

**RECONOCIMIENTO DE HUELLAS DIGITALES DESDE EL PUNTO DE  
VISTA ESTADÍSTICO**

**YAMILE FERNANDA BENITEZ ESCOBAR  
JAIR SANDINO FERNANDEZ ROJAS  
EDUARDO ANDRÉS MERA GUTIERREZ  
SILVANA ZARAMA REALPE**

**UNIVERSIDAD DEL CAUCA  
FACULTAD DE CIENCIAS NATURALES, EXACTAS Y DE LA  
EDUCACIÓN  
DEPARTAMENTO DE MATEMÁTICAS  
POPAYÁN  
2005**

**RECONOCIMIENTO DE HUELLAS DIGITALES DESDE EL PUNTO DE  
VISTA ESTADÍSTICO**

**YAMILE FERNANDA BENITEZ ESCOBAR  
JAIR SANDINO FERNANDEZ ROJAS  
EDUARDO ANDRÉS MERA GUTIERREZ  
SILVANA ZARAMA REALPE**

**TRABAJO DE GRADO**

**En la modalidad de seminario de grado presentado como requisito parcial  
para optar al título de licenciadas en educación en la especialidad  
matemáticas Yamile Fernanda Benítez Escobar, Silvana Zarama Realpe y  
matemáticos Jair Sandino Fernandez Rojas, Eduardo Andrés Mera Gutiérrez**

**Director**

**Dr. FREDY AMAYA**

**UNIVERSIDAD DEL CAUCA  
FACULTAD DE CIENCIAS NATURALES, EXACTAS Y DE LA  
EDUCACIÓN  
DEPARTAMENTO DE MATEMÁTICAS  
POPAYÁN**

**2005**

## Nota de aceptación

---

---

---

Director \_\_\_\_\_

**Doctor** Fredy Angel Amaya

## Comité evaluador

---

**Magister** Diego Ramiro Correa

---

**Magister** Orlando Rodríguez

Fecha de sustentación: Popayán, 11 de febrero de 2005

# Dedicatoria

*A mi madre por su apoyo incondicional y a mi padre con amor.*

*Eduardo*

*A mis Padres, a Carmenza, a Emilse y demás seres queridos con mucho cariño.*

*Jair*

*A mi Abuela, a mis Padres, a Paola y a Alfredo con amor.*

*Silvana*

*A mis Padres, a mi Hermano, a mi Abuela y demás Familiares con amor.*

*Yamile*

# Agradecimientos

Los autores expresan sus agradecimientos a:

**Dr. Fredy Angel Amaya, Director del Trabajo de Grado, por su apoyo, colaboración, confianza y motivación durante la realización de este trabajo.**

# Índice general

<b>Introducción</b>	<b>xiii</b>
<b>1. Sistema de reconocimiento automático de formas (SRAF)</b>	<b>1</b>
1.1. Fases de un sistema de reconocimiento automático de formas . . . . .	3
1.1.1. Fase de diseño . . . . .	3
1.1.2. Fase de operación . . . . .	5
1.2. Tipos de clasificadores . . . . .	6
1.2.1. Clasificador por regiones . . . . .	6
1.2.2. Clasificador por distancia euclidea . . . . .	8
1.3. Distinción entre las fases de diseño y de operación del clasificador euclideo	11
1.3.1. Etapas del diseño del clasificador euclideo . . . . .	11
1.3.2. Fase operativa del clasificador . . . . .	14
1.4. Reconocimiento con aprendizaje supervisado en condiciones determinísticas	15
1.4.1. Aprendizaje de funciones discriminantes por regiones . . . . .	16
1.4.2. Aprendizaje de funciones discriminantes por distancia . . . . .	20
1.4.3. Fundamentos de los algoritmos de aprendizaje . . . . .	22
1.5. Reconocimiento estadístico a priori. El rocedor bayesiano . . . . .	28
1.5.1. Teorema de bayes . . . . .	29
1.5.2. La media . . . . .	32
1.5.3. La matriz de covarianzas . . . . .	33
1.5.4. El coeficiente de correlación . . . . .	34
1.5.5. Estimación estadística de la matriz de covarianza y del vector media	34

1.5.6.	El clasificador estadístico a priori . . . . .	35
1.5.7.	Diseño de un reconocedor estadístico . . . . .	38
1.5.8.	La regla de decisión de Bayes para el riesgo mínimo . . . . .	42
1.6.	Reconocimiento automático de formas con aprendizaje en condiciones estadísticas . . . . .	43
1.6.1.	Clasificación . . . . .	44
1.6.2.	Algoritmo de aprendizaje de funciones discriminantes en condiciones estadísticas . . . . .	44
1.7.	Algoritmos de agrupación de clases (Clustering) . . . . .	50
1.7.1.	Algoritmo K-Medias . . . . .	53
<b>2.</b>	<b>Sistema de reconocimiento de huellas digitales</b>	<b>56</b>
2.1.	Nociones generales . . . . .	56
2.1.1.	Biometría o técnicas de identificación biométrica . . . . .	56
2.1.2.	Autenticación de usuarios . . . . .	57
2.1.3.	Autenticación biométrica . . . . .	59
2.1.4.	Huellas digitales . . . . .	62
2.1.5.	Individualidad de la huella digital . . . . .	66
2.1.6.	Sensores de la huella digital . . . . .	71
2.1.7.	Clasificación de huellas digitales . . . . .	73
2.1.8.	Autenticación de la huella digital . . . . .	75
2.2.	Modelos probabilísticos de identificación de la huella digital . . . . .	78
2.2.1.	Antecedentes . . . . .	81
2.2.2.	Un modelo de individualidad de la huella digital . . . . .	82
<b>3.</b>	<b>Extracción de características</b>	<b>92</b>
3.1.	Localización del punto de referencia . . . . .	98
3.2.	Algoritmo gerárquico de los campos de orientación . . . . .	99
3.3.	Algoritmo de localización de puntos de referencia . . . . .	101
3.4.	Fragmentación . . . . .	106

3.5. Filtrado . . . . .	108
3.5.1. Detección de crestas . . . . .	113
3.5.2. Detección de minucias . . . . .	114
3.6. Vector de características . . . . .	115
3.6.1. Verificación de minucias . . . . .	118
3.7. Emparejamiento . . . . .	122
3.7.1. Emparejamiento basado en la transformación de Hough (algoritmo de Hough) . . . . .	129
3.7.2. Emparejamiento basado en la distancia entre cadenas (algoritmo cadena) . . . . .	129
3.7.3. Emparejamiento basado en el banco de filtros (algoritmo filtro) . . . . .	130
3.8. Estrategias de integración . . . . .	132
3.8.1. Selección de emparejadores . . . . .	132
3.8.2. Estimación de Densidad no-paramétrica . . . . .	133
<b>4. Máxima entropía</b>	<b>135</b>
4.1. Principio de la Máxima Entropía (PME) . . . . .	136
4.1.1. El algoritmo Improve Iterative Scaling . . . . .	141
4.2. Aplicación del modelo de máxima entropía en el reconocimiento de huellas digitales . . . . .	143
<b>Conclusiones</b>	<b>147</b>
<b>Bibliografía</b>	<b>149</b>



# Índice de figuras

1.1. Diagrama de flujo indicando las etapas en el diseño. . . . .	4
1.2. Diagrama de bloques de SRAF en la fase operativa. . . . .	5
1.3. Representación en un plano de dos clases . . . . .	7
1.4. Diagrama de bloques que ilustra el proceso de reconocimiento . . . . .	9
1.5. Etapas en el diseño de un sistema de reconocimiento automático de formas.	14
1.6. Diagrama de bloques simbólico del proceso de aprendizaje. . . . .	16
1.7. Reconocedor por regiones que realiza una búsqueda en forma de árbol. . .	20
1.8. Diagrama de bloques más detallado del proceso de aprendizaje. . . . .	23
1.9. Problema de clasificación de las clases. . . . .	25
1.10. Representación en un plano abstracto de las características de $\alpha_1$ y $\alpha_2$ . . .	29
1.11. Estructura del clasificador estadístico a posteriori. . . . .	30
1.12. Estructura del clasificador estadístico a priori. . . . .	31
1.13. Representación en un plano de diferentes muestras de ambas clases. . . . .	38
1.14. Representación de la función discriminante lineal del clasificador Bayesiano.	40
1.15. $fd$ cuando las desviaciones típicas de las características difieren. . . . .	41
1.16. Esquema de bloques del reconocedor estadístico con aprendizaje. . . . .	48
1.17. Esquemmatización del proceso de aprendizaje simultáneo multiclase. . . . .	49
1.18. Esquemmatización de los cuatro metodos estudiados . . . . .	51
1.19. Representación simbólica de un algoritmo de agrupación . . . . .	54
2.1. Diagrama de bloques de un Sistemas Reconocimiento Biométrico. . . . .	61
2.2. Relación entre FAR, FRR Y ERR. . . . .	62
2.3. Minucias donde $\bigcirc$ encierra la bifurcación y $\square$ el final de la cresta. . . . .	64

2.4. Tipos de Minucias. . . . .	65
2.5. Puntos singulares en una huella digital. . . . .	66
2.6. Imágenes de la huella digital de hermanas gemelas idénticas. . . . .	68
2.7. Extracción de minucias de las gemelas. . . . .	69
2.8. Emparejamiento de minucias para (a) gemelo-no gemelo. . . . .	70
2.9. Emparejamiento de minucias para dos impresiones del mismo dedo. . . . .	70
2.10. Sensores de Huellas digitales. . . . .	73
2.11. Las Seis mejores clases de la huella digital. . . . .	75
2.12. Diagrama para un sistema de autenticación automático. . . . .	76
2.13. Una imagen de la huella digital de tipo right loop. . . . .	80
2.14. Emparejamiento automático de minucia. . . . .	86
2.15. Huella digital y minucia . . . . .	87
3.1. Patrones de flujo en una imagen de la huella digital. . . . .	93
3.2. Diagrama esquemático para la extracción genérica. . . . .	95
3.3. Huella digital de (a) un niño, y (b) un adulto. . . . .	96
3.4. Organigrama del algoritmo de extracción de minucias . . . . .	97
3.5. Las crestas cóncavos y convexos en una imagen de la huella digital. . . . .	98
3.6. Gradiente de $f(x, y)$ . . . . .	100
3.7. Estimación del punto de referencia. . . . .	102
3.8. Integrar por regiones las intensidades de los pixeles. . . . .	104
3.9. Ejemplo del algoritmo de localización de puntos de referencia. . . . .	105
3.10. El espectro de Fourier de (a) y (b) se muestra en (c) y (d), respectivamente. . . . .	109
3.11. Filtros de Gabor (Tamaño de máscara de $33 \times 33$ ). . . . .	110
3.12. Normalizado, filtrado y reconstrucción de la imagen de la huella digital. . . . .	112
3.13. Ejemplos de vectores de características de dimensión 640. . . . .	117
3.14. Campos de orientacion, crestas finas, minucias y puntos singulares. . . . .	119
3.15. Imágenes de muestra de la base datos GT. . . . .	120
3.16. Ejemplos de los perfiles de los niveles de gris en la vecidad. . . . .	121
3.17. Ejemplos de imágenes de la base de datos GT. . . . .	122

3.18. Dificultad en el emparejando de la huella digital. . . . .	125
3.19. Figercode obtenido de una huella digital por rotación . . . . .	127
3.20. Diagrama del algoritmo de emparejamiento y extracción de características. . . . .	131
4.1. Esquema del clasificador usando el principio de la máxima entropía . . . . .	144

# Resumen

Este documento contiene el informe del seminario de grado titulado “*reconocimiento de huellas digitales desde el punto de vista estadístico*” realizado dentro de la línea matemática computacional en el área de reconocimiento de formas. Inicialmente se presenta un análisis de los principios de funcionamiento de un sistema de reconocimiento automático de formas, se explica las fases de diseño y de operación, las cuales a su vez son estudiadas mediante el uso de los clasificadores, abarcando un conjunto de técnicas y herramientas tanto estadísticas como matemáticas. Luego se prosigue con el sistema de reconocimiento de huellas digitales, y se examinan algunos de los algoritmos existentes. Finalmente se estudia el principio de máxima entropía como posible herramienta para desarrollar sistemas de reconocimiento de huellas digitales.

# Introducción

Los sistemas tradicionales utilizados en el control de acceso a recintos se basan en los sistemas de tarjetas magnéticas, sistemas de tarjetas con código de barras, sistemas de captura de clave o combinación de ellos. Estos sistemas involucran el uso de una tarjeta que hay que llevar siempre consigo y la cual no está exenta de perderse, dañarse, ser robada o falsificada, con lo cual la seguridad del recinto se hace más vulnerable a fallas. Por esta razón, si se cuenta con un sistema más sólido y de mayor confiabilidad se pueden evitar los problemas antes mencionados.

Con el incremento del cálculo de las computadoras se han ido desarrollando sistemas automatizados, de este modo surge las técnicas de la identificación biométrica con el objetivo de realizar la clasificación e identificación a partir de las características propias de cada individuo, como la voz, rostro, huella dactilar, geometría de la mano, iris, retina, etc. La identificación basada en la huella digital es una de las técnicas más maduras y probadas.

Dado que el sistema de reconocimiento de huellas digitales está inmerso en el reconocimiento automático de formas, es necesario hacer un énfasis en el estudio de este último, para lograr una mejor comprensión en el momento de analizar los diferentes pasos y algoritmos, que se utilizan en el desarrollo del reconocimiento de huellas digitales.

Debido a que el proceso de formación de la huella digital es un sistema caótico en lugar de uno aleatorio, es necesario el principio de la máxima entropía puesto que construye modelos estocásticos para predecir el comportamiento como un proceso aleatorio.

# Capítulo 1

## Sistema de reconocimiento automático de formas (SRAF)

El reconocimiento automático de formas abarca un conjunto de técnicas y herramientas matemáticas y computacionales con el fin de crear sistemas computarizados con la capacidad de reconocer objetos de diferente naturaleza. El reconocimiento automático de formas tiene aplicaciones en numerosas disciplinas como acústica, sismología, ingeniería, medicina, política, etc.

Sin embargo, su desarrollo histórico ha estado estrechamente ligado a la informática, a esto se debe la procedencia y formación de los numerosos especialistas que trabajan en este campo, y de una manera muy especial, el reconocimiento de formas suele vincularse a la denominada inteligencia artificial (IA), conjunto de técnicas que hacen parte de los avances de la informática profesional. Concretamente, el reconocimiento de formas constituye la base de la visión artificial, una de las especialidades más importantes de la IA, y como consecuencia de ser un pilar de la visión computacional el reconocimiento de formas también está vinculado a la robótica, entendida no sólo como la especialidad de la ingeniería orientada a la automatización industrial, sino igualmente a la automatización del sector terciario.

A continuación se definen algunos términos que son útiles en el reconocimiento automático

de formas:

**Universo de Trabajo (UT):** Está formado por todos los posibles objetos individuales que se van a reconocer de manera automática. Estos objetos predefinidos suelen denominarse patrones o clases.

$$UT = \{patrones\}$$

Para proceder al reconocimiento automático de los objetos individualizados (segmentados) se realiza una importantísima transformación de los mismos, donde se convierten en un vector  $X$  cuyas componentes se denominan características o rasgos. Cada vector de características se compara con un conjunto de vectores preestablecidos o diccionario, compuesto por los vectores de rasgos de todos los objetos del universo de trabajo. Se emplea la notación de vector columna para el vector de características  $X$ :

$$\begin{pmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_n \end{pmatrix}$$

En donde  $X_1, X_2, \dots, X_n$  son números reales que cuantifican, para un objeto concreto, las correspondientes características.

Conviene destacar las propiedades más importantes que deben tomarse en la elección y prueba de las características.

1. Las características deben ser discriminantes: valores numéricos diferentes para objetos de clases distintas.
2. Las características deben ser fiables, es decir, con cambios numéricos pequeños para objetos de una misma clase. Posteriormente veremos que si no se cumple esta

propiedad es necesario pasar a un enfoque estadístico en el diseño de las funciones discriminantes.

3. Las características deben estar incorrelacionadas, es decir, deben ser independientes. La independencia de éstas equivale al principio de la parsimonia: decir lo mismo con la máxima economía de términos.
4. Las características deben calcularse en un tiempo aceptable. Este es un requisito que puede llegar a ser determinante en ciertas aplicaciones de tiempo real.

**Clasificador:** Es un algoritmo que dado un vector de características lo clasifica en una de las clases en que está dividido el universo de trabajo.

**Función Discriminante ( $fd$ ):** Es una función que separa en diferentes hiperplanos (en ciertos clasificadores como son: el clasificador por regiones y el clasificador por distancia que se tratan más adelante) el espacio abstracto de las clases o patrones.

## 1.1. Fases de un sistema de reconocimiento automático de formas

### 1.1.1. Fase de diseño

#### Determinación de los patrones o clases del universo de trabajo

Normalmente la determinación del universo de trabajo ( $UT$ ) se hace manualmente, aunque pueda presentarse situaciones en donde se deban aplicar herramientas automáticas como las técnicas de agrupación de datos (clustering) cuando el  $UT$  es difícil de establecer.

#### Elección y prueba de las características discriminantes

La elección de características es absolutamente vital en el diseño de todo clasificador, pues condiciona su rendimiento. Se trata de una fase abierta en la que es habitual eliminar y añadir características en función del rendimiento del clasificador. Desgraciadamente no



existen reglas formales y de validez universal para la elección de las características. Es la habilidad, la experiencia y el buen sentido del diseñador del sistema quienes pueden ayudarle en su elección.

### Diseño y prueba de las funciones discriminantes

Una vez seleccionado el vector de características del sistema, la etapa final es ya puramente mecánica y se centra en el cálculo de las correspondientes funciones de decisión o funciones discriminantes. En la figura 1.1 se ha representado las etapas de diseño.

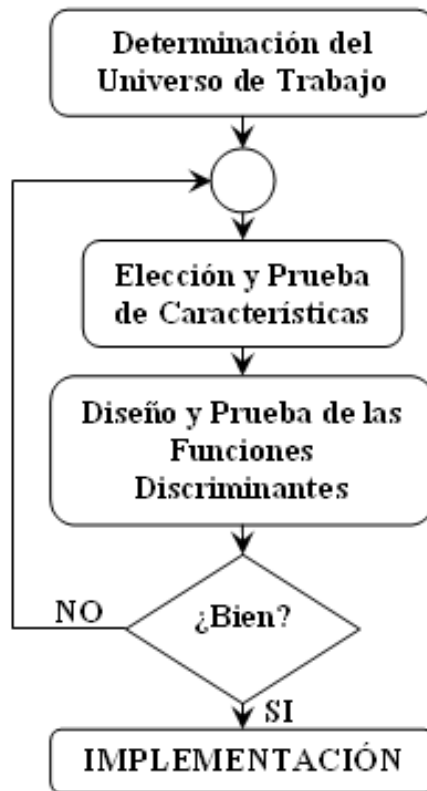


Figura 1.1: Diagrama de flujo indicando las etapas en el diseño de un sistema de reconocimiento automático de formas

### 1.1.2. Fase de operación

Una vez diseñado el sistema, como sucede en cualquier otro campo de la ingeniería o de la informática aplicada, la operación siguiente y final es su realización material o implementación, que da lugar a su fase operativa. En la fase operativa de un sistema de reconocimiento automático de formas hay que diferenciar las siguientes etapas o módulos:

#### Aislamiento de los objetos individuales (segmentación)

El aislamiento de los objetos individuales, o segmentación, es la operación más delicada en la mayoría de los casos prácticos. En realidad, se trata de una problemática ajena, en sentido estricto, al reconocimiento de formas, ya que en éste se parte siempre de la conversión de un objeto individual en un vector de características.

#### Cálculo de las características de cada objeto

La siguiente operación es el cálculo o extracción del vector de características de cada uno de los objetos individuales a reconocer.

#### Reconocimiento o clasificación automática.

Finalmente ya sólo quedaría aplicar las fd a los vectores  $X_i$  provenientes de los objetos individuales al conjunto de funciones discriminantes previamente diseñados para obtener la clasificación buscada. En la figura 1.2 se representan estos módulos.

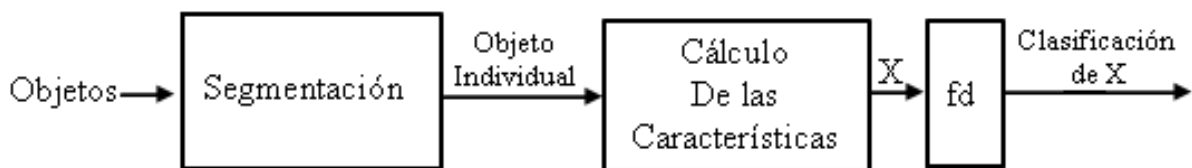


Figura 1.2: Diagrama de bloques de SRAF en la fase operativa.

## 1.2. Tipos de clasificadores

En la sección siguiente se describen dos tipos de clasificadores donde encontraremos dos familias de funciones discriminantes lineales: Las basadas en una regionalización del espacio abstracto de las clases y las basadas en el uso de una función distancia; ambas consisten en una combinación lineal de las características. Así, para  $n$  rasgos discriminantes, una función lineal se expresa como:

$$F_d(X) = W_1X_1 + W_2X_2 + \dots + W_nX_n + W_{n+1} \quad (1.1)$$

$$F_d(X) = W^T \cdot X = (W_1, W_2, \dots, W_n, W_{n+1}) \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \\ 1 \end{pmatrix} \quad (1.2)$$

Siendo  $W$  el vector de coeficientes de discriminación, indicando el superíndice  $T$  que se trata del vector transpuesto y  $X$  el vector de características ampliado en una componente unidad, para poder escribirla expresión (1.1) en forma vectorial.

### 1.2.1. Clasificador por regiones

Cuando los vectores de cada clase poseen dos características pertenecen a  $\mathbb{R}^2$ , se colocan dichos vectores en el plano cartesiano, luego se escogen parejas de puntos que no pertenezcan a las clases, de tal forma que al trazar rectas por cada una de las parejas haya separabilidad entre clases, donde cada ecuación de dichas rectas serán las funciones discriminantes. En la figura 1.3 se puede observar, lo anterior para dos clases.

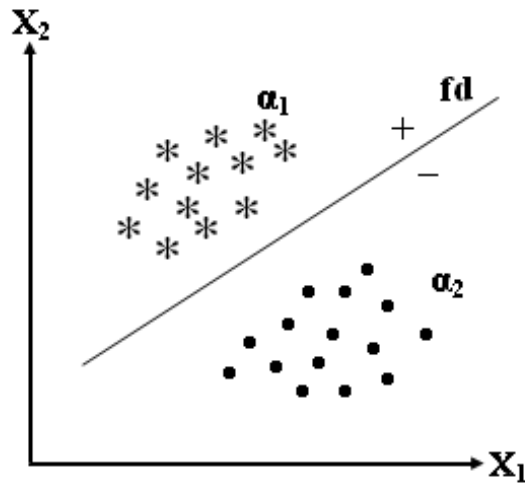


Figura 1.3: Representación en un plano de dos clases

Para el caso de vectores con tres características (vectores en  $\mathbb{R}^3$ ), las funciones discriminantes serían planos y, generalizando, en un espacio  $n - dimensional$  las funciones discriminantes ( $fd$ ) lineales son hiperplanos.

El diseño de las ( $fd$ ) consiste en encontrar hiperplanos  $W^T X$  que dividan el espacio de las clases en regiones asociadas cada una de ellas a una clase.

Para  $N$  clases  $\alpha_1, \alpha_2, \dots, \alpha_N$  no es trivial establecer el número de funciones discriminantes que se necesitarían para dividir el espacio en  $N$  regiones disjuntas, cada una asociada a una sola clase. En el caso peor, se necesitarían  $\binom{N}{2}$  funciones discriminantes al ser necesario obtener una  $fd$  para cada una de las posibles combinaciones de dos clases que se pueden formar con  $N$  clases. En el caso óptimo, se necesitaría  $N - 1$   $fd$  ya que se podría discriminar cada una de las clases con el resto, considerando éste como una sola clase. En cualquier caso, el procedimiento común es obtener una  $fd$  de dos clases. Más adelante veremos que es posible, en condiciones de aprendizaje, abordar una solución multiclase que permite discriminar simultáneamente  $N$  clases.

La función discriminante o función de decisión  $fd(\mathbf{X})$  es una aplicación escalar sobre el vector  $X$ , de tal manera que en el caso de utilizar el procedimiento de regionalización del espacio de patrones equivale a un signo: +, si  $fd(\mathbf{X}) > 0$ , y -, si  $fd(\mathbf{X}) < 0$ . Para  $fd(\mathbf{X}) = 0$  el vector queda indeterminado en cuanto a su clasificación.

Obviamente, para vectores de características superiores a tres (dimensión tres), el diseño geométrico deja de ser válido y la obtención de las funciones discriminantes se convierte en un proceso muy largo y tedioso.

### 1.2.2. Clasificador por distancia euclídea

El reconocimiento basado en la distancia euclídea parte de dos hipótesis: en primer lugar, que las clases son de naturaleza determinística (en contraposición a una hipótesis estadística) y, en segundo lugar, que toda la información necesaria y suficiente para su diseño se encuentre disponible a priori (en contraposición a un enfoque basado en el aprendizaje).

Cuando se dice que las clases presentan un comportamiento determinístico, en realidad lo que acontece es que se postula ese comportamiento puesto que en la práctica es poco menos que imposible encontrarse una clase de objetos totalmente determinísticos.

Se dice que una clase es determinística cuando sus elementos se representan por un único vector, que es el prototipo de la clase o centroide.

Dado un conjunto de elementos de una cierta clase,  $\alpha_i = \{X_{i1}, X_{i2}, \dots, X_{ip}\}$  se suele tomar como vector prototipo o centroide su media ponderada:

$$Z_i = \frac{1}{p} \sum_{j=1}^p X_{ij} \quad (1.3)$$

Cuando se dice que para el diseño del clasificador de la distancia se opera con todo el conocimiento a priori necesario y suficiente, se refiere a que es preciso conocer previamente todos los centroides de las clases.

En cuanto a la función distancia, en principio se pueden utilizar varias, aunque la más utilizada es la distancia euclídea debido a que sus propiedades matemáticas son mejores.

Entonces dado un vector a reconocer  $X$ , el principio intuitivo del clasificador mediante una función distancia es asociar el vector  $X$  a la clase cuyo centroide se encuentra más cerca de  $X$ .

Así, la distancia entre un vector a clasificar  $X$  y el centroide  $Z_i$  de una clase genérica  $\alpha_i$  viene dada por:

$$d_E(X, Z_i) = \|X - Z_i\| = \sqrt{(X - Z_i)^T \cdot (X - Z_i)} = \sqrt{\sum_{j=1}^n (X_j - Z_{ij})^2} \quad (1.4)$$

En la figura 1.4 se ilustra este proceso de reconocimiento:

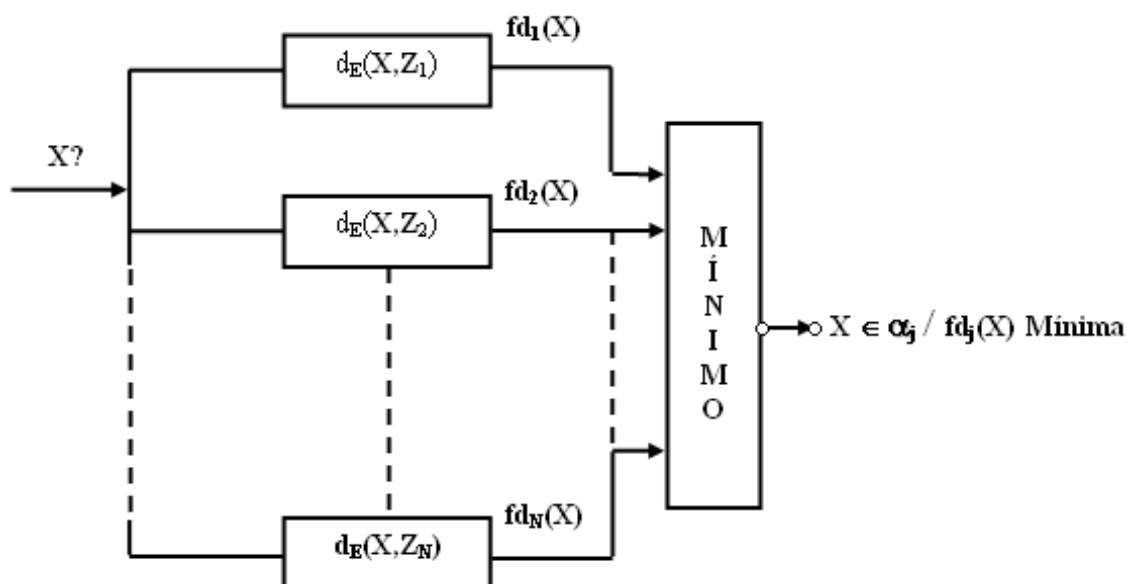


Figura 1.4: Diagrama de bloques que ilustra el proceso de reconocimiento

Nótese que se ha introducido la notación de función de decisión  $fd(\mathbf{X}) = fd(\mathbf{X}, \mathbf{Z}_i)$  asociada a cada clase. Una de las ventajas del reconocedor euclídeo respecto al clasificador por regiones es que el número de funciones de decisión o de discriminación ( $fd$ ) es siempre fijo e igual al número de clases. Otra ventaja es que el cálculo de estas ( $fd$ ) es simple y

directo suponiendo que se conocen los prototipos de las clases frente a la complejidad en el caso del discriminador por regiones.

En sentido estricto las  $fd$  del clasificador euclídeo no son lineales, puesto que aparecen raíces cuadradas y cuadrados de las características  $X_1, X_2, \dots, X_n$ , del vector a clasificar. Para llegar a funciones discriminantes lineales asociadas a la distancia euclídea, vamos a considerar su cuadrado:

$$\begin{aligned}
 d_E^2(\mathbf{X}, \mathbf{Z}_i) &= \|X - Z_i\|^2 \\
 &= (X - Z_i)^T \cdot (X - Z_i) \\
 &= X^T X - X^T Z_i - Z_i X^T + Z_i^T Z_i \\
 &= X^T X - X^T Z_i - X^T Z_i + Z_i^T Z_i
 \end{aligned}$$

De ahí que:

$$d_E^2(\mathbf{X}, \mathbf{Z}_i) = X^T X - 2X^T Z_i + Z_i^T Z_i$$

El elemento  $X^T X$  puede despreciarse ya que no discrimina y se repite en todas las clases. Si además dividimos por dos y cambiamos los signos, se obtendría la siguiente función discriminante basada en la distancia euclídea:

$$fd_i(\mathbf{X}) = X^T Z_i - \frac{1}{2} Z_i^T Z_i \quad (1.5)$$

Como  $d_E^2(\mathbf{X}, \mathbf{Z}_i)$  es una cantidad positiva, al haber eliminado el término  $X^T X$  y cambiado de signo los términos restantes, se cumple que la distancia euclídea al cuadrado mínima (entre las  $N$  distancias de  $X$  a  $Z_1, Z_2, \dots, Z_N$ ) es la que hace  $fd(\mathbf{X})$  máxima. Por último, si el vector de características está bien escogido, de tal forma que las clases son linealmente separables, este clasificador es el más recomendable por las anteriores cualidades y, además, porque en esa situación su tasa de aciertos es muy elevada.

## 1.3. Distinción entre las fases de diseño y de operación del clasificador euclideo

### 1.3.1. Etapas del diseño del clasificador euclideo

La determinación de los patrones o clases del universo de trabajo y la elección y prueba de las características discriminantes se realiza de igual forma para cualquier reconocedor. Realizadas estas dos primeras etapas se continua con las siguientes.

#### Hipótesis determinística en relación con la naturaleza de las clases

El clasificador euclideo, por definición, es determinístico, en el sentido que se representa cada clase por un vector único o prototipo. Por tanto, es preciso evaluar el cumplimiento de esta hipótesis para todas las clases existentes si se va a diseñar un clasificador euclideo, ya que cuando se adopta una hipótesis determinista se está haciendo una aproximación a la realidad. Lo que el diseñador debe evaluar es si tal aproximación es o no aceptable en la aplicación específica que esté desarrollando.

Sin existir reglas formales, un criterio que puede utilizarse se basa en el estudio de la matriz de covarianzas de la clase. Así, centrándonos en la clase genérica  $\alpha_i$  su matriz de covarianzas  $C_i$  se define como:

$$C_i = E \{(X - E \{X\})^T (X - E \{X\})\}_{\alpha_i}$$
$$C_i = \begin{pmatrix} (X_1 - \bar{X}_1)^2 & (X_1 - \bar{X}_1)(X_2 - \bar{X}_2) & \dots & (X_1 - \bar{X}_1)(X_n - \bar{X}_n) \\ (X_2 - \bar{X}_2)(X_1 - \bar{X}_1) & (X_2 - \bar{X}_2)^2 & \dots & (X_2 - \bar{X}_2)(X_n - \bar{X}_n) \\ \vdots & \vdots & \ddots & \vdots \\ (X_n - \bar{X}_n)(X_1 - \bar{X}_1) & (X_n - \bar{X}_n)^2 & \dots & (X_n - \bar{X}_n)(X_n - \bar{X}_n)^2 \end{pmatrix}_{\alpha_i} \quad (1.6)$$

Donde  $E$  es el operador esperanza matemática y  $E(X_i) = \bar{X}_i$ , es decir, la diagonal



principal de la matriz de covarianza está formada por las varianzas de cada característica y el resto de los elementos son las covarianzas. Esta matriz es simétrica. Utilizaremos por comodidad una notación más compacta de esta matriz:

$$C_i = \begin{pmatrix} C_{11} & C_{12} & \dots & C_{1n} \\ C_{21} & C_{22} & \dots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \dots & C_{nn} \end{pmatrix}_{\alpha_i} \quad (1.7)$$

De momento vamos a limitarnos a utilizar los elementos de la matriz (1.7) para validar o no la hipótesis de comportamiento determinístico de la clase  $\alpha_i$ . Para ello se calcula la media aritmética de la clase, que es un estimador de la media estadística  $E(X)$ , donde:

$$m_i = \frac{1}{p} \sum_{j=1}^p x_{ij} \quad \text{luego,} \quad \bar{X} = \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_p \end{pmatrix} \quad (1.8)$$

Que es, lógicamente, un vector columna con  $n$  elementos, cada uno de ellos la media aritmética de una característica. Obsérvese que coincide con el vector prototipo o centroide  $Z_i$ , de la clase  $\alpha_i$ . Un criterio experimental para validar las hipótesis determinísticas es que las desviaciones típicas de las características no superen el 10 % de sus correspondientes medias, es decir la clase  $\alpha_i$  se considera determinística si:

$$\frac{\sqrt{C_{ij}}}{m_{ij}} = \frac{\sigma_j}{m_{ij}} \leq 0,1 \quad j = 1, 2 \dots n \quad (1.9)$$

No obstante si las desviaciones típicas adquieren valores significativos, es decir que sean muy superiores a la cota del 10 %, pero los centroides de las diferentes clases están muy separados entre sí, entonces se puede admitir la hipótesis determinística, esto es representar los vectores de la clase por un único vector, el centroide, aunque en sentido estricto la clase presente un marcado carácter aleatorio. Más adelante se introducirá un

parámetro denominado el ratio de Fischer (cuantifica simultáneamente la separación interclases y la fiabilidad interna de las clases) que engloba los dos efectos anteriores: la dispersión interna de una clase y la dispersión externa de las clases entre sí.

### **Cálculo de los centroides de las clases**

Una vez aceptada la hipótesis determinista, el paso inmediato es calcular el centroide o prototipo de cada clase:

$$Z_i = \frac{1}{p} \sum_{j=1}^p X_{ij} \quad (1.10)$$

### **Implementación de las funciones discriminantes**

Ya que se ha demostrado que la función discriminante asociada a la distancia euclídea viene dada por:

$$f_{d_i}(\mathbf{X}) = \mathbf{X}^T Z_i - \frac{1}{2} Z_i^T Z_i \quad i = 1, 2, \dots, n \quad (1.11)$$

El cálculo de estas funciones es inmediato.

Una vez se hayan cumplido los pasos anteriores se procede con la prueba; si es satisfactoria, se pasa directamente a la implementación del reconocedor.

A continuación se muestra el diagrama de flujo del proceso de diseño expuesto en la figura 1.5.

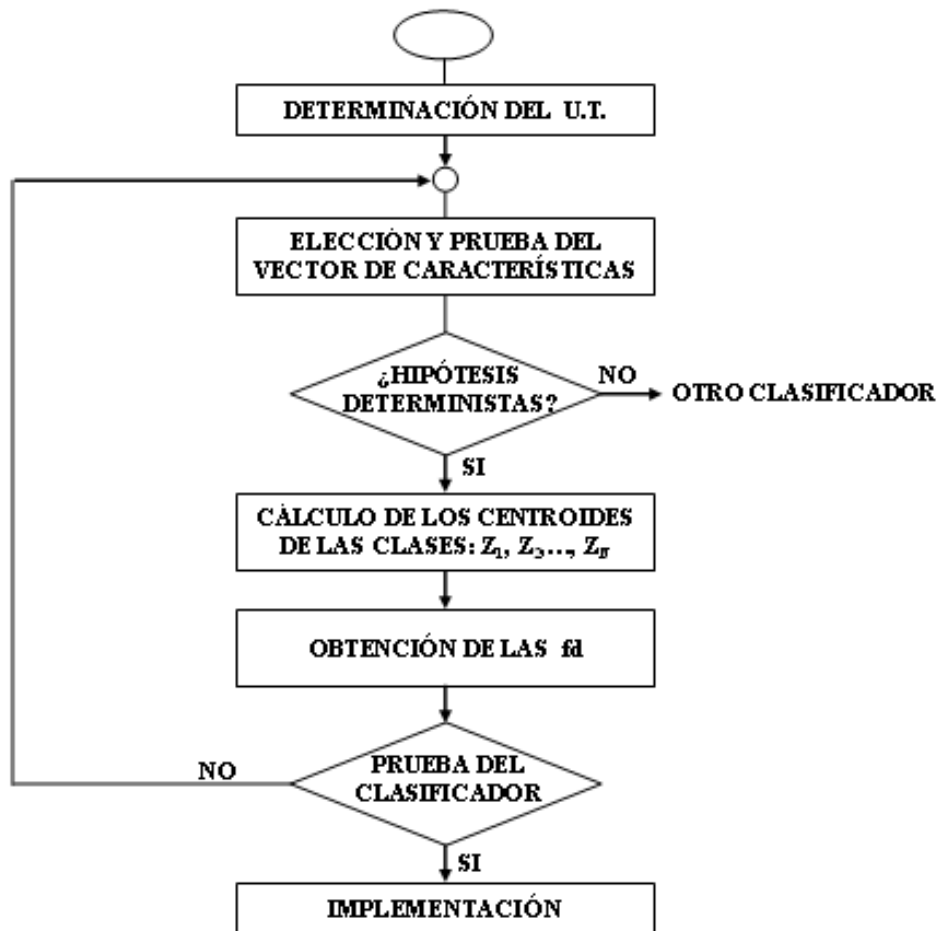


Figura 1.5: Etapas en el diseño de un sistema de reconocimiento automático de formas.

### 1.3.2. Fase operativa del clasificador

Se entra a la fase operativa del clasificador cuando habiendo finalizado con éxito el diseño de un clasificador, este se realiza físicamente. En el clasificador euclídeo se observa que el vector  $X$  correspondiente a un objeto desconocido a clasificar, es aplicado en paralelo sobre las  $N$  funciones discriminantes y de la comparación entre estas  $N$ ,  $fd$  se obtiene la solución. Es de notar que aquí se trabaja con la distancia máxima y no la mínima [3].

## 1.4. Reconocimiento con aprendizaje supervisado en condiciones determinísticas

En esta sección se planteará un enfoque distinto, basado en aprendizaje supervisado de tal manera que existe un maestro que se encarga de enseñar al sistema, corrigiéndole cada vez que se equivoque, es decir, cuando clasifique incorrectamente a un objeto.

Para ello se dispondrá de un conjunto de muestras de aprendizaje (CMA) formada por  $P$  vectores  $X_1, X_2, X_3, \dots, X_P$  cuya pertenencia a una determinada clase es desconocida. Si existen  $N$  clases  $\alpha_1, \alpha_2, \dots, \alpha_n$  entonces  $P \gg N$ . El conjunto de muestras de aprendizaje, CMA, se escoge habitualmente de forma que incluya el mismo número de muestras de aprendizaje por cada clase.

Ahora se va a plantear un modelo simple e intuitivo del aprendizaje supervisado en el diseño de un reconocedor determinista. Así, considerese la figura 1.6 en donde se extrae en un instante genérico  $k$  un vector  $X(k)$  del conjunto de aprendizaje o entrenamiento, cuya pertenencia a una clase específica es conocida por el maestro (es decir por el diseñador del proceso de aprendizaje).

El elemento  $X(k)$  es clasificado por el reconocedor automático en virtud del conjunto de funciones discriminantes que en ese instante esté manejando. Estas  $fd$  cambian obviamente, y lo hacen en función de la señal de error que se haya producido en la decisión del reconocedor; no habrá error (es decir acierto) cuando coincida su decisión con la del maestro y existe error cuando difieran.

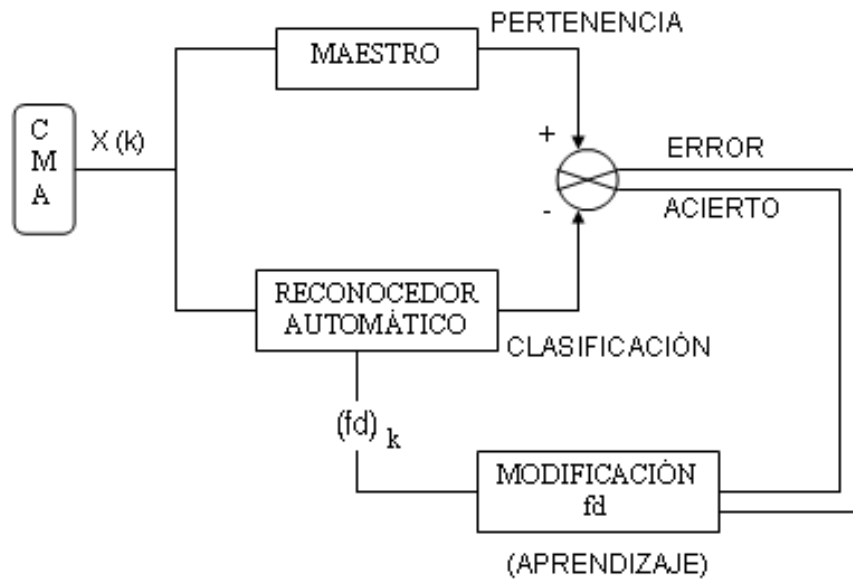


Figura 1.6: Diagrama de bloques simbólico del proceso de aprendizaje o entrenamiento de un reconocedor

### 1.4.1. Aprendizaje de funciones discriminantes por regiones

Comenzaremos por el caso más elemental de dos clases a discriminar:  $\alpha_1$  y  $\alpha_2$ . El objetivo del sistema de aprendizaje será encontrar una función discriminante  $fd$  lineal tal que:

$$FD(X) \cong W^T X = \begin{cases} > 0 & \text{si } X \in \alpha_1 \\ < 0 & \text{si } X \in \alpha_2 \end{cases} \quad (1.12)$$

el proceso de aprendizaje o entrenamiento consiste en presentarle continuamente al reconocedor adaptativo las muestras del conjunto de aprendizaje. Así, en el instante o iteración genérico  $t = k$  se presenta la muestra  $X(k)$ , de forma que la  $fd_k[X(k)]$  proporcionará la decisión correspondiente:

$$fd_k[X(k)] = W^T(k)X(k) \quad (1.13)$$

siendo  $W(k)$  el vector de ponderación en dicho instante genérico.

Al recibir la muestra  $X(k)$ , el reconocedor la clasifica en una clase u otra de acuerdo con la expresión 1.13, pudiendo acertar o equivocarse.

En caso de acierto, el algoritmo de aprendizaje no modifica el comportamiento del clasificador y por tanto se mantendrá la misma función discriminante para el siguiente vector:  $W(k+1) = W(k)$ .

Si se equivoca el algoritmo de aprendizaje modifica la función de decisión o, más exactamente, modifica el vector de ponderación de la  $fd$ .

Se pueden dar dos tipos de equivocación en el problema de la distribución biclase  $\alpha_1$  y  $\alpha_2$ , estas son:

En primer lugar, que  $X(k) \in \alpha_1$  y el reconocedor clasifique a  $X(k) \in \alpha_2$ , es decir, manejando las expresiones 1.12 y 1.13 vemos que este error se produce cuando el objetivo es que  $fd(X)$  sea positivo, pero el reconocedor maneja una  $fd_k[X(k)]$  negativa. La alternativa ahora es aumentar  $fd_{k+1}$  para la siguiente decisión.

El segundo error es el complementario del anterior. Es decir, cuando se le muestra al reconocedor un elemento  $X(k) \in \alpha_2$  que debería producir el resultado  $fd_k[X(k)] < 0$ , pero que sin embargo produce un resultado positivo. En este caso, la acción inmediata del algoritmo de aprendizaje consistirá en disminuir  $fd_{k+1}$  para la siguiente decisión. Este proceso de aumentar o disminuir la función de decisión se realiza sobre el vector de coeficientes de decisión  $W$ .

Veamos ahora en detalle como actúa el citado algoritmo de aprendizaje. Para ello supongamos que en el instante  $t = k$  el clasificador recibe una muestra  $X(k)$  perteneciente a la clase.

Los dos sucesos que pueden darse (acierto o equivocación) se sintetizan en la siguiente expresión:

$$X(k) \in \alpha_1 \begin{cases} W^T(k)X(k) > 0 & \text{ACIERTO} \Rightarrow W(k+1) = W(k) \\ W^T(k)X(k) \leq 0 & \text{ERROR} \Rightarrow W(k+1) > W(k) \end{cases} \quad (1.14)$$

Análogamente si la muestra no pertenece a  $\alpha_1$ , es decir pertenece a  $\alpha_2$ , los sucesos posibles serían los siguientes:

$$X(k) \in \alpha_2 \begin{cases} W^T(k)X(k) < 0 & \text{ACIERTO} \Rightarrow W(k+1) = W(k) \\ W^T(k)X(k) \geq 0 & \text{ERROR} \Rightarrow W(k+1) < W(k) \end{cases} \quad (1.15)$$

Existen varios procedimientos para disminuir o aumentar el vector de ponderación de la función de decisión. Vamos a plantear uno de los más utilizados, denominado algoritmo perceptrón.

El algoritmo perceptrón consiste en incrementar (disminuir) el vector de coeficientes  $W$  proporcionalmente a la muestra  $X$  que ha producido una decisión incorrecta:

$$W(k+1) = W(k) \pm \mu(k)X(k)$$

El proceso de aprendizaje consiste en los siguientes pasos:

1. Inicializar  $W(0)$
2. Repetir hasta que se dé la convergencia de las expresiones (1.14) y (1.15)
  - Clasificar los elementos del CMA
  - Calcular  $W(k)$

Hasta ahora se ha trabajado, dentro de la temática del aprendizaje de clasificadores por regiones, con reconocimiento biclase. No obstante, en la mayoría de los casos prácticos aparecen  $N$  clases a reconocer :  $\alpha_1, \alpha_2, \dots, \alpha_N$ . El análisis que sigue es todo similar al que hicimos para los clasificadores determinísticos a priori por regiones.

Para una situación genérica de  $N$  clases,  $\alpha_1, \alpha_2, \dots, \alpha_N$ , la solución óptima es conseguir para cada una de las clases un hiperplano (esto es, una función discriminante lineal) que separe esa clase del resto.

Es decir:

$$\begin{array}{l}
 \alpha_1 | \alpha_2, \alpha_3, \dots, \alpha_N; \alpha_1 | \overline{\alpha_1} \\
 \alpha_2 | \alpha_1, \alpha_3, \dots, \alpha_N; \alpha_2 | \overline{\alpha_2} \\
 \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\
 \alpha_{N-1} | \alpha_1, \alpha_2, \dots, \alpha_{N-2}, \alpha_N; \alpha_{N-1} | \overline{\alpha_{N-1}}
 \end{array}$$

En este caso óptimo se necesitarían  $N - 1$   $fd$  para resolver el problema de clasificar las  $N$  clases. Además, tendría una ventaja adicional sobre un clasificador basado en la distancia (que precisa de  $N$   $fd$ ) ya que la decisión sería más rápida.

El reconocedor por regiones debe realizar una búsqueda en forma de árbol:

En promedio, esta búsqueda tan sólo exige una carga computacional dada por :

$$\frac{N-1}{2}(n \text{ productos} + n \text{ sumas})$$

frente a la carga computacional del reconocedor por distancias:  $N(n \text{ productos} + n \text{ sumas})$  más la ordenación de  $N$  números.

En el peor de los casos, cuando no haya separabilidad lineal de ninguna de las clases con las demás, el número de  $fd$  necesarias se dispara, llegando a ser el límite superior  $N \frac{N-1}{2}$ ; es decir, es una función discriminante por cada combinación posible de dos clases entre las  $N$  existentes.

A priori no es posible encontrar la condición de separabilidad lineal de una clase con el resto para una distribución genérica de  $N$  clases. Por tanto, hay que ensayar a ciegas los algoritmos de aprendizaje de las  $fd$ . Así, se empieza por el caso óptimo; es decir, suponiendo que existe separabilidad lineal de cada clase con el resto.



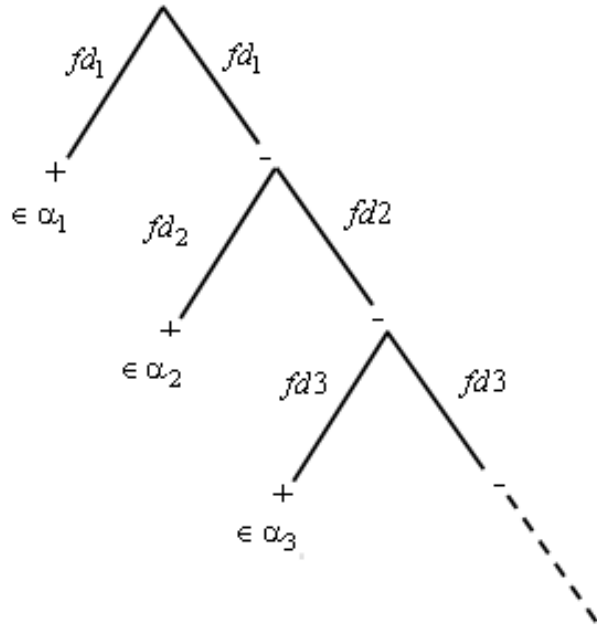


Figura 1.7: Reconocedor por regiones que realiza una búsqueda en forma de árbol.

### 1.4.2. Aprendizaje de funciones discriminantes por distancia

El objetivo del sistema de aprendizaje es converger hacia una distribución de funciones discriminantes asociada a los reconocedores por distancia.

Si para un conjunto de  $N$  clases,  $\alpha_1, \alpha_2, \dots, \alpha_N$  se aplica el reconocimiento euclídeo modificado, la solución al problema de clasificar un vector  $X$  desconocido es asociarlo a la clase cuya función discriminante  $fd(X)$  es máxima. La hipótesis de partida es que existen  $N$  funciones discriminantes,  $fd_1, fd_2 \dots fd_N$ , tales que:

$$X \in \alpha_i \text{ si, y sólo si } fd_i(X) > fd_j(X) \text{ para toda } j = 1, 2, \dots, N; j \neq i$$

El proceso de aprendizaje es muy semejante al establecido para los reconocedores por regiones. De manera sistemática consiste en los siguientes pasos.

1. El objetivo es encontrar una función discriminante por cada una de las clases presentes.

2. Cada función discriminante se considera lineal

$$fd_i(X) = W_i^T X; \quad i = 1, 2, \dots, N \quad (1.16)$$

3. se dispone de un conjunto de muestras de aprendizaje o entrenamiento:

$$CMA = Y_1, Y_2, \dots, Y_p, \quad P \gg N \quad (1.17)$$

donde,  $Y_i$  es un vector de características.

4. Se aplica un proceso iterativo de aprendizaje similar al planteado para las  $fd$  por regiones. Ahora la meta del aprendizaje viene determinada por la expresión.

$$X \in \alpha_i \text{ si, y sólo si, } fd_i(X) > fd_j(X) \text{ para todo } j = 1, 2, \dots, N; j \neq i$$

Este último punto constituye el algoritmo de aprendizaje propiamente dicho. Centrándonos en el algoritmo del perceptrón, el proceso de aprendizaje consistiría en lo siguiente:

En la interacción genérica  $t = k$  se presenta un elemento de entrenamiento  $X(k)$  tal que, para fijar ideas, pertenezca a la clase  $\alpha_1$ . Se calculan entonces todas las  $fd$  para esta muestra:

$$Fd_1[Y(k)], fd_2[Y(k)], \dots, fd_N[Y(k)] \quad (1.18)$$

de tal manera que pueden acontecer dos únicos sucesos: acierto o error, respectivamente.

ACIERTO. En tal caso se cumplirá:

$$fd_i[Y(k)] > fd_j[Y(k)] \text{ para toda } j = 1, 2, \dots, N \text{ tal que } j \neq i \quad (1.19)$$

En consecuencia, no se reajusta ninguna de las  $fd$  para la siguiente muestra a clasificar:

$$W_i(K + 1) = W_i(K), \quad i = 1, 2, \dots, N \quad (1.20)$$

ERROR. Cuando exista al menos un  $fd_j[Y(k)]$  tal que:

$$Fd_j[Y(k)] \geq fd_i[Y(k)] \text{ tal que } j \neq i \quad (1.21)$$

En este caso se ajustan únicamente las  $fd$  involucradas en la expresión 1.21. es decir, aquellas  $fd$  que den un lugar a un valor menor que el  $fd_i[X(k)]$  no se modifican.

El papel del coeficiente  $\mu(k)$  es el mismo que vimos en el aprendizaje de reconocedores por regiones.

### 1.4.3. Fundamentos de los algoritmos de aprendizaje en el reconocimiento automático de formas

Antes de pasar a una fundamentación matemática de los algoritmos de aprendizaje, vamos a resumir brevemente el planteamiento y desarrollo que hemos realizado anteriormente dentro de esta problemática.

$$Fd_i(X) = W_{i1}X_1 + W_{i2}X_2 + \dots + W_{in}X_n + W_{in+1} = W_i^T X \quad (1.22)$$

Donde,

$$W_i^T = (W_{i1}, W_{i2}, \dots, W_{in+1}) \quad y \quad X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \\ 1 \end{pmatrix}$$

Es decir, un reconocedor automático actúa sobre una combinación lineal de los valores numéricos de las características discriminantes,  $X_1, X_2, \dots, X_n$

La idea básica de este aprendizaje consiste en disponer de un conjunto de objetos  $Y_1, Y_2, \dots, Y_P$  cuya clasificación se conoce a priori y mediante un entrenamiento del reconocedor llegar a unas expresiones del tipo 1.22. este proceso lo ilustramos en la figura 1.6

Vamos ahora a profundizar en el proceso reflejado en aquella figura y manejar una más detallada (figura 1.8), en donde se especifican dos elementos clave en el aprendizaje: el error del aprendizaje, por una parte, y la acción conducente a su eliminación o, al menos, su minimización.

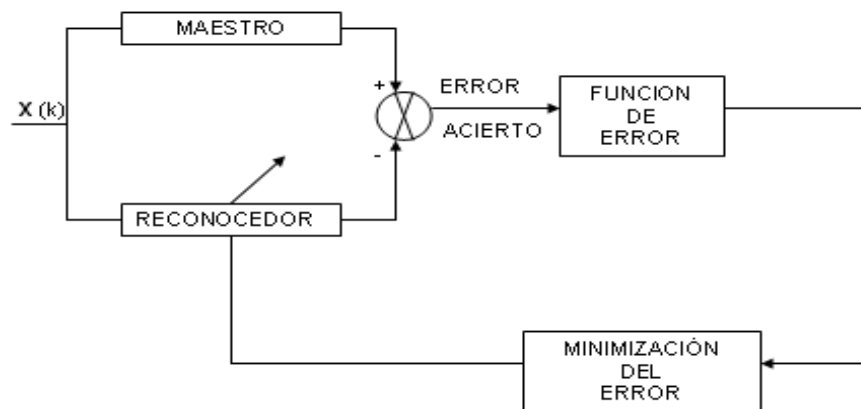


Figura 1.8: Un diagrama de bloques más detallado del proceso de aprendizaje de un reconocedor automático.

Las funciones del error que suelen manejarse son:

$$J_1 = |e|$$

$$J_2 = e^2$$

$$J_3 = E[e^2] \quad \text{valor medio}$$

Donde  $e = fd_m - W(k)X(k)$ .

Aquí hemos utilizado la notación habitual de  $J$  para representar una función o índice funcional del error. Estas funciones son, respectivamente, el modulo del error, el cuadrado

del error y el valor medio o esperanza matemática del error cuadrático.

Para los reconocedores deterministas, que son los contemplados en este capítulo, se aplican los dos primeros. En este trabajo se considerará únicamente el segundo, ya que da lugar a algoritmos de aprendizaje muy simples y robustos.

Para minimizar (o en el límite, eliminar) el error se debe obtener una expresión matemática del error o, al menos, ciertas propiedades analíticas que posibiliten su minimización.

Con este objetivo en mente, recordamos que establecíamos dos tipos o familias de reconocedores con aprendizaje dependiendo de qué meta se impusiese al reconocedor: (a) basadas en regiones y (b) basadas en distancias. En ambos casos se ensayaba una estructura lineal y dinámica,  $W^T(k)X(k)$ , del reconocedor que se modificaba en virtud de la comparación (es decir, de la creación de una función error) entre el maestro o supervisor del proceso de aprendizaje (que conocía a qué clase pertenecía la muestra de entrenamiento que se hubiese aplicado) y la clasificación del reconocedor con aprendizaje.

Por tanto, la función del error cuadrático se puede escribir como:

$$J = e^2 = (fd_m - W(k) \bullet X(k))^2 \quad (1.23)$$

En donde por  $fd_m$  simboliza la decisión del maestro.

De la expresión 1.23 se deduce que el índice de rendimiento  $J$  es como una función cuadrática de los coeficientes  $W_1, W_2, \dots, W_N$

Si este índice presenta un mínimo único se demuestra que es posible converger hacia el vector óptimo  $W_{opt}$  que minimiza  $J$  aplicando un algoritmo recursivo de actualización de los coeficientes  $W$  de la forma:

$$W(k+1) = W(k) - \mu(k) \nabla_W J(W)|_{W(k)}$$

Es decir, siguiendo una dirección opuesta al gradiente del índice del error cuadrático particularizado en el vector  $W(k)$ .

Volviendo a la problemática específica del reconocimiento automático, vamos a finalizar este capítulo con el planteamiento de un índice funcional analítico del error que nos va a conducir, aplicando la ley de adaptación del gradiente al algoritmo de perceptrón que obtuvimos de una manera intuitiva.

Así, supongamos que se trata de resolver el problema de clasificación de las dos clases  $\alpha_i, \bar{\alpha}_i$  de la figura siguiente.

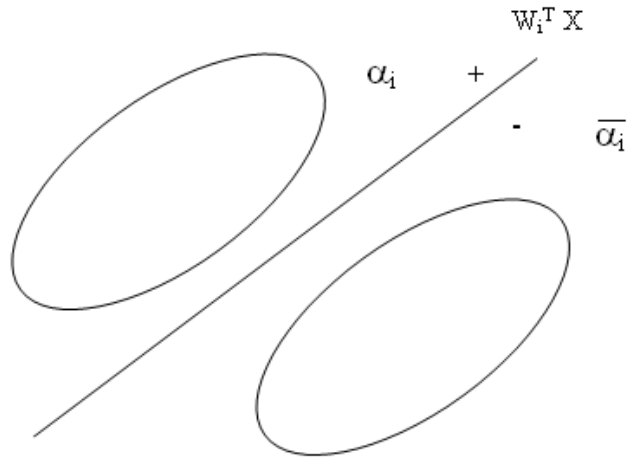


Figura 1.9: Problema de clasificación de las clases.

Consideremos el índice de error asociado al evento de que la muestra  $X$  pertenezca a la clase  $\alpha_i$ :

$$J(W_i) = \frac{1}{2} [ |W_i^T X| - W_i^T(X) ]$$

en donde por  $|\bullet|$  se indica el modulo de un escalar. Se trata, obviamente, de un procedimiento que debe converger hacia un hiperplano de separación lineal  $W_i^T X$  tal que:

$$W_i^T X > 0 \quad \text{si, y sólo si} \quad X \in \alpha_i$$

$$W_i^T X < 0 \quad \text{si, y sólo si} \quad X \in \bar{\alpha}_i$$

Lógicamente, interesa minimizar  $J(W_i)$ . El mínimo es precisamente  $J = 0$  cuando  $W_i^T X > 0$ . Por tanto, cuando se le presente al reconocedor con aprendizaje en el instante  $t = k$  un elemento  $X \in \alpha_i$ , la actualización del hiperplano de decisión para  $t = k + 1$  estará basada en el gradiente del anterior índice:

$$W_i(k+1) = W_i(k) - \mu \nabla J(W_i)|_{W_i(k)}$$

Es inmediato obtener este gradiente:

$$J = \frac{1}{2} [X(k) \text{sign}(W_i^T(k)X(k)) - X(k)] \quad (1.24)$$

Siendo la función

$$\text{sign}(\phi) = \begin{cases} 1 & \text{si, y sólo si, } \phi > 0 \\ -1 & \text{si, y sólo si, } \phi < 0 \end{cases} \quad (1.25)$$

Sustituyendo la expresión 1.25 en la 1.24, obtendremos:

$$W_i(k+1) = W_i(k) + \frac{\mu}{2} [X(k) - X(k) \text{sign}(W_i^T X)] \quad (1.26)$$

que en forma desglosada sería:

$$W_i(k+1) = W_i(k) + \mu \begin{cases} 0 & \text{si, y sólo si, } W_i^T(k)X(k) > 0 \\ X(k) & \text{si, y sólo si, } W_i^T(k)X(k) < 0 \end{cases}$$

Obsérvese que coincide con el algoritmo del perceptrón cuando la muestra que se le presenta pertenece a la clase  $\alpha_i$ .

Análogamente, para el evento de que la muestra de entrenamiento  $X(k)$  pertenezca a la clase complementaria  $\bar{\alpha}_i$ , el correspondiente índice sería:

$$J(W_i) = \frac{1}{2} [W_i^T X + W_i^T X] \quad (1.27)$$

Puede comprobarse que el mínimo  $J = 0$  se produce cuando  $W_i^T X < 0$ , lo que equivale a una clasificación correcta:  $X \in \bar{\alpha}_i$ . el gradiente de 1.26 es:

$$\nabla J(W_i) = \frac{1}{2}[X \text{sign}(W_i^T X) + X] \quad (1.28)$$

Por tanto, la actualización del hiperplano será:

$$W_i(k+1) = W_i(k) - \frac{\mu}{2}[X(k) \text{Sing}(W_i^T X) + X(k)] \quad (1.29)$$

que a su vez se puede reescribir como:

$$W_i(k+1) = W_i(k) \pm \mu \begin{cases} 0 & \text{si, y sólo si, } W_i^T(k)X(k) < 0 \\ X(k) & \text{si, y sólo si, } W_i^T(k)X(k) > 0 \end{cases}$$

que también coincide con el algoritmo del perceptrón.

Vamos a llegar al mismo resultado planteando un único índice ( en vez de los dos anteriores, que dependen de la pertenencia de la muestra de aprendizaje a una u otra clase).

Así definimos el índice de error:

$J(W_i) = \frac{1}{4}(fd_m - W_i^T X)^2 = \frac{1}{4}e^2$  en donde por  $fd_m$  indicamos la decisión del maestro:

$$fd_m = \begin{cases} 1 & \text{si, y sólo si, } x \in \alpha_i \\ -1 & \text{si, y sólo si, } x \notin \alpha_i \end{cases}$$

De forma análoga:

$$fd(X) = \begin{cases} 1 & \text{si, y sólo si, } W_i^T X > 0 \\ -1 & \text{si, y sólo si, } W_i^T X < 0 \end{cases}$$

El gradiente de éste índice es:

$$\nabla J(W_i) = \frac{1}{4}2(fd_m - W_i^T X)(-X)$$

Que puede escribirse de forma más compacta como:

$$\nabla J(W_i) = \frac{1}{2}e(-X)$$



Sustituyendo este gradiente en la expresión de la actualización del hiperplano de aprendizaje:  $W_i(k+1) = W_i(k) - \mu e[-X(k)]$ . Veamos las diferentes posibilidades que se pueden presentar y el correspondiente valor de error  $e$ .

1.  $X(k) \in \alpha_i$  y  $W_i^T(k)X(k) > 0$  Entonces  $e = 0$ ;  $W_i(k+1) = W_i(k)$
2.  $X(k) \in \bar{\alpha}_i$  y  $W_i^T(k)X(k) < 0$  Entonces  $e = 0$ ;  $W_i(k+1) = W_i(k)$
3.  $X(k) \in \alpha_i$  y  $W_i^T(k)X(k) < 0$  Entonces  $e = (+1 - (-1)) = 2$  Por tanto,  $W_i(k+1) = W_i(k) + \mu X(k)$
4.  $X(k) \in \bar{\alpha}_i$  y  $W_i^T(k)X(k) > 0$  Entonces,  $e = (-1 - (+1)) = -2$  por tanto  $W_i(k+1) = W_i(k) - \mu X(k)$

Que coincide como era de esperar, con la actualización del perceptrón cuando la clasificación ha sido incorrecta.

## 1.5. Reconocimiento estadístico a priori. El rococeador bayesiano

El clasificador automático basado en la distancia euclídea, en sus versiones a priori y con aprendizaje, utiliza en su diseño unos prototipos fijos  $Z_i$  que representan a cada una de las clases  $\alpha_i$ . Con esto estamos admitiendo una hipótesis determinística pues todos los elementos de una clase se asimilan a un único vector.

Supongamos que tenemos elementos con dos características  $X_1$  y  $X_2$  y que están divididos en dos clases  $\alpha_1$  y  $\alpha_2$ . También supongamos que la representación en un plano abstracto de las características de ambas clases es la que se muestra en la figura 1.10:

En la figura 1.10 vemos que los elementos de las clases presentan una dispersión significativa con respecto a su media, lo cual nos lleva a pensar que un enfoque determinístico no es adecuado; por ello es mejor abordar un enfoque estadístico.

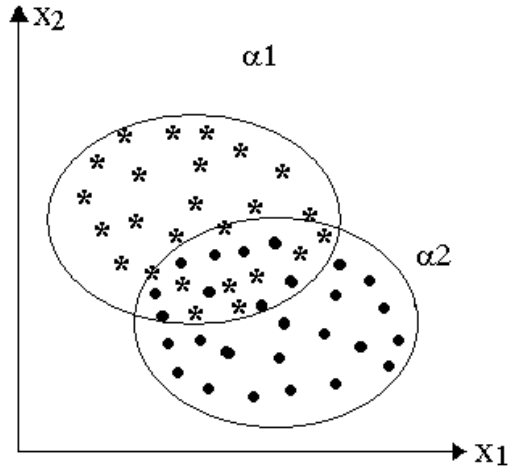


Figura 1.10: Representación en un plano abstracto de las características de  $\alpha_1$  y  $\alpha_2$

### 1.5.1. Teorema de bayes

El teorema de Bayes aplicado al reconocimiento se puede enunciar de la siguiente manera:

$$P(\alpha_i | X) = \frac{P(X | \alpha_i)P(\alpha_i)}{P(X)} \quad (1.30)$$

En donde,

- $\blacksquare$   $X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = (X_1, X_2, \dots, X_n)^T$
- $\blacksquare$   $P(\alpha_i | X)$ : Es la probabilidad de que un vector de características  $X$  pertenezca a la clase  $\alpha_i$ .
- $\blacksquare$   $P(X | \alpha_i)$ : Es la probabilidad de que dada  $\alpha_i$ , el valor de la variable aleatoria sea, precisamente  $X$ . De otra manera es la función de densidad de probabilidad (*fdp*) de la clase  $\alpha_i$  considerada como una variable aleatoria.
- $\blacksquare$   $P(\alpha_i)$ : Es la probabilidad a priori de que se presente un elemento de la clase  $\alpha_i$ .

- $P(X)$ : Es la probabilidad a priori de que se presente un objeto a clasificar con vector de características igual a  $X$  (Considerado como vector numérico concreto).

El elemento  $P(X)$  se puede despreciar puesto que no aporta información discriminante.

El termino  $P(\alpha_i | X)$ , del teorema de Bayes, aporta la solución al problema de la clasificación: Dado un vector  $X$  ¿A qué clase  $\alpha_i$  pertenece ? Entonces la clasificación será la siguiente:

$$X \in \alpha_j \iff P(\alpha_j | X) > P(\alpha_i | X) \quad \forall i \neq j, \quad i = 1, 2, \dots, N \quad (1.31)$$

La anterior clasificación se puede ver gráficamente en la figura 1.11 como sigue:

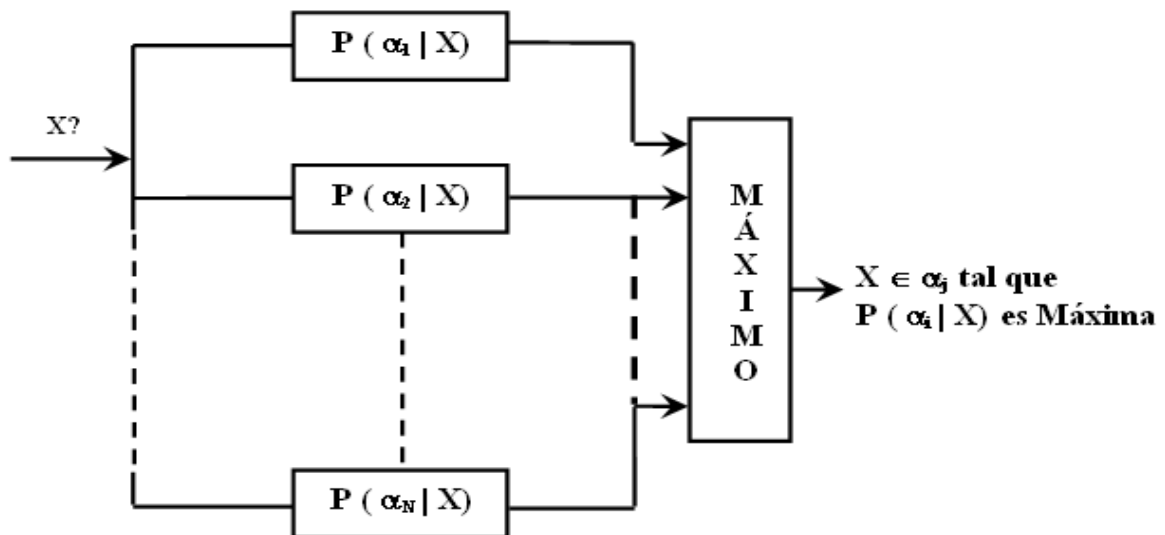


Figura 1.11: Estructura del clasificador estadístico a posteriori.

El termino del lado derecho , del teorema de Bayes, Aporta una forma opcional de clasificar un vector  $X$ ; sustituyendo dicho termino en 1.31 se obtiene:

$$X \in \alpha_j \iff P(X | \alpha_j)P(\alpha_j) > P(X | \alpha_i)P(\alpha_i) \quad \forall i \neq j, \quad i = 1, 2, \dots, N \quad (1.32)$$

La representación gráfica de esta clasificación se puede ver en la figura 1.12:

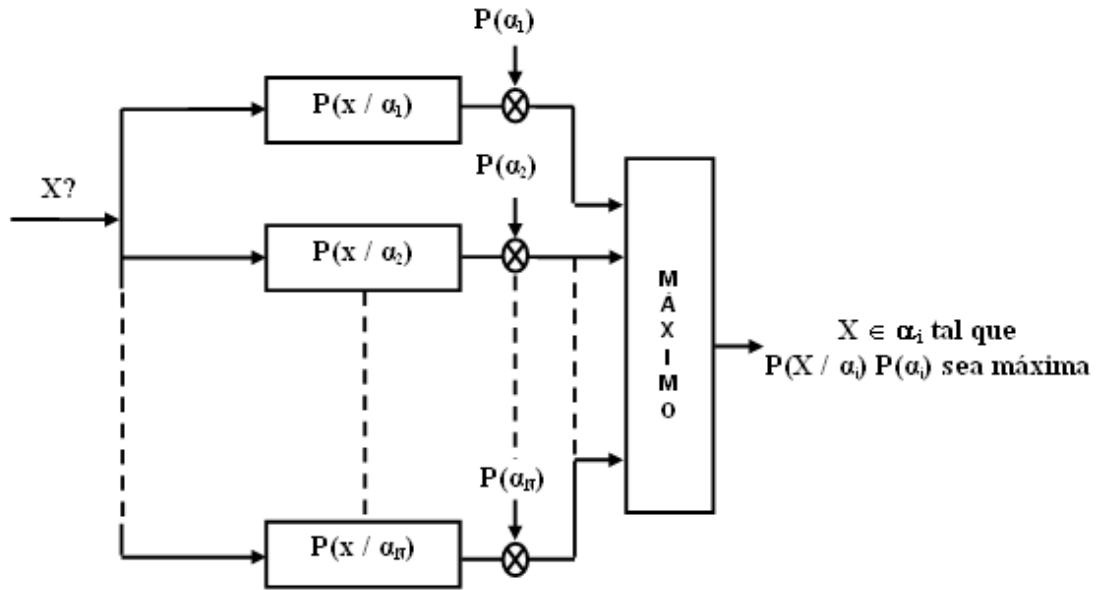


Figura 1.12: Estructura del clasificador estadístico a priori.

La Expresión 1.32 también la podemos ver como:

$$X \in \alpha_i \Leftrightarrow \frac{P(X | \alpha_i)}{P(X | \alpha_j)} > \frac{P(\alpha_j)}{P(\alpha_i)} \quad \forall i \neq j, \quad i = 1, 2, \dots, N. \quad (1.33)$$

Sea

$$L(X) = \frac{P(X | \alpha_i)}{P(X | \alpha_j)} \quad (1.34)$$

al termino  $L(X)$  se le denomina ratio de verosimilitud y al termino:

$$\frac{P(\alpha_i)}{P(\alpha_j)} \quad (1.35)$$

Se denomina valor umbral.

Algunas veces es conveniente escribirlo como  $-\text{Ln}(L(X))$ , de ahí que:

$$X \in \alpha_i \Leftrightarrow -\text{Ln}(P(X | \alpha_i)) + \text{Ln}(P(X | \alpha_j)) > \text{Ln} \left[ \frac{P(\alpha_i)}{P(\alpha_j)} \right] \quad (1.36)$$

A esta ultima se le denomina la prueba de Bayes para el mínimo error.

En general, la regla de decisión 1.36 o cualquier otra no conducen a una clasificación perfecta.

Para evaluar el funcionamiento de una regla de decisión, debemos calcular la probabilidad de error, esto es, la probabilidad de que una muestra sea asignada a la clase equivocada. Sean  $\Gamma_i$  y  $\Gamma_j$  las regiones en el dominio de  $X$  tal que  $P(\alpha_i | X) > P(\alpha_j | X)$  y  $P(\alpha_j | X) < P(\alpha_i | X)$  respectivamente. Entonces si  $X \in \Gamma_k$  asignamos la muestra a la clase  $\alpha_k$ , la probabilidad de error es:

$$\epsilon = P(\text{error}) = P(\text{error} | \alpha_i)P(\alpha_i) + P(\text{error} | \alpha_j)P(\alpha_j) \quad (1.37)$$

Si la muestra corresponde a  $\alpha_i$  hay un error siempre que  $X \in \Gamma_j$  y similarmente si la muestra corresponde a  $\alpha_j$  hay un error siempre que  $X \in \Gamma_i$  así:

$$\begin{aligned} \epsilon &= P(X \in \Gamma_j | \alpha_i)P(\alpha_i) + P(X \in \Gamma_i | \alpha_j)P(\alpha_j) \\ &= P(\alpha_i) \int_{\Gamma_j} P(X | \alpha_i) dX + P(\alpha_j) \int_{\Gamma_i} P(X | \alpha_j) dX \\ &= P(\alpha_i)\epsilon_i + P(\alpha_j)\epsilon_j \end{aligned} \quad (1.38)$$

Algunas veces es mejor integrar la función de densidad del ratio de verosimilitud  $P(L | \alpha_i)$  porque es unidimensional (mientras las otras funciones de densidad son N-dimensionales).

$$\epsilon_i = \int_0^{P(\alpha_j)/P(\alpha_i)} P(L | \alpha_i) dL \quad (1.39)$$

$$\epsilon_i = \int_{P(\alpha_j)/P(\alpha_i)}^{\infty} P(L | \alpha_j) dL \quad (1.40)$$

### 1.5.2. La media

Debido a que son muy útiles en nuestro estudio algunos términos como la media y la matriz de covarianzas, entonces se dará una pequeña definición matemática de ellos.

Tampoco queremos profundizar mucho en el tema, pues estas definiciones y muchas propiedades de la media y la matriz de covarianza se pueden encontrar en [11] o en cualquier libro de estadística y probabilidad.

Por definición tenemos que  $m = E(X)$ , entonces:

$$m = E(X) = E \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{pmatrix} \quad (1.41)$$

### 1.5.3. La matriz de covarianzas

Por definición tenemos que  $C = E \{(X - m)(X - m)^T\}$ , luego:

$$C = E \left[ \begin{pmatrix} X_1 - m_1 \\ X_2 - m_2 \\ \vdots \\ X_n - m_n \end{pmatrix} (X_1 - m_1 \quad X_2 - m_2 \quad \dots \quad X_n - m_n) \right]$$

$$C = E \left[ \begin{pmatrix} (X_1 - m_1)^2 & (X_1 - m_1)(X_2 - m_2) & \dots & (X_1 - m_1)(X_n - m_n) \\ (X_2 - m_2)(X_1 - m_1) & (X_2 - m_2)^2 & \dots & (X_2 - m_2)(X_n - m_n) \\ \vdots & \vdots & \ddots & \vdots \\ (X_n - m_n)(X_1 - m_1) & (X_n - m_n)(X_2 - m_2) & \dots & (X_n - m_n)^2 \end{pmatrix} \right]$$

Los elementos de la diagonal principal son las varianzas, así:

$$\sigma_i^2 = E(X_i - m_i)^2; i = 1, 2, \dots, n \quad (1.42)$$

$\sigma_i^2$  Son las varianzas de los elementos del vector de características.

Los demás elementos de  $C$ , que es una matriz simétrica respecto a la diagonal principal se denominan covarianzas.

### 1.5.4. El coeficiente de correlación

El coeficiente de correlación entre dos elementos  $X_i, X_j$  del vector de características se define como la razón:

$$r_{ij} = \frac{E [(X_i - m_i)(X_j - m_j)]}{E [(X_i - m_i)^2]^{1/2} E [(X_j - m_j)^2]^{1/2}} = \frac{C_{ij}}{\sigma_i \sigma_j} \quad (1.43)$$

Donde:

- $C_{ij}$  es un elemento genérico de  $C$ .
- $\sigma_i, \sigma_j$  son las desviaciones típicas de las características  $X_i, X_j$ , respectivamente.

De ahí que  $C$  se puede expresar como:

$$C = \begin{pmatrix} \sigma_1^2 & r_{12}\sigma_1\sigma_2 & \cdots & r_{1n}\sigma_1\sigma_n \\ r_{21}\sigma_2\sigma_1 & \sigma_2^2 & \cdots & r_{2n}\sigma_2\sigma_n \\ \vdots & & & \\ r_{n1}\sigma_n\sigma_1 & r_{n2}\sigma_n\sigma_2 & \cdots & \sigma_n^2 \end{pmatrix} \quad (1.44)$$

Si  $r_{ij} = 0$ , se dice que las características  $X_i, X_j$ , están incorrelacionadas, lo cual es una propiedad siempre deseable; esto es, las características deben escogerse de tal manera que  $C$  sea diagonal principal pura. La matriz de covarianza también se puede ver como:

$$\begin{aligned} C &= E [(X - m)(X - m)^T] \\ &= E [XX^T - Xm^T - mX^T + mm^T] \\ &= E(XX^T) - E(Xm^T) - E(mX^T) + E(mm^T) \\ &= E(XX^T) - m^T E(X) - mE(X^T) + mm^T \\ &= E(XX^T) - mm^T \end{aligned} \quad (1.45)$$

### 1.5.5. Estimación estadística de la matriz de covarianza y del vector media

Si tenemos una clase genérica y disponemos de  $P$  muestras de dicha clase, es decir:

$$X_1, X_2, \dots, X_p$$

la mejor estimación del vector media de la clase (considerada como una variable aleatoria) es:

$$\hat{m} = \frac{1}{p} \sum_{j=1}^p X_j \quad (1.46)$$

La estimación de la matriz de covarianza será:

$$\hat{C} = \frac{1}{p} \sum_{j=1}^p X_j X_j^T - \hat{m} \hat{m}^T \quad (1.47)$$

En donde el primer término es la estimación de la media (esperanza matemática) de  $XX^T$  y el segundo término, la estimación del término  $mm^T$ .

En sentido estricto, debe aplicarse, el estimador insesgado de la matriz de covarianza, es decir, aquel cuya media coincide con la verdadera matriz de covarianza:

$$\hat{C} = \frac{1}{p-1} \sum_{j=1}^p X_j X_j^T - \frac{1}{p-1} \hat{m} \hat{m}^T \quad (1.48)$$

### 1.5.6. El clasificador estadístico a priori

Como primer paso vamos a restringir nuestro estudio al caso de hipótesis de distribuciones normales o gaussianas, pues es la que se presenta en la mayoría de los casos prácticos.

#### Caso bidimensional

Cuando el vector característico maneja dos rasgos su distribución es bidimensional. Si suponemos que las características están incorrelacionadas la distribución tiene la forma:

$$P(X | \alpha_i) = \frac{1}{2\pi\sigma_{i1}\sigma_{i2}} \exp \left( -\frac{1}{2} \left[ \frac{(X_1 - m_{i1})^2}{\sigma_{i1}^2} + \frac{(X_2 - m_{i2})^2}{\sigma_{i2}^2} \right] \right) \quad (1.49)$$

donde



- $m_{i1}$  y  $m_{i2}$  son las medias de  $X_1$  y  $X_2$  respectivamente para la clase  $\alpha_i$ .
- $\sigma_{i1}$  y  $\sigma_{i2}$  son las desviaciones típicas de  $X_1$  y  $X_2$  respectivamente para la clase  $\alpha_i$ .

### Caso N-dimensional

Cuando el vector de características es n-dimensional, la  $fdp$  es de la forma:

$$P(X | \alpha_i) = \frac{1}{(2\pi)^{\frac{n}{2}} |C_i|^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2}(X - m_i)^T C^{-1}(X - m_i)\right) \quad (1.50)$$

donde

- $m_i$  es el vector media o esperanza matemática de  $\alpha_i$ .
- $C_i$  es la matriz de covarianzas.
- $|C_i|$  es el determinante de la matriz  $C_i$ .

Además,  $\alpha_i$  se considera como una variable aleatoria.

También,

$$P(\alpha_i)P(X | \alpha_i) = P(\alpha_i) \frac{1}{(2\pi)^{\frac{n}{2}} |C_i|^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2}(X - m_i)^T C^{-1}(X - m_i)\right) \quad (1.51)$$

Sustituyendo 1.51 en  $P(X | \alpha_i)P(\alpha_i) > P(X | \alpha_j)P(\alpha_j)$   $i \neq j$ ;  $j = 1, 2, \dots, N$  aplicando logaritmos neperianos, haciendo algunas manipulaciones algebraicas y suponiendo que todas las clases son equiprobables a priori (lo cual nos permite eliminar el término  $P(\alpha_i)$ ) la función discriminante, para la clase  $\alpha_i$ , será:

$$fd_i(X) = Ln \left[ \frac{|C_j|}{|C_i|} \right]^{1/2} + \frac{1}{2} [(X - m_j)^T C_j^{-1}(X - m_j) - (X - m_i)^T C_j^{-1}(X - m_i)] \quad (1.52)$$

El termino  $-\frac{n}{2}Ln2\pi$  no aporta información discriminante, por ello lo hemos eliminado de la expresión 1.52.

Vamos a suponer las siguientes hipótesis:

1. Las matrices de covarianza de las clases son todas diferentes:  $C_i \neq C_j \quad \forall i \neq j$ ;  $i, j = 1, 2, \dots, N$  y , en general, completas (es decir, sin elementos nulos). Debido a lo anterior se tiene que las funciones discriminantes no son lineales, ya que aparecen todas las combinaciones cuadráticas posibles de las componentes del vector de características.
2. Las matrices de covarianza de las clases son todas iguales:  $C_i = C_j \quad \forall i \neq j$ ;  $i, j = 1, 2, \dots, N$  y , en general, completas. Según esta hipótesis, podemos eliminar los términos no lineales de las funciones discriminantes, ya que se repiten en todas las clases y, por tanto, no poseen un carácter discriminante. Las funciones quedan de la siguiente manera:

$$fd_i(X) = X^T C^{-1} m_i - \frac{1}{2} m_i^T C^{-1} m_i \quad (1.53)$$

Esto implica que son lineales.

3. Las matrices de covarianza son iguales y las características son estadísticamente independientes y además tienen la misma desviación típica. Las funciones discriminantes son de la forma:

$$fd_i(X) = X^T m_i - \frac{1}{2} m_i^T m_i \quad (1.54)$$

Aquí se ha eliminado el factor  $C_i^{-1} = \frac{1}{\sigma^2} I$

4. Las matrices de covarianza son iguales, las características son estadísticamente independientes y sus desviaciones típicas son distintas. Las funciones discriminantes son de la forma:

$$fd_i(X) = X^T C^{-1} m_i - \frac{1}{2} m_i^T C^{-1} m_i \quad (1.55)$$

que siguen siendo funciones lineales.

Si tenemos dos clases  $\alpha_i, \alpha_j$  con vectores esperados  $m_i, m_j$ , matrices de covarianza  $C_i, C_j$  respectivamente y aplicamos la regla de decisión 1.36 obtenemos:

$$X \in \alpha_i \Leftrightarrow \frac{1}{2}(X - m_i)^T C_i^{-1}(X - m_i) - \frac{1}{2}(X - m_j)^T C_j^{-1}(X - m_j) + \frac{1}{2}Ln \frac{|C_i|}{|C_j|} < Ln \frac{P(\alpha_i)}{P(\alpha_j)} \quad (1.56)$$

Si las matrices de covarianza son iguales ( $C_i = C_j = C$ ) entonces:

$$X \in \alpha_i \Leftrightarrow \frac{1}{2}(X - m_i)^T C^{-1}(X - m_i) - \frac{1}{2}(X - m_j)^T C^{-1}(X - m_j) < Ln \frac{P(\alpha_i)}{P(\alpha_j)} \quad (1.57)$$

### 1.5.7. Diseño de un reconocedor estadístico

Supongamos que tenemos dos características  $X_1, X_2$  y dos clases  $\alpha_1, \alpha_2$  que de ahora en adelante las notaremos como  $A$  y  $B$  respectivamente cuya distribución de muestras en el espacio abstracto es la que se muestra en la figura 1.13.

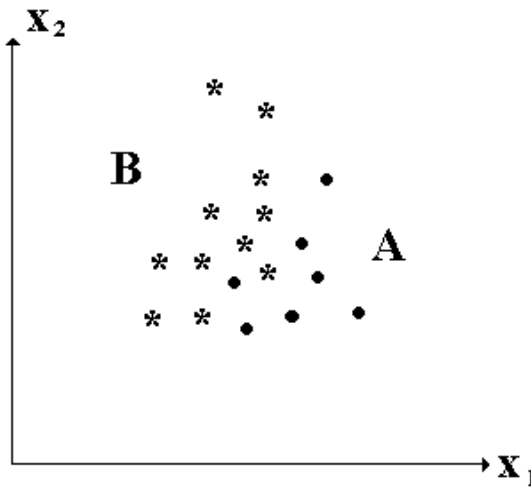


Figura 1.13: Representación en un plano de diferentes muestras de ambas clases.

Se puede observar que no hay una separabilidad lineal entre las dos clases. Supongamos que:

1. Ambas clases son equiprobables a priori  $P(A) = P(B)$ .
2.  $X_1$  y  $X_2$  son estadísticamente independientes en ambas clases. Esto significa que las matrices de varianzas son diagonales puros:

$$C_A = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}; C_B = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

3. Las desviaciones típicas de  $X_1$  y  $X_2$  son iguales tanto como para cada clase como entre si, es decir:

$$\sigma_{A_1} = \sigma_{A_2} = \sigma_{B_1} = \sigma_{B_2} = \sigma$$

Las hipótesis 2 y 3 implican que las funciones densidad de probabilidad de ambas clases son:

$$P(X | A) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2} \|(X - m_A)\|^2\right)$$

$$P(X | B) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2} \|(X - m_B)\|^2\right)$$

En donde:

$$\|(X - m_A)\|^2 = (X - m_A)^T (X - m_A) = (X_1 - m_{A_1})^2 + (X_2 - m_{A_2})^2$$

$$\|(X - m_B)\|^2 = (X - m_B)^T (X - m_B) = (X_1 - m_{B_1})^2 + (X_2 - m_{B_2})^2$$

Es decir, son los cuadrados de las distancias euclídeas entre  $X$  y  $m_A$  y  $m_B$  respectivamente. Con las tres hipótesis restrictivas, la decisión del clasificador Bayesiano se basa en lo siguiente:

$$X \in A \Leftrightarrow \|(X - m_A)\|^2 < \|(X - m_B)\|^2$$

Es decir:

$$X \in A \Leftrightarrow d_E(X - m_A) < d_E(X - m_B)$$

Y este último es el clasificador determinístico basado en la distancia euclídea a los centroides (es decir, las medias) de las dos clases a discriminar. Ver la figura 1.14.

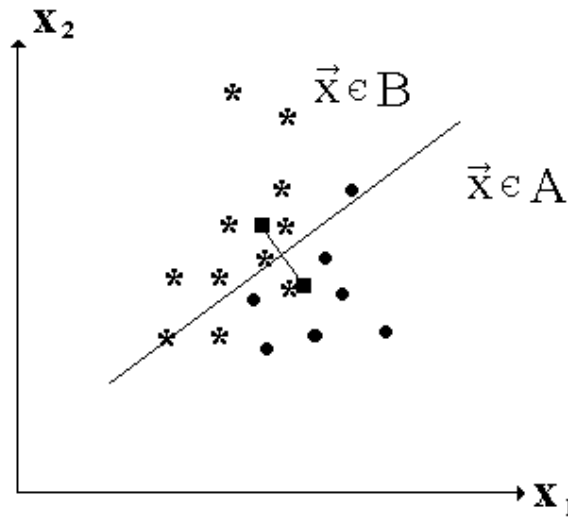


Figura 1.14: Representación de la función discriminante lineal del clasificador Bayesiano cuando se cumplen las restricciones 1,2,3.

Vemos que cuando se admite una serie de hipótesis, el clasificador Bayesiano coincide con el clasificador determinístico basado en la distancia euclídea. La razón por la cual ocurre esto es por que las hipótesis 1,2 y 3 son tan restrictivas que eliminan toda posible naturaleza aleatoria respecto al comportamiento de las clases.

Si ahora suponemos que  $\sigma_1 \neq \sigma_2$ , pero que son iguales en ambas clases, entonces las  $fdp$  en forma general serán:

$$P(X | A) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{1}{2}(X - m_A)^T C^{-1}(X - m_A)\right)$$

$$P(X | B) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{1}{2}(X - m_B)^T C^{-1}(X - m_B)\right)$$

En donde:

$$C_A = C_B = C = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

Como  $P(A) = P(B)$  entonces:

$$P(X | A) = P(X | B) \Leftrightarrow (X - m_A)^T C^{-1} (X - m_A) = (X - m_B)^T C^{-1} (X - m_B)$$

De ahí que:

$$X^T C^{-1} (m_A - m_B) = \frac{1}{2} (m_A + m_B)^T C^{-1} (m_A - m_B)$$

Y esta es la ecuación de una recta que pasa por  $\frac{1}{2}(m_A + m_B)$  (El punto medio de los centroides  $m_A$ ,  $m_B$ ) con una pendiente normal a  $C^{-1}(m_A - m_B)$  que en general es diferente a  $90^\circ$ . La igualdad se da cuando  $C = \sigma^2 I$ , siendo  $I$  la matriz identidad; es decir las varianzas de  $X_1$  y  $X_2$  son iguales.

En la figura 1.15 puede notarse la mejora que ha tenido el clasificador Bayesiano cuando se tienen las tres hipótesis sin tanta restricción como al comienzo.

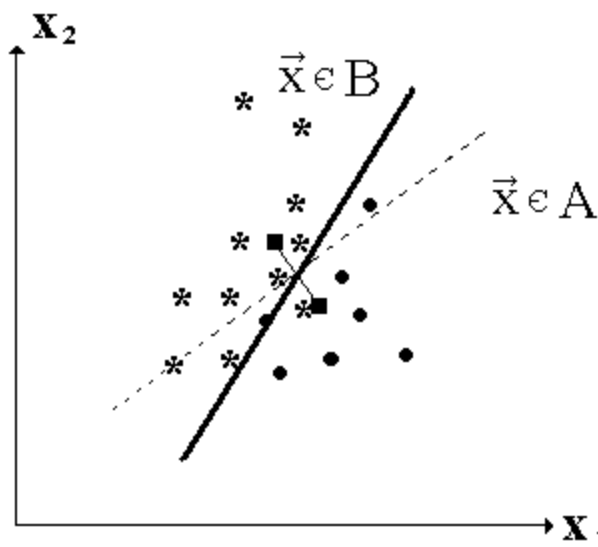


Figura 1.15: Nueva función discriminante cuando las desviaciones típicas de las características difieren.

La recta pasa por el punto medio del segmento que une los centroides con una pendiente inclinada en la dirección de la característica con mayor dispersión.

### 1.5.8. La regla de decisión de Bayes para el riesgo mínimo

Supongamos que, para una decisión dada, debemos pagar un costo que depende de la verdadera clase a la cual corresponde la muestra (o el objeto a clasificar). Entonces hay cuatro casos:

1.  $C_{11}$  = Costo al decidir que  $X \in \alpha_i$  cuando  $X \in \alpha_i$
2.  $C_{12}$  = Costo al decidir que  $X \in \alpha_j$  cuando  $X \in \alpha_i$
3.  $C_{21}$  = Costo al decidir que  $X \in \alpha_i$  cuando  $X \in \alpha_j$
4.  $C_{22}$  = Costo al decidir que  $X \in \alpha_j$  cuando  $X \in \alpha_j$

Se asume que una mala decisión es mas costosa que una decisión correcta, esto es,

$$C_{12} > C_{11} \quad y \quad C_{21} > C_{22} \quad (1.58)$$

Sea  $\Gamma_i$  y  $\Gamma_j$  las regiones en el dominio de  $X$  para las cuales decidimos que  $X \in \alpha_i$  y  $X \in \alpha_j$  respectivamente. Queremos escoger  $\Gamma_i$  y  $\Gamma_j$  de tal forma que minimicen el costo esperado o el riesgo  $r$ .

$$\begin{aligned} r &= E \{ \text{costo total} \} \\ &= \int_{\Gamma_i} C_{11} P(\alpha_i | X) P(X) dX + \int_{\Gamma_j} C_{12} P(\alpha_i | X) P(X) dX \\ &\quad + \int_{\Gamma_i} C_{21} P(\alpha_j | X) P(X) dX + \int_{\Gamma_j} C_{22} P(\alpha_j | X) P(X) dX \\ &= \int_{\Gamma_i} \{ C_{11} P(\alpha_i) P(X | \alpha_i) + C_{21} P(\alpha_j) P(X | \alpha_j) \} dX \\ &\quad + \int_{\Gamma_j} \{ C_{12} P(\alpha_i) P(X | \alpha_i) + C_{22} P(\alpha_j) P(X | \alpha_j) \} dX \end{aligned} \quad (1.59)$$

Ya que  $\Gamma_i$  y  $\Gamma_j$  no se solapan y cubren el dominio entero, entonces

$$\int_{\Gamma_i} P(X | \alpha_i) dX = 1 - \int_{\Gamma_j} P(X | \alpha_i) dX \quad (1.60)$$

Usando 1.59 y 1.60 tenemos lo siguiente:

$$r = C_{12}P(\alpha_i) + C_{22}P(\alpha_j) + \int_{\Gamma_i} \{-(C_{12} - C_{11})P(\alpha_i)P(X | \alpha_i) + (C_{21} - C_{22})P(\alpha_j)P(X | \alpha_j)\} dX \quad (1.61)$$

Ahora debemos escoger  $\Gamma_i$  de tal manera que  $r$  sea mínimo. Supongamos que para un valor dado de  $X$  el integrando en 1.61 es negativo, entonces disminuimos  $r$  asignando  $X$  a  $\Gamma_j$ . Así la regla de decisión del mínimo riesgo asigna a  $\Gamma_i$  únicamente los  $X$  para los cuales el integrando de 1.61 es negativo. Esta regla de decisión la podemos declarar de la siguiente manera:

$$(C_{21} - C_{22})P(\alpha_j)P(X | \alpha_j) \geq (C_{12} - C_{11})P(\alpha_i)P(X | \alpha_i) \Rightarrow \begin{cases} X \in \alpha_j \\ X \in \alpha_i \end{cases} \quad (1.62)$$

O también:

$$\frac{P(X | \alpha_i)}{P(X | \alpha_j)} \geq \frac{(C_{21} - C_{22})P(\alpha_j)}{(C_{12} - C_{11})P(\alpha_i)} \Rightarrow \begin{cases} X \in \alpha_i \\ X \in \alpha_j \end{cases} \quad (1.63)$$

Donde  $(C_{21} - C_{22})$  y  $(C_{12} - C_{11})$  son positivos por 1.58. Esta regla de decisión se llama prueba de Bayes para el mínimo riesgo.

## 1.6. Reconocimiento automático de formas con aprendizaje en condiciones estadísticas

Se le ha dado el calificativo de a posteriori al primer miembro de la ecuación 1.30 puesto que se trata de calcular, o al menos estimar, la probabilidad de que  $X \in \alpha_i$ , la cual sólo se puede realizar mediante un proceso de aprendizaje.

Así que a los reconocedores estadístico a priori y con aprendizaje se les puede calificar de bayesianos al estar ligados entre si, por el teorema de Bayes.



$P(\alpha_i|X)$ : Da la solución al problema de clasificación.

### 1.6.1. Clasificación

Ataquemos el problema de Reconocimiento Automático de Formas a través de la estimación de las denominadas probabilidades a posteriori  $P(\alpha_i|\mathbf{X})$ , esto es;

$$X \in \alpha_i \Leftrightarrow P(\alpha_i|\mathbf{X}) > P(\alpha_j|\mathbf{X}) \quad (1.64)$$

Para todo  $i, j = 1, 2, \dots, N, i \neq j$

### 1.6.2. Algoritmo de aprendizaje de funciones discriminantes en condiciones estadísticas

El planteamiento general de este nuevo enfoque coincide con el aprendizaje en condiciones determinísticas.

Partimos de un conjunto de elementos de entrenamiento por cada clase,

$$\{\mathbf{Y}_{11}, \mathbf{Y}_{12}, \dots, \mathbf{Y}_{1P}\}, \dots, \{\mathbf{Y}_{N1}, \mathbf{Y}_{N2}, \dots, \mathbf{X}_{NP}\} \quad (1.65)$$

Cuya pertenencia es obviamente conocida, y además la clave en el diseño de los reconocedores con aprendizaje, nace a partir de una expresión concreta de las probabilidades a posteriori  $P(\alpha_i|\mathbf{X})$  que se van a estimar (o enseñar) si se prefiere, basándose en las muestras de entrenamiento (1.65).

De las funciones de densidad de probabilidad a posteriori  $P(\alpha_i|\mathbf{X})$ , se consideran por sencillez aproximaciones lineales, esto es;

$$P(\alpha_i|\mathbf{X}) \approx W_i^T \mathbf{X} \quad (1.66)$$

Con la restricción de:

$$\sum P(\alpha_i|\mathbf{X}) = 1 \quad (1.67)$$

Y esto se da, cuando las matrices de covarianzas de todas las clases son iguales. La situación no se puede considerar como general, pero sí al menos bastante realista. La ventaja de las aproximaciones lineales, es la sencillez de los correspondientes algoritmos de aprendizaje y además la mínima carga computacional en tiempo real de operación del clasificador.

Una vez admitidas las hipótesis de aproximación lineal de las *fdp* a posteriori, el siguiente paso es establecer un índice funcional de rendimiento del reconocedor con aprendizaje, es decir, un índice de error a minimizar, donde el objetivo del proceso de aprendizaje será obtener funciones discriminantes lineales.

$$fd_i(\mathbf{X}) = W_i^T(\mathbf{X}) \approx P(\alpha_i|\mathbf{X}) \quad (1.68)$$

Tales que, la función de distribución de probabilidades sea muy próxima a 1 si  $\mathbf{X} \in \alpha_i$  y 0 si  $\mathbf{X} \notin \alpha_i$ .

El maestro o supervisor del aprendizaje se va a modelar con las probabilidades a posteriori, 1 si  $\mathbf{X} \in \alpha_i$  y 0 si  $\mathbf{X} \notin \alpha_i$ , como una variable aleatoria  $\Gamma$  tal que:

$$\Gamma_i(\mathbf{X}) = \begin{cases} 1 \Leftrightarrow \mathbf{X} \in \alpha_i \\ 0 \Leftrightarrow \mathbf{X} \notin \alpha_i \end{cases} \quad (1.69)$$

Para todo  $i, j = 1, 2, \dots, N$

Entonces existirá una variable aleatoria para cada clase.

Usando el resultado anterior, se podrán definir varios índices de error en media.

Miremos los índices de error en media:

1. El índice de error modular

$$J_{W_i}(\mathbf{X}) = E(|\Gamma_i(\mathbf{X}) - W_i^T \mathbf{X}|) \quad (1.70)$$

Para todo  $i, j = 1, 2, \dots, N$

2. El índice del error cuadrático

$$J_{W_i}(\mathbf{X}) = E \{(\Gamma_i(\mathbf{X}) - W_i^T \mathbf{X})^2\} \quad (1.71)$$

Para todo  $i, j = 1, 2, \dots, N$

La minimización de los anteriores índices es igual a la minimización de los índices de error que se manejan en la hipótesis determinística. Se aplica el algoritmo recursivo de actualización de los coeficientes  $W_i$  de la forma:

$$W(k+1) = W(k) - \mu(k) \nabla J_W(\mathbf{X})|_{W(k)} \quad (1.72)$$

Si este índice es unimodal (es decir, que presenta un mínimo único) es posible converger hacia el vector óptimo  $W_{opt}$ , que minimiza  $J_W$ , aplicando el algoritmo recursivo de actualización de los coeficientes  $W$  de la forma (1.67), es decir siguiendo la dirección opuesta del gradiente del error; además  $\mu(k)$  juega un papel vital, dado que cuantifica el grado de alejamiento hacia el mínimo y en este se exigen unas propiedades más restringidas<sup>1</sup>

$$\lim_{k \rightarrow \infty} \mu(k) = 0 \quad (1.73)$$

$$\sum_{i=1}^{\infty} \mu(k) = \infty \quad (1.74)$$

$$\sum_{i=1}^{\infty} \mu(k)^2 < \infty \quad (1.75)$$

Una función muy simple y muy utilizada que cumple estas tres restricciones es:

$$\mu(k) = \frac{1}{k} \quad (1.76)$$

La minimización de los anteriores índices es igual a la minimización de los índices de error que se manejan en la hipótesis determinística.

---

<sup>1</sup>Donde la justificación de estas restricciones no se presentará en este escrito puesto que se desborda ampliamente el contenido del mismo

Vamos a ver qué forma toma este algoritmo recursivo para cada uno de los anteriores índices de error, para ello calculamos el gradiente de los anteriores índices de error: Para el índice modular (1.70) el gradiente resulta ser:

Sea  $\epsilon(k) = \Gamma_i \mathbf{X} - W_i^T \mathbf{X}$ :

$$\nabla_{W_i}(\mathbf{X}) = -\mathbf{X} \text{sig}(|\Gamma_i(\mathbf{X}) - W_i^T \mathbf{X}|) \quad (1.77)$$

Donde el gradiente es  $-\mathbf{X}$ , de la expresión que se tiene en módulo y puesto que es modular definimos la función signo como:

$$\text{sig}(\epsilon(k)) = \begin{cases} +1 \Leftrightarrow \epsilon(k) > 0 \\ -1 \Leftrightarrow \epsilon(k) < 0 \end{cases} \quad (1.78)$$

Por lo tanto el algoritmo de aprendizaje (1.72) tomará la forma:

$$W_i(k+1) = \begin{cases} W_i(k) + \mu(k)\mathbf{X}(k) \Leftrightarrow \Gamma_i(\mathbf{X}) > W_i^T \mathbf{X} \\ W_i(k) - \mu(k)\mathbf{X}(k) \Leftrightarrow \Gamma_i \mathbf{X} < W_i^T \mathbf{X} \end{cases} \quad (1.79)$$

Como el objetivo ideal es conseguir que la función discriminante de la clase  $\alpha_i$ :

$$fd_i(X) = W_i^T X \quad (1.80)$$

coincida con la variable  $\Gamma_i(X)$  tal que:

$$\Gamma_i(X) = \begin{cases} +1 \Leftrightarrow X \in \alpha_i \\ -1 \Leftrightarrow X \notin \alpha_i \end{cases} \quad (1.81)$$

Para el segundo índice de error, basado en el error cuadrático medio,

$$J_{W_i}(X) = \frac{1}{2} E \{(\Gamma_i(X) - W_i^T X)^2\} \quad (1.82)$$

Su gradiente es:

$$\nabla_{W_i}(\mathbf{X}) = -X \{\Gamma_i(X) - W_i^T X\} = -X\epsilon(K) \quad (1.83)$$

Como el argumento de  $E$  es constante para un  $X(K)$  concreto puede eliminarse el operador esperanza matemática. De ahí, que la actualización de los coeficientes  $W_i$ , de las funciones discriminantes llega a ser:

$$W_i(k+1) = \begin{cases} W_i(k) + \mu(k)X(k)\epsilon(k) & \Leftrightarrow \Gamma_i(X) > W_i^T X \\ W_i(k) - \mu(k)X(k)\epsilon(k) & \Leftrightarrow \Gamma_i(X) < W_i^T X \end{cases} \quad (1.84)$$

Podemos ver que el algoritmo incrementa el vector de coeficientes de las funciones discriminantes  $W$ , cuando la decisión o clasificación real haya estado por debajo de la clasificación ideal o correcta. Análogamente disminuirá dicho vector, cuando la decisión o clasificación real esté por encima de la clasificación ideal o correcta.

Así que los incrementos o disminuciones de los  $W_i$ , que vienen ponderados por el error de la clasificación, vienen dados por la expresión (1.71).

En otras palabras para errores pequeños se aplican modificaciones pequeñas y para grandes errores grandes modificaciones.

La figura 1.16 representa de manera esquemática este proceso de diseño

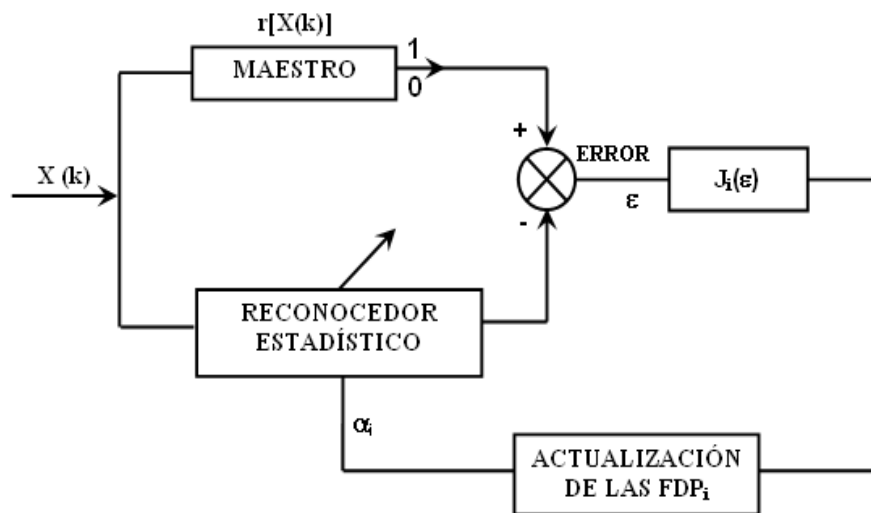


Figura 1.16: Esquema de bloques del reconocedor estadístico con aprendizaje.

El planteamiento global entre el esquema de aprendizaje estadístico y el propuesto en condiciones determinísticas son similares. Aunque presentan una diferencia importante, el reconocedor estadístico solo se refiere a la clase genérica  $\alpha_i$ . El proceso de aprendizaje o de actualización de las  $fdp$  se realiza individualmente sobre cada clase, aunque la muestra de entrenamiento  $X(K)$  se aplica simultáneamente a todas las  $fdp$ , es decir es un aprendizaje multiclase.

La figura 1.17 pretende simbolizar este aprendizaje simultáneo de todas las funciones discriminantes de todas las clases.

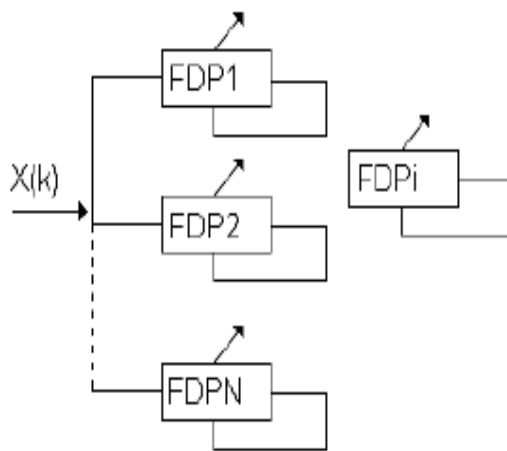


Figura 1.17: Esquemización del proceso de aprendizaje simultáneo multiclase.

Uno de los aspectos en los que se diferencia el proceso de aprendizaje determinístico con el estadístico, es que la respuesta que emite el maestro ahora es binaria y de carácter aleatorio.

De ahí actualizándose simultáneamente todas las funciones discriminantes (1.67), otro aspecto importante es la elección y el manejo del conjunto de muestras de entrenamiento. Para garantizar la convergencia correcta de los correspondientes algoritmos de aprendizaje el número de muestras de entrenamiento debe estar repartido de manera equitativa entre clases. En cuanto a la secuencia de presentación de las muestras de entrenamiento es mejor hacerlo de manera salteada y secuencial esto es:

$$\alpha_1, \alpha_2, \dots, \alpha_N, \alpha_1, \alpha_2, \dots, \alpha_N, \dots \quad (1.85)$$

Otro problema es la inicialización del vector de coeficientes  $W_i(0)$  de cada clase, en el caso estadístico no se dispone de una inicialización privilegiada como el caso determinístico, por lo que es más aconsejable comenzar por la llamada situación de indiferencia:

$$W_i(0) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (1.86)$$

## 1.7. Algoritmos de agrupación de clases (Clustering)

La técnica de agrupación se utiliza cuando no existe un conocimiento suficiente acerca de las clases en que se pueden distribuir los objetos de interés. Esto se puede dar en ciertas aplicaciones de biología, medicina, sociología, etc., en donde no se encuentran bien definidas las clases.

Los cuatro métodos que se han estudiado difieren entre si en dos puntos esenciales: La naturaleza determinística o aleatoria de las clases y el procedimiento concreto para la obtención de las funciones discriminantes, que pueden ser a priori o mediante aprendizaje.

De una manera grafica podríamos representar esta situación según el esquema siguiente.

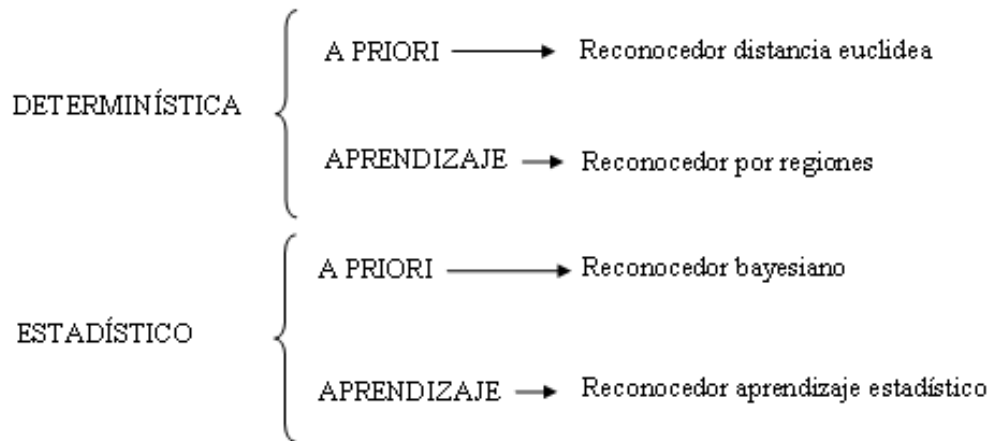


Figura 1.18: Esquemización de los cuatro métodos estudiados

En cualquier caso, la propiedad común a todos ellos es que se trata de métodos supervisados.

La única información que requiere los algoritmos de agrupación es la definición previa del vector de características. Algunos de estos algoritmos, a lo sumo precisan conocer también el número de clases.

Una vez establecido el vector de características, los procedimientos de agrupación de clases reciben como datos de entrada los objetos (convertidos lógicamente en vectores numéricos) a clasificar de modo que a partir de estos datos de entrada el algoritmo, sin supervisión de ningún tipo y de forma autónoma, agrupa esos vectores en clases (clusters o nubes). Por esta razón se les han denominado también algoritmos de clasificación autoorganizada.

También son de gran interés práctico en otras áreas (análisis automático de la voz, visión artificial, clasificación de señales, reconocimiento de huellas digitales, etc.) en las que existe un conocimiento completo de las clases y, por tanto se pueden aplicar los métodos supervisados. En estos casos, una vez diseñado el reconocimiento automático, los algoritmos de agrupación se utilizan como herramientas de verificación y prueba de la calidad del reconocedor y, en particular, del vector de características escogido.



Si se introducen como datos de entrada del algoritmo de agrupación los vectores del conjunto de muestras que se emplearon para el diseño del correspondiente clasificador supervisado, entonces los resultados que se obtengan como la salida del algoritmo de agrupación indicaran la calidad del vector de características. Si la agrupación efectuada por aquel coincide con el de las clases reales, entonces el vector estará bien escogido; en caso contrario (solapamiento o agrupaciones incorrectas) se trata de un vector de características con propiedades indeseables. Esta técnica de agrupación se ha utilizado como alternativa completa al diseño supervisado, aunque en general no dan lugar a mejores resultados, si no más bien todo lo contrario.

Los algoritmos de agrupación varían entre si por el mayor o menor grado de reglas heurísticas que utilizan e, inversamente por el nivel de procedimiento formales involucrados. Todos ellos se basan en el empleo sistemático de las distancias entre los vectores (o elementos que se van a agrupar) así como entre los clusters o grupos que se van haciendo o deshaciendo a lo largo del proceso de actuación del algoritmo en concreto.

Habitualmente se emplea la distancia euclídea entre los vectores.

$$d_E(\mathbf{X}_i, \mathbf{X}_j) = \sqrt{\sum_{K=1}^N (\mathbf{X}_{i_k} - \mathbf{X}_{j_k})^2} \quad (1.87)$$

Los algoritmos de agrupación que se estudian con mucho interes se conocen entre ellos; a saber:

1. Algoritmo de la distancia encadenada (Chaín map).
2. Algoritmo máx.-min.
3. Algoritmo de K-medias.
4. Algoritmo ISODATA y la variante del mismo K-ISODATA

Nos centraremos en la tecnica del algoritmo de agrupacion k-medias.

### 1.7.1. Algoritmo K-Medias

El nombre de este algoritmo hace referencia a que existen  $K$  clases o patrones, donde es necesario, conocer a priori el número de clases existentes.

Es un algoritmo sencillo, pero muy eficiente siempre que el número de clases se conozca a priori con exactitud. Por su sencillez y robustez, ha sido muy utilizado.

Partiendo de  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p$  un conjunto de objetos a clasificar el algoritmo de K-medias realiza las siguientes operaciones.

1. Establecido previamente el numero exacto de clases existentes, digamos  $K$ , se escoge al azar entre los elementos agrupados  $K$  vectores, de forma que van a constituir los centroides (al ser los únicos elementos) de las  $K$  clases, es decir:

$$\alpha_1 = Z_1(1), \alpha_2 = Z_2(1), \dots, \alpha_K = Z_K(1) \quad (1.88)$$

donde se ha introducido entre paréntesis el índice iterativo de este algoritmo.

2. Como se trata de un proceso recursivo con un contador  $n$ , en la iteración genérica  $n$  se disminuyen todas las muestras  $\{(\mathbf{X})\}_{1 \leq j \leq p}$ , entre las K-clases de acuerdo con la siguiente regla:

$$\mathbf{X} \in \alpha_j(n) \Leftrightarrow \|(\mathbf{X}) - Z_j(n)\| < \|(\mathbf{X}) - Z_i(n)\| \quad (1.89)$$

Para todo  $i = 1, 2, \dots, K, i \neq j$

en donde se ha indexado las clases (que son dinámicas) y sus correspondientes centroides.

3. una vez redistribuidas los elementos a agrupar entre las diferentes clases, es preciso recalcular o actualizar los centroides de las clases. El objetivo en el cálculo de los nuevos centroides es minimizar el índice de rendimiento siguiente:

$$j_i = \sum_{X \in \alpha_i(n)} \|X - Z_i(n)\| \quad (1.90)$$

Con  $i = 1, 2, \dots, K$

Este índice se minimiza utilizando la media maestra o aritmética de  $\alpha_i(n)$ :

$$Z_i(n+1) = \frac{1}{N_i(n)} \sum_{X \in \alpha_i(n)} X \quad (1.91)$$

Con  $i = 1, 2, \dots, K$

siendo  $N_i(n)$  el número de elementos de la clase  $\alpha_i$  en la iteración  $n$

4. Se comprueba si el algoritmo ha alcanzado una posición estable es decir; si se cumple:

$$Z_i(n+1) = N_i(n) \quad (1.92)$$

Para todo  $i = 1, 2, \dots, K$

Si se cumple, el algoritmo finaliza. En caso contrario, salta al paso 2.

La figura 1.19 representa el diagrama de flujos u organigrama del algoritmo de  $K$  medias



Figura 1.19: Representación simbólica de un algoritmo de agrupación

El algoritmo de las  $K$  medias, es simple y extraordinariamente eficiente dado que se conoce el número de clases a priori con exactitud. Es decir muy sensible al parámetro  $K$ . Un valor de  $K$  superior al número real de clases dará lugar a clases ficticias, mientras que

un  $K$  inferior producirá menos clases de las reales.

Una forma de detectar una posible mala elección del parámetro  $K$  es el analizar la dispersiones estadísticas de las clases formadas. Cuando estas dispersiones sean muy sensiblemente diferentes entre si, siendo algunas de ellas muy elevadas respecto a las demás, se puede sospechar que se ha manejado un valor de  $K$  bajo. Análogamente, si algunas de las distancias intercalases (separación entre centroides de dos clases) son muy pequeñas respecto a las demás distancias interclases, entonces es valido plantearse que el parámetro  $K$  introducido es alto.

## Capítulo 2

# Sistema de reconocimiento de huellas digitales

### 2.1. Nociones generales

Los sistemas biométricos se basan en características o rasgos físicos medibles ó personales de comportamiento, los cuáles son usados para reconocer o verificar la identidad de una persona a través de medios automáticos. Estos sistemas biométricos han sido un área importante de investigación en los años recientes. Un sistema biométrico en particular es aquel que utiliza la huella digital basándose en la detección de bifurcaciones y terminaciones para la verificación de personas, aun entre gemelos, representando un patrón único de identificación. Este patrón conserva la misma forma desde la formación del feto hasta la muerte de la persona con lo cual se satisface los requerimientos antes mencionados.

#### 2.1.1. Biometría o técnicas de identificación biométrica

Surge por la necesidad de autenticar de forma segura la identidad de las personas que pretenden acceder a un determinado servicio o recinto físico. Las técnicas de la biométrica son distintivas, no pueden olvidarse o perderse, ya que su objetivo es resolver este problema a partir de las características propias de cada individuo, como la voz, la huella digital, el rostro, etc.

La biométrica es inherentemente más fiable frente a otras formas de autenticación personal como el uso de tarjetas o PINes, o número de identificación personal, como el usado en cajeros automáticos, tienen la ventaja de que los patrones no pueden perderse o ser sustraídos, ni pueden ser usados por otros individuos en el caso de que lleguen a tener accesibilidad a la tarjeta personal y/o, PIN.

Una de las técnicas de la biométrica más usada, son los sistemas basados en el reconocimiento de huellas digitales, por ser relativamente económicos, en comparación con otros biométricos, como por ejemplo los basados en patrones de retinas, y es una de las técnicas más maduras y probadas.

La Biométrica se ha usado ampliamente en aplicaciones forenses tales como identificación delictiva y seguridad de prisiones. La tecnología de la biométrica está evolucionando rápidamente y tiene un mismo potencial fuerte al ser adoptado ampliamente en aplicaciones de la sociedad tales como banca electrónica, comercio electrónico y accesos de control. La biometría está volviéndose uno de las aplicaciones más utilizadas por su efectividad.

### **2.1.2. Autenticación de usuarios**

En la actualidad uno de los requisitos primordiales de los sistemas informáticos son los mecanismos de seguridad que han de incluir al menos un sistema que permita identificar a las entidades (elementos activos del sistema, generalmente usuarios) que intentan acceder a los objetos (elementos pasivos, como ficheros o capacidad de cómputo), mediante procesos tan simples como una contraseña o tan complejos como un dispositivo analizador de patrones de la retina.

El objetivo de los sistemas de identificación de usuarios no suele ser identificar a una persona, sino autenticar que esa persona es quien dice ser realmente. Aunque seguramente ambos términos nos parecerán equivalentes, para una computadora existe una gran diferencia entre ellos: Básicamente los sistemas biométricos basados en huellas digitales son de dos tipos:

1. Sistema automático de autenticación de la huella digital (SAAHD).
2. Sistema automático de identificación de la huella digital (SAIHD).

Estas características deben satisfacer los siguientes requerimientos:

1. En un SAAHD la entrada es la identidad de la persona y la imagen de la huella digital de esa persona; y la salida es una respuesta de SI ó NO, indicando si la imagen de entrada pertenece a la persona cuya identidad es proporcionada.
2. En un SAIHD la entrada es solo la imagen de la huella digital y la salida es una lista de identidades de personas que pueden tener la huella dada, además de una puntuación de cada identidad indicando el grado de similaridad entre ésta y la huella dada.

Imaginemos un sistema de identificación biométrico basado en el reconocimiento de la retina; una persona miraría a través del dispositivo lector, el sistema sería capaz de decidir si es un usuario válido y en ese caso determinar de quién se trata; esto es identificación SAIHD.

Sin embargo, lo que habitualmente hace el usuario es introducir su identidad (un número, un nombre de usuario, etc.) además de mostrar sus retinas ante el lector. En este caso, el sistema no tiene que identificar a esa persona, sino autenticarlo SAAHD: comprobar los parámetros de la retina que está leyendo con los registrados en una base de datos como identificativos del usuario. En este caso, se está reduciendo el problema de una población potencialmente muy elevada a un grupo de usuarios más reducido, el grupo de usuarios del sistema que necesita autenticación.

Los métodos de autenticación se suelen dividir en tres grandes categorías, en función de lo que utilizan para la verificación de identidad:

1. Algo que el usuario sabe.
2. Algo que éste posee.

3. Una característica física del usuario o un acto involuntario del mismo. Esta última categoría se conoce con el nombre de autenticación biométrica.

Es fácil identificar ejemplos de cada uno de estos tipos de autenticación: un password es algo que el usuario conoce y el resto de personas no, una tarjeta de identidad es algo que el usuario lleva consigo, la huella digital es una característica física del usuario.

Un sistema de autenticación para incrementar su fiabilidad, debe combinar mecanismos de diferente tipo, como en el caso de una tarjeta de crédito unida al PIN de los cajeros automáticos, teniendo así una probabilidad mas elevada de fiabilidad. Además de esta característica se tiene otra, no técnica sino humana, pero quizás la más importante: un sistema de autenticación debe ser aceptado por los usuarios, que serán al fin y al cabo quienes lo utilicen. Por ejemplo, imaginemos un potencial sistema de identificación para acceder a cierta empresa, consiste en un dispositivo que fuera capaz de realizar un análisis de sangre y así comprobar que es quien dice ser; seguramente sería barato y altamente fiable, pero nadie aceptaría dar un poco de sangre cada vez que entre a la empresa, ya que resulta ser molesto e incomodo.

### **2.1.3. Autenticación biométrica**

Estos sistemas biométricos, basados en características físicas no utilizan la criptología para la identificación de usuarios, o al menos su aplicación es secundaria. En general son más amigables para el usuario, no va a necesitar recordar passwords o números de identificación complejos y, como se suele decir, el usuario se puede olvidar de la tarjeta de identificación, pero nunca se olvidará de su mano o su ojo, y son mucho más difíciles de falsificar que una simple contraseña o una tarjeta magnética. Las principales razones por la que no se han impuesto ya en nuestros días es su elevado precio, fuera del alcance de muchas organizaciones y su dificultad de mantenimiento. Las ramas de la informática que desempeñan el papel más importante en los sistemas de identificación biométricos son: reconocimiento de formas, la inteligencia artificial y el aprendizaje.



La autenticación basada en características físicas existe desde que existe el hombre y, sin darnos cuenta, es la que más utiliza cualquiera de nosotros en su vida cotidiana: a diario identificamos a personas por los rasgos de su cara o por su voz. Obviamente aquí el agente reconocedor es una persona, lo cual es fácil, pero en el modelo aplicable a redes de comunicaciones el agente ha de ser un dispositivo que, basándose en características del sujeto a identificar, gestione el acceso a un determinado recurso.

Los dispositivos biométricos tienen tres partes principales:

1. Un mecanismo automático que lee y captura una imagen digital o analógica de las características a analizar.
2. Una entidad para manejar aspectos como la comprensión, almacenamiento o comparación de los datos capturados con los registrados en una base de datos (que son considerados válidos).
3. Una interfaz de aplicaciones.

El proceso general de autenticación sigue unos pasos comunes a todos los modelos de autenticación biométrica: captura o lectura de los datos que el usuario a validar presenta, extracción de ciertas características de la muestra (por ejemplo, las minucias de una huella digital, esto lo veremos mas adelante), comparación de tales características con las registradas en una base de datos, y decisión de si el usuario es válido o no. En la figura 2.1 se representa gráficamente el proceso de un Sistema Reconocimiento Biométrico (SRB).

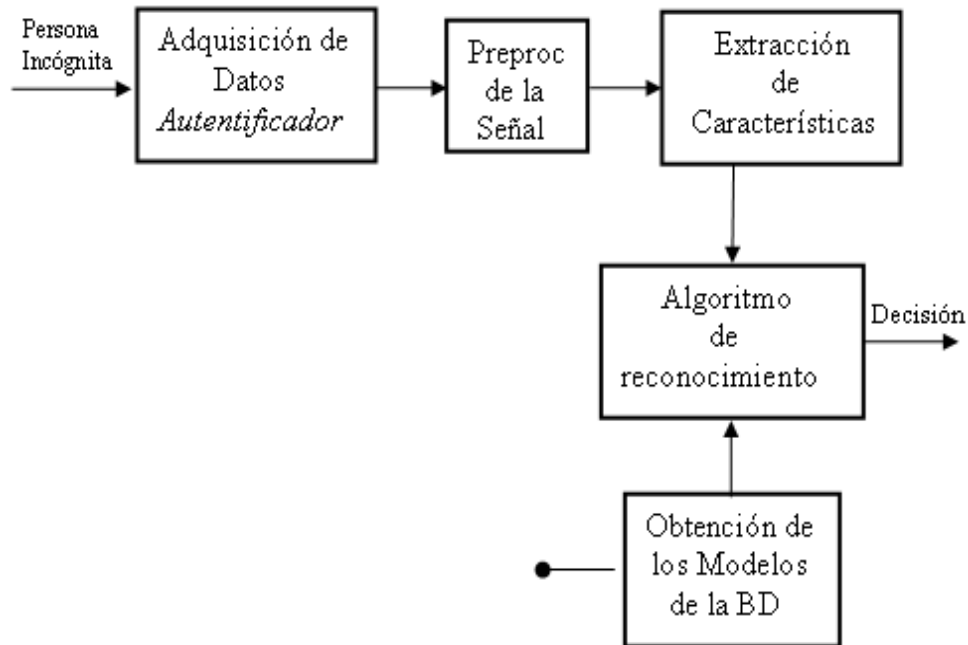


Figura 2.1: Diagrama de bloques de un Sistema de Reconocimiento Biométrico.

Es en la decisión donde principalmente entran en juego las dos características básicas de la fiabilidad de todo sistema biométrico: las tasas de falso rechazo y de falsa aceptación. Por tasa de falso rechazo (FRR) se entiende la probabilidad de que el sistema de autenticación rechace a un usuario legítimo porque no es capaz de identificarlo correctamente, y por tasa de falsa aceptación (FAR) la probabilidad de que el sistema autentique correctamente a un usuario ilegítimo; evidentemente, una FRR alta provoca descontento entre los usuarios del sistema, y una FAR elevada genera un grave problema de seguridad, proporcionando acceso a un recurso a personal no autorizado.

Para determinar las prestaciones de un sistema biométrico se suele utilizar la tasa de éxito (SR) que responde a una combinación de los dos factores anteriores:

$$SR = 1 - (FAR + FRR) \quad (2.1)$$

El FAR y el FRR responden a parámetros inversamente proporcionales, por tanto, variarán en función de las condiciones prefijadas por el programa de identificación biométrica. Así si por ejemplo se tiene que utilizar el programa en un entorno de máxima seguridad, se intentará que el FAR sea lo más pequeño posible, aunque esta acción signifique de forma explícita, el incremento drástico del factor FRR.

Por lo tanto se debe fijar un parámetro o umbral que permita igualar los dos factores, asegurando de esta manera el óptimo funcionamiento del sistema. Este umbral se denomina tasa de error igual (ERR) y es el que determinará, finalmente, la capacidad de identificación del sistema. En la figura 2.2 se muestra la relación.

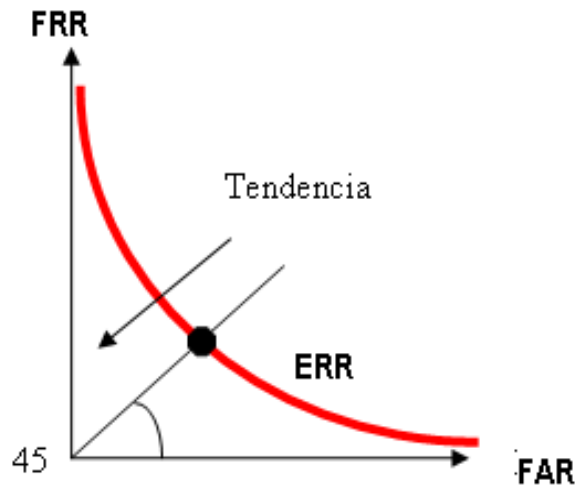


Figura 2.2: Relación entre FAR, FRR Y ERR.

#### 2.1.4. Huellas digitales

Las huellas digitales se forman totalmente en aproximadamente siete meses de desarrollo del feto y las configuraciones de la cresta del dedo no cambian a lo largo de la vida de un individuo excepto debido a los accidentes como machucaduras y cortes en las puntas de los dedos. Esta propiedad hace de las huellas digitales un identificador biométrico muy atractivo; En este caso, los genes determinan las características generales del modelo. La

formación de la huella digital es similar al crecimiento del cabello y los vasos sanguíneos.

Los detalles más finos de las huellas digitales son determinados por los cambios del microentorno. Una pequeña diferencia en el microentorno se amplifica por el proceso de diferenciación de las células. Hay muchas variaciones durante la formación de las huellas digitales que harán virtualmente imposible que dos huellas digitales sean iguales. Pero ya que la formación de las huellas digitales están determinadas por pequeñas variaciones genéticas que producen formaciones muy distintas en las huellas. Estos no serán totalmente patrones aleatorios. Se puede decir que el proceso de formación de la huella digital es un sistema caótico en lugar de uno aleatorio.

Típicamente la huella digital de un individuo ha sido un patrón bastante bueno para determinar su identidad de forma inequívoca, ya que está aceptado que dos dedos nunca poseen huellas similares, ni siquiera entre gemelos o entre dedos de la misma persona. Desde el siglo pasado hasta nuestros días se vienen realizando con éxito clasificaciones sistemáticas de huellas digitales en entornos policiales y el uso de estos patrones fue uno de los primeros en establecerse como modelo de autenticación biométrica.

Cuando un usuario desea autenticarse ante el sistema sitúa su dedo en un área determinada, el área de lectura. Aquí se toma una imagen que posteriormente se normaliza mediante un sistema de finos espejos para corregir ángulos, y es de esta imagen normalizada de la que el sistema extrae las minucias (ciertos arcos, bucles y remolinos de la huella) que va a comparar contra las que se tiene en la base de datos. Es importante resaltar que el sistema no analiza la huella en sí sino las minucias, concretamente la posición relativa de cada una de ellas. En la figura 2.3 se muestra ejemplos de minucias de huellas digitales.



Figura 2.3: Huella digital con minucias donde  $\bigcirc$  encierra la bifurcación de la cresta y  $\square$  el final de la cresta.

Las huellas de los dedos presentan como característica principal, la presencia de un conjunto de crestas (o espinazos) que son las partes donde la piel se eleva sobre las más bajas o surcos existentes entre las crestas. Con respecto a estas crestas se definen dos características particulares que obedecen al término de minucias:

- Final de la cresta. Característica definida como el punto donde la cresta acaba de forma abrupta.
- Bifurcación de la cresta. Característica definida como el punto en el que la cresta se bifurca en dos o más.

Los dos tipos anteriores de minucias son los más importantes, ya que los demás tipos se forman por la combinación de estos dos. Por esta razón, en la etapa de extracción de características se detectan estos dos tipos. En la figura 2.4 se observan diferentes tipos de minucias.








Características	
	Terminación
	Bifurcación
	Laguna
	Borde independiente
	Punto o isla
	Aguijón
	Cruce

Figura 2.4: Tipos de Minucias.

Estas dos características quedan unívocamente definidas a partir de su localización (coordenadas  $x$ ,  $y$  respecto al sistema de coordenadas central de la imagen) y de su orientación (ángulo  $\theta$ ).

Está demostrado que dos dedos nunca pueden poseer más de ocho minucias comunes, y cada uno tiene al menos entre 30 y 40 de éstas. En la figura 2.3 se muestra una imagen de una huella digitalizada con sus minucias. Si la comparación de las posiciones relativas de las minucias leídas con las almacenadas en la base de datos es correcta, se permite el acceso al usuario, denegándose obviamente en caso contrario.

Los sistemas de reconocimiento de huellas digitales, tienen en su contra la incapacidad temporal de autenticar usuarios que se hayan podido herir en el dedo a reconocer. Un pequeño corte o una quemadura que afecte a varias minucias pueden hacer inútil al sistema; también elementos como la suciedad del dedo, la presión ejercida sobre el lector o el estado de la piel pueden ocasionar lecturas erróneas.

Además de las minucias, las huellas digitales contienen dos tipos especiales de rasgos llamados puntos core y delta (Figura 2.5) Estos puntos son referidos como los puntos de singularidad de una huella digital. El punto core es definido como el punto mas alto en la cresta mas interior y el punto delta es una región en la huella donde las crestas forman

una configuración triangular. Este punto es generalmente usado como punto de referencia para la codificación de minucias.

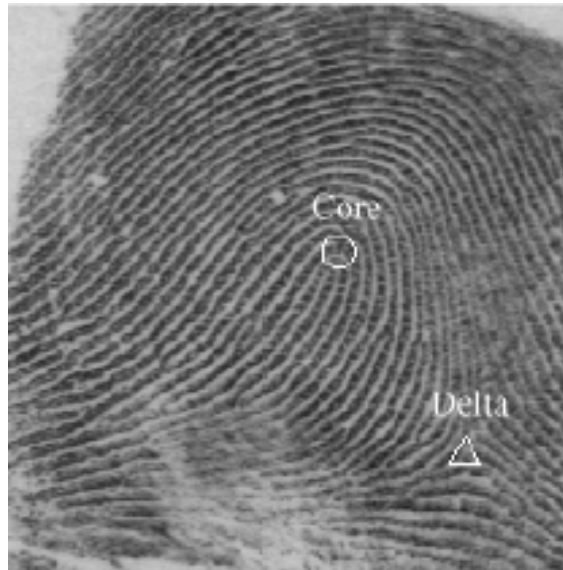


Figura 2.5: Puntos singulares en una huella digital.

### 2.1.5. Individualidad de la huella digital

La identificación de la huella digital se basa en dos premisas básicas:

- Persistencia: las características básicas de las huellas no cambian con el tiempo.
- Individualidad: la huella digital de un individuo es única.

La validez de la primera premisa ha sido establecida por la anatomía y la morfogénesis de las crestas del dedo. Mientras la segunda premisa generalmente ha sido aceptada como verdadera bajo resultados empíricos, la base científica fundamental de la individualidad de la huella digital no se ha aprobado formalmente (Ver capítulo 2.2).

Una base científica para establecer la individualidad de la huella digital no solo determinará la admisibilidad de la identificación de la huella digital sino que también establecerá una cota superior en el desempeño de un sistema automático de comprobación de

la huella digital. La naturaleza de distensión de las característica físicas de una persona se debe a la diversidad genética individual inherente dentro de la población así como a factores aleatorios que afectan el desarrollo del embrión puesto que dos individuos pueden ser arbitrariamente parecidos con respecto a su constitución genética (ejemplo: gemelos idénticos), una evaluación pesimista de la discriminación de la identidad basada en la biometría puede que sólo necesite contar con una valoración de diversidad en los rasgos debido a los procesos aleatorios que afectan al desarrollo humano. Tal estrategia de valoración necesariamente es contradictoria con las muestras biométricas de individuos que tienen su constitución genética muy parecida. Ya que los gemelos idénticos tienen relaciones genéticas similares, se espera encontrar la máxima similitud entre las huellas digitales de ellos.

### **Factores genéticos**

La magnitud de variación en un rasgo físico debido a procesos de desarrollo aleatorio difiere de rasgo a rasgo. Por definición, no se pueden distinguir gemelos idénticos bajo el ADN. Típicamente, la mayoría de las características físicas tales como el tipo de cuerpo, la voz, y la cara son muy similares para gemelos idénticos donde la identificación automática basada en la cara y la geometría de la mano es insuficiente para distinguirlos.

¿Cómo se determina si dos huellas digitales son idénticas? A fin de establecer fiablemente si dos impresiones vinieron del mismo dedo o de dedos diferentes, es necesario capturar alguna representación invariante (rasgos) de las huellas digitales: los rasgos que permanecen inalterados independiente de los cortes y magulladuras a través de los años, la orientación de la impresión con respecto al medio de captura, oclusión de una parte pequeña del dedo, la tecnología de la imagen adquirida de la huella digital del dedo, o la distorsión elástica del dedo durante la adquisición de la impresión.

Una pregunta importante en el emparejamiento de la huella digital es: ¿qué características de las huellas digitales se heredan? Varios estudios han mostrado una correlación significativa en la clase de la huella digital (es decir, whorl, right loop, left loop, arch, tented



arch) en dedos de gemelos idénticos; la correlación basada en otros atributos genéricos de la huella digital como la cresta, su anchura, la separación entre crestas, y la profundidad de la cresta que también ha sido significativa en gemelos idénticos. En estudios dermatológicos, la mejor diferencia global entre las huellas digitales se ha encontrado entre los individuos de razas diferentes. Las personas no emparentadas de la misma raza tienen muy poca similitud global en sus huellas digitales, padres e hijos tienen alguna similitud global porque comparten genes, los hermanos tienen más similitud y la similitud global máxima se observa en los gemelos monocigóticos (idénticos) que es la relación genética más íntima [9]. En la figura 2.6 se muestra las huellas digitales de hermanas gemelas idénticas.



Figura 2.6: Imágenes de la huella digital de hermanas gemelas idénticas capturadas usando un escáner óptico. (a) y (b) son dos impresiones del mismo dedo de una gemela y (c) y (d) son dos impresiones del dedo correspondiente de su hermana. Las huellas digitales tienen el mismo tipo (right loop) y parecen similares a los ojos de inexpertos.

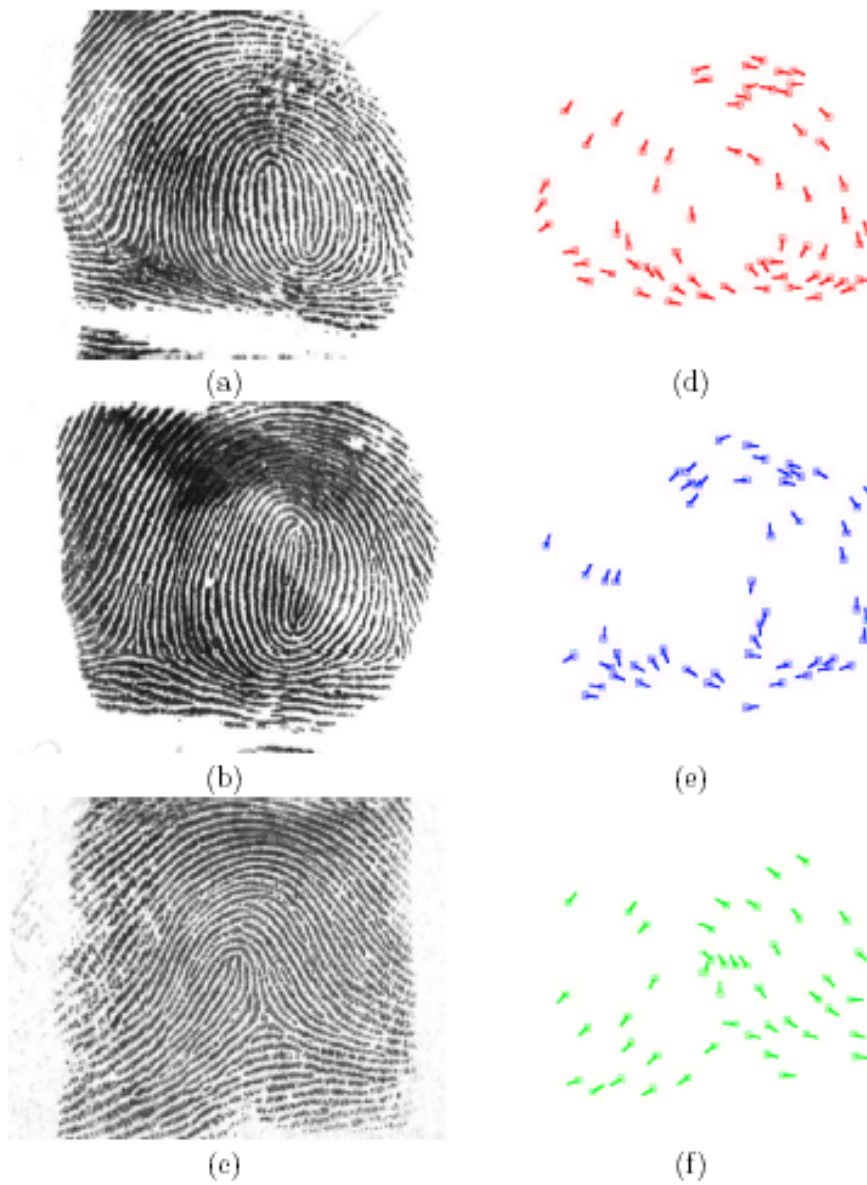


Figura 2.7: Extracción de minucias de las gemelas. (a) y (b) son las imágenes de la huella digital de una gemela idéntica y su hermana mientras la huella digital en (c) es de otra persona. (d), (e), y (f) es la minucia extraída de (a), (b), y (c), usando respectivamente el algoritmo del extracción en [2].

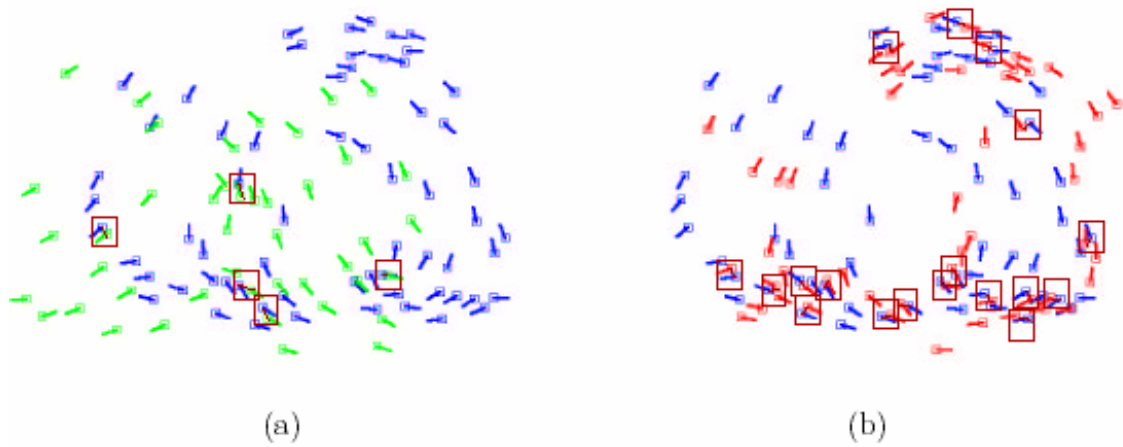


Figura 2.8: Emparejamiento de minucias para (a) gemelo-no gemelo (el emparejamiento de las Figuras 2.7(e) y 2.7(f), la cuenta de emparejamiento=3) y (b) gemelo-gemelo (el emparejamiento de las Figuras 2.7(d) y 2.7(e), la cuenta de emparejamiento = 38). Los pares de minucias "emparejadas" se muestran en las cajas de limitación.

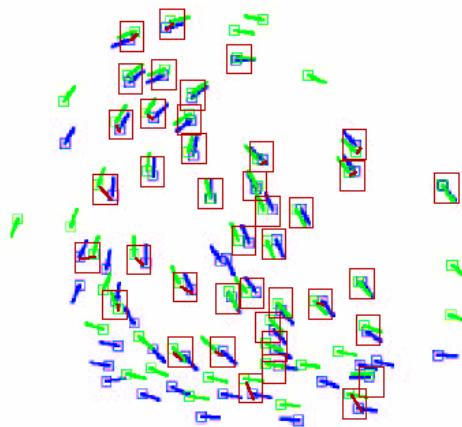


Figura 2.9: Emparejamiento de minucias para dos impresiones del mismo dedo mostradas en las Figuras 2.6(a) y 2.6(b) (la cuenta de emparejamiento = 487). Los pares de minucias "emparejadas" se muestran en las cajas de limitación.

En las figuras 2.7, 2.8 y 2.9 se muestra la similitud entre las huellas digitales que se le hizo a unas gemelas idénticas, comparándola con la similitud entre dos huellas digitales arbitrarias. Esto se realizó con un sistema biométrico automático de identificación de la huella digital en [2].

Aquí se observa que se ha confirmado la afirmación de que las huellas digitales de gemelos idénticos tienen una gran correlación de clases, es decir, si una huella digital de un gemelo idéntico es un whorl entonces es muy probable que la huella digital del otro gemelo también sea de tipo whorl. También se observa la correlación entre la clase de la huella digital y el emparejamiento de la minucia entre dos huellas digitales escogidas al azar.

Otros factores que juegan un papel importante son los factores medioambientales porque de ellos depende la calidad de la huella digital. Ver sección 2.2.

### **2.1.6. Sensores de la huella digital**

Las imágenes de la huella digital pueden ser adquiridas por un proceso offline u online. Las imágenes de la huella digital adquiridas por el proceso offline son conocidas como el "inked" de las huellas digitales mientras las imágenes adquiridas por el proceso online es conocido como "live-scan" de las huellas digitales. Las huellas digitales inked son adquiridas por tres métodos:

- Enrollado
- De toque ligero
- Latente

En el método enrollado para la adquisición de la huella digital, se aplica tinta al dedo y entonces se rueda en un papel de un lado de la uña al otro para formar una impresión. Este papel se examina a 500 dpi<sup>2</sup> de resolución por un escáner de escala de grises normal. Donde el área del surco es más grande que el de la cresta debido al proceso rodante pero

---

<sup>2</sup>Puntos por pulgada

tiene deformaciones más grandes debido a la naturaleza inherente del proceso rodante.

En el método de toque ligero para la adquisición de la huella digital, se aplica tinta al dedo y entonces se aprieta hacia un papel sin rodar. El papel es escaneado en una imagen digital. Típicamente, las huellas digitales entintadas por este método tienen menos deformación no lineal, pero el área es más pequeña que la de la huella digital entintada por enrollado.

Se forman huellas digitales latentes, cuando los dedos dejan una capa delgada de sudor y engrasan las superficies que ellos tocan, debido a la presencia de poros de sudor en nuestras yemas de los dedos. Los científicos forenses tiñen con color esta impresión que se encuentra típicamente en la escena de un crimen y entonces examinan la huella digital.

La huella digital de live-scan se obtiene directamente del dedo sin el uso intermedio del papel (a una resolución de 500 dpi). Típicamente, los sensores live-scan capturan una serie de huellas digitales de toque ligero, cuando un dedo se aprieta en la superficie del sensor. Para huellas digitales de enrollado en live-scan, el/la usuario rueda tocando de un extremo de la uña al otro en la superficie del sensor y este captura varias imágenes de la huella digital de toque ligero. La imagen de la huella digital rodada es construida por mosaicos de las múltiples imágenes de toque ligero que se capturaron durante el proceso rodante. Los sensores comercialmente disponibles live-scan son basados en varias tecnologías diferentes.



(a)



(b)

Figura 2.10: Sensores de Huellas digitales. (a) sensor Óptico de Biometricas Digitales, Inc., y (b) sensor transistorizado de Veridicom, Inc.

### 2.1.7. Clasificación de huellas digitales

Grandes cantidades de huellas digitales se han almacenado y guardado todos los días en una amplia gama de aplicaciones, incluso en forense, accesos de control, y registro de licencia de conducción. El reconocimiento automático de la identidad basado en huellas requiere que la entrada de estas sea emparejada con un número grande almacenado en un banco de datos. La clasificación de la huella digital es una técnica que la asigna a una huella uno de los tipos pre-especificados ya establecidos en la literatura (y usado en aplicaciones forenses) la cual proporciona un mecanismo indexado. Una huella digital de entrada se empareja primero a uno de los tipos pre-especificados y entonces se compara con un subconjunto del banco de datos correspondiente a ese tipo. Para aumentar la eficacia de búsqueda, el algoritmo de clasificación puede clasificar una huella digital en más de una clase. Por ejemplo, si el banco de datos de huellas digitales se divide en cinco clases whorl (espiral) (*W*), right loop (lazo derecho) (*R*), left loop (lazo izquierdo) (*L*), arch (arco)

(*A*), and tented arch (arco tendido) (*T*) ver figura 2.11, y un clasificador produce dos clases con alta exactitud, entonces el sistema de identificación sólo necesitará investigar dos de las cinco clases y así disminuirá la búsqueda. La clasificación continua de huellas digitales también es muy atractiva para el posicionamiento donde no se dividen en clases no-solapadas, pero cada huella se caracteriza con un vector numérico que resume sus rasgos principales.

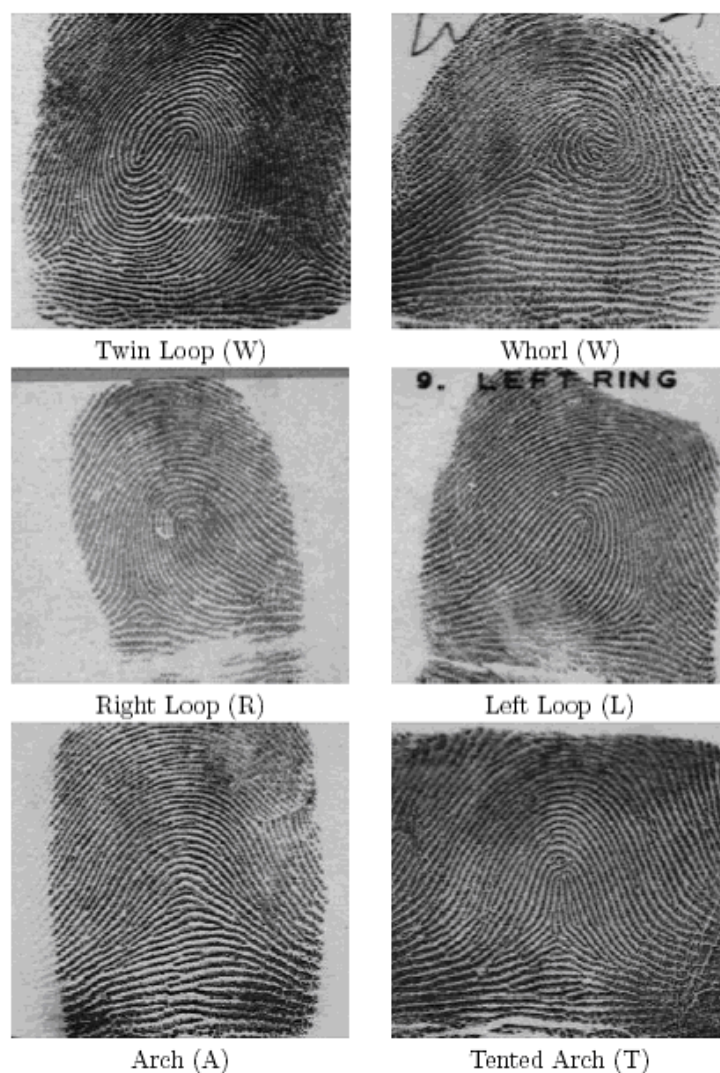


Figura 2.11: Las Seis mejores clases de la huella digital. Se etiquetan imágenes Twin loop como Whorl (espiral) en el banco de datos NIST (Instituto Nacional de Normas y Tecnología)

### 2.1.8. Autenticación de la huella digital

Un sistema biométrico puede operarse de dos modos: Modo de autenticación y modo de identificación. En el modo de autenticación, un sistema biométrico acepta o rechaza la identidad exigida de un usuario mientras el modo de identificación opera estableciendo la identidad del usuario sin una identidad exigida. La identificación con huella digital es un



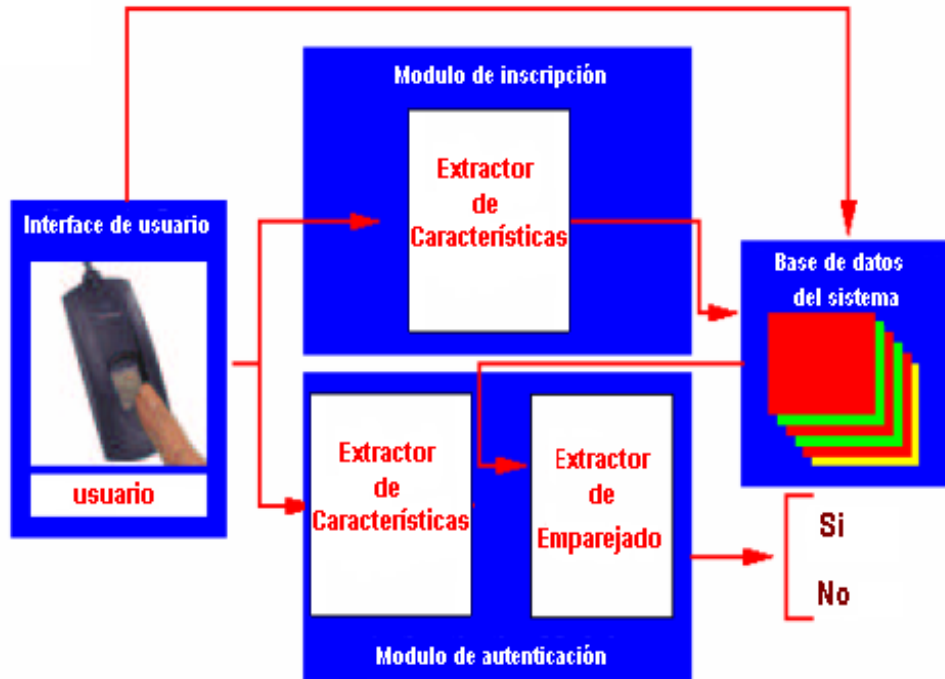


Figura 2.12: Diagrama para un sistema de autenticación automático.

problema más difícil que la autenticación de la misma porque se necesita un gran número de comparaciones al hacer una identificación. Se ha enfocado un sistema biométrico que opera en un modo de autenticación y un esquema de posicionamiento (clasificación de la huella digital) que puede ser usado en un sistema de identificación. Un buen número de aplicaciones opera en modo de autenticación de manera regular y sólo realiza identificación en el momento de registro del usuario al verificar la integridad en el banco de datos (ej., encontrando duplicados). Un sistema de autenticación típico puede ser dividido en dos módulos: inscripción y autenticación. El módulo de inscripción examina la huella digital de una persona a través de un dispositivo perceptor y entonces almacena una representación (llamada plantilla) de la huella en el banco de datos. El módulo de autenticación se invoca o se llama durante la fase de funcionamiento. La misma representación que fue usada en la fase de inscripción se extrae de la huella digital de entrada y se empareja con la plantilla de identificación exigida para dar un si/no como respuesta. Por otro lado, un sistema de

identificación empareja la huella digital de entrada con un gran número de ellas en el banco de datos y como resultado, la clasificación es eficaz en un sistema de identificación y no es un problema en el sistema de autenticación.

El problema de autenticación puede formularse como sigue. Sea  $S$  la representación de la señal biométrica guardada (plantilla) de una persona y sea  $I$  la señal adquirida (entrada) para la autenticación. Entonces las hipótesis nulas y alternas pueden declararse como:

- $H_0$ :  $I \neq S$ , la huella digital de entrada no es igual a la plantilla.
- $H_1$ :  $I = S$ , la huella digital de entrada es igual que la plantilla.

Las decisiones asociadas a estas hipótesis son:

- $D_0$ : La persona es un impostor.
- $D_1$ : La persona es auténtica.

La autenticación involucra el emparejamiento de  $S$  e  $I$  usando una medida de similitud. Si el valor de la cantidad de similitud del emparejamiento es menor que algún umbral de decisión  $T$ , entonces la decisión es  $D_0$ , sino, se decide  $D_1$ . La formulación de una prueba de hipótesis inherentemente contiene dos tipos de errores: Tipo I: aceptación falsa (se decide  $D_1$  cuando  $H_0$  es verdadera) y Tipo II: falso rechazo (se decide  $D_0$  cuando  $H_1$  es verdadera). Los dos tipos de errores también son conocidos como: FAR y FRR, definidos como:

- Porcentaje de falsa aceptación =  $P(D_1|W_0)$ .
- Porcentaje de falso rechazo =  $P(D_0|W_1)$ .

Donde  $W_0$  es la clase asociada con  $H_0 =$  verdadera y  $W_1$  es la clase asociada con  $H_1 =$  verdadera. El desempeño de un sistema biométrico normalmente se especifica en términos de su FAR. El esquema de decisión debe establecer un límite de decisión que minimiza el FRR para el FAR especificado. Hay un compromiso entre los dos tipos de

errores y ambos errores no pueden reducirse simultáneamente basados solo en el punto de operación. La aplicación biométrica dada impone el FAR y los requisitos del FRR para el sistema de autenticación. Por ejemplo, generalmente el acceso a una máquina de ATM necesita un FRR pequeño, pero el acceso a una instalación militar segura requiere un FAR muy pequeño.

## 2.2. Modelos probabilísticos de identificación de la huella digital

El propósito en el estudio de la individualidad de la huella digital tiene un doble interés. Primero, una base científica (una estimación estadística fiable del error de emparejamiento) para que la comparación de la huella digital pueda determinar la admisibilidad de la identificación de la misma en las cortes de justicia como una forma de evidencia. Segundo, se puede establecer una cota superior en el desarrollo de un sistema automático de verificación de las huellas digitales. En [24], se desarrolla un modelo de la individualidad de la huella digital que intenta estimar la probabilidad de una falsa asociación. Se usa este modelo para establecer una cota superior en el desarrollo de un sistema de comprobación de la huella digital [2].

Para resolver el problema de individualidad, se necesita primero definir a priori la representación de una huella digital (modelo) y la métrica (medida) para la similitud. Las Huellas digitales se pueden representar por un número grande de rasgos, incluso el modelo de flujo de la cresta global, la frecuencia de la cresta, la localización y posición de puntos singulares (centro(s) y delta(s)), tipo, dirección, y localización de puntos de minucia, cantidad de cresta entre los pares de minucias, y la localización de poros (vea la Figura 2.8). Todas estas características contribuyen al establecimiento de la individualidad de la huella digital. Preferiblemente se escoge la representación de la minucia de las huellas digitales porque es utilizada por expertos forenses, se ha demostrado que es relativamente estable y ha sido adoptada por la mayoría de sistemas automáticos de emparejamiento de

huellas.

Dado un esquema de representación y una métrica (medida) de similitud, hay dos aproximaciones para determinar la individualidad de las huellas digitales. En la empírica, se reúnen las muestras representativas de las huellas digitales usando un emparejador estándar, la exactitud del emparejador en las muestras proporciona una indicación de la singularidad de la huella con respecto al emparejador. Hay problemas conocidos (y costos) asociados con la recolección de las muestras representativas. En un acercamiento teórico para la estimación de la individualidad, se modelan todos los fenómenos que afectan las relaciones entre clases y las relaciones dentro de las clases. Dada la métrica de similitud, se puede estimar teóricamente la probabilidad de una falsa asociación. Los acercamientos teóricos son limitados por la medida en la cual los modelos supuestos representan la realidad.

Los patrones de minucia son generados por las huellas digitales que están fluyendo suavemente sobre texturas orientadas. Los puntos de minucia no se distribuyen aleatoriamente ya que las posiciones son determinadas por las crestas (vea la Figura 2.13). Más aún, las orientaciones de minucias cercanas están fuertemente correlacionadas. Así, el espacio de configuración examinado por los patrones de minucia es más pequeño que el examinado por un patrón de puntos aleatorios. Esto implica que la probabilidad de hallazgo de impresiones suficientemente similares de dos dedos diferentes es más alta que la de hallazgo de conjuntos suficientemente similares de patrones de puntos aleatorios.



Figura 2.13: Una imagen de la huella digital de tipo right loop. Se muestran la estructura de la cresta global, los puntos singulares, y los poros de sudor.

El número total de grados de libertad del espacio de patrones (ej., espacio de configuración de minucia) no está directamente relacionado con la discriminabilidad de los patrones diferentes (ej., minucia de los diferentes dedos). La estimación eficaz de información discernidora sólo se puede lograr tomando la cantidad de variaciones entre patrones [15]. Hay varias fuentes de variabilidad en las múltiples impresiones de un dedo [2]: contacto no-uniforme (con el sensor), contactos irreproducibles, contacto inconsistente, y distorsiones de la imagen. Esta variabilidad en las múltiples impresiones de un dedo se manifiestan así mismos en (i) detección de falsa minucia o la pérdida de minucia genuina, (ii) el desplazamiento o desorientación (también llamada deformación) de la minucia genuina detectada, y (iii) la transformación del tipo de minucia (ambigüedad conjuntiva). Esto trae consigo el diseño de una métrica de similitud (emparejador) que acomoda es-

tas variaciones intra-clase. Como resultado, la probabilidad de asociación falsa aumenta significativamente.

### 2.2.1. Antecedentes

Los estudios previos sobre la individualidad de la huella se enfocan en estudiar una representación predominantemente basada en minucia; algunos estudios se enfocan explícitamente en la información de las clases de huellas digitales (ej., right loop, left loop, whorl, arch, tented arch, etc.). El tipo, la dirección, y la localización de minucia fueron los rasgos más examinados en estudios anteriores. Los tipos de minucia usados varían de un estudio a otro: algunos estudios usan dos tipos de minucia (terminaciones y bifurcaciones) mientras que otros usan 13 tipos de eventos (ej., empty cell, ridge ending, ridge fork, island, dot, broken ridge, bridge, spure, enclosure, delta, double fork, trifurcación, múltiple events) [25].

El problema de individualidad de la huella digital fue primero estudiado por Galton en 1892 [12], quién consideró una región cuadrada expandida a seis crestas en una huella digital dada. Él asumió que, en promedio, una huella digital puede estar cubierta por 24 de tales regiones independientes. Galton estimó que él podría reconstruir correctamente cualquiera de las regiones con una probabilidad de  $\frac{1}{2}$ , mirando las crestas circundantes. De acuerdo con la configuración de una huella digital específica, dados las crestas circundantes, la probabilidad es de  $(\frac{1}{2})^{24}$ . Él multiplicó esta probabilidad condicional (en crestas circundantes) con la probabilidad de encontrar las crestas circundantes para obtener la probabilidad de ocurrencia de una huella digital como:

$$P(\text{configuración de la huella digital}) = \frac{1}{16} \times \frac{1}{256} \times \left(\frac{1}{2}\right)^{24} = 1,45 \times 10^{-11} \quad (2.2)$$

donde  $\frac{1}{16}$  es la probabilidad de ocurrencia de un tipo de huella digital específico (como arch, los tented arch, left loop, right loop, double loop, whorl, etc.) y  $\frac{1}{256}$  es la probabilidad de ocurrencia del número correcto de crestas de entrada y salida de cada una de las 24 regiones. La ecuación (2.2) da la probabilidad de configuración de la huella digital

particular en una huella de tamaño medio (conteniendo 24 regiones definidas por Galton). Roxburgh [27], Pearson [22], y Kingston [18] objetaron las suposiciones de Galton de que la probabilidad de ocurrencia de cualquier configuración de la cresta particular en un cuadrado de seis crestas es de  $\frac{1}{2}$ , y postularon que la ecuación (2.2) infravaloraba erróneamente la individualidad de la huella digital (es decir, sobrestimaba la probabilidad de ocurrencia). Pearson [98] argumentó que podría haber 36 ( $6 \times 6$ ) posibles localizaciones de minucia dentro de una de las regiones cuadradas de seis crestas de Galton, llegando a una probabilidad de ocurrencia de una configuración de la huella particular de

$$P(\text{configuración de la huella digital}) = \frac{1}{16} \times \frac{1}{256} \times \left(\frac{1}{36}\right)^{24} = 1,09 \times 10^{-41} \quad (2.3)$$

Se interrelacionaron varios modelos (Henry, Balthazard, Bose, Wentworth, Wilder, Cummins, Midlo y Gupta) subsecuentes basándose en una probabilidad fija,  $p$ , para la ocurrencia de minucias. Además, se calculó la probabilidad de configuración de N-minucias particulares para la huella digital como:

$$P(\text{configuración de la huella digital}) = p^N \quad (2.4)$$

Cada uno de los modelos escogió un valor de  $p$  diferente, en [24] se presenta una descripción mas detallada de esto. Debido a que los valores de  $p$  usados en los modelos mencionados anteriormente varían ampliamente, la probabilidad de configuración para la huella digital también varía dramáticamente de un modelo a otro.

### 2.2.2. Un modelo de individualidad de la huella digital

En [24] se desarrolló un modelo para obtener una probabilidad realista y más exacta de correspondencia entre las huellas digitales. Para estimar la probabilidad de correspondencia, se hacen las siguientes suposiciones:

1. Se consideran sólo rasgos de minucia ya que (i) la mayoría del poder discernidor del sistema automático de emparejamiento de la huella digital (SAEH) está basado en

rasgos de minucia, y (ii) para una medida objetiva de individualidad, es necesario que la representación sea de forma consistentemente reproducible, fácilmente localizable y cuantificable. En [24] se mostró que los rasgos de minucia son sistemas estables y prácticos. Sólo se consideran terminaciones y bifurcaciones de crestas porque la ocurrencia de otro tipo de minucia como islas, puntos, cercamientos, puentes, bifurcaciones dobles, trifurcaciones, etc. no son muy comunes. Adicionalmente, no se distingue entre los dos tipos de minucia porque no pueden diferenciarse con precisión las terminaciones y las bifurcaciones de crestas. Puesto que la minucia sólo puede residir en crestas que siguen ciertos modelos globales en una huella digital, las direcciones de minucia no son completamente independientes de las localizaciones de la misma.

2. Se asume una distribución uniforme de minucia con la restricción que dos de ellas no pueden ser muy cercanas a otras. Si las localizaciones de minucia no están distribuidas uniformemente, dicha suposición aproxima la distribución uniforme ligeramente modificada. No se modela explícitamente la tendencia agrupadora de la minucia. Por consiguiente, la suposición de independencia de localización de minucias podrá sesgar la estimación de la probabilidad de falsa asociación hacia valores más altos. Sin embargo, es una práctica común en estudios de individualidad de la huella digital hacer estimaciones conservativas (más altas) de la probabilidad de correspondencia.
3. La correspondencia de un par de minucias es un evento independiente y cada correspondencia es igualmente importante. Los sistemas de emparejamiento de la Huella digital ponderan de manera diferente las correspondencias basadas en su posición (ej., correspondencias que involucran minucias de patrones periféricos que tienen menor ponderación que aquellas minucias localizadas en el centro de la huella digital). Similarmente, es posible medir diversas correspondencias espaciales más que todas las correspondencias localizadas en una vecindad estrecha.
4. No se toma en cuenta la calidad de la imagen de la huella digital en la determinación



de individualidad. Es muy difícil asignar un índice de calidad fiable a una huella porque la calidad de la imagen es un concepto subjetivo. (para el emparejamiento sólo se detectará un subconjunto de la verdadera minucia en una huella digital). Todas las correspondencias son consideradas como fiables y ninguna certeza es asociada con una correspondencia basada en la calidad de imagen de la huella digital. Por consiguiente, se considera sólo evidencia positiva de una correspondencia de minucia; se ignora la información negativa de evidencia conflictiva (ej., una minucia que no empareja).

5. El ancho de las crestas es el mismo en la población y es espacialmente uniforme en el mismo dedo. Esta suposición se justifica debido a que las variaciones de presión podrían hacer no uniforme las variaciones uniformes de las crestas y viceversa. Más aun, allí puede haber sólo información limitada en la frecuencia de las crestas.
6. El análisis de emparejamiento de impresiones diferentes de dedos iguales involucran los parámetros de la probabilidad de emparejamiento de minucia en dos huellas digitales de dedos diferentes.
7. Se asume que allí existe una y sólo una alineación entre la plantilla y los conjuntos de minucia de entrada.

El problema de correspondencia de la huella digital involucra el emparejamiento de dos huellas digitales; una se llama plantilla (guardada en el sistema) y la otra se llama entrada (qué necesita ser verificada). Se asume que se ha establecido una alineación razonable entre la plantilla y la entrada. La alineación del conjunto de minucias de entrada con el conjunto de minucias de plantilla se hace de tal manera que las correspondencias de minucia puedan determinarse con una tolerancia pequeña. Un sistema automático puede buscar una alineación que maximice una función objetiva dada (como el número de minucias que concuerdan). Esta suposición puede no ser válida al emparejar una huella digital parcial (latente) con una impresión total en el banco de datos, porque allí pueden ser “razonables ” varias posibles alineaciones. Cuando las alineaciones múltiples son garanti-

zadas de hecho por una situación, la probabilidad de asociación falsa aumenta [24] (ver la ecuación (2.5)).

$$P(\text{Asociación falsa}) = 1 - (1 - P(\text{Configuración de la huella digital}))^k \quad (2.5)$$

Dada una huella digital de entrada que contiene  $n$  minucias, se debe calcular la probabilidad de que cualquier huella digital arbitraria (plantilla en un banco de datos de huellas digitales) que contenga  $m$  minucias tenga exactamente  $q$  minucias coincidentes con la de entrada. Ya que sólo se considera la minucia que es definida por su localización,  $(x, y)$ , y por el ángulo  $\theta$  de la cresta en el que ella reside, el conjunto de minucias de la entrada y la plantilla,  $E$  y  $P$ , respectivamente, se pueden definir como:

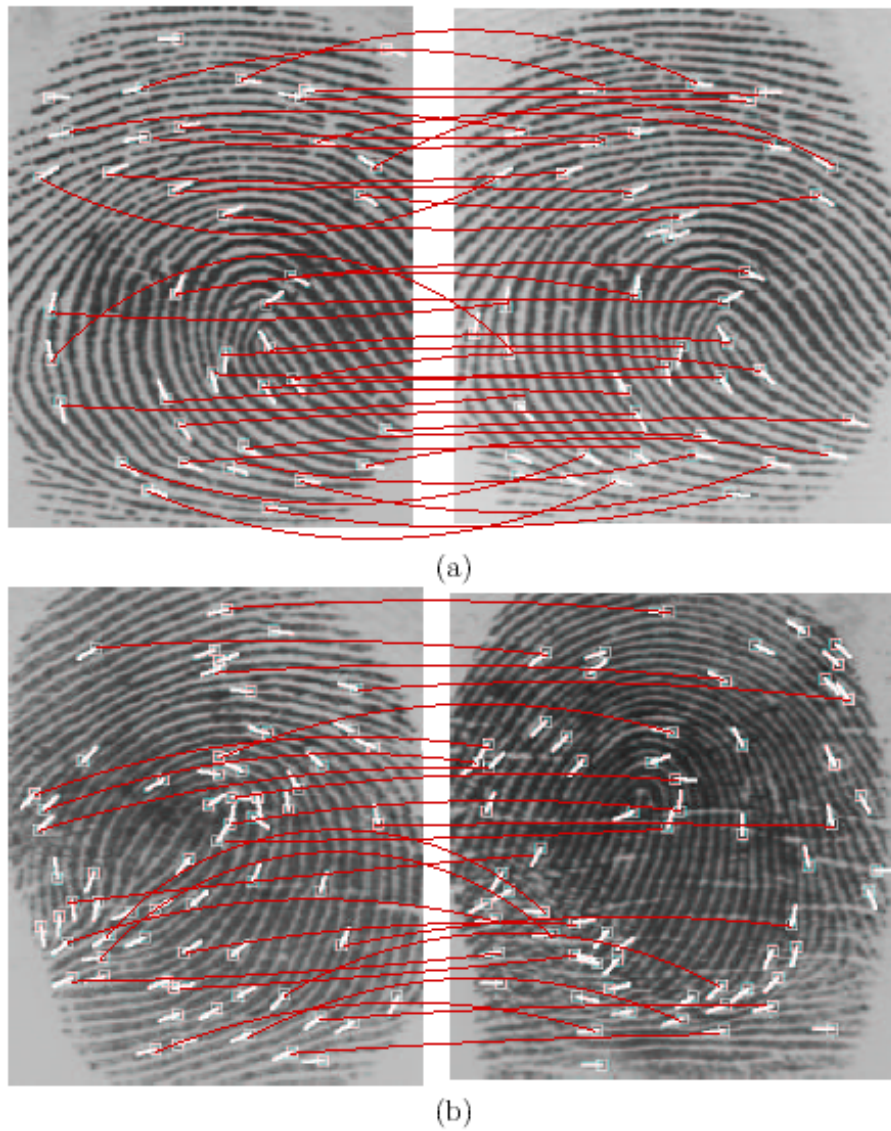


Figura 2.14: Emparejamiento automático de minucia. En (a) se emparejaron dos impresiones del mismo dedo, se detectaron 39 minucias en la entrada (izquierda), 42 en la plantilla (derecha) y 36 correspondencias verdaderas. En (b) se emparejaron dos dedos diferentes, se detectaron 64 minucias en la entrada (izquierda), 65 en la plantilla (derecha) y 25 correspondencias falsas.

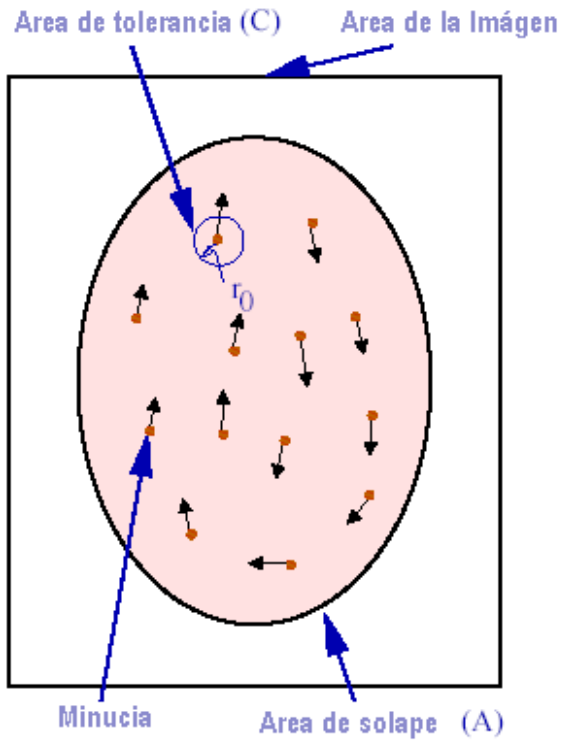


Figura 2.15: Huella digital y minucia

$$P = \{\{x_1, y_1, \theta_1\}, \{x_2, y_2, \theta_2\}, \dots, \{x_m, y_m, \theta_m\}\} \quad (2.6)$$

$$E = \{\{x'_1, y'_1, \theta'_1\}, \{x'_2, y'_2, \theta'_2\}, \dots, \{x'_m, y'_m, \theta'_m\}\} \quad (2.7)$$

Una vez se establece una alineación entre el conjunto de minucias de entrada y el conjunto de minucias de plantilla, se desarrolla un modelo de individualidad. primero se modela la variación intra-clase. Una minucia  $j$  en la huella digital de entrada se considera como “correspondiente” o “emparejada” a la minucia  $i$  en la plantilla, si y sólo si

$$\sqrt{(x'_i - x_j)^2 + (y'_i - y_j)^2} \leq r_0, \quad y \quad (2.8)$$

$$\text{mín} (|\theta'_i - \theta_j|, \quad 360 - |\theta'_i - \theta_j|) \leq \theta_0 \quad (2.9)$$

donde  $r_0$  es la tolerancia en la distancia y  $\theta_0$  es la tolerancia en el ángulo. El emparejamiento de huellas digitales manual y automático se basa en alguna tolerancia en la localización de la minucia y el ángulo a considerar para las variaciones en impresiones diferentes del mismo dedo. La ecuación (2.9) calcula el mínimo de  $|\theta'_i - \theta_j|$  y  $360 - |\theta'_i - \theta_j|$  porque los ángulos son mod 360 (la diferencia entre los ángulos de  $2^\circ$  y  $358^\circ$  es sólo  $4^\circ$ ).

Sea  $A$  el área total de solape, que se ha logrado después de una alineación razonable, entre las huellas digitales de entrada y plantilla. La probabilidad de que cualquier minucia arbitraria en la entrada emparejará con cualquier minucia arbitraria en la plantilla, sólo en términos de localización y dirección, está dada por las ecuaciones (2.10) y (2.11), respectivamente. La ecuación (2.10) asume que  $(x, y)$  y  $(x', y')$  son independientes y la ecuación (2.11) asume que  $\theta$  y  $\theta_0$  son independientes.

$$P(\sqrt{(x'_i - x_j)^2 + (y'_i - y_j)^2} \leq r_0) = \frac{\text{Área de tolerancia}}{\text{Total área sobrepuesta}} = \frac{\pi r_0^2}{A} = \frac{C}{A} \quad (2.10)$$

$$P(\min(|\theta'_i - \theta_j|, 360 - |\theta'_i - \theta_j|) \leq \theta_0) = \frac{\text{Ángulo de tolerancia}}{\text{Total ángulo}} = \frac{2\theta_0}{360} \quad (2.11)$$

Primero se desarrolla un modelo de correspondencia de huella digital cuando sólo las localizaciones individuales de minucia se emparejen y después se introducen los ángulos de la minucia en la formulación. Si la plantilla contiene  $m$  minucias, la probabilidad de que sólo una minucia en la entrada corresponda a cualquiera de las  $m$  minucias de la plantilla esta dada por  $\frac{mC}{A}$ . Ahora, dadas dos minucias de entrada, la probabilidad de que sólo la primera corresponda a una de las  $m$  minucias de la plantilla es el producto de las probabilidades que la primera minucia de entrada tiene una correspondencia ( $\frac{mC}{A}$ ) y la segunda minucia no tiene una correspondencia ( $\frac{A-mC}{A-C}$ ). Así, la probabilidad de que exactamente una de las dos minucias de entrada empareje con cualquiera de las  $m$  minucias de la plantilla es  $2 \times (\frac{mC}{A}) \times (\frac{A-mC}{A-C})$ , ya que la primera o la segunda minucia individual de entrada puede tener una correspondencia. Si la huella digital de entrada tiene  $n$  minucias, la probabilidad de que exactamente una minucia de entrada empareje con una de las  $m$  minucias de la plantilla es:

$$P(A, C, m, n) = \binom{n}{1} \left( \frac{mC}{A} \right) \left( \frac{A - mC}{A - C} \right) \quad (2.12)$$

La probabilidad de que haya exactamente  $p$  minucias coincidentes entre las  $n$  minucias de entrada y las  $m$  minucias de la plantilla esta dada por:

$$P(A, C, m, n, \rho) = \binom{n}{\rho} \underbrace{\left( \frac{mC}{A} \right) \left( \frac{(m-1)C}{A-C} \right) \dots \left( \frac{(m-\rho-1)C}{A-(\rho-1)C} \right)}_{\rho \text{ terminos}} \times \underbrace{\left( \frac{A-mC}{A-\rho C} \right) \left( \frac{A-(m-1)C}{A-(\rho+1)C} \right) \dots \left( \frac{A-(m-(n-\rho+1))C}{A-(n-1)C} \right)}_{n-\rho \text{ terminos}} \quad (2.13)$$

Los  $p$  primeros términos en la ecuación (2.13) denotan la probabilidad de emparejar  $p$  minucias entre la plantilla y la entrada; y los  $n - p$  términos restantes expresan la probabilidad de que las  $n - p$  minucias de la entrada no emparejen con ninguna minucia en la plantilla. Dividiendo el numerador y denominador de cada término en la ecuación (2.13) por  $C$ , obtenemos:

$$P(A, C, m, n, \rho) = \binom{n}{\rho} \underbrace{\left( \frac{m}{\frac{A}{C}} \right) \left( \frac{(m-1)}{\frac{A}{C}-1} \right) \dots \left( \frac{(m-\rho-1)}{\frac{A}{C}-(\rho-1)} \right)}_{\rho \text{ terminos}} \times \underbrace{\left( \frac{\frac{A}{C}-m}{\frac{A}{C}-\rho} \right) \left( \frac{\frac{A}{C}-(m-1)}{\frac{A}{C}-(\rho+1)} \right) \dots \left( \frac{\frac{A}{C}-(m-(n-\rho+1))}{\frac{A}{C}-(n-1)} \right)}_{n-\rho \text{ terminos}} \quad (2.14)$$

Haciendo  $M = \frac{A}{C}$  obtenemos,

$$P(A, C, m, n, \rho) = \binom{n}{\rho} \underbrace{\left( \frac{m}{M} \right) \left( \frac{(m-1)}{M-1} \right) \dots \left( \frac{(m-\rho-1)}{M-(\rho-1)} \right)}_{\rho \text{ terminos}} \times \underbrace{\left( \frac{M-m}{\frac{A}{C}-\rho} \right) \left( \frac{M-(m-1)}{M-(\rho+1)} \right) \dots \left( \frac{M-(m-(n-\rho+1))}{M-(n-1)} \right)}_{n-\rho \text{ terminos}} \quad (2.15)$$

Asumiendo que  $M$  es un entero (que es una suposición realista porque  $A \gg C$ ), se puede escribir la ecuación anterior en forma compacta:

$$P(A, C, m, n, \rho) = \frac{n!}{\rho!(n-\rho)!} \times \frac{(M-n)!}{M!} \times \frac{m!}{(m-\rho)!} \times \frac{(M-m)!}{((M-m)-(n-\rho))!} \quad (2.16)$$

Reestructurando los términos,

$$P(A, C, m, n, \rho) = \frac{m!}{\rho!(m-\rho)!} \times \frac{(M-m)!}{(n-\rho)!((M-m)-(n-\rho))!} \times \frac{(M-n)!n!}{M!} \quad (2.17)$$

finalmente se reduce a:

$$P(A, C, m, n, \rho) = \frac{\binom{m}{\rho} \binom{M-m}{n-\rho}}{\binom{M}{n}} \quad (2.18)$$

La ecuación (2.18) define una distribución hiper-geométrica. Para conseguir una comprensión intuitiva del modelo de probabilidad por la correspondencia de minucias en dos huellas digitales, imagine que el área solapada de la plantilla y las huellas digitales de entrada está dividida en  $M$  celdas que no se solapan. La forma de las celdas individuales no importa, sólo el número de celdas. Ahora considere una baraja que contiene  $M$  tarjetas distintas. Cada tarjeta representa una celda en el área de solapamiento. Hay una baraja para la plantilla de la huella digital y una baraja idéntica para la huella digital de entrada. Si se sacan  $m$  tarjetas de la primera baraja (plantilla) sin reemplazo, y se sacan  $n$  tarjetas de la segunda baraja (entrada) sin reemplazo, la probabilidad de emparejar  $q$  tarjetas exactamente entre las tarjetas seleccionadas esta dada por la distribución hiper-geométrica en la ecuación (2.18) [26]. El análisis anterior considera la correspondencia de minucia basada solamente en la localización de ella. Luego se considera la correspondencia de la minucia que depende de las direcciones y las posiciones de ella. Para este análisis, se asumió que las direcciones de la minucia son completamente independientes de las posiciones

de ella y el emparejamiento de la posición y dirección de la minucia son por consiguiente eventos independientes.

Sea  $l$  tal que  $P(\min(|\theta_i - \theta_j|, 360 - |\theta_i - \theta_j|) \leq \theta_0) = \frac{1}{l}$  en la ecuación (2.11). Dadas  $n$  minucias de entrada y  $m$  de la plantilla, la probabilidad de que  $p$  minucias caigan entre las posiciones similares se puede estimar por la ecuación (2.18). Una vez se emparejan  $p$  posiciones de minucia, la probabilidad que  $q$  minucias entre ellas tengan direcciones similares está dada por

$$\binom{\rho}{q} \left(\frac{1}{l}\right)^q \left(\frac{l-1}{l}\right)^{\rho-q} \quad (2.19)$$

donde  $\frac{1}{l}$  es la probabilidad de que dos posiciones de minucia emparejadas tengan una dirección similar y  $\frac{l-1}{l}$  es la probabilidad de que dos posiciones de minucias emparejadas tomen direcciones diferentes. Por consiguiente, la probabilidad de emparejar  $q$  minucias tanto en la posición como en la dirección esta dada por

$$P(A, C, m, n, \rho) = \sum_{\rho=q}^{\min(m,n)} \left( \frac{\binom{m}{\rho} \binom{M-m}{n-\rho}}{\binom{M}{n}} \times \binom{\rho}{q} \left(\frac{1}{l}\right)^q \left(\frac{l-1}{l}\right)^{\rho-q} \right) \quad (2.20)$$

Hasta ahora, se ha asumido que las posiciones de minucia están uniformemente distribuidas dentro del área completa de la huella digital. Ya que  $A$  es el área de solape entre las huellas digitales de la plantilla y la entrada, las crestas ocupan aproximadamente  $\frac{A}{2}$  del área, y la otra media es ocupada por los surcos. Ya que la minucia sólo puede quedar en las crestas, es decir, a lo largo de una curva de longitud  $\frac{A}{w}$ , donde  $w$  es el periodo de la cresta, el valor de  $M$  en la ecuación (2.20) debe cambiarse de  $M = \frac{A}{C}$  a  $M = \frac{A/w}{2r_0}$ , donde  $2r_0$  es la tolerancia de longitud en la localización de la minucia.

Este modelo de individualidad tiene varios parámetros, a saber,  $r_0, l, w, A, m, n,$  y  $q$ ; y su estimación se puede ver en [24].



## Capítulo 3

# Extracción de características

Hay dos tipos principales de rasgos tradicionales, en las huellas digitales:

1. Las crestas globales y las estructuras de surco que forman un modelo especial en la región central de las huellas digitales.
2. los detalles de minucias asociados con las crestas locales y los surcos (o valles).

Una huella digital se clasifica comúnmente, basada sólo en el primer tipo de características y se identifica únicamente, basada en el segundo tipo de características. La representación basada en minucias es la representación más común de las huellas digitales, dado que tiene una larga historia de uso por expertos forenses que las emparejan visualmente, los forenses expertos usan para los propósitos de identificación, otras características tales como el conteo de crestas entre los pares de minucia y la anchura de la cresta junto con la minucia. Sin embargo, el proceso automático de indentificación de huellas digitales permite el uso de coordenadas Cartesianas y distancia Euclidea para establecer que tenga un alto poder discriminatorio, es posible para sistemas automáticos de huellas digitales, establecer una cota superior en el desempeño de los sistemas basados en minucias debido a la cantidad limitada de información en dicha representación.

Como resultado, es deseable explorar una representación independiente alternada de huellas digitales que pueden complementar la representación basada en minucia. Esta

representación complementaria debe combinar ambas fuentes de información la global y local en una huella para obtener una mejor representación. Esta representación no sólo debe tener en cuenta las anomalías locales en la estructura de la cresta (minucia), sino también, por ejemplo, el modelo global de las crestas y surcos, distancias entre crestas, y los patrones globales de flujo de la cresta. Además, es una ventaja agregada para diseñar representaciones que pueden ser automática y fiablemente extraídas de la huella y cuya extracción se degradará con deterioración de la calidad de las huellas. Los patrones suaves de flujo de crestas y valles en una huella digital pueden verse como un campo de textura orientada (Figura 3.1). La intensidad de la imagen de superficie de la huella digital consta de crestas cuyas direcciones varían continuamente y constituyen un textura orientada. La mayoría de las texturas de las imágenes contienen un rango limitado de frecuencias espaciales, y las texturas mutuamente distintas difieren significativamente en sus frecuencias dominantes. Si la región de textura, posee frecuencia espacial diferente, orientación, o puede ser fácilmente discriminada la fase, descomponiendo la imagen en varias frecuencias espaciales y canales de orientación. Para imágenes comunes examinadas a 500 dpi, hay pequeñas variaciones en las frecuencias espaciales (determinadas por las distancias entre crestas) entre huellas digitales diferentes.

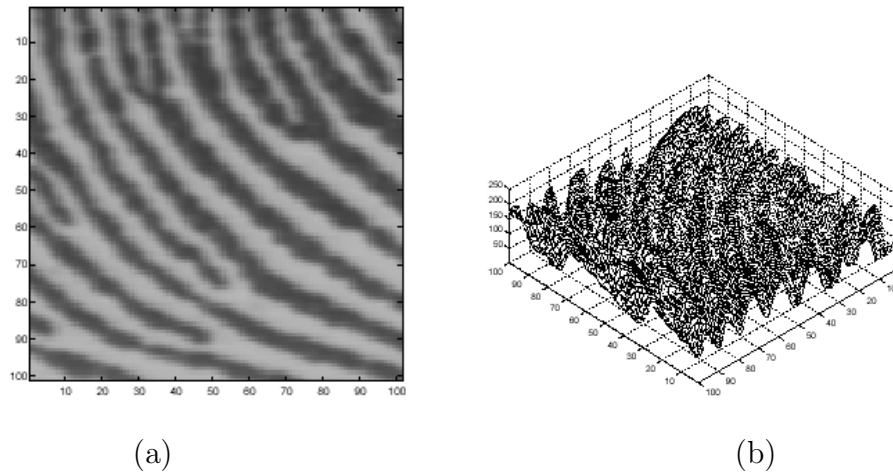


Figura 3.1: Patrones de flujo en una imagen de la huella digital. (a) Una sección de la imagen de la huella digital, (b) Región de la superficie de (a) en 3D.

Esto implica que hay una escala óptima (frecuencia espacial) para analizar la textura de la huella digital. Cada píxel en una imagen de la huella digital es asociado con una orientación local dominante y una medida local de coherencia del patrón de flujo. Una descripción simbólica de una imagen puede ser deducida calculando el ángulo y la coherencia a cada píxel de la imagen. Las huellas digitales pueden ser representadas/emparejadas usando medidas cuantitativas asociadas con el modelo de flujo (texturas orientadas) como características.

El análisis y modelado de texturas orientadas es un problema importante en el campo de investigación con una amplia variedad de aplicaciones prácticas. Los anteriores esfuerzos en describir texturas orientadas han usado características exclusivamente locales o predominantemente globales. Las limitaciones iniciales de las aproximaciones locales a la representación de una textura orientada es que ella no captura eficazmente la totalidad de la información discriminatoria. La información local también tiende a ser inestable y propensa al ruido. Los ejemplos de representaciones globales incluyen matrices de co-ocurrencia direccional y retratos de la fase de los campos orientados.

Los métodos de auto correlación son una representación global de la textura, descomponiendo la imagen de entrada en diferente frecuencia y las componentes de orientación usando un banco de filtros de Gabor [10]. Se aplica esta representación para clasificar con éxito y segmentar las texturas de imágenes.

Para representar la textura de una huella digital, una de las formas consiste en extraer uno (o más) puntos de referencia invariantes de la textura, basados en un análisis de sus campo de orientación. Una región predeterminada de interés alrededor del punto de referencia es fragmentada en celdas. Cada celda es examinada para extraer la información acerca de una o más orientaciones específicas y canales de frecuencia espaciales. Se usa como la representación de la huella digital una enumeración ordenada de las características así extraídas de cada celda (vea Figura 3.2). Así, los elementos de la representación y la enumeración ordenada de la fragmentación capturan la información local y las relaciones

globales invariantes entre los patrones locales respectivamente.

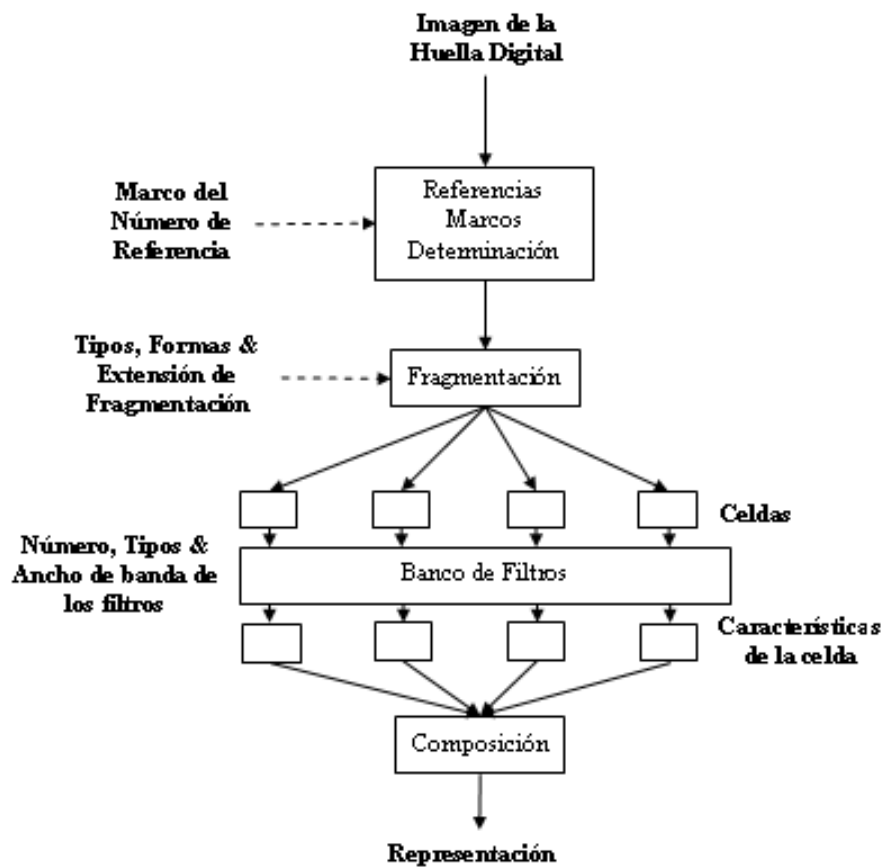


Figura 3.2: Diagrama esquemático para la extracción genérica de la representación basada en texturas para huellas digitales.

Es deseable obtener representaciones para huellas digitales invariantes (debido a la presión y a la resolución del sensor), para la traslación y rotación. La invarianza a escala (la escala a la que se está tomando la imagen influye en la identificación de la huella digital) no es un problema significativo puesto que la mayoría de las imágenes de la huella digital pueden ser escaneadas según las especificaciones del dpi del sensor. La (figura 3.3) muestra que la huella digital de un niño la cual tiene un área más pequeña que la huella digital de un adulto cuando fueron escaneadas a la misma resolución. Cuando un niño crece, la diferencia de escala entre sus huellas digitales adquiridas a diferente edad, puede ocasionar la no correspondencia entre las dos imágenes de la huella.



Figura 3.3: Huella digital de (a) un niño, y (b) un adulto. Ambas huellas digitales se examinaron a 500 dpi.

Periódicamente actualizando la plantilla de las huellas digitales se resuelve este problema. La invarianza de la traslación se lleva a cabo localizando el punto de referencia. La representación propuesta aquí no es invariante por rotación y así la rotación se maneja por una rotación de la representación en la fase de emparejando. Se define una fragmentación circular de tal manera que una rotación de la imagen de la huella corresponda a una rotación cíclica de los elementos de la representación. La información local discriminativa en el sector necesita ser descompuesta en componentes separadas. Un banco de filtros de Gabor es una de las técnicas mas conocidas para capturar la información útil en los canales específicos así como para descomponer esta información en componentes ortogonales en términos de frecuencias espaciales. Los cuatro principales pasos en el algoritmo de extracción son:

1. Determinar el punto de referencia de la imagen de la huella digital.
2. fragmentar la región alrededor del punto de la referencia.
3. Filtrar la región de interés en ocho direcciones diferentes usando un banco de filtros de Gabor (se requieren ocho direcciones para capturar las características de las

configuraciones locales completamente en una huella digital mientras sólo cuatro direcciones son requeridas para capturar la configuración global).

4. Calcular la desviación absoluta de los valores de gris (AAD) en sectores individuales en imágenes filtradas para definir el vector de características.

El organigrama total de este algoritmo se representa en la figura 3.4.

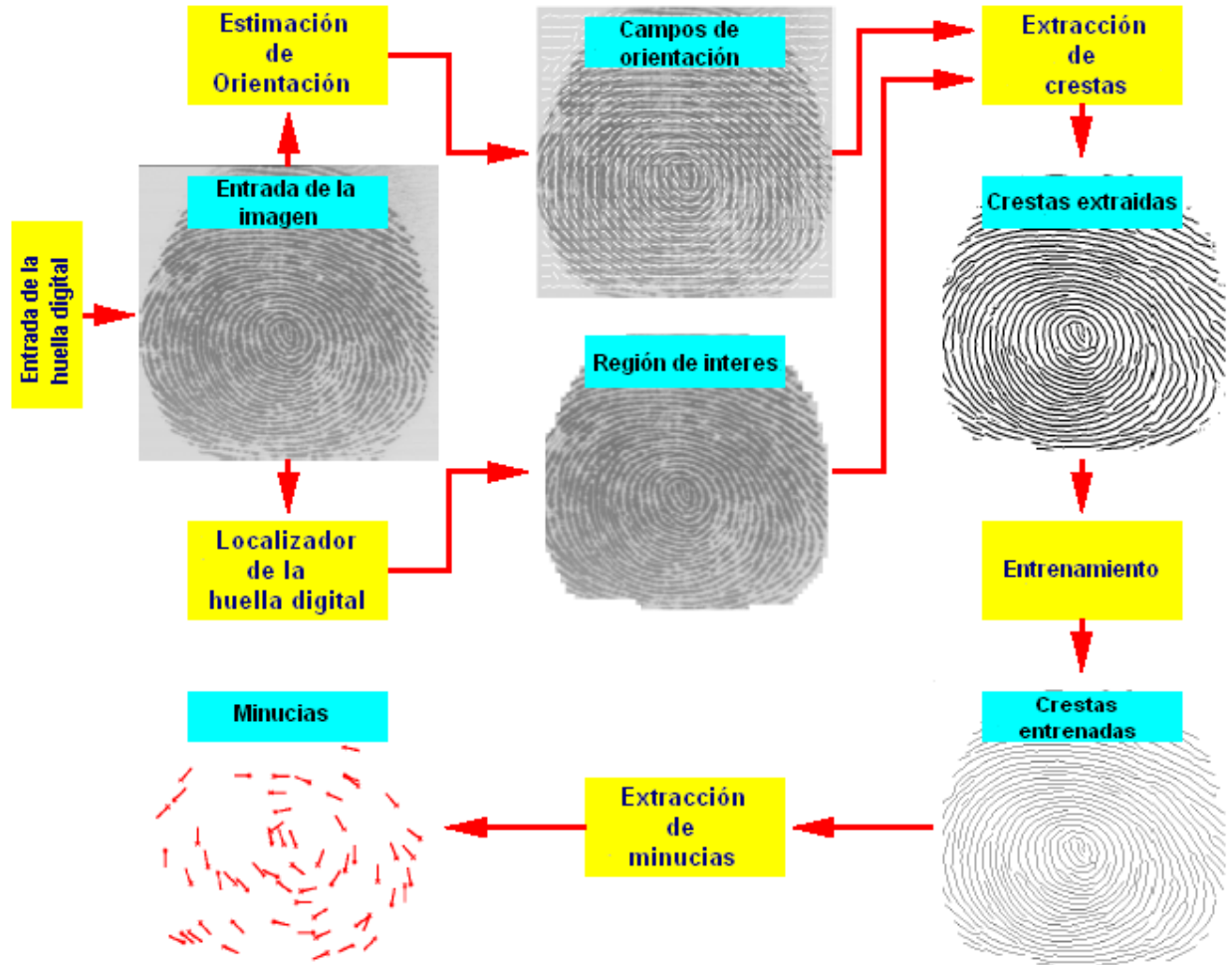


Figura 3.4: Organigrama del algoritmo de extracción de minucias

### 3.1. Localización del punto de referencia

Las huellas digitales tienen muchas señales visibles y cualquier combinación de ellas podrían ser usadas para establecer un punto de referencia. Definimos el punto de referencia de una huella digital como el punto de máxima curvatura de las crestas cóncavas (vea Figura 3.5) en la imagen de la huella digital.

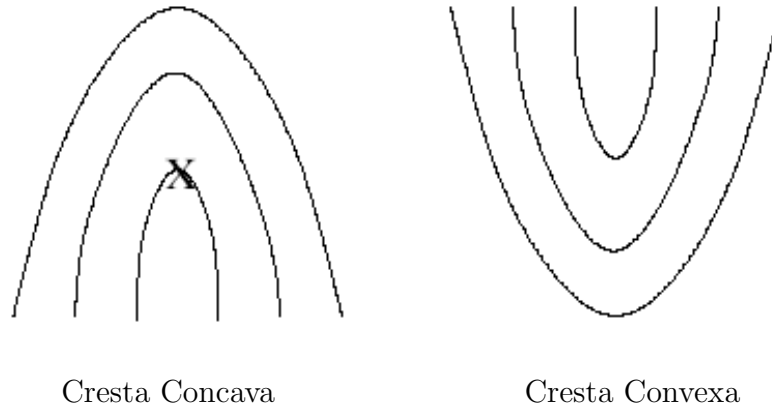


Figura 3.5: Las crestas cóncavas y convexas en una imagen de la huella digital cuando el dedo es posicionado derecho. El punto de la referencia es marcado por  $X$ .

Muchas aproximaciones previas a la determinación de un punto de referencia  $(x_c, y_c)$  confían de manera crítica en las características locales como el índice de Poincaré [23] o algunas otras propiedades similares del campo de orientación. Mientras estos métodos trabajan bien para imágenes de buena calidad, fallan para localizar correctamente puntos de referencia en imágenes de mala calidad. Para que un algoritmo calcule el punto de referencia, debe manejar el ruido local considerando un entorno grande de la imagen de baja calidad. Por otro lado, para una localización certera del punto de referencia, la aproximación debe ser sensible a las variaciones locales en pequeños entornos.

Una imagen  $I$  de tamaño  $M \times N$  píxeles con una mayor precisión de su campo de orientación  $O$ , se define como una imagen  $P \times Q$ , donde  $O(i, j)$  representa la orientación de la cresta local en el píxel  $(i, j)$ ,  $P < M$ ;  $Q < N$ . La orientación de la cresta local es normalmente especificada para un bloque en lugar de para cada píxel en la imagen  $I$ . La imagen

de la huella digital es dividida en un conjunto de  $w \times w$  bloques disjuntos y se define una única orientación para cada bloque (vea Figura 3.7(a) y (b));  $P = \lfloor M/w \rfloor$ ;  $Q = \lfloor N/w \rfloor$ , donde  $\lfloor \cdot \rfloor$  es la parte entera tomada por debajo. Note que hay una ambigüedad para un factor de  $\pi$  en la orientación de la huella digital, es decir, la orientación de las crestas locales a  $(\pi/2)$  y la orientación de las crestas a  $(3\pi/2)$  no pueden ser diferenciados entre si. Se han desarrollado varios métodos para estimar el campo de orientación en una huella digital. El algoritmo de estimación para mínimos cuadrados de la orientación utilizado aquí, tiene los siguientes pasos:

### 3.2. Algoritmo gerárquico de los campos de orientación

1. Dividir  $I$ , la imagen de huella digital de entrada, en bloques disjuntos de tamaño  $w \times w$ .

Un problema de importancia fundamental en el análisis de imagen es la detección de bordes. Los contornos caracterizan las fronteras de los objetos, y por tanto son de gran utilidad de cara a la segmentación e identificación de objetos en escenas. Los puntos de contorno son como zonas de pixeles en las que existe un cambio brusco de nivel de gris. Si pensamos en una imagen como una función continua  $f(x, y)$ , vemos que su derivada tiene un máximo local en la dirección del contorno. Es por esto que las técnicas más usadas en la detección de contornos se basen en la medida del gradiente de  $f$  a lo largo de  $r$  en una dirección  $\theta$ , (ver figura 3.6).

$$\frac{\partial f}{\partial r} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial r} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial r} = f_x \cos(\theta) + f_y \sin(\theta) \quad (3.1)$$

El máximo valor de  $\frac{\partial f}{\partial r}$  se logra cuando  $\frac{\partial}{\partial \theta} \left( \frac{\partial f}{\partial r} \right) = 0$ .

$$-f_x \sin(\theta_g) + f_y \cos(\theta_g) = 0 \quad (3.2)$$

Resolviendo para  $\theta_g$  tenemos:



$$f_x \sin(\theta_g) = f_y \cos(\theta_g)$$

$$\tan(\theta_g) = \frac{f_y}{f_x}$$

$$\theta_g = \arctan\left(\frac{f_y}{f_x}\right)$$

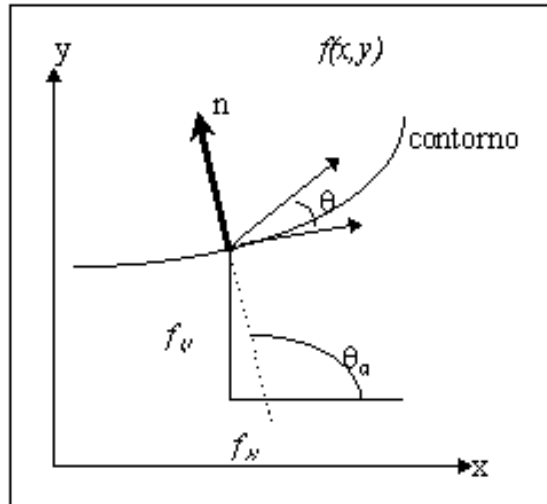


Figura 3.6: Gradiente de  $f(x, y)$

2. Calcular los gradientes  $\partial_x(i, j)$  y  $\partial_y(i, j)$  a cada píxel  $(i, j)$ . Dependiendo del requerimiento computacional, el operador de gradientes puede variar desde Operador de Sobel hasta el más complejo operador de Marr-Hildreth [20].

Formalmente, si la función de intensidad de la imagen en el espacio discreto es  $f(i, j)$  con  $1 \leq i \leq M$ ,  $1 \leq j \leq N$  el vector gradiente es:

$$\nabla f(i, j) = \frac{1}{2} ((f(i+1, j) - f(i-1, j)), (f(i, j+1) - f(i, j-1))) \quad (3.3)$$

De la ecuación (3.2) podemos definir las siguientes ecuaciones:

$$f_x = v_x(i, j)$$

$$f_y = v_y(i, j)$$

$$\theta_g = O(i, j)$$

3. Estimar la orientación local de cada bloque centrado en el píxel  $(i, j)$  usando las siguientes ecuaciones:

$$v_x(i, j) = \sum_{u=i-\frac{w}{2}}^{i+\frac{w}{2}} \sum_{v=j-\frac{w}{2}}^{j+\frac{w}{2}} 2\partial_x(u, v) \partial_y(u, v) \quad (3.4)$$

$$v_y(i, j) = \sum_{u=i-\frac{w}{2}}^{i+\frac{w}{2}} \sum_{v=j-\frac{w}{2}}^{j+\frac{w}{2}} (\partial_x^2(u, v) - \partial_y^2(u, v)) \quad (3.5)$$

$$O(i, j) = \arctan\left(\frac{v_y(i, j)}{v_x(i, j)}\right) \quad (3.6)$$

Donde  $O(i, j)$  es la estimación de mínimos cuadrados de la orientación de la cresta local del bloque centrado al píxel  $(i, j)$ . Matemáticamente, representa la dirección que es ortogonal a la dirección dominante del espectro de Fourier de la ventana  $w \times w$  [13, 33].

### 3.3. Algoritmo de localización de puntos de referencia

1. Estimar el campo de orientación  $O$  como se describió, usando un bloque de tamaño  $w \times w$ .
2. Suavizar el campo de orientación en una vecindad local. Sea  $\hat{O}$  el campo de orientación suavizada. Para suavizar (filtrado de bajo nivel), la orientación de la imagen necesita ser convertida en un campo vectorial continuo, el cual se define como sigue:

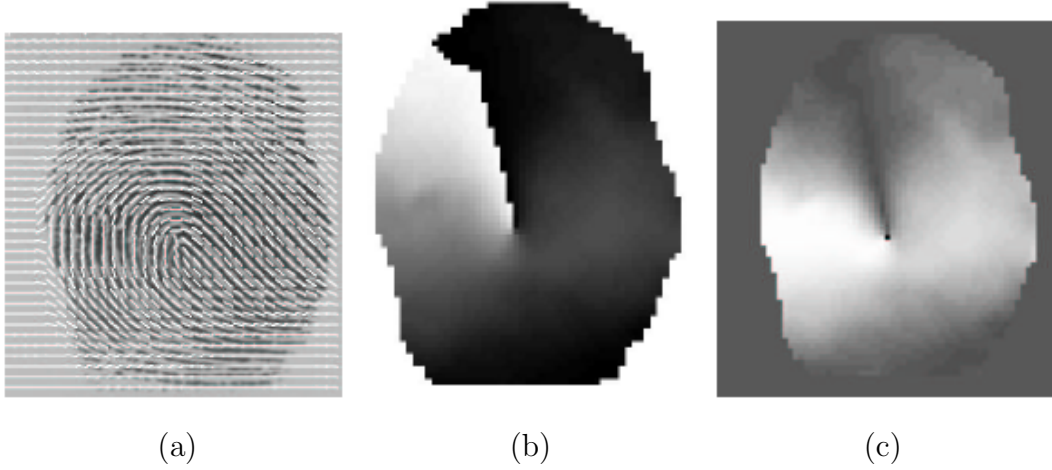


Figura 3.7: Estimación del punto de referencia. (a) Campo de orientación suavizado en la imagen original, (b) campo de orientación ( $w = 10$ ) mostrado como la distribución de intensidad; el fondo se ha segmentado, y (c) la componente sinusoidal del campo de orientación; el píxel más oscuro en el centro de la imagen marca el punto de referencia detectado.

$$\Phi_x(i, j) = \cos(2O(i, j))$$

$$\Phi_y(i, j) = \sin(2O(i, j))$$

Para ser apreciadas las imágenes han sido escalonadas de 0 – 255, donde  $\Phi_x(i, j)$  y  $\Phi_y(i, j)$ , son las componentes  $x$  e  $y$  del campo vectorial, respectivamente. Un filtrado de bajo nivel del campo vectorial resultante se ha realizado como sigue:

$$\Phi'_x(i, j) = \sum_{u=-w_\Phi/2}^{w_\Phi/2} \sum_{v=-w_\Phi/2}^{w_\Phi/2} W(u, v) \Phi_x(i - uw, j - vw) \quad (3.7)$$

$$\Phi'_y(i, j) = \sum_{u=-w_\Phi/2}^{w_\Phi/2} \sum_{v=-w_\Phi/2}^{w_\Phi/2} W(u, v) \Phi_y(i - uw, j - vw) \quad (3.8)$$

Donde  $W$  es  $w_\Phi \times w_\Phi$  es un filtro de bajo-nivel de  $x$ . Nótese que la operación de suavizado se lleva a cabo al nivel del bloque. El campo de orientación suavizada  $\acute{O}(i, j)$  es calculado como sigue:

$$\acute{O}(i, j) = \frac{1}{2} \arctan \left( \frac{\Phi'_y(i, j)}{\Phi'_x(i, j)} \right) \quad (3.9)$$

3. Calcular  $\epsilon$  como una imagen que incluya sólo la componente del seno de  $\acute{O}$ .

$$\epsilon(i, j) = \sin \left( \acute{O}(i, j) \right) \quad (3.10)$$

4. Inicializar  $A$ , una imagen de etiquetado para indicar el punto de referencia.
5. Para cada píxel  $(i, j)$  en  $\epsilon$ , integrar las intensidades del píxel (componente del seno del campo de orientación) en las regiones  $RI$  y  $RII$  mostradas en la figura 3.8 y asignar a los correspondientes pixeles en  $A$  el valor de su diferencia.

$$A(i, j) = \sum_{RI} \epsilon(i, j) - \sum_{RII} \epsilon(i, j) \quad (3.11)$$

Las regiones  $RI$  y  $RII$  (vea la figura 3.8) son empíricamente determinada aplicando el algoritmo de localización de puntos de referencia de un banco de datos grande. El radio de la región semi-circular fue fijado igual a la ventana de tamaño  $w$ . La geometría de las regiones  $RI$  y  $RII$  se diseñan para capturar la máxima curvatura en las crestas cóncavas (vea la figura 3.5). Aunque esta aproximación detecta bien el punto de referencia, incluyendo lazos dobles (vea la figura 3.9(a)), la presente aplicación no es muy precisa y consistente para las huellas digitales de tipo arch porque se dificulta localizar puntos de alta curvatura en imágenes de huellas digitales de este tipo.

6. Encontrar el valor máximo de  $A$  y asignar sus coordenadas al centro, es decir, al punto de referencia.
7. Para un número fijo de tiempos, repetir los pasos 1-6 usando una ventana de tamaño  $\acute{w} \times \acute{w}$  donde  $\acute{w} < w$  y restringir la búsqueda del punto de referencia en el paso 6 a una vecindad local del punto de referencia detectado.

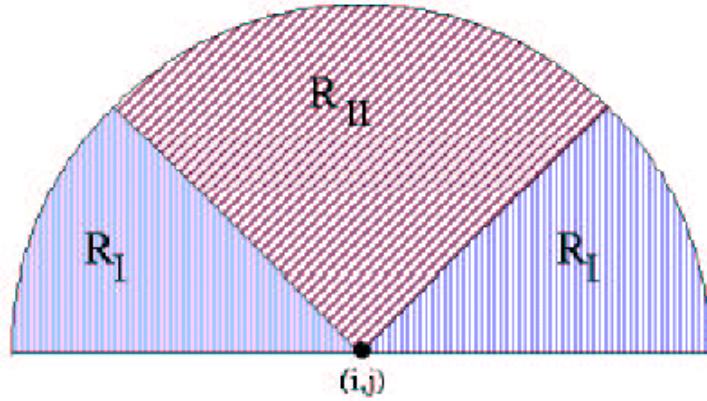


Figura 3.8: Integrar por regiones las intensidades de los píxeles en  $\epsilon$  para el cálculo de  $A(i, j)$ .

8. Calcular el nivel de consistencia de los campos de orientación en una vecindad local de un bloque de píxel  $(i, j)$  con la siguiente formula:

$$C(i, j) = \frac{1}{N} \sqrt{\sum_{(i,j) \in D} |\acute{O}(i, j) - O(i, j)|^2}, \quad (3.12)$$

$$|\acute{O} - O| = \begin{cases} d & \text{si } d = (\acute{O} - O + 360^\circ) \pmod{360^\circ} < 180^\circ \\ d - 180^\circ & \text{Otro caso} \end{cases}$$

Donde  $D$  representa la vecindad local alrededor de un bloque de píxel  $(i, j)$  (donde el tamaño de  $D$  es de  $w \times w$ );  $N$  es el número de bloques dentro de  $D$ ;  $\acute{O}(i, j)$  y  $O(i, j)$  son las orientaciones locales de las crestas en los bloques  $(i, j)$  de  $\acute{O}$  y  $O$  respectivamente.

9. Si el nivel de consistencia de la ecuación(3.12) está sobre un cierto umbral  $T_c$  la orientación local alrededor de esta región son reestimados a una baja resolución hasta que  $C(i, j)$  esté por debajo de un cierto nivel.

La figura 3.9 muestra los resultados del algoritmo de localización de puntos de referencia para cuatro imágenes diferentes. El algoritmo de localización de puntos de referencia se desempeña bien para imágenes de huellas digitales de alta calidad como whorl, left loop,

right loop, y tipos arch. Este algoritmo tiene un alto error en localizar consistentemente puntos de referencia de las huellas digitales del tipo arch debido a la ausencia de puntos singulares en la imagen de dicho tipo. El algoritmo falla para huellas digitales de mala calidad debido a los errores en la estimación de los campos de orientación.

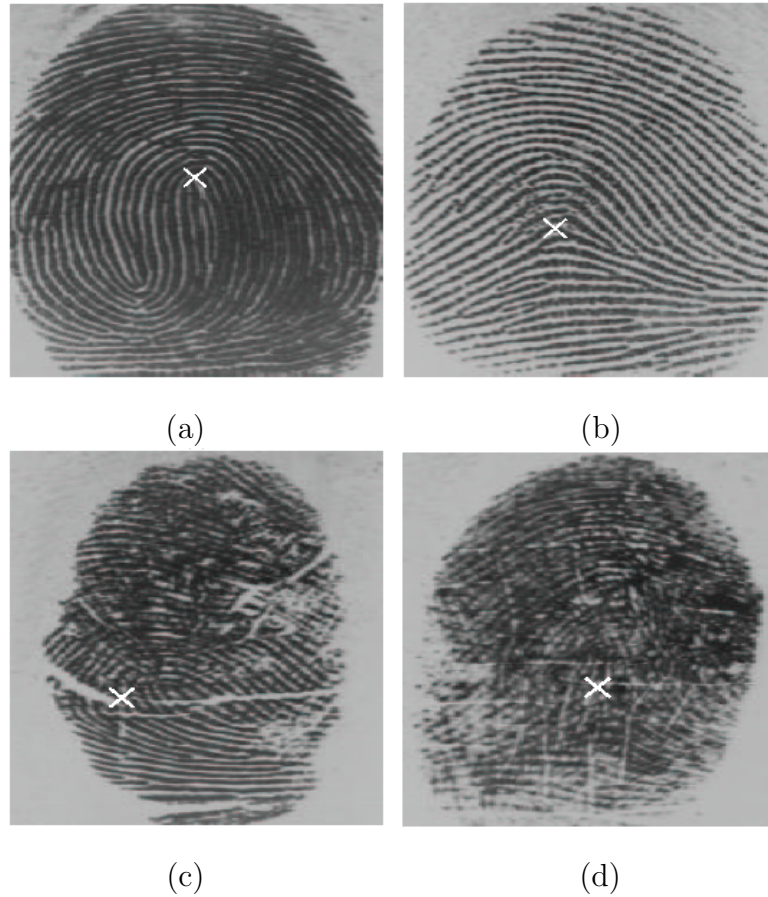


Figura 3.9: Ejemplos de los resultados del algoritmo de localización de puntos de referencia. El algoritmo falla en huellas digitales de mala calidad como (c) y (d).

Después de que se estime el campo de la orientación de una imagen de entrada de la huella digital, se utiliza un algoritmo de segmentación el cual se basa en el nivel local de certeza del campo de orientación para localizar la región de interés dentro de la imagen de la huella digital de entrada. El nivel de certeza del campo de la orientación en el píxel  $(i, j)$  es definido como sigue:

$$CL = \sqrt{\frac{1}{w \times w} \frac{(v_x(i, j)^2 + v_y(i, j)^2)}{v_e(i, j)}} \quad (3.13)$$

Donde,

$$v_e(i, j) = \sum_{u=i-\frac{w}{2}}^{i+\frac{w}{2}} \sum_{v=j-\frac{w}{2}}^{j+\frac{w}{2}} (\partial_x^2(u, v) + \partial_y^2(u, v)) \quad (3.14)$$

y  $w$  es el tamaño de una vecindad local. Para cada píxel, si el nivel de certeza del campo de orientación está por debajo de cierto umbral  $T_s$ , entonces el píxel es marcado como píxel de fondo. En este algoritmo de localización se asume que sólo una imagen de la huella digital está presente la cual se utiliza con métodos no rigurosos para encontrar la región de interés.

### 3.4. Fragmentación

Sea  $I(x, y)$  el nivel de gris del píxel  $(x, y)$  en la imagen de la huella digital  $M \times N$  y sea  $(x_c, y_c)$  el punto de referencia. La región de interés de la huella digital se define como la colección de todos los sectores  $S_i$ , donde el  $i$ -ésimo sector de  $S_i$  se calcula en términos de los parámetros  $(r, \theta)$  como sigue:

$$S_i = \{(x, y) | b(T_i + 1) \leq r < b(T_i + 2), \theta_i \leq \theta < \theta_{i+1}, 1 \leq x \leq N, 1 \leq y \leq M\} \quad (3.15)$$

Donde,

$$T_i = i \operatorname{div} k, \quad (3.16)$$

$$\theta_i = (i \operatorname{mod} k) \times (2\pi/k), \quad (3.17)$$

$$r = \sqrt{(x - x_c)^2 + (y - y_c)^2}, \quad (3.18)$$

$$\theta = \arctan\left(\frac{(y - y_c)}{(x - x_c)}\right) \quad (3.19)$$

$b$  es el ancho de cada banda,  $k$  es el número de sectores considerados en cada banda,  $i = 0, \dots, (B \times k - 1)$ , donde  $B$  es el número de bandas concéntricas consideradas alrededor del punto de referencia para la extracción de características. El parámetro  $B$  depende del área de la imagen del surco capturado. Así el parámetro  $B$  depende del tamaño de la imagen y de la resolución dpi del sensor. El ancho de las bandas concéntricas se define por el parámetro  $b$  y depende de la resolución dpi del sensor. El ancho de las bandas debe capturar en promedio una pareja de crestas y surcos. Una banda con un ancho de 20 píxeles (es el más utilizado) necesariamente captura una única minucia en un sector y permite características de bajo-nivel para capturar esta información local. Si el ancho del sector es más de 20 píxeles, entonces la información local es modulada a través de una información más global. La banda más profunda (círculo) no se usa para la extracción de características porque los sectores en la región cerca del punto de referencia contienen muy pocos píxeles y, por consiguiente, la extracción de características en esta región no es muy fiable. Se escoge una fragmentación circular porque la rotación de la huella digital corresponderá a la rotación de la fragmentación.

El valor de  $k$  controla la captura de la información global contra la información local en una huella digital, y depende de la aplicación. Por ejemplo, se requiere mucha información global para el algoritmo de clasificación de huellas, y así, se escoge el mínimo valor de  $k$ . Por otro lado, la aplicación de comprobación de la huella necesita capturar más información local y se requiere un valor de  $k$  más alto. Los valores para estos parámetros,  $B$ ,  $b$ , y  $k$  son determinados empíricamente para obtener el mejor desempeño en la clasificación y las aplicaciones de emparejamiento. Ambos algoritmos, el de clasificación y el de emparejamiento se basan en la representación de unos vectores de características que están habilitados para manejar cambios pequeños en estos parámetros sin una degradación significativa en su desempeño proporcional al cambio en los valores del parámetro. El valor de  $B$  debe fijarse de tal modo que capture el máximo número de detalles de las crestas y surcos sin rechazar un número grande de imágenes de la huella digital. El valor de  $k$  será escogido basado en la conciliación entre la información local y global requerida para una aplicación particular, el valor de  $b$  debe escogerse basado en la resolución dpi del

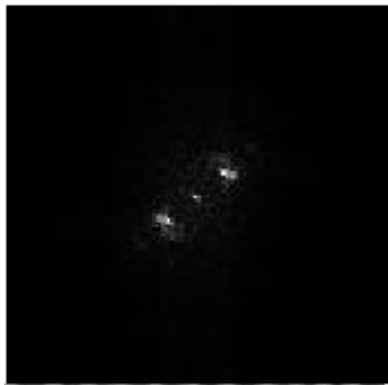


sensor y la distancia entre las crestas en la imagen de la huella. Una vez se escogen los valores del parámetro para una aplicación, ellos permanecen constantes.

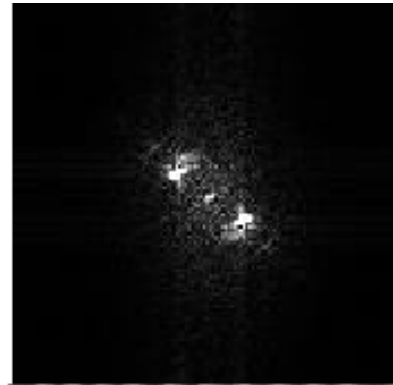
### 3.5. Filtrado

Las huellas digitales tienen crestas y surcos paralelos locales, además frecuencias y orientación local bien definidas (Ver figura 3.10). El uso apropiado de los filtros de Gabor (Ver figura 3.11) puede remover el ruido, preservar las estructuras verdaderas de crestas y surcos, y dan información contenida en una orientación particular de la imagen. Un punto de minucia puede verse como una anomalía local en las crestas paralelas y es la información que tratamos de capturar usando los filtros de Gabor.

Luego del filtrado de la imagen de la huella digital, normalizamos la intensidades de los niveles de gris en la región de interés de cada sector separadamente basados en la media y la varianza. La normalización se realiza para quitar los efectos de ruido del sensor y los niveles de gris del fondo, debido a las diferencias de presión del dedo. Sean  $I(x, y)$  los pixeles de los niveles de gris en el píxel  $(x, y)$ ,  $M_i$  y  $V_i$  la estimación de la media y la varianza de niveles de gris en el sector  $S_i$  respectivamente y  $N_i(x, y)$  el valor del nivel de gris normalizado en el píxel  $(x, y)$ .



(a)



(b)

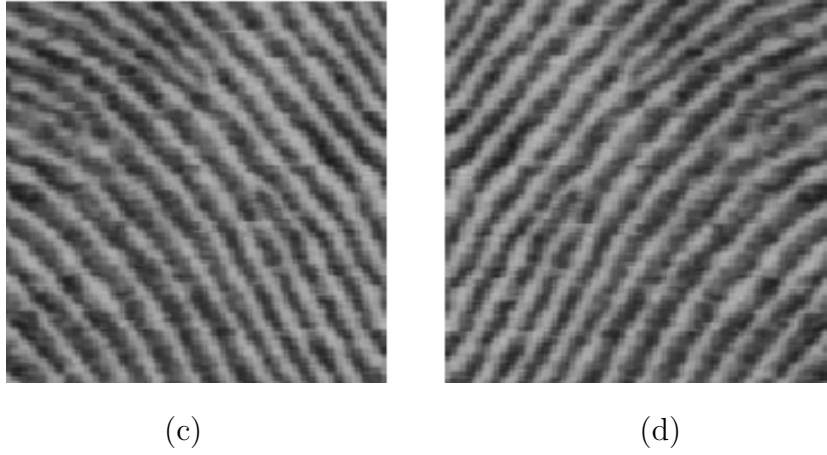


Figura 3.10: El espectro de Fourier de (a) y (b) se muestra en (c) y (d), respectivamente.

Para todos los pixeles en el sector  $S_i$ , la imagen normalizada se define como:

$$N_i(x, y) = \begin{cases} M_o + \sqrt{\frac{V_o \times (I(x, y) - M_i)^2}{V_i}} & \text{si } I(x, y) > M_i \\ M_o - \sqrt{\frac{V_o \times (I(x, y) - M_i)^2}{V_i}} & \text{Otro caso} \end{cases} \quad (3.20)$$

Donde  $M_o$  y  $V_o$  son los valores de la media y la varianza deseadas respectivamente, las operaciones de normalización del mejor píxel, la cual no cambia la claridad de las estructuras de surcos y crestas. La normalización se hace en toda la imagen y esta no puede compensar intensidades de variación en diferentes partes de ella debido a las diferencias de presión del dedo. Una normalización separada de cada sector individual resuelve el problema. La figura 3.12 muestra este esquema de normalización.

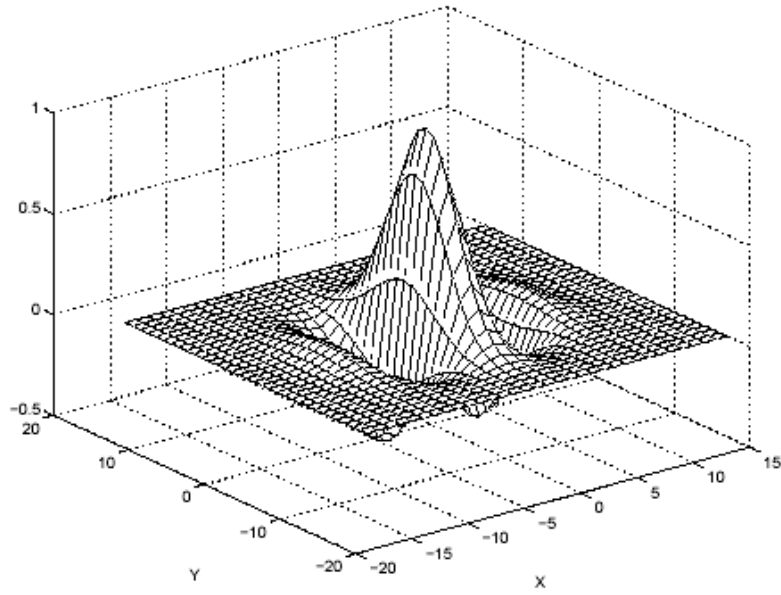
Un filtro de Gabor simétrico tiene la siguiente forma general en el dominio espacial:

$$G(x, y; f, \theta) = \exp \left\{ \frac{-1}{2} \left[ \frac{\acute{x}^2}{\delta_x^2} + \frac{\acute{y}^2}{\delta_y^2} \right] \right\} \cos(2\pi f \acute{x}) \quad (3.21)$$

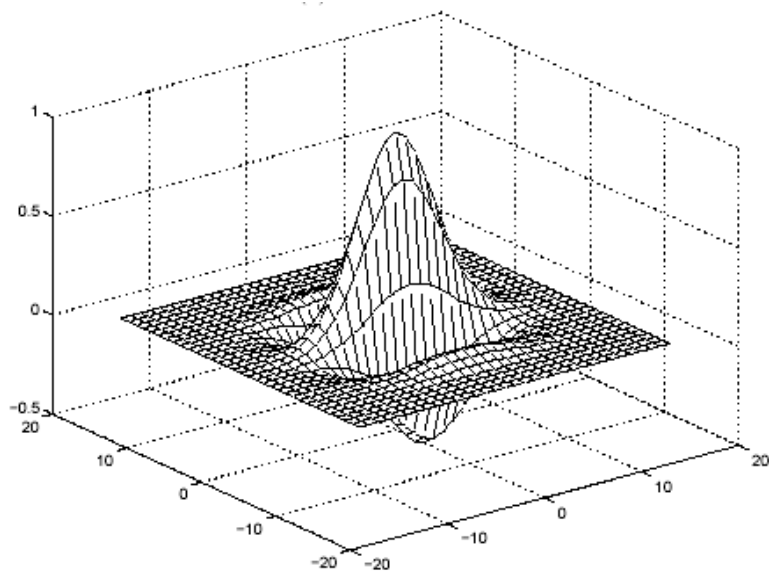
$$\acute{x} = x \sin \theta + y \cos \theta$$

$$\acute{y} = x \cos \theta - y \sin \theta$$

Donde  $f$  es la frecuencia sinusoidal ondular en dirección de  $\theta$  del eje  $x$ , y  $\delta_x$  y  $\delta_y$  son las constantes del espacio de la función gaussiana a lo largo del eje  $x$  e  $y$  respectivamente. Las características espaciales de los filtros de Gabor pueden verse en la Figura 3.11.



(a)

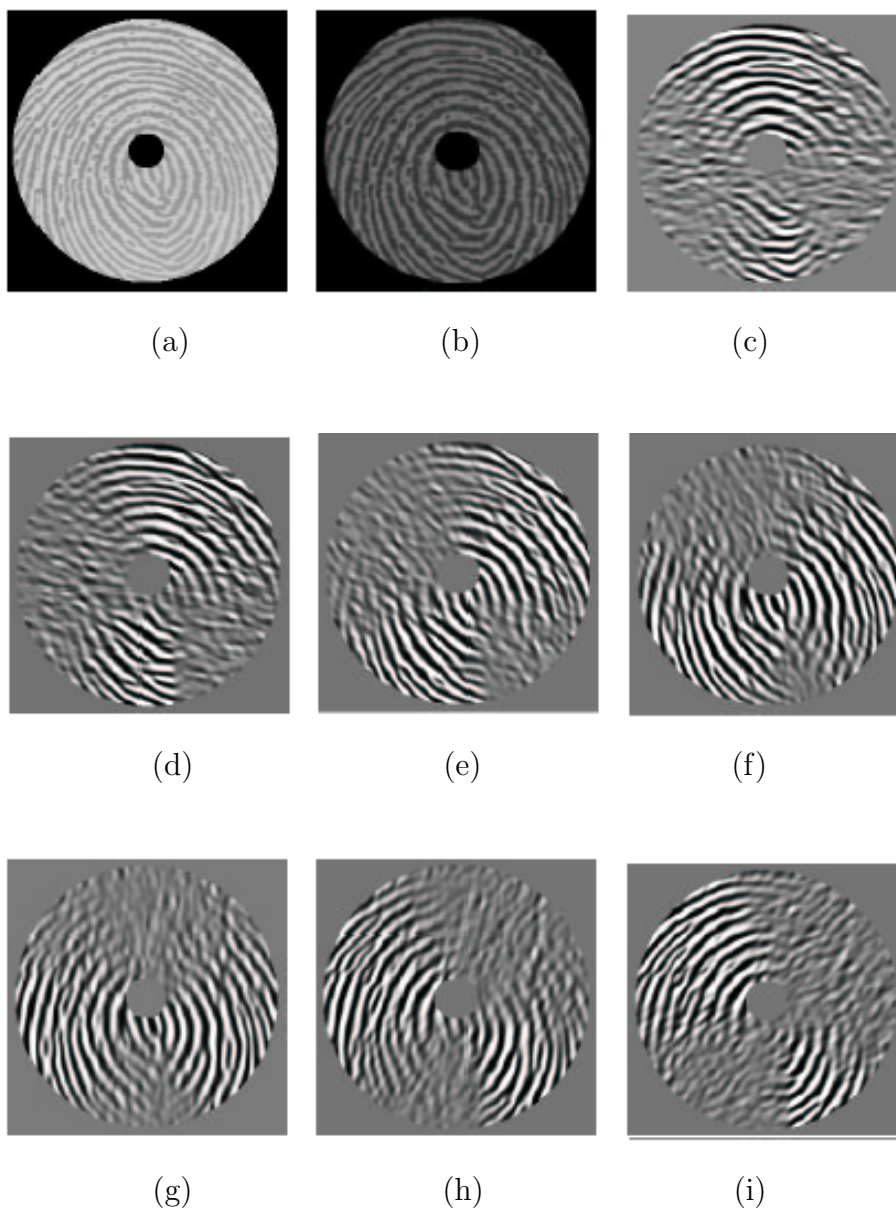


(b)

Figura 3.11: Filtros de Gabor (Tamaño de máscara de  $33 \times 33$ ,  $f = 0,1$ ,  $\delta_x = 4,0$ ,  $\delta_y = 4,0$ ). Solo se muestran aquí filtros orientados a  $0^\circ$  y  $90^\circ$

La figura 3.10 muestra que los valores del filtro fuera de esta máscara de  $33 \times 33$  tenderá a cero. Para acelerar el proceso de filtrado, convertimos un píxel solo con estos valores en la

máscara del filtro, cuyo valor absoluto sea mayor que 0.05. Esto acelera el proceso de convolución significativamente, mientras que la información contenida con valores pequeños, la convolución de la máscara del filtro no contribuye significativamente a la convolución global de entrenamiento. Hacemos uso de la simetría del filtro para acelerar la convolución [13, 33].



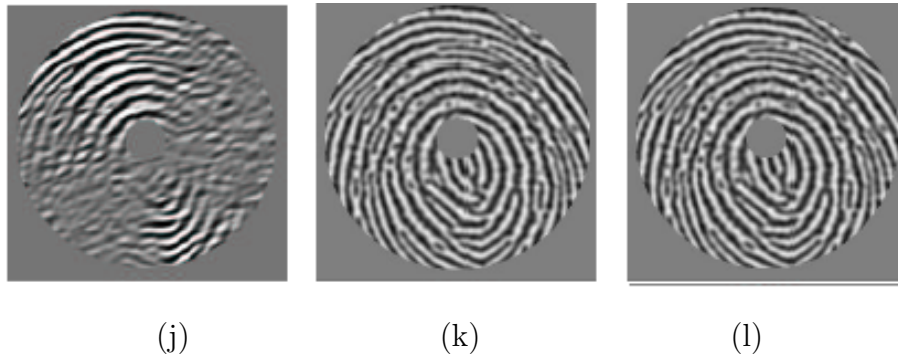


Figura 3.12: Normalizado, filtrado y reconstrucción de la imagen de la huella digital. (a) el área de interés, (b) imagen normalizada, (c)-(j)  $0^\circ, 22,5^\circ, 45^\circ, 90^\circ, 112,5^\circ, 157,5^\circ$  filtración de imágenes, respectivamente, (k) reconstrucción de la imagen con 4 filtros, y (l) reconstrucción de la imagen con 8 filtros. Mientras cuatro orientaciones del filtro son suficientes para capturar la estructura global de la huella digital, se requieren ocho orientaciones del filtro para capturar las características locales.

Sin embargo la convolución con filtros de Gabor es el proceso que mayor tiempo consume en la extracción de características pero aún es el mejor contribuyente.

En algunos experimentos se fija  $f$  a la frecuencia promedio de las crestas  $1/K$  donde  $K$  es la distancia entre crestas. La distancia promedio entre crestas es aproximadamente de 10 píxeles en una imagen de la huella digital a 500 dpi. Si  $f$  es demasiado grande, se crean falsas crestas en la imagen filtrada o si  $f$  es demasiado pequeña crestas cercanas son mezcladas en una. Las diferentes direcciones del filtro ( $\theta$ ) incluido a  $0^\circ, 22,5^\circ, 45^\circ, 67,5^\circ, 90^\circ, 112,5^\circ, 135^\circ$ , y  $157,5^\circ$  con respecto al eje  $x$ . La región de interés normalizada en una imagen de la huella digital sometida a la convolución, con cada uno de estos ocho filtros produce un conjunto de ocho imágenes filtradas. Una huella digital sometida a la convolución con un filtro orientado a  $0^\circ$  acentúa las crestas que son paralelas al eje  $x$  y suaviza las crestas en las otras direcciones. Los filtros afinados en las otras direcciones trabajan de manera similar. Estos ocho filtros direccionales sensibles capturan la mayoría de la información global direccional de las crestas, además de las características locales presentes de ellas en una huella digital. Ilustramos esto a través de la reconstru-

cción de la imagen de la huella sumando todas las ocho imágenes filtradas. La imagen reconstruida es similar a la imagen original sin pérdida de información significativa (Figura 3.12 (1)). Al capturar la información global y local, la exactitud de la comprobación se mejora aunque haya un poco de redundancia entre las ocho imágenes filtradas. Si  $\delta_x$  y  $\delta_y$  (desviaciones estándar de la envolvente Gaussiana) son también grandes, el filtro es más robusto a interferencias, pero es más probable suavizar la imagen para ampliar los detalles de crestas y surcos que se pierden en una huella digital. Si  $\delta_x$  y  $\delta_y$  son valores demasiado pequeños, el filtro no es eficiente removiendo el ruido (o interferencia). Los valores  $\delta_x$  y  $\delta_y$  son empíricamente determinados y cada uno se fija a 4.0 (alrededor de la mitad entre crestas).

### 3.5.1. Detección de crestas

Una característica importante de las crestas en la imagen de la huella digital es que los valores del nivel gris en crestas logran sus máximos locales a lo largo de una dirección normal de la orientación local de la cresta y estos pixeles son identificados por ser pixeles de la cresta basados en esta propiedad. En el algoritmo de detección de minucias, primero se convolucionan la imagen de la huella con dos máscaras,  $h_t(i, j, u, v)$  y  $h_b(i, j, u, v)$  de tamaño  $L \times H$  respectivamente. Estas dos máscaras son capaces de adaptarse acentuadamente al máximo local de los valores de nivel de gris a lo largo de una dirección de la orientación local de la cresta:

$$\begin{aligned}
 h_t(i, j, u, v) &= \begin{cases} -\frac{1}{\sqrt{2\pi}\delta} e^{-\frac{u}{\delta^2}} & \text{si } u = (v \cot(O(i, j)) - \frac{H}{2 \cos(O(i, j))}), v \in \Omega. \\ \frac{1}{\sqrt{2\pi}\delta} e^{-\frac{u}{\delta^2}} & \text{si } u = (v \cot(O(i, j))), v \in \Omega. \\ 0 & \text{Otro caso.} \end{cases} \\
 h_b(i, j, u, v) &= \begin{cases} -\frac{1}{\sqrt{2\pi}\delta} e^{-\frac{u}{\delta^2}} & \text{si } u = (v \cot(O(i, j)) + \frac{H}{2 \cos(O(i, j))}), v \in \Omega. \\ \frac{1}{\sqrt{2\pi}\delta} e^{-\frac{u}{\delta^2}} & \text{si } u = (v \cot(O(i, j))), v \in \Omega. \\ 0 & \text{Otro caso.} \end{cases} \\
 \Omega &= \left[ -\left| \frac{L \sin(O(i, j))}{2} \right|, \left| \frac{L \sin(O(i, j))}{2} \right| \right] \tag{3.22}
 \end{aligned}$$

donde  $O(i, j)$  representa la dirección de la cresta local en el píxel  $(i, j)$ . Si ambos valores del nivel de gris al píxel  $(i, j)$  de las imágenes convolucionadas son mayores que un cierto umbral  $T_{cresta}$ , entonces el píxel  $(i, j)$  se etiqueta como una cresta. Adaptando el ancho de la máscara al ancho de la cresta este algoritmo puede localizar eficientemente la cresta en una imagen de la huella digital. Sin embargo debido a la presencia de ruido, roturas y manchas, etc, en la imagen de entrada, el mapa binario de la cresta que resulta, contiene a menudo los agujeros y los puntos. Cuando los esqueletos de las crestas se usan para la detección de minucias, la presencia de tales agujeros y puntos (fragmentos falsos pequeños detectados como crestas) deterioran seriamente el funcionamiento del algoritmo de extracción de minucias porque estos agujeros y puntos pueden cambiar drásticamente el esqueleto de las crestas. Por lo tanto, un agujero y punto necesitan que se aplique un procedimiento de retiro antes del entrenamiento de la cresta.

### 3.5.2. Detección de minucias

La detección de la minucia es una tarea fácil cuando una región ideal de entrenamiento de crestas está disponible. Sin pérdida de generalidad, asumamos que, si un píxel está en la cresta obtenida después del entrenamiento ( $r$  asociaciones), entonces tiene un valor 0, y 1 en otro caso. Denotemos  $(x, y)$  el píxel en una cresta entrenada, y  $N_0, N_1, \dots, N_{r-1}$  denota sus  $r$  vecindades. Un píxel  $(x, y)$  es un final de cresta si  $(\sum_{i=0}^{r-1} N_i) = 1$  y una bifurcación de crestas si  $(\sum_{i=0}^{r-1} N_i) > 2$ . Sin embargo, la presencia no deseable de púas e interrupciones presentes en una región de entrenamiento de crestas puede llevarnos a la detección de muchas minucias falsas. Por consiguiente, antes de la detección de minucias, se aplica un procedimiento de suavizado para quitar las púas y unir las crestas rotas. El algoritmo de suavizado de crestas usando métodos heurísticos es:

- Si el ángulo formado por una rama y el tronco de la cresta es mayor que  $T_{inferior}$  y menor  $T_{superior}$  y la longitud de la rama es menor que  $T_{bifurcación}$ , entonces la rama es removida.
- Si una interrupción de una cresta es más corta que  $T_{interrupción}$  y ninguna otra cresta

pasa a travez de ella, entonces la interrupción se conecta.

Controlando el comportamiento de los parámetros de suavizado heurístico de la cresta se fijan los valores presentes más representativos. Aunque, es posible que el algoritmo de suavizado de crestas puede de vez en cuando anular la minucia genuina, más aun, anula las falsas minucias generadas por las imagenes de baja calidad, imagenes que procesan artefactos, y pliegues de la huella digital. Para cada minucia detectada, se guardan los siguientes parámetros: (i) la coordenada  $x$ , (ii) la coordenada  $y$ , (iii) orientación, la cual es definida como una orientación de la cresta local de la cresta asociada, y (iv) la segmentación de la cresta asociada. Las crestas grabadas son representadas como una señal discreta de una dimension la cual es normalizada por un parámetro de longitud predeterminado que es aproximadamente igual a la media de la distancia entre crestas del dedo. Se prueban por cada minucia aproximadamente 10 ubicaciones sobre la cresta asociada con cada cresta. La representación completa de un dedo cuando es almacenada y luego capturada en un formato comprimido tiene, un promedio, aproximado de 250 bytes. Estas crestas grabadas se usan para la alineación en la fase de emparejado de la minucia.

### 3.6. Vector de características

Es difícil confiar en características que son extraídas basándose principalmente en las características estructurales de la huella digital, sobre todo en imágenes de mala calidad. Las características de la imagen basadas en propiedades estadísticas son susceptibles a degradación con deterioración de la calidad de la imagen. Para este estudio, usamos características basados en la variación en la escala de grises. La desviación media absoluta de los niveles de gris del valor medio en un sector de la imagen indica la actividad de la cresta total en ese sector. Características similares fueron utilizadas exitosamente por Jain y Farrokhnia [14] para la clasificación y segmentación de la textura. Los resultados empíricos en aplicaciones de clasificación y verificación de la huella digital muestran que las características estadísticas simples actúan bien.



Consideremos  $F_{i\theta}(x, y)$  la imagen filtrada direccional de  $\theta$  para el sector  $S_i$ . Ahora  $\forall i \in \{1, 2 \dots 79\}$  y  $\theta \in \{0^\circ, 22,5^\circ, 45^\circ, 90^\circ, 112,5^\circ, 157,5^\circ\}$  el valor de la característica,  $v_{i\theta}$ , es la desviación media absoluta del promedio definido como:

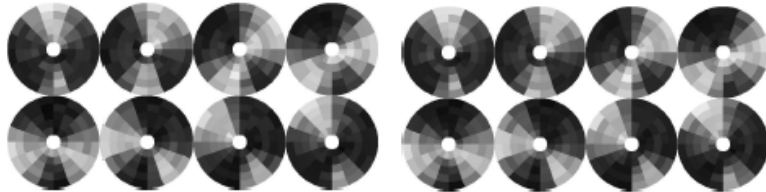
$$v_{i\theta} = \frac{1}{n_i} \left( \sum_{n_i} |F_{i\theta}(x, y) - P_{i\theta}| \right) \quad (3.23)$$

Donde  $n_i$  es el número de píxeles en  $S_i$  y  $P_{i\theta}$  es la media de los valores  $F_{i\theta}(x, y)$  en el sector  $S_i$ . La desviación media absoluta de cada sector en cada una de las ocho imágenes filtradas definen las componentes del vector de características. El número de imágenes filtradas definen las componentes del vector de características. El número de orientaciones requeridas del filtro fueron empíricamente determinadas. En la aplicación de verificación de la huella digital, usando ocho filtros orientados producían un mejor desempeño que cuando se usaron cuatro filtros de orientación (Resultados de [24]). Un incremento posterior en el número de filtros no proporcionó algún incremento en el desempeño de verificación. Similarmente, usando ocho filtros en vez de cuatro filtros no mejora el desempeño del algoritmo de clasificación de la huella digital [24].



(a)

(b)



(c)

(d)

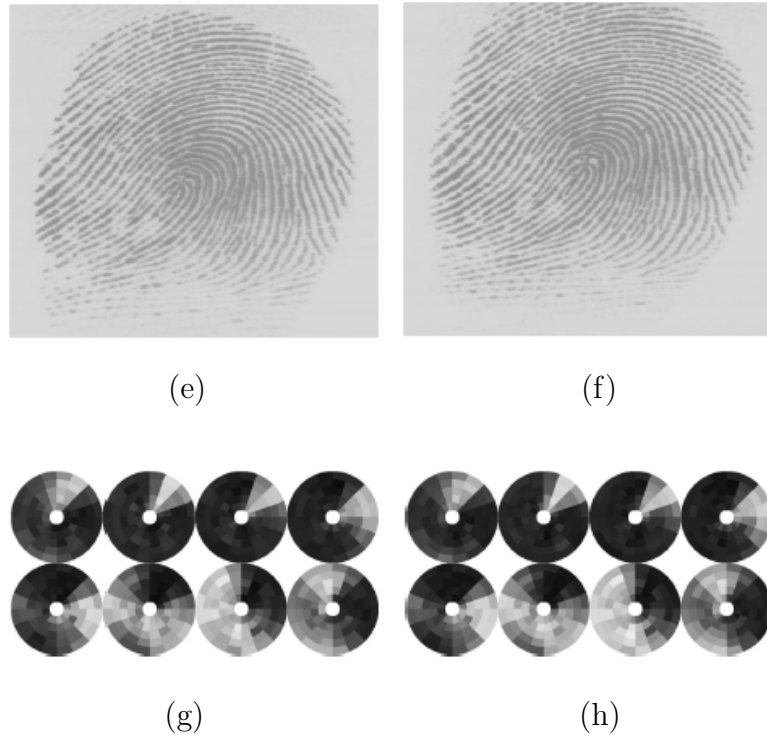


Figura 3.13: Ejemplos de vectores de características de dimensión 640. (a) Primera impresión del dedo 1, (b) Segunda impresión del dedo 1, (c) y (d) Códigos correspondientes de los dedos, (e) Primera impresión del dedo 2, (f) Segunda impresión del dedo 2, (g) y (h) Códigos correspondientes de los dedos.

Los vectores de características (FingerCodes) de dimensión 640 para imágenes de huellas digitales de dos dedos diferentes del banco de datos *MSU DBI* se muestra como imágenes niveladas de grises con ocho discos, cada disco corresponde a una imagen filtrada de la figura 3.13. El nivel gris en un sector en el disco representa el valor de las características para ese sector en la correspondiente imagen filtrada. Note que la figura 3.13(c) y (d) parece ser visualmente similar como son las figuras 3.13 (g) y (h), pero los correspondientes discos para dos dedos diferentes se ven también muy diferentes.

La traslación es dirigida para una sola localización del punto de referencia durante la fase de extracción de características. El esquema de representación puede tolerar la imprecisión en la estimación del punto de referencia de hasta 10 píxeles (aproximadamente una unidad de distancia entre crestas) lejos de su localización "verdadera". Se escoge una

fragmentación circular porque el tamaño del sector se incrementa a medida que se aleja del centro y maneja el error en la localización del mejor centro. La presente implementación de extracción de características asume que las huellas digitales están verticalmente orientadas (la yema de los dedos apuntan directamente hacia arriba). En realidad, las huellas digitales en la base de datos no están exactamente orientadas verticalmente; las huellas digitales pueden estar hasta  $\pm 45$  lejos de la orientación vertical asumida. La fragmentación circular ayuda obteniendo una representación que corresponda a una rotación cíclica de los valores del vector de características. Usamos esta rotación cíclica del vector de características para manejar parcialmente la rotación en la fase de emparejamiento.

### **3.6.1. Verificación de minucias**

Los datos de la imagen ofrecen una fuente rica de información para la extracción de características y para el emparejamiento. Por simplicidad del diseño del sistema de reconocimiento de patrones, una aproximación secuencial consiste en la detección, la extracción de características y el emparejamiento adoptado, donde cada uno de los estados transforma un componente particular de información de manera relativamente independiente. La interacción entre estos módulos es limitada a un camino en una sola dirección del flujo de control. Algunos de los errores de extremo a extremo en el proceso secuencial se pueden eliminar fácilmente en la fase de extracción de características volviendo a revisar los datos de la imagen original. Es conveniente adoptar un camino de retroalimentación en la fase de extracción de características, seguida por una fase de refinamiento de características para mejorar el desempeño del emparejamiento. Esta mejora del desempeño se ilustra en el contexto de un sistema de verificación de huellas digitales basado en minucias.

La mayoría de los sistemas automáticos de verificación de huellas digitales primero detectan las minucias en la imagen de la huella digital y después comparan el conjunto de minucias de entrada (extraídas de la imagen) con una plantilla almacenada (Base de datos). El algoritmo que describimos en la sección 3 usa una aproximación secuencial para la extracción de características. La extracción de características primero transforma

a binario las crestas de la imagen de la huella digital usando máscaras que son capaces de acentuar adaptativamente los valores máximos de la escala de grises a lo largo de una dirección normal a la dirección de la cresta local. Las minucias son determinadas como puntos que tienen: un único vecino (final de cresta) o más de dos vecinos (bifurcación de cresta) en el esquema de la imagen ver la figura 3.14. Sin embargo, la estimación de la orientación en una imagen de mala calidad no es muy confiable, resultando en la detección muchas minucias falsas ver la figura 3.15. Muchos investigadores han propuesto recortar las minucias en la fase de post-proceso para eliminar las minucias falsas pero esta fase se basa en técnicas algo heurísticas.

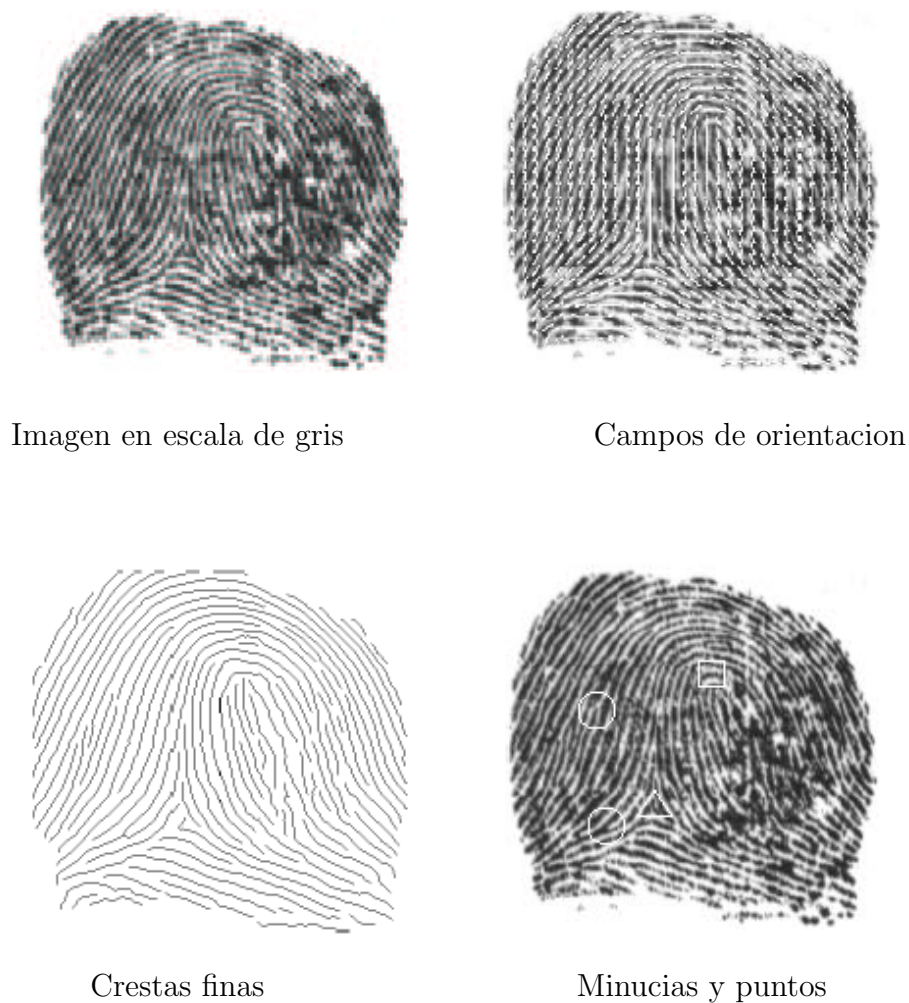


Figura 3.14: Campos de orientacion, crestas finas, minucias y puntos singulares.

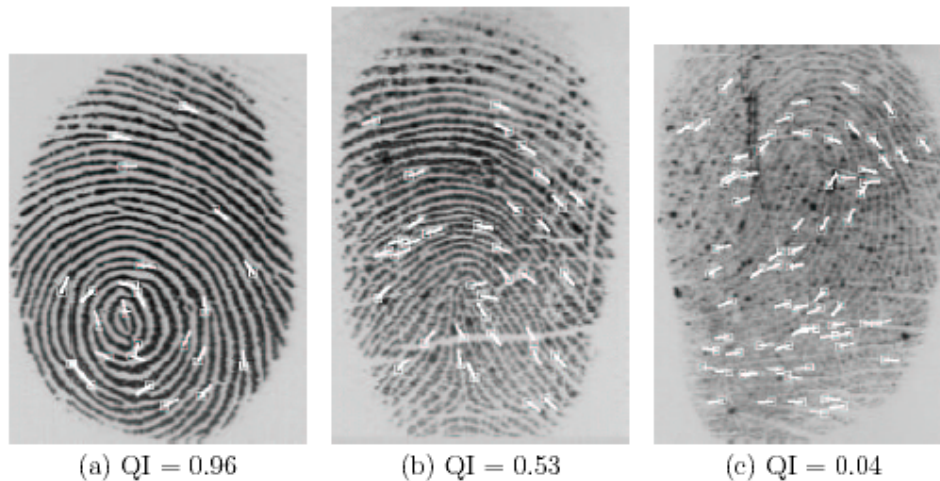


Figura 3.15: Imágenes de muestra de la base datos GT con variación en el índice de calidad (QI). no se encontraron minucias falsas en (a), 7 en (b) y 27 en (c) por el algoritmo de detección automática de minucias.

Un algoritmo de verificación de minucias se puede dividir en tres etapas que son:

- Extracción de características.
- Entrenamiento (Aprendizaje de las características de las minucias).
- Prueba o verificación.

### **Extracción de características**

Aquí se utiliza el algoritmo que se describió en la sección 3 desarrollado en [24].

### **Entrenamiento**

En la fase de entrenamiento los vectores de características de minucias y no minucias son enviados a un proceso de aprendizaje de cuantificación vectorial para que aprenda de las características en las regiones de las minucias y las falsas minucias. Para la fase de entrenamiento, se necesita establecer puntos verdaderos para las minucias y las falsas minucias en un gran número de huellas digitales. Así, se usa una base de datos que

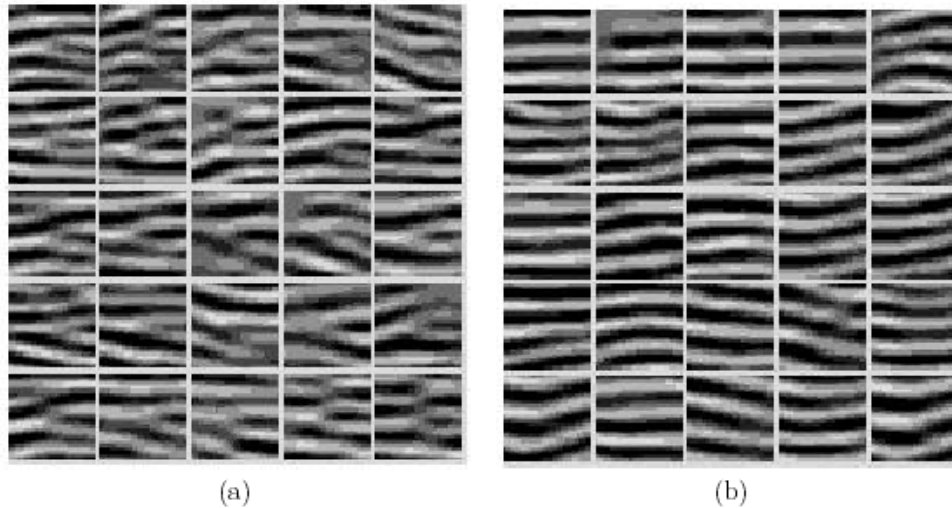


Figura 3.16: Ejemplos de los perfiles de los niveles de gris en la vecindad de (a)minucias y (b) falsas minucias. Estas subimágenes de  $32 \times 32$ , graduadas a 8 niveles de gris son usadas para el entrenamiento.

contenga una buena cantidad de imágenes de huellas digitales de dedos diferentes y tenga establecida la información verdadera acerca de las minucias provistas por un experto en huellas digitales (ver figura 3.17). Es deseable que las impresiones múltiples para cada dedo en la base de datos sean tomadas en tiempos diferentes. Si las imágenes son de diferentes tamaños se deben escanear a una resolución de 500 dpi con 256 niveles de grises. Para el entrenamiento se divide la base de datos en dos conjuntos uno para entrenamiento y otro para prueba.

NOTA: Es de notar que los vectores de características pueden ser de gran tamaño, por ejemplo en los experimentos realizados en [24] se utilizaron cada uno de tamaño 1024. Para las minucias verdaderas, se usa la dirección de la minucia provista por el experto. Para las muestras negativas se calcula la dirección del bloque de  $32 \times 32$  usando el algoritmo jerárquico de campo de orientación que se describió en la sección 3. Ver figura 3.16 para ejemplos de perfiles de minucias y no minucias en la escala de grises.

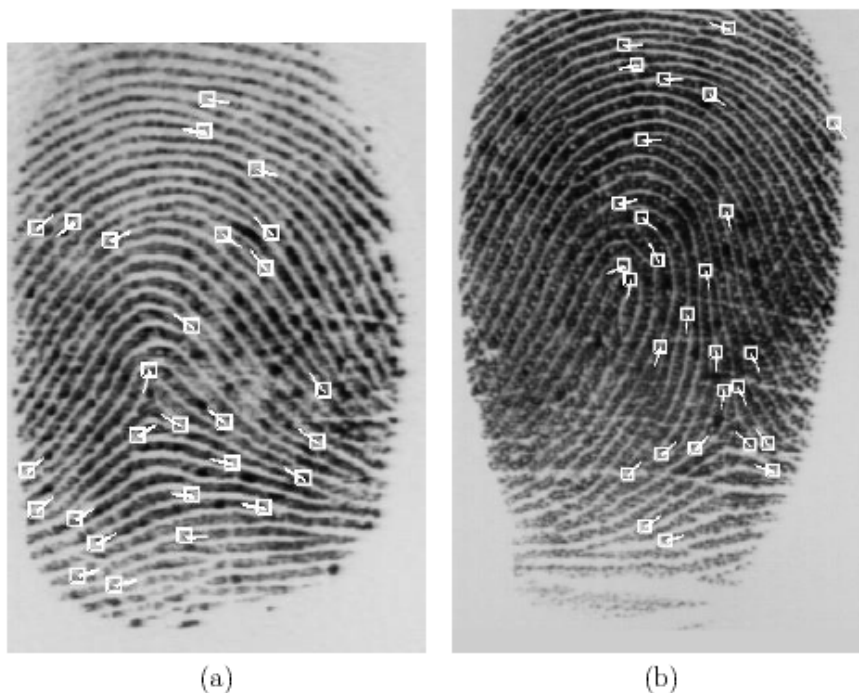


Figura 3.17: Ejemplos de imágenes de la base de datos GT. Las minucias establecidas como verdaderas provistas por un experto son marcadas en la imagen

## Prueba

Para la prueba simplemente se utiliza cualquiera de los clasificadores existentes y según los resultados arrojados se evalúa el desempeño del mismo.

## 3.7. Emparejamiento

La capacidad que tiene una huella de distinguirse puede estar determinada por el modelo global de crestas y surcos así como las anomalías de las crestas locales (los puntos de minucia). Aunque las crestas poseen la información discriminadora, el diseño de un algoritmo para emparejar huellas digitales, fiable, es muy discutido debido a la deformación no lineal y borrosa en las imágenes de huellas digitales. (Ver Figura 3.18).

Las técnicas populares existentes para emparejar huellas digitales pueden ser clasificadas

ampliamente en dos categorías: (a) basadas en minucias y (b) basadas en la correlación. La primera categoría empareja los dos conjuntos de minucias de las dos huellas digitales, alineandolos y luego contando el número de minucias que emparejan. Una técnica típica para la extracción de minucias realiza la siguiente secuencia de operaciones sobre la imagen de la huella: (i) Mejora la imagen de la huella digital, (ii) convierte a binario (segmentación en las crestas y surcos), (iii) filtrado, y (iv) detección de las minucias. Varios algoritmos comerciales y académicos siguen estos pasos secuenciales como son las técnicas alternativas, que operan directamente sobre la escala de grises en las crestas, y estas sobre las imágenes de la huella digital. La alineación entre la huella digital de entrada y la plantilla puede obtenerse usando una o más características de la huella digital. Por ejemplo, una alineación puede lograrse basada en el campo de orientación de las huellas digitales, la ubicación de puntos singulares tales como el centro y el delta, crestas, emparejamiento inexacto de las minucias, la transformada de Hough [6], patrones de puntos, etc. El número de minucias emparejadas en cierta tolerancia se normaliza generalmente con el número total de estas en los dos conjuntos a considerar que son las halladas erróneamente y las perdidas durante la extracción de características. Una de las principales dificultades en la primera categoría es que es muy difícil extraer minucias fiables en imágenes de huellas digitales de mala calidad. Se pueden usar varias técnicas para mejorar la calidad de la imagen de la huella digital previo a la extracción de las minucias.

Las técnicas basadas en la correlación emparejan el modelo global de crestas y surcos para ver si las crestas se alinean. La técnica más simple es alinear las dos imágenes de la huella digital y substrair la entrada de la plantilla para ver si las crestas corresponden. Sin embargo, tal acercamiento simplista sufre de muchos problemas incluyendo los errores en la estimación de alineación, deformación no-lineal en las imágenes de la huella digital, y ruido. Sibbald [1] ha propuesto una técnica de auto correlación que estima la correlación entre la entrada y la plantilla en los incrementos fijos de traslación y rotación. Si la correlación excede un cierto umbral, se dice que las dos huellas provienen del mismo dedo. Una variante de la técnica de correlación es realizar la correlación en el dominio de frecuencia en lugar del dominio espacial realizando una transformada rápida de Fourier



bidimensional (FFT) tanto en la entrada como en la plantilla de la huella digital. Se compara la suma de la multiplicación píxel-a-píxel de las dos representaciones de dominio de frecuencia de las imágenes de la huella digital con el umbral para tomar una decisión. Una de las ventajas de realizar la correlación con el dominio de frecuencia es que las representaciones de frecuencia de las huellas digitales son traslaciones invariantes. Sin embargo una de las desventajas, es el tiempo de cómputo extra requerido para convertir la imagen espacial en una representación de frecuencia. El emparejamiento de las frecuencias del dominio de la correlación también puede realizarse ópticamente. La entrada y la plantilla de la huella digital son proyectadas vía luz láser a través de un lente para producir su transformada de Fourier y su superposición conduce a un máximo de correlación cuya magnitud es alta para el par que coinciden y bajo en otro caso. La ventaja de realizar la correlación óptica es la velocidad; la desventaja es que el procesador óptico tiene la versatilidad muy limitada (programabilidad). Una modificación en la técnica basada en la correlación espacial es dividir la imagen de la huella digital en cuadrículas y determinar la correlación en cada sector en lugar de la imagen completa. La técnica basada en la correlación supera alguna de las limitaciones de la basada en las minucias. Por ejemplo, el algoritmo para la extracción de minucias detecta un número grande de minucias falsas e ignora las minucias genuinas, en imágenes de huellas digitales muy ruidosas. Las técnicas basadas en la correlación son menos sensibles a imágenes de huellas digitales ruidosas pero tienen sus propios problemas. Por ejemplo, las técnicas basadas en la correlación son más sensibles a un error en la estimación de la alineación entre dos huellas digitales. También, no pueden tratarse fácilmente con la deformación no lineal presente en las imágenes de las huellas digitales. Adicionalmente, tienen el tamaño de la plantilla más grande.

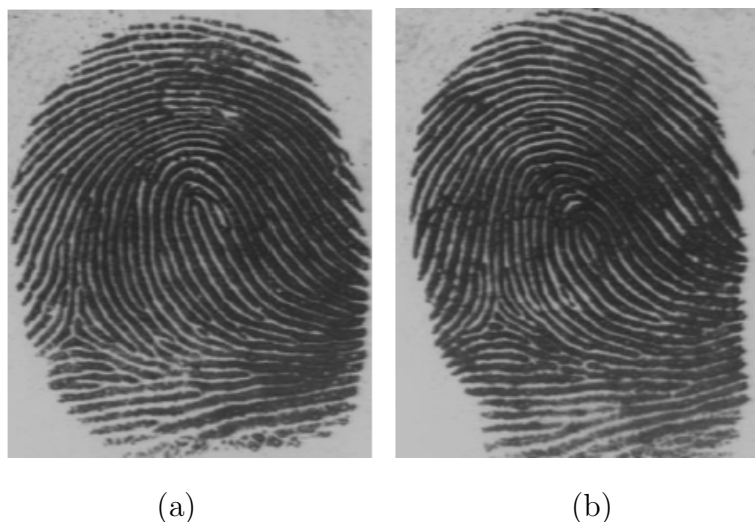


Figura 3.18: Dificultad en el emparejando de la huella digital. Las figuras (a) y (b) tiene la misma configuración global pero son imágenes de dos dedos diferentes.

La representación basada en el banco de filtros no cae en ninguna de las categorías ya sea las basadas en las minucias o las basadas en el emparejamiento por correlación. La técnica propuesta es una técnica basada en características que calculan los detalles el local y el global en una huella digital como un vector de características de longitud fijo (FingerCode). El emparejamiento de la huella digital esta basado en la distancia euclídea entre los dos fingercodes y así pues, es extremadamente rápido. Es capaz de alcanzar una calidad de verificación superior a los resultados de los algoritmos basados en las minucias más avanzados que hay actualmente. El rendimiento del emparejamiento puede mejorarse combinando las decisiones de los emparejadores basados en información complementaria (basado en minucias y basado en filtros) de la huella digital.

Es deseable explorar esquemas de representación que combinen la información global y local en una huella digital. Una nueva representación de código de longitud fija, relativamente corta, para la huella digital, llamada Fingercode es conveniente para emparejar así como para el almacenamiento en una tarjeta inteligente. El emparejamiento se reduce a encontrar la distancia Euclidea entre éstos FingerCodes de ahí que es muy rápido y la representación es fácil para indexarla.

El emparejamiento de huellas digitales esta basado en encontrar las distancias entre los fingercodes correspondientes. La traslación invariante en el fingercode se establece identificando el punto de referencia. Sin embargo los fingercodes no son invariantes por rotación. La invarianza por rotación aproximada se logra rotando las características cíclicamente en el mismo FingerCode. Un paso único para la rotación cíclica de las características en el FingerCode descrito por las ecuaciones (3.24)-(3.26) corresponde a un vector de características el cual podría ser obtenido si la imagen se rotara  $22,5^\circ$ . Una rotación por  $R$  pasos corresponde a una rotación de la imagen de  $R \times 22,5^\circ$ . Una rotación positiva implica que la rotación es en sentido contrario a las agujas del reloj mientras que una rotación negativa implica que la rotación es en el sentido de las agujas del reloj. Ver la figura 3.19 para una ilustración. El FingerCode obtenido después de  $R$  pasos de rotación esta dado por

$$V_{i\theta}^R = V_{i'\theta'}, \quad (3.24)$$

$$i' = (i + k - R) \text{ mod } k + (i \text{ div } k) \times k, \quad (3.25)$$

$$\theta' = (\theta + 180^\circ + 22,5^\circ \times (-R)) \text{ mod } 180^\circ \quad (3.26)$$

Donde  $V_{i\theta}^R$  es el fingercode rotado,  $V_{i'\theta'}$  es el fingercode original,  $k(= 16)$  es el número de sectores en una banda,  $i \in [0, 1, \dots, 79]$ , y  $\theta \in [0^\circ, 22,5^\circ, 45^\circ, 67,5^\circ, 90^\circ, 112,5^\circ, 135^\circ, 157,5^\circ]$ .

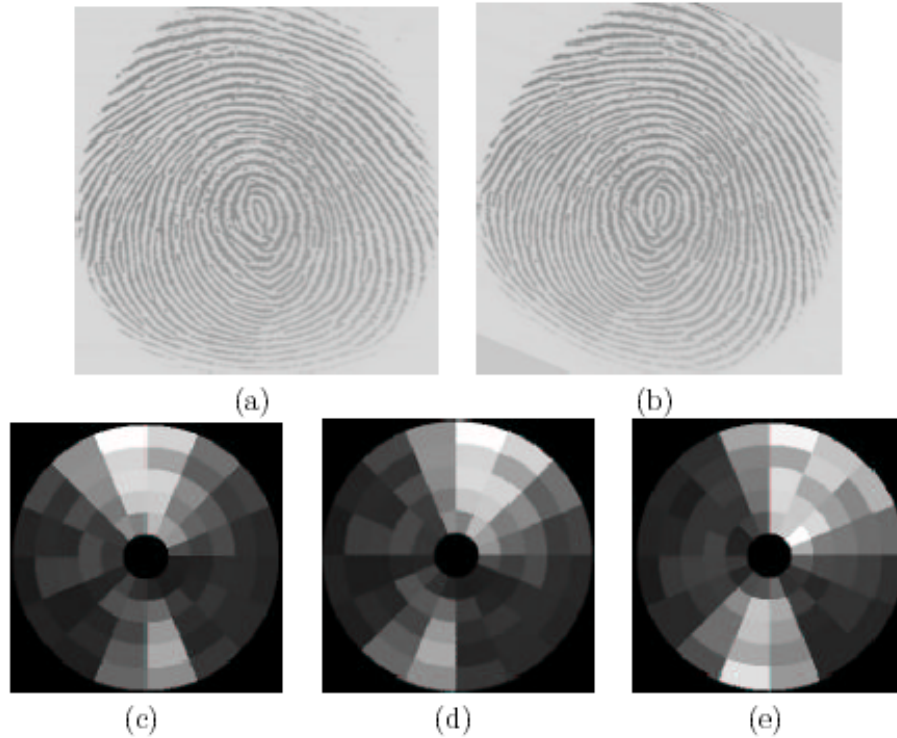


Figura 3.19: La imagen de la huella digital en (b) se obtuvo por una rotación de  $-22,5^\circ$  de (a). Una parte del vector de características correspondiente al filtro de Gabor de  $0^\circ$  de la imagen extraída de (a) se muestra en (c) como una imagen de escala de gris. El vector de características en (c) se rotó  $-22,5$  ( $R = -1$  en las ecuaciones (3.24)-(3.26)) y se muestra en (d). (e) muestra el vector de características extraído de la imagen de la huella en (b). Los vectores de características mostrados en (d) y en (e) son similares ilustrando que el vector de características para una rotación de  $-22,5^\circ$  en la imagen original aproximadamente corresponde a una unidad en rotación cíclica antihoraria del vector de características.

Para cada huella digital en la base de datos, se guardo cinco plantillas correspondientes a las siguientes cinco rotaciones del correspondiente fingerprintcode:  $V_{i\theta}^{-2}$ ,  $V_{i\theta}^{-1}$ ,  $V_{i\theta}^0$ ,  $V_{i\theta}^1$ ,  $V_{i\theta}^2$  usando sólo cinco valores de parámetros  $R$  ( $-2, -1, 0, 1, 2$ ) si la imagen de la huella digital en las dos bases de datos no tienen más que  $\pm 45^\circ$  de rotación. Para las bases de datos que tienen más rotación en la imagen de la huella digital se puede usar un rango más alto para el parámetro  $R$ . La entrada del Fingerprintcode se empareja con las cinco plantillas

guardadas en la base de datos para obtener cinco diferentes resultados de emparejamiento. El mínimo de estos cinco valores corresponde a la mejor alineación de la huella digital entrante con las huellas digitales de la base de datos. Puesto que una única rotación cíclica de las características en el Fingerprint corresponde a una rotación de  $22,5^\circ$  en la imagen original, sólo podemos generar esas representaciones de la huella digital, las cuales son múltiplos de  $22,5^\circ$ . Debido a la naturaleza de la fragmentación, las características son invariantes sólo a pequeñas perturbaciones que son menos de  $\pm 11,25^\circ$ . Por lo tanto, se genera otro vector de características para cada huella digital en el instante en que el usuario coloque su huella, la imagen original es rotada un ángulo de  $11,25^\circ$  y se genera su Fingerprint. También se guardan en la base de datos cinco plantillas correspondientes a diferentes rotaciones de este Fingerprint. Así la base de datos contiene diez plantillas para cada huella digital. Esas diez plantillas corresponden a todas las rotaciones de la imagen de la huella digital en múltiplos de  $11,25^\circ$ . Con esto se tiene cuidado de la rotación de la huella digital mientras se compara el Fingerprint entrante con las plantillas guardadas. Se toma el último valor de la distancia del emparejamiento como la mínima de los diez valores obtenidos emparejando el Fingerprint entrante con cada una de las diez plantillas. Este valor mínimo corresponde a la mejor alineación de las dos huellas digitales que están siendo emparejadas. Ya que la generación de la plantilla para el almacenamiento en la base de datos se realiza mediante un proceso fuera de línea y el proceso de emparejamiento es extremadamente rápido, el tiempo de verificación aún depende del tiempo empleado para generar una única plantilla para la imagen de prueba.

A continuación se da una descripción de algunos algoritmos para el emparejamiento de huellas digitales que pueden ser ampliamente clasificados en dos categorías: (i) Basado en minucias, y (ii) basado en filtros.

### 3.7.1. Emparejamiento basado en la transformación de Hough (algoritmo de Hough)

El problema de emparejamiento de huellas digitales puede ser considerado como emparejamiento de plantillas: dados dos conjuntos de características de minucias, calcular su cuenta de emparejamiento. Los dos pasos principales del algoritmo son: (i) calcular los parámetros de la transformación  $\delta_x$ ,  $\delta_y$ ,  $\theta$  y  $s$  en las dos imágenes, donde  $\delta_x$  y  $\delta_y$  son traslaciones que pertenecen a las direcciones  $x$  e  $y$ , respectivamente,  $\theta$  es el ángulo de rotación y  $s$  es el factor escalar; (ii) Alinear los dos conjuntos de puntos de minucias con los parámetros estimados y contar las parejas emparejadas dentro de una región acotada; (iii) repetir los dos pasos previos para el rango de transformaciones permitidas. Se supone que la transformación que resulte con la cuenta de emparejamiento más alta es la correcta. La cuenta de emparejamiento final esta entre una escala de 0 y 9.

### 3.7.2. Emparejamiento basado en la distancia entre cadenas (algoritmo cadena)

Cada conjunto de extracción de características de minucias primero se convierte en coordenadas polares con respecto aun punto fijo. Las características de las minucias bidimensionales ( $2D$ ), son reducidas a una cadena unidimensional ( $1D$ ) concatenando puntos en un orden creciente de ángulos radiales en coordenadas polares. El algoritmo de emparejamiento cadena se aplica para calcular la distancia a corregir entre las dos cadenas. La distancia a corregir se normaliza fácilmente y se convierte en una cuenta de emparejamiento. Este algoritmo se puede resumir de la siguiente manera: (i) se estiman la rotación y la traslación por el emparejamiento de segmentos de crestas (representadas como curvas planas) asociadas con cada minucia en la imagen de entrada y de la plantilla. Los parámetros de rotación y traslación que resultan en el número máximo de parejas de minucias emparejadas dentro de una región acotado se usan para definir la transformación estimada y las minucias correspondientes se etiquetan como el ancho de las minucias,  $A_1$  y  $A_2$ , respectivamente. (ii) Convertir cada conjunto de minucias en una cadena  $1D$

usando coordenadas polares fijas de,  $A1$  y  $A2$ , respectivamente; (iii) calcular la distancia a corregir entre las dos cadenas  $1D$ . Las parejas de minucias emparejadas se recuperan basándose en la corrección de la distancia mínima entre las dos cadenas; (iv) salida de la cuenta de emparejamiento normalizada (en el rango de 0-99) la cual es el radio del número de parejas emparejadas y el número de puntos de minucia.

### **3.7.3. Emparejamiento basado en el banco de filtros (algoritmo filtro)**

Los cuatro principales pasos del algoritmo de extracción de características basado en filtro son (ver Figura 3.20) (i) Determinar un punto de referencia y la región de interés para la imagen de la huella digital. El punto referencia se toma como el punto central de la huella el cual se define como el punto de máxima curvatura de las crestas en una huella digital. La región de interés es un área circular alrededor del punto de referencia. El algoritmo rechaza imágenes para las cuales el punto de referencia no puede ser establecido. (ii) Fragmentar la región de interés. La región de interés se divide en sectores y los valores de gris en cada sector son normalizados a unas constantes media y varianza (iii) Filtrar la región de interés en ocho direcciones diferentes usando un banco de filtros de Gabor. (Se requieren ocho direcciones para capturar completamente las características locales de las crestas en una huella digital mientras se necesitan solo cuatro direcciones para capturar la configuración global). El filtrado produce un conjunto de ocho imágenes filtradas. (iv) Calcular la desviación absoluta promedio a partir de la media (AAD) de valores de gris en sectores individuales en cada imagen filtrada. El valor AAD en cada sector cuantifica las estructuras fundamentales de crestas y se define como una característica. En un vector de características al cual se le llama FingerCod, está la colección de todas las características (para cada sector) de cada imagen filtrada. Así los elementos característicos capturan la información local y la enumeración ordenada de la fragmentación captura la relación global invariante entre los patrones de minucias. La representación es invariante a traslaciones de la imagen. Se asume que la huella digital se captura en una posición vertical y la invarianza de la rotación se logra

por el almacenamiento de diez representaciones correspondientes a varias rotaciones de la imagen ( $-56,25^\circ, -45^\circ, -33,75^\circ, -22,5^\circ, -11,25^\circ, 0^\circ, 11,25^\circ, 22,5^\circ, 33,75^\circ, 45^\circ$ ). Se calcula la distancia euclídea entre la representación de entrada y los diez templates para generar diez distancias de emparejamiento. Finalmente se calcula la mínima de las diez distancias y se invierte para conseguir una cantidad de emparejamiento. La cantidad de emparejamiento se mide entre 0 y 99 y se puede considerar como un valor seguro del emparejador.

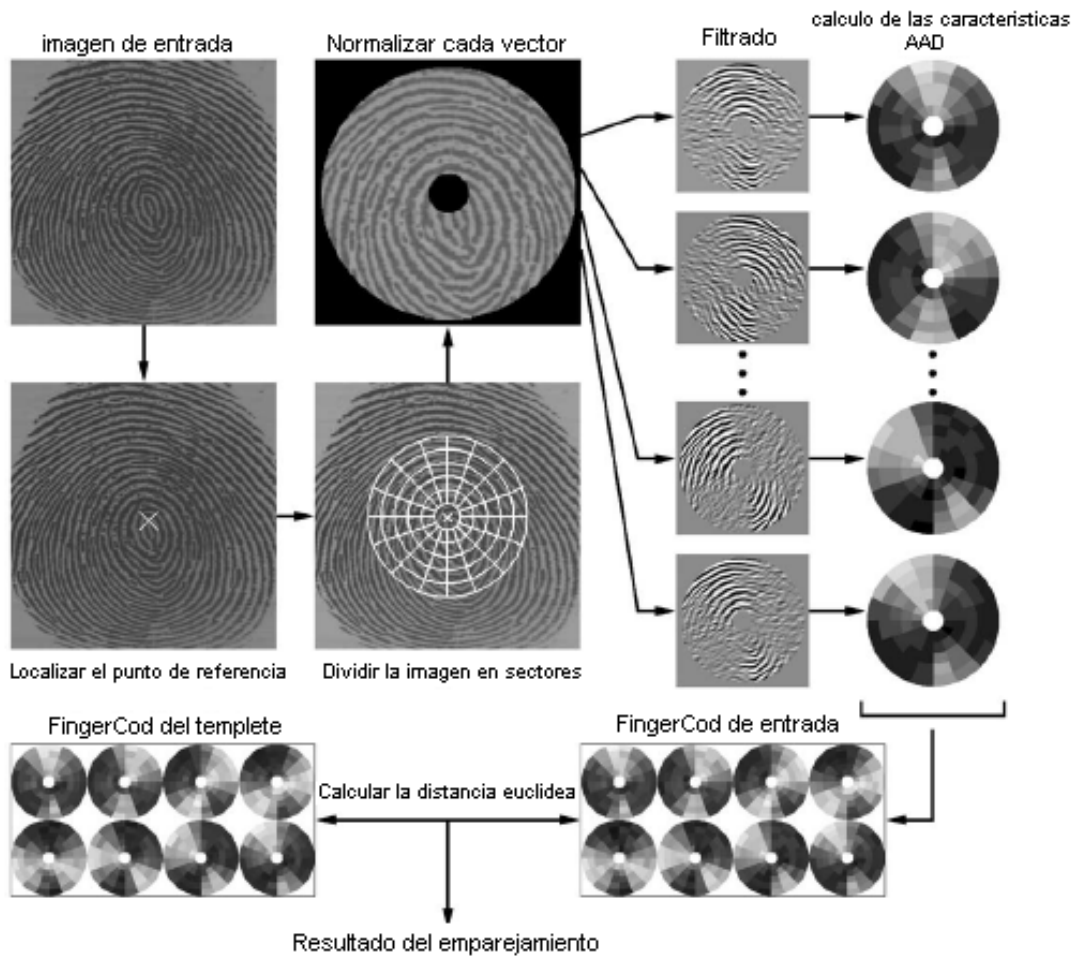


Figura 3.20: Diagrama del algoritmo de emparejamiento y extracción de características basado en el banco de filtro.



## 3.8. Estrategias de integración

Supongamos que el patrón de prueba  $Z$  será asignado a una de las dos posibles clases,  $w_0$  y  $w_1$ . Asumamos que tenemos  $N$  clasificadores, y el  $i$ -ésimo clasificador nos entrega un único valor de confianza  $i$  sobre  $w_i$  (la confianza para clase  $w_0$  será  $1 - i$ ),  $i = 1, \dots, N$ . También asumamos que las probabilidades anteriores para las dos clases son iguales. La combinación de las tareas de los clasificadores pueden ser propuestas ahora como el problema del diseño de un clasificador independiente (de los  $N$  diseños de clasificadores originales) con 2 clases y  $N$  características.

### 3.8.1. Selección de emparejadores

Es una práctica común en la combinación de clasificadores realizar un análisis extensivo de varias estrategias de combinación que envuelven los  $N$  clasificadores disponibles. En la selección de características, es bien conocido que el subconjunto de  $d$  elementos que más información da,  $N$  subconjuntos de características condicionalmente independientes, no es necesariamente la unión de las  $d$  características individuales que más información proveen. Cover [7] sostiene que ningún procedimiento de selección de  $d$  elementos secuencial no exhaustivo es óptimo, incluso para las características conjuntamente normales. Él mostró además que todas las posibles probabilidades de clasificaciones erróneas pueden ocurrir entre los subconjuntos de características sujeto a restricciones monótonas. La dependencia estadística entre las características causa además incertidumbre en los subconjuntos de  $d$  elementos compuestos por las mejores características individuales. Se podría argumentar que la propia estrategia de combinación debe escoger un subconjunto de los clasificadores que deben estar combinados. Sin embargo, se sabe por la práctica que el curso de la “dimensionalidad” hace difícil para un clasificador automático eliminar las características menos discriminantes. Por consiguiente, en [24] se propone un esquema de selección de clasificador antes de la combinación del clasificador. También se propone usar la separación estadística de clases [19] como el criterio de efectividad de características. Este estadístico,  $CS$ , mide que tanto están separadas las dos clases (errónea y genuina,

en este caso) con respecto al vector de características,  $X^d$ , en un espacio  $d$ -dimensional,  $R^d$ .

$$CS(X^d) = \int_{R^d} |p(X^d|w_0) - p(X^d|w_1)| dx \quad (3.27)$$

donde  $p(X^d|w_0)$  y  $p(X^d|w_1)$  son las distribuciones estimadas para las clases  $w_0$  (errónea) y  $w_1$  (genuina) respectivamente, note que  $0 \leq CS \leq 2$ .

En [24] se usa la separación de clases estadística para obtener el mejor subconjunto de emparejadores usando una búsqueda exhaustiva de todos los posibles  $2^N - 1$  subconjuntos de emparejadores.

### 3.8.2. Estimación de Densidad no-paramétrica

Una vez seleccionado el subconjunto de  $d$  clasificadores ( $d \leq N$ ), en [24] se desarrolló una estrategia de combinación. No se hace ninguna suposición sobre la forma de las distribuciones para las dos clases y se usa métodos no-paramétricos para estimar las dos distribuciones (errónea y genuina).

La densidad de Parzen calcula una función de densidad  $d$ -dimensional basada en  $n$  observaciones independientes e idénticamente distribuidas (muestras de entrenamiento) y el núcleo Gaussiano está dado por:

$$P(X) = \frac{1}{nh^d} \sum_{j=1}^n \left\{ \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left[ -\frac{1}{2h^2} (X - X_j)^t \Sigma^{-1} (X - X_j) \right] \right\} \quad (3.28)$$

donde  $h$  es el ancho de la ventana,  $\Sigma$  la matriz de covarianza del núcleo estimado a partir de las  $n$  muestras de entrenamiento y  $h \cong n^{-\frac{1}{d}}$ . El valor de  $h$  se determina usualmente de manera empírica. Un valor grande de  $h$  significa un grado alto de suavidad y un valor pequeño de  $h$  significa un grado bajo de suavidad de la densidad estimada. Una regla del pulgar dice que para un número pequeño (grande) de muestras de entrenamiento ( $n$ ), el ancho de la ventana debe ser grande (pequeño). Además, para un arreglo  $n$ , el ancho

de la ventana debe ser grande (pequeño) para un número de características ( $d$ ) grande (pequeño). Cuando un número grande de muestras de entrenamiento está disponible, la densidad estimada usando aproximación por la ventana Parzen se aproxima mucho a la densidad verdadera.

## Capítulo 4

# Máxima entropía

El concepto de la máxima entropía tiene una larga historia y se puede remontar al pasado a lo largo de múltiples enlaces bíblicos y en las escrituras de Herodotus (Jaynes 1990). Laplace puede ser considerado el padre de la máxima entropía, después de haber enunciado el tema subyacente 200 años atrás en su “Principio de Razón insuficiente”: cuando no se tiene información para distinguir entre la probabilidad de dos eventos, la mejor estrategia es considerarlos igualmente probables (Guisu and Shenitzer 1994). E.T. Jaynes que es un pionero más reciente de la máxima entropía también postuló (Jaynes 1990):

*“...el hecho de que una cierta distribución de probabilidad maximice la entropía sujeto a ciertas limitaciones que representan nuestra información incompleta, es la propiedad fundamental que justifica el uso de esa distribución para la inferencia; está de acuerdo con todo lo que se conoce, pero evita cuidadosamente asumir algo que no se conoce. Es una transcripción de la matemática de un antiguo principio de sabiduría...”*

Sólo recientemente las computadoras han llegado a ser bastante poderosas para permitir la aplicación a escalas amplias de este concepto a problemas del mundo real, en estimaciones estadísticas y reconocimiento de patrones.

Los modelos estadísticos van dirigidos al problema de construir un modelo estocástico

para predecir el comportamiento de un proceso aleatorio. En la construcción de este modelo, típicamente tenemos a nuestra disposición una muestra de salida del proceso. Dada esta muestra, que representa un estado incompleto de conocimiento sobre el proceso, el problema del modelado es brindar una representación del proceso haciendo uso de la información dada. Entonces podemos usar esta representación para hacer predicciones del comportamiento futuro del proceso.

Las últimas décadas han dado testimonio de progresos significantes al aumento de la capacidad de predicción de modelos estadísticos en el lenguaje natural. El modelo de lenguaje de Bahl en 1989 utilizó árboles de decisión. Della Pietra en 1994 utilizó gramáticas encadenadas para modelar la correlación en el lenguaje. Además Black en 1992 describió cómo se extraen reglas gramaticales automáticamente de un texto y cómo incorporarlas en modelos estadísticos de gramáticas.

## 4.1. Principio de la Máxima Entropía (PME)

El principio de la máxima entropía en su versión moderna fue planteado por E.T. Jaynes en [16]. En esencia se trata de determinar la distribución de probabilidad que haga la máxima entropía, haciendo uso solamente de la información de que se dispone, sin hacer suposiciones teóricas sobre la forma u otra característica de la distribución. En [16], Jaynes plantea el problema de la siguiente forma:

“Sea  $X$  una variable aleatoria discreta que toma los valores  $x_i$  para  $i = 1, 2, \dots, n$ , de la cual ignoramos las probabilidades  $p_i$ , pero conocemos el valor esperado de cierta función:  $E[f(x)] = \sum_{i=1}^n p_i f(x_i)$ . ¿Con base en esa información, es posible determinar el valor esperado de la función  $f(x)$ ?” (las funciones  $f$  y  $g$  son funciones dadas por el problema a tratar).

A primera vista el problema parece insoluble pues la información proporcionada no es suficiente. Más adelante Jaynes, apoyándose en el concepto de entropía, derivado de Shannon

desde el punto de vista de la teoría de la comunicación [28], propone la siguiente solución:

“Al hacer inferencia con base en información parcial, se debe utilizar la distribución de probabilidades que tenga máxima entropía sujeta a todo aquello que es conocido.”

Ese es el principio de la máxima entropía en los términos conocidos actualmente. Para comprender mejor el *PME*, supongamos que queremos hacer inferencia sobre la distribución de probabilidades de las frases en un lenguaje determinado. La evidencia de que se dispone para realizar tal inferencia, es decir los datos conocidos, es el corpus sobre el cual se está basando el estudio. De acuerdo con el *PME*, se debe encontrar la distribución de probabilidades con máxima entropía sobre las frases del lenguaje y que esté sujeto a las restricciones de la información contenida en el corpus. Por ejemplo, la experiencia ha mostrado que los trigramas son piezas de información importantes; entonces, se pueden incluir en el modelo características que representan a los trigramas. Las funciones encargadas de medir la ocurrencia de los trigramas deben cumplir las restricciones impuestas por el modelo.

A continuación se plantea el problema formalmente desde una perspectiva más general: la de la divergencia. Bajo el concepto de la divergencia se engloba toda una familia de distribuciones exponenciales. Como caso concreto, el modelo que habitualmente se conoce como modelo de máxima entropía, puede deducirse como un modelo de mínima divergencia.

**Definición 1. (Divergencia de Kulback-Leibler)** *Dado un conjunto finito  $X$ , y dos distribuciones de probabilidad  $p, q$  sobre  $X$ , se define la divergencia entre  $p$  y  $q$  como:*

$$D(p||q) = \sum_x p(x) \log \frac{p(x)}{q(x)} \quad (4.1)$$

Aunque, como se observa en la en la definición anterior, las distribuciones de probabilidad sobre las cuales se aplica la divergencia están definidas sobre un conjunto **finito**, la divergencia se puede aplicar a un conjunto más amplio de distribuciones de probabilidad: los campos aleatorios. Para una definición general y formal puede consultarse [32]. Para

nuestros propósitos el caso finito es suficiente, a no ser que se diga lo contrario.

En el lema siguiente se presentan algunas propiedades de la divergencia:

**Lema 1.** *Propiedades*

1.  $D(p||q)$  es una función no negativa de valor real.
2.  $D(p||q) = 0$  si, y solamente si  $p = q$ .
3.  $D(p||q)$  es estrictamente convexa en  $p$  y  $q$  separadamente.
4.  $D(p||q)$  es continuamente diferenciable en  $q$ .

Ahora bien:

$$\begin{aligned}
 D(p||q) &= \sum_{x \in X} p(x) [ \log[p(x)] - \log[q(x)] ] \\
 &= \sum_{x \in X} p(x) \log[p(x)] - \sum_{x \in X} p(x) \log[q(x)] \\
 &= -H(p) - \sum_{x \in X} p(x) \log[q(x)]
 \end{aligned}$$

Donde  $H(p)$  es la entropía de  $p$  [28, 17]. Es interesante observar que si la distribución  $q$  es la distribución uniforme,  $q(x) = \frac{1}{|x|}$  para todo  $x \in X$ , se tiene:

$$D(p||q) = -H(p) + R$$

donde  $R$  es una constante. Entonces el problema de hallar un máximo para la entropía, es equivalente hallar un mínimo para la divergencia respecto a la distribución uniforme. Bajo este marco, el problema de hallar la distribución de máxima entropía relativa a la distribución  $q$ , se plantea en términos de la siguiente definición [8]:

**Definición 2.** *Dada la familia lineal  $F$  de distribuciones de probabilidad sobre el conjunto finito  $X$*

$$F = \left\{ p : \sum_x p(x) f_i(x) = K_i \right\} \quad i = 1, 2, \dots, m \quad (4.2)$$

y una distribución de probabilidades conocida  $q$ , la distribución de probabilidades  $\hat{p} \in F$  que haga mínima la divergencia respecto a  $q$  se denomina distribución de mínima divergencia o de máxima entropía relativa a  $q$ :

$$\hat{p} = \arg \min_{p \in F} D(p||q) \quad (4.3)$$

Según el lema (1),  $D$  es no negativa y estrictamente convexa, lo cual garantiza la existencia de  $\hat{p}$  y además puede demostrarse que es la única solución de mínima divergencia [31].

Para encontrar  $\hat{p}$ , de acuerdo con el planteamiento anterior, se tiene que optimizar una función sujeta a restricciones, las impuestas por la familia lineal. El método natural a utilizar para resolver tal problema es el de los multiplicadores de Lagrange. El Lagrangiano [4], es el caso que nos ocupa, está dado por la expresión:

$$\begin{aligned} \Lambda(p, \lambda) &= D(p||q) - \sum_i \lambda_i \left[ \sum_x p(x) f_i(x) - K_i \right] \\ &= \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)} - \sum_i \lambda_i \left[ \sum_x p(x) f_i(x) - K_i \right] \end{aligned} \quad (4.4)$$

donde  $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_m\}$ . Derivando la expresión (4.4) con respecto a  $p$ , igualando a cero y despejando la expresión  $p$ , obtenemos finalmente la distribución exponencial:

$$\hat{p}(x) = \frac{1}{Z} q(x) \exp \left\{ \sum_{i=1}^m \lambda_i f_i(x) \right\} \quad (4.5)$$

donde la  $Z$  es la constante de normalización de la distribución y los  $\lambda_i$ , que son los multiplicadores de Lagrange, son parámetros que representan los pesos de aporte de cada una de las características al modelo. Las  $f_i$  serán las que determinen y ponderen la ocurrencia de cada característica dentro de los eventos concretos. La distribución  $q$



se denomina distribución a priori y fija la forma inicial de la distribución  $\hat{p}$ . Si en (4.5) definimos  $q$  como la distribución uniforme, se obtiene la expresión:

$$p(x) = \frac{1}{Z} \exp \left\{ \sum_{i=1}^n \lambda_i f_i(x) \right\}$$

que es de la forma usual conocida como distribución de máxima entropía, sin embargo los parámetros  $\lambda_i$  son desconocidos, es decir en el proceso, hasta ahora se ha optimizado respecto a  $p$  y falta hacerlo respecto a  $\lambda$ . Ahora, sustituyendo en (4.4) el  $p$  dado por (4.5) y  $K_i$  por  $E_{\hat{p}}[f_i]$  se tiene:

$$\begin{aligned} \Lambda(p, \lambda) &= \sum_{x \in X} p(x) \log p(x) - \sum_{x \in X} p(x) \log q(x) - \sum_{i=1}^m \sum_{x \in X} \lambda_i p(x) f_i(x) + \sum_{i=1}^m \lambda_i E_{\hat{p}}[f_i(x)] \\ \Lambda(p, \lambda) &= -\log Z + \sum_{i=1}^m \lambda_i E_p[f_i(x)] \end{aligned} \quad (4.6)$$

Ahora, se tiene que encontrar los valores de  $\lambda_i$  que maximizan la función mostrada en (4.6). Si definimos  $\Psi(\lambda) = \Lambda(p, \lambda)$ , podemos resumir el problema que nos resta por resolver de la siguiente manera: encontrar

$$\hat{\lambda} = \arg \max_{\lambda} \Psi(\lambda)$$

Observese que la verosimilitud de  $p$  respecto a la distribución empírica  $\tilde{p}$ , está dada por:  $\prod p(x)^{\tilde{p}(x)}$  y por tanto el algoritmo de la verosimilitud que notaremos  $L_{\tilde{p}}(\lambda)$ , tiene la forma:

$$\begin{aligned} L_{\tilde{p}}(\lambda) &= \sum_{x \in X} \tilde{p}(x) \log p(x) \\ &= \sum_{x \in X} \tilde{p}(x) \left\{ \log q(x) + \sum_{i=1}^m \lambda_i f_i(x) - \log Z \right\} \\ &= -\log Z + \sum_{i=1}^m \lambda_i E_{\tilde{p}}[f_i(x)] + \sum_{x \in X} \tilde{p}(x) \log q(x) \end{aligned} \quad (4.7)$$

Comparando (4.6) en (4.7) concluimos que maximizar  $\Psi(\lambda)$  es equivalente a maximizar  $L_{\tilde{p}}(\lambda)$ . Por tal razón, la función objetivo a optimizar con miras a estimar los parámetros del modelo es el logaritmo de la verosimilitud:  $L_{\tilde{p}}(\lambda)$ . De esa manera los parámetros del modelo se obtienen encontrando el valor de  $\lambda$  que haga máximo el logaritmo de la verosimilitud. En términos generales no es fácil encontrar analíticamente tal máximo y usualmente es necesario hacer uso de técnicas numéricas para estimar el valor de los parámetros  $\lambda$ .

Básicamente hay dos algoritmos utilizados en la estimación de los parámetros: el Generalize Iterative Scaling (GIS) [29] y el Improve Iterative Scaling (IIS) [31], siendo estos dos algoritmos iterativos. En cada iteración se estima un conjunto de parámetros  $\lambda_i$  que definen una distribución exponencial de la forma (4.5). De esa manera se va generando una distribución que converge [8] a la distribución de mínima divergencia. Puesto que el *IIS* es una generalización del *GIS*, presentaremos el *IIS*.

#### 4.1.1. El algoritmo Improve Iterative Scaling

El algoritmo *IIS* es una mejora del algoritmo *GIS* introducido por Della Pietra y otros en [31]. Como el *GIS*, las funciones  $f_i$  deben ser no negativas, pero a diferencia del *GIS* en el *IIS* no es necesario hacer restricciones sobre la suma de las  $f_i$ , en el *GIS* se debe cumplir que  $\sum_{i=1}^m f_i = 1$  para todo  $x$ ; tal condición no es necesaria en el *IIS*.

El *IIS* está basado en algunas de las propiedades del logaritmo de la verosimilitud, puede deducirse sin dificultad que:

- $L_{\tilde{p}}(\lambda) \leq 0$  siempre
- $L_{\tilde{p}}(\lambda) \leq 0$  es máximo ( $L_{\tilde{p}}(\lambda) = 0$ ), para el modelo  $p_\lambda$  que mejor ajusta a  $\tilde{p}$ , es decir, para el modelo tal que  $p_\lambda(x) = 1$  si  $\tilde{p}(x) > 0$ .

En cada iteración del *IIS* se deben encontrar los cambios  $\delta_i$ , en los parámetros  $\lambda_i$ , que produzcan una mejora en el logaritmo de la verosimilitud. Si notamos a la variación

en los parámetros como  $\Delta = \{\delta_1, \delta_2, \dots, \delta_m\}$ , el cambio en logaritmo de la verosimilitud producido por  $\Delta$  esta dado por:

$$\mathbb{L}_p(\lambda + \Delta) - L_{\tilde{p}}(\lambda) = \sum_x \tilde{p}(x) \log p_{\lambda+\Delta}(x) - \sum_x \tilde{p}(x) \log p_\lambda(x) \quad (4.8)$$

mediante manipulaciones algebraicas en la expresión anterior se puede mostrar que, [5]:

$$\mathbb{L}_p(\lambda + \Delta) - L_{\tilde{p}}(\lambda) \geq A(\Delta|\lambda) \geq B(\Delta|\lambda) \quad (4.9)$$

donde  $A(\Delta|\lambda)$  y  $B(\Delta|\lambda)$  están dados por las fórmulas:

$$A(\Delta|\lambda) = \sum_x \tilde{p}(x) \sum_{i=1}^m \delta_i f_i(x) - \sum_x p(x) \exp\left(\sum_{i=1}^m \delta_i f_i(x)\right) + 1 \quad (4.10)$$

$$B(\Delta|\lambda) = \sum_x \tilde{p}(x) \sum_{i=1}^m \delta_i f_i(x) + 1 - \sum_x p(x) \sum_{i=1}^m \delta_i f(i|x)(x) e^{\delta_i f^\#(x)} \quad (4.11)$$

$f^\#(x) = \sum_{i=1}^m f_i(x)$  y  $f(i|x) = \frac{f_i(x)}{f^\#(x)}$  son distribuciones de probabilidad definidas sobre los números naturales. Cuando las funciones características  $f_i$  son binarias,  $f^\#$  es el número de características activas de  $x$ .

En la desigualdad (4.9), al obtener máximos para  $B(\Delta|\lambda)$  se está obteniendo un modelo  $p_{\lambda+\Delta}$ , que es una mejora (desde el punto de vista de la verosimilitud) sobre el modelo  $p_\lambda$ . Para optimizar  $B(\Delta|\lambda)$  tomamos su derivada parcial respecto a cada  $\delta_i$ , así obtenemos:

$$\frac{\partial B(\Delta|\lambda)}{\partial \delta_i} = \sum_x \tilde{p}(x) f_i(x) - \sum_x p(x) f_i(x) e^{\delta_i f^\#(x)} \quad i = 1, 2, \dots, m. \quad (4.12)$$

Ahora podemos plantear el algoritmo *IIS* de la siguiente manera:

### Algoritmo 1. Improve Iterative Scaling

- *Inicie cada  $\lambda_i$  en un punto arbitrario*
- *Repita hasta que se dé la convergencia*

- a) Resolver respecto a  $\delta_i$  el conjunto de ecuaciones  $\frac{\partial B(\Delta|\lambda)}{\partial \delta_i} = 0$  para todo  $i = 1, 2, \dots, m$
- b) Hacer  $\lambda_i = \lambda_i + \delta_i$

La sucesión de distribuciones generadas por las iteraciones del *IIS*, decrece monótonamente teniendo como límite la distribución de mínima divergencia  $\hat{p}$ . Para una demostración detallada de la convergencia del *IIS* puede verse [31].

## 4.2. Aplicación del modelo de máxima entropía en el reconocimiento de huellas digitales

Como se mencionó en la sección (4.1), en la distribución exponencial (4.5) las funciones  $f_i$  son las que determinan y ponderan la ocurrencia de cada característica dentro de los eventos complejos. Para el caso de reconocimiento de huellas digitales, como se vio en los capítulos (2) y (3), las características de las huellas están determinadas por las minucias y estas últimas se pueden reducir a dos tipos: terminación y bifurcación de crestas. De ahí, que sólo se tendrán dos tipos de dichas funciones.

Antes de mencionar dichas funciones notaremos por  $B$  cuando se hable de bifurcación de cresta y  $T$  cuando hablemos de terminación de cresta.

Sea  $x$  una huella digital tomada como un conjunto de minucias (Terminaciones ( $T$ ) y bifurcaciones ( $B$ ) de crestas), ahora definimos las siguientes funciones:

$$f_B(x) = \begin{cases} 1 & \text{si } B \in x \\ 0 & \text{si } B \notin x \end{cases} \quad (4.13)$$

$$f_T(x) = \begin{cases} 1 & \text{si } T \in x \\ 0 & \text{si } T \notin x \end{cases} \quad (4.14)$$

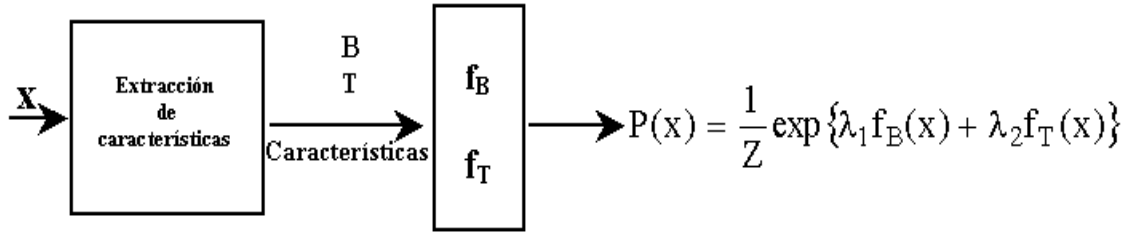


Figura 4.1: Esquema del clasificador usando el principio de la máxima entropía

La figura (4.2) muestra el proceso de reconocimiento.

Teniendo las funciones  $f_B$  y  $f_T$  entonces la expresión (4.5) queda de la siguiente forma:

$$p(x) = \frac{1}{Z} \exp \{ \lambda_1 f_B(x) + \lambda_2 f_T(x) \} \quad (4.15)$$

Se deben estimar los parámetros  $\lambda_1$  y  $\lambda_2$  pero como  $Z$  es también desconocido entonces se ha modificado el algoritmo (1) de la siguiente manera: se obtiene una muestra aleatoria de  $p$ , digamos  $s_1, s_2, \dots, s_N$  del  $\Omega^3$ . Para estimar  $E_p$ , tenemos que:

$$\begin{aligned} \frac{\partial B(\Delta|\lambda)}{\partial \delta_i} &= \sum_x \tilde{p}(x) f_i(x) - \sum_x p(x) f_i(x) e^{\delta_i f^\#(x)} \\ &= E_p [f_i e^{\delta_i f^\#}] \approx \frac{1}{N} \sum_{j=1}^N f_i(s_j) e^{\delta_i f^\#(s_j)} \quad i = 1, 2, \dots, m. \end{aligned}$$

Luego:

$$\frac{1}{N} \sum_{j=1}^N f_i(s_j) e^{\delta_i f^\#(s_j)} - \sum_{s \in \Omega} \tilde{p}(s) f_i(s) = 0 \quad i = 1, 2, \dots, m. \quad (4.16)$$

Ahora, si definimos  $\beta_i = e^{\delta_i}$ , (4.16) se transforma en:

$$\frac{1}{N} \sum_{j=1}^N f_i(s_j) \beta_i^{f^\#(s_j)} - \sum_{w \in \Omega} \tilde{p}(w) f_i(w) = 0 \quad i = 1, 2, \dots, m. \quad (4.17)$$

Lo cual introduce cambios en el algoritmo(1) de la siguiente forma:

---

<sup>3</sup>Corpus de entrenamiento

## Algoritmo 2. Improve Iterative Scaling *con muestreo.*

- Inicie cada  $\mu_i = 1$  donde  $\mu_i = e^{\lambda_i}$
- Repita hasta que se dé la convergencia
  - a) Genere una muestra  $s_1, s_2, \dots, s_N$  proveniente de  $p$
  - b) Resolver respecto a  $\beta_i$  el conjunto de ecuaciones

$$\frac{1}{N} \sum_{j=1}^N f_i(s_j) \beta_i^{f_i^\#(s_j)} - \sum_{w \in \Omega} \tilde{p}(w) f_i(w) = 0 \quad i = 1, 2, \dots, m. \quad (4.18)$$

- c) Hacer  $\mu_i = \mu_i * \beta_i$

Se hace especial referencia al paso a) del algoritmo 2 en lo que se refiere a la generación de la muestra, se observa en la ecuación(4.15) que  $p(x)$  no es totalmente conocida dado que  $Z$  es la constante de normalización que no se puede obtener sin conocer los parámetros  $\lambda$ . Los metodos Monte Carlo por cadenas de Markov (MCMC)[30, 21] permiten extraer muestras de una distribución sin que se conozca la constante de normalización.

El sistema que se propone esquematizado en la figura(4.2), tiene los siguientes pasos:

- Se introduce la huella digital  $x$ .
- Se extraen las características utilizando los métodos propuestos en el capítulo 2.
- Se estima la probabilidad  $p(x)$  con la ayuda de las funciones  $f_B$  y  $f_T$  utilizando la ecuación(4.15) y el algoritmo 2.

Este mismo procedimiento se tiene que haber realizado para generar una base de datos. Después de haber estimado la probabilidad  $p(x)$  se compara con las probabilidades de la base de datos y se elige como el representante de  $x$  aquella para la cual  $p(x)$  cumple  $\min_{p(x)} |p(x) - P_{\text{Base de datos}}(x)|$ . Dicho de otra manera, se escoge la huella (Base de datos) que tenga la probabilidad más proxima con la de la huella de entrada.

Los detalles de la implementación de las funciones, el planteamiento del modelo, su implementación computacional y puesta a punto están fuera de los objetivos del trabajo y se dejan planteados para el desarrollo de futuros trabajos.

# Conclusiones

1. En el estudio del reconocimiento automático de formas (RAF) se analizaron los reconocedores con respecto a la fase de diseño y de operación, permitiéndonos concluir que la metodología utilizada en los sistemas de RAF desempeña un papel importante para los sistemas biométricos en los reconocedores de huellas digitales.
2. Se hizo una introducción a la biometría, ya que dentro de esta área se encuentra el sistema de reconocimiento automático de huellas digitales. Se estudiaron las técnicas biométricas más importantes cuya capacidad discriminativa, basada en las características propias de cada individuo, la hacen atractiva por su seguridad en los sistemas prácticos.
3. Se estudiaron las características tanto locales como globales de la huella digital concluyendo que las bifurcaciones y finalizaciones de crestas son las más determinantes en la identificación de huellas.
4. Desde el punto de vista estadístico, la distribución hiper-geométrica ha mostrado ser la más eficiente a la hora de incorporar sistemas aleatorios al reconocedor.
5. De todas las características a considerar en el estudio de las huellas, las minucias son las que han demostrado mayor capacidad discriminativa. De modo que el estudio de la identificación de la huella se reduce usualmente al de tales características, como se hace en los sistemas más recientes.
6. Dado que existen gran variedad de algoritmos para el emparejamiento se han presentado solamente aquellos que son más popularmente usados y que han sido probados



en sistemas reales (algoritmo de Hough, algoritmo cadena, algoritmo filterbank).

7. Se analizó la posibilidad de utilizar el principio de la máxima entropía para el reconocimiento automático de huellas digitales, trabajando previamente la teoría estadística que dicho principio involucra y enfocándola a los sistemas automáticos de verificación. En este estudio solamente se planteo la posible metodología a seguir, los detalles de la implementación de las funciones, el planteamiento del modelo, su implementación computacional y puesta a punto están fuera de los objetivos del trabajo y se dejan planteados para el desarrollo de futuros trabajos.

# Bibliografía

# Bibliografía

- [1] US Patent 5633947, editor. *Method and Apparatus for Fingerprint Characterization and Recognition Using Auto-correlation Pattern*, 1997.
- [2] S. Pankanti A. K. Jain, L. Hong and Ruud Bolle, editors. *An Identity Authentication System Using Fingerprints. Proceedings of the IEEE*, 1997.
- [3] Dario Maravall Gomez Allende. *Reconocimiento de formas y visión artificial*. Addison-Wesleyes Iberoamericana, 1994.
- [4] T. M. Apostol. *Calculo con funciones de varias variables y aplicación a las ecuaciones diferenciales y la teoría de la probabilidades*. Ed. Reverte, volumen 2 edition, 1995.
- [5] A. Berger, editor. *The Improved Iterative Scaling algorithm: a gentle introduction*, Carnegie Mellon University, 1997. School of Computer Science.
- [6] N. Ratha K. Karu S. Chen and A. K. Jain, editors. *A Real-Time Matching System for Large Fringerprint Databases. Proceedings of the IEEE*, 1996.
- [7] T. M. Cover, editor. *On the Possible Ordering in the Measurement Selection Problem*. IEEE Trans. on Systems Man and Cybernetics, 1977.
- [8] I. Csiszár. A geometric interpretation of darroch and ratcliff's generalized iterative scaling, 1989.
- [9] Cummins and Charles Midlo. *Fingerprints, Palms and Soles: An Introduction to Dermatoglyphics*. Dover Publications, Inc., 1961.

- [10] Joaquim Barbé Farré. *Diseño de procesadores ópticos multicanales. Aplicación al procesamiento de texturas*. PhD thesis, Universitat Autònoma de Barcelona, Enero 2003.
- [11] Keinosuke Fukunaga. *Introduction to statistical pattern recognition*. Academic press, INC, 1972.
- [12] F. Galton. *Finger prints*. London: McMillan, 1892.
- [13] C.I. Watson G.T. Candela P.J. Grother. Comparison of fft fingerprint filtering methods for neural network classification. *U.S DEPARTMENT OF COMMERCE*, 1994.
- [14] A. K. Jain and F. Farrokhnia. Unsupervised texture segmentation using gabor filters. *Pattern Recognition*, 1991.
- [15] A. K. Jain and S. Pankanti, editors. *Biometrics Systems: Anatomy of Performance*. IEICE Trans. Fundamentals, 2001.
- [16] E. T. Jaynes, editor. *Information theory and statistical mechanics*. Physics Review, 1957.
- [17] F. Jelinek, editor. *Statistical Methods for Speech Recognition*, Massachusetts Institut of Technology. Cambridge, 1997. The MIT Press.
- [18] C. Kingston. *Probabilistic Analysis of Partial Fingerprint Patterns*. PhD thesis, University of California Berkeley, 1964.
- [19] I.-S. Oh J.-S Lee and C. Y. Suen, editors. *Analysis of Class Separation and Combination of Class-Dependent Features for Handwriting Recognition*. IEEE Trans. Patt. Anal. and Machine Intell., 1999.
- [20] D. Marr. *Vision*. W. H. Freeman, San Francisco California, 1982.
- [21] R. M. Neal, editor. *Probabilistic inference using markov chain monte carlo methods*. *Technical Report CRG-TR-93-1*, University of Toronto, 1993.

- [22] K. Pearson. Galton's work on evidential value of fingerprints. *Sankhya: Indian Journal of Statistics*, 1933.
- [23] Anil Jain Salil Prabhakar and Lin Hong. A multichannel approach to fingerprint clasification. *Computer Science and Engineering*, 1999.
- [24] Salil Prabhakar. *Fingerprint Classification and Matching Using a Filterbank*. PhD thesis, Michigan State University, 2001.
- [25] J. Osterburg T. Parthasarathy T. E. S. Raghavan and S. L. Sclove. Development of a mathematical formula for the calculation of fingerprint probabilities based on individual characteristics. *Journal of the American Statistical Association*, 1977.
- [26] J. A. Rice. *Mathematical Statistics and Data Analysis*. Duxbury Press, California, second edition edition, 1995.
- [27] T. Roxburgh. On evidential value of fingerprints. *Sankhya: Indian Journal of Statistics*, 1933.
- [28] C. E. Shannon. *From, "The Mathematical Theory of Communication "*. ©1949 by the Board of Trutees of the University of Illinois, 1949.
- [29] J: Darroch y D. Ratcliff, editor. *Generalize iterative scaling for log-linear models*. *Annals of Mathematical Statistics*, 1972.
- [30] N. Metropolis A. Rosenblut M. Rosenblut A. Teller y E. Teller. Equation of states calculations by fast computing machines. *J. Chemical Physics*, 21:1087-1091, 1953.
- [31] S. Della Pietra V. Della Pietra y J. Lafferty, editor. *Inducen features of random fields*. *Technical Report CMU-CS-95-144*. Carnegie Mellon University, 1995.
- [32] O. Haggstrom y K.Ñelander. On exact simulation of markov random fields using coupling from the past. Chalmers University of Technology, 1997.
- [33] D. Juan Luis Posadas Yagüe. *Transformada Rápida de Fourier (FFT) e Interpolación en Tiempo Real*. PhD thesis, Universidad Politécnica de Valencia, Julio 1998.