- **Falso Negativo:** La persona se encuentra en la base de datos como sospechosa, el sistema no logra identificarlo.

La situación de un determinado rostro en alguno de los cuatro cuadrantes descritos, depende del algoritmo de reconocimiento utilizado y en última instancia, del valor del umbral seleccionado (en la métrica designada) para la clasificación.

Como se recordará (sección 5.4), la clasificación final de un determinado rostro se determina utilizando la distancia a que se encuentra el rostro con respecto al “espacio de rostros”. Si dicha distancia es menor al umbral establecido, se considera “rostro identificado”, si es mayor, será clasificado como no identificado.

En la sección anterior, al describir la operación del sistema, se han presentado algunos ejemplos que muestran los resultados que indicó el programa de reconocimiento contrastados con la realidad; dando origen a una de las cuatro opciones de clasificación mencionadas con anterioridad y que se presentan también en la tabla 6.

TABLA 6. RESULTADOS POSIBLES CONSIDERANDO LA DECISION DEL SISTEMA DE RECONOCIMIENTO EN RELACION A LA VERDAD SOBRE LA FOTOGRAFIA A IDENTIFICAR.

		DECISIÓN DEL SISTEMA DE RECONOCIMIENTO DE ROSTROS SOBRE LA FOTOGRAFIA	
		NO RECONOCIDA	RECONOCIDA
VERDAD SOBRE LA IMAGEN A RECONOCER	NO SE ENCUENTRA EN LA BASE DATOS (INOCENTE)	VERDADERO NEGATIVO (INOCENTE)	FALSO POSITIVO (INOCENTE IDENTIFICADO COMO SOSPECHOSO)
	SE ENCUENTRA EN LA BASE DATOS (SOSPECHOSO)	FALSO NEGATIVO (SOSPECHOSO NO IDENTIFICADO)	VERDADERO POSITIVO (SOSPECHOSO IDENTIFICADO)

El principal problema en los sistemas de reconocimiento de rostros, radica en obtener el menor número, tanto de falsos positivos –FP- (inocentes clasificados como sospechosos), como de falsos negativos –FN- (sospechoso no identificado). Sin embargo, lo anterior no es posible, debido a que FP es función de los FN (lo cual se relaciona directamente con el umbral seleccionado).

Por tanto, si se desea disminuir el porcentaje de falsos negativos (para tratar de identificar al mayor número de sospechosos), se expande el umbral, y por tanto, se incrementará el porcentaje de falsos positivos (inocentes identificados como sospechosos). Por el contrario, si se desea disminuir el número de falsos positivos, se tendrá que disminuir el umbral, provocando un aumento en el número de sospechosos no identificados (falsos negativos).

En principio, se ha propuesto tratar de disminuir en lo más posible, la identificación de inocentes como sospechosos (FP). Sin embargo, los sistemas de reconocimiento existentes, tienen la posibilidad de cambiar el valor de su “umbral” de reconocimiento, de acuerdo a lo deseado por el usuario del sistema (dentro de un cierto intervalo preestablecido).

Por tanto, el valor del umbral se seleccionará de tal forma que permita maximizar el número de verdaderos positivos, minimizando el número de falsos positivos (inocentes identificados como sospechosos) y el de falsos negativos (sospechosos no identificados).

En las siguientes páginas se muestran los resultados de los experimentos planteados en la sección 5.3, utilizando el programa MATLAB para reconocimiento de rostros que se desarrolló para el presente trabajo.

Para cada uno de los experimentos, se consideraron seis umbrales para la fase de identificación, basados en una métrica euclídeana:

- el promedio de los valores del umbral (\bar{X}),
- el valor $\bar{X} + \sigma$,
- el valor $\bar{X} + 2\sigma$,
- el valor $\bar{X} + 3\sigma$,
- el límite superior del intervalo de confianza del 95% (\bar{X})⁵,
- el valor de la mediana y
- el máximo valor que puede presentar el umbral.

El propósito de los diversos umbrales, consiste en tratar de generalizar y obtener (de ser posible), el valor promedio del umbral que sería el más indicado para el reconocimiento adecuado de rostros independientemente del número de sujetos e imágenes por sujeto en la base de entrenamiento. Tratando, como se ha mencionado en secciones anteriores, de reducir en lo más que se pueda el porcentaje de falsos positivos y falsos negativos.

Por otra parte, los experimentos se realizaron considerando los siguientes tres casos principales:

- Una base de entrenamiento de 40 sujetos de la base AT&T,
- Una base de entrenamiento de 33 sujetos de la base AT&T
- Una base de entrenamiento de 22 sujetos de la base AT&T, 10 sujetos de fotografías seleccionadas del Portal de Internet del FBI y 1 fotografía que se obtuvo de un periódico.

⁵ Se obtuvo por medio del paquete de software estadístico SPSS.

Para la base de entrenamiento de 40 sujetos, se consideraron tres opciones secundarias: experimentos en que cada sujeto presenta 9 fotografías, 8 fotografías y 5 fotografías.

Se pretende analizar, si existe diferencia en los resultados que se obtengan, con respecto al número de sujetos en la base (40 contra 33 sujetos).

En el caso de la base de entrenamiento con 33 sujetos, se realizaron seis experimentos secundarios en los que se consideraron diferentes números de fotografías para cada sujeto: 9, 7, 5, 3, 2 y 1.

Por lo que respecta a las fotografías utilizadas como entrada al sistema, se consideraron en primera instancia, la primera fotografía de cada grupo de 10 imágenes de la base de datos AT&T (figura 16 y 17).

El propósito de experimentar con diversas fotografías por sujeto, consiste en probar si el algoritmo de componentes principales, puede funcionar adecuadamente cuando la base de entrenamiento presenta una sola imagen por cada persona. O bien, encontrar cual sería el menor número de fotografías que requiere un sistema basado en componentes principales para funcionar adecuadamente.

Se considera, que en un gran número de aplicaciones comerciales, únicamente se contaría con una fotografía de cada persona (sistemas de acceso a empresas, bases de datos de licencias de tránsito, bases de datos de pasaportes, etc.)

De forma adicional, se realizó otro experimento con 33 sujetos (EXP10). De los cuales, 22 corresponden a fotografías de la base AT&T, 10 sujetos corresponden a delincuentes y

terroristas reportados en el Portal de Internet del FBI⁶ y una fotografía se obtuvo por medio de un proceso de escáner a partir del periódico.

5.5.1. Resultados de la experimentación con fotografía de la base AT&T

En las tablas 7 a 9, se presentan los resultados que se obtuvieron para una base de entrenamiento compuesta por 40 sujetos de la base AT&T, considerando 9, 8 y 5 fotografías por cada sujeto.

En las tablas 11 a 16, se pueden observar los resultados correspondientes a la base de entrenamiento con 33 sujetos (con 9,7,5,3,2 y 1 fotografía por cada persona).

En las tablas de resultados, se presenta para cada columna: el número de sujetos identificados como verdadero positivo (VP), falso negativo (FN), falso positivo (FP), verdadero negativo (VN). Entre “paréntesis”, se muestra el porcentaje de reconocimiento considerando las 53 imágenes a identificar como el 100% (reconocimiento global). Entre “corchetes”, se presenta el porcentaje de VP considerando únicamente el número de sujetos que realmente se encuentran en la base de entrenamiento (reconocimiento local). En el primer grupo de experimentos, se toma como 100% a las 40 imágenes. Se considera que un sistema de reconocimiento, de manera ideal, si se le presentan únicamente fotografías de las personas que se encuentran en la base de entrenamiento, debería identificar al 100%.

⁶ Fotografías recuperadas del Portal del FBI, sección de fugitivos y terroristas más buscados: <http://www.fbi.gov>

Experimentos con 40 sujetos en la base de entrenamiento

En el primer conjunto de experimentos, se utilizaron los 40 sujetos que configuran la base AT&T, utilizando un máximo de 9 fotografías por persona para entrenar la base. Como ya se mencionó, la primera fotografía de cada conjunto, se utilizó como la imagen a reconocer. Dicha imagen, se considera como la fotografía que se le tomaría a una persona en algún aeropuerto o cualquier sitio de acceso restringido, con el propósito de buscar en la base de datos de sospechosos (base de entrenamiento), si la persona es inocente o nó.

En las siguientes páginas (tablas 7 a 9), se presentan los resultados que obtuvo el sistema de reconocimiento, considerando diversos umbrales y número de fotografías por sujeto.

De acuerdo a los estándares de evaluación en los sistemas de reconocimiento de rostros, la eficiencia ideal del sistema, se obtiene cuando el número de “falsos positivos” es igual a cero y se tiene un máximo en el número de “verdaderos positivos”. Por tanto, observando las tablas respectivas: en el caso de 9 fotografías por sujeto (tabla 7), no se logra obtener un FP=0, por lo menos se tiene un FP=1 con un máximo de PV=37 (69.81% de reconocimiento global y un 92.50% con respecto a la identificación local de los 40 sujetos). Lo anterior se obtiene utilizando un umbral de $\bar{X} + 2\sigma$.

TABLA 7. PORCENTAJES Y NUMERO DE ROSTROS IDENTIFICADOS EN CADA RUBRO DE ACUERDO A DIVERSOS INTERVALOS DE CONFIANZA (CON 40 SUJETOS Y 9 FOTOGRAFIAS POR CADA UNO)

INTERVALO UMBRAL	VERDADERO POSITIVO	FALSO NEGATIVO	FALSO POSITIVO	VERDADERO NEGATIVO	VALOR UMBRAL
\bar{X}	25 (47.16%) [62.50%]	15 (28.30%)	1 (1.88%)	12 (22.64%)	1126.65
$\bar{X} + \sigma$	36 (67.92%) [90.00%]	4 (7.54%)	1 (1.88%)	12 (22.64%)	1611.98
$\bar{X} + 2\sigma$	37 (69.81%) [92.50%]	3 (5.66%)	1 (1.88%)	12 (22.64%)	2097.31
$\bar{X} + 3\sigma$	39 (73.58%) [97.50%]	1 (1.88%)	3 (5.66%)	10 (18.86%)	2582.65
95%(\bar{X})	30 (56.60%) [75.00%]	10 (18.86%)	1 (1.88%)	12 (22.64%)	1281.86
MEDIANA	20 (37.73%) [50.00%]	20 (37.73%)	1 (1.88%)	12 (22.64%)	1007.69
MAXIMO	40 (75.47%) [100.00%]	0 (0.00%)	3 (5.66%)	10 (18.86%)	2639.00
σ					485.33

Por lo que respecta al reconocimiento cuando se tienen 8 fotografías por sujeto (tabla 8), el máximo porcentaje de identificación es del 71.69% global (95.00% reconocimiento local), cuando FP=0. Y corresponde también, a un umbral de $\bar{X} + 2\sigma$.

TABLA 8. NUMERO DE ROSTROS IDENTIFICADOS EN CADA RUBRO DE ACUERDO A DIVERSOS INTERVALOS DE CONFIANZA (CON 40 SUJETOS EN B.D. Y 8 FOTOGRAFIAS POR CADA UNO, 53 ROSTROS PARA IDENTIFICAR)

INTERVALO UMBRAL	VERDADERO POSITIVO	FALSO NEGATIVO	FALSO POSITIVO	VERDADERO NEGATIVO	VALOR UMBRAL
\bar{X}	24 (45.28%) [60.00%]	16 (30.18%)	0 (0.00%)	13 (24.52%)	1157.74
$\bar{X} + \sigma$	35 (66.03%) [87.50%]	5 (9.43%)	0 (0.00%)	13 (24.52%)	1648.16
$\bar{X} + 2\sigma$	38 (71.69%) [95.00%]	2 (3.77%)	0 (0.00%)	13 (24.52%)	2138.58
$\bar{X} + 3\sigma$	39 (73.58%) [97.50%]	1 (1.88%)	1 (1.88%)	12 (22.64%)	2629.01
95%(\bar{X})	29 (54.71%) [72.50%]	20 (37.73%)	0 (0.00%)	13 (24.52%)	1314.58
MEDIANA	20 (37.73%) [50.00%]	11 (20.75%)	0 (0.00%)	13 (24.52%)	1056.70
MAXIMO	40 (75.47%) [100.00%]	0 (0.00%)	1 (1.88%)	12 (22.64%)	2671.00
σ					490.42

TABLA 9. NUMERO DE ROSTROS IDENTIFICADOS EN CADA RUBRO DE ACUERDO A DIVERSOS INTERVALOS DE CONFIANZA (CON 40 SUJETOS Y 5 FOTOGRAFIAS POR CADA UNO)

INTERVALO UMBRAL	VERDADERO POSITIVO	FALSO NEGATIVO	FALSO POSITIVO	VERDADERO NEGATIVO	VALOR UMBRAL
\bar{X}	24 (45.28%) [60.00%]	16 (30.18%)	0 (0.00%)	13 (24.52%)	1223.93
$\bar{X} + \sigma$	32 (60.37%) [80.00%]	8 (15.09%)	0 (0.00%)	13 (24.52%)	1760.57
$\bar{X} + 2\sigma$	39 (73.58%) [97.50%]	1 (1.88%)	0 (0.00%)	13 (24.52%)	2297.21
$\bar{X} + 3\sigma$	40 (75.47%) [100.00%]	0 (0.00%)	2 (3.77%)	11 (20.75%)	2833.85
95%(\bar{X})	28 (52.83%) [70.00%]	12 (22.64%)	0 (0.00%)	13 (24.52%)	1395.56
MEDIANA	20 (37.73%) [50.00%]	20 (37.73%)	0 (0.00%)	13 (24.52%)	1087.45
MAXIMO	40 (75.47%) [100.00%]	0 (0.00%)	2 (3.77%)	11 (20.75%)	2624.30
σ					536.64

Por último, cuando se utilizan únicamente 5 fotografías por sujeto en la base de entrenamiento, se obtiene un 73.58% de eficiencia global (reconociendo 39 personas) y un 97.50% de reconocimiento local. Con un 0.00% de FP y un umbral de $\bar{X} + 2\sigma$.

A partir de los resultados anteriores, se puede empezar a vislumbrar que un umbral apropiado, puede ser el que se configura a partir del promedio más dos desviaciones estándar ($\bar{X} + 2\sigma$).

Si bien el ideal consistiría en tener un porcentaje 0 de falsos positivos (inocentes identificados como sospechosos), implicaría que el número de falsos negativos (sospechosos no identificados), puede incrementarse.

Por tanto, en sistemas comerciales de software biométricos, se pretende reducir el porcentaje de FN aunque se incremente un poco el porcentaje de falsos positivos. Considerando lo anterior, un aumento de los positivos verdaderos se lograría al tener un umbral de $\bar{X} + 3\sigma$, tratando de minimizar los falsos positivos y los falsos negativos. En la tabla 10 se presenta un comparativo.

TABLA 10. NUMERO DE ROSTROS Y PORCENTAJES DE IDENTIFICACION CONSIDERANDO LOS UMBRALES ($\bar{X} + 2\sigma$) Y ($\bar{X} + 3\sigma$) CON LA MENOR TASA POSIBLE DE “FALSOS POSITIVOS” / “FALSOS NEGATIVOS” Y EL MAYOR PORCENTAJE POSIBLE DE “VERDADEROS POSITIVOS” PARA LOS EXPERIMENTOS REALIZADOS CON 40 SUJETOS (CON DIVERSO NUMERO DE FOTOGRAFIAS POR SUJETO) Y 53 FOTOGRAFIAS A RECONOCER

NUMERO DE FOTOS	VERDADERO POSITIVO	FALSO NEGATIVO	FALSO POSITIVO	VERDADERO NEGATIVO	VALOR UMBRAL
$\bar{X} + 2\sigma$					
9 IMAGENES	37 (69.81%) [92.50%]	3 (5.66%)	1 (1.88%)	12 (22.64%)	2097.31
8 IMAGENES	38 (71.69%) [95.00%]	2 (3.77%)	0 (0.00%)	13 (24.52%)	2138.58
5 IMAGENES	39 (73.58%) [97.50%]	1 (1.88%)	0 (0.00%)	13 (24.52%)	2297.21
$\bar{X} + 3\sigma$					
9 IMAGENES	39 (73.58%) [97.50%]	1 (1.88%)	3 (5.66%)	10 (18.86%)	2582.65
8 IMÁGENES	39 (73.58%) [97.50%]	1 (1.88%)	1 (1.88%)	12 (22.64%)	2629.01
5 IMAGENES	40 (75.47%) [100.00%]	0 (0.00%)	2 (3.77%)	11 (20.75%)	2833.85

Se observa, que a mayor número de fotografías por sujeto, el porcentaje de identificación de positivos verdaderos es menor o similar a cuando se utilizan menos fotografías por sujeto. Por tanto, utilizar un promedio de 5 fotografías por persona es suficiente. En los experimentos con la base de entrenamiento de 33 personas, se tratará de confirmar lo anterior.

Por lo que respecta al umbral (y considerando la opción de 5 fotografías), parecería que usar ($\bar{X} + 3\sigma$), sería lo indicado. Se obtendría un 75.47% de reconocimiento global, un 100% de reconocimiento local y un 0% de falsos negativos. Sin embargo, se presenta un 3.77% de falsos positivos. Si se tuviera una gran base de datos, un 3.77% es una tasa alta, y por tanto, identificaría a un gran número de personas inocentes, como sospechosos.

Debido a lo anterior, la opción de utilizar un umbral de $(\bar{X} + 2\sigma)$ sería una opción mejor, al reducir en un 3.77% los falsos positivos y los positivos verdaderos tendrán un porcentaje del 73.58% (una disminución del 1.89% con respecto a la opción de 3 desviaciones estándar).

En realidad, la mejor opción debería encontrarse en un valor de umbral situado entre $\bar{X} + 2\sigma$ y $\bar{X} + 3\sigma$. En los experimentos posteriores, se tratará de comprobar los valores más idóneos del umbral y el número de fotografías por sujeto.

Experimentos con 33 sujetos en la base de entrenamiento

En las siguientes páginas, se presentan los resultados al utilizar 33 sujetos para configurar la base de entrenamiento.

Al igual que en los experimentos anteriores, se utilizó la primera fotografía (del conjunto de 10 imágenes), como imagen a identificar por el sistema de reconocimiento.

Asimismo, como ya se ha mencionado en secciones anteriores, 7 fotografías que corresponden a los sujetos de la base AT&T que no se utilizaron para configurar la base de entrenamiento, se utilizaron como parte de las imágenes a reconocer. El sistema debería de identificarlos como personas “No reconocida”.

En los resultados que se obtengan, se tratará de comprobar si el umbral de reconocimiento es similar al de $\bar{X} + 2\sigma$, que se obtuvo en los experimentos anteriores. Asimismo, se verificará el número idóneo de imágenes que se requieren por sujeto para obtener un adecuado reconocimiento.

TABLA 11. NUMERO DE ROSTROS IDENTIFICADOS EN CADA RUBRO DE ACUERDO A DIVERSOS INTERVALOS DE CONFIANZA (CON 33 SUJETOS Y 9 FOTOGRAFIAS POR CADA UNO)

INTERVALO UMBRAL	VERDADERO POSITIVO	FALSO NEGATIVO	FALSO POSITIVO	VERDADERO NEGATIVO	VALOR UMBRAL
\bar{X}	22 (41.50%) [66.66%]	11 (20.75%)	0 (0.00%)	20 (37.73%)	1133.2
$\bar{X} + \sigma$	29 (54.71%) [87.87%]	4 (7.54%)	1 (1.88%)	19 (35.84%)	1651.97
$\bar{X} + 2\sigma$	30 (56.60%) [90.90%]	3 (5.66%)	2 (3.77%)	18 (33.96%)	2170.74
$\bar{X} + 3\sigma$	33 (62.26%) [100.00%]	0 (0.00%)	8 (15.09%)	12 (22.64%)	2689.51
95%(\bar{X})	24 (45.28%) [72.72%]	9 (16.98%)	0 (0.00%)	20 (37.73%)	1317.15
MEDIANA	17 (32.07%) [51.51%]	16 (30.18%)	0 (0.00%)	20 (37.73%)	966.75
MAXIMO	33 (62.26%) [100.00%]	0 (0.00%)	6 (11.32%)	14 (26.41%)	2650.6
σ					518.77

TABLA 12. NUMERO DE ROSTROS IDENTIFICADOS EN CADA RUBRO DE ACUERDO A DIVERSOS INTERVALOS DE CONFIANZA (CON 33 SUJETOS Y 7 FOTOGRAFIAS POR CADA UNO)

INTERVALO UMBRAL	VERDADERO POSITIVO	FALSO NEGATIVO	FALSO POSITIVO	VERDADERO NEGATIVO	VALOR UMBRAL
\bar{X}	21 (39.62%) [63.63%]	12 (22.64%)	0 (0.00%)	20 (37.73%)	1193.67
$\bar{X} + \sigma$	28 (52.83%) [84.84%]	5 (9.43%)	0 (0.00%)	20 (37.73%)	1736.53
$\bar{X} + 2\sigma$	31 (58.49%) [93.93%]	2 (3.77%)	1 (1.88%)	19 (35.84%)	2279.39
$\bar{X} + 3\sigma$	33 (62.26%) [100.00%]	0 (0.00%)	8 (15.09%)	12 (22.64%)	2822.25
95%(\bar{X})	23 (43.39%) [69.69%]	10 (18.86%)	0 (0.00%)	20 (37.73%)	1386.16
MEDIANA	16 (30.18%) [48.48%]	17 (32.07%)	0 (0.00%)	20 (37.73%)	1036.90
MAXIMO	33 (62.26%) [100.00%]	0 (0.00%)	7 (13.20%)	13 (24.52%)	2695.90
σ					542.86

TABLA 13. NUMERO DE ROSTROS IDENTIFICADOS EN CADA RUBRO DE ACUERDO A DIVERSOS INTERVALOS DE CONFIANZA (CON 33 SUJETOS Y 5 FOTOGRAFIAS POR CADA UNO)

INTERVALO UMBRAL	VERDADERO POSITIVO	FALSO NEGATIVO	FALSO POSITIVO	VERDADERO NEGATIVO	VALOR UMBRAL
\bar{X}	19 (35.84%) [57.57%]	14 (26.41%)	0 (0.00%)	20 (37.73%)	1268.56
$\bar{X} + \sigma$	28 (52.83%) [84.84]	5 (9.43%)	0 (0.00%)	20 (37.73%)	1843.14
$\bar{X} + 2\sigma$	31 (58.49%) [93.93%]	2 (3.77%)	3 (5.66%)	17 (32.07%)	2417.72
$\bar{X} + 3\sigma$	33 (62.26%) [100.00%]	0 (0.00%)	9 (16.98%)	11 (20.75%)	2992.30
95%(\bar{X})	23 (43.39%) [69.69%]	10 (18.86%)	0 (0.00%)	20 (37.73%)	1472.30
MEDIANA	17 (32.07%) [51.51%]	16 (30.18%)	0 (0.00%)	20 (37.73%)	1130.4
MAXIMO	33 (62.26%) [100.00%]	0 (0.00%)	6 (11.32%)	14 (26.41%)	2612.80
σ					574.58

TABLA 14. NUMERO DE ROSTROS IDENTIFICADOS EN CADA RUBRO DE ACUERDO A DIVERSOS INTERVALOS DE CONFIANZA (CON 33 SUJETOS Y 3 FOTOGRAFIAS POR CADA UNO)

INTERVALO UMBRAL	VERDADERO POSITIVO	FALSO NEGATIVO	FALSO POSITIVO	VERDADERO NEGATIVO	VALOR UMBRAL
\bar{X}	19 (35.84%) [57.57%]	14 (26.41%)	0 (0.00%)	20 (37.73%)	1424.06
$\bar{X} + \sigma$	26 (49.05%) [78.78%]	7 (13.20%)	0 (0.00%)	20 (37.73%)	2104.55
$\bar{X} + 2\sigma$	32 (62.26%) [96.96%]	1 (1.88%)	6 (11.32%)	14 (26.41%)	2785.04
$\bar{X} + 3\sigma$	33 (62.26%) [100.00%]	0 (0.00%)	12 (22.64%)	8 (15.09%)	3465.53
95%(\bar{X})	23 (43.39%) [69.69%]	10 (18.86%)	0 (0.00%)	20 (37.73%)	1665.35
MEDIANA	18 (33.96%) [54.54%]	15 (28.30%)	0 (0.00%)	20 (37.73%)	1258.70
MAXIMO	33 (62.26%) [100.00%]	0 (0.00%)	6 (11.32%)	14 (26.41%)	2879.6
σ					680.49

TABLA 15. NUMERO DE ROSTROS IDENTIFICADOS EN CADA RUBRO DE ACUERDO A DIVERSOS INTERVALOS DE CONFIANZA (CON 33 SUJETOS Y 2 FOTOGRAFIAS POR CADA UNO)

INTERVALO UMBRAL	VERDADERO POSITIVO	FALSO NEGATIVO	FALSO POSITIVO	VERDADERO NEGATIVO	VALOR UMBRAL
\bar{X}	19 (35.84%) [57.57%]	14 (26.41%)	0 (0.00%)	20 (37.73%)	2098.16
$\bar{X} + \sigma$	26 (49.05%) [78.78%]	7 (13.20%)	2 (3.77%)	18 (33.96%)	2932.07
$\bar{X} + 2\sigma$	33 (62.26%) [100.00%]	0 (0.00%)	13 (24.52%)	7 (13.20%)	3765.98
$\bar{X} + 3\sigma$	33 (62.26%) [100.00%]	0 (0.00%)	15 (28.30%)	5 (9.43%)	4599.89
95%(\bar{X})	21 (39.62%) [63.63%]	12 (22.64%)	0 (0.00%)	20 (37.73%)	2393.86
MEDIANA	17 (32.07%) [51.51%]	16 (30.18%)	0 (0.00%)	20 (37.73%)	2043.50
MAXIMO	33 (62.26%) [100.00%]	0 (0.00%)	11 (20.75%)	9 (16.98%)	3502.90
σ					833.91

TABLA 16. NUMERO DE ROSTROS IDENTIFICADOS EN CADA RUBRO DE ACUERDO A DIVERSOS INTERVALOS DE CONFIANZA (CON 33 SUJETOS Y 1 FOTOGRAFIAS POR CADA UNO)

INTERVALO UMBRAL	VERDADERO POSITIVO	FALSO NEGATIVO	FALSO POSITIVO	VERDADERO NEGATIVO	VALOR UMBRAL
\bar{X}	19 (35.84%) [57.57%]	14 (26.41%)	0 (0.00%)	20 (37.73%)	1528.76
$\bar{X} + \sigma$	26 (49.05%) [78.78%]	7 (13.20%)	1 (1.88%)	19 (35.84%)	2173.79
$\bar{X} + 2\sigma$	32 (60.37%) [96.96%]	1 (1.88%)	6 (11.32%)	14 (26.41%)	2818.82
$\bar{X} + 3\sigma$	33 (62.26%) [100.00%]	0 (0.00%)	11 (20.75%)	9 (16.98%)	3463.85
95%(\bar{X})	23 (43.39%) [69.69%]	10 (18.86%)	0 (0.00%)	20 (37.73%)	1757.49
MEDIANA	17 (32.07%) [51.51%]	16 (30.18%)	0 (0.00%)	20 (37.73%)	1386.40
MAXIMO	33 (62.26%) [100.00%]	0 (0.00%)	6 (11.32%)	14 (26.41%)	2868.30
σ					645.03

Por lo que respecta a los resultados de reconocimiento utilizando una base de entrenamiento de 33 personas y 53 fotografías de personas para el reconocimiento, se pueden obtener las siguientes conclusiones:

a) Estimación del umbral promedio

A partir de la revisión de las 6 tablas anteriores y utilizando el criterio planteado en secciones anteriores, con respecto a minimizar los porcentajes de Falsos Positivos y Falsos Negativos maximizando los porcentajes de Positivos Verdaderos. Se puede observar que dicho criterio presenta las mejores alternativas cuando el umbral se encuentra entre $(\bar{X} + \sigma)$ y $(\bar{X} + 2\sigma)$.

Si bien, el ideal sería tener un porcentaje del 0.0% en el rubro de falsos positivos (lo cual se obtiene en la mitad de los casos, en ellos, el porcentaje de positivos verdaderos no rebasa el 53%. Por tanto, se considera como mejor alternativa, elevar el porcentaje de positivos verdaderos sacrificando el número de falsos positivos que se obtendrán. Lo anterior se logra al considerar un umbral de $(\bar{X} + 2\sigma)$. En la tabla 17 se presentan los porcentajes que se obtienen al considerar el umbral mencionado, en los 6 experimentos realizados.

**TABLA 17. NUMERO DE ROSTROS Y PORCENTAJES DE IDENTIFICACION
CONSIDERANDO UN UMBRAL DE $(\bar{X} + 2\sigma)$ PARA LOS SEIS EXPERIMENTOS REALIZADOS
CON 33 SUJETOS (CON DIVERSO NUMERO DE FOTOGRAFIAS POR SUJETO) Y 53
FOTOGRAFIAS A RECONOCER**

NUMERO IMAGENES	VERDADERO POSITIVO	FALSO NEGATIVO	FALSO POSITIVO	VERDADERO NEGATIVO	VALOR UMBRALES
9 IMAGENES	30 (56.60%) [90.90%]	3 (5.66%)	2 (3.77%)	18 (33.96%)	2170.74
7 IMAGENES	31 (58.49%) [93.93%]	2 (3.77%)	1 (1.88%)	19 (35.84%)	2279.39
5 IMAGENES	31 (58.49%) [93.93%]	2 (3.77%)	3 (5.66%)	17 (32.07%)	2417.72
3 IMAGENES	32 (60.37%) [96.96%]	1 (1.88%)	6 (11.32%)	14 (26.41%)	2785.04
2 IMÁGENES	32 (60.37%) [96.96%]	1 (1.88%)	6 (11.32%)	14 (26.41%)	2785.04
1 IMAGENES	32 (60.37%) [96.96%]	1 (1.88%)	6 (11.32%)	14 (26.41%)	2818.82

De la tabla anterior se obtiene que el umbral promedio tiene un valor de 2542.79 con una desviación estándar de 288.79. El valor del intervalo de $(\bar{X} \pm \sigma)$ se encuentra entre 2254.00 y 2831.58. Por lo que respecta al intervalo al 95% de confianza (de la media) estará entre 2239.72 y 2845.86. Se observa, que el máximo valor de umbral que se requeriría es de 2818.82 (para el caso de tener únicamente una fotografía por sujeto en la base de entrenamiento. Por tanto, el intervalo del 95% (media) se encuentra sobrado para los posibles valores de umbrales que se requieren.

El umbral de $\bar{X} \pm \sigma$, se considera más adecuado, cubriendo los requerimientos de los diversos experimentos. Por otra parte, se puede observar, que conforme el número de

fotografías por cada sujeto disminuye, el umbral necesario para un adecuado reconocimiento se incrementa. En los experimentos, si de 9 fotografías por sujeto se disminuye a 1 fotografía únicamente, se requiere de un aumento aproximado del 29.85% en el valor del umbral para efectuar el reconocimiento de manera adecuada. En resumen, se considera como umbral adecuado, el de $\bar{X} + 2\sigma$.

b) Eficiencia general del sistema

La eficiencia del sistema estará dada por el número de sujetos identificados correctamente y con el menor número de falsos positivos. En la literatura mundial que reportan resultados de experimentos en el área de reconocimiento de rostros (y en general, en aplicaciones Biométricas), se considera como parámetro de la eficiencia del sistema, cuando el porcentaje de falsos positivos es del 0%.

Por tanto, a partir de los resultados de los seis experimentos reportados, se muestra en la tabla 18 los porcentajes máximos de reconocimiento para cada experimento cuando la tasa de falsos positivos es de 0%. En el caso de que se presenten varias opciones de resultado (para FP=0%), se indica la opción con el máximo porcentaje de identificación (Positivo verdadero).

TABLA 18. NUMERO DE ROSTROS Y PORCENTAJES DE IDENTIFICACION CONSIDERANDO UNA TASA DE “FALSOS POSITIVOS” DEL 0.0% Y MÁXIMO PORCENTAJE DE “POSITIVOS VERDADEROS” PARA LOS SEIS EXPERIMENTOS REALIZADOS CON 33 SUJETOS (CON DIVERSO NUMERO DE FOTOGRAFIAS POR SUJETO) Y 53 FOTOGRAFIAS A RECONOCER

NUMERO IMAGENES	VERDADERO POSITIVO	FALSO NEGATIVO	FALSO POSITIVO	VERDADERO NEGATIVO	VALOR UMBRAL
9 IMAGENES	22 (41.50%) [66.66%]	11 (20.75%)	0 (0.00%)	20 (37.73%)	1133.2
7 IMAGENES	28 (52.83%) [84.84%]	5 (9.43%)	0 (0.00%)	20 (37.73%)	1736.53
5 IMAGENES	28 (52.83%) [84.84]	5 (9.43%)	0 (0.00%)	20 (37.73%)	1843.14
3 IMAGENES	26 (49.05%) [78.78%]	7 (13.20%)	0 (0.00%)	20 (37.73%)	2104.55
2 IMÁGENES	21 (39.62%) [63.63%]	12 (22.64%)	0 (0.00%)	20 (37.73%)	2393.86
1 IMAGENES	23 (43.39%) [69.69%]	10 (18.86%)	0 (0.00%)	20 (37.73%)	1757.49

La eficiencia general del sistema es del 84.84% considerando únicamente el reconocimiento controlado a partir de la base de datos AT&T, y del 52.83% considerando reconocimientos generales que incluyen fotografías con diversas características de luz, acercamiento, resolución, etc. Asimismo, la eficiencia anterior se encuentra relacionada con el número de fotografías por sujeto que se utilizó. Se observa, que la eficiencia reportada se cumple cuando se utilizan 7 o 5 imágenes por persona. A continuación se enlista la eficiencia del sistema en relación con el número de fotografías utilizadas en la base de entrenamiento (tabla 19).

TABLA 19. NUMERO DE ROSTROS Y PORCENTAJES DE IDENTIFICACION CONSIDERANDO UNA TASA DE “FALSOS POSITIVOS” DEL 0.0% Y CLASIFICADAS POR MÁXIMO PORCENTAJE DE “POSITIVOS VERDADEROS” PARA LOS SEIS EXPERIMENTOS REALIZADOS CON 33 SUJETOS (CON DIVERSO NUMERO DE FOTOGRAFIAS POR SUJETO) Y 53 FOTOGRAFIAS A RECONOCER

NUMERO IMAGENES	VERDADERO POSITIVO	FALSO NEGATIVO	FALSO POSITIVO	VERDADERO NEGATIVO	VALOR UMBRAL
5 IMAGENES	28 (52.83%) [84.84]	5 (9.43%)	0 (0.00%)	20 (37.73%)	1843.14
7 IMAGENES	28 (52.83%) [84.84%	5 (9.43%)	0 (0.00%)	20 (37.73%)	1736.53
3 IMAGENES	26 (49.05%) [78.78%]	7 (13.20%)	0 (0.00%)	20 (37.73%)	2104.55
1 IMAGENES	23 (43.39%) [69.69%]	10 (18.86%)	0 (0.00%)	20 (37.73%)	1757.49
9 IMÁGENES	22 (41.50%) [66.66%]	11 (20.75%)	0 (0.00%)	20 (37.73%)	1133.2
2 IMAGENES	21 (39.62%) [63.63%]	12 (22.64%)	0 (0.00%)	20 (37.73%)	2393.86

A partir de la tabla anterior, se aprecia que utilizar un número elevado de fotografías por sujeto para configurar la base de entrenamiento (9 en el caso que nos ocupa), no representa mejoras en la eficiencia. De hecho, presenta porcentajes bajos de reconocimiento, muy similares al que se obtendría utilizando únicamente 2 fotografías por sujeto.

Asimismo, utilizando tres fotografías o únicamente una por persona, se obtienen porcentajes de eficiencia general menores al 50% y menores al 79% si se considera la situación controlada de utilizar solo la base AT&T, lo cuál no se considera adecuado.

La eficiencia ideal (considerando un porcentaje del 0% de falsos positivos), tiene un porcentaje del 52.38% de reconocimiento general y del 84.84% con respecto a la base AT&T, presentándose cuando se utilizan por lo menos 5 fotografías por sujeto (y menos de 8) para configurar la base de entrenamiento.

Recuérdese que los resultados anteriores de eficiencia, es el caso ideal. Sin embargo, en aplicaciones prácticas no es posible tener tasas de cero en el rubro de falsos positivos. A continuación se presenta la eficiencia del sistema, considerando la alternativa más práctica.

c) Eficiencia del sistema considerando el mejor umbral de reconocimiento

Para aplicaciones prácticas se tiene que considerar la interrelación de falsos positivos, positivos verdaderos, falsos negativos y el número de fotografías por sujeto en la configuración de la base de entrenamiento.

En secciones anteriores, se llegó a la conclusión de que el mejor umbral para efectuar el reconocimiento se tiene con $\bar{X} + 2\sigma$.

Por tanto, a partir de los resultados que se obtuvieron para dicho umbral se presentan los resultados en la tabla 20.

TABLA 20. NUMERO DE ROSTROS Y PORCENTAJES DE IDENTIFICACION CONSIDERANDO EL MEJOR UMBRAL ($\bar{X} + 2\sigma$) CON LA MENOR TASA POSIBLE DE “FALSOS POSITIVOS” Y EL MAYOR PORCENTAJE POSIBLE DE “POSITIVOS VERDADEROS” PARA LOS SEIS EXPERIMENTOS REALIZADOS CON 33 SUJETOS (CON DIVERSO NUMERO DE FOTOGRAFIAS POR SUJETO) Y 53 FOTOGRAFIAS A RECONOCER

NUMERO IMAGENES	VERDADERO POSITIVO	FALSO NEGATIVO	FALSO POSITIVO	VERDADERO NEGATIVO	VALOR UMBRAL
9 IMAGENES	30 (56.60%) [90.90%]	3 (5.66%)	2 (3.77%)	18 (33.96%)	2170.74
7 IMAGENES	31 (58.49%) [93.93%]	2 (3.77%)	1 (1.88%)	19 (35.84%)	2279.39
5 IMAGENES	31 (58.49%) [93.93%]	2 (3.77%)	3 (5.66%)	17 (32.07%)	2417.72
3 IMAGENES	32 (60.37%) [96.96%]	1 (1.88%)	6 (11.32%)	14 (26.41%)	2785.04
2 IMÁGENES	32 (60.37%) [96.96%]	1 (1.88%)	6 (11.32%)	14 (26.41%)	2785.04
1 IMAGENES	32 (60.37%) [96.96%]	1 (1.88%)	6 (11.32%)	14 (26.41%)	2818.82

Clasificando los porcentajes de eficiencia general de mayor a menor, se obtiene la tabla 21.

TABLA 21. NUMERO DE ROSTROS Y PORCENTAJES DE IDENTIFICACION CONSIDERANDO EL MEJOR UMBRAL ($\bar{X} + 2\sigma$) CLASIFICADOS POR MAYOR PORCENTAJE GENERAL DE RECONOCIMIENTO, MENOR PORCENTAJE DE “FALSOS POSITIVOS” Y MENOR NUMERO DE FOTOGRAFIAS POR SUJETO PARA LOS SEIS EXPERIMENTOS REALIZADOS CON 33 SUJETOS (CON DIVERSO NUMERO DE FOTOGRAFIAS POR SUJETO) Y 53 FOTOGRAFIAS A RECONOCER

NUMERO IMAGENES	VERDADERO POSITIVO	FALSO NEGATIVO	FALSO POSITIVO	VERDADERO NEGATIVO	VALOR UMBRAL
1 IMAGENES	32 (60.37%) [96.96%]	1 (1.88%)	6 (11.32%)	14 (26.41%)	2818.82
2 IMAGENES	32 (60.37%) [96.96%]	1 (1.88%)	6 (11.32%)	14 (26.41%)	2785.04
3 IMAGENES	32 (60.37%) [96.96%]	1 (1.88%)	6 (11.32%)	14 (26.41%)	2785.04
7 IMAGENES	31 (58.49%) [93.93%]	2 (3.77%)	1 (1.88%)	19 (35.84%)	2279.39
5 IMÁGENES	31 (58.49%) [93.93%]	2 (3.77%)	3 (5.66%)	17 (32.07%)	2417.72
9 IMAGENES	30 (56.60%) [90.90%]	3 (5.66%)	2 (3.77%)	18 (33.96%)	2170.74

En principio, la eficiencia general más alta del 60.37% (con una eficiencia particular de la base AT&T del 96.96%), se presenta al utilizar indistintamente 1, 2 o 3 fotografías por sujeto para la base de entrenamiento. Sin embargo, la tasa de falsos positivos en los mismos tres casos es alta con respecto a los experimentos (11.32%).

A pesar de utilizar 9 fotografías por sujeto para el entrenamiento, el porcentaje de eficiencia general es el más bajo (56.60%), aunque la tasa de falsos positivos es menor (3.77%) con respecto a las opciones de 1,2 y 3 imágenes; pero la tasa de falsos negativos se incrementa al triple (5.66%). Por tanto, al igual que en el análisis de la eficiencia ideal, la opción de utilizar 9 fotografías es mala y no contribuye a la mejora del proceso de reconocimiento. En general, se puede empezar a vislumbrar que el uso de gran cantidad de fotografías para entrenar la base, no es significativa.

Por lo que respecta a la opción de utilizar 5 o 7 fotografías para entrenamiento, presentan ambas una eficiencia general del 58.49% (93.93% con respecto a la identificación particular de la base AT&T). Si bien, el porcentaje de eficiencia de reconocimiento es menor en un 1.88% con respecto a la mejor opción, la ventaja que se observa es la disminución del porcentaje de falsos positivos: 1.88% y 5.66% en comparación al 11.32% que presentan las opciones de mejor desempeño.

En general, se considera que la opción de tener 5 o 7 fotografías, presentaría una mejor opción de reconocimiento. La utilización de 7 fotografías daría un mejor desempeño, ya que se disminuiría notablemente el número de falsos positivos (casi 6 veces menos) con un porcentaje de reconocimiento casi similar a la opción de 1 a 3 fotografías.

Si bien, el uso de únicamente una fotografía por sujeto en la base de entrenamiento indica aparentemente el mejor desempeño general en la experimentación (60.37%), se presenta un elevado porcentaje de falsos positivos (11.32%).

Por lo anterior, en combinación a los resultados que se obtuvieron en el caso ideal de reconocimiento, indican que el uso de pocas fotografías (en este caso de 1 a 3 imágenes por persona) no presentan un rendimiento adecuado en la identificación, por tener un elevado porcentaje de falsos positivos.

El uso de 5 a 7 fotografías presentan un mejor rendimiento comercial en el uso de sistemas de reconocimiento de rostros basados en análisis de componentes principales.

5.5.2. Experimentación con una base de entrenamiento con fotografías combinadas de la base AT&T y del FBI.

El último experimento se presenta en la sección siguiente y combina en la base de entrenamiento, fotografías de la base AT&T y de fugitivos/terroristas reportados en el portal del FBI.

En el presente experimento, como ya se ha mencionado, se combinan 22 fotografías de la base AT&T (parte de las mismas utilizadas en experimentos anteriores), 10 fotografías que se obtuvieron del portal internet del FBI y una fotografía que salió publicada en un periódico (la cual fue capturada por medio de un escáner).

Se considera que en una aplicación real, sería poco probable contar con una base de entrenamiento que tuviera todas sus fotografías controladas (a excepción de bases de fotografías como las que se obtienen para licencias de conducir, pasaportes, visas). Sería más factible, que la base de entrenamiento presente una combinación de fotografías controladas y otras fotografías (no controladas), que se han obtenido por diversos medios.

Por tanto, el objetivo del experimento consiste en obtener el comportamiento del sistema de reconocimiento de rostros mediante componentes principales, al tener que utilizar una base

de entrenamiento que presenta diversos tipos de resolución en sus fotografías, fondos diferentes, etc. Por lo que respecta a las imágenes a reconocer, también presentarán diversas características que no se encuentran controladas (a diferencia de la base AT&T).

El primer rostro (entrenamiento y a reconocer), corresponde al terrorista Abul Faraj Al Libi, jefe de la organización terrorista islámica Al Qaeda en Pakistán y brazo derecho de Osama Bin Laden (figura 30). Las dos fotografías se obtuvieron por medio de un escaner, a partir del artículo que con motivo de su detención se publicó en un diario nacional.⁷

Detienen al “número tres” de Al Qaeda; Osama sigue libre

LA RED SE BURLA DE EU

► Suponen que el arresto de Abul Faraj Al Libi ayudará a atrapar a Bin Laden, de quien era “brazo derecho” ► Washington ofrecía 5 millones de dólares por su captura

INFORME
Asegura OTAN haber desmantelado 85 por ciento de armas nucleares

[EFE EN BRUSELAS]
■ La OTAN desmanteló el 85 por ciento del arsenal nuclear disponible en Europa para las fuerzas aliadas los años 90, aseguraron fuentes de la Alianza a propósito de la conferencia de revisión del Tratado de No Proliferación de armas atómicas que Naciones Unidas celebra en Nueva York. El armamento restante en el continente europeo está bajo custodia permanente de Estados Unidos, que aplica un protocolo de activación dirigido a impedir su uso irregular por actores estatales u organizaciones terroristas. La decisión de desmantelar la mayor parte del arsenal nuclear fue adoptada a iniciativa del entonces presidente de EU, George Bush, y en el contexto de desarme que siguió al fin de la Guerra Fría. Desde 1970, la OTAN contaba con un arsenal atómico estadounidense desplegado en Europa con objeto de disuadir al bloque soviético de una intervención armada y que incluía bombas, misiles, torpedos y minas nucleares.

[AGENCIAS EN ISLAMABAD]
La detención en Pakistán de Abul Faraj Al Libi, jefe de la organización terrorista islámica Al Qaeda en el país y brazo derecho de Osama Bin Laden, podría suponer un paso decisivo para la captura de éste. Al Libi, de nacionalidad libia, considerado el “número tres” en la jerarquía de Al Qaeda y uno de los principales lugartenientes de Bin Laden, fue arrestado junto con otros cinco supuestos terroristas por las autoridades paquistaníes hace unos días muy cerca de Afganistán. Se espera que la información obtenida de los supuestos miembros de Al Qaeda facilite pistas a las autoridades de Islamabad sobre el paradero de Osama Bin Laden, al que Estados Unidos considera culpable de los atentados contra Washington y Nueva York en septiembre de 2001 y por el que ofrece una recompensa de 25 millones de dólares. Osama es el terrorista más buscado por EU y otros países, entre los que está su país de origen, Arabia Saudita.



CAMBIOS. El antes (izquierda) y después del supuesto terrorista Abul Faraj Al Libi.

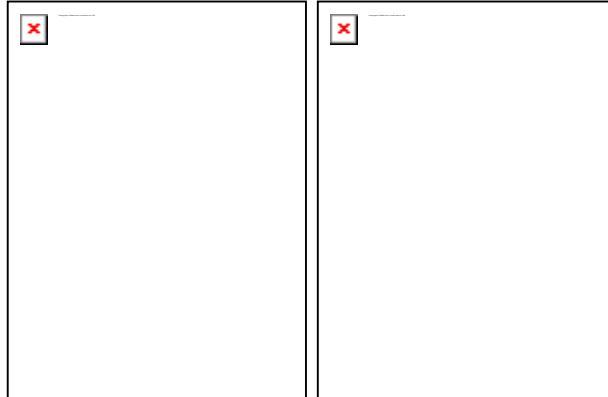
FIGURA 30. REPORTE DE PERIODICO A PARTIR DEL CUAL SE OBTUVIERON DOS FOTOGRAFÍAS DIGITALIZADAS PARA EXPERIMENTACIÓN (FUENTE: Periódico La Crónica, jueves 5 de mayo de 2005, p. 20, México, DF)

⁷ Agencias en Islamabad, “Detienen al “número tres” de Al Qaeda; Osama sigue libre”, La Crónica, jueves 5 de mayo de 2005, p. 20, México, D.F.



UNLAWFUL FLIGHT TO AVOID PROSECUTION - MURDER

WILLIAM JUNIOR JORDAN



Photograph taken in 1974

Age Enhanced Photograph 1996

Aliases: Buddy Jordan, Bill Jordan, William Jordan, William B. Jordan, William Bell Jordan, William Bill Jordan

DESCRIPTION

Date of Birth Used:	September 18, 1942	Hair:	Brown
Place of Birth:	Wadesboro, North Carolina	Eyes:	Blue
Height:	6'2"	Sex:	Male
Weight:	135 to 150 pounds	Race:	White
NCIC:	W479560391	Nationality:	American

Occupations: Heavy Equipment Operator, Brick Mason

Scars and Marks: Jordan has the following tattoos: a skeleton on his right forearm, a spider on his right upper arm, a cross with the word "Sybil" on his left forearm, and the word "Louise" on his left leg. He also has numerous scars and some discolorations on his forehead.

Remarks: Jordan has a visual defect which is corrected with glasses.

CAUTION

William Junior Jordan is wanted for the brutal murder of a man in Georgia. On March 6, 1974, James Rouse, Jr., was taken hostage by Jordan and an accomplice and was forced to drive to a nearby lake. There, after walking into the woods, he was shot in the head with a shotgun. Jordan and the accomplice were arrested and, in June of 1974, were convicted and sentenced to life in prison on charges of murder and armed robbery. On August 6, 1984, Jordan escaped from the Wayne Correctional Institution in Odum, Georgia, and has not been seen since that time.

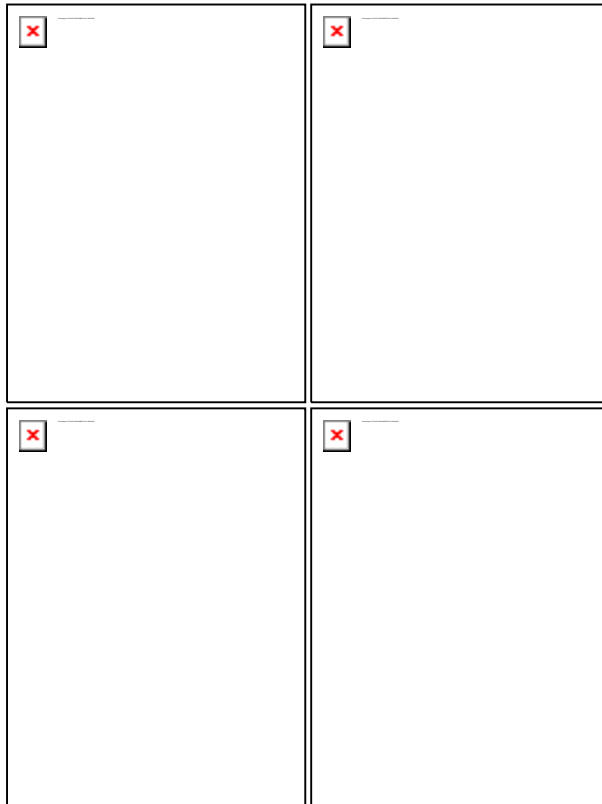
SHOULD BE CONSIDERED ARMED AND DANGEROUS

IF YOU HAVE ANY INFORMATION CONCERNING THIS PERSON, PLEASE CONTACT YOUR [LOCAL FBI OFFICE](#) OR THE NEAREST [AMERICAN EMBASSY OR CONSULATE](#).



FIGURA 31. EJEMPLO DE UN REPORTE DE DELINCIENTES PUBLICADO EN INTERNET POR EL FBI. (FUENTE: <http://www.fbi.org>)

FAHID MOHAMMED ALLY MSALAM



Aliases: Fahid Mohammed Ally, Fahid Mohammed Ali Musalaam, Fahid Mohammed Ali Msalam, Fahid Muhamad Ali Salem, Mohammed Ally Msalam, Usama Al-Kini, Fahad Ally Msalam

DESCRIPTION

Date of Birth Used:	February 19, 1976	Hair:	Black, curly
Place of Birth:	Mombasa, Kenya	Eyes:	Brown
Height:	5'6" to 5'8"	Sex:	Male
Weight:	160 to 170 pounds	Complexion:	Dark
Build:	Unknown	Citizenship:	Kenyan
Languages:	Arabic, Swahili, English		
Scars and Marks:	None known		
Remarks:	Msalam sometimes wears a light beard or moustache and has, in the past, worked as a clothing vendor.		

CAUTION

Fahid Msalam was indicted on December 16, 1998, in the Southern District of New York, for his alleged involvement in the August 7, 1998, bombings of the United States Embassies in Dar es Salaam, Tanzania, and Nairobi, Kenya, and for conspiring to kill United States Nationals.

REWARD
SHOULD BE CONSIDERED ARMED AND DANGEROUS

I

FIGURA 32. EJEMPLO DE UN REPORTE DE TERRORISTAS PUBLICADO EN INTERNET POR EL FBI. (FUENTE: Most Wanted Terrorists en: <http://www.fbi.org>)

Se realizó un escalamiento de las dos imágenes a 112x92 píxeles y se convirtieron a formato PGM. Las otras fotografías de terroristas y delincuentes, como ya se mencionó, se obtuvieron de la página Web del FBI, realizando el escalamiento de píxeles y la conversión al formato que se ha manejado.

Se utilizaron reportes del FBI que tuvieran al menos dos fotografías por sujeto. En las figuras 31 y 32 se presentan a manera de ejemplo, dos de los reportes del FBI utilizados como parte de la base de entrenamiento y de fotografías a identificar.

De las fotografías publicadas por el FBI, una de ellas se anexó a la base de entrenamiento, la otra(s) imágenes, se utilizaron como fotografías a reconocer. En la figura 33, se muestran las imágenes utilizadas como sospechosos de la base de entrenamiento. En la figura 34, se presentan las fotografías de fugitivos del FBI, utilizadas como rostros a reconocer

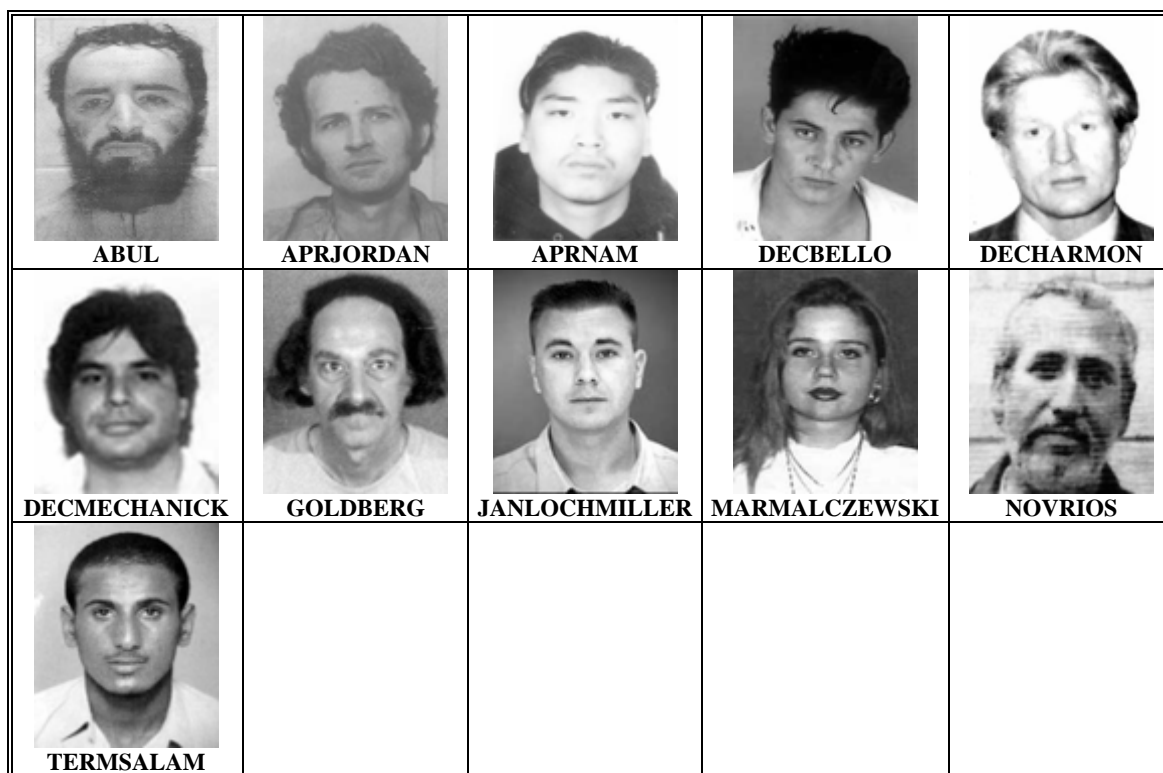


FIGURA 33. ROSTROS QUE SE OBTUVIERON DEL PORTAL DEL FBI Y SE CONSIDERAN COMO PARTE DE LA BASE DE ENTRENAMIENTO

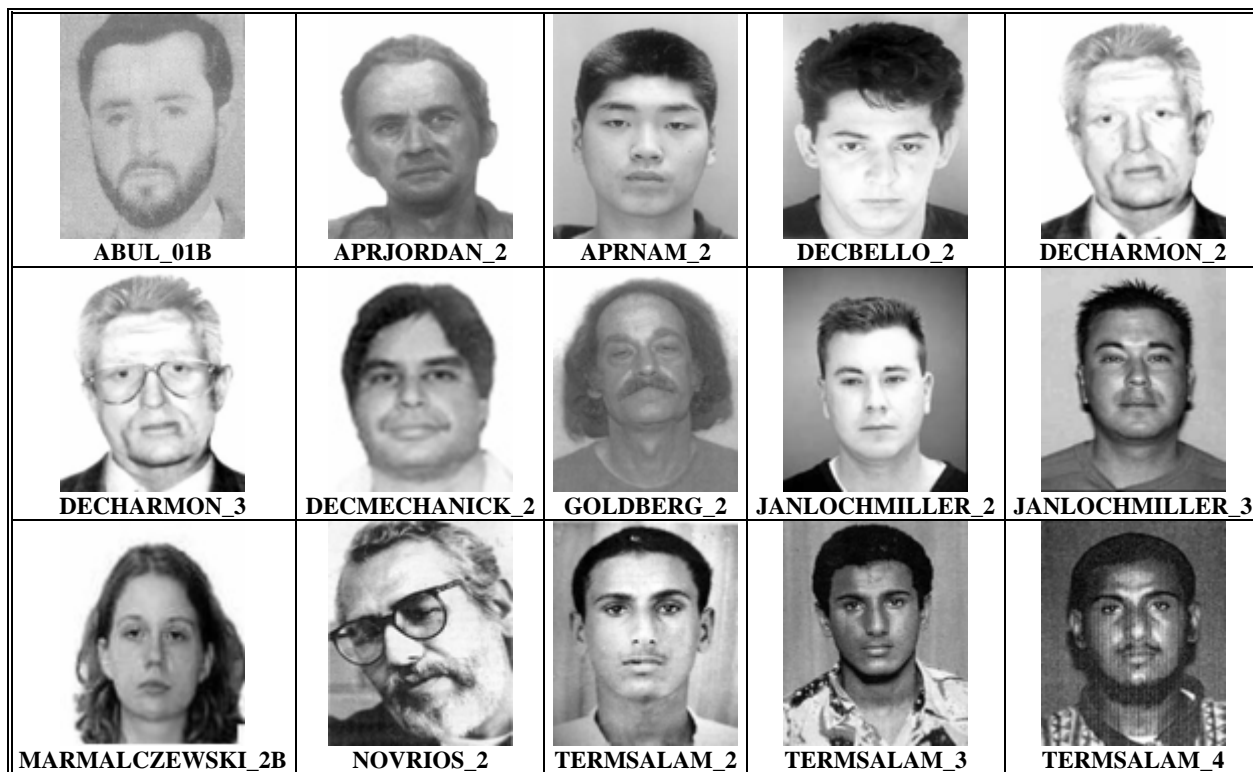


FIGURA 34. ROSTROS DE DELINCUENTES Y TERRORISTAS UTILIZADOS COMO ENTRADA AL SISTEMA (ROSTRO A RECONOCER)

En la tabla 22 se observan los porcentajes de reconocimiento que se obtuvieron en el presente experimento, utilizando diversos umbrales de reconocimiento.

TABLA 22. RESULTADOS DE RECONOCIMIENTO UTILIZANDO LA BASE COMBINADA DEL FBI CON AT&T

INTERVALO UMBRAL	VERDADERO POSITIVO	FALSO NEGATIVO	FALSO POSITIVO	VERDADERO NEGATIVO	VALOR UMBRAL
\bar{X}	17 (32.07%) [51.51%]	16 (30.18%)	5 (9.43%)	15 (28.30%)	2818.25
$\bar{X} + \sigma$	22 (41.50%) [66.66%]	11 (20.75%)	18 (33.96%)	2 (3.77%)	4305.77
$\bar{X} + 2\sigma$	24 (45.28%) [72.72%]	9 (16.98%)	19 (35.84%)	1 (1.88%)	5793.29
$\bar{X} + 3\sigma$	24 (45.28%) [72.72%]	9 (16.98%)	19 (35.84%)	1 (1.88%)	7280.80
95%(\bar{X})	20 (37.73%) [60.60%]	13 (24.52%)	12 (22.64%)	8 (15.09%)	3345.71
MEDIANA	13 (24.52%) [39.39%]	20 (37.73%)	4 (7.54%)	16 (30.18%)	2724.60
MAXIMO	24 (45.28%) [72.72%]	9 (16.98%)	19 (35.84%)	1 (1.88%)	6160.60
σ					1487.52

De la tabla anterior, se puede observar que la eficiencia ideal (cuando el porcentaje de falsos positivos es 0%), no se presenta. El sistema, utilizando la base de entrenamiento mezclada, presenta un porcentaje mínimo de falsos positivos del 7.54%. Sin embargo, con dicho porcentaje, se tiene una eficiencia general en la obtención de positivos verdaderos de solo 45.28%. Incluso a nivel particular, contemplando únicamente las fotografías de sujetos que deberían de encontrarse en la base de entrenamiento, se logra un 72.72% de reconocimiento.

En el siguiente nivel, con respecto a la menor tasa de falsos positivos, se obtiene 9.43% (cuando se utiliza como umbral la media), con un 32.07% de reconocimiento general y un 51.51% con respecto al reconocimiento particular de sujetos en la base AT&T/FBI. En ambos casos, la eficiencia es baja.

Las mejores tasas de reconocimiento comercial, se obtienen al utilizar un umbral de $\bar{X} + 2\sigma$, $\bar{X} + 3\sigma$ y al considerar el valor máximo. En los tres casos, el porcentaje de falsos positivos es del 35.84%, con un 45.28% de reconocimiento global y un 72.72% con respecto a la base AT&T/FBI en particular.

De los datos anteriores, podemos visualizar y comprobar (como ya se había planteado en secciones anteriores), que el mejor umbral para obtener los mejores porcentajes de eficiencias, corresponde a $\bar{X} + 2\sigma$.

Considerando únicamente las 11 fotografías que se obtuvieron del FBI, los porcentajes de reconocimiento (verdadero positivo) con respecto a los diversos umbrales se muestran en la tabla 23.

TABLA 23. PORCENTAJES DE RECONOCIMIENTO VERDADERO PARA LAS FOTOGRAFÍAS DE SOSPECHOSOS DEL FBI (con respecto a 11 fotografías)

INTERVALO UMBRAL	POSITIVO VERDADERO	VALOR UMBRAL
\bar{X}	0 (0.00%)	2818.25
$\bar{X} + \sigma$	2 (18.18%)	4305.77
$\bar{X} + 2\sigma$	4 (36.36%)	5793.29
$\bar{X} + 3\sigma$	4 (36.36%)	7280.80
95% (\bar{X})	0 (0.00%)	3345.71
MEDIANA	0 (0.00%)	2724.60
MAXIMO	4 (36.36%)	6160.60
σ		1487.52

Se aprecia, que el sistema únicamente logra reconocer un máximo del 36.36% de la muestra de sospechosos del FBI. El reconocimiento se logra al utilizar los valores de umbral de $\bar{X} + 2\sigma$, $\bar{X} + 3\sigma$ y con el máximo valor encontrado. Lo anterior, sigue reforzando nuestra hipótesis de utilizar como valor de umbral general: $\bar{X} + 2\sigma$

Por tanto, con el propósito de realizar una comparación entre el reconocimiento que se obtuvo en el experimento 9 (33 sujetos de la base AT&T, con una fotografía por sujeto y 53 imágenes a reconocer), con respecto al experimento anterior se presenta los resultados en la tabla 24.

TABLA 24. COMPARACIÓN DE RESULTADOS CONSIDERANDO UN UMBRAL DE $\bar{X} + 2\sigma$ ENTRE EL EXPERIMENTO “EXP09” Y EL EXPERIMENTO CON LA BASE AT&T/FBI

BASE UTILIZADA	VERDADERO POSITIVO	FALSO NEGATIVO	FALSO POSITIVO	VERDADERO NEGATIVO	VALOR UMBRAL
BASE AT&T	32 (60.37%) [96.96%]	1 (1.88%)	6 (11.32%)	14 (26.41%)	2818.82
BASE AT&T/FBI	24 (45.28%) [72.72%]	9 (16.98%)	19 (35.84%)	1 (1.88%)	5793.29

A partir de la tabla anterior, se aprecia:

- Con respecto a falsos positivos, utilizando la base AT&T el porcentaje es mucho menor (11.32%), con respecto al que obtiene el sistema si se utiliza la base mezclada (16.98%). La diferencia es alta, un 24.52%.
- En lo referente a la eficiencia global (que sería la real, en una aplicación comercial), utilizando fotografías controladas (base At&T) se obtiene un 60.37%. Si se utiliza la base mezclada, únicamente se logra un 45.28% de reconocimiento, una diferencia del 15.09%.
- Por lo que respecta al porcentaje de sospechosos reales que no se detectan, el porcentaje con la base mezclada es también alto (16.98%) en contra del 1.88% con la base original AT&T.

Sin embargo, analizando individualmente el reconocimiento de los delincuentes y terroristas que se incluyeron en la base de entrenamiento, cabe destacar:

La mayoría de las imágenes no fueron “controladas” al momento de tomarlas. Por tanto, como se puede observar en las figura 33 y 34, los fondos de las imágenes son de diverso tipo, la luz tampoco está controlada, la variación en la inclinación de algunos rostros varía, etc.

Asimismo, algunas de la fotografías fueron tomadas bastante tiempo después, incluso años.

A pesar de lo anterior, el sistema logra identificar los siguientes casos (considerados muy complejos):

- Al delincuente Jordan, la fotografía que se utilizó como entrada, corresponde a una imagen de 22 años más con respecto a la fotografía con que contaba la base de entrenamiento (figura 31).
- Al terrorista Termsalam, se logra una identificación adecuada (figura 32), a pesar de utilizar como dato de entrada una imagen más oscura y diferir en varios años, una imagen con respecto a la otra.
- Al delincuente Dercharmon, las dos fotografías utilizadas para identificación (figura 34), logran un reconocimiento adecuado con respecto a la imagen de la base de entrenamiento (figura 33). En este caso, también se observa, que las fotografías de identificación corresponden a un sujeto con mayor edad (con respecto a la de la base de datos). Incluso, en una de ellas, presenta anteojos y en ambas, el corte de cabello es diferente y se observa entrecano.

Por tanto, el funcionamiento de sistemas de reconocimiento de rostros que se basan en técnicas de componentes principales (o métodos similares); como es el caso del sistema desarrollado para la presente tesis, muestran una operación adecuada. De hecho, como se ha mencionado, los dos sistemas comerciales más eficientes que existen en la actualidad se basan en dichas técnicas.

Sin embargo, como también se ha demostrado en los experimentos anteriores, la operación de tales sistemas presenta diversas deficiencias que inciden en la disminución de su confiabilidad.

Como se ha postulado, se considera que parte de tales deficiencias en el logro adecuado de reconocimiento; se debe a la entrada de datos (rostros en base de entrenamiento y rostro a reconocer), los cuales no cumplen con el presupuesto de ajustarse a una distribución normal multivariada (elemento importante en el uso de la técnica de componentes principales).

Con el propósito de contribuir a sistemas de reconocimiento de rostros más eficientes, el autor de la presente tesis, propone un método alternativo que se basa en Algoritmos Genéticos y Aproximación Multivariada en un Espacio L_∞ . Al trabajar en dicho espacio, no se requieren los presupuestos estadísticos y matemáticos que subyacen a las que podemos denominar “técnicas tradicionales” que trabajan en un Espacio Euclidiano (L_2).

En el siguiente capítulo se describe de manera general, el método propuesto y las principales herramientas que se requieren para su implantación.

Capítulo 6

Propuesta de un método para reconocimiento holístico de rostros utilizando Aproximación Multivariada y Algoritmos Genéticos Eclécticos (AGE) en un Espacio L_∞

El contenido de este capítulo es resultado en su mayor parte, de la investigación que ha realizado el autor del presente trabajo desde hace algunos años y se ha reportado de manera parcial en las siguientes ponencias: “Reconocimiento holístico de rostros a través de análisis multivariado y algoritmos genéticos: resultados preliminares”¹ y “A Universal Eclectic Genetic Algorithm for Constrained Optimization”²

Es indudable, que el Análisis de Componentes Principales es una técnica de gran utilidad para modelar sistemas biométricos, en especial, para reconocimiento de rostros. Los resultados que se obtuvieron en los experimentos realizados en el capítulo anterior lo muestran. Asimismo, como se ha mencionado, los dos sistemas comerciales de

¹ Villegas, Carlos y Kuri, Angel. (2000). “Reconocimiento holístico de rostros a través de análisis multivariado y algoritmos genéticos: Resultados preliminares”, Congreso Internacional de Computación-CIC/2000, Instituto Politécnico Nacional, México, noviembre 15-17 de 2000..

² Kuri, Angel y Villegas, Carlos (1998) “A Universal Eclectic Genetic Algorithm for Constrained Optimization”, 6th European Congress on Intelligent techniques & Soft Computing; Aachen, Germany; September 1998.

reconocimiento de rostros más utilizados en la época actual basan sus algoritmos en la técnica de componentes principales.

Sin embargo, los dos sistemas comerciales de reconocimiento han mostrado también bajos rendimientos cuando las fotografías utilizadas presentan diversas características que las alejan de un determinado control al momento de tomar la fotografía del sujeto a reconocer o de la imagen almacenada en la base de entrenamiento.

Algunas de dichas deficiencias, se han podido observar en los resultados de la experimentación que se llevó a cabo en el presente trabajo. Se considera que entre los factores que influyen en la obtención de bajos rendimientos en el reconocimiento de rostros, se encuentra la utilización del espacio de trabajo en L_2 cuando los datos de entrada (píxeles de la configuración de rostros) no cumplen con los supuestos de normalidad que se requiere.

Por tanto, con el propósito de tratar de contribuir a la investigación en el área de reconocimiento de rostros, el autor de la presente tesis propone un método alternativo. Dicho método, trabaja en un espacio L_∞ , por lo cual no se requiere el cumplimiento de los supuestos estadísticos de las técnicas tradicionales, como sería el caso del análisis de componentes principales. Se plantea la utilización del paradigma de polinomios de aproximación y el uso de los métodos de Algoritmos Genéticos como herramienta para obtener el aproximante que caracterice al rostro en un espacio L_∞ . En las siguientes secciones se plantean las características generales del método propuesto.

6.1. Reconocimiento de rostros como un problema de aproximación multivariada.

El reconocimiento de un rostro a partir de la digitalización de fotografías frontales, se puede formular como un problema de *clasificación de patrones*. A su vez, la clasificación de patrones se puede resolver mediante el uso de una **familia de funciones de aproximación multivariada** cuyo propósito será caracterizar un rostro a partir de una serie de atributos de los píxeles de cada imagen (coordenadas, nivel de gris o color, gradiente, etc.) que se plantearán como variables independientes y una de ellas, como variable dependiente. Este conjunto tabular de variables se utilizará como un “conjunto de entrenamiento”, para qué, bajo un enfoque de aprendizaje supervisado se utilice un polinomio de aproximación que caracterice al rostro y se compare el rostro a reconocer con la base de datos de entrenamiento. Una función de aproximación multivariada es una forma de producir un clasificador a partir de un conjunto de datos de entrenamiento y proporcionar el mínimo error de ajuste con respecto al espacio muestral.

En dicho método se pretende realizar el reconocimiento de rostros en forma holística, sin recurrir a la utilización de características geométricas o plantillas (como varios de los algoritmos propuestos para reconocimiento de rostros) o la utilización del análisis de componentes principales. Por tanto, el método propone la utilización de Algoritmos Genéticos y Aproximación Multivariada en un espacio L_{∞} .

Trabajos anteriores de reconocimiento de rostros, generalmente, han basado sus investigaciones en algoritmos que reconocen características aisladas del rostro o mediante la utilización de plantillas que abarcan diversas regiones de la cara. O bien, a partir del total

de píxeles de la fotografía del rostro, reducen el número de variables utilizando el análisis de componentes principales (tema tratado a lo largo de la presente tesis).

En el método que se propone, se piensa usar un enfoque *holístico* alternativo para el reconocimiento de rostros. Se considera, reducirá algunas de las limitaciones de los métodos ya planteados anteriormente.

Así, la caracterización del rostro se realizará a partir de un muestreo aleatorio de diversos atributos seleccionados para los píxeles constituyentes de la imagen (coordenada X, coordenada Y, nivel de gris o colores, gradiente, textura, etc.). A partir de dichos atributos se obtendrá una función polinomial multivariada que se “**aproxime**” a los datos de la muestra. Debido a que se utilizará un Espacio L_∞ , con el propósito de no tener que cumplir con los requisitos estadísticos de los datos de entrada que se necesitan (o contribuyen a obtener resultados adecuados), en los métodos tradicionales; el cálculo del aproximante se considerará un problema de optimización combinatoria.

Por tanto, la herramienta que se utilizará, consiste en un algoritmo genético que servirá para encontrar la *forma del aproximante* y los *coeficientes* que intervienen en la función polinomial, de tal forma que se minimice el error de aproximación entre los datos y la función multivariada.

Cabe mencionar que lo anterior, se pudiera pensar que puede llevarse a cabo mediante el uso de la técnica de regresión no lineal (o algún método similar). Sin embargo, dichas técnicas (en su forma tradicional) trabajan bajo una métrica en L_2 . Lo que las hace susceptibles de tener las mismas limitaciones sobre los datos de entrada, de manera similar a la técnica de componentes principales.

En espacios multivariados, los métodos generalmente aplicados a la búsqueda de soluciones óptimas presentan diversas problemáticas que impiden llegar al óptimo esperado. Tales tipos de problemas, se engloban dentro del área denominada *Optimización Combinatoria*.

Como una alternativa a la mejora del proceso de búsqueda y al propio reconocimiento, se propone la utilización de un enfoque que combine Aproximación Multivariada y Algoritmos Genéticos para un mejor logro del reconocimiento de patrones. El enfoque mencionado, se utilizará para el reconocimiento holístico de rostros a partir de una base de datos.

En el método a utilizar para el Reconocimiento de Rostros, se encuentran involucradas principalmente tres áreas:

- Reconocimiento de patrones y aprendizaje supervisado,
- Algoritmos Genéticos y
- Aproximación multivariada, vista como un problema de optimización.

Las áreas mencionadas anteriormente se describen brevemente a continuación.

6.2. Reconocimiento de patrones y aprendizaje supervisado

De manera amplia, se puede decir que el Reconocimiento de Patrones es el área de la computación que se refiere a la descripción o clasificación (reconocimiento) de características (Schalkoff, 1992). El reconocimiento de patrones se puede caracterizar como

un proceso que tiene como objetivos la reducción de información, el mapeo de información o el etiquetado de información.

Las técnicas de reconocimiento de patrones son componentes importantes de los sistemas inteligentes y del área denominada Aprendizaje de Máquinas.

Un patrón, se puede conceptualizar como un conjunto de medidas u observaciones, las cuales presentan una serie de características (simbólicas o numéricas), tales como: color, peso, etc.

Una manera de concebir un reconocimiento de patrones, es mediante la utilización de un “conjunto de patrones”, en donde los atributos típicos, las clases o la estructura de cada patrón se conocen y se puede formar una base de datos denotada como “conjunto de entrenamiento”. Este conjunto de entrenamiento proporciona información significativa acerca de cómo asociar un conjunto de datos de entrada con una determinada decisión de salida. Por medio del conjunto

De acuerdo a Herbert Simon, el “aprendizaje” se puede definir como “cambios en un sistema que le permite realizar la misma tareas u otras subtareas de la misma población en forma más eficiente y más efectiva en tiempos posteriores” (Simon, 1983).

Se considera que se tienen dos formas principales mediante las cuales un sistema puede cambiar:

- El sistema puede adquirir conocimiento nuevo a partir de fuentes externas
- El sistema se puede modificar a sí mismo, para explotar en forma más efectiva su conocimiento actual.

La primera clase de aprendizaje se denomina “aprendizaje empírico” o “aprendizaje inductivo”, ya que se lleva a cabo por medio de un razonamiento a partir de ejemplos proporcionados del exterior, a partir de los cuales se producen reglas o procedimientos generales. Este tipo de aprendizaje se puede subdividir en dos tipos (Shavlik y Dietterich, 1990):

- Aprendizaje supervisado.
- Aprendizaje no supervisado

En el aprendizaje supervisado, se proporcionan ejemplos de la forma (x_i, y_i) y se asume una función de aprendizaje f tal que, $f(x_i) = y_i$. El objetivo es encontrar la función f , de tal forma que dicha función capture los “patrones generales” presentes en los datos de entrenamiento y se pueda aplicar para predecir valores y , a partir de diversos valores de x . Generalmente, cada x_i es una descripción de algún objeto, situación o evento y cada y_i es un descriptor simple (Shavlik y Dietterich, 1990):

No es necesario expresar a la función f como un conjunto de reglas, ésta, puede ser escrita en muy diversas notaciones: como una definición lógica, un procedimiento, un árbol de decisión, un polinomio, una red neuronal, etc.

Se denomina “supervisado” porque se puede pensar que los valores de y_i son proporcionados por un supervisor o profesor.

Cuando se tienen pocos valores de y_i , se les denomina clases, y la función f asignará cada x a su correspondiente clase. Si se tienen únicamente dos posibles valores para y_i , se les considera como ejemplos positivos o negativos de algún determinado concepto. Por tanto, la meta general del aprendizaje supervisado, es producir una regla de clasificación que asigne de manera correcta nuevos ejemplos a las clases propuestas.

En diversos casos, el conjunto de ejemplos de entrenamiento, representa únicamente una pequeña muestra del espacio de todos los posibles pares (x,y) . Como un ejemplo de lo mencionado anteriormente, se puede esperar que un médico aprenda a realizar diagnósticos correctos sin tener que haber atendido a todo posible paciente.

Sin embargo, esto puede provocar que el proceso de aprendizaje sea un problema “mal condicionado”, debido a que sin tener otras fuentes de restricción, no hay forma de conocer de manera totalmente correcta el valor de $f(x)$ para un valor de x que nunca se ha observado. Se desea que el algoritmo de aprendizaje encuentre definiciones de f que pueda generalizar a nuevos ejemplos, más que obtener definiciones *ad hoc* de f que únicamente se apliquen a los ejemplos observados durante el entrenamiento.

Para cumplir con lo anterior, es necesario incorporar restricciones adicionales dentro del algoritmo de aprendizaje. Este tipo de restricciones puede tomar la forma de lo que se denomina “espacio de hipótesis restringidas” (Mitchell,1980), la implementación de dicho espacio se puede formular como un problema de búsqueda en el espacio de hipótesis para una determinada f que minimiza el error sobre el conjunto de ejemplos de entrenamiento, en alguna de las siguientes formas:

- Dado un conjunto S de ejemplos de entrenamiento y un espacio de hipótesis restringidas (H). Encontrar: Una hipótesis $f \in H$ que minimice el número de ejemplos clasificados incorrectamente (S).
- Dado un conjunto S de ejemplos de entrenamiento y un ordenamiento de preferencia $el_mejor(f_1, f_2)$. Encontrar la “mejor” hipótesis f , de tal forma que se minimice el número de ejemplos clasificados incorrectamente en S .

Se requiere de la utilización de métodos eficientes de búsqueda para resolver los tipos de optimización mencionados anteriormente. Entre los métodos utilizados se encuentran: ID3, C4.5 (Quinlan,1993) para problemas que involucran conceptos lógicos. Perceptron, backpropagation y métodos de gradiente para algoritmos que involucran parámetros numéricos. Otras técnicas más recientes, susceptibles de poder emplearse, son el “recocido simulado” (simulated annealing) y algoritmos genéticos, entre otros.

6.3. Algoritmos Genéticos

Los algoritmos genéticos son técnicas de búsqueda y optimización que se guían por los principios de la Evolución y la genética natural. Son procesos de búsqueda eficientes, adaptativos y robustos que producen soluciones óptimas o cercanas al óptimo. Los trabajos iniciales fueron desarrollados por John Holland (Holland, 1995) en la Universidad de Michigan.

Los algoritmos genéticos son modelos que se basan en la utilización de una “población” (de posibles soluciones) y utilizan operadores genéticos para generar nuevos puntos de la muestra en un espacio de búsqueda. Las características fundamentales de los algoritmos genéticos son:

- La utilización de una codificación de los parámetros de un problema como una cadena de unos y ceros. Dicha codificación representa el equivalente a un cromosoma que define a un individuo. Cada bit de la cadena es equivalente a un gen y el valor particular del bit, representa un alelo.

La utilización de tres operadores básicos:

- **Selección.** Los individuos que pasarán de una generación a otra debe hacerse considerando la aptitud del individuo y la aptitud promedio de la población actual. Los individuos con una aptitud mayor que el promedio tienen una alta probabilidad de sobrevivir.
- **Cruzamiento.**- El cual permite la generación de individuos nuevos tomando características de los individuos padres. Su operación consiste en seleccionar dos cromosomas (soluciones que serán cruzadas), determinar la posición del cruce y generar a partir de lo anterior dos nuevas soluciones (figura 35).
- **Mutación.** Se considera como un operador que permite introducir información nueva no presente en una población. Opera sobre un individuo seleccionado, determina una posición aleatoria en la cual invertirá el valor del bit y reemplazará el individuo anterior con el nuevo.

El primer paso en la implantación de un algoritmo genético es la generación de una población inicial, en la cual, cada miembro o sujeto de dicha población es una cadena generalmente binaria (pueden existir otros tipos de alfabetos), de longitud “L” que corresponde a la codificación requerida por el problema en cuestión. Cada cadena se denomina como “genotipo” (Holland, 1995) o “cromosoma”. En la mayoría de los casos, la población inicial se genera en forma aleatoria. Después de contar con la población inicial, cada cadena se evalúa por medio de una función objetivo (aptitud o fitness) y se le asigna un valor de aptitud (o fitness).

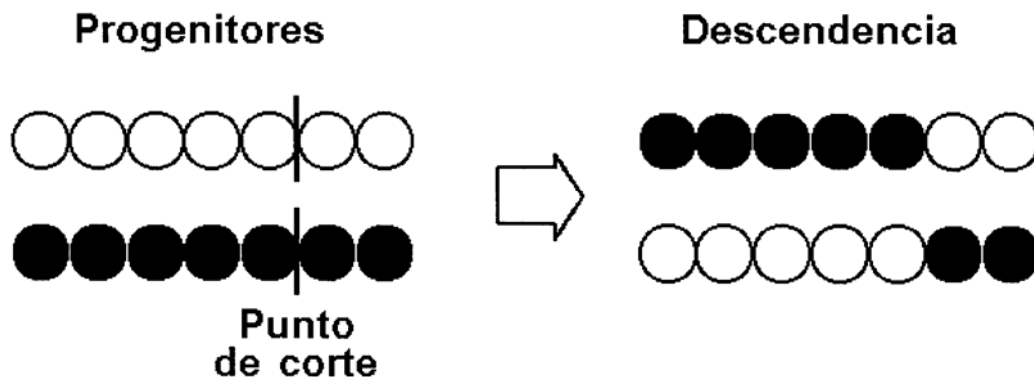


FIGURA 35. CRUZAMIENTO DE CROMOSOMAS EN UN SOLO PUNTO

El genotipo o cromosoma es un vector de la forma $\langle x_1, x_2, \dots, x_L \rangle$ en donde cada x_i es un alelo o gene. El dominio de valores a partir de los cuales x_i se selecciona es denominado el alfabeto del problema. Como se ha mencionado en párrafos anteriores, el alfabeto es binario $\{0, 1\}$, pero se pueden utilizar otros tipos de representaciones. Diversos cromosomas, forman una “población” y cada cromosoma se puede visualizar como un

“sujeto” específico de dicha población y que corresponde a una posible solución de un determinado problema. Por tanto, al tener diversas cromosomas o sujetos, implica el contar con diversos puntos simultáneos (paralelismo implícito) en el espacio de posibles soluciones. Lo anterior, es una de las ventajas de los algoritmos genéticos en contra de otros métodos, al tener la posibilidad de explorar en forma simultánea diversos puntos del espacio de búsqueda, en lugar de un solo punto, como la realizan los métodos tradicionales.

Los algoritmos genéticos difieren de los métodos tradicionales en diversas formas, entre estas:

- Un algoritmo genético trata de establecer un “compromiso” entre la exploración de nuevos puntos en el espacio de búsqueda y la explotación de la información que se ha encontrado.
- Un algoritmo genético presenta la propiedad de un paralelismo implícito. Dicho paralelismo permite que el efecto de búsqueda sea equivalente a una extensa búsqueda de hiperplanos sobre un determinado espacio, sin tener que probar directamente todos los posible hiperplanos.
- Los algoritmos genéticos operan sobre un conjunto de soluciones en forma simultánea, agrupando información a partir de puntos actuales de búsqueda, y a partir de éstos, realizar subsecuentes búsquedas. Esta forma de mantener soluciones múltiples permite que los algoritmos genéticos sean menos susceptibles de error ante problemas que presenten ruido y mínimos o máximos locales.

Funcionamiento de un algoritmo genético

La operación de un Algoritmo Genético simple se puede ilustrar con el siguiente segmento de pseudo-código.

```
Generar_Población_Inicial, G(0) ;  
Evaluar G ( 0 ) ;  
T = 0 ;  
Repetir  
    T = t + 1 ,  
    Generar G ( t ) utilizando G (t-1) ;  
    Evaluar G ( t ) ;  
Hasta encontrar_una_solución o Llegar_límite_iteraciones ;
```

En primer lugar, se genera en forma aleatoria la población inicial, la cual se constituye por un conjunto de cromosomas (cadenas de caracteres) que representan las posibles soluciones al problema. A cada uno de los cromosomas de dicha población, se le aplicará la función de aptitud o fitness, con el fin de conocer que tan buena es la solución a partir de los valores codificados en las cromosomas.

Conociendo la “aptitud” de cada cromosoma, se procede a la selección de los elementos que se cruzarán en la siguiente generación (se seleccionan a las mejores cromosomas). El método de selección más común es por medio de una “Ruleta”. Este método propuesto por Goldberg (1989) es muy simple, consiste en crear una ruleta en la cual cada cromosoma tiene asignada una fracción de la ruleta proporcional a su aptitud (fitness).

Sin referirse a ninguna función de aptitud en particular, si se supone que se tiene una población de 4 cromosomas cuyo aptitud esta dada por los valores mostrados a continuación (tabla 25):

TABLA 25. POBLACION INICIAL EN UN ALGORITMO GENETICO SIMPLE

CROMOSOMA	FUNCION APTITUD F (x)	%	NUMERO DE VECES SELECCIONADA
01101	169	14.4	1
11000	576	49.4	2
01000	64	5.5	0
10011	361	30.9	1

Con los porcentajes mostrados en la tercera columna de la tabla 25, se puede elaborar una *ruleta*, la cual se gira “n” veces para determinar de forma aleatoria que individuos (cromosomas) se seleccionarán. Debido a que los individuos más aptos se les asignó un área mayor de la ruleta, se espera que sean seleccionados un mayor número de veces que los menos aptos. Por tanto, las nuevas cromosomas seleccionadas de la tabla anterior, formarán la siguiente población:

01101
11000
11000
10011

Una vez realizada la selección, se procede a la utilización de la Reproducción o cruza de los elementos seleccionados en esta etapa, los sujetos seleccionados intercambiarán material cromosómico y sus descendientes formarán la población de la siguiente generación. Las dos formas más comunes de reproducción son:

- Utilización de un único punto de cruzamiento
- Utilización de dos puntos de cruce.

El punto de cruzamiento se selecciona de forma aleatoria sobre la cadena que representa a la cromosoma, a partir de dicho punto se intercambia el material de los dos sujetos seleccionados (figura 35).

Generalmente, el cruzamiento se maneja dentro de la implementación del algoritmo genético, como una probabilidad, con lo cual no todas las parejas de cromosomas se cruzarán, sino que habrá algunas que permanecerán intactas en la siguiente generación.

Existen otros procedimientos para manipular cromosomas, una de éstas, es que los individuos más aptos permanezcan a lo largo de las distintas generaciones, sin realizar cruce con nadie. Se mantiene intacta la cromosoma hasta que surge otro individuo mejor. A este método se le denomina elitismo.

Otro operador es la mutación, el cual realiza un cambio a uno de los genes de un cromosoma elegido de forma aleatoria. Cuando se utiliza una representación binaria, un bit se sustituye por su complemento (un cero cambia en uno y viceversa). Este operador permite la introducción de nuevo material cromosómico en la población.

Al igual que el cruzamiento, la mutación se maneja por medio de un valor probabilístico que indique la frecuencia con que se efectuará, generalmente su probabilidad es mucho menor que la probabilidad utilizada en el cruzamiento. Una forma de realizar una mutación, consiste en generar números aleatorios para seleccionar uno de los bits que componen al cromosoma y cambiar el bit correspondiente.

Con el propósito de conocer el funcionamiento de un algoritmo genético, se presenta el siguiente sencillo ejemplo:

Maximizar $f(x) = x^2$ en el intervalo de enteros $[0,3]$

Los pasos a seguir, son los siguientes:

- a) Codificar el parámetro x como una cromosoma de longitud finita que pueda representar los posibles valores, se puede proponer una longitud de 5 caracteres binarios. Con dicha longitud se pueden obtener valores que se encuentran entre: 0 (00000) y 31 (11111).
- b) Suponer una población inicial aleatoria de cromosomas, en este caso se considera una población de 4 sujetos o cromosomas.

Sujeto 1: **0 1 1 0 1**
 Sujeto 2: **1 1 0 0 0**
 Sujeto 3: **0 1 0 0 0**
 Sujeto 4: **1 0 0 1 1**

- c) Evaluar la función de aptitud (fitness). Utilizando una “ruleta”, seleccionar los cromosomas más aptos:

TABLA 26. POBLACION INICIAL Y SELECCIÓN DE CROMOSOMAS

SUJETO	CROMOSOMA	VALOR DE x	FITNESS $F(x)$	% DEL TOTAL	# VECES SELECCIONADO
1	01101	13	169	14.4	1
2	11000	24	576	49.2	2
3	01000	8	64	5.5	0
4	10011	19	361	30.9	1
TOTAL			1170	100.0	
PROMEDIO			293		
MÁXIMO			576		

d) A partir de los resultados de la ruleta (tabla 26), se copian los cromosomas más aptas el número de veces seleccionado generando la siguiente “población intermedia”:

Sujeto 1: **0 1 1 0 1**
 Sujeto 2: **1 1 0 0 0**
 Sujeto 3: **1 1 0 0 0**
 Sujeto 4: **1 0 0 1 1**

e) Se utiliza el operador de cruzamiento, seleccionando en forma aleatoria las cadenas que se utilizarán para el cruce, así como el punto del cromosoma (gen) a partir del cual se realizará dicho cruce (tabla 27).

TABLA 27. CRUZAMIENTO Y NUEVA POBLACION

SUJETO Y CROMOSOMA	SUJETO PARA CRUCE	SITIO DE CRUCE	NUEVA POBLACION	VALOR DE X	F(X)
1 0110 1	2	4	01100	12	144
2 1100 0	1	4	11001	25	625
3 11 000	4	2	11011	27	729
4 10 011	3	2	10000	16	256
TOTAL					1754
PROMEDIO					439
MAXIMO					729

Para el operador de mutación, se pueden generar números aleatorios para cada gen, de cada cromosoma, y los que tengan un valor menor que un determinado umbral (por ejemplo, los

menores a 0.003 deben tener mutación). Por ejemplo, un cromosoma 0 0 1 0 puede presentar mutación y cambiar a 0 0 1 1.

En la población de la tabla anterior, se asume que no se presentó mutación y la nueva población es la que aparece representada.

f) Con la nueva población, se evalúa la función de aptitud y se considera si es necesario detener el proceso, o bien, pasar a una nueva Generación regresando al paso (c). Se puede observar en la tabla 27, que el valor de la función de aptitud mostró una mejora en comparación al resultado que se obtuvo en la tabla 26. Después de un pequeño número de Generaciones se puede obtener la siguiente configuración (tabla 28).

TABLA 28. POBLACION FINAL

SUJETO	CROMOSOMA	VALOR DE x	FITNESS F (x)
1	11011	27	729
2	11111	31	961
3	11111	31	961
4	11111	31	961
TOTAL			3612
PROMEDIO			903
MAXIMO			961

El resultado es: $x = 31$ con $f (x) = 961$

Los modelos más sencillos de algoritmos genéticos cumplen con las características básicas mencionadas anteriormente. Sin embargo, a los algoritmos genéticos se les puede adicionar

otros operadores y diversas características adicionales con el propósito de que muestren mayor “robustez” en la solución de problemas.

Algoritmo Genético Ecléctico (AGE)

En el desarrollo que se realizará (en un futuro cercano), del método propuesto para reconocimiento de rostros, se pretende utilizar el denominado “Algoritmo Genético Ecléctico (AGE)” propuesto por Kuri (1997, 1998). En diversos experimentos realizados acerca de optimización compleja de funciones, dicho algoritmo ha presentado gran eficiencia (Kuri, 1998; Kuri y Villegas, 1998) .

El Algoritmo Genético Ecléctico incorpora los siguientes elementos:

- Elitismo total
- Selección determinística
- Cruzamiento anular
- Mutación aleatoria hillclimbing
 - Determinación adaptativa de los siguientes parámetros :
 - Número de descendientes
 - Probabilidad de cruzamiento
 - Probabilidad de mutación
 - Probabilidad de hillclimbing
 - Número de evaluaciones de la función hillclimber

6.4. Aproximación multivariada en un espacio L_∞

como un problema de optimización combinatoria

La problemática general que se presenta en la Teoría de la aproximación, consiste en poder encontrar para una serie de puntos conocidos, la interrelación que permita expresar una variable dependiente en términos de una variable independiente (aproximación univariada) o bien, dos o más variables independientes (aproximación multivariada). Por tanto, se desea encontrar una determinada función, la cual se “aproxime” lo más posible a los puntos conocidos. Para conocer qué tanto se acerca la función a los puntos verdaderos, se requiere de estar en posibilidad de “medir” la distancia de separación. Esta forma de medida se puede lograr con el uso de espacios lineales y utilizando lo que se denomina norma. La norma permite tener la noción de distancia en el espacio lineal.

Las funciones polinomiales más simples son de la forma:

$$P_n(x) = c_0 + c_1x + \dots + c_nx^n \quad (28)$$

La base de la teoría de la aproximación de funciones es un teorema descubierto por Weierstrass:

“Para funciones continuas de una simple variable real definida sobre el segmento finito $[a,b]$ se asevera que, para cualquier función $f(x)$ continua sobre $[a,b]$, existe una secuencia de polinomios ordinarios que convergen uniformemente a $f(x)$ sobre $[a, b]$.”

Esta característica, se aplica también a las funciones de muchas variables (multivariada) en un espacio *multi-dimensional*, por tanto, si $f(x_1, x_2, \dots, x_m)$ es tal función, existe una secuencia de polinomios ordinarios de la forma:

$$P_{n_1, n_2, \dots, n_m}(x_1, x_2, \dots, x_m) = \sum_{k_1=1}^{n_1} \sum_{k_2=1}^{n_2} \dots \sum_{k_m=1}^{n_m} c_{k_1, k_2, \dots, k_m} x_1^{k_1} \dots x_m^{k_m} \quad (29)$$

que convergerán en forma uniforme en una región cerrada delimitada. Un elemento importante, es la asignación del “grado” máximo a cada una del variables del polinomio con el propósito de obtener la mejor configuración y un mejor ajuste a los datos. Sin embargo, no se tienen métodos totalmente confiables para realizar dicha asignación. Asimismo, para un determinado “grado máximo” propuesto para un polinomio, el espacio de posibles alternativas que pueden presentar los coeficientes crece en forma exponencial.

Por ejemplo, para 10 variables y un grado máximo de 6, el número de posibles alternativas es de $7^{10} \approx 2.82475249 \times 10^8$, para 20 variables y un grado máximo de 10 se tiene $\approx 1.667988 \times 10^{13}$ alternativas. Se puede apreciar, que el número de combinaciones es muy grande, por lo cual, se considera que los algoritmos genéticos serán de gran utilidad en el proceso de configuración del polinomio.

Otro punto importante, es la utilización de una métrica adecuada para definir en forma precisa la interrelación entre las variables y los datos. La norma más frecuentemente utilizada es la L_2 (aproximación por Mínimos Cuadrados), muy utilizada en estadística. Sin embargo, dicha norma produce en la mayoría de los casos, sistemas *mal-condicionados*.

Asimismo, desde el punto de vista estadístico, se requiere que los datos cumplan con ciertos supuestos: Homocedasticidad, independencia y normalidad (como ya se ha mencionado en capítulos anteriores de la tesis).

La utilización de la norma L_∞ (norma **minimax** o Tchebyshev) no presenta las limitaciones mencionadas anteriormente, y por tanto, será la norma utilizada como parte del desarrollo del método propuesto.

Con base a lo planteado anteriormente, se propone que el método para reconocimiento de rostros contemplará:

“que el problema de aproximación multivariada se planteará como un problema de optimización combinatoria, en el cual para un determinado grado máximo propuesto para el polinomio que caracterizará a un determinado rostros se realice una búsqueda de la forma del aproximante y los coeficientes que mejor caractericen la interrelación entre el conjunto de variables independientes y la dependiente, bajo la norma L_∞ .

Como se ha mencionado, la solución a la problemática anterior cae dentro del área que se ha denominado optimización combinatoria, y es un problema de difícil resolución por métodos tradicionales.

Los problemas de optimización pertenecen a una clase de problemas cuya solución puede obtenerse a través de un algoritmo de búsqueda en el que se trata de encontrar un elemento particular, sujeto a ciertas restricciones, dentro de un espacio de estados definidos por una colección de parámetros, cada uno de los cuales toma valores dentro de un dominio finito.

El inconveniente de los enfoques tradicionales de optimización y de algunos enfoques heurísticos, es que en el primer caso, se requieren del conocimiento de derivadas y en el segundo caso (problemática que también afecta al primer enfoque), es que pueden quedar atrapados en un óptimo local (figura 36).

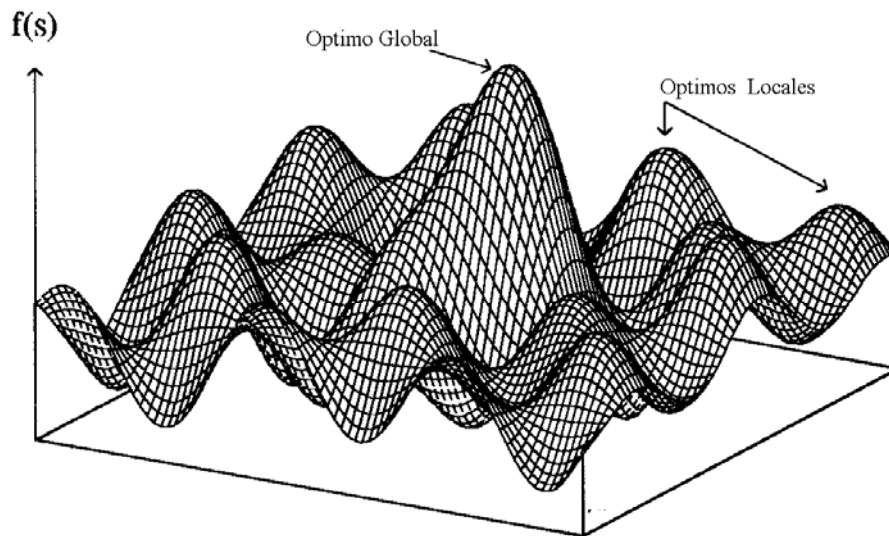


FIGURA 36. OPTIMOS LOCALES Y GLOBALES EN UN ESPACIO MULTIDIMENSIONAL

Los problemas de optimización pueden ser planteados matemáticamente mediante la consideración de la noción de distancia o criterio de mérito (fitness). En su forma más general, un problema de optimización se reduce a encontrar el valor máximo o mínimo de una función $f : D \rightarrow \mathfrak{R}$, denominada función objetivo. En los casos más sencillo D es algún subconjunto del espacio real de n dimensiones, pero puede ser cualquier conjunto de entidades.

Cuando $D \subset \mathbb{R}^n$ y en la función f las derivadas parciales de primer orden existen y son continuas, la determinación del máximo o mínimo de dicha función se reduce al análisis de los valores de la función en los puntos críticos, donde se anulan las derivadas, y por tanto:

$$\frac{\partial f}{\partial x_i} = 0 \quad \text{para } i = 1, \dots, n \quad (30)$$

La desventaja principal de este método es que emplea únicamente información local para guiar la búsqueda de valores óptimos en el espacio de parámetros, por lo que es fácil localizar mínimos locales cuando la topología de la función es complicada (lo cual es común, en problemas multivariados), después de lo cual la optimización se interrumpe. Otra desventaja del método es que requiere de información adicional al valor de la función objetivo, como sus derivadas, las cuales pueden no estar definidas en bastantes casos prácticos.

Otro enfoque consiste en emplear un procedimiento exhaustivo de enumeración de las distintas posibilidades para la optimización de funciones en espacios finitos de búsqueda. En este caso, el algoritmo de búsqueda explora los valores de la función objetivo en todos y cada uno de los puntos del espacio. Sin embargo, pese a la sencillez del procedimiento, en la mayoría de los casos el número de posibilidades será tan grande que dicho método carece de valor práctico.

Como una alternativa a las problemáticas anteriores, se postula la utilización de Algoritmos Genéticos, los cuales como se ha mencionado en secciones anteriores, son procedimientos matemáticos altamente paralelos que transforman una población de elementos del espacio de n dimensiones en una nueva población que trata de mejorar el valor de la función objetivo.

Una de las ventajas de estos algoritmos en la solución de problemas de optimización, es que requieren únicamente de información relativa a los valores de la función objetivo para llevar a cabo la optimización, y utilizan procedimientos estocásticos para guiar la búsqueda en el espacio de parámetros, llevando a cabo la búsqueda simultáneamente en diferentes puntos del espacio.

La aplicación de algoritmos genéticos a la aproximación multivariada, la cual a su vez se aplicará a la resolución de problemas de reconocimiento de patrones, vistos estos últimos como un proceso de aprendizaje supervisado, se considera que es un enfoque de gran actualidad. Con este tipo de enfoque se vislumbra la posibilidad de mejorar los resultados que se obtienen los métodos tradicionales de reconocimiento de patrones.

6.5. Método general para reconocimiento de rostros utilizando algoritmos genéticos eclécticos y aproximantes polinomiales en un espacio L_∞

Por tanto, el método general que se propone para realizar reconocimiento de rostros y se considera reducirá algunas deficiencias que se presentan en métodos como el análisis de componentes principales, utilizará principalmente las tres áreas de conocimiento mencionadas en las secciones anteriores. Específicamente, se plantean los siguientes pasos:

6.5.1. Muestreo Aleatorio de pixeles y configuración de variables

Se realizará un muestro aleatorio de los pixeles de las imágenes digitales de cada rostro que configuren la base de entrenamiento. Lo anterior tiene como propósito, reducir la información que se almacenará. Se piensa, que en principio, no se requiere de toda la información de cada fotografía (como se realiza en el uso de componentes principales).

Por ejemplo, para la base utilizada en el presente trabajo (base AT&T), cada imagen presenta un tamaño 92x112 pixeles (10,304 pixeles). El método propone la selección de una muestra aleatoria de los pixeles de la imagen, cuidando de no obtener pixeles duplicados. Utilizado la fórmula de Peña y Romo [28], se obtiene que el tamaño de la muestra será de 1,802.16 pixeles (17.49% de la muestra). Por tanto, una muestra del 17%, equivalente a 1,752 pixeles sería adecuada.

A partir de dichos pixeles, se obtendrían otras características de ellos: gradiente, máxima entropía, textura, etc. Con la información de las “m” variables de características de pixeles, se generaría un archivo (tabla 29).

**TABLA 29. MATRIZ DE CARACTERÍSTICAS DE PÍXELES
 (“m” CARACTERÍSTICAS x “Tm” PÍXELES) PARA OBTENER EL APROXIMANTE QUE
 PUEDA CARACTERIZAR AL ROSTRO**

POSICIÓN SECUENCIAL DE PÍXEL	COORDENADA “X”	COORDENADA “Y”	VALOR DE GRIS (GRAY SCALE)	ENTROPÍA DEL PÍXEL	TEXTURA DE PÍXEL	. . .	VARIABLE “m” DE CARACTERÍSTICA DEL PÍXEL
1							
2							
. . .							
Tm							

La tabla anterior se puede considerar como la matriz “Rostro”, a partir de la cual se obtendrá el aproximante que minimice el error de ajuste. Por tanto, se tendrá:

$$Rostro(VC_{i,j}) \quad i = 1, Tm; \quad j = 1, nc \quad (31)$$

donde:

- VC Es el valor numérico que representa a cada una de las variables que se utilizarán como *características* del rostro: nivel de gris, coordenada, entropía, textura, etc.
- Tm El número de pixeles que se utilizará por rostro, de acuerdo al muestreo aleatorio.
- Nc El número de variables características (VC)

De las variables (columnas) que configuran la tabla anterior, una de ellas actuará como variable dependiente y las otras, como variables independientes.

6.5.2. Reconocimiento de rostros como un problema de aprendizaje y obtención del aproximante multivariado mediante algoritmos genéticos.

Como se ha mencionado, en el aprendizaje supervisado, se proporcionan ejemplos de la forma (x_i, y_i) y se puede asumir una función de aprendizaje f , tal que, $f(x_i) = y_i$.

El objetivo consiste en encontrar la función f , de tal manera que dicha función capture los “patrones generales” presentes en los datos de entrenamiento y se pueda aplicar para predecir valores de y , a partir de diversos valores de x .

Generalmente, cada x_i es una descripción de algún objeto, situación o evento, al igual que las y_i (Hallinan, 1991). La función se puede hacer extensiva al manejo de diversas

variables en un espacio “n-dimensional”. Así, la tabla Rostro(i,j) será la que se utilizará como función de aprendizaje.

Los valores utilizados para la función que caractericen un rostro, pueden ser atributos de los píxeles de cada imagen (coordenadas, nivel de gris o color, gradiente, ruido, etc.).

Utilizando el enfoque anterior, estamos en posibilidad de caracterizar un rostro de una manera holística, sin tener que especificar elementos geométricos o plantillas de regiones.

Como se ha mencionado anteriormente, a través de una función se pretende caracterizar un determinado rostro. Específicamente, se ha seleccionado una familia de funciones polinomiales de aproximación, cuyo propósito será caracterizar un rostro a partir de una serie de atributos de los píxeles de cada imagen. De estos atributos, alguno se seleccionará como variable dependiente y otros atributos, se plantearán como variables independientes.

La aproximación del polinomio (a los atributos seleccionados), tradicionalmente, se ha realizado mediante la técnica de regresión lineal o regresión múltiple. Sin embargo, para aplicar dicho método se presupone que los datos cumplen con ciertas características: las variables deben tener una distribución normal, las distribuciones deben tener la misma varianza, para un valor de la variable independiente, la distribución de los valores de la variable dependiente debe tener una media que se encuentre en la línea de regresión, etc.

Bastantes problemas de la vida real, entre ellos los datos de un rostro humano, no cumplen con las características anteriormente descritas.

El objetivo del método, consiste en encontrar la *forma* y los *valores de los coeficientes* del **polinomio** que mejor caractericen la interrelación entre el conjunto de variable(s) independiente(s) y la dependiente bajo la norma L_{∞} . La solución a la problemática anterior

cae dentro del área que se ha denominado Optimización Combinatoria, y es un problema de difícil resolución por métodos tradicionales (Grötschel y Lobas, 1993).

Como ya se ha mencionado, para realizar la búsqueda y optimización respectiva, se plantea la utilización de un método propuesto por Kuri (1997, 1998), denominado Algoritmo Genético Eclético.

Utilizando como herramienta el algoritmo genético eclético, se plantea encontrar la forma y los valores de los coeficientes del polinomio de aproximación en el espacio de búsqueda, de tal forma que se minimice el máximo error absoluto de aproximación entre los datos y la función aproximante. La forma polinomial que en principio se desea obtener es del tipo:

$$f(v_1, \dots, v_p) = \sum_{i_1=0}^{g_1} \dots \sum_{i_p=0}^{g_p} C_{i_1 \dots i_p} v_1^{i_1} \dots v_p^{i_p} \quad (32)$$

Por tanto, en la fase de entrenamiento, se caracterizarán mediante polinomios los rostros de la base de datos. Esta aproximación se realizará a partir de diversos atributos de cada rostro, los cuales se obtuvieron de una muestra de pixeles de cada rostro, como se mencionó en la sección 4.3.

Posteriormente, un determinado rostro a identificar, se caracterizará también mediante una familia de polinomios, los cuales se comparan con los polinomios que corresponden a los rostros de la base de datos (entrenamiento) utilizando la métrica de Mahalanobis. La cantidad r en

$$r^2 = (x - m_x)^T C_x^{-1} (x - m_x) \quad (33)$$

es llamada la Distancia de Mahalanobis a partir del vector de características x al vector de medias m_x , en donde C_x es la matriz de covarianza para x .

Se puede utilizar la distancia de Mahalanobis en un *clasificador de distancia mínima* en la siguiente forma:

Dado que m_1, m_2, \dots, m_c sean los valores “medios” para las “c” clases, y C_1, C_2, \dots, C_c sean las correspondientes matrices de covarianza. Se puede clasificar un vector de características x al medir la distancia de Mahalanobis desde x , a cada una de las “medias” de las clases. Entonces se asignará x a la clase para la cual la distancia es mínima (figura 37).

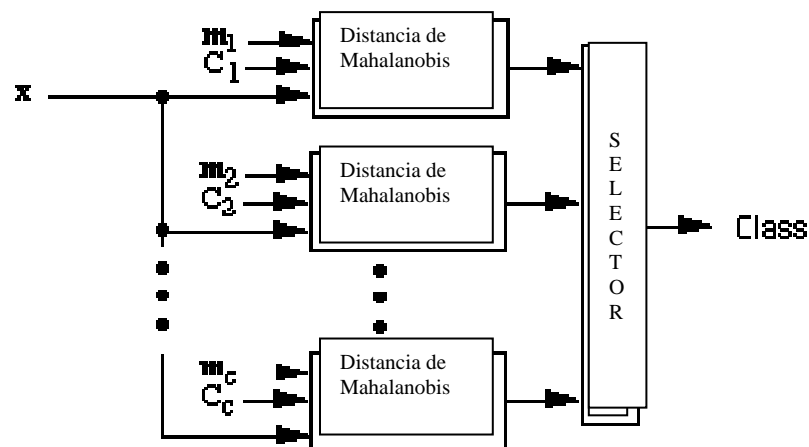


FIGURA 37 CLASIFICADOR MAHALANOBIS DE DISTANCIA MINIMA

Se propone utilizar el clasificador anterior para la clasificación del “rostro a reconocer”, en lugar del clasificador Euclidiano (utilizado en los experimentos de componentes principales).

6.5.3. El rostro y sus Espectros de Señal.

Como un ejemplo básico para empezar a verificar las posibilidades de aplicación del método propuesto, se llevó a cabo un pequeño experimento con algunos rostros de la base AT&T.

Una vez realizado el preprocesamiento de la imagen, los valores de nivel de gris correspondientes a cada pixel, se almacenan en un vector (cada renglón del rostro se coloca en forma secuencial), como se representa en la figura 38. Se aplica el muestreo aleatorio descrito en la sección 6.5.1, obteniendo de las 10,304 posiciones un vector de 1,752 posiciones (correspondientes a una muestra del 17%).

Graficando este vector, tomando como variable (x) la posición de cada elemento del vector y los niveles de gris como (y), se obtiene el equivalente a una serie de tiempo o espectro de señal. En una primera instancia, se puede pensar en tratar de obtener el polinomio de aproximación que caracterice a dichos datos que representan un determinado rostro. O bien, si utilizamos las coordenadas (x, y) y el correspondiente nivel de gris en ese punto se obtiene una gráfica del tipo (X, Y, Z). Los experimentos realizados con estos tipos de espectros no tuvieron un resultado satisfactorio en el reconocimiento del rostro. Se consideró que los espectros se deberían de simplificar para obtener un mejor resultado.

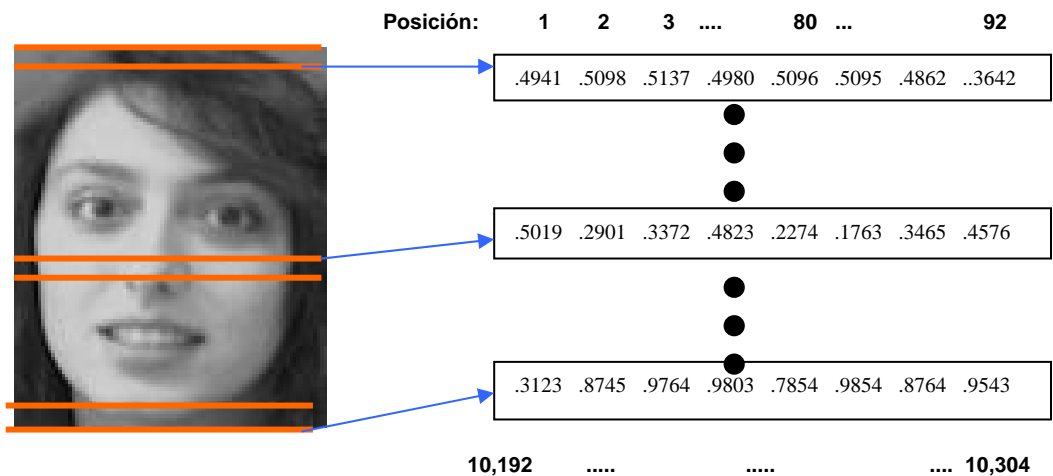


FIGURA 38 REPRESENTACION VECTORIAL DE LOS VALORES DE GRIS CORRESPONDIENTES A LA IMAGEN DE UN ROSTRO

Con base en algunas de las propuestas teóricas acerca del reconocimiento de rostros, se obtuvieron tres espectros para cada imagen: Espectro de señal de baja frecuencia, Espectro de los Gradientes y Espectro de Máxima Entropía.

Las señales de baja frecuencia se seleccionaron debido a la propuesta de diversos autores (Hancock, Burton y Bruce, 2000; Wechsler y Phillips, 1998), que proponen que el ser humano utiliza este tipo de señal para realizar el reconocimiento de rostros. El Gradiente, se encuentra relacionado con la detección de los bordes de una imagen a través de las diferencias entre los niveles de grises de la imagen. Por su parte, el espectro de máxima entropía, es una técnica para tratar de reducir “ruido” en una señal y adquirir conocimiento a partir de información incompleta. El método de máxima entropía ha sido utilizando sobre todo, en el campo de la Astronomía para procesar imágenes obtenidas por medio de radiotelescopios que presentan una gran cantidad de “ruido”.

Se obtuvieron los tres espectros mencionados anteriormente, para cada imagen, a partir de la muestra seleccionada. Por tanto, se obtienen tres vectores de 1,752 posiciones que caracterizan a cada rostro. Para cada una de estas señales, se obtuvo el *aproximante polinomial* correspondiente por medio de los algoritmos genéticos eclécticos, el cual tiene la forma presentada en la fórmula (32)

En el proceso del algoritmo genético se utilizaron poblaciones de 30 individuos con aproximadamente 50 generaciones, los polinomios de aproximación que se encontraron presentan 12 términos y un grado (exponente) máximo de 6.

En la figura 39, se presentan las gráficas de los espectros para cada una de las tres señales: Gradiente (figura 39-a), Frecuencia baja (figura 39-b) y Máxima entropía (figura 39-c). Se han representado simultáneamente 10 señales (que caracterizan a 10 rostros de la base AT&T), para cada uno de los espectros. Únicamente se presentan 350 puntos del total de la muestra.

El reconocimiento de un determinado rostro, se realizó comparando sus tres espectros contra los espectros de los polinomios de aproximación que caracterizan a la base de entrenamiento, a través de la métrica de Mahalanobis (clasificador de distancia mínima).

Se utilizaron las imágenes correspondientes a la segunda y tercera pose de cada sujeto de la base AT&T, como sujetos a identificar. En ambos casos se reconocieron 39 sujetos, correspondiendo a un 97.5% de precisión.

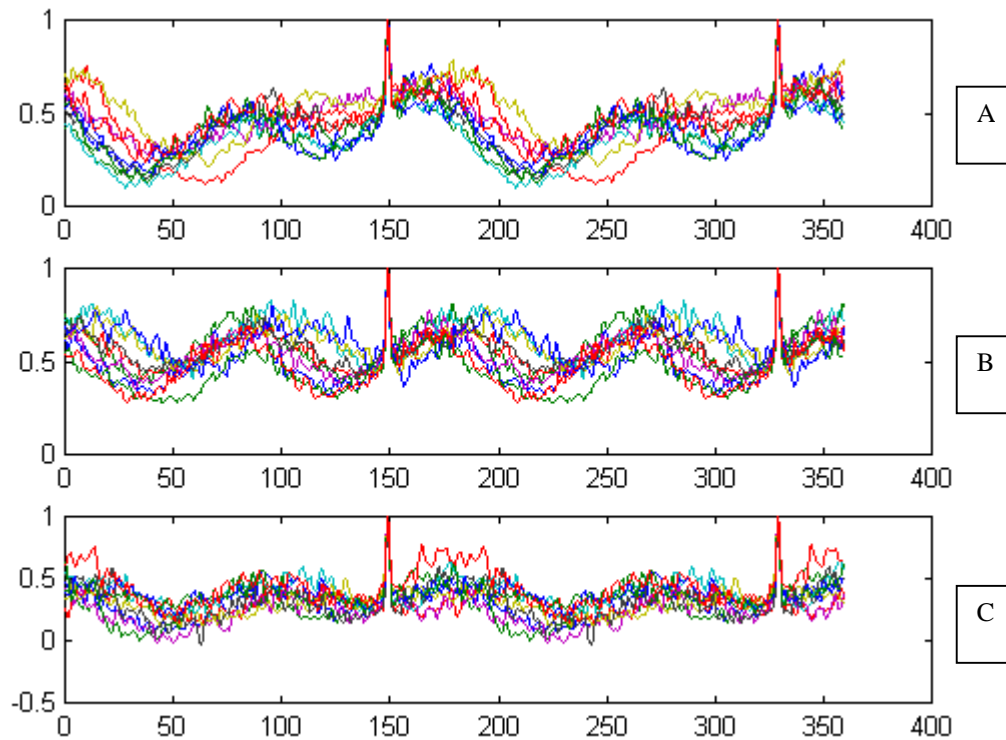


FIGURA 39. SEÑALES CORRESPONDIENTES A 10 ROSTROS DE LA BASE DE ENTRENAMIENTO “ORL”. (A) – GRADIENTE, (B) – FRECUENCIA BAJA, (C) – MAXIMA ENTROPIA

El método utilizado en este proyecto, que consiste básicamente en un muestreo aleatorio de la imagen de un rostro, su caracterización por medio de polinomios de aproximación y la búsqueda de la forma de dichos polinomios a través de un algoritmo genético, es una de las principales aportaciones de la investigación. Con la metodología anterior, no se requiere de calcular los elementos geométricos del rostro, ni obtener la posición de algún elemento básico del rostro humano (ojos, nariz, boca, etc.), ni tampoco del uso de “plantillas” que enmarquen características faciales, métodos utilizados en la mayoría de las investigaciones que se presentan en la literatura. En los métodos que se basan en la utilización de

características, se requiere de realizar búsquedas exhaustivas para identificar los ojos, la nariz o cualquier otro elemento facial a partir del cual se construye la red de puntos que configuran al rostro y a partir del cual se aplica alguno de los métodos requeridos para el reconocimiento (componentes principales, redes neuronales, estadística, etc.). Incluso, en algunos sistemas, se requiere que el usuario identifique “manualmente” algún punto del rostro humano para iniciar el proceso computacional.

El método propuesto en esta investigación será completamente automático y no requiere de conocer específicamente en dónde se encuentran los rasgos faciales de un determinado sujeto, para llevar a cabo el entrenamiento de la base de rostros y posteriormente el reconocimiento de un determinado sujeto. Asimismo, con la utilización del muestreo aleatorio, se requiere únicamente una pequeña cantidad de píxeles para llevar a cabo la aproximación polinomial en la fase de entrenamiento, al igual que en la fase de identificación. Lo anterior reditúa en un menor tiempo de procesamiento. Por otra parte, la utilización de un enfoque holístico, en el cual no se requiere de identificar específicamente rasgos faciales, permite reconocer un alto porcentaje de sujetos que presentan ciertas oclusiones faciales: barba, bigote, anteojos. Otros sistemas presentan mayor dificultad en realizar este tipo de reconocimiento e, inclusive, algunos no lo pueden realizar.

Los resultados iniciales que se obtuvieron en el reconocimiento de rostros con la base AT&T logró un 97.5% de precisión, lo cual se considera altamente satisfactorio. Sobre todo considerando que únicamente se utilizó una sola imagen por sujeto en la fase de entrenamiento (a diferencia de la mayoría de los métodos reportados en la literatura que requieren varias imágenes para el entrenamiento). El método es más robusto.

En las siguientes fases de esta investigación, se diseñará e implementará un programa computacional que se base en el método propuesto. Asimismo, realizarán más pruebas con otras bases de rostros (con mayor número de imágenes).

Además, se plantearán algunas funciones multivariadas que combinen diversas señales obtenidas a partir de los rostros. Se pretende realizar un análisis matemático acerca de los procesos que subyacen en la caracterización de rostros a través de los diversos espectros. Se planteará la utilización de otros tipos de señales, además o en lugar de las tres ya experimentadas.

Por otra parte, se considera que el sistema desarrollado mediante aproximación multivariada y algoritmos genéticos tiene altas posibilidades de ser aplicado a diversos tipos de reconocimiento de patrones: imágenes de diversos tipos, series de tiempo, reconocimiento de imágenes utilizadas en medicina, astronomía, etc.

Conclusiones

A partir de los resultados que se obtuvieron en el análisis estadístico de las imágenes de rostros y en los diversos experimentos realizados, se concluye:

- Debido a la amplia posibilidad de rangos (entre 0.0 y 1.0), que se presentan en imágenes digitales en tonos de grises de rostros humanos, la dispersión que presentan es muy grande. En el análisis realizado a la base AT&T y con fotografías de rostros que se obtuvieron en Internet, el promedio de los tonos de los píxeles que se encontró fue de 0.5208. Y la desviación estándar, de ± 0.2947 . Lo anterior, implica un amplio rango de valores. Por tanto, tratar de obtener algún tipo de análisis por medio de estadística descriptiva únicamente, se considera inadecuado. Se considera que la información del rostro, se encuentra inmersa en las diversas interrelaciones que se presentan entre los píxeles a nivel local en diversas áreas y la estructura latente que se configura por la interrelación de estas áreas.

- Por lo que respecta al supuesto de normalidad univariada y multivariada, se probó, que no se cumple. Tanto a nivel individual de una fotografía, como a nivel global de todo un conjunto de imágenes (se consideró una muestra estadística de fotografías), los rostros no se ajustan a una distribución normal multivariada.
- Por tanto, al violar este importante supuesto estadístico, se considera que la aplicación de métodos estadísticos “tradicionales”: regresión, análisis factorial, análisis de componentes principales y en general, cualquier método relacionado que trabaje bajo métricas en espacios L_2 tendrá grandes posibilidades de obtener resultados erróneos. A pesar de la “robustez” que en general, presentan dichos métodos.

Por lo que respecta al comportamiento del sistema de reconocimiento de rostros en los diversos experimentos realizados, se puede decir lo siguiente:

- Un elemento fundamental, es la asignación adecuada del “umbral” para clasificar y poder identificar adecuadamente al rostro respectivo. Dicho umbral debe reducir el número de falsos positivos (inocentes considerados como sospechosos) y el de falsos negativos (sospechosos no identificados), al mismo tiempo que trata de maximizar el número de positivos verdaderos (sospechosos identificados). En los diversos experimentos que contemplaron varias fotografías por cada sujeto para entrenamiento de la base (desde 9 imágenes a 1 imagen), se plantea, que un umbral adecuado sería el que se obtiene del promedio de las distancias del rostro al “espacio de rostros”, más dos desviaciones estándar).

- En la experimentación, se encontró que los rostros que se tomaron del portal de sospechosos del FBI, así como las fotografías que se obtuvieron por medio de un scanner (las cuales, no tuvieron un proceso controlado en su adquisición como sería el caso de la base AT&T), presentan distancias muy superiores al promedio de las distancias de fotografías controladas (con respecto al “espacio de rostros”) de las fotografías controladas.
- La conclusión anterior implica, que el sistema de reconocimiento de rostros, requiere de un “umbral grande” para poder realizar un adecuado reconocimiento de las fotografías de los sospechosos. Sin embargo, al tener dicho umbral, una gran cantidad de fotografías “controladas”, serían clasificadas como “falsos positivos”. Aparentemente, las fotografías que no presentan un control en el momento de tomarlas, requieren de umbrales superiores.
- El método de componentes principales mostró comportamientos adecuados. El porcentaje de reconocimiento de rostros en condiciones controladas (base AT&T) con respecto a si mismas, obtuvo porcentajes elevados. Sin embargo, se presentan serias deficiencias cuando se mezclan imágenes que no presentan un control en la iluminación, fondo de imagen, resolución, etc.
- En diversos casos, la técnica de componentes principales, mostró gran eficiencia. Por ejemplo, al identificar algunos sospechosos de la base del FBI, cuya edad era

muy superior en la fotografía a identificar con respecto a su fotografía de la base de entrenamiento.

- Se encontró, que el método de componentes principales no funciona de forma adecuada con bases de entrenamiento que únicamente cuenten con una fotografía por persona. El número ideal para este tipo de sistemas, sería entre 5 y 7 imágenes por sujeto.
- En relación a la conclusión anterior, se puede mencionar, que la hipótesis del trabajo de la tesis no se cumplió totalmente. Pues si bien, el sistema sí logra realizar reconocimientos de rostros a partir de solo una imagen, la confiabilidad no es adecuada en la mayoría de los experimentos que se realizaron.
- El incumplimiento parcial de la hipótesis, contribuye a reforzar el postulado del autor de la tesis, respecto a las deficiencias que presentan las técnicas basadas en espacios euclidianos y por tanto, investigar métodos en otro tipo de Espacios métricos.

En conclusión, el método de componentes principales; se considera una técnica que permite realizar reconocimientos de rostros sobre todo, cuando se tienen diversas imágenes por persona. Sin embargo, a pesar de lo anterior, la confiabilidad no es completamente adecuada.

Se considera que al tener datos (los rostros), que no cumplen con el supuesto de normalidad multivariada, inciden en las diversas fallas y errores de tales tipos de sistemas.

Por tanto, se plantea la opción de trabajar en un espacio L_∞ , el cual no requiere los supuestos de normalidad, linealidad y homoscedasticidad. Este nuevo método para el reconocimiento de rostros, utilizará un enfoque de aproximación multivariada, utilizando el denominado Algoritmo Genético Ecléctico para obtener bajo un paradigma de optimización combinatoria, la *forma y orden* del polinomio de aproximación que caracterizará a los rostros.

Asimismo, dicho método requerirá únicamente, de una muestra de los píxeles que configura cada rostro, lo cual reducirá en una menor cantidad de almacenamiento y tiempo de procesamiento computacional. Además, al utilizar un enfoque holístico, no requiere de la utilización de características geométricas o “plantillas” del rostro.

Referencias Bibliográficas

Abdi, Hervé; Valentin, Dominique; Edelman, Betty. (1997). “Eigenfeatures as intermediate level representations: The case for PCA models”; *Brain and Behavioural Sciences*; v. 4; December

Abdi, Hervé; Valentin, Dominique; O’Toole, Alice (1997). “A generalized autoassociator model for face processing and sex categorization: From principal components to multivariate analysis”; en Levine, D. S.; Elsberry, W. R. (Eds.); *Optimality in biological and artificial networks*; Erlbaum; Mahwah, NJ; pp. 317-337

Aizawa, K. Et al. (1993). “Human Facial Motion Analysis and Synthesis with Application to Model-Based Coding”, en Sezan, M., Lagendijk, R., *Motion Analysis and Image sequence Processing*, Kluwer, Boston, MA, pp. 317-348

Akamatsu, S., Sasaki, T., Fukamachi, H., Suenaga, Y. (1991). “A Robust Face Identification Scheme – KL Expansion of an Invariant Feature Space”, *SPIE Proceedings* vol, 1607: *Intelligent Robots and Computer Vision X: Algorithms and Techniques*, pp. 71-84.

Atick, Joseph, Griffin, Paul A., Redlich, A. Norman (1996). “Statistical Approach to Shape from Shading: Reconstruction of Three-Dimensional face Surfaces from Single Two-Dimensional Images”, *Neural Computation*, v.8, 6, august, p. 1321-1340.

- Bala, J.; Huang, J.; Vafaie, H.; De Jong, K.; Wechsler, H.** (1995). “Hybrid Learning Using Genetic Algorithms and Decision Trees for Pattern Classification”; International Joint Conference on Artificial Intelligence; Montreal, Canada; August.
- Bala, J.; De Jong, K.; Huang, J.; Vafaie, H.; Wechsler, H.** (2003). “Using Learning to Facilitate the Evolution of Features for Recognizing Visual Concepts”; To Appear in the Special Issue of Evolutionary Computation
- Baron, R.** (1981). “Mechanisms of Human Facial Recognition”, International Journal of Man-Machine Studies, vol. 15, pp. 137-178.
- Batista, Joan Manuel, Martínez, Ma. Del Rosario** (1989). Análisis Multivariante: Análisis de componentes principales, Editorial Hispano Europeo, Barcelona.
- Bellman, R.** (1960). Introduction to Matrix Analysis, New York, McGraw-Hill.
- Benton, Arthur L.** (1980). “The Neuropsychology of Facial Recognition”; American Psychologist; vol. 35; No. 2; February; pp. 176-186.
- Bhanu, Bir; Lee, Sungkee; Ming, John** (1995). “Adaptive Image Segmentation Using a Genetic Algorithm”; IEEE Trans. on Systems, Man, and Cybernetics; vol. 25; No. 12; December; pp. 1543-1567.
- Beveridge, Gordon, Schechter, Robert** (1970). Optimization: Theory and Practice, McGraw-Hill, New York.
- Beymer, David; Poggio, Tomaso** (1995). “Face Recognition From One Example View”; MIT-AI Laboratory; AI Memo No. 1536; September.
- Bruce, Vicki; Young, Andy** (1986). “Understanding face recognition”; The British Journal of Psychology; vol. 77; pp. 305-327

- Bruce, Vicki (Ed.)** (1991). Face Recognition. A special issue of The European Journal of Cognitive Psychology, Lawrence Erlbaum, Hillsdale.
- Brunelli, Roberto; Poggio, Tomaso** (1995). "Automatic Person Recognition by Acoustic and Geometric Features", MVA, v. 8, 5, p. 317-325.
- Brunelli, Roberto; Poggio, Tomaso** (1993). "Face Recognition: Features versus Templates"; IEEE Trans. on Pattern Recognition and Machine Intelligence; v. 15; No. 10; October; pp. 1042-1052
- Brunelli, R., Poggio, T.** (1992). "HyperBF Networks for Gender Classification", Proceedings DARPA Image Understanding Workshop, pp. 311-314.
- Brunelli, Roberto; Poggio, Tomaso** (1990). "Caricatural Effects in Automated Face Perception", En Internet.
- Buhmann, J., Lades, M., Malsburg, C.** (1990). "Size and Distortion Invariant Object Recognition by Hierarchical Graph Matching", Proceedings International Joint Conference on Neural Networks, pp. 411-416.
- Burr, D.** (1981). "A Dynamic Model for Image Registration", Computer Graphics and Image Processing, vol. 15, pp. 102-112.
- Canny, J.** (1986). "A Computational Approach to Edge Detection", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 8, pp. 679-689.
- Chen, J. S.; Huertas, A.; Medioni, G.** (1987). "Fast Convolution with Laplacian-of-Gaussian Masks"; IEEE Trans. on Pattern Analysis and Machine Intelligence; vol. 9; No. 4; July; pp. 584-590.

Cheng, Fang-Hsuan; Hsu, Wen-Hsing; Chen, Mei-Ying (1989). "Recognition of Handwritten Chinese Characters by Modified Hough Transform Techniques"; IEEE Trans. on Pattern Analysis and Machine Intelligence; vol. 11; No. 4; April; pp. 429-439.

Cheng, Y., Liu, K., Yang, J., Wang, H. (1992). "A Robust Algebraic Method for Human face recognition", Proceedings 11th International Conference on Pattern Recognition, pp. 221-224.

Cheng, Y., Liu, K., Yang, J., Zhuang, Y., Gu, N. (1991). "Human Face recognition Method Based on the Statistical Model of Small Sample Size", SPIE proceedings, vol. 1607: Intelligent Robots and Computer Vision X:: Algorithms and Techniques, pp. 85-95.

Christensen, Ronald (1996). Analysis of Variance, Design and Regression, Chapman & Hall, London.

Cohen, Fernand S.; Wang, Jin-Yinn (1994). "Part 1: Modeling Image Curves Using Invariant 3-d Object Curve Models - A Path to 3-D Recognition and Shape Estimation from Image Contours"; IEEE Trans. on Pattern Analysis and Machine Intelligence; vol. 16; No. 1; January; pp. 1-12.

Conlin, M.J. (1986). "A Rule Based High Level Vision System", SPIE Proceedings, v. 726: Intelligent Robots and Computer Vision, p. 314-320.

Cottrell, G.W., Fleming, M. (1990). "Face recognition using unsupervised feature extraction", International Neural Network Conference, 1, pp. 322-325.

Cox, Ingemar J.; Ghosn, Joumana; Yianilos, Peter N. (1995). "Feature-Based Face Recognition Using Mixture-Distance"; NEC Research Institute; Technical Report 95-09; Princeton, NJ; October.

- Craw, I., Tock, D., Bennett, A.** (1992). "Finding Face Features", Proceedings 2nd European Conference on Computer Vision, p. 92-96.
- Craw, I., Cameron, P.** (1992). "Face recognition by computer", British Machine Vision Conference, pp. 488-507, Springer-Verlag.
- Davis, Lawrence (Ed.)** (1987). Genetic Algorithms and Simulated Annealing, Morgan Kaufmann, Los Altos, CA.
- Dean, Edwin B.** (1988). "Linear Least Squares for Correlated Data", Tenth Annual International Conference of the International Society of Parametric Analysis, England, July.
- Devillers, J. (Ed.)** (1996). Genetic Algorithms in Molecular Modeling. Principles of QSAR and Drug Design 1., Academic Press.
- Edelman, Shimon; Reisfeld, Daniel; Yeshurun, Yechezkel** (1991). "Learning to recognize faces from examples"; Technical Report; The Weizmann Institute of Science; Israel; October 15.
- Ekman, Paul; Huang, Thomas; Sejnowski, Terrence; Hager, Joseph (Eds.)** (1992). Final Report to NSF of the Planning Workshop on Facial Expression Understanding; July 30 to August 1.
- Ellis, Hadyn D.** (1975). "Recognizing Faces"; The British Journal of Psychology; vol. 66; No. 4; pp. 409-426
- Fukunaga, Keinosuke** (1989). Statistical Pattern Recognition, Academic Press, New York.
- Fukunaga, Keinosuke** (1972). Introduction to Statistical Pattern Recognition. Academic Press, New York

Galton, Francis (1892). *Finger Prints*, Macmillan. En Internet:
<http://www.mugu.com/galton/books/finger-prints>

Golomb, B.A., Sejnowski, T.J. (1991). "SEXNET: A Neural Network Identifies Sex From Human Faces", en Touretzky, D.S., Lipmann, R. (Eds.), *Advances in Neural Information Processing Systems 3*, Morgan Kaufmann, San Mateo, CA, p. 572-577.

Goldberg, David E. (1989). *Genetic Algorithms in Search, Optimization & Machine Learning*, Addison-Wesley, Reading, MA.

Goshtasby, A. (1985). "Description and Discrimination of Planar Shapes Using Shape Matrices", *IEEE Trans. Pattern Analysis and Machine Intelligence*, v. 7, p. 738-743.

Goudail, Francois; Lange, Eberhard; Iwamoto, Takashi; Kyuma, Kazuo; Otsu, Nobuyuki (1996). "Face Recognition System Using Local Autocorrelations and Multiscale Integration"; *IEEE Trans. on Pattern Analysis and Machine Intelligence*; vol. 18; No. 10; October; pp. 1024-1028

Govindaraju, V., Srihari, S.N., Sher, D.B. (1990). "A Computational Model for Face Location", *Proceedings Third International Conference on Computer Vision*, p. 718-721.

Gray, Stephen B. (1971). "Local Properties of Binary Images in Two Dimensions"; *IEEE Transactions on Computers*; vol. c-20; No. 5; May; pp. 551-561.

Grenander, U., Chow, Y., Keenan, D. (1991). *Hands: A Pattern Theoretic Study of Biological Shapes*, Springer-Verlang, New York.

Grötschel, Martin, Lovász, Laszló (1993). *Combinatorial Optimization: A Survey*, DIMACS Technical Report 93-29, Princeton University, May.

Gutta, S.; Huang, J.; Takacs, B.; Wechsler, H. (1996). "Face Recognition Using Ensembles of Networks"; International Conference on Pattern Recognition (August 25-30); Vienna, En: <http://chagall.gmu.edu/FORENSIC/pub.html>

Gutta, Srinivas; Wechsler, Harry (1996). "Face Recognition Using Hybrid Classifier Systems"; International Conference on Neural Networks; Washington, DC.

Gutta, Srinivas; Huang, Jeffrey; Singh, Dig; Wechsler, Harry (1995). "The FERET Facial Database"; National Institute of Standards and Technology Workshop; Maryland.

Gutta, Srinivas; Huang, Jeffrey; Singh, Dig; Shah, Imran (1995). "Benchmark Studies on Face Recognition"; Proceedings of International Workshop on Automatic Face and Gesture Recognition (June 26-28); Zurich.

Hair, J., Anderson, R. (2004). Multivariate Data Analysis, Prentice Hall, Saddle River, NJ.

Hallinan, P. (1991). "A low-dimensional representation of human faces for arbitrary lighting conditions", CVPR Proc., p. 945-949.

Hilera, José, Martínez, Victor (1995). Redes Neuronales Artificiales: Fundamentos, modelos y aplicaciones, Addison-Wesley/ Ra-Ma, Madrid.

Hancock, Peter, Bruce, V., Burton, A.M. (1998). "Testing Principal Component Representation for Faces", Technical Report, University of Stirling, UK. En Internet.

Hancock, Peter J. B.; Burton, A. Mike; Bruce, Vicki (1996). "Face processing: human perception and principal component analysis"; Memory and Cognition; vol. 24; No. 1; pp 26-40, En: <http://nevis.stir.ac.uk/pjh>

Hancock, Peter J. B.; Baddeley, Roland J.; Smith, Leslie (1991). "The principal components of natural images".

Haralick, Robert; Sternberg, Stanley; Zhuang, Xinhua (1987). "Image Analysis Using Mathematical Morphology"; IEEE Trans. on Pattern Analysis and Machine Intelligence; vol. 9; No. 4; July; pp. 532-550.

Harmon, L., Khan, M., Lasch, R., Ramig, P. (1981). "Machine Identification of Human Faces", Pattern Recognition, v. 13, p. 97-110.

Harmon, L., Hunt, W. (1977). "Automatic Recognition of Human Face Profiles", Computer Graphics and Image Processing, v. 6, p. 135-156.

Harmon, Leon D. (1973). "The Recognition of Faces"; Scientific American; vol. 229; No. 5; November; pp. 70-82

Heo, J., Abidi, B., Paik, J. (2003). "Face Recognition: Evaluation Report For FaceIt®", Proc. Of SPIE 6th International Conference on Quality Control by Artificial Vision QCAV03, Gatlinburg, TN, USA, May .

Holland, John H. (1995). Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence, The MIT Press, Cambridge, MA.

Hong, Z. (1991). "Algebraic Feature of Image for Recognition", Pattern Recognition, v. 24, p. 211-219,

Hotteling, H. (1933). Análisis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24, 417-441; 498-520.

Huang, Ren-Jay (1998). Detection Strategies for Face Recognition using Learning and Evolution, Ph.D. Dissertation, George Mason University, Abstract. En Internet.

Intrator, Nathan; Reisfeld, Daniel; Yeshurun, Yehezkel (1995). "Face Recognition using a Hybrid Supervised/Unsupervised Neural Network"; Technical Report; Department of Computer Science; Tel-Aviv University; June.

Jain, A.K. (1989). *Fundamentals of Digital Image Processing*, Prentice-Hall, Englewood Cliffs, NJ.

Jia, Xiaoguang; Nixon, Mark S. (1995). "Extending the Feature Vector for Automatic Face Recognition"; *IEEE Trans. on Pattern Analysis and Machine Intelligence*; vol. 17; No. 12; December; pp. 1167-1176.

Jonson, J. D. (1991). *Applied Multivariate Data Analysis. Volume I: Regression and Experimental Design*, Springer-Verlag, New York.

Kaiser, H.F. (1958). The Varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187-200.

Kanade, T. (1977). "Computer recognition of human faces", en *Interdisciplinary Systems Research*, Birkhäuser Verlag.

Katz, A. J.; Thrift, P. R. (1994). "Generating Image Filters for Target Recognition by Genetic Learning"; *IEEE Trans. on Pattern Analysis and Machine Intelligence*; vol. 16; No. 9; september; pp. 906-910

Kaufman, G.J., Breeding, K.J. (1976). "The Automatic Recognition of Human Faces from Profile Silhouettes", *IEEE Trans. Systems, Man, and Cybernetics*, v. 6, p. 113-121.

Kaya, Y., Kobayashi, K. (1972). "A basic study of human face recognition", en Watanabe, S. (Ed.), *Frontiers of Pattern Recognition*, Academic Press, New York, p. 265-289.

Keren, Daniel; Cooper, David; Subrahmonia, Jayashree (1994). "Describing Complicated Objects by Implicit Polynomials"; *IEEE Trans. on Pattern Analysis and Machine Intelligence*; vol. 16; No. 1; january; pp. 38-53.

Kirby, M.; Sirovich, L (1990). “Application of the Karhunen-Loève Procedure for the Characterization of Human Faces”; IEEE Trans. on Pattern Analysis and Machine Intelligence; vol. 12; No. 1; January; pp. 103-108.

Kohonen, T. (1988). Self-Organization and Associative Memory, Springer-Verlag, Berlin.

Kuri, Angel (1998). Prediction of Dynamic Systems through the Identification of Basic Patterns in Multivariate Mathematical Landscapes Using a Genetic Algorithm”, CIC-IPN, México.

Kuri, Angel (1998). A Universal Eclectic Genetic Algorithm, Reporte, Centro de Investigación en Computación- IPN.

Kuri, Angel (1998). Un método de aprendizaje no supervisado usando algoritmos genéticos de orden, CIC-IPN.

Kuri, Angel (1997). “Pattern Recognition via a Genetic Algorithm”, en Guzmán, A., Shulcloper, J.R., Sossa, J.H., et al. (Comp.), II Taller Iberoamericano de Reconocimiento de Patrones-La Habana, Cuba, ICIMAF-CIC-IPN, pp. 345-356.

Kuri, Angel y Villegas, Carlos (1998) “A Universal Eclectic Genetic Algorithm for Constrained Optimization”, 6th European Congress on Intelligent techniques & Soft Computing; Aachen, Germany; September.

Lades, M., Vorbruggen, J., Buhmann, J., Lange, J., et. al. (1993). « Distortion Invariant Object Recognition in the Dynamic Link Architecture », IEEE Trans. Computers, v. 42, p. 300-311.

Lawrence, Steve; Giles, C. Lee; Tsoi, Ah Chung; Back, Andrew (1996). “Face Recognition: A Hybrid Neural Network Approach”; Technical Report (UMIACS-TR-96-16); Institute for Advanced Computer Studies; University of Maryland; August.

Liu, Chengjun, Wechsler, H. (1998). "Face Recognition Using Evolutionary Pursuit", Fifth European Conference on Computer Vision, University of Freiburg, Germany, En Internet.

Lucas, S.M. (1995). "Face Recognition with the continuous n.tuple classifier", En Internet.

Lucas, S.M. (1994). "The continuous n-tuple classifier and its application to face recognition", En Internet: <http://esewww.essex.ac.uk/sml/papers.html#face>

Manjunath, B.S., Chellappa, R., Malsburg, C. (1992). "A Feature Based Approach to Face Recognition", Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, p. 373-378.

Mitchell, T.M. (1980). "The Need for Biases in Learning Generalizations", Technical Report No. CBM-TR117, Rutgers University.

Moghaddam, Baback; Wahid, Wasiuddin; Pentland, Alex (1997). "Beyond Eigenfaces: Probabilistic Matching for Face Recognition"; Technical Report No. 43; MIT Media Laboratory Perceptual Computing Section.

Moghaddam, Baback; Nastar, Chahab; Pentland, Alex (1996). "Bayesian Face Recognition using Deformable Intensity Surfaces"; Technical Report No. 371; MIT Media Laboratory Perceptual Computing Section.

Nakamura, O., Mathur, S., Minami, T. (1991). "Identification of human faces based on isodensity maps", Pattern Recognition, 24 (3), pp. 263-272.

Nastar, Chahab; Mitschke, Matthias (1998). "Real-Time Face Recognition Using Feature Combination"; Proc. of the Third IEEE International Conference on Automatic Face and Gesture Recognition ; Nara, Japan; April.

Niemann, H.; Fischer, V.; Paulus, D.; Fischer, J. (1996). “Knowledge Based Image Understanding by Iterative Optimization”; Proceedings KI; Dresden, Germany.

Nixon, M. (1985). “Eye Spacing Measurement for Facial Recognition”, SPIE Proc., v. 575, p. 279-285.

O’Toole, Alice J.; Deffenbacher, Kenneth A.; Valentin, Dominique (1997). “The Perception of Face Gender: The Role of Stimulus Structure in Recognition and Classification”; Memory and Cognition; vol. 25.

O’Toole, Alice J.; Abdi, Hervé; Deffenbacher, Kenneth A.; Valentin, Dominique (1995). “ A Perceptual Learning Theory of the Information in Faces”; en Valentine, T. (De.); Cognitive and Computational Aspects of Face Recognition ; Routledge; London; pp. 159-182.

O’Toole, Alice; Abdi, Hervé; Deffenbacher, Kenneth; Valentin, Dominique (1993). “ Low-dimensional representation of faces in higher dimensions of the face space”; J. Opt. Soc. Am. A.; vol. 10; No. 3; March; pp. 405-415

Pal, Sankar K., Wang, Paul P. (1996). Genetic Algorithms for Pattern Recognition, CRC Press, Boca Raton.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. Philophical Magazine, ser 6, 2, 559-572.

Penev, Penio S.; Atick, Joseph J. (1996). “Local Feature Analysis: A general statistical theory for object representation”; Technical Report; Computational Neuroscience Laboratory; The Rockefeller University; New York; En: <http://venezia.rockefeller.edu/group/papers/full/LFA/PenevPS.LFA.ps>

Pentland, A., Moghaddam, B., Starner, T., Turk, M. (1994). “View-based and Modular Eigenspaces for Face Recognition”, Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Pérez, C. (2004). Técnicas estadísticas con SPSS, Prentice-Hall, Madrid.

Pinto-Eliás, R., Sossa-Azuela, J.H. (1998). “Human Face Identification Using Invariant Descriptions and Genetic Algorithm”, en Coelho, H. (Ed.), Progress in Artificial Intelligence-IBERAMIA 98 (6th. Ibero-American Conference on AI, Lisbon, Portugal), Springer-Verlag, Lecture Notes in AI-No. 1484, Germany, p. 293-302.

Poggio, T., Girosi, F. (1990). “Networks for Approximation and Learning”, Proc. IEEE, v. 78, p. 1481-1497.

Quinlan, J.R. (1993). C4.5 Programs for Machine Learning, Morgan Kaufmann, San Mateo, CA.

Rahardja, A., Sowmya, A., Wilson, W. (1991). “A Neural Network Approach to Component Versus Holistic Recognition of Facial Expressions in Images”, SPIE Proc., v. 1607: Intelligent Robots and Computer Vision X-Algorithms and Techniques, p. 62-70.

Rao, Rajesh P. N.; Ballard, Dana (1995). “Natural Basis Functions and Topographic Memory for Face Recognition”; Proc. International Joint Conf. on Artificial Intelligence; pp. 10-17.

Reisfeld, D., Yeshurun, Y. (1992). “Robust detection of Facial Features by Generalized Symmetry”, Proc. 11 th. International Conference on Pattern Recognition, p. 117-120.

Ripley, B.D. (1997). Pattern Recognition and Neural Networks, Cambridge University Press, U.K.

Rosenfeld, Azriel; Thurston, Mark (1971). "Edge and Curve Detection for Visual Scene Analysis"; IEEE Trans. on Computers; vol. C-20; No. 5; pp. 562-569

Roth, Gerhard; Levine, Martin D. (1994). "Geometric Primitive Extraction Using a Genetic Algorithm"; IEEE Trans. on Pattern Analysis and Machine Intelligence; vol. 16; No. 9; september; pp. 901-905.

Samaria, F. (1994). Face Recognition Using Hidden Markov Models, Ph. D., Trinity College, Cambridge University.

Schalkoff, Robert J. (1992). Pattern Recognition: Statistical, Structural and Neural Approaches, Wiley, New York.

Seibert, M., Waxman, A. (1991). "Recognizing Faces from their Parts", SPIE Proc., v. 1611: Sensor Fusion IV-Control Paradigms and Data Structures, p. 129-140.

Sergent, J. (1986). "Microgenesis of Face Perception", en Ellis, H.D., Jeeves, M., Newcombe, F., Young, A. (Eds.), Aspects of Face Processing, Dordrecht, Nijhoff.

Shavlik, Jude W., Dietterich, Thomas G. (Eds.) (1990). Readings in Machine Learning, Morgan Kaufmann, San Mateo, CA.

Shepherd, J.W. (1986). "An Interactive Computer Systems fro Retrieving Faces", en Ellis, H.D., Jeeves, M., Newcombe, F., Young, A. (Eds.), Aspects of Face Processing, Dordrecht, Nijhoff, p. 398-409.

Simon, Herbert A. (1983). "Why Should Machines Learn?", en Michalski, R.S., Carbonell, J.G. (Eds.), Machine Learning: An Artificial Intelligence Approach, v. I, Morgan Kaufmann, San Mateo, CA.

Sirovich, L., Kirby, M. (1987). "Low-dimensional Procedure for the Characterization of Human Face", Journal of the Optical Society of America, v. 4, p. 519-524.

- Sklansky, Jack** (1978). "Image Segmentation and Feature Extraction"; IEEE Trans. on Systems, Man, and Cybernetics; vol. SMC-8; No. 4; april; pp. 237-247
- Stonham, T.J.** (1986). "Practical Face Recognition and Verification with WISARD", en Ellis, H.D., Jeeves, M., Newcombe, F., Young, A. (Eds.), Aspects of Face Processing, Dordrecht, Nijhoff.
- Sung, Kah-Kay; Poggio, Tomaso** (1994). "Example-based Learning for View-based Human Face Detection"; MIT-Artificial Intelligence Laboratory/ Center for Biological and Computational Learning; AI Memo No. 1521; December.
- Sze, T. W.; Yang, Y. H.** (1981). "A Simple Contour Matching Algorithm"; IEEE Trans on Pattern Analysis and Machine Intelligence; vol. 3; No. 6; november; pp. 676-678
- Szeliski, Richard** (1990). "Fast Surface Interpolation Using Hierarchical Basis Functions"; IEEE Trans. on Pattern Analysis and Machine Intelligence; vol 12; No. 6; june; pp. 513-528.
- Takács, Barnabás; Wechsler, Harry** (1995). "Face Location Using A Dynamic Model of Retinal Feature Extraction"; International Workshop on Automatic Face and Gesture Recognition; Zurich.
- Takács, Barnabás; Wechsler, Harry** (1994). "Locating Facial Features Using SOFM"; International Conference on Pattern Recognition; Jerusalem, Israel.
- Tamburelli, Giovanni** (1981). "Some Results in the Processing of the Holy Shroud of Turin "; IEEE Trans. on Pattern Analysis and Machine Intelligence; vol. 3; No. 6; november; pp. 670-675

Taubin, Gabriel; Cukierman, Fernando; Sullivan, Steven; Ponce, Jean; Kriegman, David (1994). "Parameterized Families of Polynomials for Bounded Algebraic Curve and Surface Fitting"; IEEE Trans. on Pattern Analysis and Machine Intelligence; vol. 16; NO. 3; march, pp. 287-303.

Thurstone, L.L. (1935). The Vectors of the Mind. Chicago. University of Chicago Press.

Thurstone, L.L. (1947). Multiple Factor Analysis. Chicago. University of Chicago Press.

Troje, Nikolaus F.; Vetter, Thomas (1996). "Representations of Human Faces"; Technical Report No. 41; Max-Planck-Institut für biologische Kybernetik; October 16; Germany.

Troje, Nikolaus; Vetter, Thomas (1996). "Pixel-based versus correspondence-based representations of human faces: Implications for sex discrimination"; ECVF meeting; Strasbourg, En: <http://www.mpik-tueb.mpg.de/people/personal/niko/ECVP96/ECVP96.html>

Turk, M., Pentland, A. (1991). "Face Recognition Using Eigenfaces", Proc. IEEE Conference on Computer Vision and Pattern Recognition, p. 586-591.

Turk, M., Pentland, A. (1991). "Eigenfaces for recognition", Journal of Cognitive Neuroscience, 3 (1), pp. 71-86.

Valentin, Dominique; Abdi, Hervé; Edelman, Betty (1997). "What Represents a Face: A Computational Approach for the Integration of Physiological and Psychological Data"; Perception; vol. 26.

Valentin, Dominique; Abdi, Herve; O'Toole, Alice J.; Cottrell, Garrison W. (1994). "Connectionist Models of Face Processing: A Survey"; Pattern Recognition; vol. 27; pp. 1209-1230

Villegas-Quezada, Carlos (1999). Reconocimiento holístico de rostros utilizando aproximación multivariada y algoritmos genéticos de orden, Reporte de Investigación, Laboratorio de Inteligencia Artificial, IPN-Centro de Investigación en Computación, diciembre, México.

Wang, Jin-Yinn; Cohen, Fernand S. (1994). "Part II: 3-D Object Recognition and Shape Estimation from Image Contours Using B-Splines, Shape Invariant Matching, and Neural Network"; IEEE Trans. on Pattern Analysis and Machine Intelligence; vol. 16; No. 1; January, pp. 13-23

Wang, Shyuan; Wu, Angela Y; Rosenfeld, Azriel (1981). "Image Approximation from Gray Scale "medial Axes"; IEEE Trans on Pattern Analysis and Machine Intelligence; vol. 3; No. 6; vovember.

Wechsler, H., Phillips, P.J., Bruce, V. (Eds.) (1998). Face Recognition from Theory to Applications, Springer/NATO, Germany.

Weng, J.J., Ahuja, N., Huang, T.S. (1993). "Learning Recognition and Segmentation of 3D Objects from 2D Images", Proc. International Conference on Computer Vision, p. 121-128.

Wilde, D. (1964). Optimum Seeking Methods, Prentice-Hall, Englewood Cliffs, NJ.

Wilde, D., Beightler, C. (1976). Teoría de Optimización, Ediciones URMO, Bilbao.

Wiskott, Laurenz (1997). "Phantom Faces for Face Analysis"; Pattern Recognition, vol. 30; No. 6.

Wiskott, Laurenz; Fellous, Jean-Marc; Kruger, Norbert; von der Malsburg, Christoph (1996). "Face Recognition by Elastic Bunch Graph Matching"; Internal Report 96-08; Institut fur Neuroinformatik; Ruhr-Universitat Bochum.

Wiskott, Laurenz; von der Malsburg, Christoph (1996). "Face Recognition by Dynamic Link Matching"; Internal Report 96-05; Institut für Neuroinformatik; Ruhr-Universität Bochum; Bochum.

Wu, C. J., Huang, J. S. (1990). "Human face profile recognition by computer", Pattern Recognition, 23 (3 / 4), pp. 255-259.

Würtz, Rolf (1995). "Background Invariant Face Recognition"; en Kappen, Bert; Gielen, Stan (Eds.); Neural Networks: Artificial Intelligence and Industrial Applications; Springer-Verlag, pp. 140-143

Yang, G., Huang, T.S. (1993). "Human Face Detection in a Scene", Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, p. 453-458.

Yuille, A., Cohen, D., Hallinan, P. (1989). "Feature Extraction from Faces Using Deformable Templates", Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, p. 104-109.

Anexos

Anexo A

Base de Rostros de AT&T

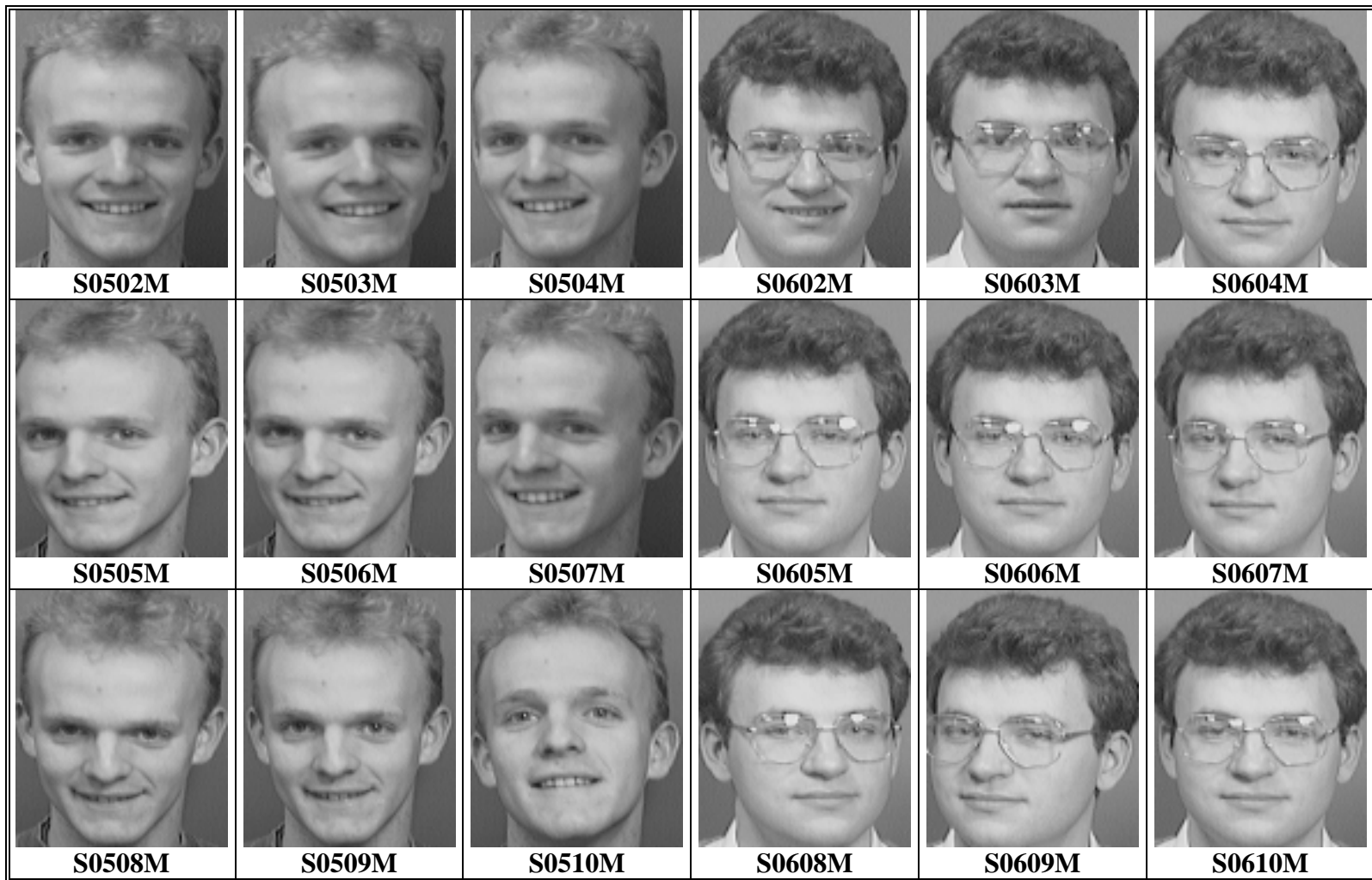
FOTOGRAFIAS DE LOS SUJETOS S01 Y S02



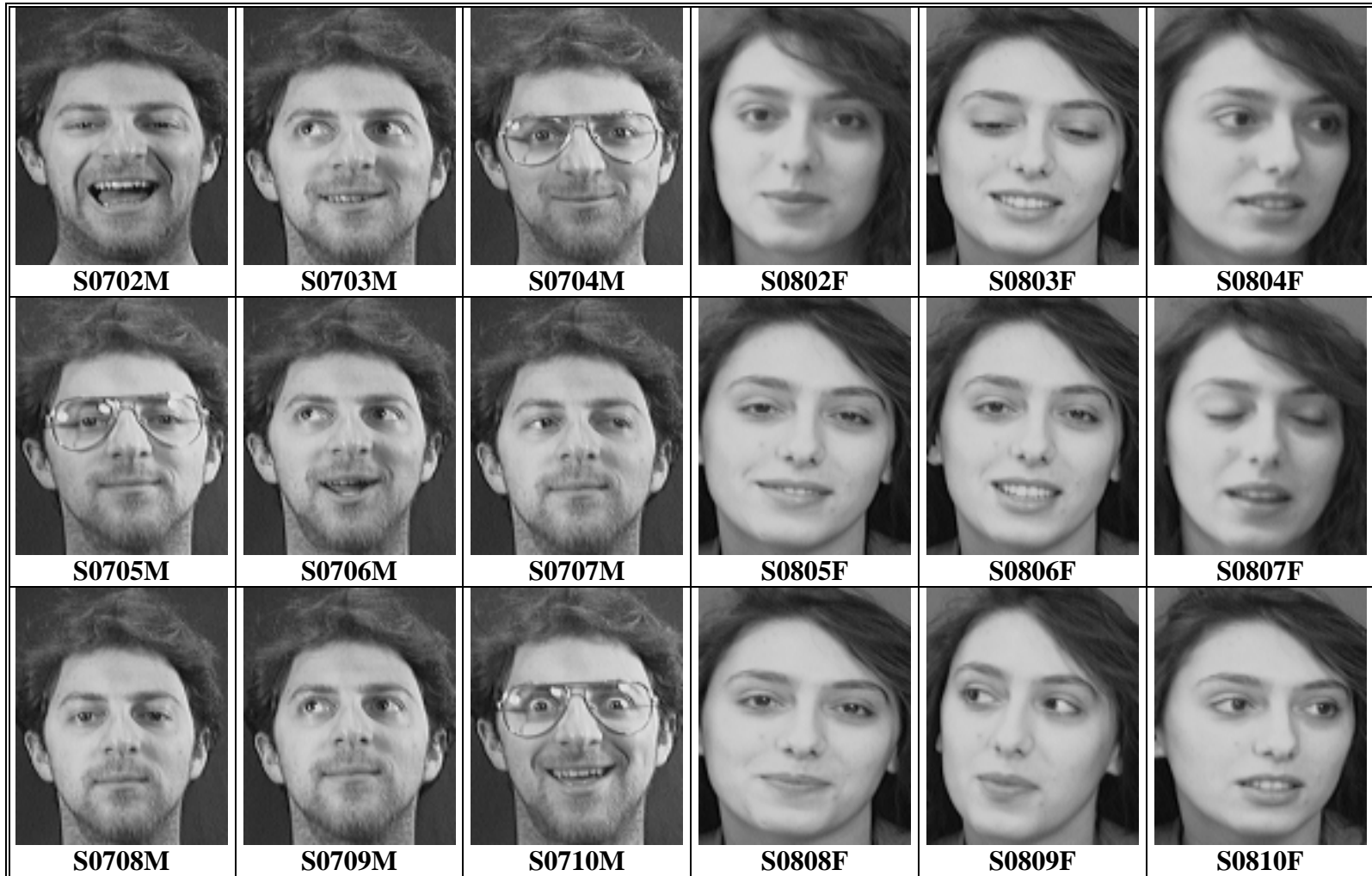
FOTOGRAFIAS DE LOS SUJETO S03 Y S04



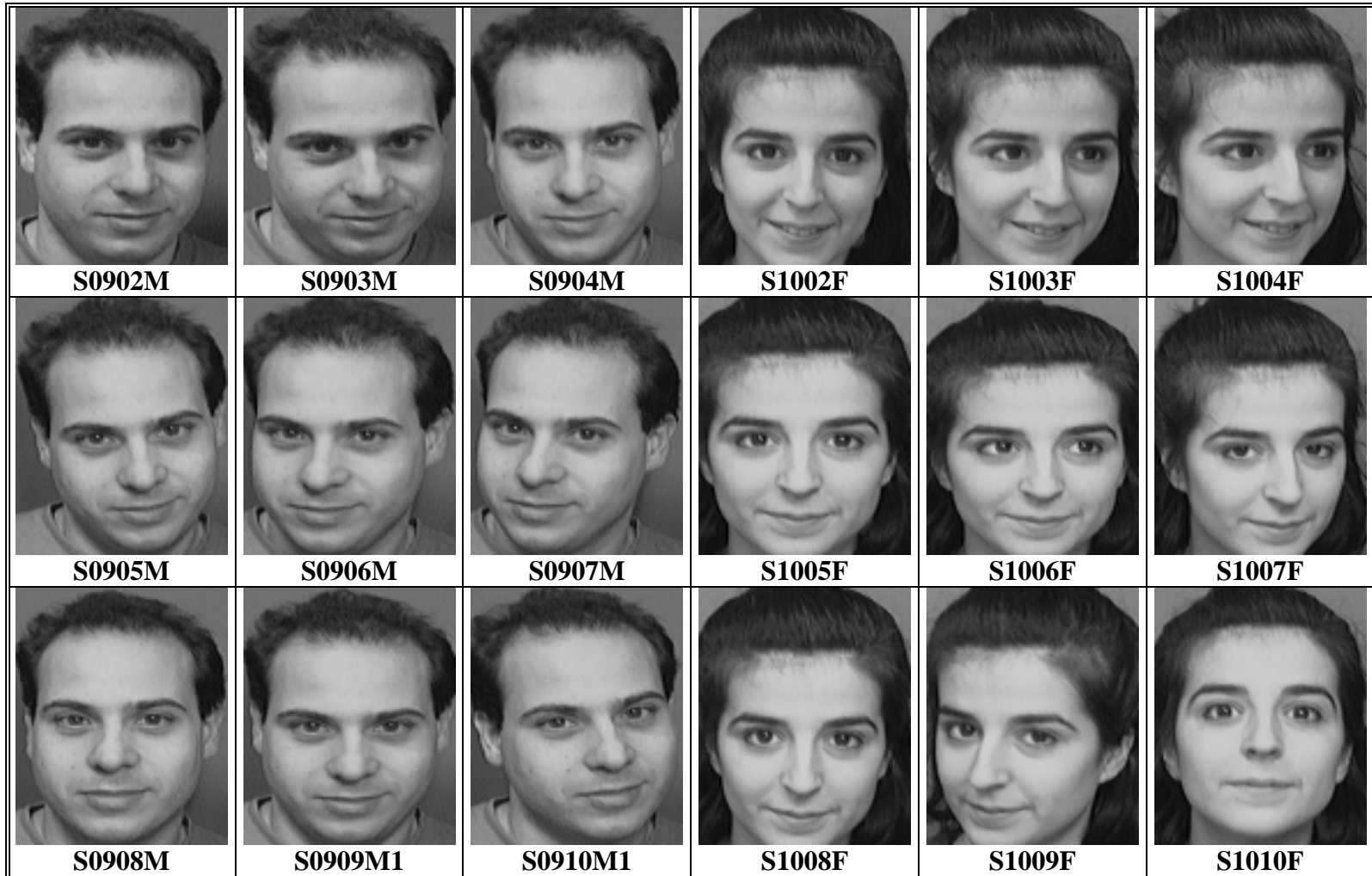
FOTOGRAFIAS DE LOS SUJETOS S05 Y S06



FOTOGRAFIAS DE LOS SUJETOS S07 Y S08



FOTOGRAFIAS DE LOS SUJETOS S09 Y S10



FOTOGRAFIAS DE LOS SUJETOS S11 Y S12



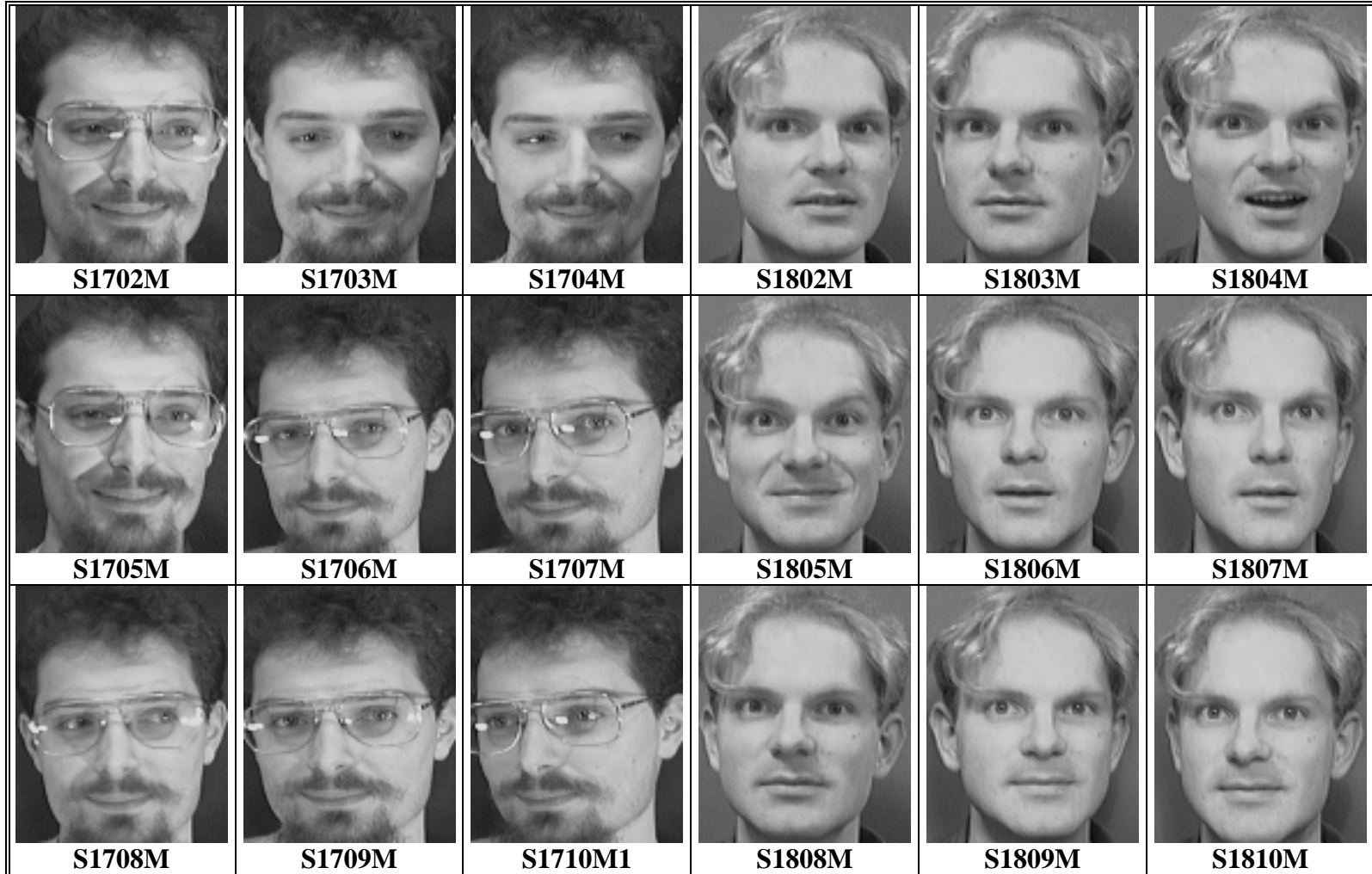
FOTOGRAFIAS DE LOS SUJETOS S13 Y S14



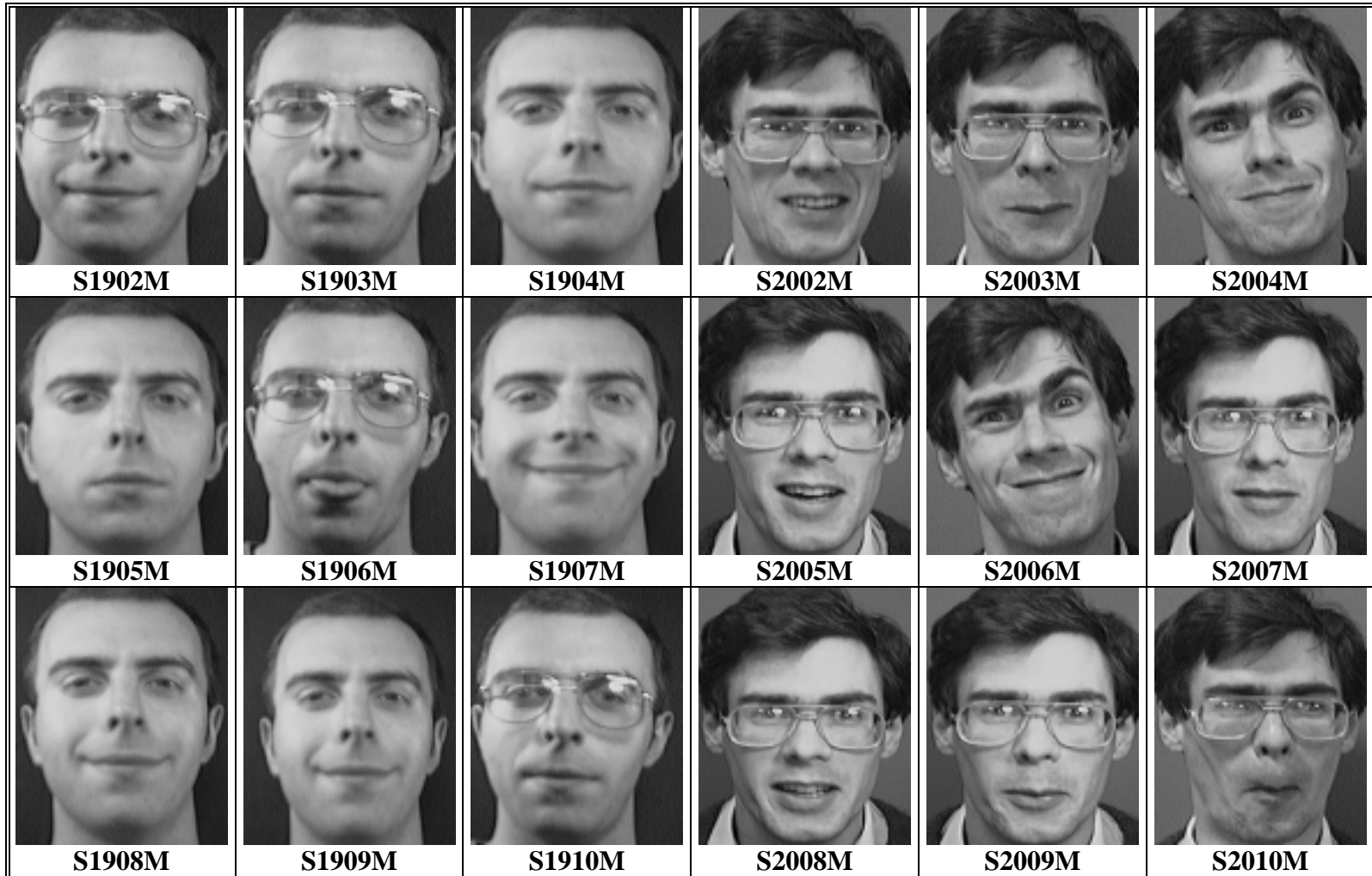
FOTOGRAFIAS DE LOS SUJETOS S15 Y S16



FOTOGRAFIAS DE LOS SUJETOS S17 Y S18



FOTOGRAFIAS DE LOS SUJETOS S19 Y S20



FOTOGRAFIAS DE LOS SUJETOS S21 Y S22



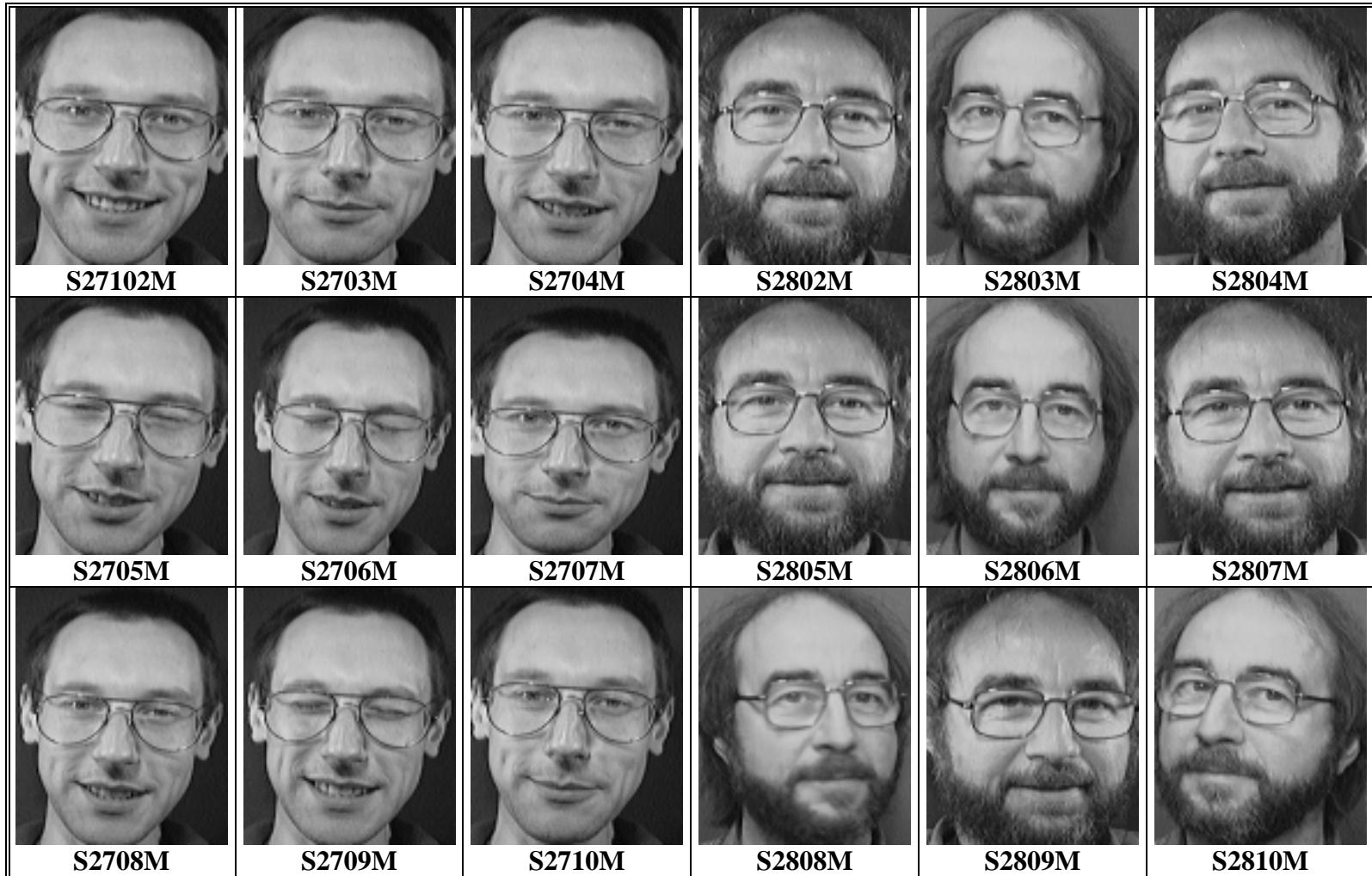
FOTOGRAFIAS DE LOS SUJETOS S23 Y S24



FOTOGRAFIAS DE LOS SUJETOS S25 Y S26



FOTOGRAFIAS DE LOS SUJETOS S27 Y S28



FOTOGRAFIAS DE LOS SUJETOS S29 Y S30



FOTOGRAFIAS DE LOS SUJETOS S31 Y S32



FOTOGRAFIAS DE LOS SUJETOS S33 Y S34



FOTOGRAFIAS DE LOS SUJETOS S35 Y S36



FOTOGRAFIAS DE LOS SUJETOS S37 Y S38



FOTOGRAFIAS DE LOS SUJETOS S39 Y S40



Anexo B

Prueba de Normalidad Univariada para los rostros de la base AT&T

Tests of Normality

	Kolmogorov-Smirnov ^a		
	Statistic	df	Sig.
FOTO0101	.182	10304	.000
FOTO0102	.209	10304	.000
FOTO0103	.208	10304	.000
FOTO0104	.142	10304	.000
FOTO0105	.136	10304	.000
FOTO0106	.210	10304	.000
FOTO0107	.193	10304	.000
FOTO0108	.128	10304	.000
FOTO0109	.145	10304	.000
FOTO0110	.130	10304	.000
FOTO0201	.138	10304	.000
FOTO0202	.132	10304	.000
FOTO0203	.121	10304	.000
FOTO0204	.135	10304	.000
FOTO0205	.140	10304	.000
FOTO0206	.129	10304	.000
FOTO0207	.116	10304	.000
FOTO0208	.116	10304	.000
FOTO0209	.121	10304	.000
FOTO0210	.116	10304	.000
FOTO0301	.065	10304	.000
FOTO0302	.056	10304	.000
FOTO0303	.049	10304	.000
FOTO0304	.056	10304	.000
FOTO0305	.052	10304	.000
FOTO0306	.058	10304	.000
FOTO0307	.055	10304	.000
FOTO0308	.058	10304	.000
FOTO0309	.055	10304	.000
FOTO0310	.057	10304	.000
FOTO0401	.061	10304	.000
FOTO0402	.063	10304	.000
FOTO0403	.062	10304	.000
FOTO0404	.065	10304	.000
FOTO0405	.066	10304	.000
FOTO0406	.070	10304	.000
FOTO0407	.069	10304	.000
FOTO0408	.057	10304	.000
FOTO0409	.059	10304	.000
FOTO0410	.075	10304	.000
FOTO0501	.037	10304	.000
FOTO0502	.042	10304	.000
FOTO0503	.038	10304	.000
FOTO0504	.035	10304	.000
FOTO0505	.055	10304	.000
FOTO0506	.052	10304	.000
FOTO0507	.034	10304	.000
FOTO0508	.051	10304	.000
FOTO0509	.057	10304	.000
FOTO0510	.074	10304	.000

a. Lilliefors Significance Correction

Tests of Normality

	Kolmogorov-Smirnov ^a		
	Statistic	df	Sig.
FOTO0601	.066	10304	.000
FOTO0602	.064	10304	.000
FOTO0603	.062	10304	.000
FOTO0604	.065	10304	.000
FOTO0605	.057	10304	.000
FOTO0606	.066	10304	.000
FOTO0607	.058	10304	.000
FOTO0608	.060	10304	.000
FOTO0609	.060	10304	.000
FOTO0610	.066	10304	.000
FOTO0701	.123	10304	.000
FOTO0702	.114	10304	.000
FOTO0703	.114	10304	.000
FOTO0704	.106	10304	.000
FOTO0705	.099	10304	.000
FOTO0706	.125	10304	.000
FOTO0707	.126	10304	.000
FOTO0708	.145	10304	.000
FOTO0709	.126	10304	.000
FOTO0710	.092	10304	.000
FOTO0801	.108	10304	.000
FOTO0802	.115	10304	.000
FOTO0803	.111	10304	.000
FOTO0804	.100	10304	.000
FOTO0805	.114	10304	.000
FOTO0806	.111	10304	.000
FOTO0807	.101	10304	.000
FOTO0808	.111	10304	.000
FOTO0809	.112	10304	.000
FOTO0810	.113	10304	.000
FOTO0901	.050	10304	.000
FOTO0902	.059	10304	.000
FOTO0903	.053	10304	.000
FOTO0904	.046	10304	.000
FOTO0905	.049	10304	.000
FOTO0906	.048	10304	.000
FOTO0907	.051	10304	.000
FOTO0908	.052	10304	.000
FOTO0909	.048	10304	.000
FOTO0910	.051	10304	.000
FOTO1001	.125	10304	.000
FOTO1002	.124	10304	.000
FOTO1003	.117	10304	.000
FOTO1004	.102	10304	.000
FOTO1005	.124	10304	.000
FOTO1006	.128	10304	.000
FOTO1007	.135	10304	.000
FOTO1008	.138	10304	.000
FOTO1009	.133	10304	.000
FOTO1010	.125	10304	.000

a. Lilliefors Significance Correction

Tests of Normality

	Kolmogorov-Smirnov ^a		
	Statistic	df	Sig.
FOTO1101	.165	10304	.000
FOTO1102	.148	10304	.000
FOTO1103	.145	10304	.000
FOTO1104	.133	10304	.000
FOTO1105	.136	10304	.000
FOTO1106	.165	10304	.000
FOTO1107	.163	10304	.000
FOTO1108	.151	10304	.000
FOTO1109	.143	10304	.000
FOTO1110	.172	10304	.000
FOTO1201	.074	10304	.000
FOTO1202	.057	10304	.000
FOTO1203	.056	10304	.000
FOTO1204	.047	10304	.000
FOTO1205	.051	10304	.000
FOTO1206	.071	10304	.000
FOTO1207	.055	10304	.000
FOTO1208	.072	10304	.000
FOTO1209	.069	10304	.000
FOTO1210	.065	10304	.000
FOTO1301	.053	10304	.000
FOTO1302	.054	10304	.000
FOTO1303	.058	10304	.000
FOTO1304	.063	10304	.000
FOTO1305	.048	10304	.000
FOTO1306	.057	10304	.000
FOTO1307	.057	10304	.000
FOTO1308	.055	10304	.000
FOTO1309	.049	10304	.000
FOTO1310	.046	10304	.000
FOTO1401	.097	10304	.000
FOTO1402	.102	10304	.000
FOTO1403	.095	10304	.000
FOTO1404	.093	10304	.000
FOTO1405	.093	10304	.000
FOTO1406	.113	10304	.000
FOTO1407	.112	10304	.000
FOTO1408	.108	10304	.000
FOTO1409	.150	10304	.000
FOTO1410	.151	10304	.000
FOTO1501	.085	10304	.000
FOTO1502	.051	10304	.000
FOTO1503	.078	10304	.000
FOTO1504	.088	10304	.000
FOTO1505	.042	10304	.000
FOTO1506	.085	10304	.000
FOTO1507	.049	10304	.000
FOTO1508	.077	10304	.000
FOTO1509	.056	10304	.000
FOTO1510	.051	10304	.000

a. Lilliefors Significance Correction

Tests of Normality

	Kolmogorov-Smirnov ^a		
	Statistic	df	Sig.
FOTO1601	.129	10304	.000
FOTO1602	.146	10304	.000
FOTO1603	.140	10304	.000
FOTO1604	.111	10304	.000
FOTO1605	.107	10304	.000
FOTO1606	.136	10304	.000
FOTO1607	.166	10304	.000
FOTO1608	.154	10304	.000
FOTO1609	.153	10304	.000
FOTO1610	.153	10304	.000
FOTO1701	.083	10304	.000
FOTO1702	.085	10304	.000
FOTO1703	.112	10304	.000
FOTO1704	.121	10304	.000
FOTO1705	.086	10304	.000
FOTO1706	.155	10304	.000
FOTO1707	.126	10304	.000
FOTO1708	.145	10304	.000
FOTO1709	.125	10304	.000
FOTO1710	.133	10304	.000
FOTO1801	.076	10304	.000
FOTO1802	.052	10304	.000
FOTO1803	.052	10304	.000
FOTO1804	.041	10304	.000
FOTO1805	.053	10304	.000
FOTO1806	.060	10304	.000
FOTO1807	.073	10304	.000
FOTO1808	.051	10304	.000
FOTO1809	.071	10304	.000
FOTO1810	.069	10304	.000
FOTO1901	.096	10304	.000
FOTO1902	.091	10304	.000
FOTO1903	.083	10304	.000
FOTO1904	.084	10304	.000
FOTO1905	.087	10304	.000
FOTO1906	.115	10304	.000
FOTO1907	.097	10304	.000
FOTO1908	.104	10304	.000
FOTO1909	.147	10304	.000
FOTO1910	.117	10304	.000
FOTO2001	.092	10304	.000
FOTO2002	.087	10304	.000
FOTO2003	.087	10304	.000
FOTO2004	.073	10304	.000
FOTO2005	.074	10304	.000
FOTO2006	.075	10304	.000
FOTO2007	.072	10304	.000
FOTO2008	.082	10304	.000
FOTO2009	.078	10304	.000
FOTO2010	.082	10304	.000

a. Lilliefors Significance Correction

Tests of Normality

	Kolmogorov-Smirnov ^a		
	Statistic	df	Sig.
FOTO2101	.039	10304	.000
FOTO2102	.040	10304	.000
FOTO2103	.059	10304	.000
FOTO2104	.067	10304	.000
FOTO2105	.073	10304	.000
FOTO2106	.078	10304	.000
FOTO2107	.079	10304	.000
FOTO2108	.075	10304	.000
FOTO2109	.069	10304	.000
FOTO2110	.074	10304	.000
FOTO2201	.064	10304	.000
FOTO2202	.049	10304	.000
FOTO2203	.068	10304	.000
FOTO2204	.070	10304	.000
FOTO2205	.044	10304	.000
FOTO2206	.090	10304	.000
FOTO2207	.084	10304	.000
FOTO2208	.077	10304	.000
FOTO2209	.067	10304	.000
FOTO2210	.064	10304	.000
FOTO2301	.100	10304	.000
FOTO2302	.106	10304	.000
FOTO2303	.054	10304	.000
FOTO2304	.064	10304	.000
FOTO2305	.057	10304	.000
FOTO2306	.062	10304	.000
FOTO2307	.060	10304	.000
FOTO2308	.097	10304	.000
FOTO2309	.110	10304	.000
FOTO2310	.106	10304	.000
FOTO2401	.143	10304	.000
FOTO2402	.144	10304	.000
FOTO2403	.107	10304	.000
FOTO2404	.087	10304	.000
FOTO2405	.115	10304	.000
FOTO2406	.136	10304	.000
FOTO2407	.140	10304	.000
FOTO2408	.136	10304	.000
FOTO2409	.096	10304	.000
FOTO2410	.101	10304	.000
FOTO2501	.067	10304	.000
FOTO2502	.062	10304	.000
FOTO2503	.068	10304	.000
FOTO2504	.091	10304	.000
FOTO2505	.077	10304	.000
FOTO2506	.074	10304	.000
FOTO2507	.069	10304	.000
FOTO2508	.069	10304	.000
FOTO2509	.065	10304	.000
FOTO2510	.063	10304	.000

a. Lilliefors Significance Correction

Tests of Normality

	Kolmogorov-Smirnov ^a		
	Statistic	df	Sig.
FOTO2601	.059	10304	.000
FOTO2602	.062	10304	.000
FOTO2603	.048	10304	.000
FOTO2604	.063	10304	.000
FOTO2605	.061	10304	.000
FOTO2606	.068	10304	.000
FOTO2607	.073	10304	.000
FOTO2608	.069	10304	.000
FOTO2609	.060	10304	.000
FOTO2610	.067	10304	.000
FOTO2701	.103	10304	.000
FOTO2702	.103	10304	.000
FOTO2703	.101	10304	.000
FOTO2704	.098	10304	.000
FOTO2705	.103	10304	.000
FOTO2706	.143	10304	.000
FOTO2707	.142	10304	.000
FOTO2708	.138	10304	.000
FOTO2709	.130	10304	.000
FOTO2710	.124	10304	.000
FOTO2801	.084	10304	.000
FOTO2802	.096	10304	.000
FOTO2803	.081	10304	.000
FOTO2804	.098	10304	.000
FOTO2805	.094	10304	.000
FOTO2806	.081	10304	.000
FOTO2807	.096	10304	.000
FOTO2808	.088	10304	.000
FOTO2809	.086	10304	.000
FOTO2810	.073	10304	.000
FOTO2901	.063	10304	.000
FOTO2902	.062	10304	.000
FOTO2903	.078	10304	.000
FOTO2904	.073	10304	.000
FOTO2905	.064	10304	.000
FOTO2906	.063	10304	.000
FOTO2907	.066	10304	.000
FOTO2908	.067	10304	.000
FOTO2909	.070	10304	.000
FOTO2910	.062	10304	.000
FOTO3001	.068	10304	.000
FOTO3002	.045	10304	.000
FOTO3003	.063	10304	.000
FOTO3004	.054	10304	.000
FOTO3005	.043	10304	.000
FOTO3006	.063	10304	.000
FOTO3007	.064	10304	.000
FOTO3008	.060	10304	.000
FOTO3009	.056	10304	.000
FOTO3010	.070	10304	.000

a. Lilliefors Significance Correction

Tests of Normality

	Kolmogorov-Smirnov ^a		
	Statistic	df	Sig.
FOTO3101	.054	10304	.000
FOTO3102	.066	10304	.000
FOTO3103	.059	10304	.000
FOTO3104	.054	10304	.000
FOTO3105	.054	10304	.000
FOTO3106	.054	10304	.000
FOTO3107	.050	10304	.000
FOTO3108	.053	10304	.000
FOTO3109	.047	10304	.000
FOTO3110	.058	10304	.000
FOTO3201	.078	10304	.000
FOTO3202	.069	10304	.000
FOTO3203	.055	10304	.000
FOTO3204	.094	10304	.000
FOTO3205	.074	10304	.000
FOTO3206	.084	10304	.000
FOTO3207	.065	10304	.000
FOTO3208	.074	10304	.000
FOTO3209	.076	10304	.000
FOTO3210	.077	10304	.000
FOTO3301	.053	10304	.000
FOTO3302	.069	10304	.000
FOTO3303	.060	10304	.000
FOTO3304	.064	10304	.000
FOTO3305	.053	10304	.000
FOTO3306	.046	10304	.000
FOTO3307	.055	10304	.000
FOTO3308	.055	10304	.000
FOTO3309	.050	10304	.000
FOTO3310	.068	10304	.000
FOTO3401	.065	10304	.000
FOTO3402	.073	10304	.000
FOTO3403	.073	10304	.000
FOTO3404	.061	10304	.000
FOTO3405	.056	10304	.000
FOTO3406	.070	10304	.000
FOTO3407	.059	10304	.000
FOTO3408	.081	10304	.000
FOTO3409	.087	10304	.000
FOTO3410	.084	10304	.000
FOTO3501	.055	10304	.000
FOTO3502	.051	10304	.000
FOTO3503	.058	10304	.000
FOTO3504	.051	10304	.000
FOTO3505	.045	10304	.000
FOTO3506	.062	10304	.000
FOTO3507	.057	10304	.000
FOTO3508	.063	10304	.000
FOTO3509	.049	10304	.000
FOTO3510	.048	10304	.000

a. Lilliefors Significance Correction

Tests of Normality

	Kolmogorov-Smirnov ^a		
	Statistic	df	Sig.
FOTO3601	.113	10304	.000
FOTO3602	.121	10304	.000
FOTO3603	.109	10304	.000
FOTO3604	.113	10304	.000
FOTO3605	.121	10304	.000
FOTO3606	.120	10304	.000
FOTO3607	.117	10304	.000
FOTO3608	.121	10304	.000
FOTO3609	.111	10304	.000
FOTO3610	.097	10304	.000
FOTO3701	.073	10304	.000
FOTO3702	.112	10304	.000
FOTO3703	.071	10304	.000
FOTO3704	.105	10304	.000
FOTO3705	.129	10304	.000
FOTO3706	.118	10304	.000
FOTO3707	.069	10304	.000
FOTO3708	.120	10304	.000
FOTO3709	.076	10304	.000
FOTO3710	.075	10304	.000
FOTO3801	.093	10304	.000
FOTO3802	.049	10304	.000
FOTO3803	.061	10304	.000
FOTO3804	.102	10304	.000
FOTO3805	.050	10304	.000
FOTO3806	.081	10304	.000
FOTO3807	.055	10304	.000
FOTO3808	.103	10304	.000
FOTO3809	.082	10304	.000
FOTO3810	.052	10304	.000
FOTO3901	.046	10304	.000
FOTO3902	.061	10304	.000
FOTO3903	.053	10304	.000
FOTO3904	.050	10304	.000
FOTO3905	.064	10304	.000
FOTO3906	.056	10304	.000
FOTO3907	.048	10304	.000
FOTO3908	.053	10304	.000
FOTO3909	.081	10304	.000
FOTO3910	.045	10304	.000
FOTO4001	.041	10304	.000
FOTO4002	.062	10304	.000
FOTO4003	.049	10304	.000
FOTO4004	.040	10304	.000
FOTO4005	.051	10304	.000
FOTO4006	.048	10304	.000
FOTO4007	.062	10304	.000
FOTO4008	.048	10304	.000
FOTO4009	.056	10304	.000
FOTO4010	.045	10304	.000

a. Lilliefors Significance Correction