

Autenticación de Personas Usando la Geometría de las Manos

Tesis para obtener el grado de Maestro en
Inteligencia Artificial

Presenta
Roberto Vásquez Jiménez
Maestría en Inteligencia Artificial
Universidad Veracruzana

Tutor: Dr. Homero V. Ríos Figueroa
Universidad Veracruzana

Revisor: Dr. A. Raúl Hernández Montoya
Universidad Veracruzana

Revisor: M.C. V. Angélica García Vega
Universidad Veracruzana

Septiembre de 2001.

Resumen

Hay muchas tareas que para su realización requieren inteligencia y que los seres humanos y animales realizan todos los días sin prácticamente ningún esfuerzo. Por ejemplo el reconocimiento de patrones es muy difícil de automatizar debido a diversos factores como son el ruido, las diferentes posiciones del objeto, los dispositivos mal calibrados, etcétera. Una de las áreas que actualmente está siendo objeto de gran interés y estudio, es la identificación/autenticación automática de personas utilizando alguna o varias características del individuo. Esta técnica es llamada biometría. Los sistemas biométricos están siendo empleados en una gran variedad de dominios que requieren alguna identificación o verificación de individuos como es el control de accesos, la identificación de personas, etc. Se han usado numerosas características de las personas para realizar aplicaciones, tales como las huellas digitales, el rostro, la voz, el iris, la retina, la mano, la palma de la mano y la escritura, pues otorgan mayor seguridad que los métodos tradicionales de identificación personal que tienen un número importante de inconvenientes capaces de frustrar la adecuada identificación de personas, tales como las claves o los números de identificación personal.

Cada mano de una persona es única, la longitud de los dedos, ancho, grosor, curvaturas y las relativas ubicaciones de estas características distinguen a cada humano de otras personas, aunque no es la medida biométrica ideal, cada aplicación tiene sus ventajas y sus limitaciones y éstas dependen del tipo de aplicación en el que se usará. La geometría de las manos ha sido aplicada para el control de accesos, el control de inmigración, el control para el reparto de comidas, el control de algunos beneficios como pensiones, salud, etc. La geometría de las manos es muy efectiva debido a que las personas tenemos las manos disponibles y que pueden ser utilizadas en cualquier momento. Las medidas de las manos se pueden computar y obtener un vector de datos más fácilmente y con menor costo que para otras partes del cuerpo como las huellas digitales o el iris, que necesitan de aparatos más costosos e involucran un proceso más complejo de identificación debido a la necesidad de obtener detalles más minuciosos.

Aquí se presenta un trabajo realizado para la autenticación de personas utilizando la geometría de las manos. Una aplicación biométrica consta básicamente de dos módulos: el módulo de registro y el módulo de autenticación. El módulo de registro se encarga de registrar al usuario en una base de plantillas que posteriormente servirán para comparar peticiones de autenticación de usuarios que pueden ser genuinos o impostores. El módulo de autenticación se encarga de hacer la comparación de una muestra o vector de características de un usuario que intenta autenticarse contra las que se encuentran almacenadas en la base de plantillas. La tarea principal de este módulo, es la de encontrar la plantilla que corresponda al vector entrante utilizando una función de distancia y restringido por medio de un valor de umbral, dando como resultado si el usuario es aceptado o rechazado.

En ambos módulos se requiere de un proceso preliminar para la extracción de características. Para lograr este objetivo se usaron técnicas de visión por computadora para la eliminación del

ruido por medio del filtrado de imágenes, seguimiento de contornos y finalmente, la obtención de un vector de medidas que es la representación que sirve como materia prima para el funcionamiento de ambos módulos.

Las redes neuronales han sido usadas en un diverso número de aplicaciones. Para este trabajo se usaron los mapas autoorganizativos desarrollado por Kohonen. Este es un método de aprendizaje no supervisado para la clasificación que a partir de los datos de entrada forma grupos por medio de la creación de mapas bidimensionales en donde los datos que estén más relacionados aparecerán más juntos.

Con esta técnica fue posible lograr una buena clasificación de los vectores de las características extraídas de las imágenes. Este tipo de red neuronal tuvo un buen desempeño en el registro de individuos donde las plantillas son almacenadas en forma de pesos de las conexiones entre las neuronas. Asimismo, para la etapa de autenticación, la red fue capaz de asociar adecuadamente las muestras biométricas.

Se comprobó que el uso de las dos áreas correspondientes a la inteligencia artificial (visión por computadora y redes neuronales) se combinaron muy bien, logrando como resultado una aplicación biométrica de calidad. En el mejor de los casos, un 3% para los falsos positivos y un 6.6% para los falsos negativos, tomando en consideración, únicamente el poder discriminativo de las características de las manos. Con esta tasa de rendimiento, se considera que la autenticación de individuos fue alta.

Palabras claves

Biometría, Geometría de las Manos, Inteligencia Artificial, Redes Neuronales, Mapas Auto-organizativos, Kohonen.

Dedico éste trabajo a mis padres quienes me enseñaron a luchar por lo que se quiere y quienes siempre me apoyaron en todas las cosas que fueron importantes para obtener el grado. De la misma forma, a mis hermanos: Juan Antonio y José Alfredo.

Agradezco a:

- *Dios por permitirme la vida y la salud.*
- *A mis padres por todo el apoyo recibido durante toda mi vida.*
- *Al Dr. Homero V. Ríos por sus extensivas sugerencias y continua asesoría prestada durante el proyecto.*
- *A mis compañeros de clases, de manera muy especial a Rosario Aldana por el apoyo escolar y extraescolar.*
- *A las personas que amablemente se prestaron para utilizar sus manos en los experimentos.*
- *Al Dr. Raúl y a la Maestra Angélica por dedicar un poco de su tiempo a la revisión de la tesis.*
- *Al CONACyT por el apoyo recibido mediante una beca para poder realizar la Maestría en Inteligencia Artificial en la Universidad Veracruzana.*

Índice General

1 Marco Teórico	1
1.1 Introducción	1
1.2 Biometría	2
1.2.1 Identificadores biométricos	4
1.3 Identificación contra verificación	7
1.4 Donde se aplica la biometría	8
1.5 Planteamiento del problema	10
2 Análisis de imágenes	12
2.1 Visión por computadora	12
2.1.1 Procesamiento digital de imágenes	12
2.1.2 Reconocimiento de patrones	13
2.1.3 Inteligencia artificial	14
2.2 Análisis de imágenes	14
2.2.1 Adquisición de imágenes	15
2.2.2 Preprocesado de la imagen	15
2.2.3 Segmentación de la imagen	17
2.2.4 Extracción de características	23
2.2.5 Reconocimiento, clasificación o interpretación	25

3	Redes neuronales	31
3.1	Redes neuronales artificiales	31
3.2	Estructura y función de una simple neurona	34
3.2.1	Neurona biológica	34
3.2.2	Modelo de una neurona artificial	36
3.3	Tipos de aprendizaje en las redes neuronales	36
3.4	Un modelo básico: El <i>Perceptron</i>	37
3.5	Red de Retropropagación de Errores	38
3.5.1	Deficiencias de la Retropropagación de errores	42
3.6	Mapas Autoorganizativos de Kohonen	42
4	Implementación del sistema	46
4.1	Metodología Empleada	46
4.1.1	Adquisición de Imágenes	48
4.1.2	Preprocesamiento	48
4.1.3	Segmentación	50
4.1.4	Seguimiento del contorno	50
4.1.5	Registro y autenticación	56
5	Resultados y conclusiones	63
5.1	Resultados	63
5.2	Discusión	65
5.3	Conclusiones	71
5.4	Trabajo Futuro	72

Índice de Figuras

1.1	Sistema biométrico para autenticación.	4
1.2	Efectos en el espacio de representación imperfecta y semejanza métrica en la precisión del sistema. RSL (Representación y Semejanza Limitada). SL (Semejanza Limitada). RL (Representación Limitada). IL (Información limitada). Cuando la curva es alineada con los ejes X e Y, ésta representa una correspondencia perfecta [11].	6
1.3	Comportamiento de un clasificador [11].	8
2.1	Etapas fundamentales del análisis de imágenes.	15
2.2	Pixel $h[i,j]$ y sus 8 vecinos.	17
2.3	Histogramas de la imagen original y de la modificada por la ecualización.	18
2.4	Imagen binaria.	19
2.5	Notación usada para el direccionamiento en el seguimiento de contornos.	22
2.6	Frontera entre dos clases cuya función de densidad condicional es normal.	28
3.1	Neuronas biológicas.	34
3.2	Sinapsis.	35
3.3	Modelo de neurona artificial.	36
3.4	Perceptron y las funciones de activación y de transferencia.	38
3.5	Arquitectura de una red de retropropagación de errores.	39
3.6	Gráfica de la función sigmoide.	41
3.7	Función que define el tipo de interacción entre neuronas.	43

3.8	Topología de la red de Kohonen.	44
4.1	Prototipo donde los usuarios ubican las manos para capturar las imágenes. . . .	47
4.2	Ruido existente en la imagen.	49
4.3	Proceso que se realiza con las imágenes hasta convertirlas a binarias.	51
4.4	Contorno de la imagen antes y después de aplicar el algoritmo de suavizado del contorno.	52
4.5	Contorno de la imagen procesada.	53
4.6	Pasos para obtener medidas de los dedos.	54
4.7	Medidas obtenidas por la aplicación. Cada línea representa una medida extraída de la mano que es representada por el vector de la figura (b) y se puede relacionar por el índice asociado.	57
4.8	Topología de la red empleada.	59
4.9	Creación de una muestra por cada individuo.	60
4.10	Dos únicas posibilidades que presenta el sistema.	62
5.1	Comportamiento del sistema con la modificación del umbral en la etapa de autenticación.	66
5.2	Fijando un umbral de 13, la tasa de falsos positivos es cero, pero la de falsos negativos se incrementa a un 13.33%.	66
5.3	Fijando un umbral de 15, la tasa de falsos positivos es de 3% y la de falsos negativos se incrementa a un 6.6%. Este es el que se considera como el umbral que reporta los mejores resultados debido a que los falsos negativos es bajo y los falsos positivos no es muy alto.	67
5.4	Con un umbral de 17, la tasa de falsos positivos es de 5%, por lo que ya no es muy conveniente para aplicaciones civiles.	67
5.5	Con un umbral de 20, para aplicaciones civiles, ya no es conveniente, porque el porcentaje de falsos positivos es de 8% y es alto.	68
5.6	Manos procesadas con diferentes rotaciones.	69
5.7	Ejemplo de manos con uñas largas bien procesadas.	70

Capítulo 1

Marco Teórico

1.1 Introducción

La autenticación automática de personas es en la actualidad una importante materia de investigación. Para lograr este objetivo, existe una cantidad variada de métodos, y entre ellos se encuentra el estudio de la geometría de las manos. Las manos de los humanos son versátiles y nos permite asir, lanzar, hacer herramientas, etcétera, sin embargo, actualmente las manos tienen otro uso, para obtener un vector de medidas que nos permite verificar identidad. La técnica para verificar automáticamente la identidad de un individuo basado en las características físicas o de comportamiento que pueden ser medidas, contadas y que son únicas en las personas, es llamada biometría (ver sección 1.2).

Los datos biométricos pueden ser extraídos y analizados en un buen número de formas y están siendo usados actualmente como un método de identificación/verificación personal (ver sección 1.3) por medio del cual las personas son reconocidas por alguna de sus partes físicas o características de comportamiento. Se han usado una gran cantidad de características para estos estudios como son: las huellas digitales, la voz, el rostro, la retina, el iris, las firmas, y la geometría de las manos. Idealmente, lo que se quiere hacer, es realizar el mismo trabajo que el cerebro para reconocer y distinguir una persona de otra.

El uso de la biometría para la identificación y la autenticación de personas ofrece algunas ventajas sobre los métodos tradicionales. La autenticación biométrica está basada en la identificación de una característica intrínseca que un humano tiene y que lo diferencia de sus congéneres. Objetos, tales como son las tarjetas inteligentes, las tarjetas de franja magnética y las llaves físicas, pueden ser perdidas o duplicadas. Las claves o contraseñas pueden ser olvidadas o intencionalmente observados por terceras personas.

La biometría puede ser integrada dentro de aquellas aplicaciones que requiere seguridad, control de accesos e identificación o verificación de usuarios. Con la seguridad biométrica, podemos prescindir de las llaves, las claves y códigos PIN¹, de esta manera el acceso es denegado a las personas ajenas que no tengan el permiso necesario.

Recientemente, en Sandia National Laboratories se ha reportado que la geometría de las manos ha tenido la más alta aceptación por parte de los usuarios de todas las tecnologías biométricas probadas. Con un alto nivel de seguridad, fácil uso y sin peligro para el usuario. La geometría de las manos ha llegado a ser la tecnología biométrica más ampliamente usada hoy en día [22].

En los años recientes, la drástica reducción en costos de los microprocesadores ha llevado la tecnología de la geometría de manos a los mercados comerciales. Este tipo de tecnología puede ser encontrada en plantas nucleares, centros de salud, clubes, bancos de sangre y de esperma y en departamentos de policía entre otros.

Las aplicaciones de biometría que utilizan la geometría de las manos han sido explotadas por más de tres décadas. Estas han sido en su gran mayoría aplicaciones comerciales y por lo tanto la literatura sobre investigaciones en este tipo de aplicaciones es escasa. Por tal motivo, para lograr una solución satisfactoria a nuestro problema, en esta tesis se desarrolla un método utilizando técnicas de inteligencia artificial como son la visión por computadora y las redes neuronales. La visión por computadora se encarga de procesar las imágenes de las manos y de extraer las características que nutren a la red neuronal que es la encargada de registrar y de autenticar a los usuarios. Las redes neuronales con aprendizaje no supervisado² ha resultado una buena herramienta para asociar datos con características similares que forman los grupos que le permiten resolver adecuadamente problemas de biometría como el de la geometría de las manos.

1.2 Biometría

La biometría toma como base la siguiente pregunta. ¿Que características distinguen a una persona de sus semejantes?

La verificación biométrica es un método automatizado por el cual la identidad de un individuo es confirmada examinando uno o más de sus rasgos fisiológicos o características de comportamiento, como podrían ser las huellas digitales, las retinas o las firmas.

Los rasgos fisiológicos son características físicas estables, como aquellas que presentan las palmas de las manos o el patrón del iris que son esencialmente inalterables. Las características de comportamiento tales como las firmas y la voz son influenciadas por acciones controlables porque

¹Del inglés *Personal Identification Number*.

²Es aprendizaje supervisado porque no se conoce que neurona responderá a que patrón de entrada.

Concepto	H. Digitales	ojos	Manos	Rostro	Voz	Firma
Correspondencia: a) 1 a 1 b) 1 a muchos	si, si	si, si	si, no	si, no	si, no	si, no
Variación: a) tiempo de vida b) día a día	nada, poco	nada, muy poco	mucho, muy poco	mucho, medio	mucho, medio	mucho, medio
Muestras máximas por persona	10	2	2	1	1	1
Costo del sensor \$US	$10 - 10^2$	$10^2 - 10^3$	10^2	10^2	$0 - 10^2$	10^2
Tamaño del sensor	muy pequeño	medio	medio	pequeño	muy pequeño	medio

Tabla 1.1: Características de diferentes modalidades biométricas [16].

cambian con el tiempo y los factores fisiológicos son menos controlables. Por lo tanto, algunas pueden servir para verificación, otras para identificación y otras para ambas (ver tabla 1.1).

La plantilla de referencia biométrica registrada debe ser actualizada cada vez que ésta sea usada por las variaciones que pueden existir al paso del tiempo. Aunque la biometría basada en el comportamiento puede ser menos cara y menos amenazante para los usuarios, los rasgos fisiológicos tienden a ofrecer mayor precisión y seguridad. En este caso ambas técnicas proveen un alto nivel de identificación/verificación comparadas con los métodos tradicionales como son las claves o las tarjetas. Los identificadores biométricos actualmente disponibles o bajo desarrollo incluyen huellas digitales, aroma del cuerpo, ADN, forma de la oreja, reconocimiento de rostros, termograma facial, geometría de la palma, escaneo retinal, patrón del iris, geometría de las manos, firma y reconocimiento de voz. Cada técnica biométrica mencionada tienen sus propias ventajas y desventajas tomando en cuenta los requerimientos que se desee incluir. Esto se puede ver en la tabla 1.2 donde se hace una breve comparación entre ellas.

Un sistema biométrico es esencialmente un sistema de reconocimiento de patrones y su trabajo es la identificación de una persona estableciendo la autenticidad de una característica física o de comportamiento específica poseída por el usuario [8]. Un sistema biométrico puede ser dividido en dos módulos (ver figura 1.1):

Módulo de registro: Es el responsable del registro de individuos dentro del sistema biométrico. Durante la fase de registro, las características biométricas de un individuo, son escaneadas por un sensor para adquirir una versión digital. Para facilitar la comparación y reducir los requerimientos de almacenamiento, la representación digital es procesada por un extractor de características que genera una compacta pero expresiva representación

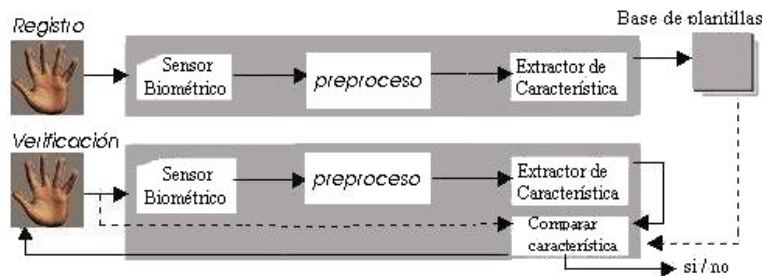


Figura 1.1: Sistema biométrico para autenticación.

llamada *plantilla*³. Dependiendo de la aplicación, la plantilla puede ser almacenada en una base central del sistema biométrico o ser guardada en una tarjeta magnética o tarjeta inteligente distribuida para cada individuo.

Módulo de identificación: Es el responsable para identificar o verificar (ver sección 1.3) a los individuos en el punto de acceso. Durante la fase de operación, los lectores biométricos capturan las características del individuo y los convierte a formato digital el cual es procesado por el extractor de características para producir la misma representación de la plantilla. El resultado de la representación sirve para alimentar a un submódulo que se encarga de compararlo contra las plantillas almacenadas para identificar o autenticar al individuo. Si la tarea esencial del sistema es la autenticación de individuos, el resultado emitido por este módulo debe ser aceptado o rechazado o equivalente a falso o verdadero, permitiendo tomar una decisión apropiada que ayude a mantener la seguridad en donde el sistema se está empleando. Si la tarea es la de identificación, el resultado debe ser la identidad del individuo; por ejemplo en aplicaciones forenses.

1.2.1 Identificadores biométricos

Un identificador biométrico ideal debería tener los siguientes requerimientos[8] (Ver tabla 1.2 donde se hace una comparación entre identificadores biométricos tomando en cuenta estos requerimientos):

Universalidad: Cada persona debería poseer la característica en cuestión.

³En inglés es *template*.

Id. Biométrico	Universalidad	Singular	Permanencia	Colectabilidad	Desempeño	Aceptación	Seguridad
Rostro	alto	bajo	medio	alto	bajo	alto	bajo
Huellas digitales	medio	alto	alto	medio	alto	medio	alto
Geometría de manos	medio	medio	medio	alto	medio	medio	medio
Ritmo de escritura	bajo	bajo	bajo	medio	bajo	medio	medio
Venas en manos	medio	medio	medio	medio	medio	medio	alto
Iris	alto	alto	alto	medio	alto	bajo	alto
Retina	alto	alto	medio	bajo	alto	bajo	alto
Firmas	bajo	bajo	bajo	alto	bajo	alto	bajo
Voz	medio	bajo	bajo	medio	bajo	alto	bajo
Termograma facial	alto	alto	bajo	alto	medio	alto	alto

Tabla 1.2: Comparación entre tecnologías biométricas basado en la percepción de tres expertos [8].

Singularidad: Dos personas no deberían ser la misma en términos de la característica.

Permanencia: La característica no debería cambiar y no debería poderse alterar.

Colectabilidad: La característica puede ser medida cuantitativamente.

En la práctica, una característica biométrica que satisfaga todos los requerimientos antes mencionados no siempre se cumple para los sistemas biométricos.

En sistemas biométricos prácticos hay otros puntos que deben de considerarse:

1. *Desempeño.* Si la identificación/verificación se realiza con precisión, velocidad y robustez.
2. *Aceptabilidad.* El grado en que las personas están dispuestas a aceptar un identificador biométrico particular en su vida diaria.
3. *Seguridad.* Refleja que tan fácil es para engañar al sistema con métodos fraudulentos.

Otros factores como el costo y la facilidad de uso también afectan la eficacia de un sistema biométrico.

Típicamente, un sistema biométrico no es perfecto y comete errores al no reconocer correctamente las identidades genuinas y al no rechazar correctamente a los impostores. El error del primer tipo en conjunto es llamado *falsos negativos* y el segundo tipo de error es

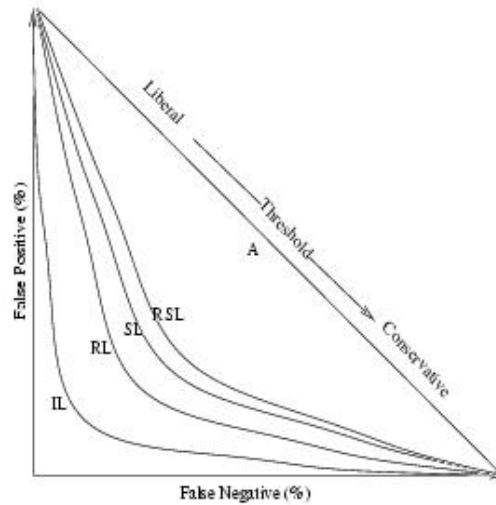


Figura 1.2: Efectos en el espacio de representación imperfecta y semejanza métrica en la precisión del sistema. RSL (Representación y Semejanza Limitada). SL (Semejanza Limitada). RL (Representación Limitada). IL (Información limitada). Cuando la curva es alineada con los ejes X e Y, ésta representa una correspondencia perfecta [11].

llamado *falsos positivos*. La probabilidad de cometer estos dos tipos de errores está en función de los términos FRR^4 y FAR^5 respectivamente.

EL FAR indica la probabilidad de que un dispositivo biométrico permita el acceso o acepte a un intruso o persona no autorizada. Se debe procurar que este indicador sea bajo para mantener a los impostores fuera de los beneficios y privilegios que no tienen permitidos. El FRR es la probabilidad de que un dispositivo biométrico no reconozca o no permita el acceso a un usuario genuino. Un bajo FRR es muy importante para la mayoría de aplicaciones debido a que los usuarios se llegan a frustrar si se les niega el acceso por un sistema o dispositivo que antes los reconoció. Las magnitudes de esos errores dependen de que tan liberal o conservador es el sistema para determinar si dos medidas son originadas por el mismo individuo (ver figura 1.2).

Un pequeño FRR (un sistema más tolerante) usualmente conduce a un gran FAR, mientras que un pequeño FAR (un sistema menos tolerante) usualmente implica un gran FRR, es decir, existe un compromiso entre estas dos tasas. La figura (1.3(a)) describe una caracterización del desempeño de un sistema hipotético. Donde G1 y G2 (distribuciones genuinas) pertenece

⁴Por sus siglas en inglés *False Reject Rate*.

⁵Por sus siglas en inglés *False Acceptance Rate*.

a las distribuciones de las marcas de correspondencias entre patrones de la misma clase; I1 y I2 (distribución impostora) de diferentes clases. Las marcas que constituyen la distribución genuina son en promedio, más altas que la de la distribución del impostor. Las dos regiones sombreadas indican el mínimo porcentaje de error total del sistema. Sin embargo, se puede eliminar la dependencia utilizando un umbral T .

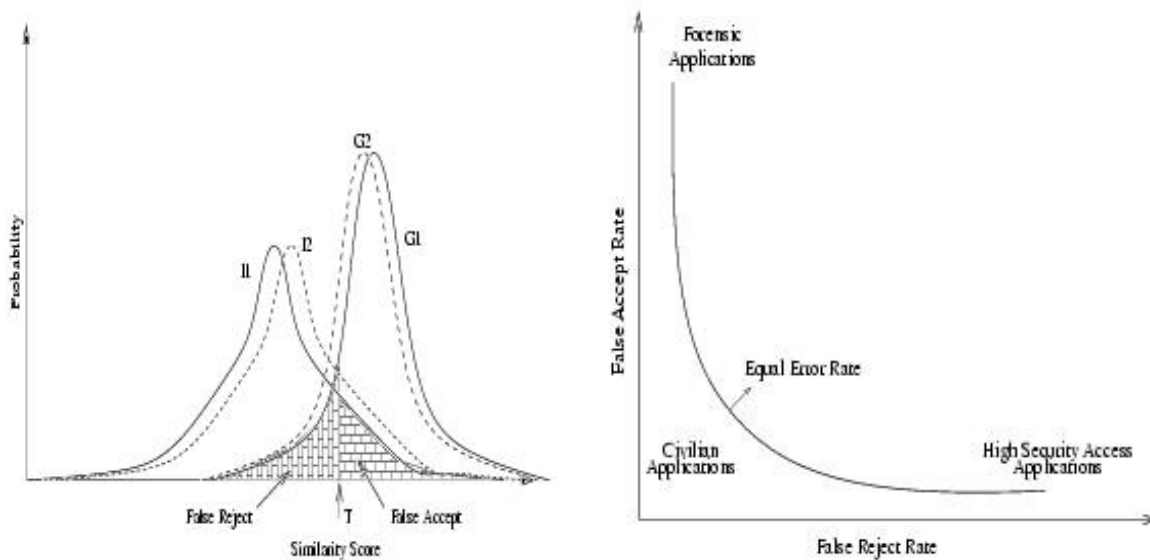
La figura (1.3(b)) muestra la relación que existe entre el FAR y el FRR de un sistema en diferentes puntos de operación, esto es llamada ROC⁶ que es una medida comprensible de la precisión del sistema. Las aplicaciones con alta seguridad de acceso deben operar a un punto ROC con pequeño FAR. Las aplicaciones forenses intentan apresar a un criminal no importando si se tiene que examinar un gran número de falsos candidatos y por lo tanto operan con un alto FAR. Finalmente, las aplicaciones civiles esperan operar en un punto medio, es decir con un bajo FRR y un bajo FAR. El porcentaje de error del sistema en un punto de operación donde el FAR es igual a FRR es llamado EER⁷ el cual es utilizado como una somera descripción de la precisión del sistema [11].

1.3 Identificación contra verificación

Esencialmente hay dos funciones claves ofrecidas por los sistemas biométricos. Una es la *identificación*, que es un proceso de correspondencia de una a muchas, en el cual la muestra biométrica es comparada contra una base de plantillas donde están almacenadas las características de los usuarios previamente registrados. La otra es la *verificación*, que es un proceso de correspondencia una a una en el cual el sistema biométrico verificará si una muestra biométrica de un individuo corresponde con alguna previamente registrada en el sistema. Generalmente en esta función se hace uso de cierta métrica entre muestras y es manejada por un umbral. En la tabla 1.1 se indican algunas partes físicas del cuerpo que son utilizadas para identificación y/o verificación.

⁶Por sus siglas en inglés *Receiver Operating Characteristics*.

⁷Por sus siglas en inglés *Equal Error Rate*.



(a) Distribución de probabilidad entre clasificación genuina e impostora.

(b) Curva que ilustra el FRR y FAR en un clasificador en todos sus puntos.

Figura 1.3: Comportamiento de un clasificador [11].

1.4 Donde se aplica la biometría

La seguridad en términos biométricos se refiere a las técnicas de autenticación que confían en las medidas de características biológicas que pueden ser automáticamente revisadas. Ejemplos de lo anterior se encuentran las huellas digitales o el reconocimiento de voz. La biometría ha sido por décadas usada en el sector público y adoptada por la milicia y las agencias de aplicación de leyes. En la actualidad las agencias públicas están usando la biometría para prevenir fraudes o para determinar que personas pueden gozar de los beneficios de un determinado servicio tales como el médico. Actualmente, un amplio rango de proyectos pilotos que usan la biometría, están desarrollándose en todo el mundo. El reconocimiento del rostro está siendo incorporado dentro de cajeros automáticos en Alemania. El reconocimiento del iris está siendo incorporado en cajeros automáticos en Japón. Debido a la pérdida de grandes cantidades de dinero por fraudes, MasterCard y Visa están explorando el uso de la biometría para incrementar la seguridad en sus tarjetas de crédito. Cuando los usuarios desean obtener dinero, realizar depósitos o consultar su balance deben de proporcionar sus huellas digitales ya que el Standard Bank de Sudáfrica los

está usando en sus cajeros automáticos en vez de usar los PINs. Recientemente, el Chemical Bank anunció que ha seleccionado un sistema para la verificación de voz que le permite la identificación de clientes en sus transacciones. Charles Schwab and Co., están usando un sistema de verificación de huellas digitales con empleados y jefes para conformar los requerimientos de seguridad y cambios de comisiones. Otros negocios como la Coca-Cola y Woolworth Australia, usan la biometría para sus tiempos y sistemas de servicio. Los visitantes de Walt Disney World en Florida están siendo registrados en un sistema de huellas digitales. En Texas, EUA, los cajeros automáticos de algunos bancos están incorporando reconocimiento de iris. Se ha estado iniciando el uso en la administración o programas de derecho como beneficios económicos, salud o pensiones. No solamente en las grandes corporaciones está siendo usada la biometría, sino también, ya existen aplicaciones para usuarios individuales. Casas en Japón ya están siendo aseguradas por dispositivos que reconocen las huellas digitales.

Ninguno de los ejemplos antes mencionados incluyen aplicaciones gubernamentales de la biometría, sin embargo, en el gobierno de Estado Unidos consideran que la tecnología es lo suficientemente avanzada para formar un panel de expertos para estudiar los últimos desarrollos en la industria. Asimismo, el gobierno Estadounidense utiliza la verificación biométrica para acceso a redes de computadoras y acceso físico, con el objeto de monitorear los beneficios y reducir el fraude de esos programas. Algunos otros programas de gobierno que hacen uso de la biometría son, el Servicio de Inmigración y Naturalización (INS), el Sistema de Servicio Acelerado a Pasajeros (INPASS) el cual permite a visitantes frecuentes de los Estados Unidos pasar rápidamente los puntos de inspección utilizando la geometría de las manos. También, a lo largo de la frontera de Canadá, el gobierno usa un sistema llamado CANPASS que es similar a INPASS pero usa las huellas digitales en vez de la geometría de las manos.

En el aeropuerto de la ciudad de México está siendo usada la geometría de las manos para el control de taxis. Cuando existen personas que necesitan el servicio, el taxista tiene que pasar por la autenticación para que se le permita prestar el servicio al usuario.

Los atletas del mundo y otros participantes en los juegos olímpicos de 1996 en Atlanta, tuvieron la seguridad en la palma de sus manos. Start trek-like, como inicialmente llamaron a la tecnología biométrica de la geometría de las manos, fue desarrollado por Recognition System, Inc. (RSI) de Campbell, California. Ellos leyeron el tamaño y forma de la mano de cada individuo participante en los juegos olímpicos para conceder acceso a áreas de alta seguridad. Asimismo, se registró a los

entrenadores y otros individuos que formaban parte del grupo de los atletas. También en estos juegos olímpicos, el hotel Omni usó la tecnología de la geometría de las manos para hospedar a jugadores de los Estados Unidos. De este modo fueron registrados los equipos de basquetbol de hombres y mujeres, entrenadores, familiares e invitados. El plan de seguridad consistió en registrar a las personas antes de los juegos olímpicos. El total de personas registradas en este hotel fue aproximadamente de dos mil.

1.5 Planteamiento del problema

Debido a la gran demanda de sistemas de seguridad que sean rápidos y baratos, así como la poca riqueza de metodologías públicas en aplicaciones biométricas que usen la geometría de las manos, se realiza un estudio para resolver el problema de autenticación de individuos utilizando esta tecnología biométrica.

La geometría de las manos de personas ha sido utilizada para la autenticación, mayormente en sistemas comerciales y basado principalmente en patentes⁸. Aunque tales sistemas han sido comercializados desde hace aproximadamente tres décadas, la literatura en esta área aún es muy pobre.

Este trabajo pretende contribuir en el diseño de un método para la autenticación de personas mediante la geometría de las manos lo suficientemente robusto que utilice técnicas de inteligencia artificial. Para extraer las características de las manos se usan técnicas de visión por computadora (ver sección 2.1) y para la clasificación se hace uso de las redes neuronales artificiales con aprendizaje competitivo, específicamente los mapas autoorganizativos de Kohonen (ver sección 3.8). El problema que se pretende resolver con las técnicas mencionadas, consiste en verificar si un individuo está previamente registrado, si es así, el sistema será capaz de aceptarlo, en caso contrario rechazarlo. Esto tomando en cuenta las dimensiones físicas en la estructura de la mano derecha de la persona. Esta estructura incluye el ancho y largo de los dedos y el ancho de la palma de la mano.

El problema de la clasificación de las manos entre diferentes individuos se vuelve interesante

⁸Se refiere a un certificado otorgado por un gobierno al autor de un invento industrial para asegurarle su propiedad y la explotación exclusiva durante cierto tiempo.

debido a que las medidas de éstas no varían significativamente entre la población y se puede encontrar una gran cantidad de grupos, por lo tanto, la distancia entre éstas es muy pequeña haciendo difícil su discriminación.

Capítulo 2

Análisis de imágenes

En este capítulo se presenta de forma general las técnicas involucradas para el procesamiento y análisis de imágenes tales como el filtrado, ecualización y segmentación. Además, se presenta un bosquejo de algunas técnicas tradicionales para la clasificación de datos como son las estadísticas.

2.1 Visión por computadora

El propósito de la Visión por computadora es el de crear modelos del mundo real a partir de imágenes visuales. Un sistema de visión recupera información útil de la escena de su proyección en dos dimensiones. Estos sistemas han sido utilizados exitosamente en aplicaciones de inspección industrial, agricultura, control de calidad de productos, hojas para turbinas y otras más.

Muchos campos están relacionados con la visión por computadora debido a que reciben información proveniente de imágenes.

2.1.1 Procesamiento digital de imágenes

Las técnicas de procesamiento de imágenes, usualmente transforman imágenes dentro de otras imágenes. La tarea de recobrar información para los humanos es muy fácil y la realizamos

todos los días sin darnos cuenta. En este campo se incluyen tópicos como el realce de imágenes, compresión de imágenes y corrección de imágenes borrosas por dispositivos mal calibrados. Los algoritmos toman imágenes de entrada y producen otro tipo de representación como salida, tales como los contornos de los objetos existentes en una imagen. El objetivo principal de la visión por computadora es recobrar información automáticamente, con la mínima intervención del humano.

Una de las aplicaciones iniciales de la primera categoría de técnicas de tratamiento de imágenes consistió en mejorar las fotografías digitalizadas de periódico enviadas por un cable submarino interoceánico entre Londres y Nueva York. La introducción del sistema Bartlane de transmisión de imágenes por cable a principios de la década de 1920 redujo el tiempo necesario para enviar una fotografía a través del Atlántico de más de una semana a menos de tres horas [6].

2.1.2 Reconocimiento de patrones

El objetivo del reconocimiento de patrones, es el de clasificar datos numéricos y simbólicos. Muchas técnicas estadísticas y sintácticas han sido desarrolladas para la clasificación de patrones. Las técnicas de reconocimiento de patrones, hoy en día, juegan un papel importante en la visión por computadora para el reconocimiento de objetos.

El diseño de un sistema de reconocimiento de patrones está basado en la modelación de *(i)* la información que se espera permanezca invariante en diferentes instancias de una entidad dada para ser reconocida y *(ii)* la información suficiente para discriminar una entidad de otras entidades extrañas o de otras entidades válidas. Un típico sistema de reconocimiento de patrones simultáneamente espera capturar la información invariante en la representación de patrones de la misma entidad, así como la disponibilidad de información para la discriminación de patrones que representan diferentes entidades en dos estados distintos:

- *(a)* Cantidad de información extraña captada por el sensor. El extractor de características debe ser bien diseñado para poder extraer una representación precisa.
- *(b)* Un estado de decisión que sea tolerante a fallas para poder realizar una correspondencia precisa entre la muestra de entrada y la muestra almacenada.

Por ejemplo, en un sistema de autenticación de personas basado en la geometría de las manos, el módulo de extracción de características puede localizar el dedo índice en la imagen de la mano capturada por una cámara y determinar su longitud y ancho para obtener un vector de características (l, w) ; El módulo que se encarga de compararlos puede usar la norma Euclidiana como distancia métrica:

$$D_i = \sqrt{(l - l_i)^2 + (w - w_i)^2} < T. \text{ donde } (l, w) \in C_i \quad (2.1)$$

Donde T es un umbral, (l_i, w_i) es un patrón prototipo que representa a la clase C_i para identificar a los patrones originados por la misma mano.

2.1.3 Inteligencia artificial

La Inteligencia Artificial se encarga del diseño de sistemas que intentan emular el comportamiento y comprensión humana y que sean capaces de aprender, reconocer y pensar. La inteligencia artificial es usada para analizar escenas procesando una representación simbólica del contenido de la escena después de que las imágenes han sido procesadas para obtener sus características. La inteligencia artificial puede ser vista en tres etapas principales: *percepción*, *cognición* y *acción*. La *Percepción* traslada las señales del mundo a símbolos, la *Cognición* manipula símbolos y la *Acción* traslada símbolos a señales o efectos que cambian al mundo. Muchas técnicas de inteligencia artificial juegan un papel importante en todos los aspectos de visión por computadora.

2.2 Análisis de imágenes

El análisis de la imagen se realiza en las siguientes fases: Adquisición, preprocesado, segmentación, extracción de características y reconocimiento (ver figura 2.1).

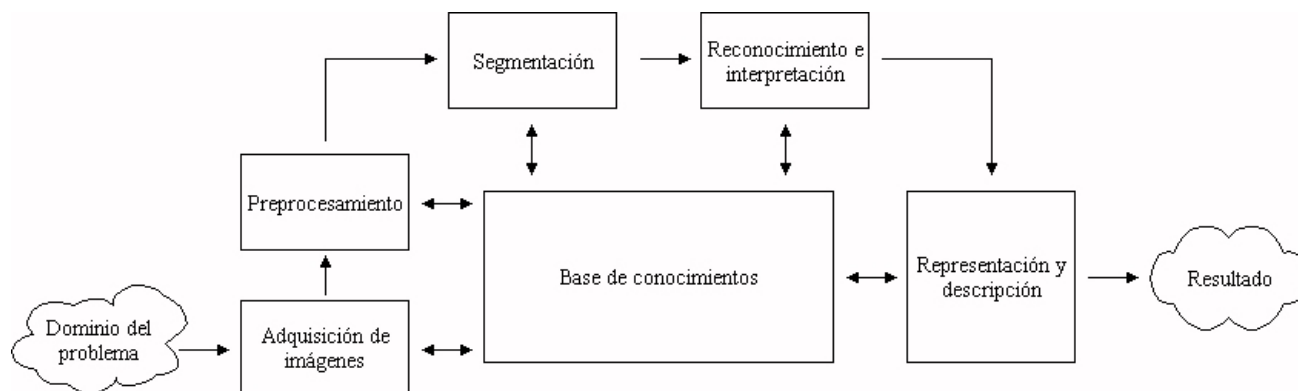


Figura 2.1: Etapas fundamentales del análisis de imágenes.

2.2.1 Adquisición de imágenes

La adquisición de la imagen se efectúa por medio de una cámara de video o digital que codifica la imagen formada en el plano del sensor en forma de una señal analógica que es función de la intensidad luminosa recibida. Esta se envía a un conversor analógico/digital, que la convierte en una matriz de valores numéricos, llamados píxeles, dentro de un rango denominado *niveles de grises*. Las imágenes a color son la agrupación de tres imágenes como la anterior filtradas en rojo, verde y azul¹. Los valores de cada píxel son almacenados en una memoria sobre la que se efectuarán todas las operaciones posteriores.

2.2.2 Preprocesado de la imagen

El preproceso incluye operaciones de realce, reconstrucción y filtrado. El realce consiste en disminuir o eliminar los efectos negativos que se producen por la variación de la iluminación en la imagen o el mal calibrado de la lente de la cámara. La reconstrucción se lleva a cabo cuando se ha perdido parte de la información que ha llegado en mal estado por efecto de las interferencias. El filtrado de la imagen se emplea generalmente para eliminar el ruido debido a las interferencias.

Atendiendo a la información que se emplea, se puede clasificar a las técnicas de tratamiento de la imagen mediante los siguientes criterios [15]:

¹Denominadas en la literatura como **RGB** por sus iniciales en inglés.

Basadas en operadores puntuales: se trata de operaciones que modifican el valor de los pixeles individualmente sin tener en cuenta el valor de los pixeles vecinos.

Basadas en histogramas: Emplean la información global proporcionada por el histograma de frecuencias de los niveles de grises para generar una nueva imagen.

Basadas en operadores locales: Utilizan información de los pixeles vecinos para cada uno de los que forman la imagen para generar una nueva imagen.

Filtrado de imágenes.

Cuando una imagen es adquirida por una cámara u otro sistema de captura de imágenes, es muy común que un sistema de visión sea incapaz de utilizarlas directamente de manera correcta. La imagen puede estar corrompida por variaciones aleatorias en las intensidades, variaciones en la iluminación o un pobre contraste y que deben ser distribuidos en etapas tempranas del sistema de visión, por lo tanto el filtrado de imágenes forma parte del preprocesado de la imagen.

Filtro del promedio.

Es uno de los más simples filtros lineales porque es implementado por una operación de promedio local donde el valor de cada pixel es reemplazado por el promedio de todos los valores de una vecindad local. El procedimiento consiste en dada una imagen $f(i, j)$, generar una imagen filtrada $g(i, j)$ cuya intensidad para cada punto $h[i, j]$ se obtiene promediando los valores de intensidades de los pixeles de f incluidos en el entorno de vecindad de (i, j) , es decir, para obtener la imagen filtrada se utiliza la fórmula 2.2:

$$h[i, j] = \frac{1}{M} \sum_{(k,l) \in N} f[k, l] \quad (2.2)$$

Donde M es el número total de pixeles en la vecindad N (ver figura 2.2). Por ejemplo, tomando una vecindad de 3 x 3 y como base el punto [i,j] la fórmula que nos permite aplicar este filtro es la siguiente:

1	2	3
4	$h[i,j]$	5
6	7	8

Figura 2.2: Pixel $h[i,j]$ y sus 8 vecinos.

$$h[i, j] = \frac{1}{9} \sum_{k=i-1}^{i+1} \sum_{l=j-1}^{j+1} f[k, l] \quad (2.3)$$

Ecuación de histograma

Cuando no existe un ambiente adecuado para adquirir las imágenes, éstas pueden contener valores de grises mal distribuidos. Es muy común encontrarse con imágenes en las cuales sus niveles de intensidades se distribuyen en un pequeño rango, es decir, tienen un pobre contraste. La ecualización del histograma es un método para aumentar el contraste para tales imágenes, redistribuyendo uniformemente los valores de grises. En general, la modificación del histograma realza la calidad de una imagen y es usado cuando la imagen es intentada ser vista por un humano. Este método es muy efectivo cuando posteriormente se requiere el uso de un umbral. En la figura 2.3(a) se puede observar el histograma de una imagen con un pobre contraste y en la figura 2.3(b) se muestra el histograma de la imagen que ha sido modificado por la ecualización. Para realizar esta tarea es necesario aplicar la fórmula 4.1.

2.2.3 Segmentación de la imagen

El primer paso en el análisis y comprensión de las imágenes consiste en particionar la imagen de manera que las regiones que representan diferentes objetos sean explícitamente marcadas, es decir, se extraen los objetos de interés de la imagen y se separan del fondo. Las particiones pueden ser obtenidas a partir de las características de los valores de grises de los píxeles de

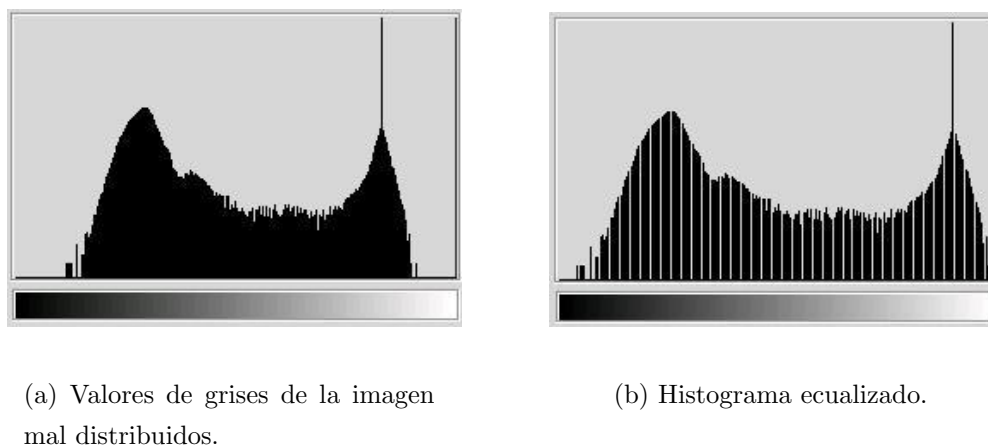


Figura 2.3: Histogramas de la imagen original y de la modificada por la ecualización.

la imagen. Hay que recalcar que una imagen es un arreglo bidimensional y los valores de este arreglo son los valores de grises. Durante la segmentación, la imagen se divide en regiones de interés según una serie de criterios definidos por la persona encargada de realizar este trabajo, en función de los objetos que se desean encontrar. La segmentación más sencilla consiste en establecer un nivel de gris como umbral en imágenes de 256 niveles. Cada conjunto de píxeles que sean mayores que dicho umbral pertenece a una región y los que no, a la otra. La segmentación por color es mucho más complicada al tener que manejar no solo una característica por píxel (nivel de gris), sino tres, correspondientes a los valores de las componentes RGB.

La segmentación autónoma es una de las labores más difíciles del tratamiento digital de imágenes. Por una parte, un procedimiento de segmentación fino dilata la solución satisfactoria de un problema. Por la otra, un algoritmo de segmentación débil o errático casi siempre garantiza que en cualquier momento habrá un fallo. El papel fundamental de la segmentación es el de extraer caracteres individuales y palabras de fondo [6].

Un método para segmentar una imagen es el *basado en regiones*, donde todos los píxeles que corresponden a un objeto son agrupados y son marcados para indicar que ellos pertenecen a una misma región. Los píxeles son asignados a regiones usando algún criterio que los distingue del resto de la imagen. Dos importantes principios en segmentación son *semejanza de valores* y *proximidad espacial*. Los píxeles pueden ser asignados a la misma región si ellos tienen características de intensidad semejantes o si ellos están cerca el uno del otro. Por ejemplo, una

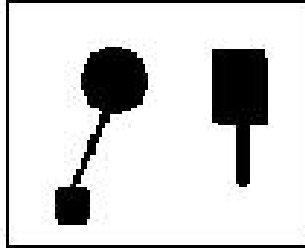


Figura 2.4: Imagen binaria.

medida específica de la semejanza de valores entre dos píxeles es la diferencia entre los valores de grises y una medida específica de proximidad espacial puede ser la distancia Euclidiana (ver ecuación 2.1).

El resultado del proceso de segmentación arroja datos de los píxeles que no tienen significado para una máquina. Por ejemplo, pueden constituir el contorno de un objeto o bien un conjunto de puntos que conforman un objeto. Por lo tanto, es necesario convertirlos adecuadamente a datos que sea comprensible y a la vez factible de ser interpretado por una computadora utilizando algún algoritmo para tal propósito.

Imagen binaria.

Una imagen contiene un continuo de valores de intensidades pero antes debe ser cuantificada² para obtener una imagen digital. La información de una imagen se encuentra en los valores de grises originados por la discretización. Para interpretar una imagen, las variaciones en los valores de intensidad deben de ser analizados. El número más comúnmente usado de niveles de cuantificación para representar las intensidades de la imagen es de 256 valores diferentes de grises. No es común, pero se pueden ver imágenes digitales cuantificadas a 32, 64, 128 o 512 niveles de intensidad para ciertas aplicaciones y a veces hasta a 4096 (12 bits) son usados en medicina [9].

En los primeros días de la visión por computadora, la memoria y el poder de computación disponible fue muy limitado y caro. Esas limitaciones estimularon a los diseñadores de aplicaciones de visión a centrar sus esfuerzos en sistemas de visión binaria. Una imagen binaria

²Cuantificación se entiende en el sentido de discretizar y no de contar.

contiene solamente dos niveles de grises (ver figura 2.4). Además, los diseñadores notaron que las personas no tenían dificultad para entender las líneas dibujadas, siluetas, y otras imágenes formadas usando solamente dos niveles de grises. Estimulados por esta capacidad humana, se usaron imágenes binarias en muchas aplicaciones.

A pesar de que las computadoras han llegado a ser mucho más poderosas, los sistemas de visión binaria aún siguen usándose. Una razón es que los algoritmos para computar las propiedades de imágenes binarias son bien entendidos. Utilizar este tipo de imágenes tiende a ser menos caro y más rápido que utilizando sistemas de visión que trabajan sobre niveles de grises o en imágenes a color; esto es debido al pequeño uso de memoria y de requerimientos de procesamiento.

Los pequeños requerimientos de memoria y rápido tiempo de ejecución no son las únicas razones para el estudio de sistemas de visión binaria. Muchas técnicas desarrolladas para estos sistemas son también aplicables a los sistemas de visión que usan imágenes en escala de grises. La máscara de un objeto es un elemento binario en donde los pixeles del objeto tienen valores de unos y los otros puntos de ceros. Una forma conveniente para representar un objeto en niveles de grises o imágenes a color es usar esta máscara. Después de que el objeto ha sido separado del fondo, las propiedades geométricas y topológicas pueden ser computadas de la imagen binaria.

Segmentación por umbral.

Uno de los más importantes problemas en un sistema de visión es identificar las subimágenes que representan a los objetos. Esta operación, que es muy natural y muy fácil para las personas, es sorprendentemente difícil para las computadoras. Al proceso de particionar una imagen dentro de regiones es llamado segmentación. Idealmente, una partición representa un objeto o partes de un objeto. Formalmente, la segmentación puede ser definida como un método para particionar una imagen, $F[i,j]$, dentro de subimágenes llamadas regiones P_1, \dots, P_k , donde cada subimagen es un objeto candidato.

Una imagen binaria es obtenida usando una apropiada segmentación de una imagen en escala de grises. Si los valores de las intensidades de un objeto están en un intervalo y los valores de las intensidades de los pixeles del fondo están fuera del intervalo, una imagen binaria puede ser obtenida usando una operación llamada umbral (*Thresholding*) que ponen los pixeles en el intervalo a uno y a cero los que se encuentran fuera. Para la visión binaria, la segmentación y

la aplicación de un umbral son sinónimos.

La segmentación por umbral es un método para convertir una imagen en escala de grises a una imagen binaria, así que los objetos de interés son separados del fondo. El umbral debe ser efectivo en la separación entre los objetos y el fondo, así, es necesario que los objetos y el fondo tengan suficiente contraste que nos permita conocer los niveles de intensidades de los objetos o del fondo.

Asumimos que la imagen binaria $B[i,j]$, es la misma imagen a la original en niveles de grises $F[i,j]$ pero que se le aplicó un umbral T dando como resultado $F_T[i,j]$, tal que [9]:

$$B[i, j] = F_T[i, j] \quad (2.4)$$

Donde para un objeto oscuro sobre un fondo claro

$$F_T[i, j] = \begin{cases} 1 & \text{si } F[i,j] \leq T \\ 0 & \text{en otro caso} \end{cases} \quad (2.5)$$

Si se conoce que los valores de intensidad de un objeto se encuentran en un rango $[T_1, T_2]$, entonces podemos usar:

$$F_T[i, j] = \begin{cases} 1 & \text{si } T_1 \leq F[i,j] \leq T_2 \\ 0 & \text{en otro caso} \end{cases} \quad (2.6)$$

Un esquema general del uso de un umbral en donde los niveles de intensidad para un objeto pueden venir de varios intervalos disjuntos, puede ser representado como:

$$F_T[i, j] = \begin{cases} 1 & \text{si } F[i,j] \in Z \\ 0 & \text{en otro caso} \end{cases} \quad (2.7)$$

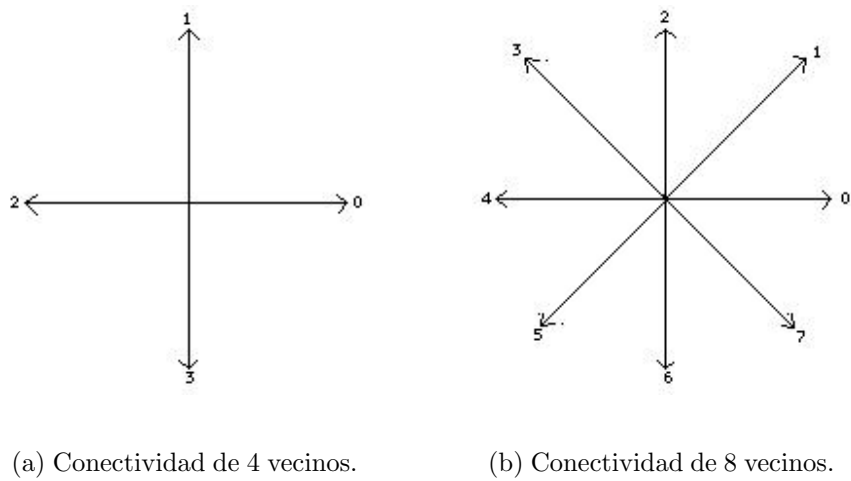


Figura 2.5: Notación usada para el direccionamiento en el seguimiento de contornos.

Donde Z es un conjunto de valores de intensidades para los componentes del objeto. El resultado de la aplicación de un umbral a una imagen en escala de grises, es una imagen binaria (ver figura 2.4).

Seguimiento de Contorno

El contorno del objeto S es un conjunto de pixeles que pertenecen a S pero que se encuentran en la frontera con los pixeles del fondo de la imagen.

Una simple operación local en imágenes binarias puede ser la de encontrar pixeles entre los límites de los objetos y el fondo. En muchas aplicaciones, uno quiere trazar todos los pixeles de una región para posteriormente realizar algún tipo de interpretación sobre el contorno. A continuación se presenta el algoritmo para hacer lo anterior.

Algoritmo:

1. • Buscar en la imagen un pixel que se encuentre en el borde P_0 , que puede ser el pixel de la esquina superior izquierda. Este es el pixel de inicio del contorno de la región.
2. • Definir una variable dir donde se almacenará la dirección de la operación previa a lo largo del contorno.
 - Asignar a la variable el valor asociado al grado de conectividad.
 - a) $dir = 0$ si el contorno es detectado en una conectividad de 4 vecinos. (ver figura 2.5(a)).
 - b) $dir = 7$ si el contorno es detectado en una conectividad de 8 vecinos. (ver figura 2.5(b)).
- Buscar en una vecindad de 3 x 3 tomando en cuenta el pixel actual en una dirección contraria a la del reloj, iniciando la búsqueda en el pixel en la posición indicada por dir de la vecindad.
4. - $(dir+3) \bmod 4$ si es una vecindad de 4.
- $(dir+7) \bmod 8$ si dir es par en una vecindad de 8.
- $(dir+6) \bmod 8$ si dir es impar para una vecindad de 8.
5. • Si el pixel actual del borde P_n es igual al segundo pixel del borde P_1 y si el pixel anterior del borde P_{n-1} es igual a P_0 ^c, entonces, terminar. En caso contrario repetir a partir del paso 4.
6. • El borde detectado es representado por los pixeles P_0, \dots, P_n

^cEste algoritmo es en un inicio para contornos cerrados, pero puede adaptarse para bordes no cerrados.

2.2.4 Extracción de características

A cada región se le extrae sus características, que son un conjunto de parámetros tales como el área, perímetro, momentos de inercia, concavidades o convexidades, etc., que forman el llamado vector de características. Estas características deben ser importantes para la interpretación posterior y son obtenidas en forma de alguna *estructura de descriptores*. Por lo general deben

ser invariantes bajo las operaciones de traslación, rotación y escalamiento.

A continuación se presentan algunos **Descriptores de frontera** comúnmente utilizados.

Códigos de cadena: Se usan para representar un contorno como un conjunto de segmentos con longitud y dirección específica. Son mucho más útiles para la representación de formas que para el reconocimiento de formas. Típicamente esta representación está basada en la *conectividad* de 4 y 8 vecinos, donde la dirección de cada vecino está codificada usando un esquema numérico.

Firmas: Es una representación funcional unidimensional de una frontera. Existen muchas formas de obtener firmas. Una de las más sencillas es la polar que se usa para representar la distancia desde el centro hasta la frontera como una función de ángulo. Las firmas creadas mediante este método dependen del tamaño y del punto de comienzo. Se puede hacer una normalización de tamaño haciendo que la curva $r(\theta)$ tenga un valor máximo de una unidad. El problema de encontrar el punto de comienzo se puede resolver obteniendo primero el código de cadena de la frontera y usando a continuación el procedimiento discutido anteriormente [1].

Aproximaciones poligonales: Una frontera digital se puede aproximar con una exactitud arbitraria mediante un polígono. Para una curva cerrada, la aproximación es exacta cuando el número de segmentos del polígono es igual al número de puntos de la frontera de forma que cada par de puntos adyacentes defina un segmento del polígono. En la práctica, el objetivo de la aproximación poligonal es capturar la *esencia* de la frontera con el menor número posible de segmentos poligonales.

Números de contornos: Una frontera codificada en cadena tiene varias diferencias, dependiendo del punto de comienzo. El número de contorno de una frontera, basándose en el código de cuatro direcciones, se define como la primera diferencia de menor valor. El orden n de un número de contornos se obtiene como el número de dígitos de su representación. El orden n es par para las fronteras cerradas y su valor nos da el número de posibles contornos diferentes.

2.2.5 Reconocimiento, clasificación o interpretación

Si hacemos una pequeña revisión del sistema de visión biológico, somos capaces desde temprana edad, de reconocer con gran facilidad una amplia variedad de objetos. Incluso, los insectos tienen esta capacidad muy desarrollada. Por ejemplo las abejas son capaces de realizar reconocimiento visual. Sin embargo, los sistemas artificiales actuales que realizan reconocimiento se quedan muy atrás con respecto a los sistemas biológicos [1]. Nuestro sistema de visión nos permite ser capaces de generalizar a partir de un conjunto pequeño de vistas de un mismo objeto. Por ejemplo, somos capaces de ver a una persona pocas veces y después poder reconocerla aunque esté en posiciones diferentes a las previamente vistas, sin importar que esté vestida de forma diferente o esté peinada con un estilo diferente. Desde el punto de vista computacional, existen métodos que son capaces de reconocer objetos, los cuales, tanto por las partes que los componen como por su forma no cambian sustancialmente.

El reconocimiento consiste en analizar la región y poder clasificarla. Para ello se utilizan diferentes tipos de operadores (bayesianos, lineales, cuadráticos, redes neuronales, etc.). El sistema más simple solo maneja dos clases posibles: objetos y fondo.

Es importante hacer notar que cuanto más características y clases se manejen, el procedimiento se complica en proporciones no lineales. Para efectos prácticos, la detección se suele realizar únicamente a partir de unas pocas características atribuidas a dos clases.

La elección del clasificador y del vector de características depende de cada problema particular.

El cálculo de las coordenadas espaciales del objeto detectado no es inmediato, incluso, muchas veces es imposible. Así ocurre en los sistemas de visión que trabajan con una sola imagen, ya que como ésta se encuentra definida en un plano, solo se obtiene información de dos de ellas. La tercera se debe encontrar por otros procedimientos, tales como sensores infrarrojos o ultrasónicos.

Decisión de Bayes

Suponiendo que los vectores son aleatorios y que su función de densidad condicional depende de la clase a la que pertenece es conocida, entonces el problema de la clasificación de objetos se

convierte en una prueba de hipótesis.

Sea x un vector n -dimensional y w_i ($i = 1, \dots, m$) una de las m clases en que el objeto que representa puede ser incluido. Sea $P(w_i)$ la probabilidad a priori de que un objeto pertenezca a w_i , es decir, sin conocer su vector. Sea $p(x|w_i)$ la función de densidad condicional del vector x en la clase w_i , es decir, la función de densidad de la distribución de los vectores observación que pertenece a la clase w_i .

Sea $P(w_i|x)$ la probabilidad de que un objeto pertenezca a la clase w_i , dado que su vector toma los valores representados por x . según la regla de Bayes [15]:

$$P(w_i|x) = \frac{p(x|w_i)P(w_i)}{\sum_{j=1}^m p(x|w_j)P(w_j)} \quad (2.8)$$

Por lo que dado un objeto, cuyo vector observación es x , decidimos que pertenece a la clase w_i si:

$$\forall j \neq i, P(x|w_i) > P(x|w_j) \quad (2.9)$$

Puesto que el denominador en 2.8 es común en ambos términos de la comparación, la regla de decisión puede expresarse como:

$$\forall j \neq i, p(x|w_i)P(w_i) > p(x|w_j)P(w_j) \quad (2.10)$$

El modo de estimar, o los supuestos de los que se parte para obtener $p(x|w_i)$ y $P(w_i)$ dan lugar a los diversos métodos de análisis discriminante.

El cálculo o estimación de las probabilidades a priori de las clases, $P(w_i)$, en muchos casos es sencillo. Si además, se asume que las distribuciones condicionales son normales, resulta:

$$p(x|w_i) = \frac{\exp^{-\frac{1}{2}(x-\mu_i)^T C_i^{-1}(x-\mu_i)}}{(2\pi)^{\frac{n}{2}} |C_i|^{\frac{1}{2}}} \quad (2.11)$$

La media de cada clase μ_i y la matriz de covarianza de la misma, C_i se pueden estimar empleando un conjunto de entrenamiento T_i de N_i elementos de la clase w_i . Para ello se utilizan las ecuaciones 2.12.

$$\mu_i = \frac{1}{N_i} \sum_{x \in T_i} x \quad (2.12)$$

$$C_i = \frac{1}{N_i} \sum_{x \in T_i} (x - \mu_i)(x - \mu_i)^T$$

Análisis discriminante lineal

El análisis discriminante lineal puede considerarse como un caso particular de la regla de Bayes. Como resultado del mismo, se encuentra la combinación lineal de variables independientes que minimice el error estimado de clasificación. Se asume que las variables independientes se distribuyen normalmente y tienen similar matriz de covarianzas en cada clase ($C = C_i$) donde $i = 1, \dots, m$. Por lo tanto, la regla de decisión que determina a que clase se asigna un vector x , viene dada por la ecuación 2.13.

$$x \in w_i \leftrightarrow p(x|w_i)P(w_i) > p(x|w_j)P(w_j) \quad \forall j \quad (2.13)$$

Donde, explícitamente, la desigualdad se escribe:

$$\frac{\exp^{-\frac{1}{2}(x-\mu_i)^T C^{-1}(x-\mu_i)}}{(2\pi)^{\frac{n}{2}} |C|^{\frac{1}{2}}} P(w_i) > \frac{\exp^{-\frac{1}{2}(x-\mu_j)^T C^{-1}(x-\mu_j)}}{(2\pi)^{\frac{n}{2}} |C|^{\frac{1}{2}}} P(w_j)$$

Si se extraen logaritmos neperianos a ambos lados de la inecuación y se realizan las simplificaciones oportunas resulta que x pertenece a la clase w_i cuando [15]:

$$z_i = -\frac{1}{2}(x - \mu_i)^T C^{-1}(x - \mu_i) + \ln P(w_i) > -\frac{1}{2}(x - \mu_j)^T C^{-1}(x - \mu_j) + \ln P(w_j) = z_j \quad (2.14)$$

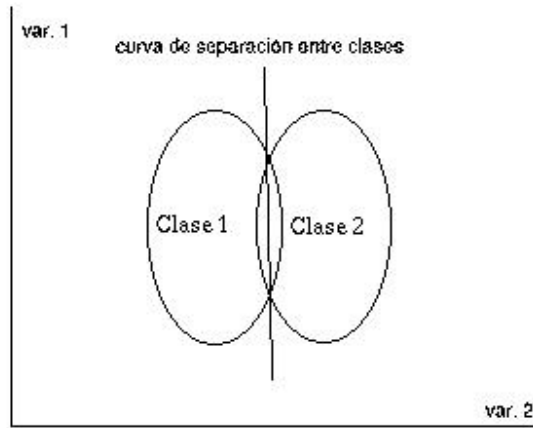


Figura 2.6: Frontera entre dos clases cuya función de densidad condicional es normal.

Donde z_i y z_j se le conoce como coeficientes de discriminación.

Los puntos del espacio con la misma probabilidad de pertenecer a w_i y a w_j , y por lo tanto, con igual coeficiente de discriminación, definen la frontera entre ambas clases. El lugar geométrico de dichos puntos es definido por la ecuación 2.15.

$$z_i = z_j \rightarrow (\mu_j - \mu_i)^T C^{-1} x + \frac{1}{2} (\mu_i^T C^{-1} \mu_i - \mu_j^T C^{-1} \mu_j) = K \quad (2.15)$$

Siendo K una constante, lo que indica que dicho lugar geométrico es un hiperplano de dimensión $n-1$. En el caso bidimensional ($n = 2$) resultará que la frontera entre dos clases es una línea recta (ver figura 2.6). Se puede comprobar de manera sencilla que si las matrices de covarianzas de las distribuciones condicionales no son iguales, entonces la frontera entre los grupos es una forma cuadrática.

Análisis discriminante paso a paso

La aplicación más comúnmente usada es el análisis discriminante, que consiste en incluir múltiples medidas en un estudio y determinar cuales son las que otorgan mayor discriminación entre las clases.

Una variable tiene un gran poder discriminante entre dos clases si la varianza de los elementos que pertenecen a cada una de las clases es pequeña (varianza dentro de la clase). De la misma manera, una variable es muy discriminante si la varianza entre las medias de las clases es grande.

Se define la función λ de Wilks, que mide el poder discriminante total de un conjunto de variables como:

$$\lambda = \frac{|W|}{|B + W|} \quad (2.16)$$

Donde W es la matriz de covarianzas dentro de las clases y B es la matriz de covarianzas entre las medias de los grupos. La función λ de Wilks tiene valores entre 0 y 1, y cuanto menor sea su valor, mayor discriminación existe entre clases. Asimismo, se define estadísticamente la λ parcial de Wilks para la variable i -ésima, λ_i como:

$$\lambda_i = \frac{\lambda_p}{\lambda_a} \quad (2.17)$$

Siendo λ_p la lambda de Wilks cuando la variable está en el modelo y λ_a cuando no está.

Podemos definir estadísticamente F para el caso de una sola variable como:

$$F = \frac{\text{varianza entre clases}}{\text{varianza dentro de las clases}} \quad (2.18)$$

Cuanto mayor es F , más poder de separación tiene dicha variable. Del mismo modo que para la lambda de Wilks se puede calcular una F parcial para cada variable, encontrando el cociente entre la F del modelo posterior y anterior a la inclusión de la misma o a la exclusión de la variable del modelo.

El análisis discriminante paso a paso tiene como objetivo crear una combinación lineal de variables que maximiza la razón entre la varianza entre clases y la varianza dentro de las clases. Se comienza eligiendo la variable que mejor discrimina (mayor F). Luego se emparejará ésta

con cada una de las restantes y se elige la combinación que mejor discrimina (mayor F o mayor λ_i) repitiéndose el proceso con tres variables y continuando del mismo modo sucesivamente. A medida que se van incluyendo nuevas variables, otras pueden salir del modelo, si la información que contienen puede ser explicada mediante una combinación lineal de las variables incluidas en el modelo, es decir, tiene una F para exclusión mayor que un valor prefijado. El proceso se termina cuando las variables que no están en el modelo no añaden capacidad discriminativa al clasificador, porque las F para la inclusión son inferiores generalmente a cierto valor prefijado. Al final del proceso se consigue una combinación lineal que incluye menos variables que el conjunto inicial, pero con similar capacidad de discriminación.

Clasificadores basados en redes neuronales

El estudio y diseño de las redes neuronales fue motivado por las redes neuronales biológicas y son buenas herramientas para clasificación y clustering. Estas simulan una estructura computacional interconectada y en paralelo, cuyos componentes más elementales se denominan neuronas que, por lo general se agrupan en capas (ver sección 3.1).

Las redes neuronales se han descrito como sistemas estadísticos adaptativos y no lineales debido a que su comportamiento está basado en los puntos siguientes:

Estadísticas: Su comportamiento es muy bueno si los datos que se emplean lo son. Estos deben de ser representativos para el problema que se intenta resolver.

Adaptativas: Se pueden adaptar o pueden aprender de los datos que se introducen como patrones de entrenamiento mediante la modificación de los pesos entre las conexiones.

No lineales: No están limitados en asociaciones lineales entre las señales de entrada y salida.

El estudio, diseño y análisis de las redes neuronales han llegado a ser un campo muy activo y están siendo ampliamente aplicadas para resolver problemas de visión por computadora. Aunque existe una gran variedad de topologías, hay algunas que son más especializadas que otras. Un ejemplo claro es la red *Neocognitron* que ha sido ampliamente usada para el reconocimiento de patrones.

Capítulo 3

Redes neuronales

En este capítulo se hace una introducción sobre redes neuronales, así como una reseña histórica donde se enumeran los hechos más importantes desde su inicio hasta la actualidad. Se describe brevemente la primer red neuronal multicapa llamada retropropagación de errores y por último la descripción de la red neuronal de Kohonen. En particular los mapas autoorganizativos que es la que se usó en el presente trabajo.

3.1 Redes neuronales artificiales

Las redes neuronales están constituidas por un conjunto de neuronas unidas entre ellas mediante conexiones. A cada conexión se le asigna un peso numérico que indica el grado de sinapsis que pueden ser excitatorias, inhibitorias o irrelevantes. Los pesos constituyen el principal recurso de memoria a largo plazo de las redes neuronales y el aprendizaje usualmente se realiza con la actualización de éstos. Las redes neuronales están organizadas en subconjuntos denominadas capas. La capa de entrada contiene a las neuronas que reciben las señales del ambiente externo y las distribuye hacia otro grupo de neuronas; la capa de salida está formada por una o más neuronas que se encargan de emitir los resultados al mundo externo o hacia otras redes neuronales. Algunos tipos de redes neuronales tienen capas entre las dos anteriores y son denominadas como capas ocultas. Los pesos son modificados utilizando una regla de aprendizaje de tal manera que la conducta de entrada-salida de la red esté más acorde con la del ambiente

que produce las entradas. La regla de aprendizaje depende del tipo de red que se use.

Las neuronas constan de un conjunto de conexiones de entrada provenientes de otras unidades, un conjunto necesario de conexiones que van hacia otras unidades, un nivel de activación y recursos para calcular cuál será el nivel de activación del paso siguiente, con base en sus entradas y pesos respectivos. Lo importante es que en cada una de las unidades se efectúa un cálculo local con base en las entradas que le proporcionan sus vecinas, pero sin que sea necesario un control global en todo el conjunto de unidades.

Las redes neuronales artificiales presentan un conjunto de características que las hacen atractivas en la solución de problemas que para otras técnicas son complicadas. Las características son:

Paralelismo masivo: Una vez entrenada, la red neuronal puede trabajar muy adecuadamente en tiempo real. Si el problema es muy complejo, la red puede solucionarlo con un número adecuado de capas, por ejemplo 2 ó 3.

Tolerancia a fallas: El fallo de una neurona puede influir muy poco en el funcionamiento global, debido a que el funcionamiento depende de todas las neuronas.

Aprendizaje mediante ejemplos: Para que una red realice lo que se requiere, solo es necesario suministrarle un conjunto de ejemplos que debe ser lo suficientemente representativo para que ésta se entrene, a diferencia de las máquinas de Neuman que hay que programarlas metódicamente.

Adaptabilidad: La red aprende continuamente y se adapta a los casos nuevos que se le presentan. Con esto se logra que no solo responda bien a los ejemplos presentados o a los muy similares, sino a unos completamente nuevos. Esto se logra con un algoritmo de entrenamiento adecuado que debe de ejecutarse continuamente durante el aprendizaje.

Generalización: Una red entrenada con patrones adecuadamente seleccionados es capaz de responder a otros que no existían durante la fase de entrenamiento.

A continuación se relatan brevemente los principales acontecimientos históricos en el área.

- 1943. Mc-Culloch y Pitts (Mc-Culloch y Pitts, 1943) publican el artículo “A Logical Calculus of Ideas Immanent in Nervous Activity”. La colaboración de un neurobiólogo y

un matemático genera un modelo abstracto de neurona en el que la probabilidad de que ésta se activase dependía de la señal de entrada y de la sinapsis de conexión.

- 1949. Hebb publica el libro “The organization of the Behavior” (Hebb, 1949) donde se describe como pueden aprender las neuronas. Si la neurona x y la neurona y estaban continuamente activadas, entonces se reforzaba la conexión sináptica de ambas, ésto fue llamado aprendizaje Hebbiano.
- 1951. Marvin Minsky y Dean Edmons fabrican con tubos, motores y dispositivos mecánicos una máquina capaz de aprender. Para ello se basaron en las ideas de MacCulloch y Pitts. Aunque rudimentariamente, esta máquina fue fascinante en su tiempo.
- 1956. Organizada por Minsky, John McCarthy, Nathaniel Rochester y Claude Shannon se celebró la primera conferencia sobre Inteligencia Artificial.
- 1959. Frank Rosenblatt (Rosenblatt, 1959) desarrolla su concepto de Perceptron. Un sistema que permitía interpretar patrones tanto abstractos como geométricos. El Perceptron supone un gran avance en la IA y el inicio de las redes neuronales artificiales.
- 1960. Widrow y Hoff. Desarrollan un modelo llamado ADALINE. Este modelo de redes es capaz de clasificar los datos en espacios linealmente separables. El ADALINE era extremadamente similar al perceptron y pronto se generalizaron estos resultados al modelo de Rosenblatt. ADALINE es una red neuronal entrenada por una regla del gradiente descendente para minimizar el error cuadrático medio.
- 1969. Minsky y Papert (Minsky and Papert, 1969) publican el libro llamado “Perceptron” en el que presenta el principal problema del perceptron, el problema del XOR o el No exclusivo. El XOR es un operador lógico en el que la conclusión es falsa cuando ambos son falsos o verdaderos, en caso contrario es verdadero. Este problema no es linealmente separable y el Perceptron fue incapaz de resolverlo. Lo anterior produjo que se redujera la investigación en las redes neuronales artificiales.
- 1986. Galland introduce el desarrollo de sistemas híbridos, combinando redes neuronales y componentes no conexionistas.

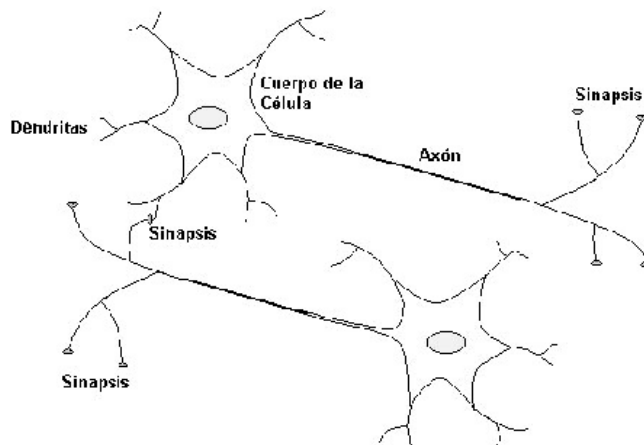


Figura 3.1: Neuronas biológicas.

- 1988. Baum y Haussler llevan a cabo análisis teóricos para establecer cuando las redes neuronales pueden brindar una solución aproximadamente correcta con una gran probabilidad.

3.2 Estructura y función de una simple neurona

3.2.1 Neurona biológica

Una neurona biológica típica está compuesta por una *soma o cuerpo*, un *axón* y *dendritas* (ver figura 3.1).

La parte central de la célula es llamada cuerpo de la célula o *soma*, de él se proyectan numerosas extensiones en forma de raíz llamadas *dendritas*. La neurona también está formada por una simple fibra tubular, el *axón*, el cual se divide en numerosas ramificaciones. El tamaño típico del soma es de unos 10 a 80 micrómetros. Las dendritas y el axón tienen un diámetro de apenas unos cuantos micrómetros.

El propósito de las dendritas es servir como elementos receptores para las señales provenientes de neuronas adyacentes.

Las dendritas y el soma reciben mensajes de otras células. La información es procesada y el

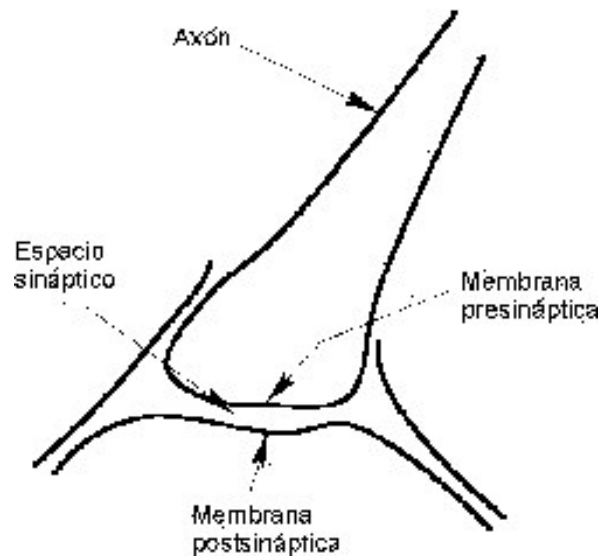


Figura 3.2: Sinapsis.

mensaje resultante, si lo hay, pasa a lo largo del axón a otro conjunto de neuronas, por lo que el propósito del axón es la transmisión de la actividad neuronal generada a otra célula nerviosa o fibra muscular. Para el caso de las células nerviosas se utiliza el término interneuronal, mientras que para las fibras musculares se emplea el nombre de neurona motora. Existe un tercer tipo de neuronas, las que reciben la información desde los músculos o de los órganos sensoriales, las cuales son llamadas neuronas receptoras.

La unión entre el final de una rama del axón, la cuál tiene forma de plato, y otra neurona o músculo es llamada "sinapsis" (ver figura 3.2). Cada sinapsis entre dos células pueden ser excitatorias o inhibitorias y está separada por una pequeña brecha de aproximadamente 200 nanómetros de ancho. Dicha brecha es conocida como *brecha sináptica*. Existen dos estructuras relacionadas con la sinapsis denominadas *presinapsis* y *postsinapsis* en relación a si se encuentran antes o después de la unión sináptica. Las sinapsis pueden estar localizadas en cualquier parte del cuerpo de la célula o en las dendritas. El grado de sinapsis generalmente disminuye con el incremento de la distancia, el cual se mide desde el cuerpo de la célula hasta la unión sináptica.

La señal nerviosa es transmitida por medio de una reacción electro-química. La transmisión eléctrica prevalece en el interior de la neurona, mientras que el mecanismo químico opera entre neuronas, por ejemplo en las sinapsis.

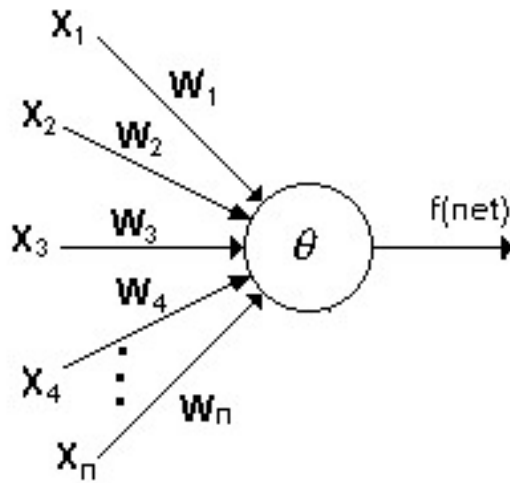


Figura 3.3: Modelo de neurona artificial.

3.2.2 Modelo de una neurona artificial

Este modelo está basado en las principales características de la neurona biológica. Cada unidad desempeña un trabajo relativamente simple: recibe entradas de sus vecinas o de fuentes externas y usa estas señales para calcular una salida, la cual es propagada hacia otras neuronas (ver figura 3.3). Aparte de este proceso, su segunda tarea es la de ajustar sus pesos. El sistema es inherentemente paralelo en el sentido de que muchas unidades pueden realizar el cómputo al mismo tiempo.

En los sistemas neuronales artificiales se pueden distinguir tres tipos de unidades o neuronas: *unidades de entrada*, las que reciben los datos del exterior de la red neuronal. *Unidades de salida*, las que envían los datos resultantes hacia afuera de la red. *Unidades ocultas*, las cuales reciben señales de entrada de neuronas de la misma red y la salida es también utilizada por otras neuronas de la red.

3.3 Tipos de aprendizaje en las redes neuronales

Dentro de las redes neuronales artificiales se pueden crear dos grandes grupos de acuerdo a la forma en que se realiza el aprendizaje. Si la red recibe información sobre si su funcionamiento

es correcto o incorrecto, basado en alguna medida de error entre la salida generada por la red y la salida deseada para un determinado patrón de entrada, entonces, se está usando una red con *aprendizaje supervisado*. Si la red no cuenta con un maestro y aprende heurísticamente examinando las muestras de datos o el medio ambiente, se está hablando de una red con *aprendizaje competitivo* [14].

3.4 Un modelo básico: El *Perceptron*

El perceptron, fue expuesto en 1959 por Frank Rosenblatt (Rosenblatt, 1959). El modelo está compuesto de dos capas: una capa de entrada de datos y una capa de salida con una única neurona (ver figura 3.4).

El aprendizaje en un perceptron es extremadamente sencillo, la capa de salida suma todas las entradas ponderadas por sus pesos. Al resultado de esa suma se le aplica una función escalón dando como resultado 1 si es mayor que cero y -1 en caso contrario.

El aprendizaje se realiza con la actualización de los pesos con la siguiente regla:

$$W_{ij}^{t+1} = W_{ij}^t + \Delta W_{ij} \quad (3.1)$$

donde:

$$\Delta W_{ij} = \alpha(O_{jp}^* - O_{jp})O_{ip} \quad (3.2)$$

O_{jp}^* = La salida deseada.

O_{jp} = La salida real.

O_{ip} = Valor que llega de una neurona de la capa anterior.

α = factor de aprendizaje.

Esta es la regla de aprendizaje más básica de un perceptron. Lo importante del perceptron es que con esta regla extremadamente simple es posible clasificar datos linealmente separables. De hecho, un perceptron puede ser considerado como una recta de regresión donde los coeficientes de la recta son los pesos de las conexiones.

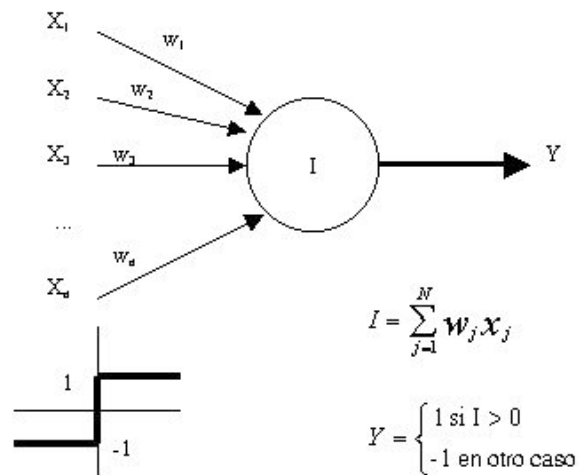


Figura 3.4: Perceptron y las funciones de activación y de transferencia.

Algoritmo:

- Activar el estrato de entrada de la red con un patrón de la muestra.
- Propagar la activación por toda la red hasta obtener la salida.
- Aplicar la regla de aprendizaje (fórmula 3.1) para la modificación de los pesos en la neurona o neuronas de salida a partir del patrón presentado.
- Decrementar α en un valor pequeño.
- Evaluar el criterio de parada. Si se cumple el criterio, entonces detener el entrenamiento, en caso contrario, repetir el algoritmo.

3.5 Red de Retropropagación de Errores

La red con retropropagación de errores (*backpropagation*) está formada por una capa de unidades o neuronas de entrada, una capa de neuronas de salida y cero o más capas ocultas o internas de neuronas. Las capas consecutivas están conectadas completamente (*fullyconnected*), es decir, cada unidad de la capa k recibe entradas de cada una de las unidades de la capa $k-1$. Las conexiones van siempre hacia adelante (*feedforward*), en la dirección entrada-salida (ver figura 3.5). Cada neurona calcula su salida (excepto en la capa de entrada, debido a que únicamente se

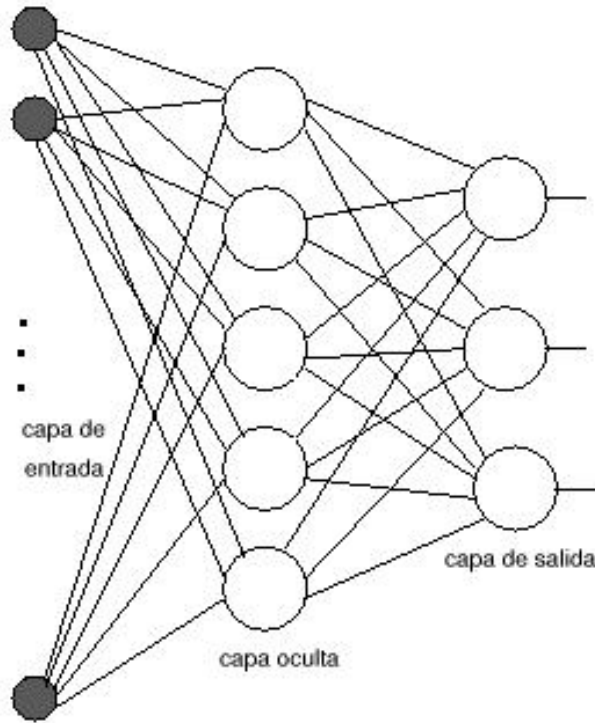


Figura 3.5: Arquitectura de una red de retropropagación de errores.

encarga de proyectar las señales de entrada hacia la siguiente capa) como una suma ponderada por pesos $\omega_1, \dots, \omega_n$ de las n entradas X_1, \dots, X_n que recibe, a la que se aplica una función real f , llamada *función de activación* (fórmula 3.3):

$$net = \sum_{i=1}^n X_i \omega_i \quad (3.3)$$

Los valores de los pesos pueden ser fijos u obtenerse por medio de un proceso llamado aprendizaje que minimiza cierto criterio de error. Esta red realiza un aprendizaje supervisado, es decir, las salidas que produce la red en la capa de salida deben de ser lo más parecidas posible a unas salidas prefijada para unos patrones de entrada. La red busca ajustar lo mejor posible (mínimos cuadrados) una función prefijada $O_i = F(y_i)$, con $O_i \in \mathfrak{R}^m$, $y_i \in \mathfrak{R}^n$, $1 \leq i \leq p$ si hay p patrones, n unidades de entrada y m de salida. Por lo tanto, el aprendizaje de la red debe llevar los pesos a unos valores tales que $E_p = \frac{1}{2} \sum_k (O_{pk} - F(y_p))^2$ sea mínimo. Esta red aprende utilizando una regla de aprendizaje llamada *regla delta generalizada*.

La red es capaz de construir una representación interna por medio de sus pesos que puede aproximar la función $F : \Re^n \longrightarrow \Re^m$.

La aplicación de la regla delta generalizada envuelve dos fases:

- Durante la primer fase, la entrada X es presentada y propagada a través de la red para computar el valor de salida O_i de cada unidad. La salida es comparada con el valor de salida deseada O_i^* resultando un error de salida δ_i para cada unidad de salida.
- La segunda fase involucra una pasada hacia atrás en toda la red durante el cual el error es pasado por cada unidad y los apropiados cambios en los pesos son calculados.

La regla de aprendizaje (delta generalizada) para la red de retropropagación de errores es:

$$\omega_{ij}^{l,t+1} = \omega_{ij}^{l,t} + \alpha \phi_j^{l+1} O_i^l \quad (3.4)$$

donde:

$$\phi_j^l = \begin{cases} (O_i^{*l} O_i^l) (F_i^l(A_i^l))', & l=\text{índice del estrato de salida.} \\ (F_i^l(A_i^l))' \sum_j \phi_j^{l+1} \omega_{ij}^l & \text{para los estratos ocultos} \end{cases} \quad (3.5)$$

$$(F_i^l(A_i^l))' = O_p^k (1 - O_p^k) \quad (3.6)$$

La función de transferencia $f(net)$ para cada una de las unidades de las capas ocultas y la capa de salida se usa la función sigmoide, representada por la ecuación 3.7. (ver figura 3.6).

$$O_i^l = f(net) = \frac{1}{1 + \exp^{-net}} \quad (3.7)$$

El algoritmo para realizar el entrenamiento de la red de retropropagación de errores es el siguiente:

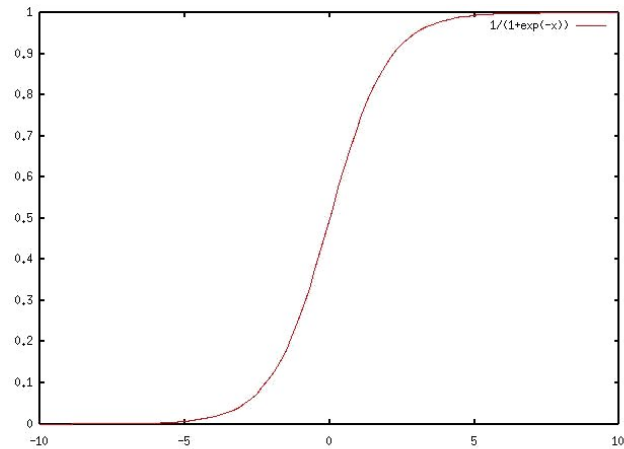


Figura 3.6: Gráfica de la función sigmoide.

Algoritmo:

- Activar el estrato de entrada de la red con un patrón de la muestra.
- Propagar la activación por toda la red hasta obtener la salida O_i^l .
- Calcular el valor de ϕ para todas las neuronas de la capa de salida.
- Calcular el valor de ϕ para todas las neuronas de las capas inferiores u ocultas, guardando la relación de orden descendente respecto al índice del estrato.
 A este proceso se le conoce como retropropagación de ϕ . El mismo puede ser efectuado ya que el cálculo de ϕ para una capa superior posibilita su cálculo para la capa inmediatamente inferior. El resto de la información necesaria ya está contenida en las conexiones y en las activaciones de las neuronas al presentarse el patrón de entrada.
- Actualizar los pesos con la regla de aprendizaje.
- decrementar el factor de aprendizaje en un valor pequeño.
- Si el criterio de parada no se cumple, repetirlo. En caso contrario terminar.

3.5.1 Deficiencias de la Retropropagación de errores

A pesar del aparente éxito del algoritmo de aprendizaje de la Retropropagación de errores, hay algunos aspectos que hacen que el algoritmo no garantice ser universalmente útil. La deficiencia más difícil es el largo proceso de aprendizaje. Este puede ser el resultado de un factor de aprendizaje no óptimo. Durante el entrenamiento de la red, los pesos pueden ser ajustados a valores grandes. La entrada total de una unidad oculta o unidad de salida puede por consiguiente tener valores muy altos porque de la función sigmoide se tendría un valor muy cercano a cero o muy cercano a uno para la unidad. La otra deficiencia con que se encuentra es el mínimo local. La superficie de error de una red compleja está llena de colinas y valles y el gradiente descendente puede quedar atrapado en un mínimo local cuando hay un mínimo profundo cercano.

3.6 Mapas Autoorganizativos de Kohonen

Para los sistemas que procesan información, los mapas son una parte importante tanto para los sistemas naturales como para los artificiales. Ejemplos de mapas en el sistema nervioso son los mapas retinotópicos (mapas en la corteza visual), mapas tonotópicos (mapas de la corteza auditiva) y mapas somatotópicos (sensibilidad en la piel) [3]. Los mapas auto-organizativos son un método de aprendizaje no supervisado para la clasificación de datos multidimensionales y fueron introducidos por Teuvo Kohonen inspirado en el comportamiento del cerebro humano. La principal característica de un SOM¹ es la creación de mapas bidimensionales donde los patrones de entrenamiento que estén más relacionados aparecerán más próximos, es decir, la similitud en los patrones de entrada es preservada en la salida del proceso que realiza la red. La preservación topológica de un SOM, lo hacen especialmente usado en la clasificación de datos que contienen un gran número de clases.

Este modelo está basado en las propiedades que presenta el cerebro humano, donde ciertas partes del cerebro son excitadas o producen una respuesta para una determinada característica en la señal de entrada. Cada neurona está conectada con otras por su entorno de manera que produce una excitación en las más próximas y una inhibición en las más alejadas produciendo una interacción lateral. Tanto la excitación como la inhibición lateral son gradualmente más

¹Por las siglas en inglés de *Self Organizing Map*.

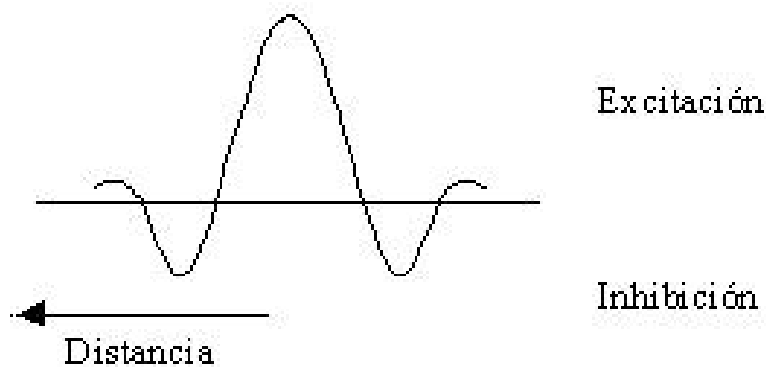


Figura 3.7: Función que define el tipo de interacción entre neuronas.

débiles a medida que nos alejamos de la neurona que está respondiendo más fuertemente a la señal de entrada. La función que establece esta interacción lateral viene definida por una función que tiene la forma del sombrero Mexicano (ver figura 3.7).

Esta red se autoorganiza debido a que los nodos tienden a alcanzar vectores de pesos que capturan las características en el espacio del vector de entrada. En el *clustering o agrupamiento*, los vectores de pesos asociados con nodos en estas redes son interpretados como el centroide del cluster. En el contexto de *vector quantization*, cada vector es un vector del codebook en donde los vectores de entrada pueden ser mapeados.

La topología de la red se define de la siguiente manera: La primer capa debe contener n neuronas de entrada, donde n corresponde a los datos de entrada para la red. La capa de salida, que es la competitiva, puede ser n -dimensional. Comúnmente se utiliza bidimensional y puede ser rectangular o hexagonal. La elección de una u otra topología tiene un efecto despreciable sobre los resultados (ver figura 3.8).

El objetivo de la red es el de adaptar sus pesos de manera que cada unidad esté especialmente sensibilizada a un dominio de la señal de entrada. El proceso consiste en presentar un estímulo y propagarlo a través de la red. Según una métrica (por ejemplo *Euclidiana*. Ecuación: 2.1), elegir como ganadora a la célula que produzca una mejor señal en la capa de competición.

Se pretende que la red responda a estímulos similares de manera muy parecida. Para lograr lo anterior es necesario utilizar una regla de aprendizaje que es la hebbiana [Hebb, 1949]. El

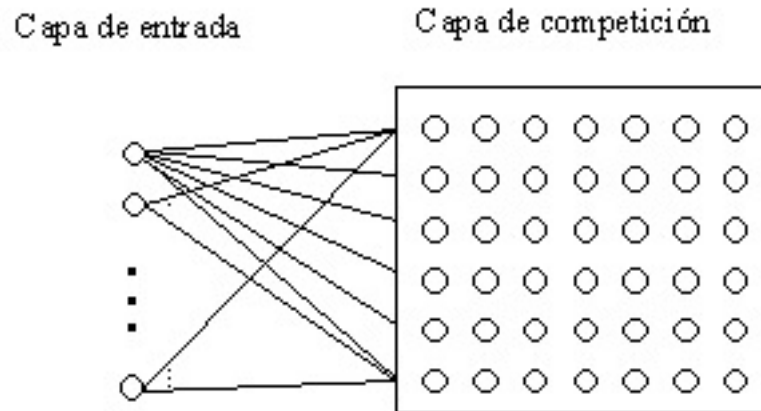


Figura 3.8: Topología de la red de Kohonen.

algoritmo cambia los vectores de pesos para que se ajusten mejor a las señales de entrada en una cantidad proporcional a la distancia que tengan con la neurona ganadora. Es decir, se modifican los pesos de la neurona ganadora y todos los vectores de pesos pertenecientes a las neuronas que se encuentran en la vecindad. El alcance de la función de vecindad comienza siendo grande y va decreciendo al paso del tiempo de entrenamiento.

Como en el Perceptron y la Retropropagación de errores, la red de Kohonen tiene un factor de aprendizaje que se encarga de modificar la cantidad en que los vectores de pesos cambiarán. En un principio este factor es grande pero va disminuyendo progresivamente hasta llegar a ser muy pequeño, muy similar a la función de vecindad.

Los valores correspondientes a los vectores de pesos son inicialmente asignados de manera aleatoria y como se dijo anteriormente, éstos cambian conforme la red se va entrenando y las neuronas se especializarán en algún subconjunto de los patrones de entrada. El número de nodos en la capa de salida debe ser mayor que el número máximo de posibles clusters del problema pero más pequeño que el número de muestras de entrenamiento [14].

Algoritmo de mapas autoorganizativos de Kohonen:

- Inicializar los pesos de la red con valores pequeños.
- **Hacer**
 - Seleccionar un patrón de entrada.
 - Calcular las distancias entre el vector de pesos de cada neurona y el patrón de entrada utilizando la distancia Euclidiana (ver ecuación 2.1).
 - Encontrar a la neurona ganadora (con la menor distancia).
 - Actualizar los pesos de la neurona ganadora y la de sus vecinas.
$$\text{peso}_i^{(t+1)} = \text{peso}_i^{(t)} + \alpha * \text{vecindad}(i, k) * (x^{(t)} - \text{peso}_i^{(t)})$$
donde:
 α = factor de aprendizaje.
 $\text{vecindad}(i, k)$ = función que determina el radio en que las neuronas vecinas actualizan sus pesos tomando como base la neurona ganadora.
 - Decrementar α a cierto número de épocas.
 - Decrementar el tamaño de la vecindad a cierto número de épocas.
- **Mientras no se obtengan las salidas deseadas**

Se ha presentado en este capítulo una introducción a las redes neuronales artificiales para que el lector tenga una visión general de como están constituidas, cual es su funcionamiento y como se utilizan. Se incluyeron las primeras redes más importantes como el Perceptron y la red de Retropropagación de Errores que fueron las que desencadenaron la investigación en este campo prometedor, sobre todo la segunda. Aunque en el presente trabajo la red utilizada fue la de Kohonen.

Hasta este capítulo se ha descrito de forma general algunas técnicas de procesamiento de imágenes y de clasificación y de éstas algunas son utilizadas en el presente trabajo y que se verán a más detalle en el siguiente capítulo.

Capítulo 4

Implementación del sistema

En el presente capítulo se describe de manera más precisa la metodología empleada para autenticar personas utilizando la geometría de las manos en una aplicación biométrica que es el objetivo principal de este trabajo. Las herramientas utilizadas y descritas de manera general son, en primer lugar, las técnicas de visión por computadora usadas para extraer las características de las manos de cada individuo. En segundo lugar, para la clasificación, es utilizada una red neuronal, específicamente los mapas autoorganizativos de Kohonen.

4.1 Metodología Empleada

El problema consiste en autenticar a un individuo tomando en cuenta las dimensiones físicas de la mano derecha del mismo. La clasificación entre manos de diferentes individuos se vuelve interesante debido a que se puede encontrar una gran cantidad de clases y la distancia entre éstas es muy cercana lo que hace que la clasificación sea más complicada. Este es uno de los principales problemas que se enfrentó debido a la falta de variabilidad discriminante de las manos, esto no quiere decir que no sea factible desarrollar una aplicación biométrica. Si se desea tener más seguridad, se puede combinar con otro identificador biométrico. Por lo anterior, se buscan características que sean lo suficientemente discriminantes.

Cada mano de las personas es única debido a las diferentes medidas y posiciones de las



Figura 4.1: Prototipo donde los usuarios ubican las manos para capturar las imágenes.

características usadas. Sin embargo, estas medidas no varían significativamente entre la población, pero debido a que no se requieren aparatos sofisticados pueden usarse para aplicaciones que requieren una verificación rápida y a un bajo costo. Las manos son las que tenemos más inmediatamente disponibles para obtener una imagen y la que requiere menos procesos para poder extraer datos característicos a diferencia de otras partes del cuerpo humano que son usados en biometría.

Para este trabajo únicamente se tomaron medidas de los dedos, de la palma de la mano y la distancia entre la parte baja entre los dedos pulgar e índice y la punta del dedo índice, además de la dimensión de la mano cuantificada en pixeles. De los dedos se tomaron medidas del largo y el ancho de cada uno, en dos de sus partes específicas. De la palma de la mano se tomaron dos medidas en el ancho de ésta. Por el momento, no se tomaron en cuenta otras características de las manos como son el color de la piel, textura y el grosor de los dedos porque creemos que las medidas seleccionadas son las más discriminatorias. Los resultados arrojados por estas medidas forman un conjunto de datos unidimensionales. Los vectores obtenidos de cada una de las manos se utilizaron como patrones de entrenamiento para la red neuronal.

4.1.1 Adquisición de Imágenes

Las imágenes fueron tomadas con una cámara fotográfica digital Canon Power Shot A50 sin ningún tipo de adecuación en el medio ambiente. Solamente se usaron lámparas comunes de alumbrado casero. Los puntos de control fueron puestos estratégicamente para que el usuario ubicara la mano correctamente sobre la superficie designada para tal propósito, tratando de evitar una rotación y traslación fuerte (ver figura 4.1). Las imágenes originales tomadas por la cámara fueron 640 x 480 en RGB, posteriormente fueron convertidas a escala de grises. Estas son las las imágenes de las que finalmente se procesan para obtener los datos que forman el vector de características de las manos. Se lograron eliminar los puntos de control del prototipo pintándolos del mismo color que el fondo. El proceso de calibración tiene un efecto directo sobre las imágenes tomadas, es por eso que se adquirieron las imágenes a la misma distancia y con la misma cámara.

4.1.2 Preprocesamiento

Ecualización de histograma

La mayoría de las imágenes obtenidas para la realización de este trabajo tenían los niveles de grises mal distribuidos como producto de una mala iluminación, es decir, con pobre contraste. Por consiguiente, se realizó una ecualización del histograma para realizar una redistribución de los valores de grises de la imagen y de esta manera obtener mejores resultados al ser convertida a binaria utilizando un umbral. Este método ayudó mucho a la mejora de la imagen debido a que a los valores blancos (fondo) son vueltos más blancos y los negros (mano) más negros.

Si los pixeles de la imagen se encuentran distribuidos en el rango $[X_1, X_2]$ y se quieren expandir al rango $[X'_1, X'_2]$, la fórmula utilizada para mapear los valores de cada uno de los pixeles p del rango original a un nuevo valor de p' dentro del nuevo rango es:

$$p' = \frac{X'_2 - X'_1}{X_2 - X_1}(p - X_1) + X'_1 \quad (4.1)$$

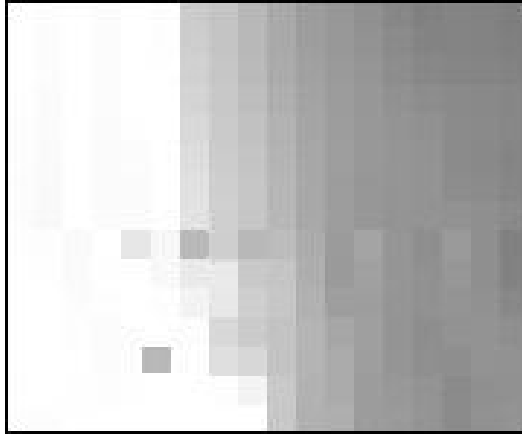


Figura 4.2: Ruido existente en la imagen.

$$= \frac{X'_2 - X'_1}{X_2 - X_1}p + \frac{X'_1X_2 - X'_2X_1}{X_2 - X_1}$$

Filtrado de la imagen

Las imágenes obtenidas son corrompidas por el ruido, que son variaciones aleatorias en los valores de intensidad. Algunos tipos comunes de ruido son *sal y pimienta*, *de impulsos* y *ruido gaussiano* [9]. El ruido sal y pimienta son las ocurrencias aleatorias en los valores de las intensidades tanto en las blancas como en las negras. El ruido de impulsos contiene solo ocurrencias aleatorias en los valores de intensidades de blanco. El ruido gaussiano contiene variaciones en intensidades con distribución gaussiana o distribución normal.

De las imágenes obtenidas para este estudio, se encontró que el tipo de ruido que mayormente se presentaba era el de *impulsos* (ver imagen 4.2) por lo que se optó por el *filtro del promedio* (ver sección 2.2.2). Debido a que las imágenes con que se trabajó, solamente constan del objeto (la mano) y el resto forma parte del fondo, este filtro dio buenos resultados para eliminar el ruido por impulsos y a la vez resaltar el objeto de interés. Como este filtro típicamente usa el mismo patrón de pesos para cada ventana, significa que este filtro lineal es espacialmente invariante y puede ser implementado usando una máscara de convolución. Para la aplicación del filtro se utilizó una máscara de 8 vecinos del pixel $h[i, j]$ (ver figura 2.2), lo que permitió disminuir o aumentar el valor de éste. El filtro ayudó en resaltar adecuadamente el objeto,

es decir, poder identificar muy bien los límites entre el objeto y el fondo para posteriormente obtener una imagen binaria de calidad.

4.1.3 Segmentación

Una vez que el ruido fue eliminado en la imagen, se procedió a segmentarla utilizando un umbral simple (ver sección 2.2.3), para generar una imagen binaria (ver figura 4.3(d)), la cual finalmente se utilizó para extraer un vector de características que sirvió para registrar o autenticar a la persona. Realizando múltiples pruebas con la segmentación de imágenes, se llegó a la conclusión de que el valor más adecuado para el umbral era de 130, lo que permitió binarizar la imagen adecuadamente.

4.1.4 Seguimiento del contorno

Como se puede ver en la imagen segmentada (ver figura 4.3(d)), el objeto de interés (la mano) es completamente identificable debido al contraste con el resto de la imagen (fondo), lo que hace muy fácil el seguimiento del borde del objeto. Para este trabajo se utilizó un algoritmo de seguimiento de contorno (ver sección 2.2.3).

En la práctica el ruido puede dar lugar a valores muy diferentes de intensidad, deformando el objeto debido a las fluctuaciones que producen los valores altos de intensidad. Para poder obtener un contorno de calidad fue necesario realizar los siguientes procesos:

1. Aplicar el algoritmo de seguimiento del contorno (ver sección 2.2.3). Se obtuvo una imagen con solo el límite del objeto. Por el ruido que no fue posible eliminar en el proceso de filtrado, se generó un contorno poco uniforme (ver figura 4.4(a)).
2. Para hacer el contorno un poco más uniforme (ver figura 4.4(b)), se aplica el algoritmo de suavizamiento de contorno que consiste en tomar los píxeles vecinos al píxel $h[i, j]$ y se promediaron sus coordenadas para obtener la nueva posición de este píxel dentro del vector que define el contorno. Se resume el procedimiento con la fórmula 4.2. Este algoritmo no modifica la conectividad. En la figura 4.5 se puede ver el resultado.



(a) Imagen original.



(b) Imagen ecualizada.



(c) Imagen filtrada.



(d) Imagen Binaria.

Figura 4.3: Proceso que se realiza con las imágenes hasta convertirlas a binarias.

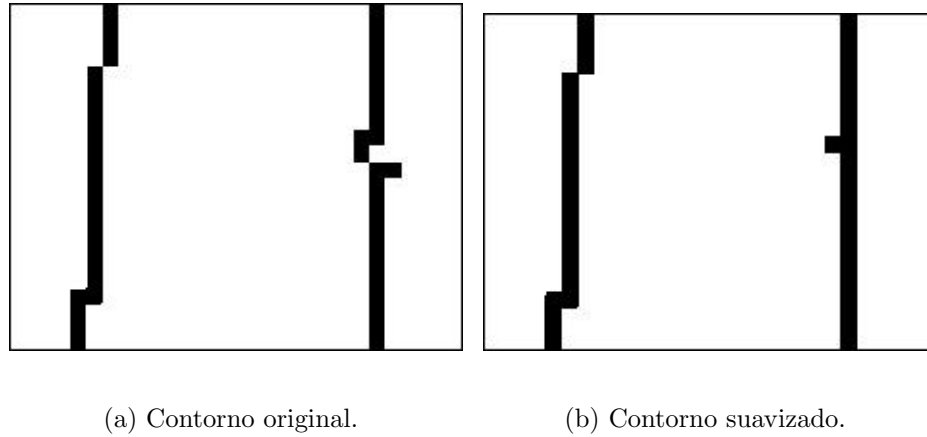


Figura 4.4: Contorno de la imagen antes y después de aplicar el algoritmo de suavizado del contorno.

$$i' = \frac{1}{3} \sum_{k=i-1}^{i+1} \text{coordenada } x \in k, \quad j' = \frac{1}{3} \sum_{l=j-1}^{j+1} \text{coordenada } y \in l \quad (4.2)$$

El objetivo primordial del seguimiento del contorno es el de encontrar las coordenadas donde la imagen llega a su máximo y vuelve a decrecer (punta de los dedos), es decir, donde haya un cambio en la dirección del contorno (ver figura 4.6(a)). De la misma forma, se encuentran las coordenadas donde el contorno del dedo baja y luego vuelve a subir (parte baja entre dedos). Una vez obtenidas estas coordenadas, es posible calcular la inclinación de cada uno de los dedos y su longitud (ver figura 4.6(c)).

Como las coordenadas que indican el cambio de dirección en la parte baja del dedo se encuentran sobre el contorno, se procede a calcular las nuevas coordenadas que nos indiquen el centro del dedo. La forma de encontrarlo consiste en recorrer la misma fila hasta encontrar el otro extremo del contorno, con esto obtenemos el número de píxeles que posteriormente se divide entre dos (ver figura 4.6(b)). El objetivo de lo anterior, es la de obtener la distancia más precisa al medir la longitud de los dedos. Una vez que se cuentan con los dos puntos, una que indica aproximadamente el centro del dedo en la parte baja y la otra en la punta, se procede a obtener la medida de éste, trazando una línea entre los dos puntos (ver figura 4.6(c)). El problema se limita a calcular el ángulo θ para formar el eje de mínima inercia de cada uno de los dedos

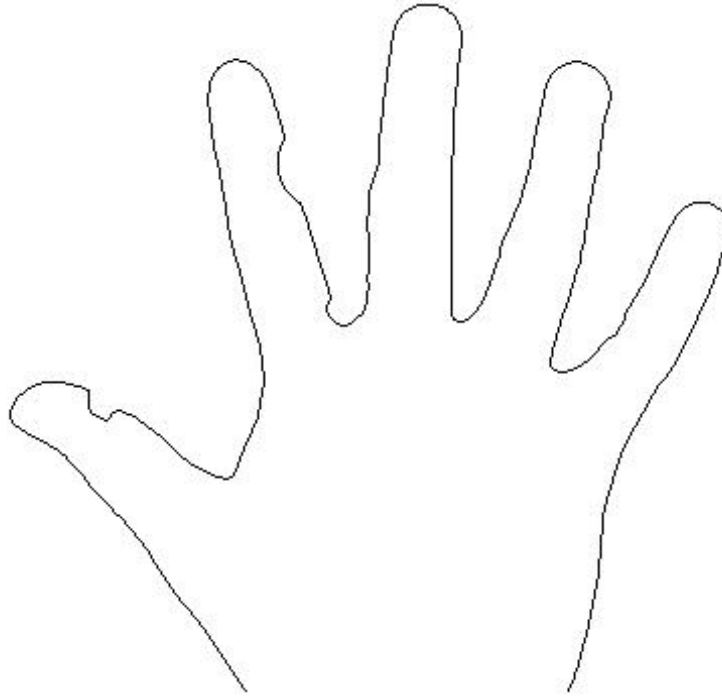


Figura 4.5: Contorno de la imagen procesada.

utilizando las coordenadas obtenidas.

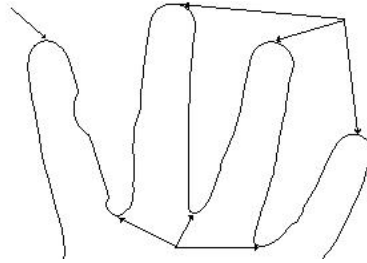
Así, básicamente se utiliza la ecuación de una recta que pasa por un punto genérico (α, β) formando un ángulo θ con un eje y viene dada por:

$$(y - \beta) \cos \theta = (x - \alpha) \operatorname{sen} \theta \quad (4.3)$$

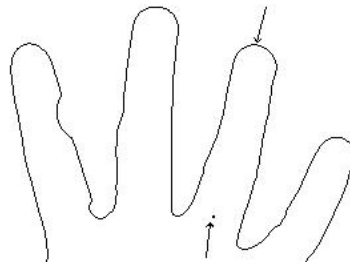
Para poder trazar la línea que indican los dos puntos obtenidos $(x_1, y_1$ y $x_2, y_2)$, fue necesario aplicar la fórmula 4.4 que da como resultado el grado de inclinación de la recta.

$$\theta = \tan^{-1} \left(\frac{y_2 - y_1}{x_2 - x_1} \right) \quad (4.4)$$

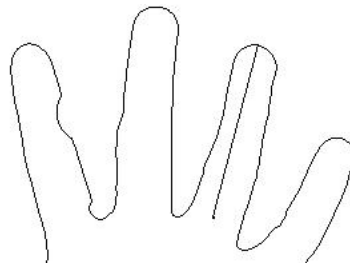
Una vez teniendo el grado de inclinación de la línea que pasa por el centro del dedo, fue necesario trazar dos líneas perpendiculares a la línea central. Toda línea perpendicular es definida de forma



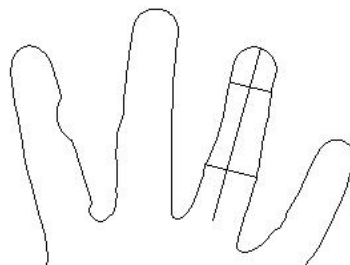
(a) Coordenadas donde existe cambio de dirección.



(b) Coordenadas que indican aproximadamente el centro del dedo.



(c) Largo del dedo.



(d) Ancho del dedo en dos diferentes lugares.

Figura 4.6: Pasos para obtener medidas de los dedos.

general por las fórmulas 4.5, donde θ el grado de inclinación de la recta de la cual se trazará la línea perpendicular.

$$\beta = \theta + 90^\circ \quad (4.5)$$

o

$$\beta = \theta - 90^\circ$$

Usando un punto fijo (x', y') sobre el contorno del dedo y con ayuda de las fórmulas 4.6 y 4.7 se puede encontrar las coordenadas del otro punto (x, y) de la recta perpendicular a la línea que pasa por el centro del dedo, ésta se corta cuando la línea se cruza con el contorno al lado opuesto. Lo anterior se hace con la finalidad de obtener medidas que indican el ancho de los dedos en dos lugares definidos (ver figura 4.6(d)).

Las fórmulas que obtienen el grado de inclinación de una perpendicular son las siguientes:

$$\tan \theta = - \left(\frac{x_2 - x_1}{y_2 - y_1} \right) \quad (4.6)$$

$$\theta = \tan^{-1} \left(- \frac{x_2 - x_1}{y_2 - y_1} \right)$$

Donde : (x_1, y_1) y (x_2, y_2) son los coordenadas de los dos puntos de la línea central de donde se definen las líneas perpendiculares.

Asimismo, para encontrar el punto (x, y) de la línea perpendicular una vez teniendo θ , son las siguientes:

$$y - y' = \tan \theta (x - x') \quad (4.7)$$

$$y = y' + \tan \theta(x - x')$$

$$x - x' = \frac{y - y'}{\tan \theta}$$

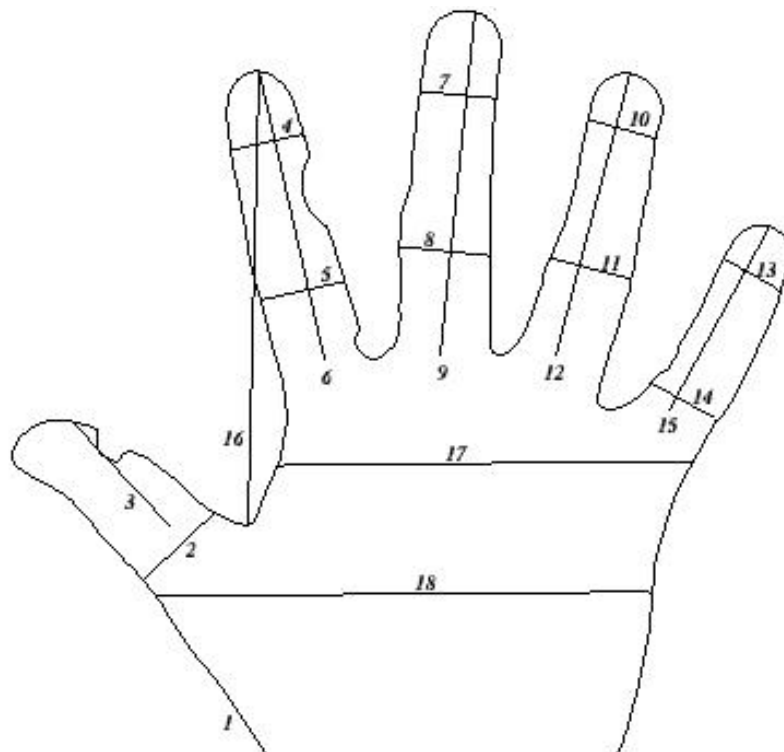
$$x = \frac{y - y'}{\tan \theta} + x'$$

Una vez que este proceso fue realizado para todos los dedos, se anexaron otras medidas como el número de píxeles que constituyen el contorno de la mano, la distancia entre la parte baja entre los dedos pulgar e índice y la punta del dedo índice y dos medidas más de la palma de la mano. Tomando en cuenta todas las medidas anteriores, se obtuvo un vector de 18 valores que representa las características que posteriormente se usaron como patrones de entrenamiento para la red de Kohonen (ver figura 4.7(b)). En la figura 4.7(a) puede observarse más claramente de que partes de la mano son extraídas dichas medidas.

4.1.5 Registro y autenticación

Una típica arquitectura de un sistema biométrico (ver figura 1.1) obtiene los patrones usando algún tipo de sensor. Una representación de la entrada adquirida es extraída usando un algoritmo de obtención de características y finalmente una decisión es tomada con base en la representación de entrada y los patrones de representaciones previamente almacenados en el sistema. Como puede verse el sistema consta de dos módulos: El módulo de registro (*o entrenamiento*) y el módulo de autenticación (*prueba, verificación o reconocimiento*). Se describe la forma en que estos módulos fueron implementados para este trabajo haciendo una equivalencia de términos usados en biometría y en redes neuronales.

La función del módulo de registro es la misma que el *entrenamiento o aprendizaje* en redes neuronales e incluso en reconocimiento de patrones. El registro o asociación de identidades



(a) Medidas obtenidas de la imagen.

1621	32	48	32	36	132	33	40	157	29	35	129	25	29	84	208	189	225
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18

(b) Ejemplo de un vector de medidas.

Figura 4.7: Medidas obtenidas por la aplicación. Cada línea representa una medida extraída de la mano que es representada por el vector de la figura (b) y se puede relacionar por el índice asociado.

de personas inicia con representaciones o vectores como resultado de las medidas biométricas. Cuando una señal de entrada es presentada debe ser alimentada al módulo de registro: Primero un extractor de características debe ser aplicado a la señal biométrica de entrada y una representación de las características es extraída (ver sección 4.1) y almacenada en la base de plantillas. En este caso particular, y dado que se usó una red neuronal, éstas son guardadas en forma de pesos entre las conexiones que relacionan a las neuronas de las diferentes capas.

El módulo de *autenticación* es similar al *reconocimiento o fase de prueba* en redes neuronales o reconocimiento de patrones. Este autentifica la identidad de la persona que intenta acceder al sistema. La persona a ser autenticada indica su identidad presentando al sistema su característica biométrica a ser medida (la mano); Un sensor se encarga de capturar la señal de entrada presentada por el usuario; el extractor de características se encarga de extraer una representación o un vector de datos que es comparado contra las representaciones de los individuos almacenados en el sistema para verificar la identidad. Si se encuentra un vector igual o muy similar con una distancia determinada, entonces, el usuario es aceptado. En caso contrario es rechazado.

Como se mencionó anteriormente, para la clasificación se utilizó una red autoorganizativa de Kohonen (ver sección 3.8). Para la capa de entrada se definieron 18 neuronas que es el mismo número de medidas obtenidas de las manos y que conforman los valores de un patrón de entrenamiento. La capa competitiva es rectangular donde se definieron 500 neuronas las que compiten entre sí para responder a un patrón de entrada determinado. Se definió esa cantidad previendo hasta 500 usuarios en el entrenamiento, es decir, si se desea agregar a un usuario más, únicamente se tiene que entrenar nuevamente a la red.

La métrica utilizada para definir a la neurona ganadora en la capa de salida fue la Euclidiana (ver sección 2.1). La neurona ganadora es la que se toma como referencia para definir la vecindad de neuronas que actualizan su vector de pesos incluyendo a ésta (ver figura 4.8).

Fase de registro

El proceso de registro fue hecho tomando una muestra de 5 imágenes de 30 personas. Primero, el individuo ubica la mano correctamente y se toma la primer imagen; el usuario remueve completamente la mano y vuelve a ubicarla y es tomada la segunda imagen y así sucesivamente

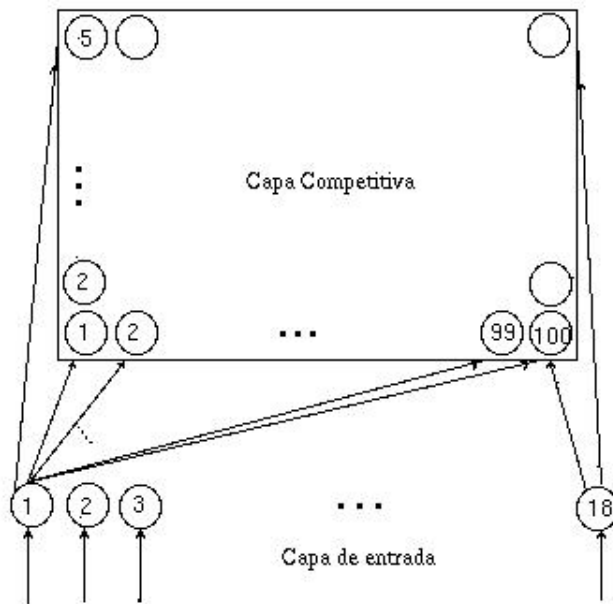


Figura 4.8: Topología de la red empleada.

hasta completar las cinco. Se tomaron tres imágenes aleatoriamente de las cuales se computó un vector de características de cada una. Una vez obtenidos los tres vectores, se procedió a promediarlos y obtener un vector resultado que es el que finalmente sirve como muestra o patrón de entrenamiento para la red, o lo que es lo mismo, el centro del cluster (ver figuras 4.9(a) y 4.9(b)).

Una vez que se tienen todas las muestras de cada uno de los individuos se procede a entrenar la red neuronal, haciendo modificaciones en los pesos de las neuronas en la capa competitiva, es decir, la neurona que tenga los pesos más parecidos al del vector de entrada será la que se entrene. Para este trabajo únicamente se entrenó a las neuronas vecinas a la ganadora y con el paso de las épocas los pesos fueron modificados en menor medida llegando el momento de que solo la neurona ganadora era modificada. Esta neurona ganadora es la que se encarga de responder más fuerte cuando un patrón de entrenamiento o muy parecido a éste es presentado a la red. Este proceso requiere muy poco tiempo de entrenamiento para todos los patrones. Una vez que la red ya cuenta con una neurona ganadora por cada patrón de entrenamiento se procede a inhibir en mayor medida a las neuronas perdedoras modificando los pesos con tendencia a cero.



(a) Diálogo que permite promediar los vectores de tres imágenes.

```

1541 28 46 28 34 130 28 35 154 24 29 119 20 22 63 204 171 198
1535 26 44 29 35 130 29 36 153 24 29 119 19 22 62 204 171 199
1544 28 45 28 34 130 27 35 153 23 27 119 19 22 62 205 171 200
█
1540 28 45 29 35 130 28 36 154 24 29 119 20 22 63 205 171 199

```

(b) Ejemplo del promedio de tres vectores.

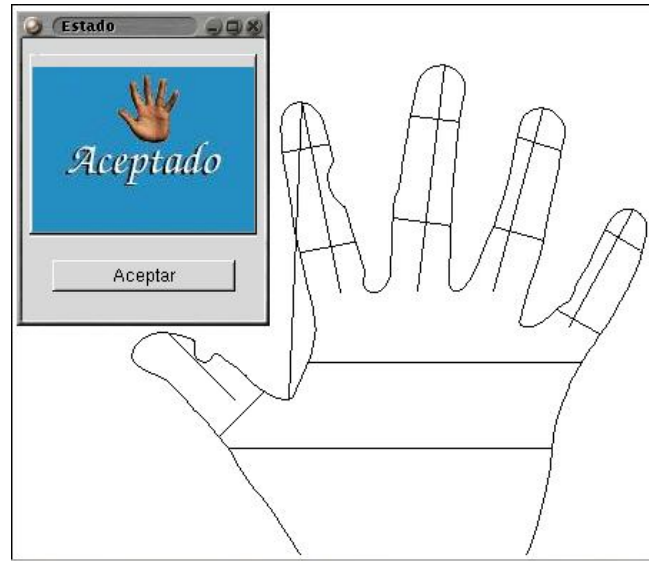
Figura 4.9: Creación de una muestra por cada individuo.

Fase de verificación

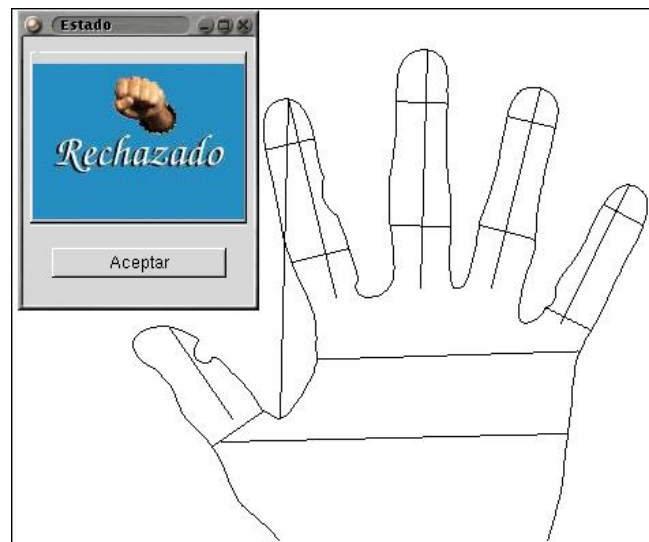
En la capa competitiva de la red neuronal existen neuronas ganadoras que son el centroide de cada una de las clases, en este caso las manos de cada uno de los individuos, por lo tanto, cuando se quiere identificar la mano de una persona, el vector de pesos de una neurona que se parece más al vector de datos¹ debe de responder más fuerte que el resto de las neuronas, debido al entrenamiento hecho en la primer fase de la red. Si la neurona que respondió coincide con alguna que respondió en la fase de registro, entonces, quiere decir que pertenece a una persona registrada y que cuenta con autorización para el acceso o uso de algún servicio restringido. En otro caso la petición de acceso es rechazada debido a que no responde ninguna neurona al vector de características de entrada por existir una distancia entre los vectores² que rebasa el umbral definido como distancia mínima (ver figuras 4.10(a) y 4.10(b)).

¹El vector corresponde a las medidas tomadas de la mano.

²Se compara un vector de entrada y el vector de pesos de la neurona que responde a la señal de entrada.



(a) Ejemplo cuando una muestra es aceptada.



(b) Ejemplo cuando una muestra es rechazada.

Figura 4.10: Dos únicas posibilidades que presenta el sistema.

Capítulo 5

Resultados y conclusiones

5.1 Resultados

Se realizaron diversos experimentos con las imágenes de las manos y aquí se presentan los más relevantes. Como se comentó anteriormente, para este trabajo se tomaron cinco imágenes de las manos de cada persona, en promedio. Este proceso fue hecho en condiciones ambientales de iluminación normal lo que dio como resultado imágenes con una cantidad considerable de ruido, por lo tanto, el proceso de identificación entre imágenes de manos de la misma persona se hizo más compleja. Cabe mencionar que se tomaron en cuenta, también, aquellas imágenes en donde el usuario no ubicó correctamente la mano sobre los puntos que servían como guía en el prototipo, existiendo una traslación de algunos de los dedos e incluso de la mano.

Experimento 1.

En este experimento se prueba la respuesta del sistema, registrando a todos los individuos y probándolo con imágenes pertenecientes a éstos y de esta manera obtener el porcentaje de falsos negativos.

Se tomaron a los 30 individuos que constituyen la muestra total y fueron registrados en el sistema. Posteriormente se probó la respuesta de la aplicación con 150 imágenes y con un

umbral cuyo valor fue de 17 tomado como base para la clasificación de acuerdo a una métrica vista en la sección 2.1.2. Con esta prueba sólo se logró obtener el porcentaje de falsos negativos, porque todas las imágenes pertenecían a los individuos registrados. El resultado arrojado por este experimento fue de 5 imágenes rechazadas; en términos de porcentajes es de 3.33% del total. Este experimento no es muy significativo porque en la vida real el sistema se probará con personas que no se encuentran registradas y por esa razón se realiza el experimento dos.

Experimento 2.

En este experimento se prueba la respuesta del sistema pero tomando en cuenta los dos tipos de errores en que toda aplicación biométrica incurre. Estos son el porcentaje de falsos negativos y falsos positivos. Las pruebas son realizadas dividiendo la muestra total y probadas con umbrales diferentes.

Del total de 30 individuos que constituyen la muestra, se dividió para poder obtener el porcentaje de falsos negativos y falsos positivos, distribuyéndolos de la siguiente manera: 10 individuos para registrarlos en el sistema y de éstos obtener el porcentaje de falsos negativos; 20 individuos para realizar las pruebas sobre los registrados y así obtener el porcentaje de falsos positivos. Lo anterior quiere decir que se usaron 50 y 100 imágenes respectivamente.

En este experimento el umbral determina si la distancia de un vector de características corresponde con alguna previamente registrada o no, es decir, si la distancia entre el vector de características de un usuario y un vector de pesos de alguna neurona, es menor o igual al umbral, entonces, el usuario es considerado como aceptado, en caso contrario, es rechazado. En la tabla 5.1 se hace un resumen de los resultados obtenidos modificando el umbral a diferentes valores y donde se puede observar que para un umbral con valor de 15 se obtiene el mejor resultado porque existe un equilibrio entre la tasa de falsos negativos y la de falsos positivos, esto es, tomando como referencia las aplicaciones civiles. Para otro tipo de aplicaciones debe ser considerado de forma diferente. En la figura 5.1 puede observarse de manera gráfica los umbrales más significativos y su comportamiento.

Realizando un análisis basado en la distancia entre los vectores de características que pertenecían a la misma persona y los que no, fue como se determinó el rango en donde el umbral podía recorrerse para realizar la clasificación. Si el umbral tomaba valores fuera de ese rango, los

Umbral	Falsos Positivos	Falsos Negativos	% F.P.	% F.N.
13	0	20	0.0	13.33
14	2	13	2.0	8.66
15	3	10	3.0	6.66
17	5	5	5.0	3.33
20	8	3	8.0	2.00

Tabla 5.1: Comparación de resultados con diferentes umbrales.

	Promedio
Promedio de distancia entre manos de la misma persona	8.75
Promedio de distancias entre manos de diferentes personas	31.16

Tabla 5.2: Distancias promedio entre manos de la misma y entre diferentes personas.

porcentajes se disparaban, tanto para la tasa de falsos positivos o para la tasa de falsos negativos. En la tabla 5.2 se pueden observar las distancias promedio entre manos del mismo individuo, así como la distancia entre diferentes usuarios. Como se puede observar en las figuras (5.2, 5.3, 5.4 y 5.5), algunos vectores pertenecientes a las manos se traslapan y esto trae como consecuencia que al definir un umbral pueden aceptarse individuos no estando registrados o viceversa. Es importante mencionar que en el eje Y se grafican las distancias entre los centroides de cada una de las manos, tanto de las registradas como de las no registradas y éstas son separadas por un umbral que es representado por una línea que las clasifica.

5.2 Discusión

Se ha presentado un proyecto para la autenticación de individuos utilizando la geometría de manos como tecnología biométrica, usando técnicas de visión por computadora para la extracción de características y para la clasificación de individuos mediante vectores de características se usó una red neuronal.

Las pruebas de datos biométricos es una tarea difícil, porque se envuelve porcentajes de error muy bajos. Para lograr un resultado estadístico creíble, deben ser examinadas miles de transacciones.

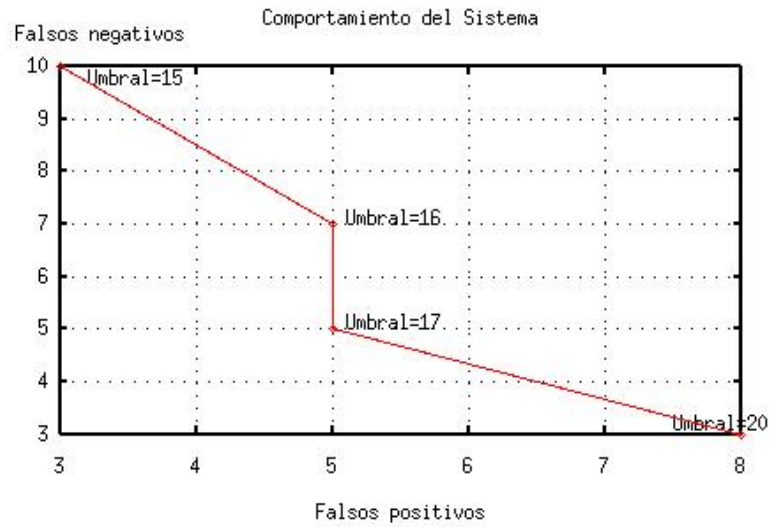


Figura 5.1: Comportamiento del sistema con la modificación del umbral en la etapa de autenticación.

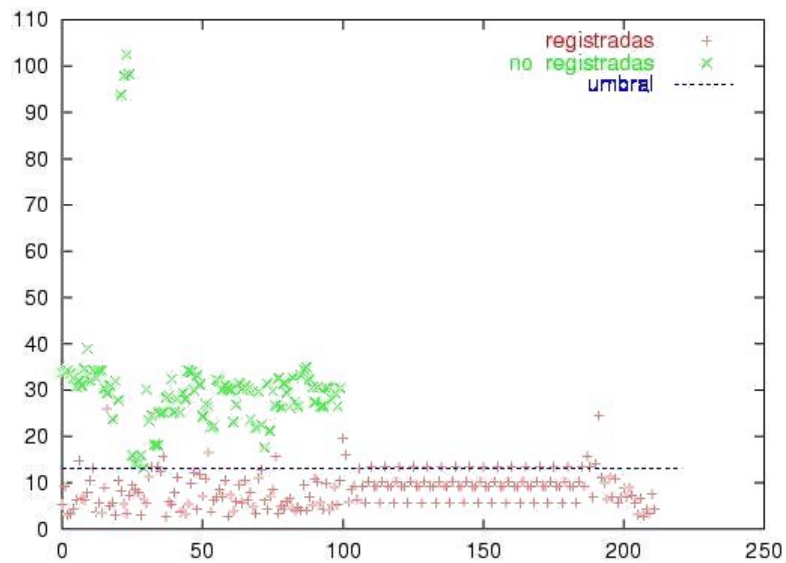


Figura 5.2: Fijando un umbral de 13, la tasa de falsos positivos es cero, pero la de falsos negativos se incrementa a un 13.33%.

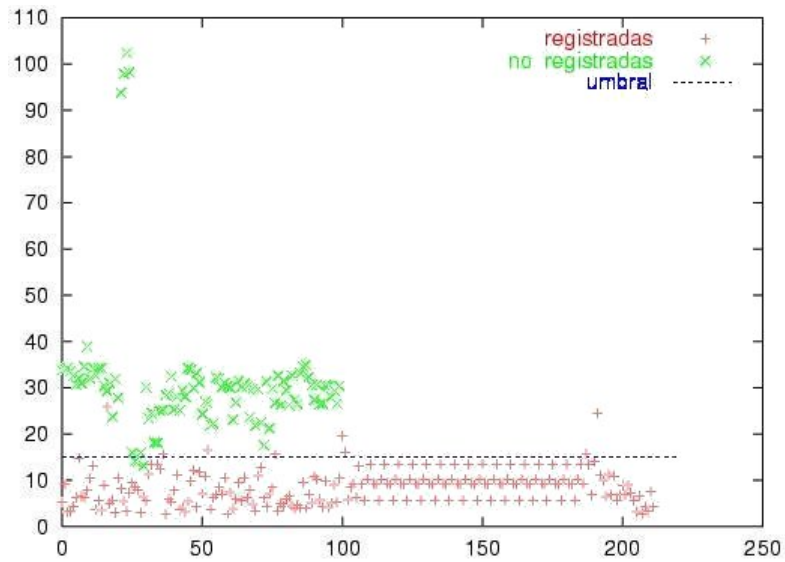


Figura 5.3: Fijando un umbral de 15, la tasa de falsos positivos es de 3% y la de falsos negativos se incrementa a un 6.6%. Este es el que se considera como el umbral que reporta los mejores resultados debido a que los falsos negativos es bajo y los falsos positivos no es muy alto.

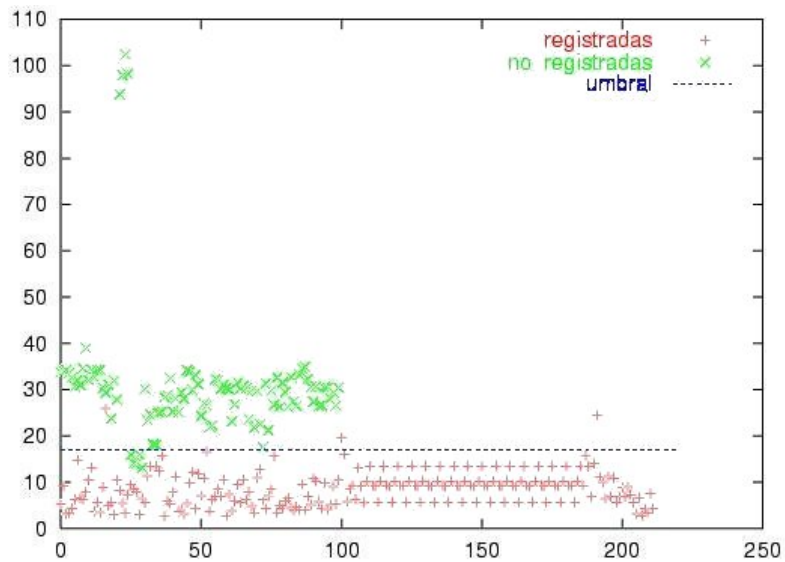


Figura 5.4: Con un umbral de 17, la tasa de falsos positivos es de 5%, por lo que ya no es muy conveniente para aplicaciones civiles.

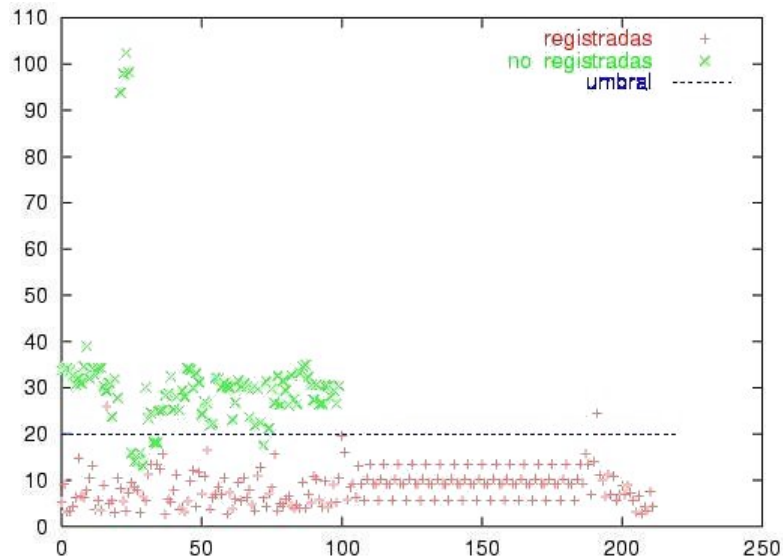
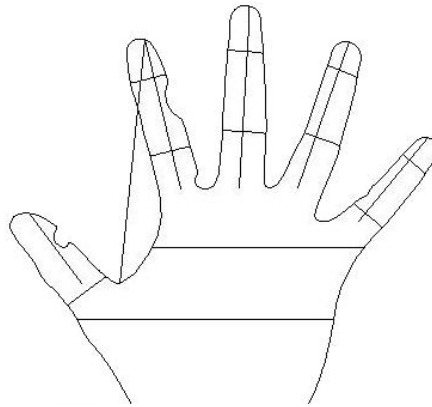


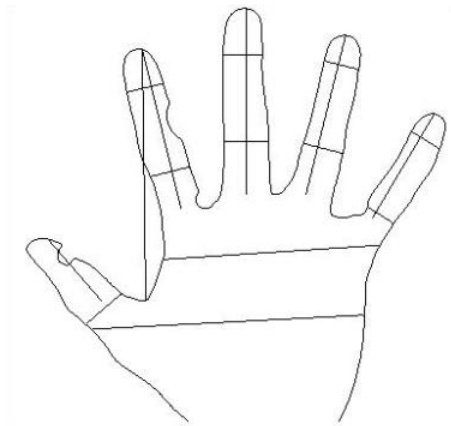
Figura 5.5: Con un umbral de 20, para aplicaciones civiles, ya no es conveniente, porque el porcentaje de falsos positivos es de 8% y es alto.

El módulo de extracción de características arrojó buenos resultados. Utilizando apropiadamente las técnicas de procesamiento de imágenes se ha logrado procesar adecuadamente imágenes con cierto grado de dificultad por diversos factores, como son el ruido, cierto grado de rotación (ver figuras 5.6(b) y 5.6(c)) e incluso, se logro una buena autenticación en manos de personas del sexo femenino que contaban con las uñas largas que brillaban o eran muy similares en color al fondo de la imagen (ver figuras 5.7(a) y 5.7(b)).

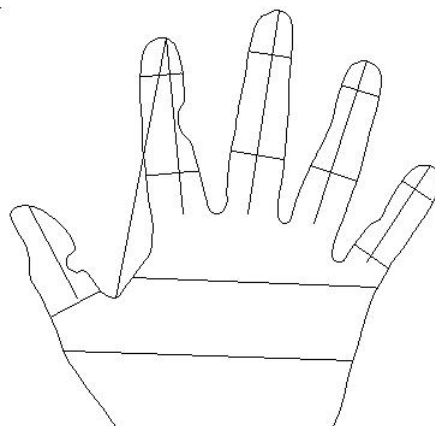
Una vez más se ha comprobado que las redes neuronales son una herramienta muy importante para la clasificación de datos, dando resultados muy buenos en este trabajo (ver tabla 5.1). Aunque los porcentajes reportados en algunos parezcan altos, cabe mencionar que las imágenes reportadas como falsos positivos pertenecen a una misma persona por lo que se puede considerar a los resultados como bastante aceptables. Se debe tener en cuenta que algunos porcentajes de error citados por fabricantes son basados en cálculos teóricos. Otros porcentajes son obtenidos en el campo de aplicación; éstos cuentan con datos más fidedignos. En el caso de falsos negativos, solo en el campo de aplicación pueden ser considerado exacto, debido a que una aplicación biométrica requiere interacción humana.



(a) Mano correctamente ubicada.



(b) Mano rotada a la izquierda.

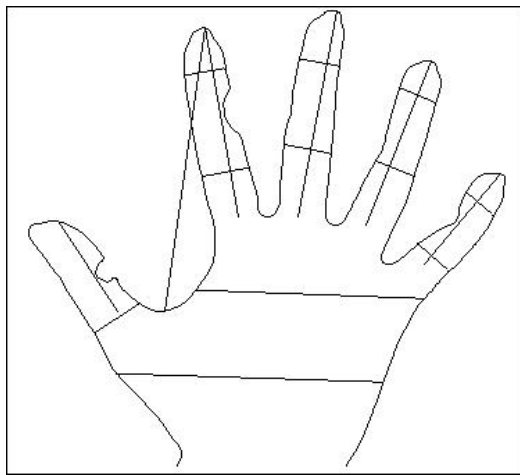


(c) Mano rotada a la derecha.

Figura 5.6: Manos procesadas con diferentes rotaciones.



(a) Mano original.



(b) Mano procesada por el sistema.

Figura 5.7: Ejemplo de manos con uñas largas bien procesadas.

5.3 Conclusiones

La necesidad de seguridad en diferentes dominios y la capacidad de las tecnologías biométricas de otorgarlas, hacen un buen juego para seguir ampliando y mejorando estas técnicas. En un mundo que está siendo cada día más automatizado crece el deseo de aumentar y mejorar los métodos de identificación y de autenticación debido al vigente fraude de identidad que tiene un costo muy alto. Se ha estimado que el restablecer contraseñas olvidadas tiene un costo de \$50 dólares Norteamericanos por persona [11].

Un punto tan importante como el desempeño de un sistema biométrico es la aceptación en la sociedad. Una de las más fundamentales preguntas que se hace es acerca de que tan *práctico* debe ser un sistema biométrico o que tan discriminante es la información disponible en la señal de entrada. Desafortunadamente, estas preguntas han sido contestadas en una forma muy limitada en todas las modalidades biométricas. Por lo tanto, los usuarios finales y el dominio en donde se aplicará cualquiera de estas tecnologías, deben de ser quienes decidan cual de todos los sistemas existentes es el que conviene más poner en práctica.

El objetivo de este trabajo fue el de mostrar una metodología completa para desarrollar un sistema biométrico utilizando como parte fundamental la mano derecha de las personas. El objetivo se cumplió satisfactoriamente, pues como se puede ver en los resultados, el sistema fue confiable y muy fácil de implementar, además de que no requirió de tecnología electrónica muy cara. Otro objetivo fue el de despertar la inquietud de desarrollar más esfuerzos en esta área (esto es decisión del lector), debido a que no existe una cantidad considerable de literatura pública o gratuita para el desarrollo de biometría utilizando la geometría de las manos dado que la mayoría son sistemas comerciales o patentados.

El sistema de extracción de características modela perfectamente las medidas censadas que mantiene toda la información que ayuda a tener una mejor discriminación y poder tener una mejor clasificación entre las clases.

Los mapas autoorganizativos de Kohonen tuvieron un buen desempeño desarrollando el papel de clasificador ya que cuentan con un amplio poder de abstracción y de generalización lo que permite que sean más tolerantes a fallas causadas por el ruido que pueda ser adquirido al obtener la imagen.

Se abre la pauta para que la geometría de las manos se combine con cualquier otra técnica biométrica para poder tener una mayor seguridad, tales como huellas digitales, reconocimiento de rostros y otras. Actualmente, la geometría de las manos solo se usa en combinación con una clave numérica que se tiene que teclear cuando el individuo desea autenticarse.

5.4 Trabajo Futuro

Se ha identificado algunos aspectos que ayudarán a mejorar el desempeño de la aplicación:

- Se usaron pocos individuos como prueba, pero se piensa trabajar con una cantidad de muestras más numerosas para la prueba de la aplicación.
- De cada una de las manos se tomaron solo 18 medidas pero es posible computar un número más grande para lograr una discriminación mayor entre las manos de diferentes individuos.
- En este trabajo y en otros reportados en la literatura, no se ha tomado en consideración la coloración de la piel, por lo que se probará diseñar un método que tomará en cuenta las intensidades de los pixeles pertenecientes a la mano utilizando una iluminación uniforme.
- Aunque esta aplicación permite cierto grado de rotación y traslación, es posible hacerla invariante a rotación y a traslación en un grado mayor.
- Tener un ambiente más controlado, es decir, con una iluminación uniforme al tomar las imágenes de los individuos.
- Implementar el sistema con respuesta en tiempo real.

Bibliografía

- [1] M. A. Cazorla, O. Colomina, F. Escolano, D. Gallardo, R. Rizo, R. Satorre. Técnicas de Inteligencia Artificial. Publicaciones de la Universidad de Alicante.
- [2] Miguel Angel Castellanos y Pedro Concejero. Introducción a las Redes Neuronales y su Aplicación a las Búsquedas de Documentos. Grupo factores Humanos telefónica I+D.
- [3] Gengxin Chen. Simulation of Auditory Cortical Feature Map Formation with Kohonen Network Model. May, 1998.
- [4] Joydeep Ghosh and Kagan Tumer. Integration of Neural Classifiers for Passive Sonar Signals. Dept. of Elect. and Comp. Engineering, university of Texas.
- [5] Dario Maravall Gómez Allende. Reconocimiento de Formas y Visión Artificial. Universidad Politécnica de Madrid. Addison-Wesley.
- [6] Rafael C. González and Richard E. Woods. Tratamiento Digital de Imágenes. Addison-Wesley/Diaz de Santos.
- [7] A. K. Jain and Nicolae Duta. Deformable Matching of Hand Shapes for Verification. Michigan State University, East Lansing, MI 48824-1026 and IBM T. J. Watson Research Center.
- [8] Anil Jain, Lin Hong, Sharath Pankanti. Biometrics: Promising Frontiers for Emerging Identification market.
- [9] Ramesh Jain, Rangachar Kasturi, Brian G. Schunck. Machine Vision. Mc.Graw-Hill, Inc.

- [10] A. K. Jain, M. N. Murty, P. J. Flynn. Data Clustering: A Review.
- [11] Anil Jain, Sharath Pankanti. Biometrics Systems: Anatomy of Performance.
- [12] A. K. Jain, Arun Ross, Sharath Pankanti. A Prototype Hand Geometry-Based Verification System. Michigan State University, East Lansing, MI 48824-1026.
- [13] Steve Lawrence, C. Lee Giles, Ah Chung Tsoi and Andrew D. Back. Face Recognition: A Convolutional Neural Network Approach. NEC Research Institute, 4 Independence Way, Princeton, NJ 08540. Electrical and Computer Engineering, University of Queensland, St. Lucia, Australia.
- [14] Mehrotra, Kishan. Elements of Artificial Neural Networks. Massachusetts Institute of Technology.
- [15] Enrique Moltó García. Introducción a la Visión Artificial. Universidad Politécnica de Valencia.
- [16] Lawrence O’Gorman. Fingerprint Verification. Veridicom Inc.
- [17] Berthold Klaus Paul Horn. Robot Vision. Massachusetts Institute of Technology. Mc. Graw Hill.
- [18] Miguel E. Ruiz, Padmini Srinivasan. Automatic Text Categorization Using Neural Network. School of Library and Information Science, The University of Iowa.
- [19] Rafael Sánchez, Ariel J. Barreiro, Pedro Pablo González. Un Modelo Competitivo Supervisado de Redes Neuronales. Centro de Neurociencias de Cuba.
- [20] Rafael Sánchez, Pedro Pablo González, Alejandro Riquenes, Edgar Leonel Chavez. A Neural Network to Learn Patterns from Non-Stationary Environments. Maestría en Ciencias de la Computación, IIMAS-UNAM. CU.
- [21] Richard L. Zunkel. Hand Geometry Based Verification. Recognition System Inc. Campbell, CA.
- [22] Dick Zunkel. Hand Geometry Today.
- [23] Marcos Faúndez Zanuy. Tratamiento Digital de Voz e Imagen. Alfaomega Grupo Editor.