

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

DEPARTAMENTO DE SEÑALES Y COMUNICACIONES



TESIS DOCTORAL

**RECONOCIMIENTO DE FORMAS MANUSCRITAS CON MODELOS
OCULTOS DE MARKOV**

CARLOS MANUEL TRAVIESO GONZÁLEZ

Las Palmas de Gran Canaria, Mayo de 2002

59/2001-02
UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA
UNIDAD DE TERCER CICLO Y POSTGRADO

Reunido el día de la fecha, el Tribunal nombrado por el Excmo. Sr. Rector Magfco. de esta Universidad, el/a aspirante expuso esta TESIS DOCTORAL.

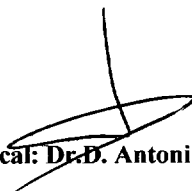
Terminada la lectura y contestadas por el/a Doctorando/a las objeciones formuladas por los señores miembros del Tribunal, éste calificó dicho trabajo con la nota de SOBRESALIENTE

Las Palmas de Gran Canaria, a 15 de mayo de 2002.

El/a Presidente/a: Dr.D. Rafael Pérez Jiménez,

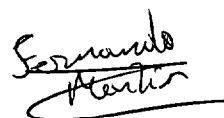


El/a Secretario/a: Dra.Dña. Itziar Goretti Alonso González,



El/a Vocal: Dr.D. Antoni Gasull Llampallas,

El/a Vocal: Dr.D. Fernando Martín Rodríguez,



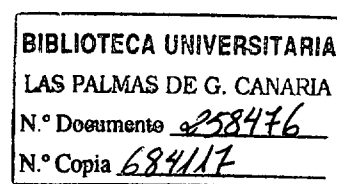
El/a Vocal: Dr.D. Fernando Díaz de María,



El Doctorando: D. Carlos Manuel Travieso González,



UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA
DEPARTAMENTO DE SEÑALES Y COMUNICACIONES
PROGRAMA DE DOCTORADO
TECNOLOGÍAS DE LAS TELECOMUNICACIONES



TESIS DOCTORAL

RECONOCIMIENTO DE FORMAS MANUSCRITAS
CON MODELOS OCULTOS DE MARKOV

AUTOR

A handwritten signature in black ink, consisting of several overlapping loops and a long horizontal stroke extending to the right.

Fdo.: Carlos Manuel Travieso González

DIRECTOR

A handwritten signature in black ink, featuring a stylized 'M' and 'F' with a long horizontal stroke underneath.

Fdo.: Miguel Ángel Ferrer Ballester

FECHA: Mayo de 2002

*A mis padres,
y para Tas.*

Agradecimientos

Esta tesis no sólo ha sido realizada por mí, sino por muchas personas que se encuentran en mí alrededor, y que gracias a ellos, ésto ha llegado ha ser posible.

En primer lugar quisiera agradecer a Miguel Ángel Ferrer, todo el tiempo, la dedicación, y la paciencia que me ha tenido en estos años de trabajo.

También quisiera acordarme de mis compañeros, toda esa gente que ha pasado por el grupo de investigación, proyectandos y otros; y por los fijos, los que siempre han estado en él, Itziar y la inestimable ayuda de José Cruz.

A mi familia, a mis padres, Manuel y Pilar, y a mis hermanas, Carolina y Nayra.

A Tasmi, que sin sus ánimos, su comprensión y sus correcciones, esta tesis hubiese sido un poco más dura.

A todos mis amigos, como no los del Lomo, a todos los perretillas (Gran Canaria), y por sí a caso, a los de Tenerife también, por hacerme éstos años más llevaderos.

A todos, todos, ¡¡¡GRACIAS!!!

Índice General

Capítulo 1. Introducción

1.1	Introducción	1
1.2	Antecedentes	2
1.2.1	Antecedentes históricos.....	2
1.2.2	Análisis de documentación	3
1.2.3	Reconocimiento de las aplicaciones manuscritas	4
1.3	Reconocimiento automático de formas.....	5
1.3.1	Establecimiento de patrones.....	6
1.3.2	Extracción de parámetros.....	6
1.3.3	Funciones discriminantes o de decisión.....	7
1.4	Propuesta: Planteamiento de la teoría	8
1.5	Objetivos de la tesis	9
1.6	Estructura de la memoria	9

Capítulo 2. Estado de la técnica en el reconocimiento de aplicaciones manuscritas

2.1	Introducción	11
2.2	Revisión del estado de la técnica	12
2.2.1	Revisión del estado de la técnica en el reconocimiento de los dígitos manuscritos	12
2.2.2	Revisión del estado de la técnica en el reconocimiento de las letras manuscritas.....	14
2.2.3	Revisión del estado de la técnica en el reconocimiento de las palabras manuscritas.....	16
2.2.4	Revisión del estado de la técnica en el reconocimiento de las firmas.....	18
2.3	Estado actual de las funciones discriminantes o de decisión	19
2.3.1	Clasificadores determinísticos	21
2.3.2	Clasificadores estadísticos	23
2.3.2.1	Clasificador Bayesiano.....	23
2.3.2.2	Clasificadores estadísticos supervisados.....	24
2.3.3	Funciones de Discriminación.....	24
2.3.3.1	Función de Discriminación Lineal.....	25
2.3.3.2	Clasificador de Mínima Distancia.....	25
2.3.3.3	Clasificador según el Vecino más Cercano.....	26
2.3.3.4	Función de Discriminación Polinomial.....	26
2.3.3.5	Redes Neuronales.....	26
2.3.3.6	Modelos ocultos de Markov	27
2.3.3.6.1	Procesos de Markov discretos.....	27
2.3.3.6.2	Ampliación a Modelos Ocultos de Markov	30
2.3.3.6.3	Elementos de los Modelos Ocultos de Markov.....	32
2.3.3.6.4	Los tres problemas fundamentales de un Modelo Oculto de Markov	33

2.3.3.6.5 Tipos de Modelos Ocultos de Markov	34
2.3.3.6.6 Multietiquetado en HMM discreto	35
2.3.3.6.7 Multiparametrización	35

Capítulo 3. Elaboración de la base de datos para manuscritos simples: dígitos y letras

3.1 Introducción	37
3.2 Base de datos de dígitos y firmas manuscritas	38
3.2.1 Parámetros de la base de datos	38
3.2.2 Elección del sistema de captura de datos	39
3.2.3 Diseño y construcción de las bases de datos	41
3.2.4 Instrucciones para rellenar las plantillas	41
3.2.4.1 Base de datos de dígitos manuscritos	42
3.2.4.1.1 Perfil de los escritores	42
3.2.4.1.2 Formato de la plantilla	44
3.2.4.2 Base de datos de letras manuscritas	47
3.2.4.2.1 Perfil de los escritores	47
3.2.4.2.2 Formato de la plantilla	49
3.2.5 Dificultades en la creación de las bases de datos	54
3.3 Preprocesado de imagen de las bases de datos	54
3.3.1 Digitalización de las bases de datos	54
3.3.2 Corrección de la inclinación	56
3.3.3 Separación de las muestras	57
3.3.3.1 Binarización de los datos	58
3.3.3.2 Eliminación del ruido	59
3.3.4 Proceso de normalización	61
3.3.5 Operador morfológico: Dilatación	62
3.4 Conclusiones	64

Capítulo 4. Clasificación de manuscritos simples: dígitos y letras

4.1 Introducción	65
4.2 Parametrización de los patrones manuscritos simples	66
4.2.1 Procedimiento para detectar y vectorizar el contorno	66
4.2.1.1 Desarrollo de la técnica del contorno	67
4.2.1.2 Proceso de vectorización del contorno por la secuencia	69
4.2.1.3 Proceso de vectorización del contorno en coordenadas polares	72
4.2.2 Procedimiento para detectar y vectorizar la esqueletización	73
4.2.2.1 Proceso de la técnica de esqueletización	74
4.2.2.2 Vectorización de la esqueletización	75
4.2.3 Discusión entre las técnicas de esqueletización y contorno	78
4.3 Clasificación de los patrones simples manuscritos	78
4.3.1 Etapa de entrenamiento para el clasificador HMM discreto	79
4.3.2 Etapa de test para el clasificador HMM	80
4.4 Optimización de los parámetros del clasificador	80
4.4.1 Optimización del clasificador para dígitos manuscritos	82
4.4.2 Optimización del clasificador para letras manuscritas	84
4.5 Base de datos NIST-19	87
4.5.1 Características de la base de datos NIST-19	88
4.5.2 Clasificación con la base de datos NIST-19	88
4.6 Conclusiones	91

Capítulo 5. Clasificación de cadenas manuscritas simples: palabras y códigos postales

5.1	Introducción	93
5.2	Bases de datos de palabras	94
5.2.1	Base de datos de palabras manuscritas.....	94
5.2.1.1	Perfil de los escritores	95
5.2.1.2	Formato de la plantilla	97
5.3.	Procesado de imagen aplicado a la base de datos	106
5.3.1	Unión de las componentes de una palabras manuscritas	106
5.3.2	Corrección de la inclinación de palabras	108
5.4	Parametrización de las palabras manuscritas	109
5.4.1	Proceso de la vectorización del contorno.....	111
5.5	Clasificación de las palabras manuscritas	113
5.6	Resultados de la clasificación de palabras manuscritas	114
5.7	Aplicación: segmentación y clasificación.....	117
5.7.1	Clasificación de códigos postales.....	118
5.7.2	Clasificación de palabras por segmentación	121
5.8	Conclusiones	127

Capítulo 6: Clasificación de manuscritos complejos: firmas

6.1	Introducción	129
6.2	Base de datos de firmas manuscritas.....	130
6.2.1	Perfil de los escritores	131
6.2.2	Formato de la plantilla	132
6.2.3	Dificultades en la creación de las bases de datos.....	138
6.3	Preprocesado de imagen de la base de datos.....	138
6.4	Parametrización de los patrones manuscritos complejos	138
6.5	Vectorización de los patrones manuscritos complejos.....	140
6.6	Clasificación de las firmas manuscritas	143
6.7	Resultados: optimización de los parámetros del clasificador	144
6.8	Verificación en las firmas manuscritas	148
6.9	Conclusiones	148

Capítulo 7. Conclusiones y línea futuras de trabajo

7.1	Introducción	151
7.2	Conclusiones	152
7.3	Líneas futuras de trabajo	153
7.3.1	Sistemas clasificadores híbridos	153
7.3.2	Expansión del ámbito de aplicación.....	156

Bibliografía

Bibliografía	163
--------------------	-----

Anexo A. Pruebas realizadas con dígitos manuscritos

A.1	Introducción	169
A.2	Optimización del clasificador para dígitos manuscritos.....	170
A.2.1	Optimización realizada para HMM con detección de contornos interior y exterior para dígitos de tamaño original.....	171

A.2.2 Optimización realizada para HMM con detección de contorno exterior para dígitos de tamaño original	174
A.2.3 Optimización realizada para HMM con detección de contorno exterior para dígitos de tamaño reducido	177
A.2.4 Optimización realizada para HMM con detección de contornos exteriores parametrizado con coordenadas polares para dígitos de tamaño reducido	181
A.2.5 Optimización realizada para HMM con la técnica de esqueletización para dígitos de tamaño reducido	183
A.2.6 Optimización realizada para NN con la técnica del gradiente	186
A.3 Conclusiones	190

Anexo B. Pruebas realizadas con las letras manuscritas

B.1 Introducción	193
B.2 Optimización del clasificador para letras manuscritas	194
B.2.1 Optimización realizada para HMM con detección de contorno exterior para letras de tamaño normalizado vectorizado por su secuencia	195
B.2.2 Optimización realizada para HMM con detección de contorno exterior para letras de tamaño normalizado vectorizado por coordenadas polares	198
B.2.3 Optimización realizada para HMM con la técnica de esqueletización para letras de tamaño normalizado	199
B.2.4 Optimización realizada para NN con la técnica del gradiente	200
B.3 Conclusiones	202

Anexo C. Pruebas realizadas con palabras manuscritas

C.1 Introducción	203
C.2 Optimización del clasificador para palabras manuscritas	204
C.2.1 Análisis de las técnicas de extracción de parámetros según el HMM	204
C.2.2 Análisis de las técnicas de extracción de parámetros según el entrenamiento	208
C.2.3 Optimización de la extracción de parámetros y su generalización	210
C.3 Conclusiones	214

Anexo D. Pruebas realizadas con firmas manuscritas

D.1 Introducción	215
D.2 Optimización del clasificador para firmas manuscritas	216
D.2.1 Optimización realizada para HMM con detección de contorno exterior para firmas	218
D.2.2 Optimización realizada para HMM con la técnica de esqueletización para firmas	221
D.2.3 Optimización realizada para NN con la técnica del gradiente	224
D.3 Conclusiones	228

Anexo E: Técnica del gradiente para clasificación de parámetros geométricos

E.1 Introducción	229
E.2 Extracción de características: Técnica del gradiente	230
E.3 Vectorización de los parámetros	231
E.4 Clasificador de parámetros geométricos	233

Índice de Figuras

Figura 1.1: Procesado de documentación y sus categorías.	4
Figura 1.2: Diagrama de bloques de un sistema de reconocimiento automático de formas.	5
Figura 1.3: Diagrama de flujo que verifica el buen funcionamiento de las etapas, en el diseño de un sistema de reconocimiento automático de formas.	6
Figura 2.1: Sistema reconocedor de patrones.	19
Figura 2.2: Diagrama de bloques simbólico del proceso de aprendizaje o entrenamiento de un reconocedor.	22
Figura 2.3: Representación en un plano de dos patrones del ejemplo gráfico.	23
Figura 2.4: Clasificador lineal de dos clases.	25
Figura 2.5: Una neurona abstracta.	26
Figura 2.6: Cadena de Markov ergódica.	28
Figura 2.7: Diferentes modelos de la secuencia cara/cruz con HMM.	31
Figura 2.8: Ejemplo de proceso a modelar con un HMM.	32
Figura 2.9: Ejemplo de modelo oculto de Markov ergódico y de izquierda a derecha (Bakis).	34
Figura 3.1: Dispositivos de entrada para los sistemas on-line y off-line.	40
Figura 3.2: Porcentaje de Hombres vs. Mujeres de la base de datos de dígitos manuscritos.	44
Figura 3.3: Porcentajes de Diestros vs. Zurdos de la base de datos de dígitos manuscritos.	44
Figura 3.4: Distribución por edades en la base de datos de dígitos manuscritos.	44
Figura 3.5: Plantilla de la base de datos de los dígitos manuscritos.	45
Figura 3.6: Plantilla de la base de datos de los dígitos manuscritos rellena.	46
Figura 3.7: Porcentaje de Hombres vs. Mujeres de la base de datos de letras manuscritas.	48
Figura 3.8: Porcentajes de Diestros vs. Zurdos de la base de datos de letras manuscritas.	49
Figura 3.9: Distribución por edades en la base de datos de letras manuscritas.	49
Figura 3.10: Plantilla de la base de datos de las letras minúsculas manuscritas.	50
Figura 3.11: Plantilla de la base de datos de las letras mayúsculas manuscritas.	51
Figura 3.12: Plantilla de la base de datos de las letras minúsculas manuscritas rellena.	52
Figura 3.13: Plantilla de la base de datos de las letras mayúsculas manuscritas rellena.	53
Figura 3.14: Procesado de las plantillas.	55
Figura 3.15: Inclinación de la línea de referencia de la plantilla y cálculo del ángulo.	56
Figura 3.16: Preprocesado de la muestra de una letra mayúscula.	57
Figura 3.17: Histograma para detectar las rectas horizontales y verticales.	58
Figura 3.18: Histograma lineal y logarítmico del dígito manuscrito 7.	59
Figura 3.19: Ejemplos de máscaras para el filtrado kfill.	60
Figura 3.20: Eliminación del ruido salpicado.	61

Figura 3.21: Escritura por parte de diferentes escritores del patrón "a".	61
Figura 3.22: Proceso de normalización.	62
Figura 3.23: Ejemplo de dilatación.	63
Figura 3.24: Dilatación del número nueve.	63
Figura 4.1: Técnicas de contorno y esqueletización.	66
Figura 4.2: Contorno exterior e interior de la letra "A".	66
Figura 4.3: Puntos críticos.	67
Figura 4.4: Matrices para detectar el contorno (X: no importa el valor).	67
Figura 4.5: Matrices para detectar puntos críticos.	68
Figura 4.6: Transiciones en los puntos de cruce.	68
Figura 4.7: Ejemplo de píxeles esquina redundante.	69
Figura 4.8: Matrices para detectar píxeles esquina redundantes.	69
Figura 4.9: Convenio para aplicaciones con letras manuscritas.	70
Figura 4.10: Convenio para aplicaciones con dígitos manuscritos.	70
Figura 4.11: Matrices de prioridad.	71
Figura 4.12: Diagrama de flujo de la vectorización del contorno.	71
Figura 4.13: Origen de los sistemas de coordenadas polares para vectorizar.	72
Figura 4.14: Barrido en coordenadas polares para determinar la secuencia de la muestra.	73
Figura 4.15: Máscara del proceso de esqueletización para detectar el 4.	74
Figura 4.16: Máscara del proceso de esqueletización para detectar el 2.	74
Figura 4.17: Máscara del proceso de esqueletización para detectar el 3.	75
Figura 4.18: Proceso de esqueletización.	75
Figura 4.19: Preprocesado para la mejora de las imágenes esqueletizadas.	76
Figura 4.20: Matriz de saltos y prioridad en la vectorización del esqueleto.	76
Figura 4.21: Proceso de vectorización del esqueleto.	77
Figura 4.22: Matriz de confusión para el reconocimiento de letra minúscula manuscrita.	81
Figura 5.1: Porcentaje de Hombres vs. Mujeres de la base de datos de palabras manuscritas.	96
Figura 5.2: Porcentajes de Diestros vs. Zurdos de la base de datos de palabras manuscrita.	97
Figura 5.3: Distribución por edades en la base de datos de palabras manuscritas.	97
Figura 5.4: Primera plantilla de la base de datos de palabras manuscritas.	98
Figura 5.5: Segunda plantilla de la base de datos de palabras manuscritas.	99
Figura 5.6: Tercera plantilla de la base de datos de palabras manuscritas.	100
Figura 5.7: Cuarta plantilla de la base de datos de palabras manuscritas.	101
Figura 5.8: Primera plantilla de la base de datos de palabras manuscritas rellena.	102
Figura 5.9: Segunda plantilla de la base de datos de palabras manuscritas rellena.	103
Figura 5.10: Tercera plantilla de la base de datos de palabras manuscritas rellena.	104
Figura 5.11: Cuarta plantilla de la base de datos de palabras manuscritas rellena.	105
Figura 5.12: Preprocesado de imagen aplicado a la base de datos de palabras manuscritas.	106
Figura 5.13: Unión de las componentes del número seis.	107
Figura 5.14: Proceso de unión de las componentes de una palabra.	107
Figura 5.15: Ejemplos de las inclinaciones que puede presentar una palabra en su escritura.	108
Figura 5.16: Método de corrección de la inclinación.	110
Figura 5.17: Método de parametrizar con altos, anchos y transiciones.	111
Figura 5.18: Convenio para los patrones de palabras manuscritas.	111

Figura 5.19: Matriz de saltos.....	112
Figura 5.20: Matrices de prioridad.....	112
Figura 5.21: Vectorización del contorno con coordenadas polares.	112
Figura 5.22: Estructura de datos de los parámetros altos, anchos y transiciones.....	113
Figura 5.23: Ejemplo de sobre con código postal 51005 digitalizado.	118
Figura 5.24: Zona de la dirección y líneas separadas extraídas del sobre con código postal 51005.	119
Figura 5.25: Primeras palabras en blanco y negro sin ruido de las dos últimas líneas del sobre con código postal 51005.	119
Figura 5.26: Primeras palabras de las dos últimas líneas del sobre con código postal 51005 segmentadas.....	120
Figura 5.27: Líneas del sobre con código postal 44001 cuyas primeras palabras de las dos últimas líneas presentan múltiples posibilidades de segmentación.....	120
Figura 5.28: Ejemplo de un formulario de un texto manuscrito.	122
Figura 5.29: Descripción del proceso de reconocimiento.	123
Figura 5.30: Ejemplos de medidas usando cajas acotadas y distancia horizontal.	124
Figura 5.31: Ejemplos de medidas basadas en polígonos convexos.....	125
Figura 5.32: Detección de los signos de puntuación.....	125
Figura 5.33: Aplicación del método de segmentación de palabras.....	127
Figura 6.1: Ejemplos de recogida de firmas en documentos oficiales.....	130
Figura 6.2: Porcentaje de Hombres vs. Mujeres de la base de datos de firmas.	132
Figura 6.3: Porcentajes de Diestros vs. Zurdos de la base de datos de firmas.....	132
Figura 6.4: Distribución por edades en la base de datos de firmas.	132
Figura 6.5: Plantilla de la base de datos de firmas.....	133
Figura 6.6: Plantilla de la base de datos de firmas rellena.	134
Figura 6.7: Muestrario de las firmas de la base de datos.....	137
Figura 6.8: Imagen de una firma y su contorno.	139
Figura 6.9: Proceso del cálculo de la envolvente.....	139
Figura 6.10: Proceso del cálculo de la envolvente con varias regiones aisladas.	140
Figura 6.11: Determinación de los centroides a partir de su envolvente.	140
Figura 6.12: Técnica de vectorización por coordenadas polares de la envolvente.	141
Figura 6.13: Número de píxeles a negros por ángulo para las coordenadas polares. ..	141
Figura 6.15: Cálculo de los altos, anchos y transiciones de una firma.	142
Figura 6.16: Firma original y esqueletizada.....	142
Figura 6.17: Aplicación de la técnica del gradiente en las firmas.	143
Figura 6.18: Combinación lineal de centroides y coordenadas polares de la envolvente para reconocer firmas manuscritas.....	146
Figura 6.19: Combinación lineal de altos, anchos y transiciones con coordenadas polares y píxeles negros de la envolvente para reconocer firmas manuscritas.	147
Figura 6.20: Curvas de error en verificación.....	148
Figura 7.1: Comparativa en las aplicaciones manuscritas según su tasa de reconocimiento.....	152
Figura 7.2: Combinación de sistemas de clasificadores de diferente tipo de información.....	154
Figura 7.3: Sistema de reconocimiento biométrico mediante labios.	157
Figura 7.4: Efecto de independencia de luminosidad y bello facial.....	158
Figura 7.5: Detección de la envolvente de los labios.....	158
Figura 7.6: Extracción de parámetros en las manos.....	159

Figura 7.7: Sistema clasificador de especies arbóreas por su hoja.	160
Figura 7.8: Cálculo del contorno de las hojas.	160
Figura 7.9: Procesado de imagen aplicado a la letra "C"	161
Figura 7.10: Sistema de reconocimiento del alfabeto dactilológico.	162
Figura A.1: Tasa de reconocimiento en función del nº de estados (3 muestras por escritor).	172
Figura A.2: Tasa de reconocimiento en función del nº de estados (9 muestras por escritor).	172
Figura A.3: Tasa de reconocimiento en función del nº de escritores (3 muestras por escritor).	173
Figura A.4: Tasa de reconocimiento en función del nº de escritores (9 muestras por escritor).	174
Figura A.5: Tasa de reconocimiento en función del nº de estados (3 muestras por escritor).	175
Figura A.6: Tasa de reconocimiento en función del nº de estados (9 muestras por escritor).	176
Figura A.7: Tasa de reconocimiento en función del nº de escritores (3 muestras por escritor).	177
Figura A.8: Tasa de reconocimiento en función del nº de estados (3 muestras por escritor).	178
Figura A.9: Tasa de reconocimiento en función del nº de estados (9 muestras por escritor).	179
Figura A.10: Tasa de reconocimiento en función del nº de escritores (3 muestras por escritor).	180
Figura A.11: Tasa de reconocimiento en función del nº de escritores (9 muestras por escritor).	180
Figura A.12: Tasa de reconocimiento en polares con origen en el centro de masas... ..	182
Figura A.13: Tasa de reconocimiento en polares con origen en la parte inferior izquierda.	182
Figura A.14: Tasa de reconocimiento en función del nº de estados (3 muestras por escritor).	184
Figura A.15: Tasa de reconocimiento en función del nº de estados (9 muestras por escritor).	184
Figura A.16: Tasa de reconocimiento en función del nº de escritores (3 muestras por escritor).	185
Figura A.17: Tasa de reconocimiento en función del nº de escritores (9 muestras por escritor).	186
Figura A.18: Tasas de reconocimiento en función del número de neuronas de la capa oculta y el coeficiente de aprendizaje para 10 escritores y 3 muestras por escritor.	187
Figura A.19: Tasas de reconocimiento en función del número de neuronas de la capa oculta y el coeficiente de aprendizaje para 10 escritores y 9 muestras por escritor.	188
Figura A.20: Tasas de reconocimiento en función del número de escritores y el coeficiente de aprendizaje para 100 neuronas en la capa oculta.	189
Figura A.21: Tasas de reconocimiento en función del número de escritores y el coeficiente de aprendizaje para 300 neuronas en la capa oculta.	189
Figura B.1: Tasa de reconocimiento en función del nº de escritores con contornos para letra minúscula.	196

Figura B.2: Tasa de reconocimiento en función del nº de estados con contornos para letra minúscula.	197
Figura B.3: Tasa de reconocimiento en función del nº de estados con contornos para polares en letra minúscula.	198
Figura B.4: Tasa de reconocimiento en función del nº de estados con contornos para polares en letra mayúscula.	199
Figura B.5: Tasa de reconocimiento en función del nº de escritores con esqueletización para letra minúscula.	200
Figura B.6: Tasa de reconocimiento de la técnica del gradiente para letras minúsculas.	201
Figura C.1: Tasa de reconocimiento de la secuencia con y sin corrección de la inclinación.	205
Figura C.2: Tasa de reconocimiento de la secuencia en coordenadas polares a partir del centro de masas, corrigiendo y sin corregir el efecto de cursiva.	206
Figura C.3: Tasa de reconocimiento de la secuencia en coordenadas polares a partir del margen inferior izquierdo, corrigiendo y sin corregir el efecto de cursiva.	206
Figura C.4: Tasa de reconocimiento para los altos y anchos del contorno, corrigiendo y sin corregir el efecto de cursiva.	207
Figura C.5: Tasa de reconocimiento para los altos, anchos y transiciones del contorno, corrigiendo y sin corregir el efecto de cursiva.	208
Figura C.6: Tasa de reconocimiento para los altos, anchos y transiciones del contorno, al variar el porcentaje de entrenamiento.	209
Figura C.7: Tasa de reconocimiento para la secuencia del contorno codificando sus direcciones, al variar el porcentaje de entrenamiento.	210
Figura C.8: Tasa de reconocimiento para las variaciones del multietiquetado aplicadas a altos, anchos y transiciones del contorno, corrigiendo el efecto de cursiva.	211
Figura C.9: Tasa de reconocimiento para las variaciones del multietiquetado aplicadas a altos, anchos y transiciones del contorno, con el efecto y sin el efecto de la corrección de la cursiva.	212
Figura C.10: Tasa de reconocimiento para la agrupación de palabras.	213
Figura D.1: Tasa de reconocimiento para 10 firmantes con esqueletización para firmas.	221
Figura D.2: Tasa de reconocimiento para 10 firmantes con esqueletización para firmas con una reserva.	223
Figura D.3: Tasa de reconocimiento para 10 firmantes con la técnica del gradiente para firmas.	225
Figura D.4: Tasa de reconocimiento para 10 firmantes con la técnica del gradiente para firmas con una reserva.	227
Figura E.1: Definición de la vecindad de un píxel.	224
Figura E.2: Máscara de Kirsch usadas para la extracción de parámetros direccionales.	225
Figura E.3: Proceso de extracción de las características de una firma.	226

Índice de Tablas

Tabla 2.1: Tasas de reconocimiento del clasificador en la referencia [Seiler96].....	15
Tabla 2.2: Tasas de reconocimiento del clasificador en la referencia [Brakensiek99].....	16
Tabla 2.3: Tasas de reconocimiento del clasificador en la referencia [Seiler96].....	16
Tabla 3.1: Muestrario de escritores de la base de datos de dígitos manuscritos.....	43
Tabla 3.2: Muestrario de los escritores de la base de datos de letras.....	48
Tabla 3.3: Tamaños de los ficheros de las bases de datos en píxeles.....	62
Tabla 4.1: Tasas de reconocimiento según el tamaño de los dígitos.....	82
Tabla 4.2: Tasas de reconocimiento según las técnicas de vectorización de la información secuencial para dígitos.....	83
Tabla 4.3: Tasas de reconocimiento según el tipo de información para dígitos.....	83
Tabla 4.4: Establecimiento de los reconocimientos medios para clasificador HMM y NN con dígitos.....	84
Tabla 4.5: Tasas de reconocimiento para optimizar el clasificador HMM con letras manuscritas.....	85
Tabla 4.6: Tasas de reconocimiento según el tipo de información para letra manuscrita.....	86
Tabla 4.7: Establecimiento de los reconocimientos medios para clasificador HMM y NN con letras minúsculas.....	87
Tabla 4.8: Establecimiento de los reconocimientos medios para clasificador HMM y NN con letras mayúsculas.....	87
Tabla 4.9: Tipos de patrones y cantidad de muestras de la base de datos NIST-19.....	89
Tabla 4.10: Clasificación de dígitos de la base de datos NIST-19.....	89
Tabla 4.11: Comparación de los sistemas reconocedores actuales para dígitos manuscritos.....	90
Tabla 4.12: Clasificación de letras de la base de datos NIST-19.....	90
Tabla 4.13: Comparación de los sistemas reconocedores actuales para letras manuscritas.....	91
Tabla 5.1: Muestrario de los escritores de la base de datos de palabras manuscritas.....	96
Tabla 5.2: Resultados de las diversas técnicas planteadas para reconocer palabras, corrigiendo y sin corregir el efecto de cursiva.....	114
Tabla 5.3: Resultados de los diversos entrenamientos aplicados a la técnica de secuencia, corrigiendo el efecto de cursiva.....	115
Tabla 5.4: Resultados de los diversos entrenamientos aplicados a la técnica de altos, anchos y transiciones, corrigiendo el efecto de cursiva.....	116
Tabla 5.5: Resultados del multietiquetado aplicado a la técnica de altos, anchos y transiciones, corrigiendo el efecto de cursiva.....	116
Tabla 5.6: Resultados del multietiquetado aplicado a la técnica de altos, anchos y transiciones, con y sin la corrección del efecto de cursiva.....	116
Tabla 5.7: Resultados de los diferentes agrupamientos de las palabras para su reconocimiento.....	117

Tabla 5.8: Ejemplo de reconocimiento del código postal.....	121
Tabla 5.9: Resultados del reconocedor de códigos postales manuscritos.....	121
Tabla 5.10: Medidas de la separación de las palabras.....	125
Tabla 5.11: Distintos tipos de clases utilizadas.....	126
Tabla 5.12: Tasas de reconocimiento de los signos de puntuación.....	126
Tabla 5.13: Comparación de los sistemas reconocedores actuales para palabras manuscritas.....	128
Tabla 6.1: Muestrario de los escritores de la base de datos de firmas.....	131
Tabla 6.2: Estudio sobre el tamaño de la envolvente para 20 firmas.....	144
Tabla 6.3: Tasas de reconocimiento para optimizar el clasificador HMM con la técnica del contorno para firmas manuscritas.....	145
Tabla 6.4: Tasa de reconocimiento para optimizar el clasificador NN con la técnica del gradiente para firmas manuscritas.....	145
Tabla 6.5: Tasas de reconocimiento según el tipo de información para firmas.....	145
Tabla 6.6: Tasas de reconocimiento del sistema combinado de polares y centroides.....	146
Tabla 6.7: Tasas de reconocimiento del sistema combinado de altos, anchos y transiciones con coordenadas polares y píxeles negros.....	147
Tabla 6.8: Tasas de reconocimiento del sistema combinado de altos, anchos y transiciones con coordenadas polares y píxeles negros variando el etiquetado.....	147
Tabla 6.9: Comparación de los sistemas reconocedores actuales para firmas.....	149
Tabla 7.1: Sistema combinado para los clasificadores HMM y NN con dígitos.....	155
Tabla 7.2: Sistema combinado para los clasificadores HMM y NN con letras minúsculas.....	155
Tabla 7.3: Sistema combinado para los clasificadores HMM y NN con letras mayúsculas.....	155
Tabla 7.4: Sistema combinado I para el clasificador HMM para la técnica de contorno para firmas manuscritas.....	156
Tabla 7.5: Sistema combinado II para el clasificador HMM para la técnica de contorno para firmas manuscritas.....	156
Tabla 7.6: Tasas de reconocimiento de los diversos sistemas para clasificar labios.....	159
Tabla 7.7: Tasas de reconocimiento de los diversos sistemas para clasificar árboles endémicos de Canarias.....	161
Tabla 7.8: Tasas de reconocimiento de los diversos sistemas para clasificar árboles endémicos de Canarias con multietiquetado.....	161
Tabla 7.9: Tasas de reconocimiento sobre el alfabeto dactilológico.....	162
Tabla A.1: Tasas de reconocimiento para 3 muestras en el entrenamiento variando el número de estados.....	171
Tabla A.2: Tasas de reconocimiento para 9 muestras en el entrenamiento variando el número de estados.....	171
Tabla A.3: Tasas de reconocimiento para 3 muestras en el entrenamiento variando el número de escritores.....	173
Tabla A.4: Tasas de reconocimiento para 9 muestras en el entrenamiento variando el número de escritores.....	173
Tabla A.5: Mejor resultado para la optimización realizada para HMM con detección de contornos interior y exterior para dígitos de tamaño original.....	174
Tabla A.6: Tasas de reconocimiento para 3 muestras en el entrenamiento variando el número de estados.....	175

Tabla A.7: Tasas de reconocimiento para 9 muestras en el entrenamiento variando el número de estados.....	175
Tabla A.8: Tasas de reconocimiento para 3 muestras en el entrenamiento variando el número de escritores.....	176
Tabla A.9: Mejor resultado para la optimización realizada para HMM con detección de contorno exterior para dígitos de tamaño original.....	177
Tabla A.10: Tasas de reconocimiento para 3 muestras en el entrenamiento variando el número de estados.....	178
Tabla A.11: Tasas de reconocimiento para 9 muestras en el entrenamiento variando el número de estados.....	178
Tabla A.12: Tasas de reconocimiento para 3 muestras en el entrenamiento variando el número de escritores.....	179
Tabla A.13: Tasas de reconocimiento para 9 muestras en el entrenamiento variando el número de escritores.....	180
Tabla A.14: Mejor resultado para la optimización realizada para HMM con detección de contorno exterior para dígitos de tamaño reducido independiente de escritor.....	181
Tabla A.15: Tasas de reconocimiento para 12 muestras en el entrenamiento variando el número de escritores en coordenadas polares con referencia al centro de masas.....	181
Tabla A.16: Tasas de reconocimiento para 12 muestras en el entrenamiento variando el número de escritores en coordenadas polares con referencia a la parte inferior izquierda.....	182
Tabla A.17: Tasas de reconocimiento para 3 muestras en el entrenamiento variando el número de estados.....	183
Tabla A.18: Tasas de reconocimiento para 9 muestras en el entrenamiento variando el número de estados.....	183
Tabla A.19: Tasas de reconocimiento para 3 muestras en el entrenamiento variando el número de escritores.....	185
Tabla A.20: Tasas de reconocimiento para 9 muestras en el entrenamiento variando el número de escritores.....	185
Tabla A.21: Mejor resultado para la optimización realizada para HMM con la técnica de esqueletización para dígitos de tamaño reducidos.....	186
Tabla A.22: Tasas de reconocimiento con 10 escritores y 3 muestras por escritor para el entrenamiento para dígitos.....	187
Tabla A.23: Tasas de reconocimiento con 10 escritores y 9 muestras por escritor para el entrenamiento para dígitos.....	187
Tabla A.24: Tasas de reconocimiento variando el número de escritores y con 100 neuronas de capa oculta.....	188
Tabla A.25: Tasas de reconocimiento variando el número de escritores y con 300 neuronas de capa oculta.....	189
Tabla A.26: Mejor resultado para la optimización realizada para NN con la técnica del gradiente para dígitos manuscritos.....	190
Tabla A.27: Establecimiento de los reconocimientos medios para los clasificadores HMM y NN.....	191
Tabla B.1: Tasas de reconocimiento con vectorización del contorno para letras minúsculas variando el número de escritores.....	196
Tabla B.2: Tasas de reconocimiento con vectorización del contorno para letras minúsculas variando el número de estados.....	197

Tabla B.3: Mejor resultado para la optimización realizada para HMM con detección de contornos exterior para letras minúsculas normalizadas.	197
Tabla B.4: Tasas de reconocimiento vectorizando con contornos para las combinaciones de agrupamiento de los patrones de letras manuscritas.	198
Tabla B.5: Tasas de reconocimiento con vectorización del contorno para polares en letra minúscula variando el número de estados.....	198
Tabla B.6: Tasas de reconocimiento con vectorización del contorno para polares en letra mayúscula variando el número de estados.	199
Tabla B.7: Tasas de reconocimiento con vectorización por esqueletización para letra minúscula variando el número de escritores.	200
Tabla B.8: Tasas de reconocimiento según las variaciones de número de escritores y neuronas en la capa oculta para letras minúsculas.	201
Tabla B.9: Mejor resultado para la optimización realizada para NN con la técnica del gradiente para letras minúsculas normalizadas.	201
Tabla B.10: Tasas de reconocimiento para NN de las combinaciones de agrupamiento de los patrones de letras manuscritas.	202
Tabla C.1: Variaciones del número de estados para la secuencia codificada por la matriz de saltos, corrigiendo y sin corregir el efecto de cursiva.....	204
Tabla C.2: Variaciones del número de estados para la secuencia en coordenadas polares a partir del centro de masas, corrigiendo y sin corregir el efecto de cursiva.	205
Tabla C.3: Variaciones del número de estados para la secuencia en coordenadas polares a partir del margen inferior izquierdo, corrigiendo y sin corregir el efecto de cursiva.	205
Tabla C.4: Variaciones del número de estados para los altos y anchos del contorno, corrigiendo y sin corregir el efecto de cursiva.....	207
Tabla C.5: Variaciones del número de estados para los altos, anchos y transiciones del contorno, corrigiendo y sin corregir el efecto de cursiva.....	207
Tabla C.6: Variaciones del porcentaje de entrenamiento para los altos, anchos y transiciones del contorno, corrigiendo el efecto de cursiva.	209
Tabla C.7: Variaciones del porcentaje de entrenamiento para la secuencia del contorno codificando sus direcciones, corrigiendo el efecto de cursiva.....	210
Tabla C.8: Variaciones del multietiquetado para altos, anchos y transiciones del contorno, corrigiendo el efecto de cursiva.	211
Tabla C.9: Resultados del multietiquetado para altos, anchos y transiciones del contorno, sin corregir el efecto de cursiva.	212
Tabla C.10: Resultados del multietiquetado para altos, anchos y transiciones del contorno, corrigiendo el efecto de cursiva.	212
Tabla C.11: Resultados de los diferentes agrupamientos de las palabras para su reconocimiento.....	213
Tabla D.1: Estudio sobre el tamaño de la envolvente para la secuencia con 20 firmas.	217
Tabla D.2: Tasas de reconocimiento con la detección de envolvente por agrupación de regiones.	218
Tabla D.3: Tasas de reconocimiento con la detección de envolvente por coordenadas polares.	219
Tabla D.4: Tasas de reconocimiento con la detección de envolvente por coordenadas polares y el número de píxeles a negro.	219
Tabla D.5: Tasa de reconocimiento de la envolvente vectorizada por la secuencia. ...	220

Tabla D.6: Tasas de reconocimiento con la detección de envolvente por altos, anchos y transiciones.	220
Tabla D.7: Tasas de reconocimiento de 10 firmantes según el número de estados y el número de repeticiones por firmantes.	221
Tabla D.8: Tasas de reconocimiento de 30 firmantes según el número de estados y el número de repeticiones por firmantes.	222
Tabla D.9: Tasas de reconocimiento de 60 firmantes según el número de estados.	222
Tabla D.10: Mejor resultado para la optimización realizada para HMM con la técnica esqueletización para reconocimiento de firmas.	222
Tabla D.11: Tasas de reconocimiento con reservas de 10 firmantes según el número de estados y el número de repeticiones por firmantes.	223
Tabla D.12: Tasas de reconocimiento con reservas de 30 firmantes según el número de estados y el número de repeticiones por firmantes.	224
Tabla D.13: Tasas de reconocimiento con reservas de 60 firmantes según el número de estados.	224
Tabla D.14: Mejor resultado para la optimización realizada para HMM con la técnica esqueletización para reconocimiento con dos reservas de firmas.	224
Tabla D.15: Tasas de reconocimiento de 10 firmantes según el número de neuronas en la capa oculta y el número de repeticiones por firmantes.	225
Tabla D.16: Tasas de reconocimiento de 30 firmantes según el número de neuronas en la capa oculta y el número de repeticiones por firmantes.	225
Tabla D.17: Tasas de reconocimiento de 60 firmantes según el número de neuronas en la capa oculta.	226
Tabla D.18: Mejor resultado para la optimización realizada para NN con la técnica del gradiente para reconocimiento de firmas.	226
Tabla D.19: Tasas de reconocimiento con reservas de 10 firmantes según el número de neuronas en la capa oculta y el número de repeticiones por firmantes.	226
Tabla D.20: Tasas de reconocimiento con reservas de 30 firmantes según el número de neuronas en la capa oculta y el número de repeticiones por firmantes.	227
Tabla D.21: Tasas de reconocimiento con reservas de 60 firmantes según el número de neuronas en la capa oculta.	227
Tabla D.22: Mejor resultado para la optimización realizada para NN con la técnica del gradiente para reconocimiento con reservas de firmas.	228

Capítulo 1

Introducción

1.1 Introducción

Desde que el hombre es capaz de usar su inteligencia, materializa sus ideas en palabras o expresiones, que expresa mediante su aparato de fonación y representa mediante signos gráficos. En la actualidad, el proceso de escritura y lectura implica el análisis de la serie de signos gráficos, lo que permite reconstruir su aspecto auditivo y visual, análisis que en su conjunto sugiere el concepto o la idea.

El problema del reconocimiento de formas es variable según el planteamiento propuesto. Pero, ¿se podría crear una técnica universal, que construya un operador para reconocer formas? Esto se creía en los años cincuenta, en los que se veían florecer propuestas de sistemas autoorganizadores y de aprendizajes automáticos basado en perceptrones. En realidad, no existe ninguna aplicación capaz de construir un operador que resuelva cualquier problema del reconocimiento de formas. Por eso, se recomienda indagar en la modestia y en la paciencia, ya que en la actualidad, estamos muy lejos de hacerlo tan bien con nuestras potentes máquinas “de quinta generación”; como lo hacen

los seres vivos con sus sentidos del olor y de la vista. Por tanto, se nos presenta un campo amplio y en plena expansión, como se demuestra con las miles de complejas publicaciones por año. Este reconocimiento es un acercamiento de las distintas técnicas de la programación y la matemática, y no por la síntesis del algoritmo o de la teoría, sino por todas las formas que nos podemos encontrar en nuestro entorno más cercano.

En particular, el reconocimiento de las aplicaciones manuscritas se presenta como un tipo de forma a reconocer, que se puede desarrollar gracias a los avances tecnológicos actuales, lo que nos permite pasar de la idea a la realidad, de la teoría a los resultados prácticos.

En este primer capítulo se va a describir, en general, los orígenes del reconocimiento de patrones, profundizando en las aplicaciones manuscritas. También, se va a realizar un recorrido por los distintos aspectos de un reconocedor de patrones: la segmentación de los elementos a discriminar, la extracción de características y la toma de decisiones. Seguidamente, se pasará a dar las ideas del planteamiento que sostiene esta tesis, y finalmente, se cerrará el capítulo, con el planteamiento de los objetivos a lograr y la descripción del resto de la memoria.

1.2 Antecedentes

1.2.1 Antecedentes históricos

El proceso de escritura no ha sido siempre el mismo, ya que muchas veces no se sigue una única línea, ni se cubren todas las etapas de realización. Así, la génesis de la escritura puede sintetizarse en los siguientes cuatro periodos:

1. El hombre primitivo utilizó diversos medios para su comunicación gráfica, sintetizándolo en el empleo de un signo material, que inicialmente fue un símbolo y más tarde un dibujo o una pictografía.
2. Una nueva época viene dada por el surgimiento de la escritura sintética, según la cual un signo gráfico o un grupo de signos sugiere toda una frase. El problema radica en las infinitas frases o pensamientos.
3. Otro periodo en la génesis de la escritura consistió en la evocación mediante un signo de una sola palabra, ya no de una frase. Así nacieron las ideográficas o analíticas que elaboraron una serie de signos gráficos constates, y con las que se podía constituir frases.
4. Las escrituras silábicas y fonéticas suponen la génesis decisiva de la escritura. Éstas nacen de un análisis de los elementos de las palabras, sílabas y fonemas, reduciéndose así, de frases a palabras y de palabras a letras.

Realizando una síntesis histórica de estos cuatro periodos, se establece que el nacimiento de la escritura alfabética actual parte a mediados del segundo milenio A.C., posiblemente en Siria y en el medio cultural que descubrieron las excavaciones arqueológicas, las cuales dieron a conocer las escrituras ugaríticas, protosinaíticas, pseudojeroglíficas de Biblos, o bien de los egeas. Estos precedentes hicieron posible el

nacimiento del alfabeto fenicio, consonántico en un comienzo, lo mismo que otros alfabetos semíticos, y cuya lenta evolución dio lugar a la aparición de los grandes sistemas gráficos griego y latino [Salvat65].

El uso de la información documental ha sido escrita, copiada, almacenada y manipulada, directamente por la acción humana desde el origen de su creación. Pero a finales de los años sesenta nació el reconocimiento de formas, al mismo tiempo que los potentes ordenadores llamados “de la tercera generación”. La llegada de estas máquinas permitió la experimentación sobre las representaciones de sensores físicos, como imágenes, palabras, datos físicos, médicos, económicos, lingüísticos, etc.; y por tanto, un nuevo campo de observación y de reflexión fue transferido de la Filosofía a la Ciencia experimental. Al igual que del telescopio nace la Astronomía, y de las técnicas metalúrgicas, químicas, electrónicas y del vacío nace la Física; no es ninguna paradoja el hecho de que el reconocimiento de formas y la inteligencia artificial hayan nacido a partir de los ordenadores, máquinas concebidas para la gestión contable y el cálculo científico.

1.2.2 Análisis de documentación

A finales de los ochenta, la supremacía de los rápidos ordenadores, con grandes memorias y escáner más baratos, provocaron el interés por el análisis de documentos. En esos tiempos y hasta hoy en día, muchos documentos han sido enviados y recibidos por fax, y almacenados en interminables bases de datos, y de ahí que haya crecido el interés por tratar a estas imágenes de forma más simple tanto para su visión como para su impresión. La cuestión era extraer la información igual que lo hacen los humanos, con una lectura sencilla. El área de investigación estaba servida y los sistemas comerciales eran construidos para leer texto en una página, encontrar registros en una forma, y localizar líneas y símbolos en un diagrama.

Desde el nacimiento de nuestro alfabeto hasta ahora, la transmisión y almacenamiento de la información se viene haciendo mediante diferentes tipos de documentos. Sin embargo, estos avances en la informática han marcado la tendencia de digitalizar esta información. Por tanto, el objetivo del análisis gráfico de documentos es reconocer las componentes de texto y gráficos en las imágenes, y extraer la información como si de un humano se tratase. De este modo, han surgido dos categorías en el procesado de documentos: el procesado textual y el procesado gráfico, tal y como se indica en la figura 1.1.

El procesado de texto engloba el reconocimiento de texto o reconocimiento de caracteres ópticos (OCR), y también otras tareas tales como determinar inclinación en el documento, detección de palabras, frases, líneas de texto, columnas de texto,...

El procesado gráfico contiene componentes de símbolos y líneas no textuales, líneas rectas entre secciones de texto y logotipos. También se incluyen imágenes, que es la tercera componente más frecuente en los documentos.

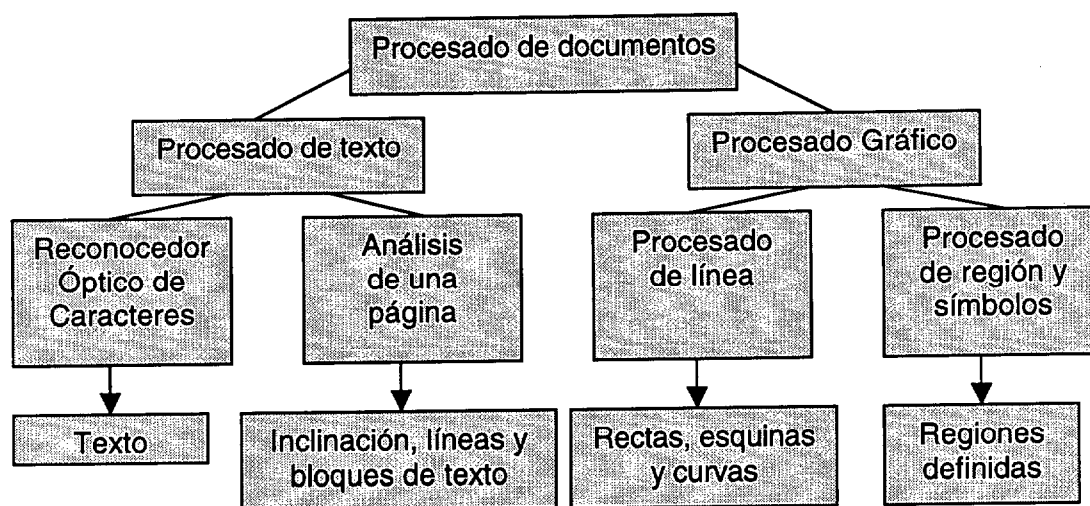


Figura 1.1: Procesado de documentación y sus categorías.

Hoy en día, los resultados de los trabajos de investigación en el análisis de documento se pueden ver y sentir cada día. Por proponer algunos ejemplos se podría mencionar:

- a) Los escáneres de reconocimiento de caracteres son usados en las oficinas de correos para una separación automática del correo.
- b) Los diagramas de bloques son extraídos desde los documentos para almacenarlos y modificarlos desde el ordenador.
- c) Las imágenes de las huellas dactilares son analizadas por el ordenador para su posterior reconocimiento.

En el futuro, las técnicas usadas serán mejoradas y aparecerán más aplicaciones en el análisis de documentos. Por comentar otro ejemplo, los millones de viejos libros que hay en las librerías serán reemplazados por ficheros informatizados, que podrán ser buscados por contenidos y su acceso podrá ser simultáneo para mucha gente. La gente del mundo de negocios podrá tener sus documentos en sus ordenadores portátiles, así como recibirlos y transmitirlos. Y las firmas podrán ser analizadas en los ordenadores para su verificación y acceso de seguridad. Las notas musicales y otros símbolos podrán ser leídos y reconocerse su significado.

1.2.3 Reconocimiento de las aplicaciones manuscritas

Pero todavía se podría llegar más lejos, y distinguir entre documentos impresos por máquinas o aquellos realizados por la propia mano humana. Muchos de los métodos empleados para el reconocimiento en ambos tipos de documentos son similares, pero hay importantes diferencias entre ellos.

Otro matiz significativo en las aplicaciones manuscritas, aparece en la captura de los datos, que se puede plantear de dos formas diferentes: *on-line* y *off-line*. En los sistemas de reconocimiento *on-line*, los datos capturados son representados como una secuencia temporal de la posición marcada por un lápiz óptico sobre una tableta digitalizadora cuando se realiza la escritura. En los sistemas *off-line*, los datos se

capturan mediante un escáner óptico convencional y procesados con un software específico.

Para estas aplicaciones manuscritas sobresale la variedad de formas en que el texto es escrito, básicamente es ilimitada. Incluso se podría observar que el texto escrito por un mismo escritor en dos ocasiones diferentes no es totalmente idéntico. En este momento, el reconocimiento de aplicaciones manuscritas es un área de investigación activa, con una multitud de publicaciones y congresos internacionales por año.

1.3 Reconocimiento automático de formas

Un sistema de reconocimiento automático de formas opera siempre con un universo de trabajo previamente definido. Este universo de trabajo está formado por todos los posibles objetos individuales que se van a reconocer de manera automática. Estos objetos predefinidos suelen denominarse patrones o clases.

Los objetos aparecen normalmente mezclados entre sí, presentando incluso asociaciones e interacciones muy complejas. La operación de extraer los objetos individuales puede llegar a ser extremadamente difícil. Por ejemplo, en visión artificial esta operación se conoce como segmentación, y constituye por regla general la etapa más delicada y difícil de todo el sistema de visión.

Para proceder al reconocimiento automático de los objetos individualizados (segmentados) se realiza una importantísima transformación de los mismos, para convertirlos en vectores cuyas componentes se denominan características o parámetros. A cada vector de características se le aplican unas funciones discriminantes o de decisión, para asignarle al conjunto de vectores preestablecidos de todos los objetos del universo de trabajo, un patrón determinado. Este sistema de reconocimiento presenta tres etapas bien diferenciadas, y con unos conceptos concretos: patrón o clase, parámetro o características, y funciones de discriminación o de decisión. En la siguiente figura se presenta el proceso de reconocimiento automático que se ha descrito.

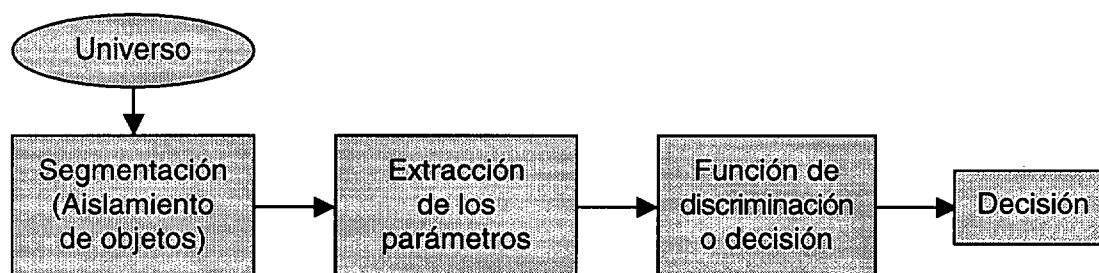


Figura 1.2: Diagrama de bloques de un sistema de reconocimiento automático de formas.

Por tanto, se establecen tres etapas en el diseño de un sistema de reconocimiento automático de formas, y que éstas siguen un curso secuencial a la hora de ponerlo en funcionamiento. Para conseguir éxito en el sistema de reconocimiento de formas hay que revisar en cada momento el buen funcionamiento de la etapa, por eso sería conveniente seguir el siguiente diagrama de flujo:

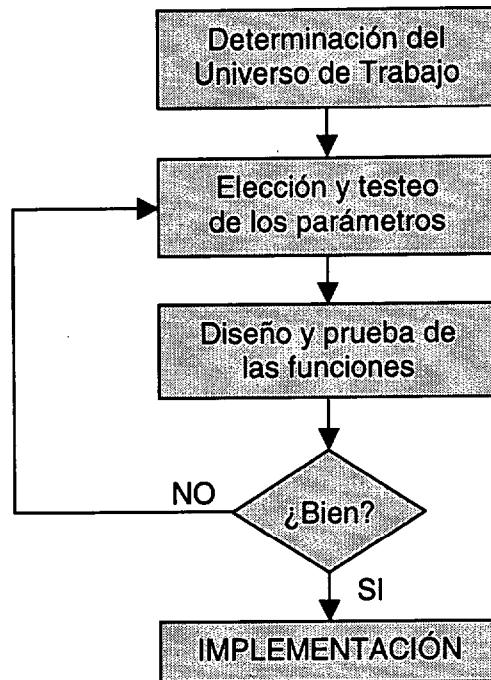


Figura 1.3: Diagrama de flujo que verifica el buen funcionamiento de las etapas, en el diseño de un sistema de reconocimiento automático de formas.

1.3.1 Establecimiento de patrones

El primer paso debe de definir el universo de trabajo del sistema. En la mayoría de los casos reales esta etapa es directa y trivial, ya que el diseñador del sistema conoce perfectamente las clases de objetos que han de ser reconocidos.

Pero también puede ser que los patrones sean desconocidos a priori. Suele aparecer en ciertos campos de la medicina o la biología, en donde no está clarificado el universo de clases. En esta situación se recurre a las técnicas denominadas de agrupación, o expresado más gráficamente, de reconocimiento sin supervisión. Estas técnicas difieren de las expuestas con anterioridad, ya que no existe una supervisión o reconocimiento externo, que guía el diseño de las funciones de discriminación.

Estas técnicas de reconocimiento sin supervisión también se pueden emplear como alternativa en situaciones con conocimiento previo de los patrones.

1.3.2 Extracción de parámetros

La extracción de parámetros consiste en la elección de las características que se almacenarán en el vector de parámetros. Esta etapa es crítica y la bondad del sistema final estará completamente determinada por los rasgos escogidos.

El vector de características constituye el elemento clave en un sistema de reconocimiento automático de formas. El concepto mismo de vector de parámetros es lo que dota a este reconocimiento su enorme potencial práctico y su eficacia, al reducir de

forma extraordinaria la diversa complejidad de los objetos, a la muy manejable información condensada en un vector de datos numéricos.

Al ser esta etapa, la fundamental en el diseño del sistema, es al mismo tiempo la que menos se presta a la formalización y a la aplicación de reglas generales. La elección de parámetros es muy dependiente de la aplicación concreta que se tenga entre manos; se trata más de un arte que de una ciencia, y es la intuición y la experiencia quiénes pueden guiar su elección.

Por último, conviene destacar las propiedades más importantes que deben barajarse en la elección de los parámetros:

1. Deben ser discriminantes: valores numéricos diferentes para objetos de patrones distintos.
2. Deben ser fiables, es decir, con cambios numéricos pequeños para objetos de un mismo patrón. Si esta propiedad no se cumple es necesario pasar a un enfoque estadístico en el diseño de las funciones discriminantes.
3. Deben estar incorrelados. La independencia de los parámetros equivale al principio de la parsimonia, decir lo mismo con la máxima economía de términos.
4. Deben calcularse en un tiempo aceptable. Este es un requisito que puede llegar a ser determinante en ciertas aplicaciones de tiempo real.

1.3.3 Funciones discriminantes o de decisión

El problema del reconocimiento de patrones se caracteriza por la discriminación o clasificación de un conjunto de objetos. El número de tipos de patrones se determina según sea la aplicación, como puede ser el caso de las letras del alfabeto del Español que contiene veintisiete letras. También se puede dar el caso en que se quieran incluir cada vez más patrones a reconocer, como puede ser la unión de los alfabetos rusos y español. Esta tarea que los humanos realizamos en casi todos los instantes de nuestras labores, es lo que los ingenieros y científicos empiezan a usar en máquinas para esta tarea de reconocer patrones.

Una manera intuitiva para el reconocimiento de patrones es el uso de plantillas para comparar. Se debe disponer de una para cada tipo de patrón que se quiere identificar, y la clasificación se basa en un criterio de similitud o preselección. Este sistema ha sido usado en caracteres impresos y en lectoras de tarjetas, pero su desventaja es la dificultad para seleccionar una buena plantilla para cada patrón y definir un buen criterio para el caso de grandes variaciones y distorsiones en un patrón determinado, aunque se hayan propuesto el uso de técnicas de plantillas flexibles.

Posteriormente se han utilizado un sin fin de diferentes técnicas matemáticas que han llegado hasta nuestros días. Estas se han agrupado en dos enfoques principales, el enfoque teórico de decisión o estadístico y el enfoque sintáctico o lingüístico. En el enfoque teórico de decisión, los parámetros son extraídos desde los patrones, el reconocimiento de cada patrón se hace por un agrupamiento de las características, es decir, una división según cada patrón que responden a secciones estadísticas. Estas técnicas ya vienen usándose desde los años sesenta, aplicándose al reconocimiento de

caracteres, la clasificación de cultivos, el diagnóstico médico, la clasificación de electrocardiogramas, etc.

1.4 Propuesta: Planteamiento de la teoría

Hasta ahora se ha hecho un planteamiento general de los conceptos teóricos que hay que tener en cuenta dentro del reconocimiento de formas, pero ahora se va a dar el enfoque que ha dado sentido al nacimiento de esta tesis.

Si se indaga en la literatura y en las publicaciones existentes sobre el reconocimiento de aplicaciones manuscritas, se puede llegar a la conclusión de que las principales herramientas usadas en la extracción de los parámetros de estas aplicaciones son de carácter geométrico, y por tanto se usa un clasificador que se ajusta mejor a la discriminación de este tipo de parámetros, como son las redes neuronales y cualquier otro tipo de transformadas espaciales. La idea que se propone es usar un clasificador que siga las secuencias, rompiendo con los moldes del uso exclusivo de la información geométrica, e introduciendo este tipo de información, la contenida en la envolvente, a partir del trazo de la secuencia de escritura, para establecer un posible punto de partida en el reconocimiento de las aplicaciones manuscritas.

Por tanto, hay que indagar en la literatura para encontrar una función de decisión eficaz ante las secuencias. En la bibliografía de reconocimiento, se pueden hallar aplicaciones desarrolladas en procesado de voz, que usan los modelos ocultos de Markov (HMM) como reconocedor. Estos vectores que se poseen no distan mucho de la información contenida en un vector de voz, ya que es una información secuencial, de una dimensión medida a lo largo del tiempo, registrando las variaciones de la secuencia que se producen. Por tanto, se utilizará los modelos ocultos de Markov como reconocedor de estas aplicaciones manuscritas.

Tal y como se ha comentado con anterioridad, la extracción de las características de los patrones es un campo muy amplio, y casi es cuestión de un híbrido entre la intuición y la experiencia, la que nos podrá indicar el éxito de los vectores de parámetros, y a partir de ahí, utilizar una función de decisión que discrimine bien los diferentes patrones. En esta tesis se ha decidido aplicar una extracción de parámetros a las aplicaciones de documentación manuscrita, pero es difícil afirmar si por la aplicación surgió la idea de la extracción, o si por la extracción de parámetros que se propondrá se le buscó la aplicación. Quizás se podría mantener como una asimilación de ideas duales, es como entrar en el dicho, *“que fue primero sí el huevo o la gallina”*. Pero lo importante es que ha surgido esta idea de extracción de parámetros para aplicaciones manuscritas.

Por tanto, la idea que se propone va dirigida a las aplicaciones manuscritas, lo que se había denominado patrones o universo de trabajo. Veamos que en este caso se va a aplicar a dígitos manuscritos, letras minúsculas y mayúsculas manuscritas, a palabras manuscritas y a firmas manuscritas; por lo que se han creado una base de datos para cada una de estas aplicaciones. Los patrones para cada una de ellas vienen determinados por la propia aplicación. En el caso de los números, el espacio de las muestras cubre desde el “0” hasta el “9”; para las letras minúsculas van desde la “a” hasta la “z”, y otro tanto de lo mismo para las mayúsculas, desde la “A” hasta la “Z”. Para las palabras se ha usado un número de palabras restringido, que son todas las posibles que pueden

indicar una cantidad en un cheque bancario, y por último las firmas también se han tomado un número de firmas restringido.

Para este tipo de patrones se ha escogido un sistema de reconocimiento *off-line*, ya que ofrece una captura de los datos mucha más económica mediante un escáner óptico convencional.

Una vez definido los patrones, la extracción de parámetros se realiza usando la información que existe en el trazado de la propia escritura. Se va a crear un vector con la información secuencial de cada patrón, que se tiene de cada una de las aplicaciones manuscritas. Un detalle importante es que los datos de los patrones son recibidos en forma de imagen es decir bidimensional. En esta tesis se propone que las imágenes sean vectorizadas, es decir, pasadas a una dimensión, por medio de una codificación aplicada a los datos en dos dimensiones. De esta forma, se obtienen unos vectores que siguen la secuencia de escritura realizada sobre un plano bidimensional.

En resumen, se propone modelar las aplicaciones manuscritas mediante un HMM, ya que clasifica muy bien las secuencias, utilizando para ello, el carácter secuencial que tienen las formas manuscritas a ser escritas.

1.5 Objetivos de la tesis

El principal objetivo de esta tesis es estudiar el alcance los modelos ocultos de Markov para modelar patrones de formas manuscritas para el reconocimiento de los mismos. Para conseguir este objetivo se deberán alcanzar también otros objetivos secundarios, pero que en su conjunto engloban este objetivo principal.

Entre estos objetivos secundarios se describe en primer lugar el realizar unas bases de datos que sean fiables e independientes de los escritores para mostrar las tasas de reconocimiento. Se han construido cuatro bases de datos, una de dígitos, otra de letras mayúsculas y minúsculas, otra de palabras, y una última firmas. Aunque dentro del desarrollo del reconocimiento de las palabras se harán mención a las cadenas de palabras tales como determinación de códigos postales manuscritos en cartas o la segmentación de palabras en letras. Todas estas muestras de entrada, nos permitirá probar el objetivo principal.

Otro objetivo secundario es adecuar unas técnicas estándar para todas las aplicaciones en la extracción de los parámetros aplicadas al HMM, con la finalidad de generalizar en la medida de lo posible una misma herramienta de clasificación.

Estos objetivos se materializarán mediante diferentes tipos de algoritmos y programas, desde la lectura de los datos capturados en las bases de datos, pasando por las extracciones de parámetros y finalmente implementando los clasificadores. Todos estos algoritmos fueron realizados para trabajar en el entorno de trabajo de MATLAB.

1.6 Estructura de la memoria

En este primer capítulo, se recoge una amplia visión del reconocimiento de formas hasta llegar al reconocimiento de aplicaciones manuscritas. También se ha planteado el

motivo de la tesis y las investigaciones que se pretenden estudiar, así como los objetivos que se persiguen. Y por último, se ha descrito esta estructura de la memoria de la tesis.

En el segundo capítulo se ha tratado el estado del arte actual del reconocimiento de formas manuscritas en general, y se ha particularizado para cada una de las aplicaciones comentadas en los objetivos. También contiene una breve descripción de los clasificadores actuales, profundizando en los modelos ocultos de Markov.

El tercer capítulo recoge la creación de las bases de datos de formas manuscritas simples como son los dígitos y las letras, tanto mayúsculas como minúsculas. Este proceso se explica detenidamente ya que servirá para las siguientes bases de datos. También incluye el preprocesado de imagen desde su digitalización hasta la preparación para la extracción de parámetros, para estas aplicaciones de manuscritos simples.

En el siguiente capítulo se ha tratado la extracción de parámetros de los dígitos y letras, en particular, como se trabaja con imágenes. Se plantea todo el recorrido del procesado de imagen que se realiza hasta llegar a la obtención de los vectores de los patrones. Es decir, describe el procesado de imagen para extraer los parámetros secuenciales, relatando los métodos de parametrización utilizados, así como la vectorización final de los patrones. Finalmente, se exponen los modelos ocultos de Markov como clasificador de las aplicaciones de dígitos y letras manuscritas, así como el ajuste del mismo para mejorar la tasa de reconocimiento.

El capítulo quinto trata sobre las cadenas manuscritas, que son las palabras y son tratadas como un sólo elemento. Se realiza desde la creación de la base de datos, el preprocesado de imagen, la extracción de parámetros, la vectorización de los mismo, hasta el ajuste del clasificador para obtener la mejor tasa de reconocimiento. Se incluye también en este capítulo, las aplicaciones de reconocimiento de código postal manuscrito y reconocimiento de palabras mediante la segmentación de cada uno de los miembros de la cadena.

En el sexto capítulo se incluyen las formas manuscritas complejas, en particular se han estudiado las firmas. El proceso para su reconocimiento es el mismo que el capítulo anterior, creación de la base de datos, parametrización y ajuste del clasificador para obtener el mayor éxito.

En el último y séptimo capítulo, se analizan las conclusiones obtenidas durante el desarrollo de la tesis y se plantean las posibles líneas futuras de trabajo.

Capítulo 2

Estado de la técnica en el reconocimiento de aplicaciones manuscritas

2.1 Introducción

Este capítulo trata de describir el estado de la técnica de las aplicaciones manuscritas, centrándose en dos vertientes claramente definidas. Por un lado, un sondeo general de cual son las técnicas y métodos en la extracción de parámetros o características que se emplean para el reconocimiento de las diferentes aplicaciones, que esta tesis pretende estudiar, así como la tasa de éxito que se está obteniendo. Y por otro lado, una descripción de las funciones de decisión existentes y su empleo para el reconocimiento de patrones, particularizándolo para el reconocimiento de las aplicaciones manuscritas.

2.2 Revisión del estado de la técnica

Desde los años setenta se pueden encontrar diferentes tipos de técnicas utilizadas para el reconocimiento de patrones. A partir de entonces, este campo se ha diversificado extendiéndose a diversas áreas, como el reconocimiento de objetos, el análisis del movimiento, el reconocimiento biométrico, el análisis de documentación, etc. Y cada vez, los métodos utilizados para la extracción de la información están siendo más perfeccionados y variados, entre ellos se puede citar como extracción de selección de características los diversos dominios transformados, las medidas geométricas, las texturas, los colores, la morfología, etc. También se han utilizado diferentes tipos de técnicas para manipular esta etapa del proceso de reconocimiento, como la segmentación, la codificación e incluso la compresión de la información.

Centrando la atención en el reconocimiento de las aplicaciones manuscritas y para analizar el estado actual de las investigaciones, se pueden empezar por distinguir entre los tipos de sistemas *on-line* y los sistemas *off-line*. Prácticamente las técnicas que se utilizan en las funciones de toma de decisión son utilizadas por igual en ambos sistemas, pero debido a las diversas formas de captar los datos, la extracción de parámetros en ambos sistemas es diferente.

Para las aplicaciones manuscritas que se conocen, cuyos datos han sido adquiridos con un sistema *on-line*, se usan los parámetros dinámicos de las formas, tales como la duración del proceso, la inclinación del bolígrafo óptico al efectuar la escritura, la presión ejercida por el elemento usado para escribir, la velocidad del trazo, etc. Para obtener estos parámetros se utilizan tabletas digitalizadoras y bolígrafos ópticos [O'Gorman95][Maravall93][Knerr92].

En las aplicaciones provenientes de sistemas *off-line*, la extracción de parámetros se centra principalmente en la extracción de características geométricas que intente definir un tipo de clase determinado para su discriminación con respecto al resto de las clases de patrones.

Para desarrollar esta tesis, se ha fijado cuatro aplicaciones manuscritas *off-line* diferenciadas por su grado de complejidad. Por un lado se tienen dos aplicaciones referidas a patrones sencillos como las letras y los dígitos. Si se aumenta la dificultad aparece otra aplicación, como combinación de las dos anteriores, dando lugar a las cadenas manuscritas para formar palabras. Y por último la aplicación más complejas son las firmas, como un agrupamiento de palabras o trazos, o el conjunto de ambos. Un detalle a destacar se tiene con el número de patrones que componen cada aplicación, ya que es diferente. En las letras, dígitos y palabras se proponen conjuntos cerrados y limitados de patrones, y en firmas este grupo es ilimitado.

Antes de pasar a ver con detalle el desarrollo de las aplicaciones, se va a realizar un repaso del estado del arte de cada una de estas cuatro aplicaciones.

2.2.1 Revisión del estado de la técnica en el reconocimiento de los dígitos manuscritos

Para el caso particular de reconocimiento de dígitos manuscritos las técnicas que se han venido utilizando hasta ahora, están caracterizadas por el uso de la información

geométrica que contiene el propio dígito. La forma de extraer los parámetros es muy diversa y se pueden encontrar infinidad de métodos para obtenerla.

En conjunto de patrones para esta aplicación está limitado a diez valores, que son cada uno de los dígitos que se conocen en la actualidad en nuestro sistema decimal de representación numérico.

La extracción de parámetros por medio de las máscaras de Kirsch, es un método bastante usado, como se puede obtener en las referencias [Knerr92][Cho97]. Finalmente estos parámetros se introducen en una red neuronal, que puede estar compuesta por una sola capa como en [Knerr92], o por varias capas como en [Cho97]; pero lo que si es común es que la red neuronal es del tipo perceptrón multicapa o MLP. La tasa de reconocimiento en estos casos suele estar en torno al 96% de éxito.

En otras ocasiones se utiliza una reducción del propio dígito como vector de parámetros a la entrada del propio clasificador. Este sistema es utilizado en las redes neuronales convolucionales, que se obtiene a partir de los perceptrones, y a los que se le suele denominar Neocognitrón y Neoperceptrón. Este tipo de clasificador devuelve resultados entorno al 98% de éxito en la decisión [Neubauer96].

El vector de parámetros de dígitos manuscritos también se puede calcular por medio de un procesamiento de imagen, que calcula el contorno de dicho dígito; y a partir de dicho contorno se determina cuales son los descriptores de Fourier. Estos descriptores forman el vector de parámetros que es discriminado mediante una función que compara el modelo prototipo con el posible dígito a reconocer. Distingue hasta un total de 212 modelos de patrones diferentes, para reconocer los diez dígitos, y con una tasa de reconocimiento ligeramente mayor al 99%. Hay que hacer mención que esta clasificación se ha hecho con dígitos que dependían de los modelos de patrones con los que se habían creado [Jeong99].

Otro parámetro utilizado en la creación de los vectores de características es la geometría espacial mediante las variaciones en la dirección, y después se introducen en un clasificador de redes conexionadas donde se incluyen aprendizajes complejos para optimizar el reconocimiento, y alcanzar tasas de hasta el 96% [Shastri95].

También se ha utilizado la transformada Karhunen-Loève para extraer los parámetros para los vectores. Estos están compuestos por los coeficientes de la transformada y finalmente son clasificados mediante una red neuronal perceptrón multicapa, obteniendo una tasa de reconocimiento de entorno al 96% [Grother92].

Existe otro método que dilata o erosiona las diferentes áreas de una imagen de un dígito. El resultado se obtiene mediante la creación de una rejilla de muestreo, que define el nuevo tamaño de las áreas que lleguen a contactar con esa rejilla, una vez superpuesta a la imagen del dígito. El vector de parámetros se obtiene a partir de las concavidades que se presentan en el nuevo dígito, que se depositarán a la entrada de una red neuronal del tipo perceptrón multicapa. Los resultados obtenidos en la etapa de decisión rondan entorno al 98% [Gosselim97].

No solamente se profundiza en la extracción de parámetros geométricos, sino también, en la optimización de los clasificadores de vectores con parámetros

geométricos. De esta forma, se puede encontrar en la literatura, métodos para mejorar las estructuras de redes neuronales, como el uso de algoritmos genéticos que optimizan los pesos calculados para cada red. En [Schäfer95], se puede observar que la tasa de reconocimiento para este caso es superior al 98%. Otro método que optimiza el funcionamiento es mediante la creación de subclases antes de pasarle los vectores de entrada a la red neuronal. La forma de crear las subclases es a partir de los vectores de entrada, según se contengan en unas ciertas máscaras, se crearán un vector a partir de términos definidos, y éstos son enviados a la red neuronal. En esta aplicación la tasa de reconocimiento encontrada superaba el 99% [Kawajiri91].

En otros métodos, se intenta mejorar la función de decisión, como es el caso de [Schölkopf95], donde se implementa el algoritmo del vector de apoyo para minimizar el riesgo en las funciones de los clasificadores polinomiales, RBF y perceptrón multicapa. Este algoritmo se basa en una estructura de conjuntos de hiperplanos, donde cada uno corresponde a la pareja entre un conjunto de vectores y un cierto espacio, como cada hiperplano es una función de decisión, se toma la que minimiza el riesgo de la decisión. En este caso la tasa de reconocimiento está entorno al 96% [Schölkopf95].

Las redes neuronales de tipo RBF también se han utilizado para el reconocimiento de números manuscritos [Hwang97]. Se han determinado los vectores de parámetros mediante una función que calcula las variaciones en las direcciones, y se han creado varios subconjuntos para cada dígito. Entre mayor es el número de subconjuntos mejor resultado se obtenía. La variación de la tasa de reconocimiento va entre el 95% y el 98%, según dicha agrupación [Hwang97].

Según la clasificación que se le aplique, se pueden encontrar otro método para reconocer los dígitos manuscritos. Este método usa unos prototipos mediante curvas que establecen los diferentes dígitos, y mediante un clasificador estadístico, implementado por medio de una función de decisión, realiza la clasificación [Hastie94], estableciendo unas tasas de reconocimiento entorno al 97%.

Los modelos ocultos de Markov, también se han venido utilizando como clasificador en el reconocimiento de los dígitos manuscritos en el sistema on-line de captura de datos, y que han venido desarrollando su gran utilidad en el reconocimiento de voz. Aunque de forma mínima se ha utilizado para en los sistemas off-line [Proter98]. La extracción de parámetros la realiza tomando los valores en filas y después en columna de la imagen del dígito. Estos valores se introducen al clasificador HMM y posteriormente, se realiza un sistema híbrido multiplicando las probabilidades de salida, y se logra un reconocimiento del 93%.

Después de leer estos pequeños resúmenes de cada referencia, la principal idea que surge es el uso casi masivo las aplicaciones de NN con vectores de parámetros geométricos, para estimar el reconocimiento de los dígitos. Y es casi insignificante el uso del clasificador HMM, en aplicaciones manuscritas para los sistemas *off-line*.

2.2.2 Revisión del estado de la técnica en el reconocimiento de las letras manuscritas

Los trabajos realizados sobre el reconocimiento de letras manuscritas con un sistema *off-line*, se pueden separar en estudios de letras solamente mayúsculas, solamente

minúsculas ó el sistema mixto, con mayúsculas y minúsculas a la vez. En las siguientes referencias aparecerán las diferentes formas de abordar el reconocimiento de las letras, y un aspecto a destacar es la diversidad de las técnicas abordadas para la clasificación. Un ejemplo de uso del reconocimiento de letras es para reconocer palabras, segmentándola y aislando sus letras para su clasificación.

En esta aplicación, el universo del espacio de patrones se ve limitado al número total que representan las letras, que para el alfabeto español, se centran en 27 letras, tanto para mayúsculas como para minúsculas. Aunque hay que recordar que otros alfabetos tienen distinto número de letras, en particular la lengua inglesa tiene 26 elementos, una menos que el castellano, nuestra “ñ”.

En [Rizki90] se puede encontrar un detector de parámetros geométricos aplicado a las letras mayúsculas. Se aplican distintos tipos de operadores morfológicos, de erosión a las imágenes de las letras y posteriormente se calcula las coordenadas en polares de los valores de los píxeles resultantes. Finalmente se calcula la función de distribución de la suma de los dos términos en polares, angular y radial. La evaluación de estos parámetros registra una tasa de reconocimiento del 75%.

Otra técnica para el reconocimiento de letras mayúscula, minúscula y mixto se introduce en [Seiler96]. Las técnicas de extracción de características utilizadas son con histogramas verticales, cruces de líneas, lazos en la parte superior, media y baja, ciruelas y extremos en la parte izquierda y derecha, cortes en la parte izquierda y derecha y doce características de las direcciones. Estas características son modeladas por un clasificador geométrico híbrido para establecer las tasas de reconocimiento, que se recogen en la siguiente tabla.

Sistema	Tipo de letra	Reconocimiento
<i>Off-line</i>	Mayúscula	93.6 %
	Minúscula	67.5 %
<i>On-line</i>	Mayúscula	92.5 %
	Minúscula	72.3 %

Tabla 2.1: Tasas de reconocimiento del clasificador en la referencia [Seiler96].

Se recoge un sistema de extracción de parámetros a partir de las transformadas de Karhunen-Loève, clasificándose con Redes Neuronales, este sistema de clasificación es capaz de reconocer letras mayúsculas con una tasa del 84.5%, y letras minúsculas con un éxito del 75.3% [Garris95], a partir de la base de datos del NIST SD-19 [Grother95].

En [Provan95] se utiliza un algoritmo que optimiza el aprendizaje de redes bayesianas para el reconocimiento de las letras, lo hace a partir de escoger el máximo de las probabilidades de todas las características que se introducen. De esta forma llegar a reconocer al 82.5% de las letras manuscritas mayúsculas.

La selección de los parámetros es un punto importante para discriminar las diferentes clases a reconocer. En [Kumar99], se utiliza una función para ver cuales son las características más relevantes, y a partir de las más discriminantes se forma un vector, que finalmente a partir de la función de densidad de probabilidad de cada conjunto de características se discrimina. Este sistema da una tasa de reconocimiento

para letras mayúsculas del 79.3% para una red neuronal del tipo perceptrón multicapa (MLP), y un 86.6% para el clasificador Bayesiano.

Los clasificadores basados en modelos ocultos de Markov (HMM), también se han utilizado para el reconocimiento de letras manuscritas, de forma mixta, en mayúsculas y en minúsculas [Brakensiek99]. Se ha realizado para el idioma alemán y con cuatro escritores, usando una extracción de características basadas en la cuantificación del mapa de bits y en la transformada coseno discreto (DCT). Y como clasificador se ha utilizado HMM discreto y continuo. Las tasas de reconocimiento que se han obtenido se muestran en la siguiente tabla;

Clasificador	Extracción de parámetros	Tasa de éxito
HMM Discreto	Cuantificación	87.1
	TCD	86.7
HMM Continuo	Cuantificación	79.7
	TCD	83.7

Tabla 2.2: Tasas de reconocimiento del clasificador en la referencia [Brakensiek99].

En general, se puede comentar del estado actual de la tendencia del reconocimiento de letras, que se utilizan en su generalidad, la extracción de parámetros geométricos, y casi siempre aplicados a redes neuronales, y casi en menor medida el uso de clasificadores basados en HMM.

2.2.3 Revisión del estado de la técnica en el reconocimiento de las palabras manuscritas

Este campo está mucho más en expansión que los anteriores, debido al uso más práctico de la aplicación del OCR (*Object Character Recognition*), y prueba de ello es la cantidad de información que se puede encontrar al respecto.

En la referencia [Seiler96], se plantea el reconocimiento de palabras a partir del mismo sistema, mediante la extracción de parámetros de tipo geométrico y utilizando un clasificador híbrido. Este sistema propone el uso de diccionario y sin él, para su reconocimiento, de esta forma, las tasas de éxito obtenidas fueron;

Sistema	Tipo de letra	Sin Diccionario	Con Diccionario
Off-line	Mayúscula	70.5 %	96.7 %
	Minúscula	16.5 %	73.1 %
On-line	Mayúscula	61.4%	96.5 %
	Minúscula	21.3 %	82.2 %
Combinado	Mayúscula	---	87.5 %
	Minúscula	---	85.5 %

Tabla 2.3: Tasas de reconocimiento del clasificador en la referencia [Seiler96].

Igualmente, en la referencia [Garris95], el artículo se plantea el método no sólo para reconocimiento de letras comentado anteriormente, sino también para el de palabras, extrayendo los parámetros con la transformada Karhunen-Loève, y clasificándolos con redes neuronales. De esta forma se obtiene una tasa de éxito del 60.5% para un total de 38 palabras de lengua inglesa.

Otro sistema de reconocimiento es mediante el uso del filtrado de Gabor [Busel94], que pasa al dominio de la frecuencia y es un filtro paso banda que toma orientaciones específicas que lo diferencia de los parámetros lineales. El clasificador es conocido por CRG y es una estructura de árbol binario de decisión. Las tasas de reconocimiento encontradas están entre el 54% y el 81%, según el ajuste del clasificador para un total de 12 palabras diferentes.

Hay dos referencias que utilizan como clasificador un HMM, aunque los parámetros escogidos se basan en la geometría de la palabra. En [Chen94] se utilizan 35 características diferentes, tales como momentos de la imagen en dos dimensiones, características topológicas y geométricas, distribución de píxeles y líneas de referencia. Posteriormente, se segmenta la palabra y se le aplica un diccionario, ya que el conjunto de palabras a reconocer es de 627 palabras, realizándolo con un éxito del 89%. En cambio, en [El-Yacoubi99], se utilizan como parámetros los lazos ascendentes y descendentes y el histograma bidimensional. Esta aplicación se realizó para ayudarse del código postal en el reconocimiento de ciudades, y utiliza a éste último como diccionario para estimar la palabra. La tasa de reconocimiento oscila entre 87% para 1000 palabras y el 99% para 10 palabras.

Entre los diversos clasificadores, se ha encontrado el uso del sistema híbrido formado por HMM y MRF (Campo aleatorio de Markov), que devuelve una tasa de éxito entre el 80.5% y el 85.7% [Saon97]. Este sistema fue implementado para una base de datos formada por 26 palabras de lengua francesa, formada por la escritura de números para cheques bancarios.

Otro sistema que utiliza MRF [Choisy00], extrae los parámetros a partir de los vecinos más cercano (píxeles), una vez se haya escalado en tamaño y se haya corregido su inclinación. No se basa en la segmentación de la palabra sino que utiliza la palabra completa para extraer los parámetros. Sobre un conjunto de 25 palabras diferentes, es capaz de tener un éxito del 83.1%.

La mayoría de las referencias que se pueden encontrar entorno al reconocimiento de palabras, se centran en lenguaje hablado, alfabeto escrito impreso y el alfabeto escrito manuscrito con la captura de datos mediante los sistemas *on-line*. Pero sigue existiendo una cierta cantidad de referencias sobre el reconocimiento manuscrito con sistema *off-line*. Para este tipo de caso, se utilizan parámetros geométricos, y en menor medida parámetros de secuencias temporales; y aplicados a diversos tipos de clasificadores, entre los que destacan los modelos ocultos de Markov (HMM).

Dentro de los diferentes planteamientos de reconocimiento, se puede realizar según el tamaño del vocabulario, planteado principalmente para dos casos: un número de palabras reducidas sin utilización de diccionario y un gran tamaño de vocabulario para aplicaciones con diccionario. Incluso se han encontrado referencias, que utilizan técnicas sintácticas y de léxicas para el reconocimiento de texto manuscrito [Srihari95].

Otra opción es cómo se ha detectado la palabra, si segmentando letra a letra o por el contrario, por el conjunto de la misma. De una forma u otra, se hace notar que es un campo todavía bastante latente y en constante perfeccionamiento, ya que las tasas de reconocimiento todavía no llegan a garantizar buenos resultados en el conjunto de las referencias.

2.2.4 Revisión del estado de la técnica en el reconocimiento de las firmas

La última de las aplicaciones manuscritas, en las que se va a revisar el estado de su investigación, es el reconocimiento de firmas manuscritas. Para este caso de reconocimiento con sistemas *off-line* se siguen dos vertientes: la identificación, que consiste en reconocer a un firmante de entre un grupo de posibles firmantes. Y la verificación, que se parte de la firma y la identidad del firmante, y se debe decidir si es auténtica o no. Para las firmas, el conjunto de patrones a discriminar, puede ser tan grande como se quiera, y viene dado por el programador que define este conjunto, en función de la potencia computacional disponible y la eficiencia de su sistema automático de reconocimiento o verificación.

En [Kaewkongka99], se puede encontrar un sistema de verificación de firmas a partir de la transformada Hough. Se parte de la esqueletización de la firma, y se le aplica dicha transformada, tomando los parámetros en coordenadas polares. La clasificación se realiza con una red neuronal back-propagation, obteniendo una tasa del 95.28%.

Una propuesta de verificación de firmas con un sistema *off-line*, se basa en la transformada wavelet de la descomposición de los datos del contorno de las curvas en señales multiresolucionales. Esta transformada nos sirve para marcar el umbral de las medidas estadísticas tomadas. Los resultados experimentales dan una tasa de éxito del 91.71% para firmas con simbología del alfabeto latino y un 93% para firmas con simbología del alfabeto chino, para una base de datos de 25 escritores por simbología [Deng98].

Una de las formas de extraer parámetros de una firma *off-line* para su posterior verificación, es mediante una escala de grises aplicados al trazo que se ocupa en un conjunto cuadrado de 10 píxeles de lado. Según sea el trazo se le asigna un valor y éste posteriormente forma el vector de características. Este vector se sitúa en la entrada de HMM. Siendo la base de datos de 14 escritores, con 20 repeticiones por escritor, se ha obtenido un éxito del 98.1% [Rigoll98]. También introduce un análisis *on-line* de verificación con el 99% de éxito.

Para los sistemas *on-line*, Bell Laboratories ha desarrollado tres bases de datos, DB1 de 59 firmantes, DB2 de 102 firmantes y DB3 de 43 firmantes. Ha utilizado para parametrizar polígonos normalizados y los tiempo de duración de la escritura, de tal forma que al promediar cinco firmas obtiene una función característica. La clasificación la realiza con el error cuadrático medio con respecto a estas funciones. De esta forma, lograron reconocer un 97%, 98% y 95% respectivamente a cada base de datos [Jain01].

En resumen, se puede concluir que tanto se usa el sistema *on-line* como el sistema *off-line*. Las diferencias entre unos sistemas y otros, suelen darse en los sistemas de vectorización dependiendo del tipo de información que se analice o se procese. Algunos sistemas analizan las características estructurales, más generales en las firmas, tales como tamaño, inclinación, densidad de píxeles en negro. Y otros trabajan con parámetros geométricos, que son más específicos, tales como huecos, curvas, rectas, trazos; así hasta transformar éstas características en un vector. Posteriormente, estos parámetros se sitúan en la entrada de un clasificador, que generalmente es del tipo red neuronal, aunque también se utilizan otro tipo de funciones de decisión.

2.3 Estado actual de las funciones discriminantes o de decisión

El problema del reconocimiento de patrones se caracteriza por la discriminación o clasificación de un conjunto de objetos. El número de tipos de patrones se determina según sea la aplicación, como puede ser el caso de las letras del alfabeto de la lengua castellana, que contiene veintisiete letras. También se puede dar el caso que se desee incluir cada vez más patrones a reconocer, como puede ser la unión de los alfabetos ruso y castellano. Esta labor que los humanos realizamos en casi todos los instantes de nuestras tareas, es lo que los ingenieros y científicos empiezan a usar en máquinas para el reconocimiento de patrones.

Una primera idea intuitiva para este reconocimiento de patrones es una plantilla para comparar. Se dispone de una plantilla para cada tipo de patrón que se quiere identificar, y la clasificación se basa en un criterio de similitud o preselección. Este sistema ha sido usado en caracteres impresos y en lectoras de tarjetas. Pero su desventaja es la dificultad para seleccionar una buena plantilla para cada patrón y definir un buen criterio para el caso de grandes variaciones y distorsiones en un patrón determinado, aunque se han propuesto el uso de técnicas de plantillas flexibles.

Posteriormente se han utilizado un sin fin de diferentes técnicas matemáticas que han llegado hasta nuestros días. Éstas se han agrupado en dos enfoques principales, el enfoque teórico de decisión o estadístico y el enfoque sintáctico o lingüístico.

En el enfoque teórico de decisión, los parámetros son extraídos desde los patrones, el reconocimiento de cada patrón se hace por un agrupamiento de las características, es decir, una división según cada patrón que responden a secciones estadísticas. Estas técnicas ya vienen siendo usadas desde los años sesenta, y aplicadas al reconocimiento de caracteres, clasificación de cultivos, diagnóstico médico, clasificación de electrocardiogramas, etc.

En este enfoque teórico de decisión, la clasificación está basada en un conjunto de medidas, extraídas desde el patrón de entrada. Estos parámetros seleccionados se suponen que deben ser invariantes o menos sensibles a las variaciones o distorsiones que se puedan encontrar en otros patrones. Bajo este propósito se puede plantear un problema a superar. El problema en el reconocimiento de patrones que viene dado por la clasificación, es decir, crear un elemento de decisión en la asignación de cada clase para los patrones de entrada, basado en los parámetros tomados a partir de las medidas seleccionadas. El proceso de la extracción de parámetros ya ha sido comentado en el apartado 2.4.2, y el elemento que realiza la función de clasificación se denomina clasificador. De esta forma, en la siguiente figura se puede observar un simple diagrama de bloques de un sistema de reconocimiento de patrones:

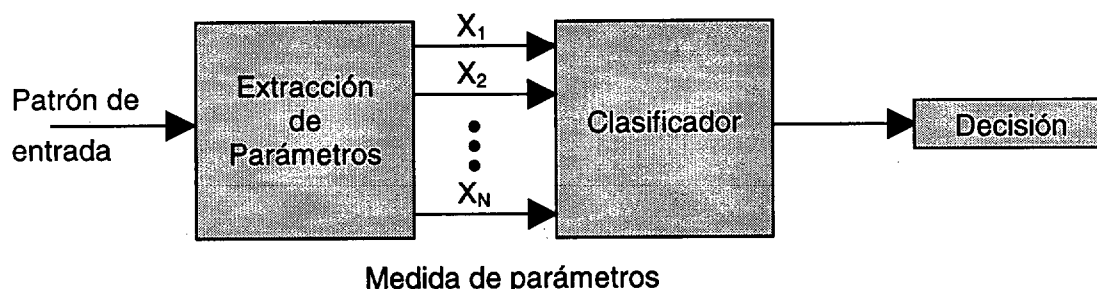


Figura 2.1: Sistema reconocedor de patrones.

El enfoque de plantillas de comparación se debe de contemplar como un caso particular del enfoque sintáctico. En la extracción de parámetros, las plantillas se almacenan según las medidas de los parámetros y se utiliza un cierto criterio para la clasificación.

En el reconocimiento de patrones la información estructural, la cual describe cada patrón es muy importante. Esta información no sólo incluye la capacidad de asignar el patrón a cada clase en particular, sino también la capacidad de describir aspectos del patrón que son inteligibles para la asignación a otra clase. Un ejemplo de esto, es el reconocimiento de figuras o de la voz. Bajo estas condiciones los patrones se consideran bastantes complejos y el número de características a veces suele ser muy grande, que produce la idea de describir los patrones complejos en términos de una composición de subpatrones más simples. Tal es el caso del reconocimiento de huellas dactilares, reconocimiento de rostros humanos, reconocimiento continuo del habla, caracteres chinos, etc. En orden a representar subpatrones más simples se le asigna al estructural o sintáctico.

De forma sencilla, se puede hacer una estructuración de los clasificadores en función de la forma de plantear la decisión. En la literatura se puede encontrar los clasificadores paramétricos y no paramétricos [Fu80][Banks90][Bishop95][Fukunaga90]. Los clasificadores no paramétricos no presupone la forma de la distribución de probabilidad por lo que no requieren la estimación de los parámetros. A cambio, necesitan estimar la densidad de probabilidad de cada observación usando las muestras disponibles en el conjunto de entrenamiento. Para el caso de los clasificadores paramétricos, éstos proporcionan un modelo robusto ya que suponen una distribución de probabilidad subyacente conocida para las clases y tan solo hay que estimar los parámetros de la distribución de probabilidad.

En otro tipo de referencia se puede encontrar los clasificadores como estadísticos y determinísticos [Maravall93][Bishop95][Bow92], según sea la naturaleza de la función de decisión y por tanto, donde se tendrá en cuenta el valor de los parámetros en la propia función de discriminación para tomar la decisión del reconocimiento, o simplemente se usa una función para marcar la frontera de decisión entre las regiones de los diferentes patrones, respectivamente [Maravall93].

Además se puede introducir otra variable, que los clasificadores estén supervisado o no. El concepto de supervisado implica un aprendizaje que requiere de un conjunto de prototipos, que serán el conjunto de entrenamiento, a partir del cual se construye y se evalúa el éxito del clasificador. Para el caso de no ser supervisado, esta construcción no conoce previamente ninguna información del patrón a evaluar, encasillándolo en las regiones establecidas o creando una nueva clase. En resumen, se tiene la existencia de un maestro o supervisor que indica las variaciones en la función discriminante.

La descripción de los clasificadores se establecerán según sean determinísticos o estadísticos, y añadiéndole la variable de supervisados o no.

2.3.1 Clasificadores determinísticos

El concepto de clasificación determinístico de patrones, se expresa en términos de partición del espacio de los parámetros, es decir el traslado desde el espacio de los parámetros hasta el espacio de la decisión. Se puede suponer que N características son extraídas de un cierto patrón, y cada conjunto de N características será el vector X que denominamos vector de parámetros o espacio de parámetros de N dimensiones (Ω_X). El problema de clasificación es asignar cada posible vector o espacio de parámetros a una clase de patrones. Esto se puede interpretar como una partición del espacio de características en regiones exclusivas, y cada región corresponderá a una clase particular. Matemáticamente el problema de la clasificación puede ser formulado en términos de "funciones de discriminación". Se puede designar w_1, w_2, \dots, w_m como los m posibles clases de patrones para ser reconocidos, y en el vector X estará el vector de parámetros, según la siguiente ecuación;

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} \quad (2.1)$$

Donde X_N representa el parámetro N -ésimo medido. Si se parte de la función de discriminación $D_j(X)$ asociada con la clase del patrón w_j , para $j=1, \dots, m$, y si el patrón de entrada representado por el vector de característica X es de la clase w_i , se entiende que X proviene de la clase w_i , y el valor de $D_i(X)$ debe ser el mayor en discriminar, ésto es, para todos los X equivalentes a w_i , como se representa a continuación considerando i diferente de j ;

$$D_i(X) > D_j(X), \quad i, j = 1, \dots, m, \quad i \neq j \quad (2.2)$$

De esta manera, en el espacio de parámetros Ω_X , el límite de partición, también llamado la frontera de decisión, entre las regiones asociadas con clase w_i y clase w_j , se podría expresar según la ecuación 2.3, y se podrían considerar de la misma clase.

$$D_i(X) - D_j(X) = 0 \quad (2.3)$$

En resumen, la función de discriminación determinística plantea que las clases se modelen como vectores o prototipos únicos, en los que se dispone a priori de toda la información necesaria y suficiente para su diseño.

Pero este tipo de clasificador puede llevar un aprendizaje supervisado, de tal manera que existe un maestro que se encarga de enseñar al sistema, corrigiéndolo cada vez que se equivoque; es decir, cuando clasifique incorrectamente a un objeto.

Para ello se dispone de un conjunto de muestras de aprendizaje formado por, P vectores X_1, X_2, \dots, X_P cuya pertenencia a una determinada clase es conocida. Si existen N patrones, $\alpha_1, \alpha_2, \dots, \alpha_N$, entonces $P \gg N$. El conjunto de muestras de aprendizaje se escoge habitualmente de forma que incluya el mismo número de muestras de aprendizaje por cada clase.

Se extrae un conjunto de muestras de aprendizaje o entrenamiento, cuya pertenencia a una clase específica es conocida por el maestro, es decir, por el diseñador del proceso de aprendizaje. La muestra es clasificada por el reconocedor automático en virtud del conjunto de funciones de discriminación que en ese instante esté manejando. Estas funciones discriminantes cambian y lo hacen en función de la señal de error que se haya producido en la decisión del reconocedor: no habrá error (acierto) cuando coincida su decisión con la del maestro y sí existirá error cuando difieran. Esta descripción se expresa en la siguiente figura;

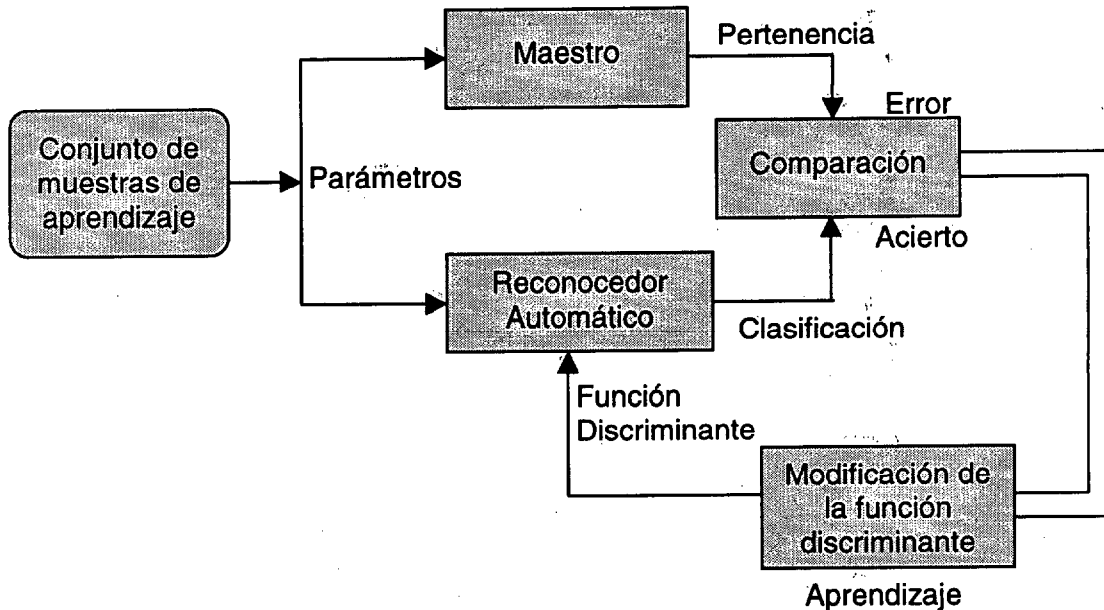


Figura 2.2: Diagrama de bloques simbólico del proceso de aprendizaje o entrenamiento de un reconocedor.

La idea clave para plantear los procesos o algoritmos de aprendizaje consiste en el establecimiento de una meta u objetivo concreto del reconocedor que aprende. Se puede considerar cuando se trabaja con hipótesis de patrones o clases con comportamiento determinístico, dos grandes grupos de algoritmos de aprendizaje:

- a) Aquellos que presentan funciones de decisión basadas en la regionalización del espacio de patrones.
- b) Aquellos cuyas funciones discriminantes manejan el concepto de distancia.

En definitiva, se trata de la misma división que se encuentran para los clasificadores determinísticos en condiciones de conocimiento total a priori. El grupo (a) sólo puede resolver el problema de las distribuciones biclase, mientras que el grupo (b) permite una solución más general de tipo multiclase.

Los mismos clasificadores que se plantean para los tipos no supervisados, se pueden plantear en los clasificadores supervisados, pero lógicamente, con funciones de discriminación que hayan sido previamente enseñadas por el maestro.

2.3.2 Clasificadores estadísticos

En el diseño de un clasificador automático basado en la distancia euclídea se utilizan unos prototipos fijos, representativos de cada una de las clases o patrones, con lo que se está admitiendo una hipótesis determinística en el comportamiento de los elementos de una clase, de tal manera que se asimilan todos sus elementos a un vector único.

En aquellas situaciones en las cuales los vectores de algunas clases presentan una dispersión significativa con respecto a su media es conveniente abandonar la hipótesis determinista en favor del enfoque estadístico.

Por notar un ejemplo gráfico podríamos comentar la siguiente figura:

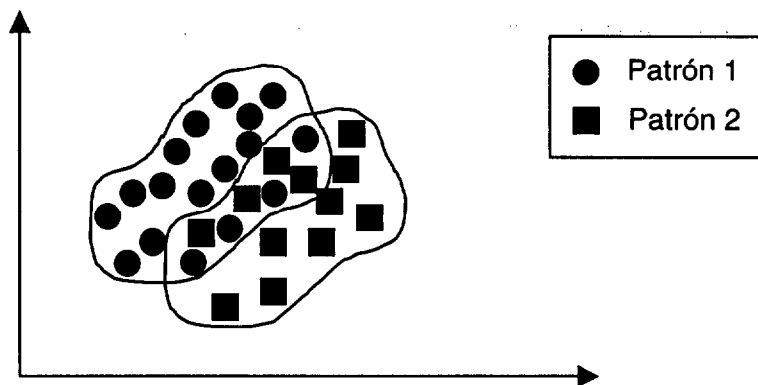


Figura 2.3: Representación en un plano de dos patrones del ejemplo gráfico.

De la figura 2.3 se observa dos peculiaridades, la primera es la elevada dispersión de las muestras lo que presupone un mal uso de las técnicas determinísticas, en detrimento de las estadísticas. Y en segundo lugar, se observa un solape de las clases, lo que hace presagiar otra vez la poca eficiencia de las técnicas determinísticas.

Si se define las distribuciones de estas muestras, mediante su media y su varianza, quedarían totalmente definidos su comportamiento estadístico, es decir se tendría su función de densidad de probabilidad.

Este tipo de clasificador estadístico es fundamentalmente la estimación de las funciones de densidad de probabilidad, ya que la clasificación de los objetos se basa en estas funciones.

2.3.2.1 Clasificador Bayesiano

Se va a considerar X como el vector de parámetros de los N parámetros medidos, lo mismo que se indica en la ecuación 2.1. Y cada clase de patrón se representará como $P(w_j)$, para $j=1, \dots, m$, la cual asume la multivariabilidad (N -dimensiones) en la función de densidad de probabilidad o la función de densidad de distribución del vector de parámetros X , $p(X/w_j)$. Todos estos valores son conocidos al igual que la probabilidad de ocurrencia de w_j , $P(w_j)$. En base a la información conocida a priori de $p(X/w_j)$ y $P(w_j)$, donde $j=1, \dots, m$, la función de un clasificador estadístico, es producir la tarea de clasificación para minimizar la probabilidad del error de reconocimiento. Por tanto, el problema en los clasificadores estadísticos puede radicar en un problema de decisión

estadística (testar las m hipótesis estadísticas) debido a la definición de una función de decisión $d(X)$, donde $d(X)=d_i$ quiere decir que la hipótesis que X es equivalente a w_i es aceptada. La regla de la decisión óptima que minimiza el posible error que se pueda cometer, esta decisión se denomina regla de Bayes.

2.3.2.2 Clasificadores estadísticos supervisados

Dentro de los clasificadores supervisados se pueden encontrar los clasificadores estadísticos, cuya diferencia radica en que el problema del reconocimiento de formas se aborda a través de la estimación de las denominadas probabilidades a posteriori y no a priori como en los casos anteriores.

El aspecto clave en el diseño de los reconocedores estadísticos con aprendizaje estriba en partir de una expresión concreta para las probabilidades a posteriori que se van a estimar o aprender, basándose en las muestras de entrenamiento.

El objetivo del proceso de aprendizaje será obtener unas funciones discriminantes, que devuelvan una probabilidad a posteriori de 1 cuando sea éxito la forma a reconocer, y 0 en caso contrario. El maestro o supervisor del aprendizaje se va a modelar como una variable aleatoria, que tenderá a minimizar el índice de error o el índice de error cuadrático. A esta minimización de los índices de error se le aplica una actualización recursiva de los coeficientes de las funciones discriminantes basadas en el gradiente del error.

En los sistemas clasificación estadística todas las N características se observan por el clasificador en un estado. El coste de la medida de los patrones no ha sido tomada en cuenta, y es evidente que un número insuficiente de parámetros no será capaz de dar resultados satisfactorios. Al igual que un número elevado de parámetros, podría crear un estado de impracticabilidad, debido a un excesivo coste temporal y computacional. Por tanto, en este tipo de clasificador existe el problema del ajuste de la cantidad de parámetros para obtener dicho reconocimiento de patrones.

2.3.3 Funciones de Discriminación

Se pueden encontrar una amplia variedad de funciones de discriminación. En particular se van a describir algunos modelos que se pueden considerar como clásicos, como es el caso de las siguientes funciones:

- Funciones de Discriminación Lineal
- Clasificadores de Mínima Distancia
- Clasificadores por el Vecino más cercano
- Funciones Discriminantes Polinomiales

Aunque podrían incluirse algunas más dentro de este grupo, tanto por su simplicidad, como por su uso. En un segundo término se va a detallar otra función de decisión, como son las redes neuronales, muy utilizadas por las referencias bibliográficas en el estado del arte de las aplicaciones manuscritas. Y por último, y por el uso dado en esta tesis para clasificar, se comentarán los modelos ocultos de Markov.

2.3.3.1 Función de Discriminación Lineal

En este caso, una combinación lineal de los parámetros x_1, x_2, \dots, x_N se selecciona por $D_i(X)$, como se observa en la siguiente ecuación:

$$D_i(X) = \sum_{k=1}^N w_{ik} x_k + w_{i,N+1} \quad i = 1, \dots, m. \quad (2.4)$$

La frontera de decisión entre dos regiones Ω_N asociada con w_i y w_j es de la forma:

$$D_i(X) - D_j(X) = \sum_{k=1}^N w_k x_k + w_{N+1} = 0 \quad (2.5)$$

con $w_k = w_{ik} - w_{jk}$ y $w_{N+1} = w_{i,N+1} - w_{j,N+1}$. La ecuación anterior es la ecuación de un hiperplano en el espacio de los parámetros Ω_X .

Para el caso particular de dos patrones ($m=2$), el umbral lógico del clasificador sigue una función de discriminación lineal según el siguiente esquema, donde la separación de las clases viene dada por los niveles de salida +1 o -1.

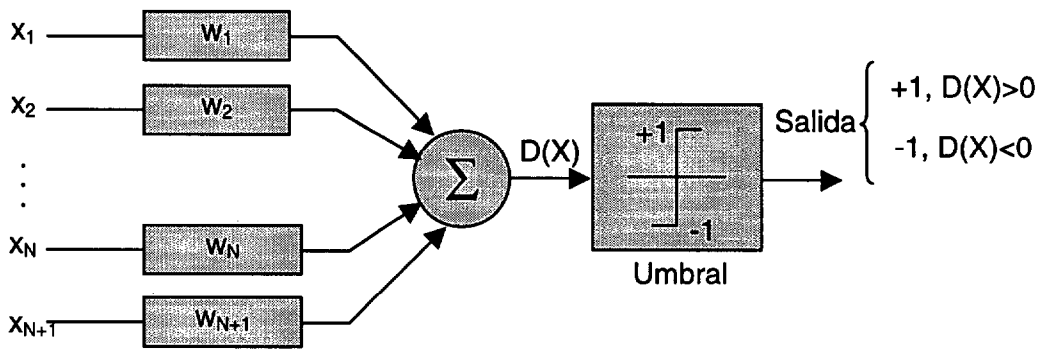


Figura 2.4: Clasificador lineal de dos clases.

Para un número mayor de clases de patrones a clasificar, los umbrales lógicos se conectan en paralelo a partir de las combinaciones de las salidas. Así M umbrales lógicos pueden llegar a distinguir m clases de patrones siguiendo la relación $2^M \geq m$.

2.3.3.2 Clasificador de Mínima Distancia

Un tipo importante de clasificadores lineales no paramétricos son los que usan las distancias entre las entradas de los patrones y un conjunto de vectores de referencia o puntos de prototipos en el espacio de los parámetros como criterio de clasificación. Supongamos que los m vectores de referencia R_1, R_2, \dots, R_m , son dados con R_j asociados con la clase de patrón w_j . Un esquema del clasificador de mínima distancia con respecto a R_1, R_2, \dots, R_m , es para clasificar la entrada de parámetros X a partir de la clase w_i ;

$$X \sim w_i \quad \text{si } |X - R_i| \text{ es el mínimo} \quad (2.6)$$

donde $|X - R_i|$ es la distancia definida entre X y R_i , definida en la siguiente ecuación.

$$|X - R_i| = \sqrt{(X - R_i)^T (X - R_i)} \quad (2.7)$$

donde el superíndice T representa la operación transpuesta a un vector. Así, la función de discriminación $D_i(X)$ viene dada por el producto de los términos en el interior de la raíz, siempre que dependan de la referencia R_i .

2.3.3.3 Clasificador según el Vecino más Cercano

El concepto adoptado en el Clasificador de Mínima Distancia es aplicado para al Vecino más cercano. En este caso, la función de discriminación viene dada por la mínima distancia entre el vector de parámetros de patrón de entrada X , y el vector de referencia R_j . Por tanto, dado el conjunto de los vectores de referencia R_1, R_2, \dots, R_m , asociados a las clases w_1, w_2, \dots, w_m , respectivamente; y dado el vector de referencia R_j expresado por $R_j^{(k)}$ como;

$$R_j^{(k)} \in R_j, \quad k = 1, \dots, u_j \quad (2.8)$$

donde u_j es el número de vectores referencia en el conjunto R_j . Se define la distancia entre el vector de parámetros X y R_j como;

$$d(X, R_j) = \min_{k=1, \dots, u_j} |X - R_j^{(k)}| \quad (2.9)$$

2.3.3.4 Función de Discriminación Polinomial

Una función de discriminación polinomial de orden n -ésimo está formada por un polinomio de orden n -ésimo y se utiliza para establecer la frontera de decisión entre dos patrones. Para el caso particular de $n=2$, la función de discriminación se denomina función de discriminación cuadrática y se representa por una hiperboloide, en casos especiales se pueden incluir hiperesféricas, hiperelipsoides e hiperelipsoides cilíndricos.

2.3.3.5 Redes Neuronales

Las redes neuronales (NN) estiman la función de densidad de probabilidad de los vectores de entrada de los patrones a reconocer, estableciendo regiones de decisión que permiten separar estas funciones de densidad de probabilidad. Es un sistema que establece una relación no lineal entre la salida y la entrada, y cuyas características es la simulación abstracta del sistema nervioso.

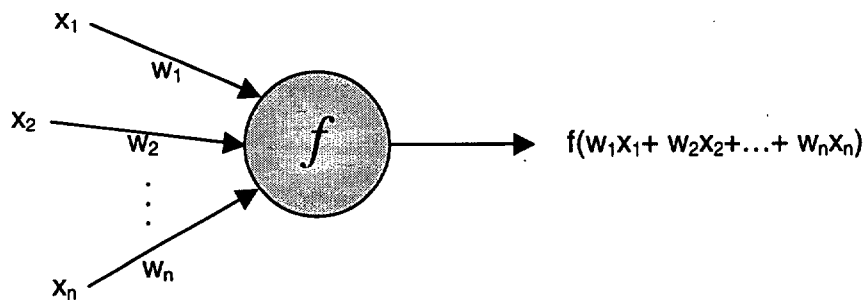


Figura 2.5: Una neurona abstracta.

En la figura 2.5, se muestra la estructura de una neurona con “ n ” entradas. Cada canal de entrada “ i ” puede transmitir un valor real x_i . La función primitiva f del cuerpo

de la neurona abstracta se puede seleccionar arbitrariamente. Generalmente, los canales de entrada tienen asociado un peso " w_i ", al que se multiplica por la información entrante por dicha rama x_i . La información transmitida se integra en la neurona, sumándose las " n " entradas y siendo evaluadas por la función primitiva (f).

Si se concibe cada nodo de una red como una función primitiva capaz de transformar sus entradas en una salida definida, entonces se obtiene una red neuronal artificial. Los diferentes modelos de redes neuronales artificiales se obtienen al variar las funciones primitivas utilizadas, la interconexión entre sus nodos, y el tiempo de transmisión de la información. Pero fundamentalmente, hay tres elementos de particular importancia en un modelo de red neuronal artificial;

1. La estructura de los nodos
2. La topología de la red
3. El algoritmo de aprendizaje utilizado para encontrar los pesos de la red

2.3.3.6 Modelos ocultos de Markov

El acrónimo anglosajón HMM proviene de *Hidden Markov Models* (Modelos Ocultos de Markov). Estos modelos se empezaron a desarrollar a finales de los 60 y, desde entonces, se han aplicado a problemas de clasificación de señales.

Los modelos ocultos de Markov tratan de caracterizar alguna propiedad estadística de la señal. En estos modelos se asume que la señal se puede caracterizar perfectamente como un proceso aleatorio paramétrico y que los parámetros se pueden calcular o estimar.

Los modelos ocultos de Markov (HMM) crean secuencias de estados que forman un modelo, a partir de los vectores de parámetros de la entrada del HMM de cada uno de los patrones que se quieren clasificar. Para llegar a reconocer un vector se calcula la probabilidad de que el vector de entrada corresponda a cada una de las secuencias del modelo, asignando el vector de entrada al modelo que mayor probabilidad obtuvo [Rabiner89].

Las características del Modelo oculto de Markov son, en primer lugar la cadena oculta, que es la estructura oculta; y en segundo lugar el conjunto de observaciones asociadas a cada estado de la cadena, que es la estructura visible.

A continuación se va a realizar una breve descripción de los modelos ocultos de Markov, basada en [Rabiner89].

2.3.3.6.1 Procesos de Markov discretos

Sea un sistema que se puede caracterizar en cualquier instante de tiempo por el estado en que se encuentra, que pertenece a un conjunto de N estados: S_1, S_2, \dots, S_N , según se muestra en la figura 2.6 (donde $N=5$ por sencillez).

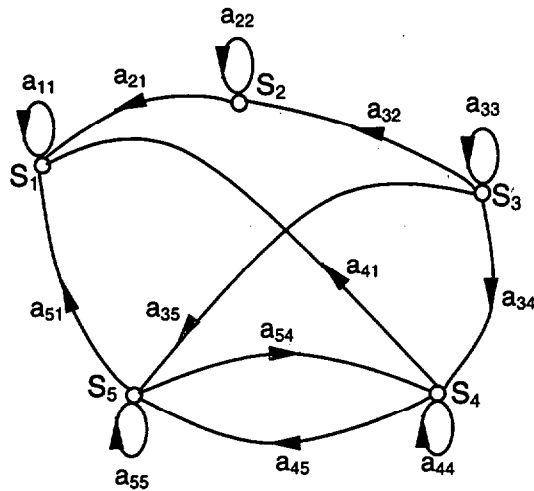


Figura 2.6: Cadena de Markov ergódica.

En instantes discretos de tiempo equidistantes, el sistema cambia de estado, o permanece en el mismo, según un conjunto de probabilidades asociado a cada estado. Los instantes de tiempo en los que se produce un cambio de estado se representan así: $t=1, 2, \dots$; y el estado en el que el sistema se encuentra en el instante "t" se representa por q_t . Una descripción estadística completa de tal sistema requiere, en general, conocer el estado actual (en el instante t) y todos los estados anteriores. Para el caso concreto de una cadena de Markov discreta de primer orden, basta con conocer el estado actual y el predecesor, es decir, cumple la propiedad markoviana:

$$P[q_t = S_j / q_{t-1} = S_i, q_{t-2} = S_k, \dots] = P[q_t = S_j / q_{t-1} = S_i] \quad (2.10)$$

Además, sólo se consideraran aquellos procesos en los que la parte derecha de la ecuación 2.10 es independiente del tiempo. Se define la probabilidad de transición desde el estado S_i al estado S_j como:

$$a_{ij} = P[q_t = S_j / q_{t-1} = S_i] \quad 1 \leq i, j \leq N \quad (2.11)$$

la probabilidad de transición, para ser probabilidad, debe cumplir:

$$a_{ij} \geq 0 \quad (2.12a)$$

$$\sum_{j=1}^N a_{ij} = 1 \quad (2.12b)$$

Se puede decir que el proceso estocástico presentado hasta ahora es un modelo observable de Markov, porque la salida del proceso es un conjunto de estados en cada instante de tiempo, donde cada estado corresponde a un suceso físico observable.

A continuación se muestra un ejemplo de lo dicho hasta ahora: considera un modelo de Markov de tres estados, que indican el tiempo meteorológico. Se observa el tiempo una vez al día, y se asume que las posibles observaciones pueden ser:

- Estado 1: lluvioso.
- Estado 2: nublado.
- Estado 3: soleado.

Se supone que, en un día, se puede hacer una de estas tres observaciones y que la matriz de transición es:

$$A = \{a_{ij}\} = \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$

Se considera que el tiempo el día 1 ($t=1$) es soleado (estado 3), y se pregunta cuál será la probabilidad, de acuerdo con el modelo, de que durante los siete días siguientes el tiempo sea: 'soleado, soleado, soleado, lluvioso, lluvioso, soleado, nublado, soleado'. Se define la secuencia de observación como $O = \{S_3, S_3, S_3, S_1, S_1, S_3, S_2, S_3\}$ correspondiente a los instantes $t=1, 2, \dots, 8$, y se desea determinar la probabilidad de que ocurra O , dado el modelo. Esta probabilidad se puede expresar como:

$$\begin{aligned} P(O/\text{Modelo}) &= P(S_3, S_3, S_3, S_1, S_1, S_3, S_2, S_3/\text{Modelo}) \\ &= P(S_3) \cdot P(S_3/S_3) \cdot P(S_3/S_3) \cdot P(S_1/S_3) \cdot P(S_1/S_1) \cdot P(S_3/S_1) \cdot P(S_2/S_3) \cdot P(S_3/S_2) \\ &= \pi_3 \cdot a_{33} \cdot a_{33} \cdot a_{31} \cdot a_{11} \cdot a_{13} \cdot a_{32} \cdot a_{23} \\ &= 1 \cdot 0.8 \cdot 0.8 \cdot 0.1 \cdot 0.4 \cdot 0.3 \cdot 0.1 \cdot 0.2 \\ &= 1.536 \cdot 10^{-4} \end{aligned}$$

donde se ha presentado la probabilidad inicial de estados como:

$$\pi_i = P(q_1 = S_i) \quad 1 \leq i \leq N \quad (2.13)$$

Otra interesante pregunta que se puede hacer es que, dado que el modelo se encuentra en un estado conocido, S_i , ¿cuál es la probabilidad de que continúe en ese estado exactamente 'd' días? Esta probabilidad se puede calcular a partir de la probabilidad de observación de la secuencia $O = \{S_i(t=1), S_i(t=2), S_i(t=3), \dots, S_i(t=d), S_i(t=d+1) \neq S_j\}$ dado el modelo, cuya expresión es:

$$P(O/\text{Modelo}, q_1 = S_i) = (a_{ii})^{d-1} (1 - a_{ii}) = p_i(d) \quad (2.14)$$

Se conoce a $p_i(d)$ como la función de densidad de probabilidad de permanencia durante un tiempo 'd' en el estado 'i'. Utilizando $p_i(d)$ se puede calcular el número medio de observaciones (duración) de un estado, condicionado a empezar en ese estado:

$$\bar{d}_i = \sum_{d=1}^{\infty} d \cdot p_i(d) = \sum_{d=1}^{\infty} d \cdot a_{ii}^{d-1} \cdot (1 - a_{ii}) = \frac{1}{1 - a_{ii}} \quad (2.15)$$

De este modo, el número medio de días soleados consecutivos es $1/0.2=5$, el de días nublados es 2.5 y el de días lluviosos es 1.67.

2.3.3.6.2 Ampliación a Modelos Ocultos de Markov

Hasta ahora se han considerado modelos de Markov, en los que cada estado corresponde a un suceso físico observable. Este modelo es demasiado restrictivo para poderlo aplicar a muchos problemas de interés. En este apartado se amplía el concepto de los modelos de Markov para incluir aquellas situaciones en las que la observación es una función de probabilidad del estado en que se encuentra el modelo. Este tipo de modelos constituye los Modelos Ocultos de Markov. Los HMM consisten en un proceso estocástico doble en el que uno de los procesos está oculto, y es no observable directamente, y otro proceso que permite observar el anterior [Rabiner89].

Véase un ejemplo: una persona, que no se puede ver porque está situada detrás de una cortina, está realizando un experimento aleatorio que consiste en lanzar una o más, no se sabe, monedas trucadas, e informa del resultado del experimento cuya secuencia de observación podría haber sido:

O=O₁O₂O₃...O_T=CCXXXCXXC...C
 X: cruz, C: cara.

Para construir el HMM que explique la secuencia observada, lo primero es determinar qué significan los estados del modelo para después decidir el número de estados. Una opción es considerar que sólo se lanza una moneda trucada. En esta situación se puede realizar un modelo con dos estados, en el que uno corresponde a salir cara, y el otro a salir cruz (véase la figura 2.7a). Este modelo es observable y el único elemento que queda por determinar es la probabilidad de la cara y de la cruz, que no es 0.5 debido a que la moneda está trucada.

Otro modelo para explicar la secuencia observada se representa en la figura 2.7b, en él hay dos estados y cada uno representa el resultado de lanzar una moneda trucada. Cada estado se caracteriza por la función de distribución de probabilidad de caras y cruces, y las transiciones entre estados se determinan mediante la matriz de transición. El lanzamiento de otra moneda, u otro experimento aleatorio independiente, podría proporcionar la forma de pasar de un estado a otro.

En la figura 2.7c se representa otro modelo para explicar la secuencia observada. En este modelo se emplean tres monedas trucadas y se elige una de ellas a través de otro experimento aleatorio independiente.

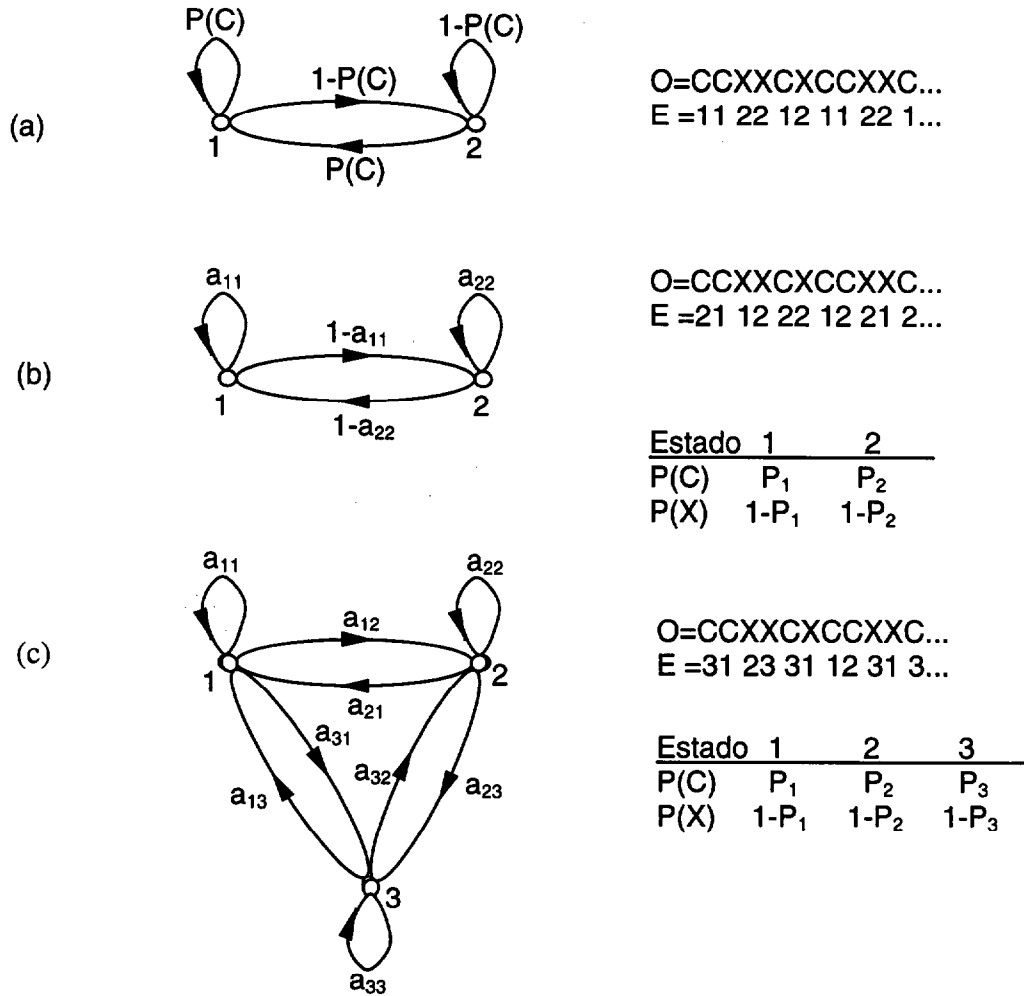


Figura 2.7: Diferentes modelos de la secuencia cara/cruz con HMM.

Se han planteado tres modelos (ver figura 2.7) para explicar el experimento, pero ¿cuál es el más adecuado para explicar la secuencia observada? El modelo de la figura 2.7a tiene un parámetro desconocido, el de la figura 2.7b tiene cuatro parámetros desconocidos, y el de la figura 2.7c tiene nueve. Es decir, según aumenta el grado de libertad del HMM, éste aumenta su tamaño y es capaz de modelar, además, los modelos de tamaño inferior. Aunque esta afirmación es teóricamente correcta, existen limitaciones prácticas que se contarán más adelante. Por otro lado, si la situación física real fuera la correspondiente a la figura 2.7a, se cometería un error si se emplea el modelo de la figura 2.7c.

Véase un ejemplo más complicado: el modelo de las urnas y las bolas (véase la figura 2.8). Considérense N urnas situadas en una habitación, dentro de cada urna hay 'infinitas' bolas de M colores diferentes. El experimento aleatorio consiste en elegir una urna (no se sabe cuál, puesto que están dentro de otra habitación y no se puede ver) y, en otro experimento aleatorio independiente, elegir una de las bolas de la urna. Se anota el color de la bola como resultado de la observación y se devuelve la bola a su urna. Este proceso se repite indefinidamente.

El modelo más simple para justificar este experimento es aquel que tiene tantos estados como urnas y en el que está definida la probabilidad de cada color dentro de cada estado. La elección de la urna viene determinada por la matriz de transición.

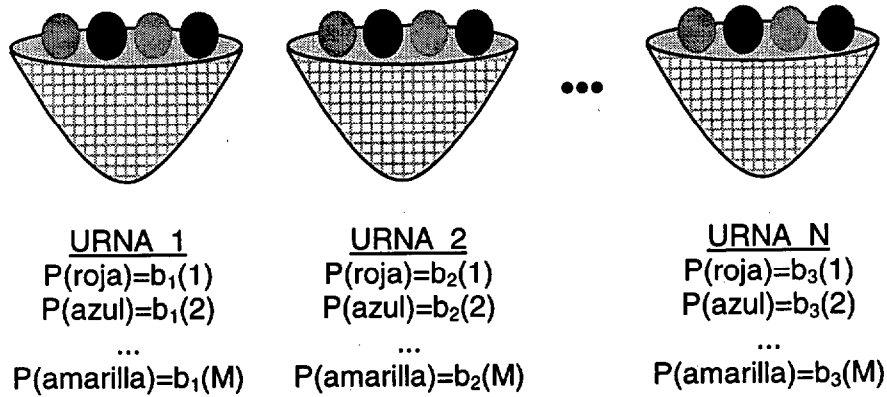


Figura 2.8: Ejemplo de proceso a modelar con un HMM.

2.3.3.6.3 Elementos de los Modelos Ocultos de Markov

Un modelo oculto de Markov queda caracterizado mediante los siguientes parámetros:

1. N , número de estados del modelo. A pesar de que los estados están ocultos, se les puede asociar un significado físico y, por lo tanto, determinar el número de estados. En el ejemplo del lanzamiento de monedas, cada estado corresponde a una moneda. En el ejemplo de las urnas y las bolas, los estados corresponden con las urnas. Se emplea la siguiente nomenclatura para designar los estados $S=\{S_1, S_2, \dots, S_N\}$ en el instante t : q_t .
2. M , número de símbolos distintos que se pueden observar en cada estado. Los símbolos observados se corresponden con la salida física del sistema que se desea modelar. En el ejemplo del lanzamiento de monedas los símbolos son cara, C, o cruz, X. En el ejemplo de las urnas y las bolas, los símbolos son los colores de las bolas. Para designar los símbolos se empleará la siguiente nomenclatura: $V=\{v_1, v_2, \dots, v_M\}$.
3. $A=\{a_{ij}\}$ es la función de distribución de probabilidad de transición entre estados, donde:

$$a_{ij} = P(q_{t+1} = S_j / q_t = S_i) \quad 1 \leq i, j \leq N \quad (2.17)$$

4. $B=\{b_j(k)\}$ es la función de distribución de probabilidad de observación de símbolos en el estado j , donde:

$$b_j(k) = P(v_k \text{ en el instante } t / q_t = S_j) \quad \begin{matrix} 1 \leq j \leq N \\ 1 \leq k \leq M \end{matrix} \quad (2.18)$$

5. $\pi=\{\pi_i\}$ es la distribución inicial de estados, donde:

$$\pi = P[q_1 = S_i] \quad 1 \leq i \leq N \quad (2.19)$$

Resumiendo, un modelo oculto de Markov queda determinado mediante los parámetros N , M , los símbolos de observación v_k , y las probabilidades A , B y π . Para hacer referencia al HMM se utilizará la notación compacta $\lambda=(A,B,\pi)$.

El modelo $\lambda=(A,B,\pi)$ se puede emplear para generar una secuencia de observación utilizando el siguiente procedimiento:

1. Elegir un estado inicial $q_1=S_i$ de acuerdo con la función de distribución inicial de estados π .
2. Asignar $t=1$.
3. Elegir $O_t=v_k$ de acuerdo con la función de distribución de probabilidad de observación de símbolos en el estado S_i , $b_i(k)$.
4. Ir al estado $q_{t+1}=S_j$ de acuerdo con la función de distribución de probabilidad de transición del estado S_i , a_{ij} .
5. Asignar $t=t+1$; si $t < T$ ir al paso III, si no Fin.

2.3.3.6.4 Los tres problemas fundamentales de un Modelo Oculto de Markov

A continuación se plantean tres problemas que deben de ser resueltos para que el modelo sea útil en aplicaciones reales [Rabiner89]:

- Problema 1: Dada la secuencia de observación $O=O_1O_2O_3...O_T$ y el modelo $\lambda=(A,B,\pi)$ ¿cómo se calcula la probabilidad de la secuencia observada dado el modelo, es decir, $P(O/\lambda)$?
- Problema 2: Dada la secuencia de observación $O=O_1O_2O_3...O_T$ y el modelo $\lambda=(A,B,\pi)$ ¿cuál es la secuencia de estados $Q=q_1q_2...q_T$ que mejor se adapta a la secuencia observada según un criterio de similitud?
- Problema 3: ¿Cómo ajustar los parámetros del modelo para maximizar $P(O/\lambda)$?

El problema 1 es conocido como el problema de evaluación. Al calcular $P(O/\lambda)$ se está determinando la posibilidad de que la secuencia O se haya producido por el modelo λ . En esta tesis se emplea la solución de este problema para, dada una secuencia de observación, correspondiente a un patrón manuscrito, y dados los modelos de dicho patrón, creados mediante el problema 3, determinar cuál de ellos pudo haber producido esa secuencia y, por lo tanto, identificarlo.

El problema 2 quiere calcular la secuencia de estados $Q=q_1q_2...q_T$ que mejor se adapta a la secuencia observada según un criterio de similitud, dada la secuencia de observación $O=O_1O_2O_3...O_T$ y el modelo $\lambda=(A,B,\pi)$. La solución no es única, ya que existen múltiples criterios a seguir para definir la secuencia de estados óptima, por ejemplo:

- ✓ Elegir aquellos estados individuales que son más probables.
- ✓ Elegir aquellas parejas, ternas ... de estados que son más probables.
- ✓ Elegir la secuencia de estados más probable.

El criterio más empleado es el de buscar la secuencia de estados más probable, para calcularla se emplea el algoritmo de Viterbi. Este algoritmo busca el camino que empieza con el primer estado más probable hasta el último, yendo de izquierda a derecha, e indicando finalmente, el camino más probable.

El problema 3 permite entrenar al modelo. Se emplean las secuencias de observación para ajustar los parámetros del modelo (entrenamiento). Por ejemplo, se utilizan muchas muestras de un mismo patrón, para que al entrenar el modelo, tenga

diferentes formas del dicho patrón, de forma que, al recibir una secuencia que no se utilizó para ajustar el modelo, ésta pueda ser identificada al resolver el problema 1.

2.3.3.6.5 Tipos de Modelos Ocultos de Markov

Hasta ahora se han presentado modelos ergódicos o completamente conectados (véase la figura 2.9a), llamados así porque desde cada estado se puede llegar a todos los demás. Es obvio que se cumple la relación en los elementos de la matriz de transiciones (A):

$$a_{ij} > 0 \quad 1 \leq i, j \leq N \quad (2.20)$$

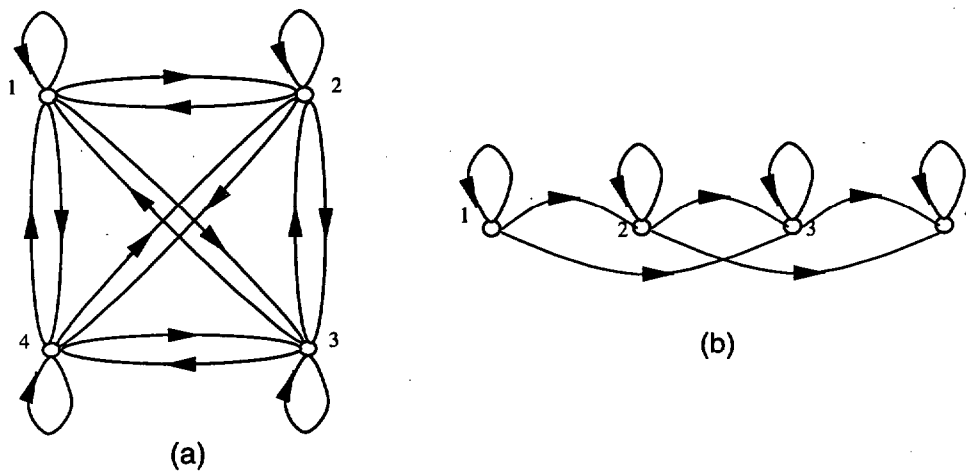


Figura 2.9: Ejemplo de modelo oculto de Markov ergódico y de izquierda a derecha (Bakis).

En la figura 2.9b se presenta un modelo izquierda-derecha o de Bakis. Este modelo se caracteriza porque, según aumenta el tiempo, aumenta el índice del estado, es decir, los estados se recorren de izquierda a derecha, sin poder volver hacia atrás. Presenta la propiedad de modelar fácilmente señales cuyas propiedades evolucionan en el tiempo.

En los modelos de Bakis la probabilidad de transición cumple que:

$$a_{ij} = 0 \quad \forall j < i \quad (2.21)$$

además, las probabilidades iniciales tienen la propiedad siguiente:

$$\pi_i = \begin{cases} 0 & i \neq 1 \\ 1 & i = 1 \end{cases} \quad (2.22)$$

Por último, cuando se llega al último estado, S_N , el sistema se queda en él, por lo que su coeficiente de transición es:

$$\begin{aligned} a_{NN} &= 1 \\ a_{Ni} &= 0 \quad \forall i < N \end{aligned} \quad (2.23)$$

Para garantizar que no ocurran transiciones entre estados muy separados se puede añadir la siguiente restricción en los modelos de Bakis:

$$a_{ij} = 0 \quad \forall j < i + \Delta \quad (2.24)$$

La variable Δ es un entero que indica a qué estados se puede saltar, estando en el estado S_i . En el ejemplo de la figura 2.9b, $\Delta=2$ en dicho modelo.

2.3.3.6 Multietiquetado en HMM discreto

En el planteamiento del HMM discreto, se aplica la técnica convencional de cuantificar el vector. Para cada vector de entrada, el cuantificador toma la decisión sobre cual es el valor que más conviene, a partir de la información de como es el vector de entrada. Para evitar tomar una decisión *software*, se realiza una decisión fija sobre el valor cuantificado [Hernado97]. Para ampliar los posibles valores que vaya a tomar el cuantificador, se usa el multietiquetado, de forma que variando este parámetro se controle los posibles valores cuantificados. El número de etiquetas en el DHMM está relacionado con los valores que se puedan tomar en el número de símbolos por estado.

Los algoritmos del DHMM deben ser generalizados para acomodarse al multietiquetado de salida ($\{v_k\}_{k=1,\dots,C}$), para generar el vector de salida $\{w(x_t, v_k)\}_{k=1,\dots,C}$. Por tanto, para un estado dado j del HMM, la probabilidad que un vector x_t sea observado en el instante t , puede ser escrita como;

$$b_j(x_t) = \sum_{k=1}^C w(x_t, v_k) b_j(k) \quad (2.25)$$

donde $b_j(k)$ es la probabilidad de salida discreta asociada con el valor v_k y el estado j ; siendo C el tamaño de la librería de valores del vector.

2.3.3.6.7 Multiparametrización

Una de las ventajas que puede ofrecer los modelos ocultos de Markov es el planteamiento de ver, la secuencia de observación como una matriz. Para el caso particular de la voz, dicha secuencia es un vector definido por unos determinados parámetros, pero para otro tipo de aplicaciones se le pueden asignar varias secuencias en paralelo, formando la matriz comentada con anterioridad.

Para el caso de imágenes, permitiría seguir la secuencia en coordenadas cartesianas (componentes "X" e "Y") o en polares (módulo y argumento), mediante una matriz de dos columnas, que describiera la secuencia de una cierta región o contorno.

Por tanto, el vector puede tomar la forma;

$$x_t = [x_t^1 \quad x_t^2 \quad \dots \quad x_t^{N1}] \quad (2.26)$$

y la probabilidad de salida para este vector x_t , vendría dada por:

$$b_j(x_t) = \prod_{i=1}^N b_j(x_t^i) \quad (2.27)$$

Capítulo 3

Elaboración de la base de datos para manuscritos simples: dígitos y letras

3.1 Introducción

Debido a los objetivos planteados en esta tesis, este capítulo desarrolla el diseño y la creación de dos bases de datos para imágenes manuscritas simples, en particular de dígitos del sistema decimal y las letras del alfabeto castellano. Este proceso de captación de muestras usado para estas aplicaciones, será común para el resto de base de datos que se definan en esta tesis, como es la base de datos de palabras manuscritas que se estudiará en el capítulo quinto; y la base de datos de firmas detallada en el capítulo sexto.

Se va a desarrollar cuales son las contingencias que debe de tener una base de datos para su correcta conformación y elaboración. Posteriormente, se detalla cual ha sido el preprocesado aplicado a las muestras mediante el uso de los algoritmos realizados para un sistema *off-line*. Finalmente, se preprocesarán los datos para conformar las imágenes de los patrones.



3.2 Base de datos de dígitos y firmas manuscritas

3.2.1 Parámetros de la base de datos

La creación de una base de datos requiere de unos requisitos a superar que son los que le dan validez a dicha base. En este caso, el principal propósito radica en su tamaño, ya que debe ser lo suficientemente amplia para que se pueda considerar un patrón de esta base de datos independiente de otro patrón, y al mismo tiempo recoja sus particularidades. Este propósito lleva implícito que sea representativa de la población, en cuanto al sexo de la persona, la mano de escritura y la edad de la persona.

Por tanto, la función de la base de datos debe ser la de proporcionar la cantidad de muestras suficientes, para crear el clasificador y comprobar su comportamiento. Visto de otro modo, la base de datos servirá a los programas de clasificación con el fin de obtener un aprendizaje de los modelos mediante los parámetros adquiridos, para alcanzar un reconocimiento de los patrones, fueran cual fueran estos parámetros de entrada al clasificador para entrenar.

No se pretende tener una base de datos con muestras perfectas estéticamente, sino que éstas sean cotidianas, pero que transmitan la suficiente información como para diferenciar entre distintos patrones.

Para conseguir una base de datos como la descrita en los párrafos anteriores se fijó una serie de pautas:

1. Conseguir muestras manuscritas
2. Definir el universo de patrones
3. Conseguir diferentes tipos de muestras de un mismo patrón
4. Representar estadísticamente la población real
5. Lograr una cantidad suficiente de muestras
6. Conservar la naturalidad de la escritura

El primer punto implica la captación de patrones que sean del tipo manuscrito, ya que es la aplicación en la que se ha basado la defensa de esta tesis para mostrar el uso de HMM en la clasificación de este tipo de imágenes.

En el segundo punto se va a definir un grupo de aplicaciones para poder llevar a cabo la tarea. Este grupo de bases de datos tiene la misión de probar la teoría en varias aplicaciones diferentes y con características diferentes para demostrar el éxito del planteamiento. El grupo de patrones que se han escogido para implementar estas bases de datos ha sido:

1. Dígitos manuscritos
2. Letras manuscritas, tanto mayúsculas como minúsculas

En el siguiente punto, un patrón puede ser escrito por diferentes escritores, de tal forma que se puedan captar todas las posibles variaciones que pueda registrar dicho patrón, y generalizar en mayor medida su modelado.

El cuarto punto implica obtener una base de datos representativa de la población y por tanto, generalizable. Por eso se va a etiquetar a los escritores de la base de datos según su sexo, su edad y la mano de escritura, para que represente a cualquier persona de la población estadísticamente.

El quinto punto permite obtener muestras de diferentes escritores, y además, varias muestras del mismo escritor, de forma que éste escribe varias repeticiones de un mismo patrón y también, escribe patrones diferentes; y así con todos los escritores. La finalidad es obtener una base de datos lo suficientemente grande para que las secuencias de datos de entrenamiento sean independientes de las secuencias de test, y poder realizar experimentos con este carácter de independencia del escritor. De esta forma los clasificadores pueden aprender de forma incorrelada.

El último y sexto punto, conserva la naturalidad de la escritura, pretende que las muestras no estén condicionadas por el hecho de que van a ser utilizadas en una tesis científica. Por eso se pensó en que los voluntarios escribiesen las muestras directamente con bolígrafo sobre el papel, en su manera habitual y natural de realizarlo, para su posterior captación *off-line*. Para ello, se fijó un número reducido de repeticiones sobre el conjunto de patrones que se pretende recabar, y se fija el tamaño de los datos que se quieren obtener de cada escritor, ya que si se les piden muchas muestras a éstos, acaban siendo mal escritas.

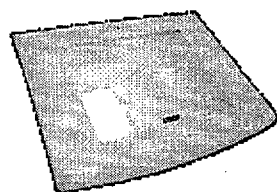
Una vez que las pautas sobre las bases de datos han quedado claras sólo queda comentar en detalle el sistema de captación *off-line*, antes de explicar cada conjunto de patrones escogidos o base de datos en su caso particular.

3.2.2 Elección del sistema de captura de datos

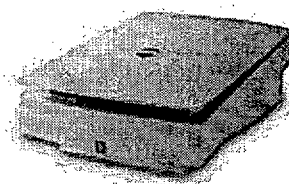
Los sistemas existentes para la captura de información escrita son dos: el sistema *on-line* y el sistema *off-line*. En apartados anteriores ya se ha mencionado, que las muestras se van a recoger mediante el sistema de captación *off-line* [O’Gorman95][Seiler96]. En este apartado se van a describir las principales características de cada uno de los sistemas y se justifica el uso del sistema de captación *off-line*.

La diferencia entre ellos se encuentra en su principio de funcionamiento, mientras en el sistema *on-line* la captura de los datos se realiza al mismo tiempo que se introducen, en el sistema *off-line* se hace después de que hayan sido introducidos. Esta misma diferencia marca las ventajas y los inconvenientes de ambos métodos. La ventaja que ofrece el sistema *on-line*, es la obtención del vector de parámetros de la información de entrada a los clasificadores en tiempo real, evitando así el procesado de la señal, y pudiéndose plantear de esta forma, que los vectores ataquen directamente a un clasificador.

Durante el desarrollo de esta tesis se han utilizado ambos métodos, manejando dispositivos de entrada a un ordenador diferente para cada método. En el sistema *on-line* se utilizó una tarjeta digitalizadora con un lápiz óptico, y para el sistema *off-line* un escáner óptico, mostrándose dichos dispositivos en la siguiente figura.



(a) Tableta digitalizadora



(b) Escáner óptico

Figura 3.1: Dispositivos de entrada para los sistemas *on-line* y *off-line*.

El manejo de la tarjeta digitalizadora presentó toda una serie de dificultades en la captación de los datos enumeradas a continuación:

1. La sensibilidad del instrumento no se adecuó a la escritura natural, aproximando las curvas a la unión de pequeñas rectas. Para evitar dicho efecto se debía de escribir muy lentamente para recoger dichos cambios, por lo que no se establecía un mecanismo normal para representar la escritura.
2. El trazo de escritura que se realizaba sobre la tableta digitalizadora no se visualizaba en la misma, y por tanto se incurría en los errores de escritura por inclinación, y de no saber dónde situar un trazo discontinuo durante el proceso de escribir.
3. No se disponía de ningún software específico que hiciera de interfaz de la tableta, y devolviera la información del trazo de la escritura.

El sistema *off-line* devuelve la información mediante el uso de un escáner. Mediante un procesado de los datos, se obtienen los vectores de los parámetros de los patrones de entrada. El inconveniente es la imposibilidad de realizar aplicaciones a tiempo real, con respecto a la captación de las muestras debido al escaneado y al retardo producido por los algoritmos de procesado.

La decisión de escoger este sistema *off-line* entre ambos, viene dada por dos cuestiones. La primera, aparte del menor coste económico, es la facilidad en su manejo y en la captación de datos. Y la segunda es que para las muestras de la base de datos, éstas ya estaban escritas, por lo que no se podría usar una tableta, pero sí un escáner.

Otro motivo para elegir este sistema, es la movilidad que permite la recogida de las muestras. Se pueden repartir plantillas entre los escritores y ser rellenas con independencia del tiempo y del lugar. El método *on-line* requiere que el escritor se desplace al laboratorio donde esté instalada la tarjeta digitalizadora, lo cual dificulta la participación en la recopilación de muestras.

Por tanto, comparando ambos sistemas, el sistema *on-line* presenta más inconvenientes que hacen muy dificultosa su implementación con los medios disponibles, por lo que se ha escogido el sistema *off-line*.

3.2.3 Diseño y construcción de las bases de datos

En este apartado se va a detallar el proceso y los parámetros de construcción de las dos bases de datos, tanto de dígitos como de letras. Además, se incluyen las instrucciones para completar cada plantilla y el formato que se establece para cada base de datos.

Un punto importante en toda la recogida de información de todas las bases de datos, es la conformación del aspecto de las plantillas y de la redacción de las instrucciones de relleno para el sistema *off-line*. El tamaño de papel para el formato de la plantilla viene dado por el escáner óptico de mesa, con un tamaño de DIN-A4.

El formato para cada plantilla cumple dos objetivos, el primero es guiar a los escritores indicándoles donde deben escribir la muestra correspondiente; y el segundo es para facilitar el preprocesado de imagen, situando las muestras de forma simétrica para reducir el proceso de separación y gestión de datos. De las muestras recogidas se analizará la estadística de los escritores participantes, así como el planteamiento de las dificultades que se han presentado en la recopilación de las mismas.

3.2.4 Instrucciones para rellenar las plantillas

Hay que tener presente que se van a plantear dos bases de datos diferentes, con un conjunto de patrones diferentes, por lo que las exigencias geométricas de las plantillas en cada caso serán diferentes. El tamaño de la cuadrícula se eligió de forma que pueda contener la muestra manuscrita, persiguiendo ante todo, que el escritor se sintiese cómodo con el espacio destinado a escribir.

Las instrucciones que se deben de tener en cuenta al cumplimentar dichas plantillas, para la posterior manipulación de las muestras adquiridas, son comunes a ambas bases de datos.

Estas instrucciones para rellenar de manera correcta las plantillas, se establecieron con dos fines principalmente. Por un lado, facilitar los subsiguientes pasos del preprocesado digital de imagen en la ordenación y manipulación de las muestras. Y por otro lado, conseguir un efecto de naturalidad en el trazado de la escritura por parte del escritor, para no fatigar, ni agilizar el trazado de las muestras manuscritas, de manera que no se limitase la forma de escribir de cada uno de los escritores.

Las instrucciones establecidas son las siguientes:

- 1ª Instrucción: Rellenar la plantilla con rotulador azul o negro, de punta normal, para poder evitar los trazos discontinuos procedentes de los bolígrafos comunes, ya que producen problemas de continuidad en el proceso de digitalización.
- 2ª Instrucción: Realizar una escritura natural, ya que en pruebas realizadas se observó que algunos escritores tendían a reproducir los ejemplos de muestra en lugar de escribir con su letra personal.

- 3ª Instrucción: Escribir cada letra dentro de su recuadro correspondiente, evitando invadir con los trazos de escritura el recinto limitado para cada muestra.
- 4ª Instrucción: Centrar la muestra dentro del recuadro, con la finalidad de no crear ruidos o interferencias inadecuadas en las muestras adyacentes.
- 5ª Instrucción: Escribir las muestras en sentido ordenado y según corresponda para evitar realizar las repeticiones de cada muestra de forma seguida, ya que este efecto provoca la pérdida de la naturalidad de la escritura, al intentar imitar la muestra anterior repetida escribiendo a mayor velocidad, con la correspondiente pérdida de los rasgos característicos de los trazos de cada escritor.
- 6ª Instrucción: No retocar las letras escritas o lo que es lo mismo, no sobrescribir o poner tachones, para evitar los trazos añadidos, que no se encuentran en la escritura natural.
- 7ª Instrucción: Evitar doblar la plantilla, para evitar posibles fuentes de ruido en la digitalización de la misma.

3.2.4.1 Base de datos de dígitos manuscritos

Esta base de datos consta de un universo de elementos limitado a diez, que son cada uno de los miembros del sistema numérico decimal, (0,1,2,3,4,5,6,7,8,9). Por tanto, ya se tiene seleccionado a priori, cuales son las posibles respuestas que debería de estimar el clasificador.

En particular, para esta base de datos se han utilizado 92 escritores, que han realizado 12 repeticiones por cada uno de los 10 dígitos, obteniendo un total de 11040 muestras ($92 \times 10 \times 12$) [Morales99-pfc].

3.2.4.1.1 Perfil de los escritores

Al realizar esta base de datos se pretendió que se recogieran la mayor variedad posible de formas de cada dígito, para de esta manera, simular una situación más real y evitar tasas de reconocimiento falseadas por la propia base de datos. Por eso, se recurrió a buscar escritores de diferentes edades, sexo y mano de escritura, para que fuera representativo de la población.

En la siguiente tabla se muestran los escritores que han compuesto esta base de datos, según los parámetros anteriormente citados.

ESCRITOR	SEXO	MANO ESCRITURA	EDAD	ESCRITOR	SEXO	MANO ESCRITURA	EDAD
1	Hombre	Diestro	26	47	Hombre	Diestro	22
2	Mujer	Diestra	24	48	Mujer	Diestra	20
3	Hombre	Diestro	28	49	Hombre	Diestro	24
4	Mujer	Diestra	49	50	Mujer	Diestra	21
5	Mujer	Diestra	37	51	Hombre	Diestro	27
6	Hombre	Diestro	30	52	Hombre	Diestro	24
7	Mujer	Diestra	26	53	Mujer	Diestra	22
8	Hombre	Diestro	40	54	Hombre	Diestro	22
9	Hombre	Diestro	25	55	Hombre	Diestro	28
10	Hombre	Diestro	25	56	Hombre	Diestro	23
11	Hombre	Diestro	27	57	Hombre	Diestro	25
12	Hombre	Diestro	28	58	Hombre	Diestro	25
13	Mujer	Diestra	32	59	Hombre	Diestro	28
14	Hombre	Diestro	24	60	Hombre	Diestro	22
15	Mujer	Diestra	29	61	Mujer	Diestra	22
16	Hombre	Diestro	59	62	Mujer	Diestra	25
17	Hombre	Diestro	29	63	Mujer	Diestra	24
18	Hombre	Diestro	32	64	Hombre	Diestro	23
19	Hombre	Zurdo	26	65	Hombre	Diestro	24
20	Hombre	Diestro	28	66	Hombre	Diestro	22
21	Hombre	Diestro	40	67	Hombre	Diestro	23
22	Hombre	Zurdo	27	68	Hombre	Diestro	24
23	Mujer	Diestra	23	69	Hombre	Diestro	26
24	Hombre	Diestro	26	70	Mujer	Diestra	18
25	Mujer	Diestro	53	71	Hombre	Diestro	18
26	Mujer	Diestra	48	72	Hombre	Diestro	18
27	Hombre	Diestro	26	73	Hombre	Diestro	18
28	Hombre	Diestro	58	74	Hombre	Diestro	20
29	Mujer	Zurda	23	75	Hombre	Diestro	22
30	Mujer	Diestra	47	76	Hombre	Diestro	21
31	Hombre	Diestro	53	77	Mujer	Diestra	18
32	Hombre	Zurdo	25	78	Mujer	Diestra	21
33	Hombre	Diestro	27	79	Hombre	Diestro	20
34	Hombre	Zurdo	15	80	Hombre	Diestro	26
35	Hombre	Diestro	19	81	Mujer	Diestra	22
36	Hombre	Diestro	31	82	Hombre	Diestro	22
37	Mujer	Diestro	31	83	Hombre	Diestro	19
38	Hombre	Diestro	26	84	Mujer	Diestra	18
39	Mujer	Zurda	14	85	Mujer	Diestra	18
40	Mujer	Diestra	28	86	Hombre	Diestro	26
41	Mujer	Zurda	21	87	Hombre	Diestro	24
42	Mujer	Diestra	21	88	Hombre	Diestro	16
43	Hombre	Diestro	25	89	Mujer	Diestra	15
44	Hombre	Diestro	25	90	Hombre	Diestro	21
45	Mujer	Diestra	24	91	Hombre	Diestro	25
46	Mujer	Diestra	27	92	Hombre	Diestro	26

Tabla 3.1: Muestrario de escritores de la base de datos de dígitos manuscritos.

De los datos expuestos en la tabla 3.1, se puede observar que los escritores van en el intervalo de 14 a 59 años de edad. Además la población de sexo masculino es algo superior al femenino, y que la inmensa mayoría de los escritores son diestros. Si extrapolamos estos datos a una gráfica para su mejor interpretación, se obtiene a modo de resumen;

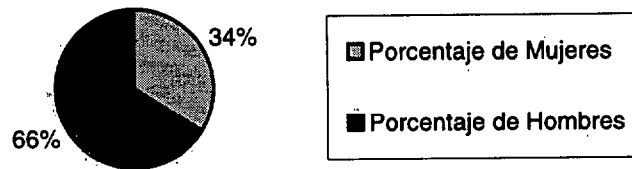


Figura 3.2: Porcentaje de Hombres vs. Mujeres de la base de datos de dígitos manuscritos.

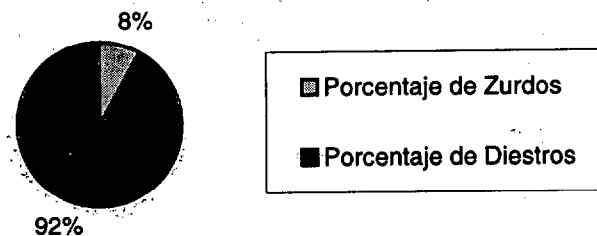


Figura 3.3: Porcentajes de Diestros vs. Zurdos de la base de datos de dígitos manuscritos.

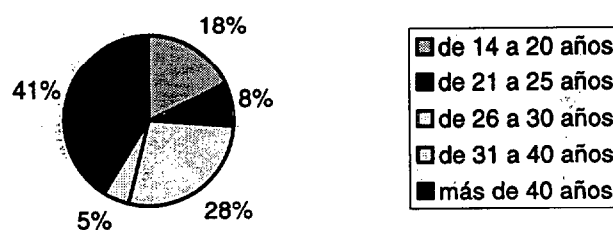


Figura 3.4: Distribución por edades en la base de datos de dígitos manuscritos.

3.2.4.1.2 Formato de la plantilla

La plantilla utilizada para la realización de esta base de datos se representa en la figura 3.5, y en la figura 3.6 una plantilla totalmente rellena. Se puede notar como dicha plantilla se divide claramente por tres grandes rasgos geométricos. Por un lado las líneas verticales y horizontales que acompañan a la plantilla por el lado superior y por el lado izquierdo. Su utilidad radica en uno de los pasos del preprocesado digital, aquel en que calcula su inclinación. Otra región notoria, es el cajetín situado debajo de las líneas horizontales, cuya misión es la de recoger la estadística del escritor y alguna observación de la correspondiente plantilla. Y la última región significativa, es la cuadrícula que casi abarca toda la plantilla, donde se recogerán las muestras. Esta

N. de escritor: Datos del escritor: Derecho (53) Tipo de caracteres: NOTAS: PROFESOR										60
1	2	3	4	5	6	7	8	9	0	
1	2	3	4	5	6	7	8	9	0	
1	2	3	4	5	6	7	8	9	0	
1	2	3	4	5	6	7	8	9	0	
1	2	3	4	5	6	7	8	9	0	
1	2	3	4	5	6	7	8	9	0	
1	2	3	4	5	6	7	8	9	0	
1	2	3	4	5	6	7	8	9	0	
1	2	3	4	5	6	7	8	9	0	
1	2	3	4	5	6	7	8	9	0	
1	2	3	4	5	6	7	8	9	0	
1	2	3	4	5	6	7	8	9	0	
1	2	3	4	5	6	7	8	9	0	

© Universidad de Las Palmas de Gran Canaria. Biblioteca Digital. 2004

Figura 3.6: Plantilla de la base de datos de los dígitos manuscritos rellena.

3.2.4.2 Base de datos de letras manuscritas

La base de datos de letras manuscrita consta de un universo de elementos limitado, que son las letras de nuestro alfabeto castellano, desde la “a” hasta la “z”, y pasando por la “ñ”, como caso particular para nuestra lengua castellana. En total suman veintisiete letras, pero con la particularidad, de que se ha especificado tanto letras mayúsculas, como letras minúsculas. Por lo que el número total de patrones a diferenciar en esta base de datos es de 54 patrones diferentes. Por tanto, el clasificador a priori, conoce cual es el total de las supuestas salidas, ya que este conjunto de patrones está totalmente cerrado.

Para la construcción de esta base de datos se han utilizado 100 escritores, que han realizado 12 repeticiones por cada una de las diferentes letras, tanto de mayúsculas (27 patrones) como de minúsculas (27 patrones), obteniendo un total de 64800 muestras (100×27×2×12) [Peinado00-pfc].

3.2.4.2.1 Perfil de los escritores

La representatividad que se ha querido lograr en esta base de datos, se ha conseguido en gran medida, ya que la estadística sobre los escritores ha sido realizada antes de obtener las muestras. Los datos de los escritores se pueden encontrar en la siguiente tabla;

ESCRITOR	SEXO	MANOESCRITURA	EDAD
0	Hombre	Diestro	26
1	Hombre	Diestro	21
2	Hombre	Diestro	23
3	Hombre	Zurdo	25
4	Hombre	Diestro	21
5	Hombre	Diestro	20
6	Hombre	Diestro	21
7	Hombre	Diestro	21
8	Mujer	Diestra	21
9	Hombre	Diestro	21
10	Hombre	Diestro	22
11	Hombre	Zurdo	25
12	Hombre	Diestro	30
13	Hombre	Diestro	24
14	Mujer	Diestra	20
15	Hombre	Diestro	21
16	Hombre	Diestro	21
17	Hombre	Diestro	23
18	Hombre	Zurdo	21
19	Mujer	Diestra	24
20	Hombre	Diestro	21
21	Hombre	Diestro	20
22	Mujer	Diestra	26
23	Hombre	Diestro	28

ESCRITOR	SEXO	MANOESCRITURA	EDAD
24	Hombre	Diestro	19
25	Mujer	Diestra	33
26	Mujer	Zurda	20
27	Mujer	Diestra	20
28	Mujer	Diestra	21
29	Hombre	Diestro	19
30	Hombre	Diestro	21
31	Mujer	Diestra	13
32	Hombre	Diestro	18
33	Mujer	Diestra	22
34	Hombre	Diestro	19
35	Mujer	Diestra	21
36	Mujer	Diestra	28
37	Hombre	Diestro	19
38	Hombre	Zurdo	19
39	Mujer	Diestra	14
40	Hombre	Diestro	19
41	Mujer	Diestra	19
42	Mujer	Diestra	30
43	Hombre	Diestro	27
44	Mujer	Diestra	21
45	Mujer	Diestra	29
46	Mujer	Diestra	19
47	Hombre	Diestro	19

48	Mujer	Diestra	19
49	Hombre	Diestro	26
50	Mujer	Diestra	46
51	Hombre	Diestro	35
52	Hombre	Diestro	32
53	Mujer	Diestra	21
54	Mujer	Zurda	31
55	Mujer	Diestra	21
56	Mujer	Diestra	39
57	Mujer	Diestra	40
58	Mujer	Diestra	27
59	Hombre	Diestro	41
60	Hombre	Diestro	25
61	Mujer	Diestra	22
62	Mujer	Diestra	22
63	Hombre	Diestro	85
64	Mujer	Diestra	18
65	Mujer	Diestra	21
66	Mujer	Diestra	42
67	Mujer	Zurda	21
68	Mujer	Diestra	36
69	Mujer	Diestra	27
70	Mujer	Diestra	17
71	Hombre	Diestro	42
72	Hombre	Diestro	42
73	Mujer	Zurda	23
74	Mujer	Diestra	27
75	Mujer	Diestra	29
76	Mujer	Diestra	36
77	Hombre	Diestro	49
78	Mujer	Diestra	43
79	Mujer	Diestra	43
80	Mujer	Diestra	38
81	Hombre	Diestro	29
82	Mujer	Diestra	37
83	Hombre	Diestro	55
84	Mujer	Diestra	48
85	Hombre	Diestro	54
86	Hombre	Diestro	73
87	Hombre	Diestro	80
88	Mujer	Diestra	53
89	Mujer	Diestra	61
90	Mujer	Diestra	75
91	Mujer	Diestra	59
92	Mujer	Diestra	59
93	Hombre	Diestro	66
94	Hombre	Diestro	64
95	Hombre	Diestro	55
96	Hombre	Diestro	60
97	Hombre	Diestro	59
98	Mujer	Diestra	55
99	Hombre	Diestro	41

Tabla 3.2: Muestrario de los escritores de la base de datos de letras.

De los datos de los escritores se puede extraer, que el intervalo de edades oscila entre los 13 y los 85 años (véase la figura 3.7), hay un mismo porcentaje de hombres y de mujeres del 50% (véase la figura 3.8), y lógicamente hay una mayor cantidad de diestros frente a los zurdos (véase la figura 3.9). A modo de resumen, se pueden observar las siguientes gráficas ya indicas;

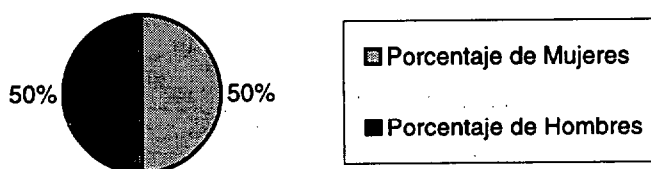


Figura 3.7: Porcentaje de Hombres vs. Mujeres de la base de datos de letras manuscritas.

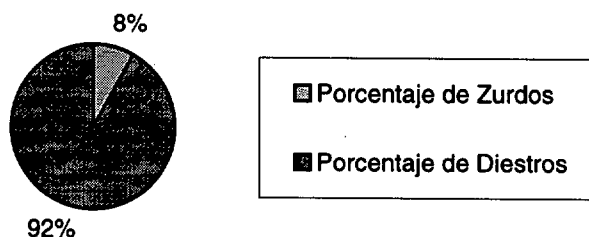


Figura 3.8: Porcentajes de Diestros vs. Zurdos de la base de datos de letras manuscritas.

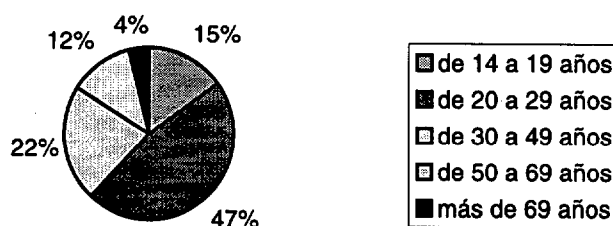


Figura 3.9: Distribución por edades en la base de datos de letras manuscritas.

3.2.4.2.2 Formato de la plantilla

El formato utilizado para esta base de datos tiene la particularidad de que consta de dos plantillas, una para letras mayúsculas y otra para letras minúsculas; ya que el número de muestras era excesivo para incorporarlo en una sola. En la figuras números 3.10 y 3.11 se representan las plantillas utilizadas, y en las figuras 3.12 y 3.13, se muestran dichas plantillas con la incorporación de las muestras.

De este formato de plantilla, el primer detalle que se advierte es la posición apaisada de dicha plantilla (véase figuras 3.10 y 3.11). De la parte superior a la inferior se puede localizar tres bloques fundamentales en la que se divide. En esa parte superior se distinguen dos líneas horizontales, cuya misión ya se había comentado con anterioridad, y es la consideración de la validez de dicha plantilla teniendo en cuenta el parámetro de inclinación, $\pm 0.5^\circ$ como máximo [O'Gorman95].

En la parte central se localiza el cajetín dividido en dos partes, una donde se sitúan las instrucciones citadas en el apartado 3.2.4 de este mismo capítulo, y otra parte destinada a la captación de la información de los escritores, para complimentar la estadística de su perfil. Y finalmente, el último bloque es la cuadrícula donde se depositarán los muestras de cada una de las letras. En esta cuadrícula se incluye a modo de ejemplo, para mayor claridad del escritor, cada una de las muestras con la que se debe de rellenar. El enrejado de la cuadrícula consta de veintisiete cuadrados horizontales, uno por cada una de las letras que determinarán esta base de datos; y por 13 cuadrados verticales, el primero donde se sitúa a modo de ejemplo cada una de las letras, y los doce restantes para incluir las muestras que se quieren captar. Lógicamente como son letras mayúsculas y minúsculas las que se quieren captar, se han realizado dos plantillas, como se ha comentado con anterioridad.

Figura 3.12: Plantilla de la base de datos de las letras minúsculas manuscritas rellena.

Nº DE MUESTRA: 66 MANO: Derecha SEXO: Mujer EDAD: 42.		INSTRUCCIONES PARA RELLENAR LA PLANTILLA: 1.- Escriba con rotulador negro o azul. 2.- Escriba naturalmente. 3.- Escriba la letra dentro de su cuadro; procure no cortar el recuadro con los trazos de la letra. 4.- Intente centrar la letra dentro del recuadro. 5.- Escriba en sentido horizontal: a → b → c → d → ... → y → z 6.- No retoque las letras escritas. 7.- No doble la plantilla ni realice borrões.																									
a	b	c	d	e	f	g	h	i	j	k	l	m	n	ñ	o	p	q	r	s	t	u	v	w	x	y	z	
a	b	c	d	e	f	g	h	i	j	k	l	m	n	ñ	o	p	q	r	s	t	u	v	w	x	y	z	
a	b	c	d	e	f	g	h	i	j	k	l	m	n	ñ	o	p	q	r	s	t	u	v	w	x	y	z	
a	b	c	d	e	f	g	h	i	j	k	l	m	n	ñ	o	p	q	r	s	t	u	v	w	x	y	z	
a	b	c	d	e	f	g	h	i	j	k	l	m	n	ñ	o	p	q	r	s	t	u	v	w	x	y	z	
a	b	c	d	e	f	g	h	i	j	k	l	m	n	ñ	o	p	q	r	s	t	u	v	w	x	y	z	
a	b	c	d	e	f	g	h	i	j	k	l	m	n	ñ	o	p	q	r	s	t	u	v	w	x	y	z	
a	b	c	d	e	f	g	h	i	j	k	l	m	n	ñ	o	p	q	r	s	t	u	v	w	x	y	z	
a	b	c	d	e	f	g	h	i	j	k	l	m	n	ñ	o	p	q	r	s	t	u	v	w	x	y	z	
a	b	c	d	e	f	g	h	i	j	k	l	m	n	ñ	o	p	q	r	s	t	u	v	w	x	y	z	
a	b	c	d	e	f	g	h	i	j	k	l	m	n	ñ	o	p	q	r	s	t	u	v	w	x	y	z	
a	b	c	d	e	f	g	h	i	j	k	l	m	n	ñ	o	p	q	r	s	t	u	v	w	x	y	z	
a	b	c	d	e	f	g	h	i	j	k	l	m	n	ñ	o	p	q	r	s	t	u	v	w	x	y	z	

Figura 3.13: Plantilla de la base de datos de las letras mayúsculas manuscritas rellena.

N° DE MUESTRA: 66 MANO: Derecha SEXO: Mujer EDAD: 42	INSTRUCCIONES PARA RELLENAR LA PLANTILLA: 1.- Escriba con rotulador negro o azul. 2.- Escriba naturalmente. 3.- Escriba la letra dentro de su cuadro; procure no cortar el recuadro con los trazos de la letra. 4.- Intente centrar la letra dentro del recuadro. 5.- Escriba en sentido horizontal: a → b → c → d → ... → y → z 6.- No retoque las letras escritas. 7.- No doble la plantilla ni realice borroneos.
---	--

A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z

3.2.5 Dificultades en la creación de las bases de datos

Las dificultades que se han presentado en la elaboración de esta base de datos han tenido que ver con el relleno de las plantillas por parte de los escritores. A continuación, se van a detallar estos puntos;

- Debido al excesivo tiempo y el proceso mecánico de poner muestras en todas las bases de datos, se procedió a realizar la recogida de las muestras en dos sesiones, para evitar el cansancio y la rapidez en la escritura. Algunos escritores que rellenaron las plantillas en una sesión, admitieron al final del mismo un efecto de cansancio de la mano.
- En muchas ocasiones los escritores se han salido del espacio destinado en las cuadrículas para cada muestra, lo que ha generado ruido, al aparecer otro trazo en la cuadrícula adyacente; e interferencia, por solapar el trazo de una muestra con el trazo de la adyacente.

Quizás el hecho de no tener ninguna compensación económica ha acarreado el fruto de estos dos puntos, pero la buena voluntad de todos, finalmente ha tomado éxito en la construcción de estas bases de datos.

3.3 Preprocesado de imagen de las bases de datos

En este apartado se detallarán los pasos a seguir para procesar la información. Todo este proceso conlleva unas tareas de preprocesado de imagen aplicado a las muestras captadas, tal y como se detalla en la figura 3.14.

3.3.1 Digitalización de las bases de datos

El proceso que se seguirá va a transformar las muestras recogidas en un papel, para almacenarlas digitalmente en un fichero gráfico. Se parte de un sistema *off-line* para la captura de los datos, es decir, que la información se debe obtener a posteriori de su escritura. Para dejar estos datos digitalizados en el ordenador se ha utilizado un escáner como el que se mostró en la figura 3.1(b). Éste resulta un medio económico y al alcance de cualquier persona, y permite recoger la información depositada en una hoja de papel. El escáner utilizado fue el *ScanJet 5p* de *Hewlett Packard*, tomando una resolución de 300 dpi. Esta cantidad se consideró la idónea, para captar las características de los datos de forma aceptable, ya que esta resolución ha sido usada también en otros trabajos [O'Gorman95][Cho97].

Se ha ajustado el software del escáner para recoger una imagen en escala de grises de 8 bits, cuantificando por tanto, con 256 niveles entre el blanco y el negro. Puede resultar de mayor ventaja realizar el proceso en monocromo, en blanco y negro, ya que ocupan menor cantidad de memoria, pero se optó por los valores establecidos en primer lugar, por necesidades software en la implementación de los algoritmos de procesado. Por ejemplo, es mejor establecer el umbral de binarización controladamente, que mediante el propio software de manipulación del escáner. Posteriormente, esta

información se almacenará en un fichero gráfico de mapas de bits, que comúnmente tienen la extensión "bmp".

En este proceso, hay parámetros que no son nada relevantes, como el tiempo de exposición del escáner, el brillo, el contraste, o la luz, ya que no se requiere una gran nitidez de la imagen. Debido a que se quiere trabajar con valores de contrastes grandes, del blanco al negro, estos parámetros pueden tomar valores típicos o una cierta variación, ya que no van a ser significativos en el resultado final de la imagen. Incluso el software utilizado para el escáner, y el modelo particular del propio escáner, no son nada significativo en estas imágenes.

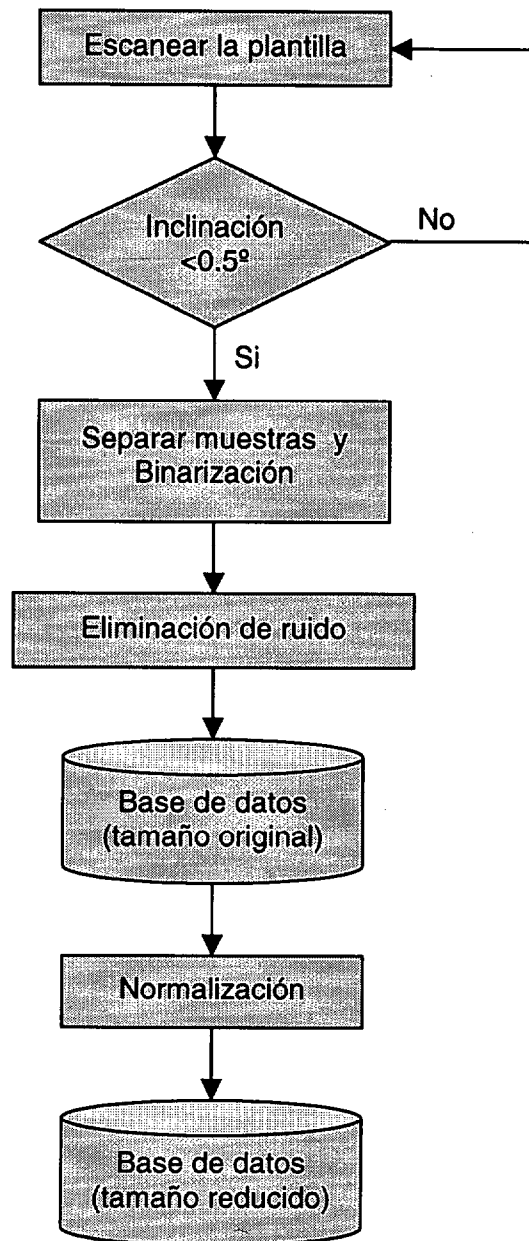


Figura 3.14: Procesado de las plantillas.

3.3.2 Corrección de la inclinación

Una vez que los datos han sido pasados del papel a un mapa de bits, hay que tener presente, que para el ordenador, tan solo es una matriz de datos numéricos. El ordenador no tiene la percepción humana para distinguir si los datos están inclinados o no, y por eso, hay que prever esta variable.

La causa de una imagen inclinada puede ser producida por dos motivos:

1. que sea una característica intrínseca de la propia imagen de su origen
2. que sea debida a un error en la adquisición de los datos mediante el escáner

En este apartado se pretende dar solución al segundo de los casos, ya que en la literatura se describe que el error máximo debido al proceso de escanear, debe de ser de $\pm 0.5^\circ$, para considerar los datos adquiridos como satisfactorios [O’Gorman95].

Para detectar este tipo de errores humanos usando el escáner, se ha implementado un pequeño algoritmo que calcula la inclinación producida. Este algoritmo se ayuda de unas líneas horizontales y verticales que poseen las plantillas que se han utilizado, como se pueden apreciar en las figuras 3.6, 3.12, y 3.13. Estas líneas han sido introducidas en la plantilla para la detección de la inclinación, para no tener que corregirla toda, lo que implicaría un gran coste computacional. Con la detección de la pendiente de una de una de estas líneas se puede averiguar si la inclinación supera el umbral o no. Si se observa la figura 3.15, se muestra como utilizando la razón trigonométrica de la tangente, relacionando el cociente de la diferencia en píxeles con respecto a los puntos (x_1, y_1) y (x_2, y_2) , se puede calcular el ángulo de inclinación, tal y como se recoge en la expresión del cálculo del ángulo en la ecuación 3.1.

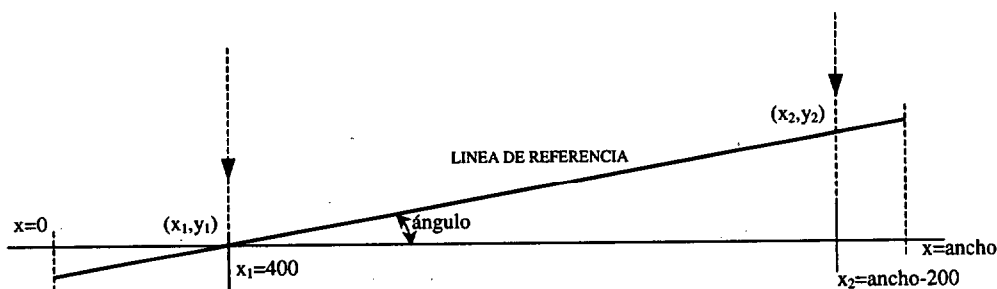


Figura 3.15: Inclinación de la línea de referencia de la plantilla y cálculo del ángulo.

$$\text{ángulo} = \frac{180}{\pi} \arctg\left(\frac{y_2 - y_1}{x_2 - x_1}\right) \quad (3.1)$$

En el supuesto de que la imagen quede por encima del umbral fijado por [O’Gorman95], se desecha la plantilla escaneada y se vuelve a repetir el proceso. Como cuestión práctica, cuando se realiza el escaneado de una plantilla, generalmente no se suele sobrepasar este umbral. De hecho, en la elaboración de estas bases de datos, nunca ocurrió que se superase.

3.3.3 Separación de las muestras

Una vez ya se tengan todas las plantillas escaneadas correctamente, el siguiente paso es la separación de las muestras. En los distintos formatos de plantillas, se fijan unos patrones comunes a todos:

- las líneas horizontales y verticales
- el cajetín para obtener la información de los escritores y las instrucciones
- la rejilla, donde están todas las cuadrículas con las muestras

Por tanto, es muy fácil estimar la posición de cada uno de estos elementos dentro de cada formato. Una vez determinada la posición de las muestras, éstas son binarizadas y finalmente se les elimina el posible ruido existente, residuo intrínseco de escanear. Este proceso es descrito en la siguiente figura, mediante un ejemplo con letras mayúsculas;

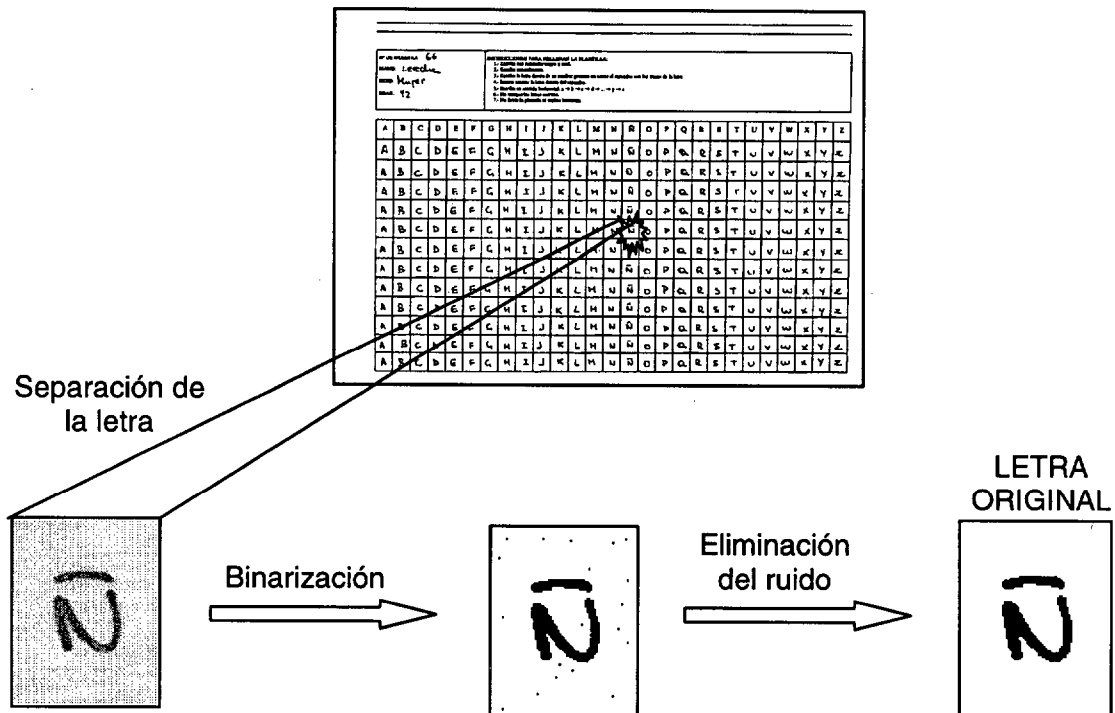


Figura 3.16: Preprocesado de la muestra de una letra mayúscula.

El primer paso será por tanto, extraer el enrejado de las muestras del resto de la plantilla. Para ello, se utiliza una herramienta muy básica como es el histograma de la imagen de la plantilla binarizada, para destacar los valores de blanco (no escrito) y negro (escrito). De esta forma, se marcarán las líneas horizontales y verticales como rectas, ya que tienen una inclinación menor a 0.5° . Estas rectas que se usan para aislar las muestras, se utilizarán para determinar la posición de cada una de ellas. En la figura 3.17, se muestra el histograma vertical y horizontal para el ejemplo de la plantilla de letras mayúsculas, cuyos máximos devuelven la posición de las cuatros esquinas de la cuadrícula de cada letra. Como el formato de las plantillas es conocido, entonces se puede eliminar las líneas de corrección de inclinación y el cajetín de datos, para retener

el enrejado. La combinación de todos estos máximos referentes al enrejado, como filas y columnas de una matriz, determinan la posición de cada muestra.

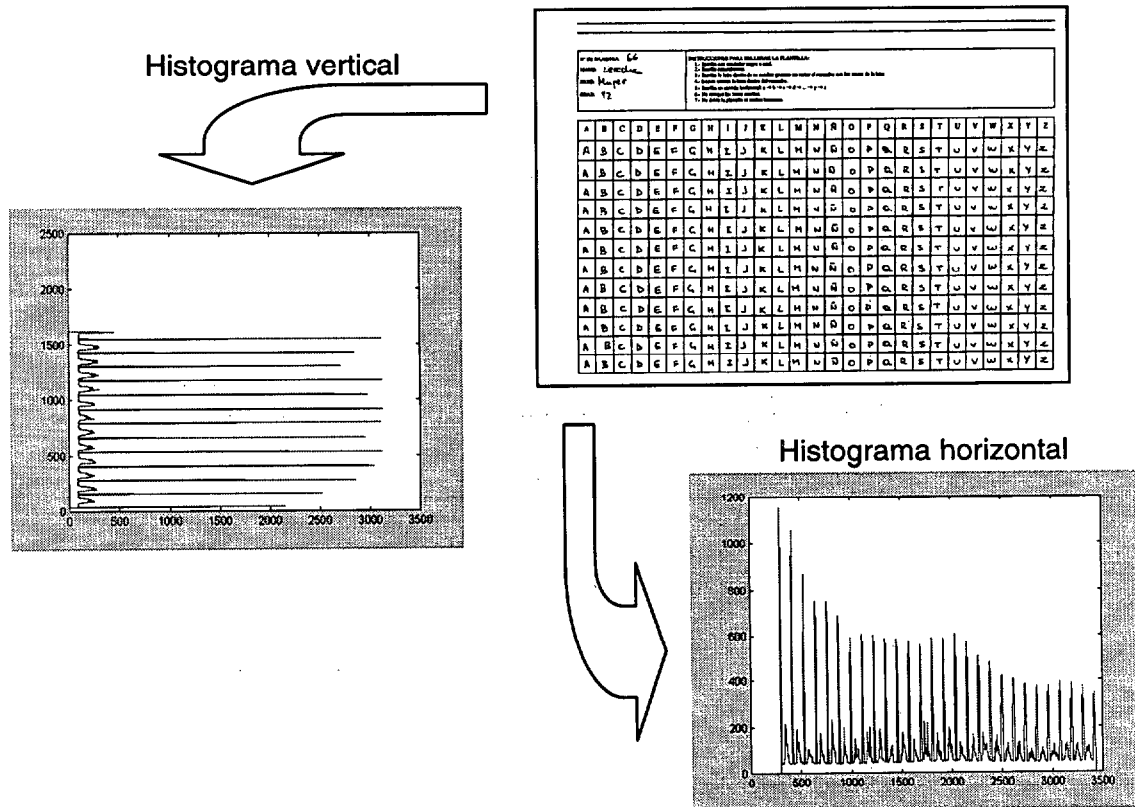


Figura 3.17: Histograma para detectar las rectas horizontales y verticales.

3.3.3.1 Binarización de los datos

Una vez separadas las muestras se realiza su binarización. Mediante este paso se convierte el mapa de bits, de escala de grises, como fue escaneado; a valores de blanco y negro. De esta forma se consigue representar la imagen con dos valores de intensidad, un '1' para el color negro y un '256' para el color blanco, eliminando la información de los valores intermedios de escala de grises, logrando una representación más simple de las imágenes.

Para realizar la binarización se eligió un umbral entorno al valor de intensidad 200. Por debajo del umbral todas las intensidades de grises son consideradas como color negro y las que toman valores superiores a 200 se hacen corresponder con el color blanco. En el histograma de una muestra (véase la figura 3.18), se observa que se incluyen todos los valores de grises menores a 200 como señal escrita. El efecto de incluir estos valores es engrosar las muestras de dígitos y letras, beneficioso en cualquier caso, aunque aparezcan ciertos píxeles negros aleatoriamente en la imagen, que no plantean ningún problema porque serán eliminados en la siguiente etapa. Se

muestra los histogramas con ejes lineal y con eje logarítmico en accisas, para observar la existencia de estos píxeles.

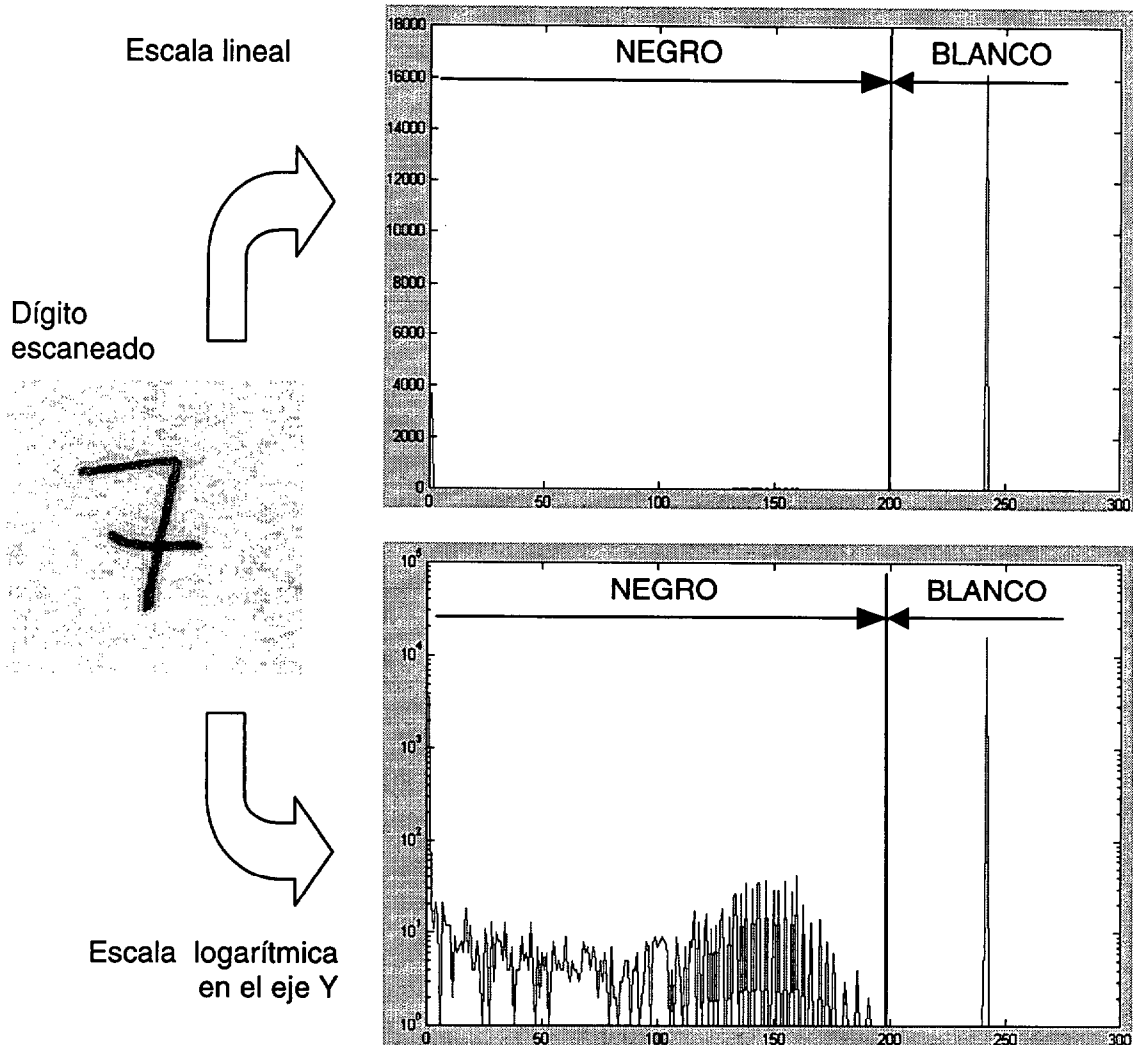


Figura 3.18: Histograma lineal y logarítmico del dígito manuscrito 7.

3.3.3.2 Eliminación del ruido

La principal fuente de ruido que se ha encontrado en las muestras binarizadas es el ruido salpicado [O’Gorman92][Chinnasarn98], denominado ‘*salt and pepper*’ en la lengua inglesa, que es el ruido que prevalece en los documentos binarios escaneados. Este tipo de ruido consiste en píxeles negros que aparecen distribuidos de forma aleatoria, salpicados, sobre zonas blancas de la imagen (el fondo de la muestra); y píxeles blancos que salpican zonas negras (el trazo de la muestra).

El filtrado del ruido ‘*salt and pepper*’ se realiza mediante la técnica “kfill”. Se ha elegido este filtro, frente a otros filtros morfológicos, porque se trata de un filtro conservativo: cuando se trata de elegir entre eliminar un píxel que puede ser ruido salpicado o puede pertenecer a un trazo deseado, el filtro kfill no actúa. Para conseguirlo, mantiene aquellos píxeles que están situados en una esquina y forman un ángulo de 90° o menos.

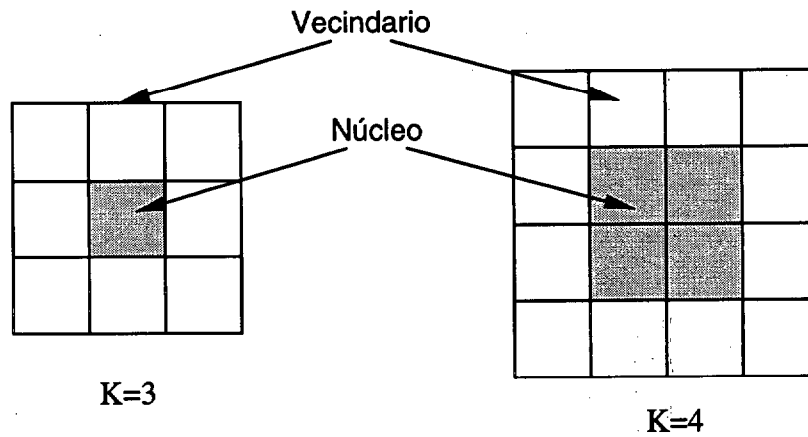


Figura 3.19: Ejemplos de máscaras para el filtrado kfill.

El filtro kfill utiliza una máscara (véase la figura 3.19) de tamaño $k \times k$ píxeles para recorrer toda la imagen y detectar los puntos que son ruido salpicado, para finalmente eliminarlos.

La máscara consta de un núcleo formado por los $(k-2) \times (k-2)$ píxeles interiores y un vecindario formado por los $4 \times (k-1)$ píxeles que rodean el núcleo. El parámetro k define el tamaño del ruido salpicado: para $k=3$ el tamaño del núcleo es de un píxel por lo que se procesarán todos individuales, permitiendo eliminar el ruido de tamaño uno. Para $k=4$ el tamaño del núcleo es 2×2 píxeles por lo que se procesará todo el ruido salpicado de este tamaño. En este proyecto se estableció $k=3$ porque se observó que la gran mayoría del ruido salpicado estaba formado por un único píxel.

La imagen es recorrida con esta máscara de izquierda a derecha y de arriba abajo, buscando los píxeles debidos al ruido salpicado, que cumplen la condición indicada por la ecuación 3.2. Una vez detectado uno de estos píxeles se procede al relleno del núcleo de la máscara, 'filling' en inglés, que consiste en cambiar su valor: de blanco a negro o viceversa. El relleno a negro (blanco) se produce cuando todos los píxeles del núcleo son blancos (negros), y los píxeles del vecindario cumplen la siguiente condición;

$$(c = 1) \text{ and } \{ \{ (n > 3) \text{ or } [(n = 3k - 4) \text{ and } (r = 2)] \} \} \quad (3.2)$$

donde:

- c: número de grupos de píxeles negros (blancos) conectados en la vecindad.
- n: número de píxeles negros (blancos) en la vecindad.
- r: número de píxeles negros (blancos) en las esquinas de la vecindad.

La ecuación 3.2 asegura que el filtro kfill sea conservativo. El valor $c=1$ garantiza que el proceso de relleno no cambie la conectividad, es decir, que no una dos trazos diferentes de una muestra, ni separe dos trazos que estén unidos.

El proceso de relleno, que se encuentra en la figura 3.20, se realiza hasta que no se produce ningún cambio. Cada iteración de relleno se divide en dos subiteraciones: en la primera se rellena con píxeles blancos y en la segunda se hace con píxeles negros.

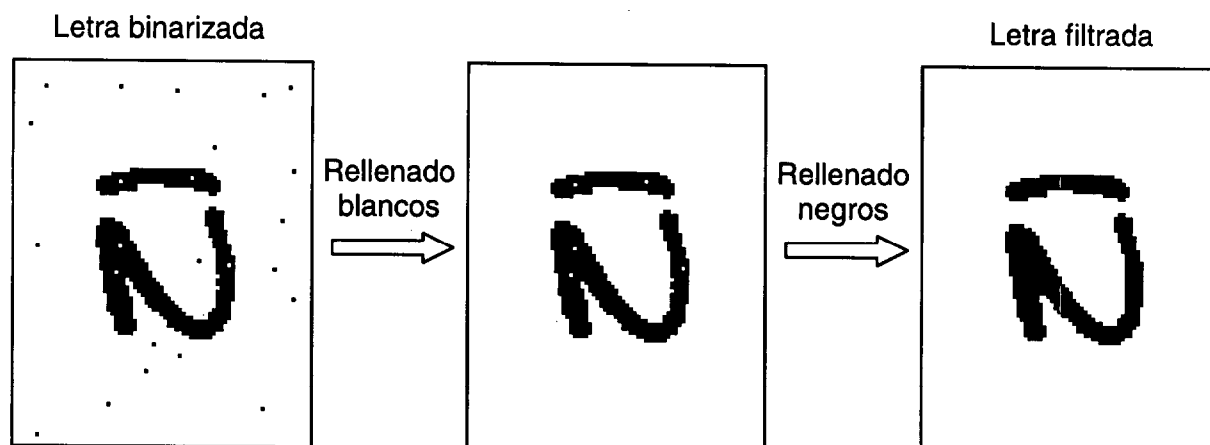


Figura 3.20: Eliminación del ruido salpicado.

3.3.4 Proceso de normalización

Se van a crear dos bases de datos para experimentar, una de muestras de tamaño original, según ha sido dispuesta por cada escritor, y otra base de datos que esté normalizada. Es decir, cada escritor no escribe de la misma forma, ni tampoco del mismo tamaño, por tanto, en aquellas bases de datos donde se precisa reconocer un patrón que sea general para todos los escritores, interesa que estos patrones se parezcan lo más posibles, en cuanto a forma y tamaño, disminuyendo así su variabilidad. Por el contrario, también puede producir un efecto de semejanza entre diferentes patrones. La ventaja o no de este proceso se observará en los resultados alcanzados, según la mayor discriminación en el reconocimiento.

Este efecto se presenta en nuestras dos bases de datos, tanto para dígitos como para letras manuscritas, lo importante es que este factor puede ser discriminante en el reconocimiento. Por tanto, para estas bases de datos, se va a crear los dos tipos paralelos, una para la muestra original del escritor y otra para la muestra reducida, pero ambas con un formato preestablecido. En la siguiente figura se puede apreciar como los escritores realizan los mismos patrones de forma y tamaño original diferentes;



Figura 3.21: Escritura por parte de diferentes escritores del patrón "a".

Los tamaños que se van a utilizar para la base de datos normalizada son de dos tipos:

- aproximado al tamaño original
- formato más reducido

Con ésto se logra que la varianza de un tipo de patrón sea menor, por lo que es previsible que los clasificadores precisen un entrenamiento menor. Además, la correlación cruzada de los patrones iguales aumenta, por lo que se podría mejorar el

porcentaje de acierto del clasificador. Y con el tamaño más reducido se intenta disminuir más aun esta varianza en un tipo de patrón, así como agilizar más los procesos, desde el punto de vista de coste computacional.

El proceso para llegar a estas bases de datos, pasa por realizar una tarea de ajuste a partir de la letra ya filtrada, de tal forma que la muestra disminuye el tamaño del archivo, por lo que se requiere menos memoria para almacenar los datos y procesar cada muestra, disminuyendo además, el tiempo de procesado, ya que existen menos píxeles. Después de la normalización, se puede almacenar la base de datos de las muestras originales de los escritores, y por último, queda la reducción y el almacenamiento de los datos normalizados. En la figura 3.22 se puede observar el proceso de normalización y el almacenamiento de la base de datos.

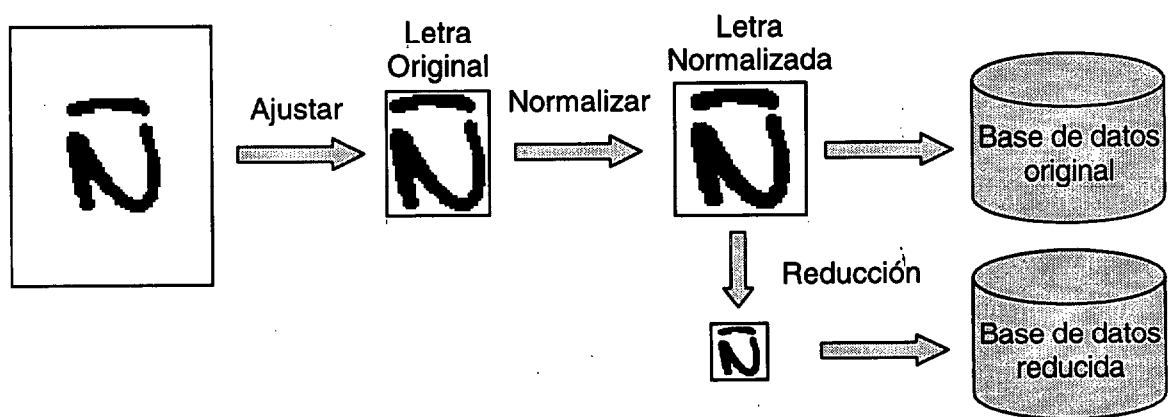


Figura 3.22: Proceso de normalización.

En este proceso de reducción se han fijado una serie de valores para determinar el tamaño de las muestras normalizadas. Lógicamente, este parámetro varía en función de la base de datos que se esté tratando [Travieso99]. En la siguiente tabla, se observan los valores en píxeles del mapa de bits de la muestra original y de la muestra normalizada, referenciadas en la figura 3.22;

Tipo de patrón	Tamaño original	Tamaño reducido
Dígitos manuscritos	162×162 píxeles	32×32 píxeles
Letras manuscritas	162×162 píxeles	50×50 píxeles

Tabla 3.3: Tamaños de los ficheros de las bases de datos en píxeles.

3.3.5 Operador morfológico: Dilatación

La palabra morfología usualmente denota la rama de la Biología que se dedica al estudio de la forma y estructura de los animales y plantas. Se usará aquí la misma palabra en el contexto de morfología matemática.

La morfología matemática se basa en geometría y forma, las operaciones morfológicas simplifican imágenes y conservan las principales características de formas de los objetos. Los puntos en un conjunto sobre los que se aplica la transformación son

el conjunto de puntos seleccionado y el complementario, el no seleccionado. En las imágenes binarias los puntos seleccionados son los que no pertenecen al fondo. Las operaciones primarias morfológicas son la erosión y la dilatación. En este caso, se va a hacer uso para la ampliación del grosor de la muestra y por tanto, se utilizará la dilatación.

La dilatación es la transformación morfológica que combina dos vectores utilizando la suma. La dilatación binaria fue usada primero por Minkowski, y en la literatura matemática recibe el nombre de suma de Minkowski. Si A y B son conjuntos en un n-espacio E_n con elementos $a = (a_1, \dots, a_n)$ y $b = (b_1, \dots, b_n)$, respectivamente, siendo ambos n-espacio, entonces la dilatación de A por B es el conjunto de todos los posibles vectores que son suma de pares de elementos, uno de A y otro de B [Serra82].

En la práctica los conjuntos A y B no son simétricos. El primer elemento de la dilatación, A, está asociado con la imagen que se está procesando y el segundo recibe el nombre de elemento estructural (B), la forma que actúa sobre A en la dilatación para producir $A \oplus B$.

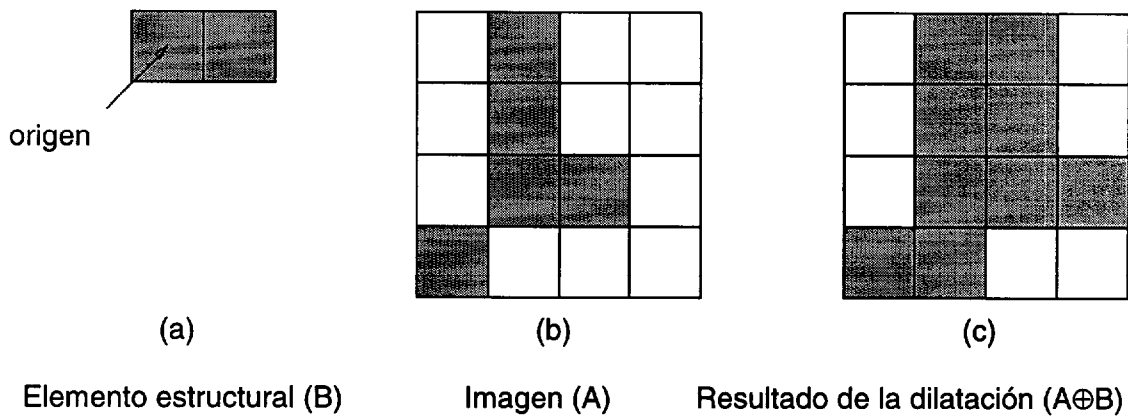


Figura 3.23: Ejemplo de dilatación.

Aplicado a nuestras bases este concepto, se obtiene la ventaja de hacer que las muestras sean más iguales y las muestras tengan un mayor grosor. En la siguiente figura, se puede observar el efecto que produce y la ventaja que se obtiene.

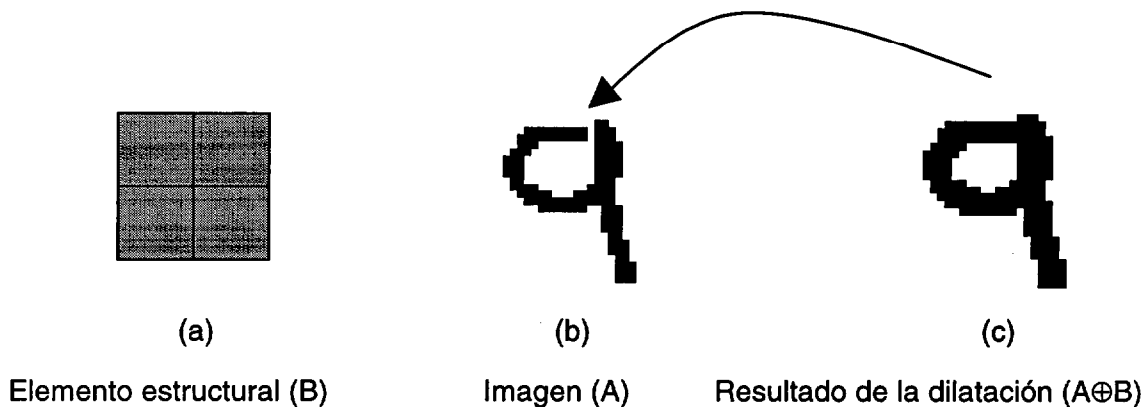


Figura 3.24: Dilatación del número nueve.

En esta figura 3.24, se observa que después de realizar la dilatación morfológica, el dígito queda totalmente cerrado en todo su trazo, logrando el efecto de que todos los dígitos se parezcan más.

3.4 Conclusiones

Este proceso va a permitir generar las bases de datos, en este caso de dígitos y letras manuscritas, transportando la información de un papel a una imagen, que ya estará dispuesta para su transformación o procesado. En este caso, será para la extracción de los parámetros que definirán a cada patrón para su posterior clasificación.

Este proceso seguido en la creación de la bases de datos y el preprocesado de imagen aplicado, se podrá extrapolar al resto de las bases de datos que se describen en los siguientes capítulos, creando un sistema de igual condiciones para todas.

Capítulo 4

Clasificación de manuscritos simples: dígitos y letras

4.1 Introducción

En el capítulo anterior se ha definido la base de datos de dígitos y letras manuscritas, y se ha caracterizado para que sean imágenes. En este capítulo, se van a parametrizar dichas imágenes para adecuarlas al clasificador HMM.

Teniendo en cuenta que los modelos ocultos de Markov, es una técnica muy usado en procesado de voz, ya que realizar un buen seguimiento del modelo de dicha secuencia natural; se van a implementar unas técnicas para la parametrización para se amolden al seguimiento de la secuencia de estas aplicaciones manuscritas, para su posterior reconocimiento.

4.2 Parametrización de los patrones manuscritos simples

En este apartado se presentan los principales procesos de extracción de la información secuencial de un patrón manuscrito. Estas técnicas se dividen en dos grupos, por un lado los procesos de detección de contorno y por el otro los de esqueletización.

La técnica de detección consiste en eliminar los píxeles del interior del trazo y mantener los extremos sin perder la conectividad entre ellos, y la técnica de esqueletización, adelgaza el trazo eliminando los píxeles más externos del trazo hasta que su grosor sea de solo uno, manteniendo también la continuidad de la secuencia [Travieso99]. En la siguiente figura se puede apreciar el resultado visual del efecto de ambas técnicas, aplicado a una muestra de la base de datos de dígitos manuscritos.



Figura 4.1: Técnicas de contorno y esqueletización.

4.2.1 Procedimiento para detectar y vectorizar el contorno

Dentro de la detección de los contornos de una muestra, se pueden encontrar dos casos en los patrones que se están analizando; que sólo posean un contorno exterior, como es el caso de la figura anterior, o que tengan contornos exterior e interior, como es el caso de la siguiente figura;

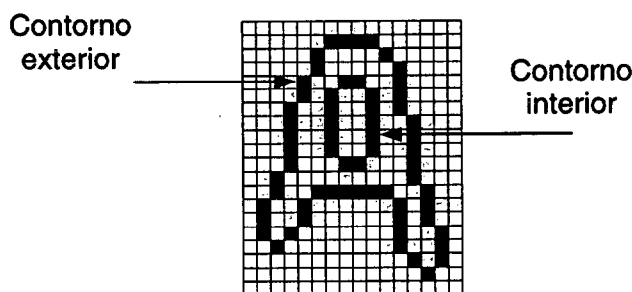


Figura 4.2: Contorno exterior e interior de la letra "A".

Y por tanto, con este planteamiento se puede diferenciar estos dos casos con la consideración del contorno, como composición del contorno interior y exterior, o tan sólo contar con el contorno exterior. Este aspecto es importante a la hora de vectorizar, ya que se puede incluir o no los interiores, según sea más o menos discriminantes para

el patrón. Pero de una forma o de otra, el proceso para el cálculo del contorno es común en ambos casos.

4.2.1.1 Desarrollo de la técnica del contorno

En el barrido para detectar el contorno se marcan los píxeles exteriores, y además, se detectan los puntos críticos o puntos de cruce entre trazos. Los puntos críticos son aquellos píxeles esquinas que si se eliminan separarían dos trazos (véase la figura 4.3 a), o aquellos píxeles que forman parte de un trazo de grosor de un píxel (ver figura 4.3 b).

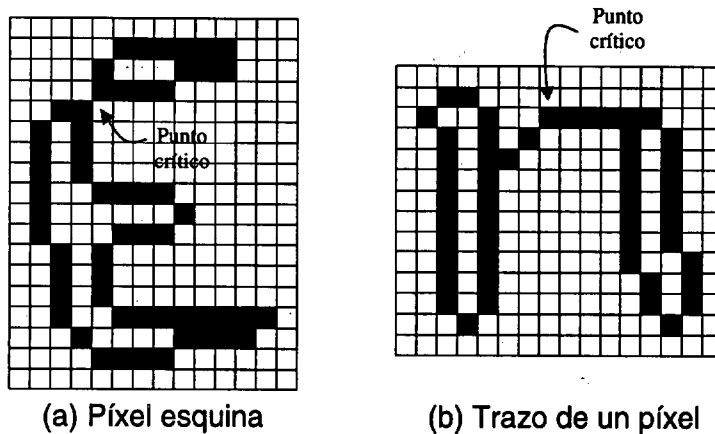


Figura 4.3: Puntos críticos.

Para ello, se recorre el patrón manuscrito mediante una máscara de tamaño 3×3, buscando grupos de píxeles que cumplan cualquiera de las condiciones de las matrices de la figura 4.4. Cuando se produce esta coincidencia es porque el píxel central de la máscara pertenece al contorno, y por indicar tal hecho, éste es marcado con un valor de intensidad de 3. Estas máscaras siempre buscan en vecinos cuatro, hacia la horizontal y la vertical en ambos sentidos.

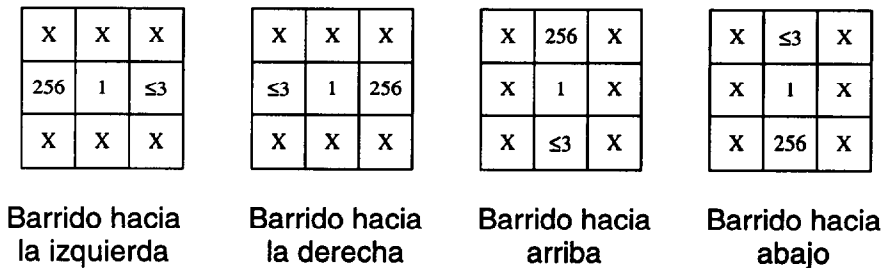


Figura 4.4: Matrices para detectar el contorno (X: no importa el valor).

Para detectar los puntos críticos, durante el barrido se aplican las condiciones indicadas por las matrices de la figura 4.5. Todos ellos deberán ser leídos dos veces para garantizar la continuidad del contorno por lo que se marcan con un valor de intensidad de 2.

Matrices para detectar puntos críticos en las esquinas.

X	X	X
256	1	X
≤3	256	X

≤3	256	X
256	1	X
X	X	X

X	256	≤3
X	1	256
X	X	X

X	X	X
X	1	256
X	256	≤3

Matrices para detectar trazos de un píxel de ancho.

X	256	X
≤3	1	≤3
X	256	X

X	≤3	X
256	1	256
X	≤3	X

256	X	≤3
X	1	X
≤3	X	256

≤3	X	256
X	1	X
256	X	≤3

Figura 4.5: Matrices para detectar puntos críticos.

En estos momentos del procesado, se tienen los siguientes valores de intensidad en los píxeles en la matriz de la imagen:

- Píxel de valor 1: píxel del interior del trazo
- Píxel de valor 2: píxel crítico del contorno del trazo.
- Píxel de valor 3: píxel del contorno
- Píxel de valor 256: píxel en blanco.

A la vista de estos valores queda claro que si se marca con el píxel de valor blanco (256), los píxeles del interior del trazo, de valor 1, llega a eliminarse el interior del trazo. A los píxeles del contorno con valor 3 se les asigna ahora el píxel de valor 1, correspondiente al color negro. Se mantiene el valor de los puntos críticos, 2, ya que se pasa dos veces por ellos, durante la vectorización.

El siguiente paso es detectar los puntos de cruce que pueden existir en el contorno para marcarlos como puntos críticos. Para ello, se recorre la muestra con una máscara de tamaño 3x3 en cuyo centro se sitúa el píxel candidato a punto de cruce, si en los píxeles de la máscara que lo rodean hay 6 ó más transiciones de color (de blanco a negro o de negro a blanco) se trata de un punto de cruce (véase figura 4.6).

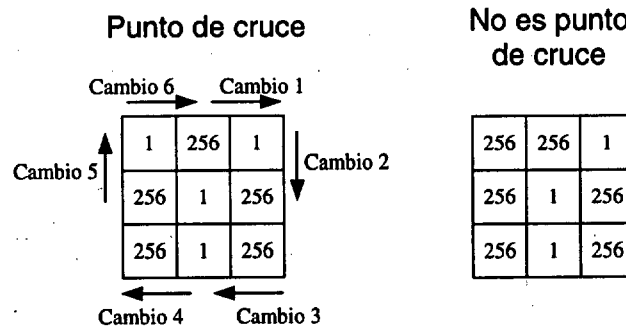


Figura 4.6: Transiciones en los puntos de cruce.

El último paso para obtener el contorno, consiste en eliminar los píxeles de esquina redundante. Un píxel esquina redundante es aquel píxel del contorno que se puede eliminar sin pérdida de continuidad en el trazo, como se muestra en la figura 4.7. Para eliminarlos se recorre la letra con una máscara de tamaño 3x3 y se compara ésta con el juego de matrices de la figura 4.8 y cuando se da una coincidencia el píxel del centro de la máscara es borrado.

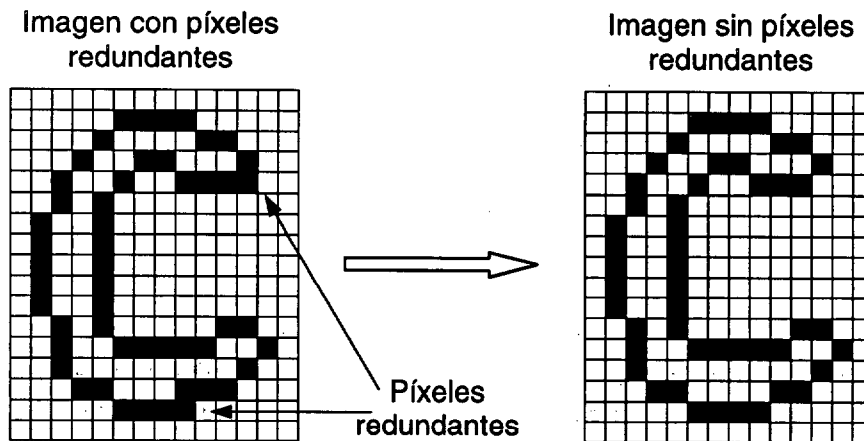


Figura 4.7: Ejemplo de píxeles esquina redundante.

256	256	X
256	1	1
X	1	256

X	256	256
1	1	256
256	1	X

X	1	256
256	1	1
256	256	X

256	1	X
1	1	256
X	256	256

Figura 4.8: Matrices para detectar píxeles esquina redundantes.

4.2.1.2 Proceso de vectorización del contorno por la secuencia

El proceso de vectorización del contorno consiste en pasar la imagen del patrón manuscrito a un vector, es decir, pasar de las dos dimensiones que tiene una imagen a una sola dimensión, que tiene un vector. Este proceso se realiza para captar la información de la secuencia de la escritura de cada patrón, y es la idea propuesta como forma de parametrizar los patrones manuscritos.

Partiendo del contorno de la imagen, lo primero es proveerse de un punto de inicio. El punto de inicio es el píxel del contorno por el que se comienza la vectorización. Su correcta elección es muy importante para conseguir que el mismo patrón, procedente de distintas realizaciones, siempre produzca la misma secuencia de observación.

Se pueden aplicar diferentes convenios para la elección del punto de inicio. En la técnica de parametrización por la técnica del contorno, según sea la aplicación se ha escogido un convenio u otro. En las letras manuscritas se ha tomado el convenio que propone elegir como punto de inicio a aquel píxel negro más próximo a la esquina inferior derecha de la matriz, y en el caso de que haya dos o más píxeles equidistantes se escoge el situado más a la derecha [Peinado00-pfc]. Para ello, se emplea la distancia euclídea, para representar esta mínima distancia, tal como se indica en la ecuación 4.1.

$$d = \sqrt{x'^2 + y'^2} = \sqrt{(N^\circ \text{ columnas} - x)^2 + (N^\circ \text{ filas} - y)^2} \quad (4.1)$$

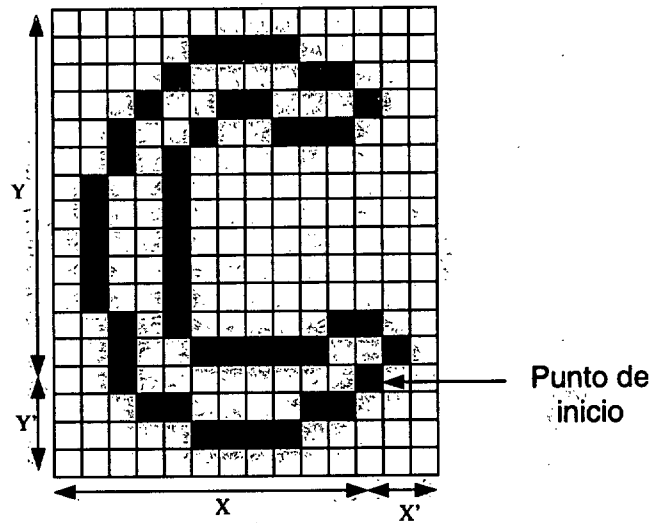


Figura 4.9: Convenio para aplicaciones con letras manuscritas.

Se escoge el punto de inicio en esta posición porque se ha observado que, para letras con varios extremos (A, H, h, m, n, R,...), la técnica clásica de elegir el punto de inicio como el primer píxel negro que se encuentra al barrer la imagen de izquierda a derecha y de arriba a abajo, unas veces proporciona el punto de inicio en el extremo derecho y otras en el izquierdo. Este fenómeno se produce porque el escritor hace unas veces el extremo derecho más largo y otras el izquierdo. Eligiendo el punto de origen como el píxel negro más cercano a la esquina inferior derecha se consigue evitar este problema, con lo que las secuencias de observación producidas, por letras iguales, tendrán mayor correlación y, por tanto, aumenta la probabilidad de reconocer la letra.

Para los patrones de dígitos manuscritos, debido a que son menos numerosos y por tanto, más fáciles de discriminar, se ha tomado como convenio del punto de inicio, el primer píxel negro que se encuentra al barrer la imagen de izquierda a derecha y de arriba abajo [Travieso99].

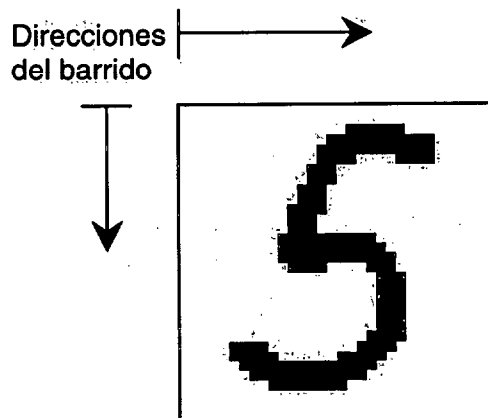


Figura 4.10: Convenio para aplicaciones con dígitos manuscritos.

Pero los valores que toma el vector vienen dados por las matrices de prioridad, para mantener siempre el mismo sentido de giro, y se activan según el valor del salto anterior. Si éste fue codificado con 1, 2, 3, 4 ó se trataba de un punto de inicio, se elige la matriz de prioridad 1. Pero si los valores del salto anterior fueron 5, 6, 7 ó 8, se utiliza

la matriz de prioridad 2, consiguiéndose que el giro se realice en el sentido de las agujas del reloj. Estas matrices se encuentran en la figura 4.11 [Morales99-pfc]. El desarrollo detallado del proceso del algoritmo se encuentra en la figura 4.36, que es el diagrama de flujo del proceso de vectorización del contorno.

4	5	6
3	P	7
2	1	8

8	1	2
7	P	3
6	5	4

Matriz de prioridad 1 Matriz de prioridad 2

Figura 4.11: Matrices de prioridad.

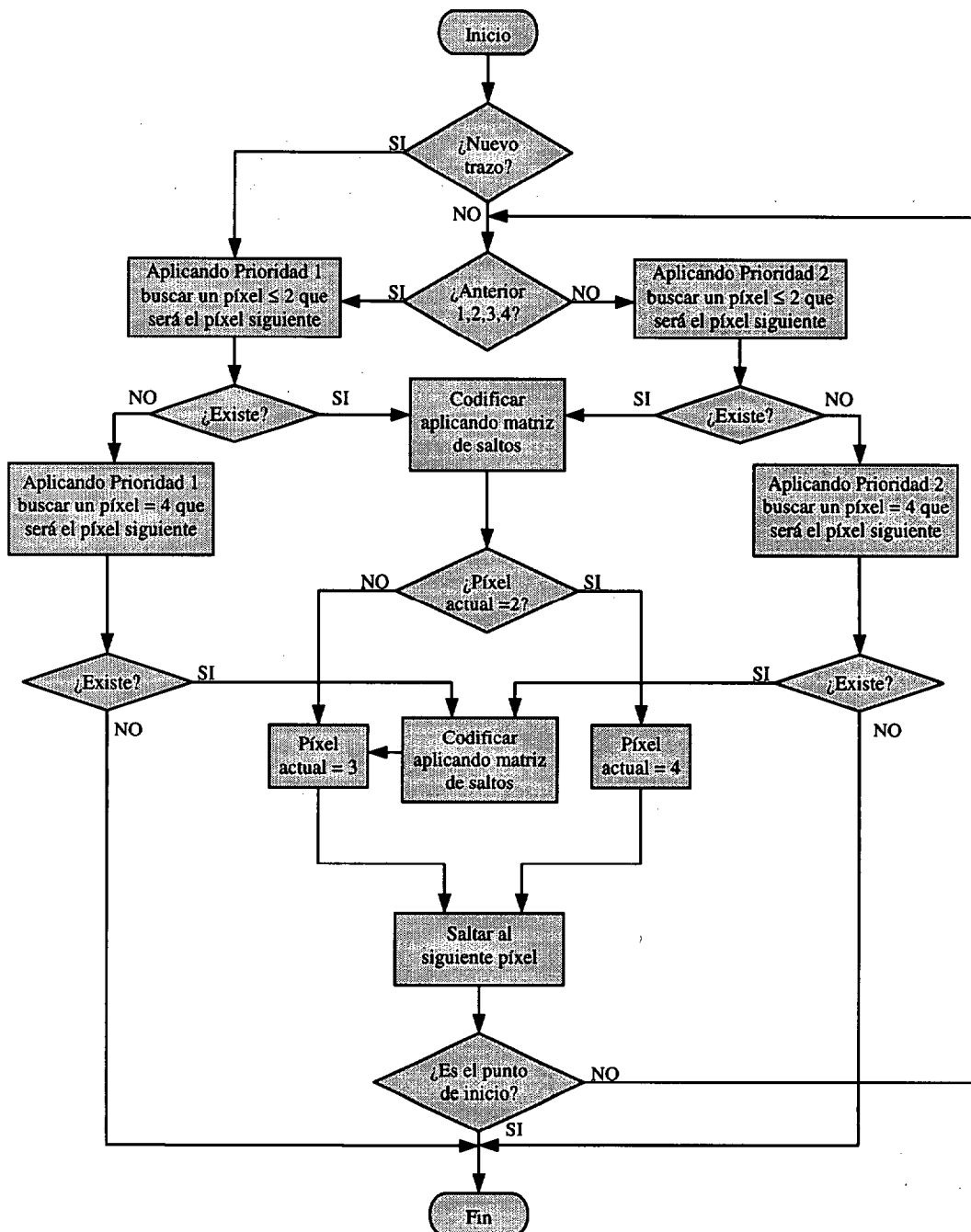


Figura 4.12: Diagrama de flujo de la vectorización del contorno.

La vectorización del proceso termina cuando se llega al punto de inicio, pero quedaría comprobar si quedan más contornos en la imagen. En caso afirmativo, los que quedan serán los correspondientes a los contornos interiores de los patrones manuscritos. El proceso a seguir es el mismo, con la particularidad, de que los valores de la vectorización se sitúan al final del vector calculado para el contorno exterior. De esta forma, se completa el proceso de vectorización.

4.2.1.3 Proceso de vectorización del contorno en coordenadas polares

Otra forma de realizar el paso de las dos dimensiones de la imagen (matriz) a una única dimensión (vector) es mediante la transformación de cada píxel del contorno externo de una muestra en su coordenada polar. La cuestión que se plantea es con respecto a que punto se puede realizar el paso a coordenadas polares, módulo y argumento (ρ, θ). En un principio, se van a proponer dos puntos como puntos de inicio, uno es la esquina inferior izquierda, igual que si se representase en los ejes de referencia, y la otra es con respecto al centro de masas de la muestra que se esté procesando. En la siguiente figura, se puede observar cuáles son estos dos puntos de referencia.

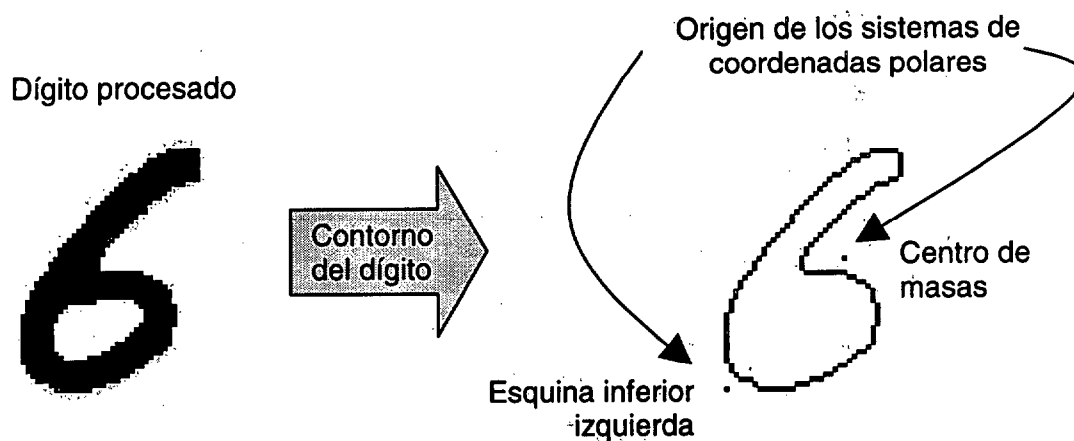


Figura 4.13: Origen de los sistemas de coordenadas polares para vectorizar.

Una vez establecido cual es el eje de referencia para estas coordenadas sólo hay que establecer el barrido por cada píxel de contorno. Llegados a este punto se puede plantear otro inconveniente con respecto a cual va a ser el punto de inicio del barrido. Realmente es invariante del punto de inicio, ya que una vez calculadas todas las coordenadas, se ordenan por el ángulo de inclinación de la componente polar (θ), empezando por aquellos que sea nula.

De esta forma, conservamos la información de la secuencia pero esta vez, codificada en coordenadas polares. En la figura 4.14, se indica cual es el proceso de barrido para una muestra manuscrita, distinguiendo el caso según sea el punto de origen del sistema.

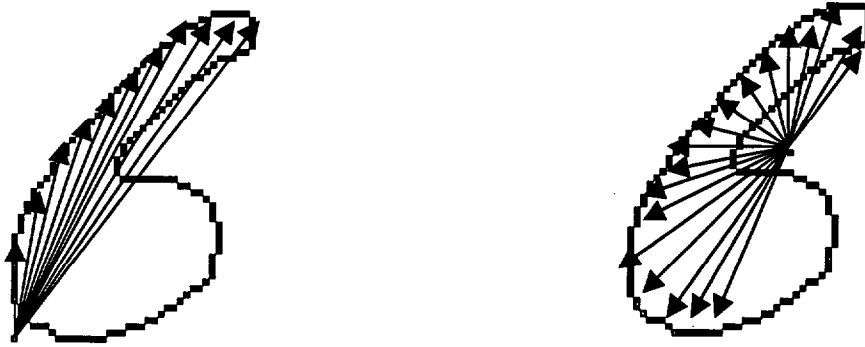


Figura 4.14: Barrido en coordenadas polares para determinar la secuencia de la muestra.

4.2.2 Procedimiento para detectar y vectorizar la esqueletización

Este proceso de esqueletización ha sido bastante utilizado para el reconocimiento de patrones de forma automática. La necesidad de reducir la información y comprimir los datos, ha dado con la aparición del proceso de esqueletización [O’Gorman95]. Esta técnica ya se ha venido usando en aplicaciones de biomedicina (análisis automáticos de imágenes de rayos X [Santos99]), en aplicaciones biométricas como son la clasificación de huellas dactilares [Weber92][Emiroglu97], y en música (para el cálculo del pentagrama y las notas musicales [Randriamahefa93]).

El propósito de la esqueletización es reducir los píxeles de la imagen hasta quedarse con la información esencial contenida en ésta y manteniendo las propiedades geométricas y topológicas, para de esta forma facilitar y acelerar los procesos de análisis y reconocimiento, quedándose la información en un solo trazo conectado por un píxel de grosor.

Los algoritmos de esqueletización o adelgazamiento se pueden clasificar según la referencia [Lam92], de la siguiente manera:

- Iterativos: Son aquellas técnicas que eliminan los píxeles exteriores tras sucesivos barridos o iteraciones, hasta que las líneas de la imagen tienen un ancho de un solo píxel. Según la forma de examinar los píxeles se pueden dividir en:
 - Secuenciales: con esta técnica los píxeles son examinados según una secuencia establecida en cada iteración, y la decisión se toma en función de las operaciones realizadas en la iteración anterior y en los píxeles procesados en la iteración actual, por tanto se examina los píxeles dentro de su contexto de su región de vecindad.
 - Paralelos: se realiza esta técnica mediante la decisión de la eliminación de todos los píxeles que no son necesarios para la conectividad del trazo, como resultado de las operaciones de la iteración anterior, por lo que se puede trabajar con toda la imagen.
- No iterativos: calculan los ejes de simetría de los trazos que componen la muestra manuscrita.

- Para imágenes 3D.
- Para imágenes en escala de grises.

Los requisitos que debe cumplir un buen algoritmo de esqueletización son [Morales99-pfc]:

- a) Mantener la conectividad: es decir, aquellas regiones que en la imagen original estaban conectadas deben estar conectadas en la imagen esqueletizada.
- b) La imagen esqueletizada debe contener el mínimo número de píxeles para que se mantenga la conectividad.
- c) La localización de los puntos finales debe ser mantenida para no reducir el tamaño de la imagen.
- d) El resultado obtenido del proceso de esqueletización debe ser lo más cercano posible a las líneas centrales de la imagen.
- e) Evitar introducir espurios extraños.

4.2.2.1 Proceso de la técnica de esqueletización

La elección de la técnica a utilizar se ha basado en la sencillez del algoritmo, y se ha optado por el algoritmo de esqueletización iterativo secuencial [Morales99]. La implementación de este algoritmo se basa en unas máscaras de 3x3 que barren toda la imagen. Estas matrices describen las formas geométricas de los trazos por lo que determinan si el píxel central de la máscara es un píxel esqueleto o no. El valor del píxel central será cambiado si cumple las condiciones indicadas por alguna de las matrices de las máscaras. Los valores de los píxeles afectados en este proceso deben seguir el orden indicado a continuación, primero los píxeles de valor 4, después los de valor 2 y finalmente los de valor 3;

4: para los píxeles que cumplen con las combinaciones de las máscaras

256	256	256
256	1	X
X	≤3	X

≤3	≤3	256
≤3	1	256
256	256	256

Figura 4.15: Máscara del proceso de esqueletización para detectar el 4.

2: para los píxeles que cumplen con las siguientes máscaras

X	X	X
256	1	256
X	X	X

X	256	X
X	1	X
X	256	X

Figura 4.16: Máscara del proceso de esqueletización para detectar el 2.

3: para los píxeles que cumplan con las siguientes máscaras y condiciones en cada una de las máscaras:

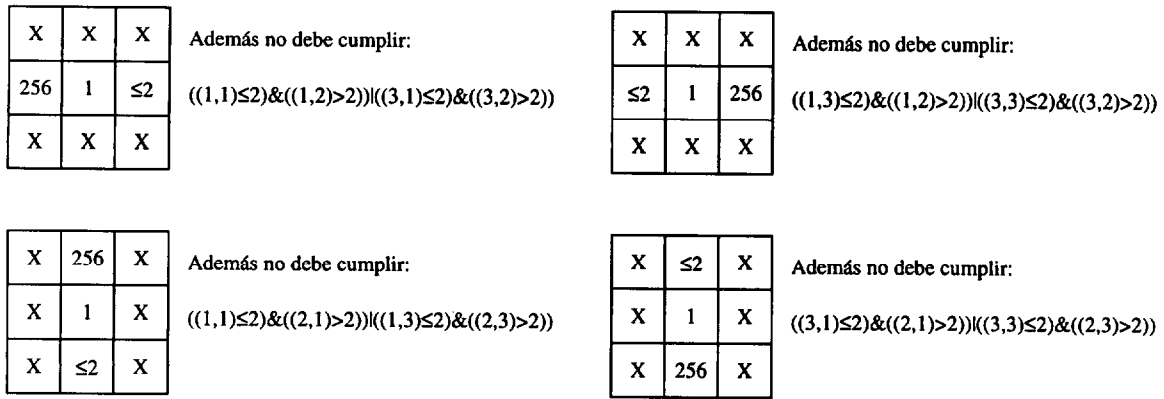


Figura 4.17: Máscara del proceso de esqueletización para detectar el 3.

El siguiente paso a seguir, es la eliminación de los valores marcados con 3 y con 4, ya que se tratarán de píxeles exteriores al esqueleto, y prevalecen los marcados con 2, ya que son píxeles del esqueleto, que garantizan la conectividad. Una vez realizado este segundo paso, se repiten el barrido de las matrices y la eliminación de los píxeles marcados con 3 y 4, hasta cuando la imagen esqueletizada de la iteración sea igual a la actual, con lo que se ha logrado la esqueletización de la imagen.

En la siguiente figura se muestra la secuencia de pasos para obtener el patrón esqueletizado de la muestra dada. En este caso en particular, se presenta como ejemplo la esqueletización del patrón numérico siete, en un total de cinco iteraciones hasta su conversión final.

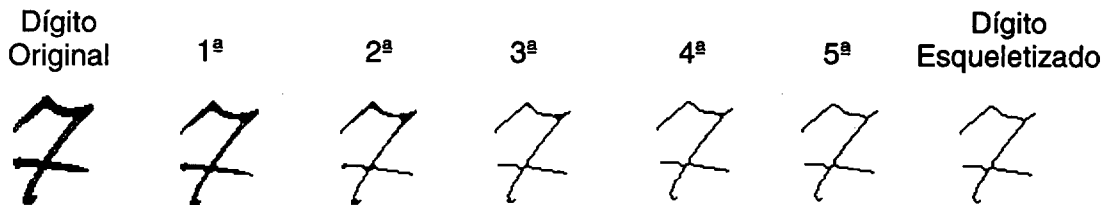


Figura 4.18: Proceso de esqueletización.

4.2.2.2 Vectorización de la esqueletización

Partiendo de estas imágenes esqueletizadas, ya se puede crear el vector que contenga las variaciones del trazo de los patrones manuscritos. Al igual que se realizó en la técnica del contorno, se va a pasar de dos dimensiones que poseen las imágenes esqueletizadas, a una dimensión que posee un vector, obteniéndose como resultado los parámetros que se situarán en la entrada del clasificador.

En algunas aplicaciones, se ha encontrado el caso, de tener que realizar un preprocesado para subsanar, algunos inconvenientes propio de la escritura manuscrita. En particular, se han encontrado dos casos una vez obtenido la imagen esqueletizada. El primero es la separación del trazo en dos partes, debido a la propia escritura del patrón manuscrito, y la segunda es la aparición de coletillas en el trazo del patrón. La corrección de estos casos, facilita en gran medida la realización del proceso de vectorización de las imágenes esqueletizadas, y por eso se realiza su eliminación. En la siguiente figura, se pueden observar estos dos casos y el resultado de su corrección;

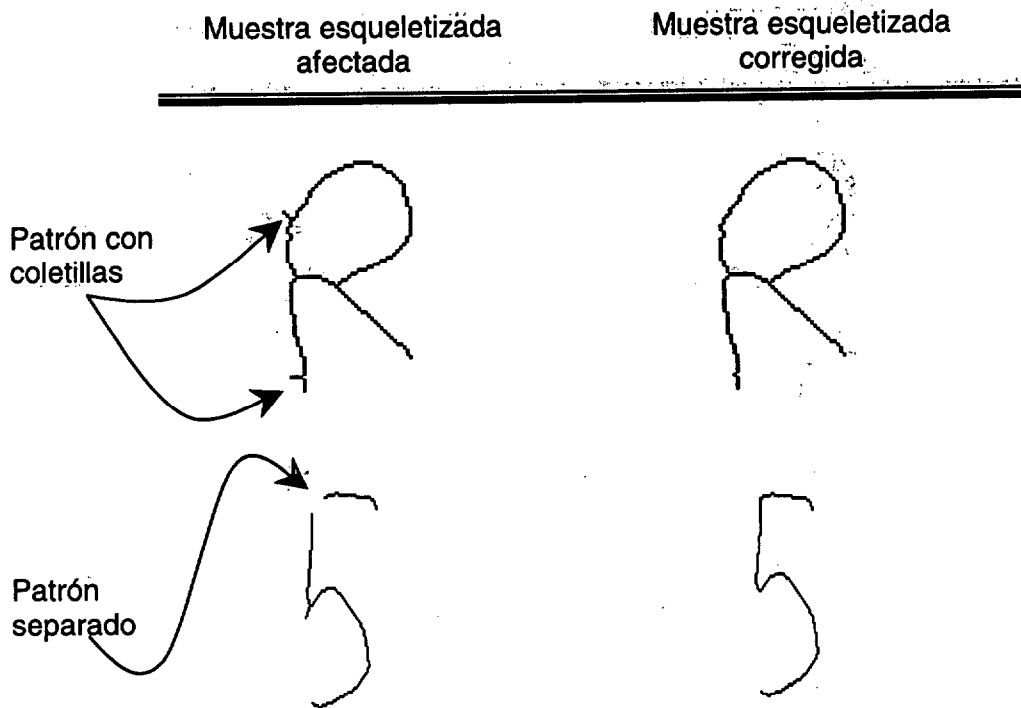


Figura 4.19: Preprocesado para la mejora de las imágenes esqueletizadas.

En el caso de mantener estos espúrios, se complicaría la realización del proceso de vectorización de los patrones ya esqueletizados. Los pasos a seguir a continuación siguen las mismas pautas que en el caso de la vectorización de los contornos. Primeramente se determina el punto de inicio, y a continuación se vectoriza la imagen, pasando de dos dimensiones (imagen) a una sola dimensión (vector).

La elección del punto de inicio se dispone a partir de una serie de elementos geométricos que pueden ser elegidos, como son los extremos del esqueleto; siendo el extremo el punto donde termina el trazo del esqueleto. Se puede tomar como referencia entre los extremos, a aquel punto de inicio que esté más cercano a la esquina inferior derecha. Interesa que este punto sea siempre el mismo para patrones iguales, ya que de ello depende, que los vectores codificados sean lo más parecido posible, para su posterior clasificación.

La vectorización del esqueleto de un patrón utiliza los mismos conceptos que la vectorización del contorno, hay una matriz de saltos y otra de prioridad, y en ambos casos la matriz es la misma.

3	4	5
2	P	6
1	8	7

Figura 4.20: Matriz de saltos y prioridad en la vectorización del esqueleto.

El diagrama de flujo de la vectorización del esqueleto se explica en la figura 4.21, siguiendo cada uno de los pasos de su secuenciación.

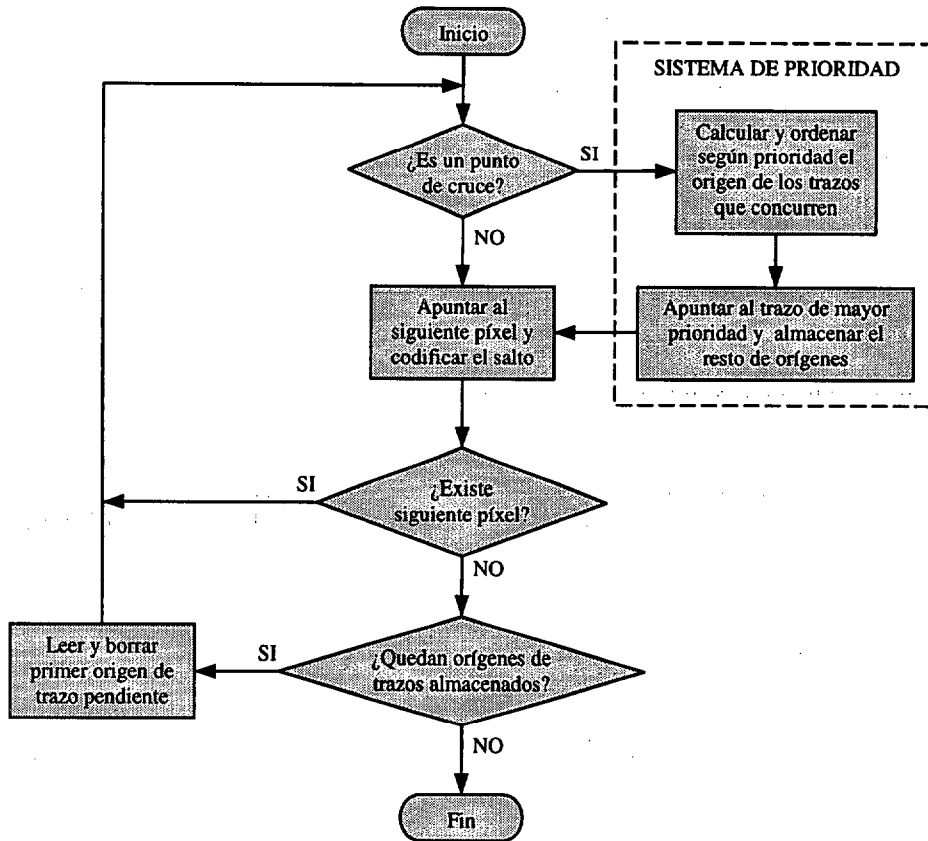


Figura 4.21: Proceso de vectorización del esqueleto.

La matriz de prioridad busca los trazos concurrentes en el cruce, calcula sus pendientes y comprueba si alguna de las pendientes es aproximadamente igual a la del trazo por el que se ha llegado al cruce. De ser así, ese será el de mayor prioridad y al resto se le aplica la matriz de prioridad de la figura 4.20. En caso contrario, se aplica la matriz de prioridad para elegir el orden en que se seguirán todos los trazos.

Los valores de los parámetros del sistema de prioridad que se presentan a continuación han sido obtenidos experimentalmente, demostrando ser los más adecuados:

- Si hay dos cruces cuya distancia sea inferior a tres píxeles, serán considerados el mismo punto.
- Para calcular la pendiente de un trazo se emplean los 5 píxeles siguientes a partir del punto de cruce. Si el trazo tiene sólo 4 píxeles se emplearán éstos. Nunca habrá trazos de menos de 3 píxeles ya que éstos se eliminan en el preprocesado.
- Para considerar que el trazo por el que se llega al cruce continúa en uno de los trazos concurrentes, la diferencia entre sus ángulos deberá ser inferior a $\pm 45^\circ$.

Finalmente, se obtiene el vector con la información de cómo ha sido realizado el trazo de los patrones, obteniendo estos datos secuenciales reflejados en ese vector.

4.2.3 Discusión entre las técnicas de esqueletización y contorno

Si se plantea cual de las dos técnicas ofrecerán mejor resultado, es difícil de estimar, ya que depende de la comprobación experimental de esta prueba. Pero a priori, y según la propia naturaleza de las técnicas si se puede presentar un planteamiento de las ventajas y los inconvenientes de su uso.

Los puntos a comparar son la dificultad del proceso de contorno o esqueletización, la determinación del punto de inicio, y finalmente la vectorización de ambos métodos.

En cuanto al primero punto, la implementación del algoritmo del contorno o la esqueletización, ambos métodos resultan sencillos de realizar, sin ofrecer mayor dificultad.

En la elección del punto de inicio, las imágenes esqueletizadas se prestan a una mayor eficacia a la hora de determinar siempre el mismo punto, ya que los extremos en un patrón son relativamente escasos, y fijar un criterio es más sencillo que en el caso de las imágenes con contorno. Ya que éstas, siempre dependen de cómo ha sido su escritura, y se aproxima el punto de inicio a una cierta región, pero no se puede dar la certeza de que sea siempre en el mismo lugar.

Por último, la vectorización de ambos procesos sufre diferencias. Mientras con la técnica del contorno, se empieza en un punto y termina en el mismo, ya que barre un trazo cerrado, en la esqueletización se pueden encontrar puntos de cruce, que hay que almacenar y criterios que hay que tomar para que los vectores de los patrones sean lo más iguales posible. Para este caso, se complica más el proceso de la vectorización para el esqueleto, que para el contorno.

Si se hace un balance de todos los puntos planteados, se observa que tanto hay ventajas como inconvenientes para cada uno de las técnicas de vectorización, y por tanto será el clasificador, el que decida cual de los dos procesos propone vectores más discriminadores.

4.3 Clasificación de los patrones simples manuscritos

Una vez calculado el vector del patrón, éste se situará a la entrada del clasificador. Teniendo en cuenta la propiedad de este vector, que ha captado la información que sigue el trazo cuando se escribe el patrón manuscrito, conviene usar un clasificador que detecte bien las variaciones secuenciales-temporales, es decir, la secuencia que sigue el trazo del patrón manuscrito a lo largo del tiempo de su barrido.

Un clasificador muy utilizado y con éxito en el reconocimiento de la señal de voz, son los modelos ocultos de Markov [Rabiner93]. Por tanto, se probará con este mismo clasificador para la información contenida en los parámetros secuenciales-temporales, que se puedan encontrar en las aplicaciones manuscritas simples, ya que se puede presuponer un buen resultado en la clasificación de estos patrones.

La clasificación se realiza en dos etapas. En la primera, se crea el modelo de las secuencias de observación, y se denomina etapa de entrenamiento. La segunda etapa

consiste en usar el modelo creado en la anterior etapa, para reconocer con él, las secuencias de observación, se trata de la etapa de test. Se utiliza por tanto, un clasificador estadístico supervisado, ya que calcula la probabilidad de las diferentes clases frente a una serie de parámetros, siendo previamente entrenado. Tanto los programas utilizados en la clasificación para la etapa de entrenamiento como de test, han sido desarrollados por el Grupo de investigación de Procesado de Digital de la Señal del Departamento de la Señales y Comunicaciones de la Universidad de Las Palmas de Gran Canaria.

4.3.1 Etapa de entrenamiento para el clasificador HMM discreto

En primer lugar se definen una serie de variables para caracterizar el tipo de cadena oculta de Markov. Estas variables son;

- Número de estados del modelo. Este número de estados está interconectado entre sí, y varía con los parámetros del patrón analizado. Según la aplicación se juega con este parámetro para optimizar la tasa de reconocimiento.
- Número de símbolos de observaciones por estado. Indican los símbolos de observación correspondiente a la salida física del sistema que se está modelando. Por ejemplo, para la vectorización del contorno codificado por la matriz de 3×3 , el valor viene dado por la dirección que sigue el trazo y es la matriz de saltos la que nos indica los valores de los símbolos, del 1 al 8, en total ocho símbolos.
- Sistema BAKIS. Con este sistema se desea un modelo de izquierda a derecha. Por tanto, sólo se permiten la transiciones en este sentido entre los estados, buscando dar más importancia a la información que se obtiene en la secuencia de la escritura.
- Finalización por umbral. Indica el valor mínimo de la distancia entre dos modelos consecutivos, para que se produzca otra iteración, ya que se consideran que estos modelos son muy parecidos y ofrecerán similares tasas de reconocimiento. En el caso de que se cumpla un número máximo de iteraciones y no se haya alcanzado este umbral, también finalizará el proceso. Por tanto, hay dos formas de acabar las iteraciones para el cálculo de cada modelo.

Una vez que se han fijado estas variables se necesita introducir la secuencia de vectores de entrada y su resultado correcto de reconocimiento para generar el modelo oculto. Los vectores de entrada serán finitos, pero deben ser las suficientes ocurrencias de los diferentes sucesos para que se obtenga una buena estimación de los parámetros del modelo. Si es necesario se deberá de reducir o aumentar el número de observaciones, para crear el entrenamiento más óptimo, pero son cuestiones que se deberán aplicar sobre la interpretación del resultado de los experimentos.

Después de la captación inicial de estas variables, se genera de manera aleatoria las siguiente matrices:

- La matriz de transición de estados (A), que indica la probabilidad de que cualquier estado puede alcanzar otro estado.

- La matriz de probabilidad de símbolos (B), que determina la probabilidad de que estando en un cierto estado ocurra una observación determinada.
- La matriz de la probabilidad de estados iniciales (π).

Como se puede concluir, para definir un modelo HMM son necesarios los parámetros de número de estados, número de símbolos por observación, y las matrices A , B y π para determinar el modelo $\lambda(A, B, \pi)$. Además, para asegurar que la inicialización es buena, se toma como criterio, que todos los estados se generen al menos el doble de símbolos que el número de símbolos de observación por estado, si no se cumple esta condición se repetirá la inicialización.

A continuación se utilizará el procedimiento de Baum-Welch para que ajuste los parámetros del modelo $\lambda(A, B, \pi)$, con la intención de maximizar la probabilidad de la secuencia de observación dada. Y se repite el proceso hasta que llegue a la finalización por umbral, o alcance el número máximo de iteraciones. Una vez se salga de este bucle, el modelo resultante se almacenará, y estará listo para ser testado, y comprobar de esta forma su éxito en el reconocimiento.

4.3.2 Etapa de test para el clasificador HMM

En esta etapa se calcula la probabilidad de la secuencia de observación para cada modelo creado en la etapa de entrenamiento, escogiendo aquel que proporciona la mayor probabilidad. De esta forma se puede obtener la matriz de confusión, donde se presentan las tablas de los aciertos y los errores cometidos durante la clasificación, para finalmente calcular la tasa media de reconocimiento.

Pero para obtener estos resultados, esta etapa de test necesita como entrada las secuencias de los vectores de los patrones que se quieren testar, los parámetros del modelo oculto creado (A, B, π), y las salidas del modelo para verificar automáticamente el resultado del reconocimiento.

En la figura 4.22, se muestra un ejemplo de la matriz de confusión, para el caso particular del reconocimiento de letra minúscula manuscrita.

4.4 Optimización de los parámetros del clasificador

En este apartado se determinarán cuales son los valores prácticos de los métodos de extracción de parámetros y de los clasificadores para las dos diferentes bases de datos. Partiendo que estas bases son diferentes, y que de una naturaleza común como es la escritura, han surgido diferentes agrupamientos de patrones, es preciso determinar y definir el valor de cada una de las variables de los métodos de extracción de parámetros y de los clasificadores, que debe de amoldarse a estas distintas condiciones de las diferentes muestras. Por tanto las variables que se analizarán son:

- El número de estados de la cadena
- El método de extracción de los parámetros para HMM (esqueletización, contorno por codificación de las direcciones de la secuencia o contorno por coordenadas polares)

- El tamaño del patrón en la base de datos (original o reducido)
- El número de escritores empleados durante el entrenamiento del clasificador
- El número de muestras por escritor empleadas durante el entrenamiento

Se ha especificado el uso de los clasificadores, los modelos ocultos de Markov (HMM) se usarán para la información proveniente de los trazos de las imágenes manuscritas, y las redes neuronales (NN) para las variaciones en la dirección de trazo manuscrito.

En el estudio de la optimización de los HMM para su clasificación, se va a probar ambas técnicas de extracción de características planteadas, la detección de contorno y la esqueletización. Y dentro de la detección de contorno, sus diferentes composiciones, que pueden ser el contorno interior y exterior del patrón manuscrito, o el contorno exterior del mismo. Además la vectorización del contorno, se puede diferenciar en la vectorización secuencial de la codificación de las direcciones, o por sus coordenadas polares.

Numero de palabras: 27
 Numero de repeticiones: 480
 Numero de modelos de Markov: 27
 Numero de estados: 60
 Numero de simbolos por estado: 8
 Parametros de salida: logPO
 FIN test. Tiempo: 28572.1

MD.	a	b	c	d	e	f	g	h	i	j	k	l	m	n	ñ	o	p	q	r	s	t	u	v	w	x	y	z	Prob.	
a	303	16	0	15	0	3	0	27	0	0	13	1	6	20	1	5	16	42	1	0	5	0	1	0	3	2	0	63.13 %	
b	3	266	24	72	0	2	14	0	0	23	0	0	0	1	3	10	5	5	0	41	0	2	0	0	0	2	7	55.42 %	
c	0	0	450	0	2	1	1	0	0	0	3	0	0	0	0	16	0	4	3	0	0	0	0	0	0	0	0	93.75 %	
d	1	20	3	136	0	0	6	0	1	50	0	0	0	0	8	3	152	39	1	0	43	0	9	0	1	0	6	1	28.33 %
e	0	22	17	14	366	25	4	0	17	1	0	0	0	0	0	2	1	2	0	2	0	0	0	0	0	0	7	76.25 %	
f	1	16	1	11	2	383	7	0	11	0	5	0	0	0	0	0	2	14	9	1	7	10	0	0	0	0	0	79.79 %	
g	1	8	21	23	2	1	276	3	0	37	4	0	0	14	0	18	3	42	10	7	0	6	1	0	0	1	2	57.50 %	
h	35	0	0	2	0	2	0	335	15	0	10	0	46	2	3	0	2	2	1	0	12	0	0	1	0	12	0	69.79 %	
i	5	11	0	1	4	12	9	1	253	11	5	3	34	0	0	0	0	120	0	2	3	0	0	0	0	6	0	52.71 %	
j	0	9	0	10	0	0	59	2	3	251	1	0	0	3	12	1	2	0	0	110	1	13	0	0	0	3	0	52.29 %	
k	7	3	9	1	3	1	0	6	0	0	391	2	6	1	0	0	2	21	12	0	0	0	0	0	11	0	4	81.46 %	
l	0	1	24	0	19	8	0	0	5	0	10	393	1	0	0	4	0	7	0	0	0	1	1	1	0	0	5	81.88 %	
m	9	0	0	0	0	0	60	2	0	6	0	376	7	2	0	0	2	0	0	0	0	0	0	1	1	14	0	78.33 %	
n	29	0	0	26	0	0	0	1	0	0	8	0	12	208	63	0	1	7	1	1	3	2	0	115	1	2	0	43.33 %	
ñ	12	1	0	49	0	0	5	0	1	1	2	0	2	2	373	3	2	3	2	0	3	0	0	15	4	0	0	77.71 %	
o	1	2	68	5	1	0	30	0	1	33	0	0	0	5	0	315	3	6	0	1	0	8	0	0	0	0	1	65.63 %	
p	2	13	1	45	0	1	1	0	6	0	0	0	0	0	1	2	389	4	6	2	1	0	0	0	1	5	0	81.04 %	
q	15	6	9	8	0	1	7	5	0	0	9	1	0	6	0	23	2	376	9	1	0	0	0	0	0	0	2	78.33 %	
r	10	5	13	11	0	3	0	3	0	0	3	20	7	3	1	3	1	85	290	0	0	0	0	0	0	8	1	13	60.42 %
s	0	1	0	8	0	0	32	0	0	1	0	0	0	3	0	0	0	0	2	433	0	0	0	0	0	0	0	90.21 %	
t	4	3	0	3	0	1	13	13	9	0	0	1	3	0	0	0	4	0	0	1	424	0	0	0	0	0	1	88.33 %	
u	1	1	0	1	1	0	1	2	1	0	12	2	25	1	0	1	0	28	0	0	0	389	7	7	0	0	0	81.04 %	
v	14	1	0	0	0	0	4	0	0	1	7	0	1	1	3	0	10	0	1	0	0	116	280	19	0	22	0	58.33 %	
w	1	1	0	12	0	0	3	2	0	0	0	0	7	25	24	3	0	2	0	0	0	1	1	385	0	10	3	80.21 %	
x	3	2	0	0	0	1	0	1	0	0	52	0	3	0	0	0	4	5	0	0	0	0	0	1	407	0	1	84.79 %	
y	9	11	0	1	0	0	2	9	1	1	5	0	4	3	7	1	6	6	1	15	0	0	0	6	6	379	7	78.96 %	
z	1	10	0	11	1	0	0	0	9	8	0	1	0	0	0	0	0	5	1	2	0	0	0	0	1	0	430	89.58 %	

RECONOCIMIENTO MEDIO: 71.4275 %

Figura 4.22: Matriz de confusión para el reconocimiento de letra minúscula manuscrita.

Centrados en la estructura del HMM, se pretende optimizar la tasa de reconocimiento de los clasificadores, para ellos se va a determinar los valores que

definen a cada clasificador. En el caso del modelo oculto de Markov, los parámetros a optimizar son:

- Número de estados del modelo
- Número de muestras por escritor usadas para el entrenamiento de la cadena
- Número de escritores empleados en el entrenamiento

Con respecto a las redes neuronales, se va utilizar un único sistema de extracción de características, que es la técnica del gradiente, con las máscaras de Kirsch. Los parámetros de la NN que se van a optimizar son:

- Número de muestras por escritor usadas durante el entrenamiento
- Número de neuronas de la capa oculta
- Coeficiente de aprendizaje inicial

4.4.1 Optimización del clasificador para dígitos manuscritos

Partiendo del punto que todas las bases de datos son diferentes, habrá que definir para cada una de ellas los parámetros del clasificador para mejorar el éxito del reconocimiento. En particular, para la base de datos de dígitos manuscritos, se van a seguir todas las pautas indicadas anteriormente, con una doble finalidad. Primero, caracterizar el clasificador para la extracción de la información geométrica y secuencial, y segundo, empezar a definir cuales son las posibles técnicas que ofrecen un mejor comportamiento del clasificador [Travieso99].

En el anexo A, se puede seguir cada uno de las pruebas trazadas. Para la información de carácter secuencial, hay dos parámetros básicos a distinguir en primer lugar, el uso de una base de datos con el tamaño original de su trazado o el uso de una base de datos con un tamaño reducido. De las pruebas realizadas se obtuvieron los siguientes resultados:

Tamaño del patrón	Técnica de extracción de parámetros	Número de estados	Escritores para entrenar	Número de muestras por escritor	Tasa de reconocimiento
Original	Contorno exterior e interior	70	30	3	97,22 %
Original	Contorno exterior	50	10	9	91,77 %
Reducido	Contorno exterior	50	30	3	96,22 %

Tabla 4.1: Tasas de reconocimiento según el tamaño de los dígitos.

El mejor resultado se ha obtenido para un tamaño original del dígito, y usando una técnica de contorno exterior e interior. Pero puede resultar más gráfico, observar las tablas y gráficas descritas en el anexo A, concernientes a estos experimentos, para concluir que los valores en media corresponden a una mayor tasa de reconocimiento si se trata de dígitos manuscritos de tamaño reducido y caracterizados por el contorno exterior. Aunque en la tabla 4.1, se refleje con un punto menos que la citada anteriormente. Por tanto se puede considerar como un valor de pico, y no un resultado de la tendencia de los valores. Por eso, se considerará mejor método, si se utiliza un dígito reducido y vectorizado mediante el contorno externo.

La explicación del uso del tamaño reducido en los dígitos, es debido a que la reducción del tamaño hace perder características particulares de cada escritor en su escritura, y generaliza los trazados establecidos para cada dígito, haciendo que cada tipo de patrón de dígito se parezca más entre sí. Esta reducción equivale a realizar un filtrado de la imagen paso bajo, donde se eliminan las variaciones rápidas del trazado, y éste es suavizado, produciendo este efecto de mayor parecido, y que las tasas de reconocimiento en media, sean más elevadas y con menor varianza.

Pero se han ofrecido dos técnicas para vectorizar los dígitos, usando la ya mencionada técnica del contorno, frente a la técnica de esqueletización. Como ya se ha mencionado, la mejor tasa de resultados es para el uso de dígitos reducidos, y por tanto, sólo se empleará este tipo de tamaño para los patrones. Las pruebas realizadas con ambas técnicas han dado como resultado:

Tamaño del patrón	Técnica de extracción de parámetros	Número de estados	Número de escritores	Número de muestra por escritor	Tasa de reconocimiento
Reducido	Esqueletización	30	30	9	93.55 %
Reducido	Contorno exterior secuencia	50	30	3	96.22 %
Reducido	Contorno exterior polares	50	40	12	74.28 %

Tabla 4.2: Tasas de reconocimiento según las técnicas de vectorización de la información secuencial para dígitos.

Con esta tabla 4.2, la conclusión es clara, la vectorización de los dígitos por coordenadas polares es el peor sistema de discriminación, pero mirando en general los resultados, la técnica de vectorización del contorno exterior por su secuencia es mejor que la de esqueletización. Esta diferencia en la tasa de reconocimiento indica que los modelos ocultos de Markov, modelan mejor el contorno por secuencia que la esqueletización.

En las aplicaciones de dígitos manuscritos, en referencia a la información secuencial, se puede concluir que el uso del tamaño reducido de los patrones con el uso de la vectorización por contorno exterior de la secuencia, ofrecen la mejor combinación para un HMM. Los valores internos del propio clasificador, se han discutido en el anexo A, y quedan recogidos en las tablas anteriores 4.1 y 4.2.

Con respecto a la información geométrica, ésta ha sido vectorizada por la técnica del gradiente [Cho97], y clasificada mediante redes neuronales. La comparación entre los resultados obtenidos con los clasificadores de HMM y NN, se puede contemplar en la siguiente tabla.

Tipo de información	Tasa de reconocimiento
Geométrica	91.67 %
Secuencial	96.22 %

Tabla 4.3: Tasas de reconocimiento según el tipo de información para dígitos.

De esta tabla se obtiene dos conclusiones. La primera es que la información secuencial debe de ser tenida en cuenta para el reconocimiento de dígitos manuscritos, ya que los valores en la tasa de éxito son comparables a los de la información geométrica. Y la segunda conclusión, es que esta información además de ser tenida en cuenta, en esta aplicación de dígitos manuscritos, responde mejor que la información de carácter geométrica.

Por ultimo, quedaría hacer constancia, que estas conclusiones a las que se han llegado no responden a valores aislados y a picos de éxito en la tasa de reconocimiento. Por tanto, se va a realizar una prueba de promediado de los resultados. A partir de toda la base de datos, dividida al azar en un 40% de los escritores para entrenar y un 60% de los mismo para el test, se han realizado diez veces variando de esta forma aleatoria el contenido de los vectores de entrenamiento y de test, para finalmente obtener un resultado promedio del éxito del reconocimiento.

Para el clasificador de HMM se ha fijado como extractor de parámetros la técnica de contorno exteriores, aplicados a la base de datos de dígitos reducidos, y se ha escogido los parámetros de optimización del clasificador HMM dados en la tabla 4.1. Para el clasificador con NN los datos se han obtenido del anexo A. El resultado de estos valores promedios se recogen en la siguiente tabla.

Nº de Test de dígitos	Clasificador HMM	Clasificador NN
Test 1	94.48%	87.33%
Test 2	93.89%	89.51%
Test 3	94.21%	86.94%
Test 4	93.42%	85.96%
Test 5	94.48%	86.59%
Test 6	94.98%	89.51%
Test 7	93.86%	86.94%
Test 8	94.35%	85.96%
Test 9	94.24%	88.10%
Test 10	94.77%	86.71%
Media ± Varianza	94.27% ± 0.21	87.35% ± 1.23

Tabla 4.4: Establecimiento de los reconocimientos medios para clasificador HMM y NN con dígitos.

Como se puede comprobar, para un volumen grande de dígitos, el comportamiento en media de los clasificadores vuelve a destacar las conclusiones anteriores, el buen y mejor comportamiento de la información secuencial con el uso de una vectorización de contorno exterior, frente a la información geométrica vectorizada con la técnica del gradiente.

4.4.2 Optimización del clasificador para letras manuscritas

La optimización del clasificador se ha planteado para tres supuestos diferentes, dadas las combinaciones de las técnicas de extracción y los clasificadores que se han planteado para esta tesis. Las pruebas que se han realizado para esta optimización han sido la realizada para HMM con detección de contorno exterior para letras de tamaño

normalizado, para HMM con la técnica de esqueletización para letras de tamaño normalizado y la realizada para NN con la técnica del gradiente.

En esta aplicación de letras manuscritas ya se han utilizado directamente la base de datos con imágenes reducidas y normalizadas, ya explicadas en el capítulo tercero; porque se ha visto que existe una mayor correlación entre las imágenes con este tamaño, y por tanto, la tasa de reconocimiento a obtener tendrá un valor superior, a la base de datos de tamaño original [Morales99-pfc], tal y como se ha estudiado en el apartado anterior y en el Anexo A. En este mismo estudio se ha concluido también que la vectorización con el contorno exterior devuelve mejores resultados que el contorno exterior e interior, y por ese motivo, se ha excluido de este estudio. Aunque se ha incluido la diferencia de vectorización por coordenadas polares y por la secuencia codificada por la matriz de direcciones.

Con relación a las pruebas efectuadas para las letras manuscritas, se ha planteado el estudio detallado en el anexo B de esta tesis. Estas pruebas han sido realizadas independientes de escritor, es decir, las muestras de los escritores usados para el entrenamiento son diferentes, en muestras y en escritores; a las usadas para la etapa de evaluación o test. Las únicas pruebas efectuadas con forma dependiente de escritor, usando diferentes muestras de un mismo escritor, tanto para entrenar como para evaluar, fueron para toda la base de datos de mayúsculas y minúsculas. Con la finalidad de establecer un criterio de fiabilidad de la base de datos, y observar si ésta y sus modelos creados, son una generalidad o no, de las muestras y sus escritores.

Un detalle importante de estas pruebas se haya unido intrínsecamente a la propia base de datos. Ya que en ésta, se pueden distinguir tanto letras mayúsculas, como letras minúsculas, y por tanto, también se puede establecer el criterio, de realizar las pruebas con las letras de forma aisladas según sean mayúsculas o minúsculas; o establecer las pruebas de forma mixta, tanto para letras mayúsculas como de letras minúsculas. El criterio seguido, ha sido establecer las pruebas para la base de datos de las letras minúsculas y según la optimización dada, hacerlo extensible a las letras mayúsculas y al modelo mixto.

La primera comparación que se puede establecer, es entre pruebas para un mismo clasificador, de forma que se puede captar qué técnica de vectorización ofrece mejor tasa de reconocimiento. En este caso se habla de vectorizar con la técnica de esqueletización o la técnica de contorno exterior, los resultados de la tasa de reconocimiento y de valores óptimos del clasificador, se muestran en la siguiente tabla.

<i>Técnica de extracción de parámetros</i>	<i>Tipo de letra</i>	<i>Número de estados</i>	<i>Número de escritores</i>	<i>Tasa de reconocimiento</i>
Contorno exterior secuencial	Minúscula	60	40	75.04%
Contorno exterior polares	Minúscula	40	50	51.51%
Esqueletización	Minúscula	60	60	66.77%

Tabla 4.5: Tasas de reconocimiento para optimizar el clasificador HMM con letras manuscritas.

La diferencia referenciada en esta tabla 4.5 es clara, el contorno de la secuencia codificada por direcciones ofrece la mejor respuesta en un clasificador HMM. Los sistemas de esqueletización y contorno por coordenadas polares distan en bastantes puntos del caso citado con anterioridad. Quedaría observar de la otra prueba establecida, que clasificador ofrece mejor tasa de resultados, enfrentando al HMM con la información secuencial a la NN, con la información geométrica.

<i>Tipo de información</i>	<i>Tipo de letra</i>	<i>Tasa de reconocimiento</i>
Secuencial	Minúscula	75.04%
Secuencial	Mayúscula	86.06%
Secuencial	Mixto	69.59%
Geométrica	Minúscula	68.22%
Geométrica	Mayúscula	77.62%
Geométrica	Mixto	68.88%

Tabla 4.6: Tasas de reconocimiento según el tipo de información para letra manuscrita.

Se observa de esta tabla 4.6, que en todos los casos, la información secuencial clasificada con HMM, ofrece una mejor tasa de reconocimiento que la información geométrica. Viendo los tipos de letras de forma aislada, mayúscula y minúscula, la tasa de éxito se diferencian en más de ocho puntos, aunque la mixta lo hace de forma más discreta. Los valores más óptimos encontrados para el HMM son 60 estados y 40 escritores con sus 12 repeticiones para el entrenamiento. Y para NN, se hallaron que los valores óptimos se componían de 120 neuronas en la capa oculta y 50 escritores con sus 12 repeticiones para el conjunto de datos de entrenamiento. En ambos casos, las pruebas son independiente de escritor.

Por tanto, la conclusión que se puede deducir, es que para la aplicación de las letras manuscritas, el uso de la información secuencial se puede emplear para su reconocimiento, incluso del mismo orden que los encontrados en la literatura actual [Rizki90] [Seiler96] [Kumar99]. Otra conclusión extraíble, es el mejor reconocimiento de las técnicas con información secuencial frente a las técnicas de información geométrica, según se señala, la tabla 4.6.

Por último, queda realizar una prueba para promediar estos resultados, para notar que estos resultados no responden a picos de éxitos. Para ello, se escoge la mejor optimización para cada uno de los dos clasificadores, y se van a realizar cinco veces las pruebas, tanto para letras mayúsculas como para letras minúsculas. Para estas cinco pruebas, se dotará en las muestras de entrenamiento el 40% de la base de datos, quedando el 60% restante para el test, cogiendo todas las repeticiones de cada escritor para la tarea que se le asigne, entrenamiento o evaluación, haciendo un test independiente de escritor. En cada una de las pruebas, los escritores han sido escogidos aleatoriamente.

<i>Nº de Test de letras</i>	<i>Clasificador HMM</i>	<i>Clasificador NN</i>
Test 1	73.63 %	61.06 %
Test 2	73.61 %	59.38 %
Test 3	71.22 %	59.02 %
Test 4	70.85 %	60.14 %
Test 5	74.62 %	60.61 %
Media ± Varianza	72.79 % ± 1.65	59.84 % ± 0.84

Tabla 4.7: Establecimiento de los reconocimientos medios para clasificador HMM y NN con letras minúsculas.

<i>Nº de Test de letras</i>	<i>Clasificador HMM</i>	<i>Clasificador NN</i>
Test 1	88.03 %	67.40 %
Test 2	86.54 %	69.86 %
Test 3	83.94 %	68.83 %
Test 4	83.81 %	68.63 %
Test 5	85.27 %	69.14 %
Media ± Varianza	85.52 % ± 1.79	68.77 % ± 0.90

Tabla 4.8: Establecimiento de los reconocimientos medios para clasificador HMM y NN con letras mayúsculas.

Con el volumen de toda la base de datos promediada se vuelve a destacar las dos conclusiones anteriormente citadas, en definitiva, la buena respuesta de la información secuencial vectorizada mediante el contorno.

4.5 Base de datos NIST-19

Con la finalidad de validar el trabajo realizado hasta ahora, se ha utilizado otra base de datos muy usada de forma internacional [LeCun99]. Recibe el nombre de NIST número 19, y ha sido desarrollada por “*National Institute of Standards and Technology*”, en particular por la División de Sistemas Avanzado, en la figura del Grupo de Procesado de Imagen Avanzada [Grother95].

Esta base de datos viene en formato CD ROM, y contiene imágenes binarias extraídas de 3699 plantillas, que divididas en muestras tiene un total de 814255, entre dígitos y caracteres alfabéticos. NIST número 19 ha seguido el formato de ISO-9660 para su archivo y distribución.

Las intenciones que se tienen con respecto a esta base de datos son dos. En primer lugar validar la forma de construir bases de datos que se ha fijado en esta tesis, para así como considerarlas como aptas. Y en segundo lugar, contrastar que los datos de reconocimiento logrados, son del mismo orden de magnitud que los que se obtendrán al ejecutar los mismos algoritmos en NIST-19.

4.5.1 Características de la base de datos NIST-19

Cada uno de los caracteres segmentados es una imagen binaria de 128×128 píxeles. Estos caracteres forman un total de 62 patrones o clases, divididos en tres subclases, correspondientes a;

- “0” al “9”
- “A” a la “Z”
- “a” a la “z”

Todas estas clases han sido chequeadas manualmente por el “*National Institute of Standards and Technology*”, con un error de clasificación por caracteres residuales de aproximadamente el 0.1%. La compilación para descomprimir la base de datos del NIST-19 se ha realizado en un entorno Solaris, bajo una estación de trabajo Ultra-10 de Sun.

Si se engloban las muestras según los dígitos, se recogen un total de 402953 muestras, para letras minúsculas hay 81096, para mayúsculas 81463, y finalmente letras extraídas de un texto de la constitución norteamericana, se tienen 248743 muestras entre minúsculas y mayúsculas. En la tabla 4.9 se detallan los patrones y las muestras que contiene para cada uno de ellos.

4.5.2 Clasificación con la base de datos NIST-19

El empleo de la base del NIST-19, como ya se ha comentado, se ha hecho para lograr dos objetivos, verificar que las bases de datos creadas son robustas y por tanto, el método para realizar ha sido satisfactorio. Y en segundo lugar, comprobar que el método de parametrización y vectorización planteado devuelve buenos resultados independientemente de la base de datos utilizada.

Las pruebas a implementar se realizarán con los patrones manuscritos simples vistos hasta ahora, los dígitos y las letras. Para ello, se han escogido los mejores resultados para distinguir las técnicas de parametrización que se aplicarán al NIST-19. Tanto para los dígitos como para las letras, se ha escogido la detección de contorno externo siendo vectorizado con la codificación dada por la matriz de direcciones que sigue la secuencia, que es la que mejores tasas de clasificación ofrecen. Esta técnica es aplicada a las muestras ya reducidas y normalizadas.

La única variable que se ha tenido en cuenta para realizar los modelos es el número de muestras a utilizar para crear el modelo y para realizar el test. El término común entre las dos aplicaciones fue la división de la base de datos para realizar estos dos procesos. Para realizar las pruebas independientes del escritor se determinó que las muestras dadas por el conjunto de escritores agrupados de *hsf_0* al *hsf_3*, se dedicasen al entrenamiento, y los escritores de *hsf_4* a *hsf_7* al test, donde *hsf_i* es la forma del NIST de agrupar a los escritores, yendo “i” desde el 0 hasta el 7.

Para los dígitos manuscritos, se escogió aleatoriamente las muestras para entrenar de los grupos *hsf_0* al *hsf_3*, que se fijó en un 10% del total, y para el test se utilizaron 10000 muestras entre los grupos *hsf_4* al *hsf_7*. Para validar y promediar todo este proceso se repitieron las pruebas en cuatro ocasiones.

Dígitos		Letras Minúsculas		Letras Mayúsculas		Constitución			
Clases	Muestras	Clases	Muestras	Clases	Muestras	Clases	Muestras	Clases	Muestras
0	40363	a	3210	A	3033	a	8467	A	4436
1	44704	b	3136	B	3026	b	2876	B	1500
2	40072	c	3286	C	3344			C	8489
3	41112	d	3111	D	3020	d	8749	D	2321
4	39154	e	3084	E	2892	e	25639	E	2893
5	36606	f	2961	F	3053			F	7569
6	39937	g	2966	G	2964	g	1310		
7	41893	h	3253	H	2925	h	6964	H	748
8	39579	i	3152	I	4490			I	9504
9	39533	j	2213	J	2958			J	1430
		k	2957	K	2850				
		l	4454	L	3375	l	13399	L	2511
		m	3109	M	3077			M	7410
		n	3150	N	2128	n	10166	N	6460
		o	3215	O	3176			O	25963
		p	2816	P	3127			P	6617
		q	2648	Q	3018	q	851		
		r	3113	R	3061	r	13312	R	2821
		s	3136	S	3129			S	21143
		t	3058	T	2981	t	18169	T	8415
		u	3312	U	3143			U	11461
		v	3378	V	3237			V	2196
		w	3164	W	3122			W	2379
		x	3292	X	3203				
		y	2746	Y	2966			Y	2575
		z	3176	Z	3165				

Tabla 4.9: Tipos de patrones y cantidad de muestras de la base de datos NIST-19.

Los parámetros del clasificador que se han utilizado, además de las condiciones de entrenamiento y test descritas, son las mismas que se fijaron para realizar la clasificación con HMM de la tabla 4.4, en definitiva, 50 estados para el clasificador. Bajo estas condiciones se obtuvieron los siguientes resultados:

Nº de Test de dígitos (NIST-19)	Clasificador HMM
Test 1	91.42 %
Test 2	92.71 %
Test 3	91.20 %
Test 4	92.95 %
Media ± Varianza	92.07 % ± 0.78

Tabla 4.10: Clasificación de dígitos de la base de datos NIST-19.

Si se compara estos resultados con los obtenidos para reconocer dígitos, con la base de datos creada en esta tesis, se observan que la variación es de cuatro puntos de

empeoramiento para el uso de la base de datos del NIST-19. Además estos valores obtenidos, se pueden comparar con los que nos ofrece la literatura, recogidos en la revisión del estado de la técnica de los dígitos del capítulo segundo. En particular en [Garris95], que utiliza la misma base de datos, obtiene resultados del mismo orden de magnitud (92.9%), aunque en [Cho97] se ve superado (96.05%).

<i>Reconocedor de dígitos manuscritos</i>	<i>Tasas de reconocimiento</i>
[Cho97]	96 %
[Neubauer96]	98 %
[Jeong99]	99 %
[Shastri95]	96 %
[Garris95]	92 %
[Grother92]	96 %
[Gosselim97]	98 %
[Schäfer95]	98 %
[Kawajiri91]	99 %
[Schölkopf95]	96 %
[Hwang97]	98 %
[Hastie94]	97 %
[Proter98]	93 %
Esta tesis en media	94 %
Esta tesis en pico	97 %
Esta tesis con NIST	92 %

Tabla 4.11: Comparación de los sistemas reconocedores actuales para dígitos manuscritos.

Para las letras manuscritas se ha desarrollado el entrenamiento con el 20% de las muestras de los grupos *hsf_0* al *hsf_3* escogidas aleatoriamente de entre estos grupos, y el test con 10000 muestras de los grupos *hsf_4* al *hsf_7* extraídas también al azar, de esta manera se consigue un reconocimiento independiente de escritor. Con respecto al clasificador HMM se utilizaron 60 estados, ya que es donde mejores resultados se ha obtenido, recabada esta información de las tablas 4.7 y 4.8. En la siguiente tabla se observa los resultados obtenidos para esta base de datos del NIST-19, implementados con los mismos algoritmos usados para parametrizar las letras manuscritas. Estas pruebas se han repetido en tres ocasiones para verificar su funcionamiento.

<i>Nº de Test de letras (NIST-19)</i>	<i>Letras minúsculas</i>	<i>Letras mayúsculas</i>
Test 1	73.46 %	78.08 %
Test 2	77.49 %	80.78 %
Test 3	74.98 %	79.56 %
Media ± Varianza	75.31 % ± 2.03	79.47 % ± 1.35

Tabla 4.12: Clasificación de letras de la base de datos NIST-19.

Si se extiende el grado de comparación hasta la literatura, se puede resumir en la siguiente tabla las tasas del estado actual del reconocimiento de letras con los resultados logrados en esta tesis.

<i>Reconocedor de dígitos manuscritos</i>	<i>Tasas de reconocimiento</i>	
	<i>Minúscula</i>	<i>Mayúscula</i>
[Rizki90]	---	75 %
[Seiler96]	67.5 %	93.6 %
[Garris95]	75.3%	84.5 %
[Kumar99]	---	86.6 %
Esta tesis en media	72.79 %	85.52 %
Esta tesis en pico	75.04 %	86.06 %
Esta tesis con NIST	75.31 %	79.47 %

Tabla 4.13: Comparación de los sistemas reconocedores actuales para letras manuscritas.

4.6 Conclusiones

La principal conclusión e idea que se ha ido repitiendo en estos últimos apartados es el mejor funcionamiento de la extracción de parámetros a partir del contorno de las muestras manuscritas simples, tales como dígitos y letras. Y según sea para una aplicación u otra, varían los parámetros del clasificador.

Si se contrastan estos resultados con los de la literatura actual sobre este tipo de reconocimiento, se observa que los resultados para los dígitos son ligeramente inferiores. Para las letras, tiene dos tendencias diferentes, mientras que para las letras minúsculas los resultados son bastantes parecidos, para las mayúsculas funciona ligeramente peor.

La gran diferencia reside que la técnicas utilizadas en el estado del arte, para esta discriminación de patrones, se basa en parámetros geométricos usando clasificadores adaptados a este tipo de datos, como son las redes neuronales, en la mayoría de los casos. En cambio, la introducción de parámetros secuenciales para clasificar estos patrones es diferente a la utilizada, y los resultados son aproximadamente del mismo orden que los actuales.

Capítulo 5

Clasificación de cadenas manuscritas simples: palabras y códigos postales

5.1 Introducción

Este quinto capítulo abarca la clasificación de cadenas manuscritas integradas a partir de letras y números, en particular, de palabras y códigos postales numéricos. Por tanto, se tendrán que plantear nuevas bases de datos que contengan este tipo de información para clasificar estos patrones. Los pasos a seguir, ya se han marcado en capítulos anteriores, se realizará el procesado de imagen para posteriormente extraer la información de los parámetros contenidas en las muestras. Éstos se sitúan a la entrada del clasificador para su posterior reconocimiento.

En este capítulo se va a explicar cuales han sido las técnicas utilizadas para realizar todos estos procesos, y llegar a reconocer palabras manuscritas. En principio, se pretende emplear la misma técnica que para las letras y dígitos manuscritos, de tal forma que este sistema se pueda generalizar a cualquier grupo de patrones manuscritos.

La particularidad que se puede encontrar con las cadenas de manuscritas es que se conforman a partir de una serie de símbolos alfanuméricos, letras y números. Esta forma natural de creación de las cadenas nos puede sugerir otra forma de reconocimiento a parte del contorno de la toda la cadena, y es la clasificación de cada uno de los elementos que la conforman.

5.2 Bases de datos de palabras

Como se puede obviar, las palabras se pueden formar a partir de variaciones de un número determinado de letras y con repetición, formando tantas palabras como permita esta combinación del alfabeto castellano. En este caso, sólo se va a trabajar con palabras aisladas, y no extraídas de un texto, ya que el objetivo de esta tesis es realizar un primer planteamiento sobre técnicas de reconocimiento de formas manuscritas, sin entrar en campos como la segmentación de las palabras de un texto, ya que se sale del objetivo marcado.

Debido a la gran cantidad de palabras que se pueden formar, se va a crear una base de datos que incluya un número determinado de patrones. Quizás se esté sesgando el total del conjunto de patrones, pero va a permitir trazar el proceso de reconocimiento de forma más cómoda.

Para diseñar y crear esta base de datos, se va a partir de las condiciones impuestas en el capítulo tercero, ya que se ha demostrado que es eficiente el sistema de construcción utilizado, con el contraste realizado frente a la base de datos NIST-19. A continuación se resume esquemáticamente los principales puntos que se han tenido en cuenta:

- Lograr una base de datos que sea representativa de cualquier tipo de población, mediante un cierto número de escritores y patrones
- Se utilizará un sistema de captación *off-line*, utilizando un escáner óptico para tal efecto
- El tamaño de la plantilla se fija en el DIN-A4, determinado por el escáner usado
- Las instrucciones para rellenar la plantilla son las mismas, que están explícitas en el apartado 3.4.1

5.2.1 Base de datos de palabras manuscritas

El objetivo es obtener patrones que estén formados por palabras manuscritas y por tanto, se va a diseñar y construir la base de datos con éstas. Teniendo en cuenta la posibilidad de combinar las letras de nuestro alfabeto, agrupadas según su longitud y con repetición, la cantidad de palabras que se pueden generar es enorme, aunque en nuestro idioma se puede limitar las palabras con significado en más de treinta mil. Ésto implica tener un universo de patrones bastante grande. Para la siguiente base de datos, ha limitado este universo de patrones, y se ha fijado en 61 palabras, que son las que se pueden obtener si queremos generar una palabras escrita en castellano, que se contenga en un cheque bancario.

Para la construcción de esta base de datos se han utilizado un total de 160 escritores, que han realizado 3 repeticiones por cada una de las 61 palabras que se han tenido previsto. No todos los escritores han escrito las mismas palabras, sino que se ha dividido el total de los escritores en cuatro grupos, segmentando la escritura de toda la base de datos. Ésto se ha realizado por mantener la naturalidad en la escritura, ya que tantas palabras escritas por una sola persona puede desfigurar el trazo. Por tanto, es como si 40 escritores (una cuarta parte de los 160), hubiesen hecho toda la base de

datos. Las muestras totales obtenidas en esta base de datos se han realizado con 61 patrones diferentes, y 3 repeticiones por cada patrón, y con 40 escritores, resultando un total de 7320 (40x61x3) muestras.

5.2.1.1 Perfil de los escritores

Los escritores miembros de esta base de datos quedan representados según los parámetros preestablecidos en la siguiente tabla;

ESCRITOR	SEXO	MANO ESCRITURA	EDAD
1	Mujer	Diestra	17
2	Mujer	Diestra	18
3	Mujer	Diestra	17
4	Mujer	Diestra	18
5	Mujer	Diestra	17
6	Hombre	Diestro	17
7	Hombre	Diestro	16
8	Hombre	Diestro	18
9	Hombre	Diestro	17
10	Hombre	Diestro	18
11	Mujer	Diestra	26
12	Mujer	Diestra	23
13	Mujer	Diestra	25
14	Mujer	Diestra	29
15	Mujer	Diestra	26
16	Mujer	Diestra	25
17	Mujer	Diestra	28
18	Mujer	Diestra	27
19	Mujer	Diestra	29
20	Mujer	Diestra	29
21	Hombre	Diestro	25
22	Hombre	Diestro	26
23	Hombre	Diestro	27
24	Hombre	Diestro	26
25	Hombre	Diestro	23
26	Hombre	Diestro	29
27	Hombre	Zurdo	24
28	Hombre	Diestro	27
29	Hombre	Diestro	23
30	Hombre	Diestro	27
31	Mujer	Diestra	50
32	Mujer	Diestra	42
33	Mujer	Diestra	41
34	Mujer	Diestra	45
35	Mujer	Diestra	44
36	Hombre	Diestro	65
37	Hombre	Diestro	46
38	Hombre	Diestro	43
39	Hombre	Diestro	52
40	Hombre	Diestro	51
41	Mujer	Diestra	17
42	Mujer	Diestra	18
43	Mujer	Zurda	17
44	Mujer	Diestra	15
45	Mujer	Diestra	15

ESCRITOR	SEXO	MANO ESCRITURA	EDAD
47	Hombre	Diestro	15
48	Hombre	Diestro	15
49	Hombre	Diestro	16
50	Hombre	Zurdo	16
51	Mujer	Diestra	27
52	Mujer	Diestra	33
53	Mujer	Diestra	21
54	Mujer	Diestra	24
55	Mujer	Diestra	22
56	Mujer	Diestra	23
57	Mujer	Diestra	25
58	Mujer	Diestra	29
59	Mujer	Diestra	22
60	Mujer	Diestra	32
61	Hombre	Diestro	29
62	Hombre	Diestro	30
63	Hombre	Diestro	28
64	Hombre	Diestro	34
65	Hombre	Diestro	22
66	Hombre	Diestro	27
67	Hombre	Diestro	28
68	Hombre	Zurdo	25
69	Hombre	Diestro	22
70	Hombre	Zurdo	22
71	Mujer	Diestra	53
72	Mujer	Diestra	45
73	Mujer	Diestra	41
74	Mujer	Diestra	47
75	Mujer	Diestra	40
76	Hombre	Diestro	46
77	Hombre	Diestro	55
78	Hombre	Diestro	45
79	Hombre	Diestro	46
80	Hombre	Diestro	52
81	Mujer	Diestra	20
82	Mujer	Diestra	19
83	Mujer	Diestra	18
84	Mujer	Diestra	18
85	Mujer	Diestra	18
86	Hombre	Diestro	14
87	Hombre	Zurdo	16
88	Hombre	Diestro	16
89	Hombre	Zurdo	16
90	Hombre	Diestro	19
91	Mujer	Diestra	29

46	Hombre	Diestro	17	92	Mujer	Diestra	29
93	Mujer	Diestra	26	127	Hombre	Diestro	17
94	Mujer	Diestra	34	128	Hombre	Zurdo	19
95	Mujer	Diestra	24	129	Hombre	Diestro	19
96	Mujer	Diestra	31	130	Hombre	Diestro	18
97	Mujer	Diestra	22	131	Mujer	Diestra	29
98	Mujer	Diestra	22	132	Mujer	Diestra	26
99	Mujer	Diestra	23	133	Mujer	Diestra	26
100	Mujer	Diestra	25	134	Mujer	Diestra	34
101	Hombre	Diestro	30	135	Mujer	Diestra	25
102	Hombre	Diestro	29	136	Mujer	Diestra	29
103	Hombre	Zurdo	29	137	Mujer	Diestra	29
104	Hombre	Diestro	24	138	Mujer	Diestra	23
105	Hombre	Diestro	22	139	Mujer	Diestra	24
106	Hombre	Zurdo	25	140	Mujer	Diestra	22
107	Hombre	Diestro	22	141	Hombre	Zurdo	25
108	Hombre	Diestro	24	142	Hombre	Zurdo	23
109	Hombre	Diestro	26	143	Hombre	Zurdo	22
110	Hombre	Diestro	26	144	Hombre	Diestro	22
111	Mujer	Diestra	42	145	Hombre	Diestro	26
112	Mujer	Diestra	45	146	Hombre	Diestro	30
113	Mujer	Diestra	42	147	Hombre	Diestro	22
114	Mujer	Diestra	40	148	Hombre	Diestro	24
115	Mujer	Diestra	41	149	Hombre	Zurdo	29
116	Hombre	Diestro	46	150	Hombre	Diestro	28
117	Hombre	Diestro	45	151	Mujer	Diestra	44
118	Hombre	Diestro	47	152	Mujer	Diestra	46
119	Hombre	Diestro	45	153	Mujer	Diestra	49
120	Hombre	Diestro	40	154	Mujer	Diestra	47
121	Mujer	Zurda	17	155	Mujer	Diestra	41
122	Mujer	Diestra	17	156	Hombre	Diestro	47
123	Mujer	Diestra	17	157	Hombre	Diestro	56
124	Mujer	Diestra	17	158	Hombre	Diestro	46
125	Mujer	Diestra	18	159	Hombre	Diestro	47
126	Hombre	Diestro	18	160	Hombre	Diestro	44

Tabla 5.1: Muestrario de los escritores de la base de datos de palabras manuscritas.

De la anterior tabla se pueden extraer el sexo de los escritores, la mano de escritura, y que el rango de edades de los escritores está entre los 15 y los 56 años. Hay justamente la mitad de hombres que de mujeres, y en su mayoría casi todos los escritores son diestros. Cuantitativamente, estos datos se pueden reflejar en las siguientes figuras;

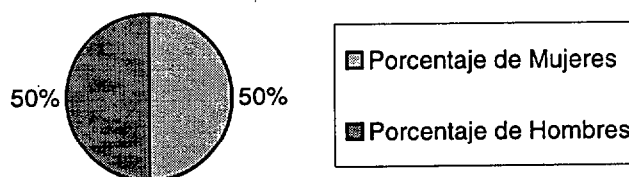


Figura 5.1: Porcentaje de Hombres vs. Mujeres de la base de datos de palabras manuscritas.

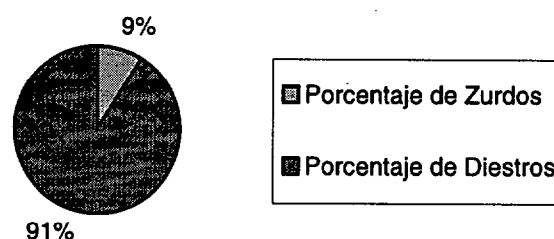


Figura 5.2: Porcentajes de Diestros vs. Zurdos de la base de datos de palabras manuscrita.

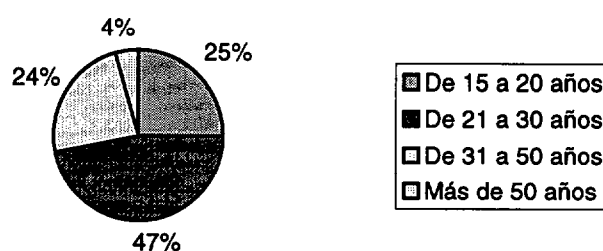


Figura 5.3: Distribución por edades en la base de datos de palabras manuscritas.

5.2.1.2 Formato de la plantilla

En esta base de datos se han realizado varias plantillas debido a que las palabras son más costosas temporalmente, en escribir, y el escritor termina por variar su proceso de escritura normal, acelerando este proceso, y deformando su escritura natural. Por esta razón se ha dividido la base de datos en cuatro grupos de plantillas, y los escritores que las han realizado han sido diferentes. Cada plantilla ha sido realizada por 40 escritores diferentes y por tanto, al haber cuatro plantillas, se han llegado a contabilizar un total de 160 escritores para conformar esta base de datos.

El formato que se le ha dado a las plantillas sigue los cánones anteriores, y se pueden distinguir tres partes fundamentales. Las líneas horizontales y verticales para la estimación de la inclinación, el cajetín, que ésta vez está dividido en dos, una parte para recoger la información del escritor, y otra parte para explicar las instrucciones de relleno de la plantilla. Y la tercera última parte, es el enrejado de las cuadrículas donde se sitúan las muestras. Este enrejado consta de 20 por 4 rectángulos, las 20 filas van dirigidas a veinte palabras, y las 4 columnas, van dirigidas a las repeticiones. La primera muestra la palabra que hay que seguir escribiendo, y los otros 4 huecos son para las repeticiones que tiene que hacer el escritor.

A continuación, se muestran las cuatro plantillas diseñadas para la recolección de las muestras de las palabras manuscritas (véase las figuras 5.4, 5.5, 5.6, y 5.7); y seguidamente, las plantillas totalmente rellenas con los datos capturados (figuras 5.8, 5.9, 5.10 y 5.11).

<p>N° de muestra:</p> <p>Mano:</p> <p>Sexo:</p> <p>Edad:</p>	<p>INSTRUCCIONES PARA RELLENAR LA PLANTILLA:</p> <p>1. Escribir con rotulador negro o azul.</p> <p>2. Escribir con naturalidad y con todas las letras en minúsculas.</p> <p>3. Escribir las palabras dentro de su recuadro, sin alcanzar el margen.</p> <p>4. Intenta centrar la palabra dentro del recuadro.</p> <p>5. Escribir en sentido vertical.</p> <p>6. No toque las palabras escritas.</p> <p>7. No doblar la plantilla ni realizar cortes.</p>		
uno			
dos			
tres			
cuatro			
cinco			
seis			
siete			
ocho			
nueve			
diez			
once			
doce			
trece			
catorce			
quince			
dieciséis			
diecisiete			
dieciocho			
diecinueve			
veinte			

Figura 5.4: Primera plantilla de la base de datos de palabras manuscritas.

N° de muestra: Mano: Sexo: Edad:	INSTRUCCIONES PARA RELLENAR LA PLANTILLA: 1. Escribir con rotulador negro o azul. 2. Escribir con naturalidad y con todas las letras en mayúsculas. 3. Escribir las palabras dentro de su recuadro, sin alcanzar el margen. 4. Intente centrar la palabra dentro del recuadro. 5. Escribir en sentido vertical. 6. No toque las palabras escritas. 7. No doblar la plantilla ni realizar borrones.		
veintiuno			
veintidós			
veintitrés			
veinticuatro			
veinticinco			
veintiséis			
veintisiete			
veintiocho			
veintinueve			
treinta			
cuarenta			
cincuenta			
sesenta			
setenta			
setenta			
ochenta			
noventa			
cién			
doscientas			
trescientas			

Figura 5.5: Segunda plantilla de la base de datos de palabras manuscritas.

Nº de muestra:	INSTRUCCIONES PARA RELLENAR LA PLANTILLA:		
Mano:	1.	Escribir con rotulador negro o azul.	
Sexo:	2.	Escribir con naturalidad y con todas las letras en minúsculas.	
Edad:	3.	Escriba las palabras dentro de su cuadrado, sin alcanzar el margen.	
	4.	Intente centrar la palabra dentro del recuadro.	
	5.	Escribir en sentido vertical.	
	6.	No toque las palabras escritas.	
	7.	No doblar la plantilla ni realizar borrones.	

doscientas			
trescientas			
cuatrocientas			
quinientas			
seiscientas			
setecientas			
ochocientas			
novecientas			
mil			
millón			
un			
ciento			
doscientos			
trescientos			
cuatrocientos			
quinientos			
seiscientos			
setecientos			
ochocientos			
novecientos			

Figura 5.6: Tercera plantilla de la base de datos de palabras manuscritas.

N° de muestra: Mano: Sexo: Edad:	INSTRUCCIONES PARA RELLENAR LA PLANTILLA: 1. Escribir con rotulador negro o azul. 2. Escribir con naturalidad y con todas las letras en minúsculas. 3. Escribir las palabras dentro de su recuadro, sin alcanzar el margen. 4. Intente centrar la palabra dentro del recuadro. 5. Escribir en sentido vertical. 6. No toque las palabras escritas. 7. No doblar la plantilla ni realizar cortes.		
millones			
una			
y			
pesetas			
\$			

Figura 5.7: Cuarta plantilla de la base de datos de palabras manuscritas.

N° de muestra: Mano: Derecha Sexo: Hembra Edad: 17.		INSTRUCCIONES PARA RELLENAR LA PLANTILLA: 1. Escribir con rotulador negro o azul. 2. Escribir con naturalidad y con todas las letras en minúsculas. 3. Escriba las palabras dentro de su recuadro: sin alcanzar el margen. 4. Intente centrar la palabra dentro del recuadro. 5. Escribir en sentido vertical. 6. No retoque las palabras escritas. 7. No doblar la plantilla ni realizar borrones.	
uno	uno	uno	uno
dos	dos	dos	dos
tres	tres	tres	tres
cuatro	cuatro	cuatro	cuatro
cinco	cinco	cinco	cinco
seis	seis	seis	seis
siete	siete	siete	siete
ocho	ocho	ocho	ocho
nueve	nueve	nueve	nueve
diez	diez	diez	diez
once	once	once	once
doce	doce	doce	doce
trece	trece	trece	trece
catorce	catorce	catorce	catorce
quince	quince	quince	quince
dieciséis	dieciséis	dieciséis	dieciséis
diecisiete	diecisiete	diecisiete	diecisiete
dieciocho	dieciocho	dieciocho	dieciocho
diecinueve	diecinueve	diecinueve	diecinueve
veinte	veinte	veinte	veinte

Figura 5.8: Primera plantilla de la base de datos de palabras manuscritas rellena.

N° de muestra: Mano: <i>Derecha</i> Sexo: <i>macho</i> Edad: <i>24</i>		INSTRUCCIONES PARA RELLENAR LA PLANTILLA: 1. Escribir con rotulador negro o azul 2. Escribir con naturalidad y con todas las letras en minúsculas 3. Escriba las palabras dentro de su recuadro, sin alcanzar el margen 4. Intente centrar la palabra dentro del recuadro 5. Escribir en sentido vertical 6. No retoque las palabras escritas 7. No doblar la plantilla ni realizar borrones	
veintiuno	<i>veintiuno</i>	<i>veintiuno</i>	<i>veintiuno</i>
veintidós	<i>veintidós</i>	<i>veintidós</i>	<i>veintidós</i>
veintitrés	<i>veintitrés</i>	<i>veintitrés</i>	<i>veintitrés</i>
veinticuatro	<i>veinticuatro</i>	<i>veinticuatro</i>	<i>veinticuatro</i>
veinticinco	<i>veinticinco</i>	<i>veinticinco</i>	<i>veinticinco</i>
veintiséis	<i>veintiséis</i>	<i>veintiséis</i>	<i>veintiséis</i>
veintisiete	<i>veintisiete</i>	<i>veintisiete</i>	<i>veintisiete</i>
veintiocho	<i>veintiocho</i>	<i>veintiocho</i>	<i>veintiocho</i>
veintinueve	<i>veintinueve</i>	<i>veintinueve</i>	<i>veintinueve</i>
treinta	<i>treinta</i>	<i>treinta</i>	<i>treinta</i>
cuarenta	<i>cuarenta</i>	<i>cuarenta</i>	<i>cuarenta</i>
cincuenta	<i>cincuenta</i>	<i>cincuenta</i>	<i>cincuenta</i>
sesenta	<i>sesenta</i>	<i>sesenta</i>	<i>sesenta</i>
setenta	<i>setenta</i>	<i>setenta</i>	<i>setenta</i>
setenta	<i>setenta</i>	<i>setenta</i>	<i>setenta</i>
ochenta	<i>ochenta</i>	<i>ochenta</i>	<i>ochenta</i>
noventa	<i>noventa</i>	<i>noventa</i>	<i>noventa</i>
cien	<i>cien</i>	<i>cien</i>	<i>cien</i>
doscientas	<i>doscientas</i>	<i>doscientas</i>	<i>doscientas</i>
trescientas	<i>trescientas</i>	<i>trescientas</i>	<i>trescientas</i>

Figura 5.9: Segunda plantilla de la base de datos de palabras manuscritas rellena.

N° de muestra: Mano: <i>DE LA IZA</i> Sexo: <i>MUJER</i> Edad: <i>25</i>		INSTRUCCIONES PARA RELENAR LA PLANTILLA: 1. Escribir con rotulador negro o azul. 2. Escribir con naturalidad y con todas las letras en minúsculas. 3. Escribir las palabras dentro de su recuadro sin alcanzar el margen. 4. Intente centrar la palabra dentro del recuadro. 5. Escribir en sentido vertical. 6. No retoque las palabras escritas. 7. No doblar la plantilla ni realizar borrones.	
doscientas	<i>doscientas</i>	<i>doscientas</i>	<i>doscientas</i>
trescientas	<i>trescientas</i>	<i>trescientas</i>	<i>trescientas</i>
cuatrocientas	<i>cuatrocientas</i>	<i>cuatrocientas</i>	<i>cuatrocientas</i>
quinientas	<i>quinientas</i>	<i>quinientas</i>	<i>quinientas</i>
seiscientas	<i>seiscientas</i>	<i>seiscientas</i>	<i>seiscientas</i>
setecientas	<i>setecientas</i>	<i>setecientas</i>	<i>setecientas</i>
ochocientas	<i>ochocientas</i>	<i>ochocientas</i>	<i>ochocientas</i>
novcientas	<i>novcientas</i>	<i>novcientas</i>	<i>novcientas</i>
mil	<i>mil</i>	<i>mil</i>	<i>mil</i>
millón	<i>millón</i>	<i>millón</i>	<i>millón</i>
un	<i>un</i>	<i>un</i>	<i>un</i>
ciento	<i>ciento</i>	<i>ciento</i>	<i>ciento</i>
doscientos	<i>doscientos</i>	<i>doscientos</i>	<i>doscientos</i>
trescientos	<i>trescientos</i>	<i>trescientos</i>	<i>trescientos</i>
cuatrocientos	<i>cuatrocientos</i>	<i>cuatrocientos</i>	<i>cuatrocientos</i>
quinientos	<i>quinientos</i>	<i>quinientos</i>	<i>quinientos</i>
seiscientos	<i>seiscientos</i>	<i>seiscientos</i>	<i>seiscientos</i>
setecientos	<i>setecientos</i>	<i>setecientos</i>	<i>setecientos</i>
ochocientos	<i>ochocientos</i>	<i>ochocientos</i>	<i>ochocientos</i>
novcientos	<i>novcientos</i>	<i>novcientos</i>	<i>novcientos</i>

Figura 5.10: Tercera plantilla de la base de datos de palabras manuscritas rellena.

Nº de muestra: Mano: <i>DERECHA</i> Sexo: <i>FEMENINO</i> Edad: <i>22</i>	INSTRUCCIONES PARA RELLENAR LA PLANTILLA: 1. Escribir con rotulador negro o azul. 2. Escribir con naturalidad y con todas las letras en minúsculas. 3. Escriba las palabras dentro de su recuadro: sin alcanzar el margen. 4. Intente centrar la palabra dentro del recuadro. 5. Escribir en sentido vertical. 6. No retoque las palabras escritas. 7. No doblar la plantilla ni realizar borrones.
--	--

millones	<i>millones</i>	<i>millones</i>	<i>millones</i>
una	<i>una</i>	<i>una</i>	<i>una</i>
y	<i>y</i>	<i>y</i>	<i>y</i>
pesetas	<i>pesetas</i>	<i>pesetas</i>	<i>pesetas</i>
\$	<i>\$</i>	<i>\$</i>	<i>\$</i>

Figura 5.11: Cuarta plantilla de la base de datos de palabras manuscritas rellena.

5.3. Procesado de imagen aplicado a la base de datos

Las palabras manuscritas siguen el mismo proceso de digitalización que se ha establecido para las aplicaciones anteriores, ya que se tratan de palabras aisladas, y no extraídas de un texto, sin la necesidad de tener que aplicar cualquier otro procesado. Realizando un resumen de las tareas aplicadas en este procesado, se pueden sintetizar en;

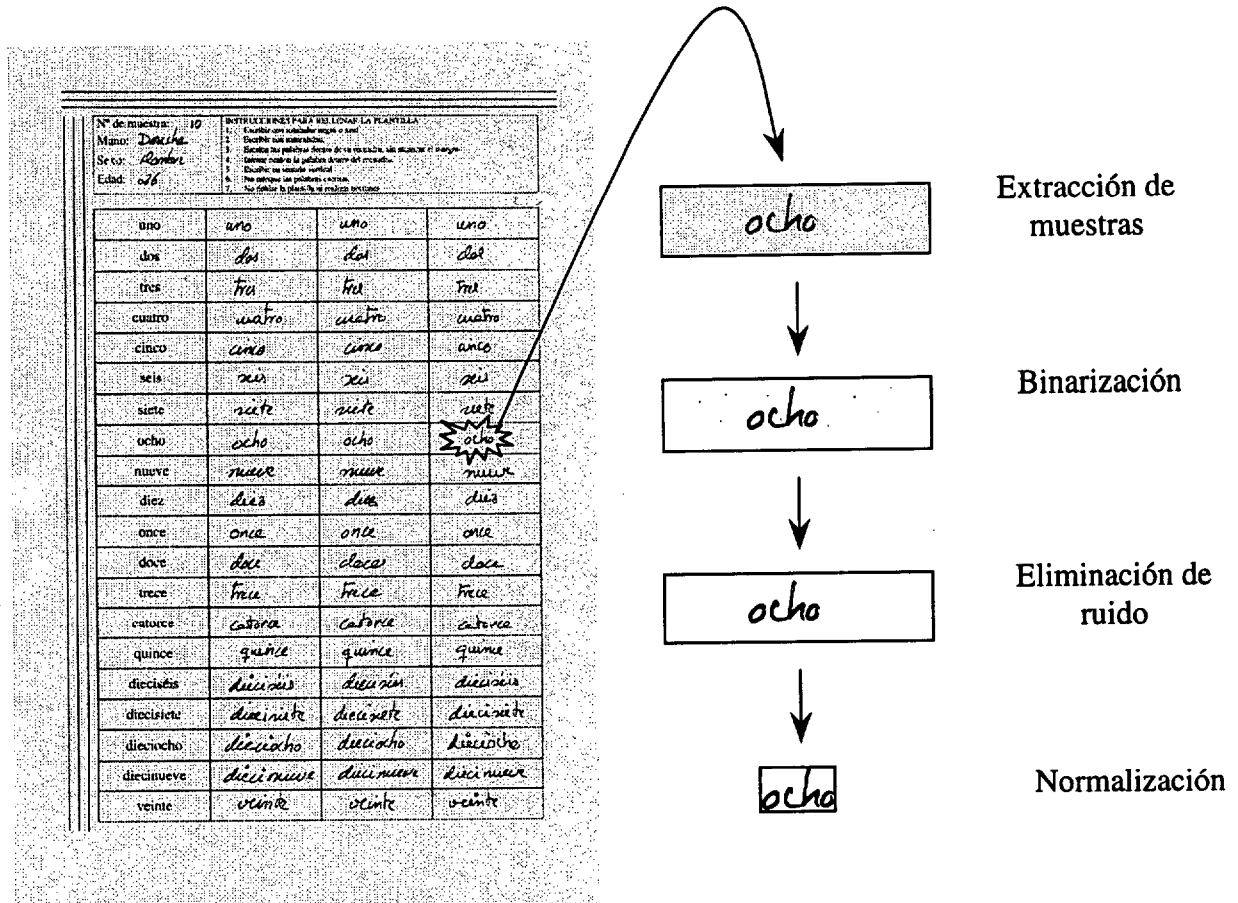
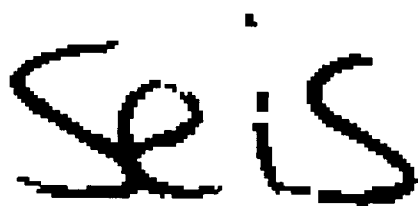


Figura 5.12: Preprocesado de imagen aplicado a la base de datos de palabras manuscritas.

5.3.1 Unión de las componentes de una palabras manuscritas

Antes de pasar a describir los métodos empleados para la extracción de parámetros de las palabras manuscritas, hay que solucionar un problema fundamental. Y es que a la hora de reconocer una palabra como un único elemento, y debido a las particularidades de los diferentes escritores, las palabras no se encuentran totalmente unidas, sino que estén formada por varios componentes, como se observa en la figura 5.13. La primera cuestión en este caso, es la de unir estas componentes para considerarla una sola palabra, antes de su reconocimiento.



(a) Palabra separada en varias componentes



(b) Palabra unida en un elemento

Figura 5.13: Unión de las componentes del número seis.

Este proceso es relativamente complejo debido a las muy variadas formas de escribir una misma palabra, por tanto se ha optado por emplear un sistema lo más robusto posible y aplicable a todas las palabras manuscritas. El siguiente organigrama describe los diferentes pasos que se han seguido.

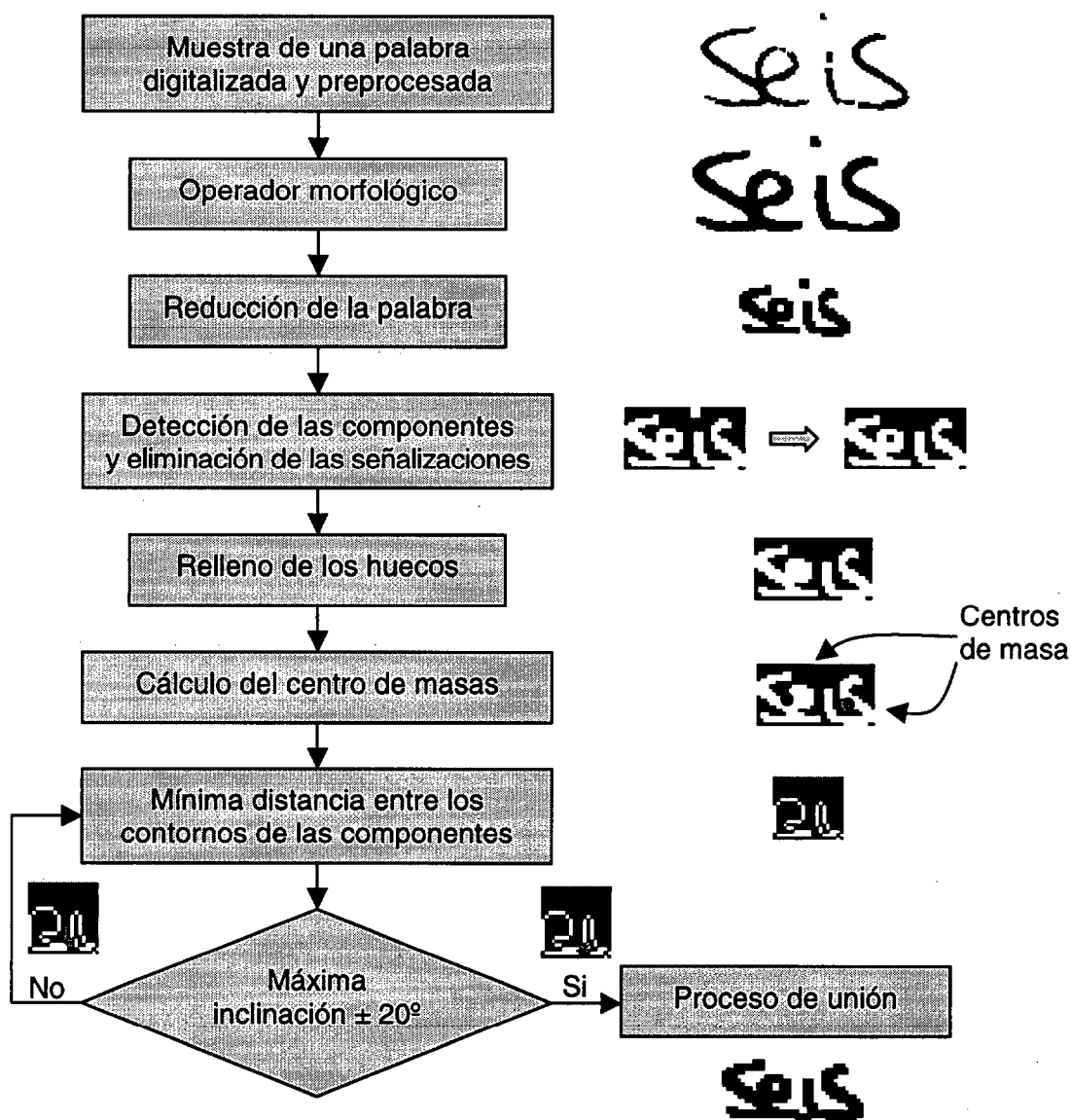


Figura 5.14: Proceso de unión de las componentes de una palabra.

Por tanto, el primer paso es hacer uso de la morfología para darle un mayor grosor al trazo de la palabra, en particular se aplica el operador de apertura y cierre y posteriormente se le aplica la dilatación de negro sobre fondo blanco. Una vez se ha engordado el trazo, la palabra se reduce, provocando el efecto ya comentado con los dígitos y letras manuscritas, de eliminar grandes variaciones y unificar la forma del patrón, con la intención de aumentar de esta forma la tasa de reconocimiento. Con ambas herramientas se puede lograr también, que las palabras se unan, reduciendo en sí el propio proceso.

En el caso de que aún quede alguna componente por unir, se aplica un algoritmo para contar y etiquetar las regiones aisladas que contenga la imagen. Si se detecta alguna región que sea pequeña en comparación con el tamaño medio de las regiones, ésta se eliminará. Con este efecto se consiguen eliminar los signos de señalización, como son los puntos en las “i” y “j”, los acentos y el apóstrofe de la “ñ”. A las regiones restantes se las rellenan del color del objeto, creando regiones macizas de contraste entre el objeto y el fondo. Ambos procesos se realizan con la imagen invertida. Esto va a ayudar para el cálculo del centro de masas de cada una de sus componentes (véase la figura 5.14).

A continuación se establece el criterio del cálculo de la unión de las componentes. Se calcula la distancia mínima entre los contornos de las dos regiones a unir. Para reducir el coste computacional se toma como valores extremos los centroides de las regiones, y se busca la distancia mínima que no tenga una inclinación superior a $\pm 20^\circ$, ya que es el trazo natural con el que se hace la unión de las letras de una palabra cuando se realiza la escritura. Cuando se haya determinado los dos puntos de unión entre las regiones, se unirán mediante una recta de la misma pendiente que se calcula de ambos puntos, siendo la recta de un grosor de tres píxeles (véase la figura 5.14).

5.3.2 Corrección de la inclinación de palabras

Una vez subsanado este problema de la unión de palabras, se termina el procesado de imagen con una particularidad, la corrección de la inclinación de la palabra. Cuando ésta es escrita, por lo general se suele realizar ligeramente ladeada. Este tipo de escritura es bastante habitual de encontrar [Bozinovic89].

Este proceso que puede ser importante para la mejor discriminación de las palabras es la corrección de la inclinación, debido a que el propio escritor intrínsecamente lo puede hacer así. A este tipo de letra se le suele conocer con el nombre de cursiva. Una característica que lleva asociada, es que se puede presentar la inclinación tanto a derechas o positiva, como a izquierdas o negativa, sobre la vertical. En la figura 5.15, se ilustran los dos casos comentados, y otro más, con la inclinación prácticamente nula, para diferentes muestras.

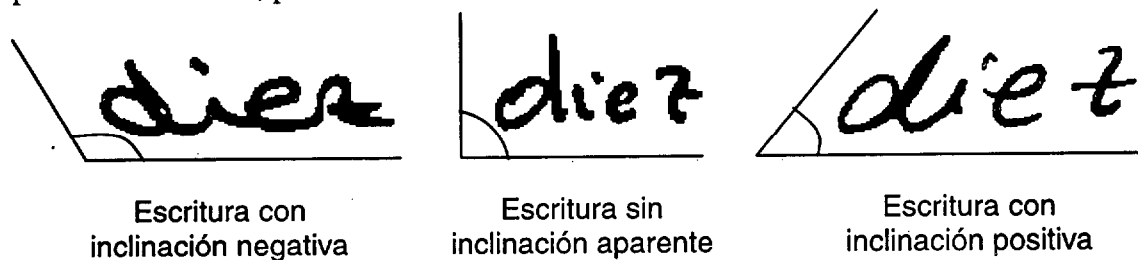


Figura 5.15: Ejemplos de las inclinaciones que puede presentar una palabra en su escritura.

En [Bozinovic89] se presenta un método para eliminar esta inclinación. El principal problema que plantea es la necesidad de establecer unos parámetros de ajuste en el sistema, que puede afectar de forma diversa, frente a las variaciones de anchura y altura de la escritura. Este problema se intenta solucionar, estableciendo un tamaño reducido y normalizado a todas las palabras, por lo que pasa a ser irrelevante, para la base de datos.

Este método consta de los siguientes pasos:

1. Se divide la imagen en filas.
2. Se eliminan todas las filas de la imagen que contengan una sucesión de píxeles a "1", que sea mayor a un parámetro preestablecido.
3. De las restantes líneas, se eliminarán aquellas que no se encuentren agrupadas en un cierto número, preestablecido por otro parámetro.
4. Los trazos contenidos en las líneas que no hayan sido eliminadas se usarán para la estimación de la inclinación de las palabras.
5. De cada trazo seleccionado, se calcula el centroide de la mitad superior y de la mitad inferior. Con ambos, se realiza una estimación local de la inclinación (I).

$$I = \frac{y_{\text{centroide superior}} - y_{\text{centroide inferior}}}{x_{\text{centroide superior}} - x_{\text{centroide inferior}}} \quad (5.1)$$

6. La media de todas las estimaciones locales se usará como estimación global.

Una vez determinada la inclinación, se procede a la aplicación de su corrección, consiguiendo finalmente, una palabra no cursiva. En la figura 5.16 se comenta un ejemplo de este método.

5.4 Parametrización de las palabras manuscritas

Si se parte del HMM y por tanto, se busca una secuencia para clasificarla, se pueden hacer dos posibles planteamientos a la hora de intentar parametrizar una palabra, por un lado se puede tomar las palabras como un único elemento, que son las técnicas que se han ido planteando hasta ahora, es decir, cada palabra es un patrón. Pero otro posible caso, es determinar su clasificación a través del reconocimiento aislado de cada una de las componentes de una palabra, a través de la estimación de cada una de sus letras, donde en este caso los patrones vienen dados por las 27 letras del alfabeto latino [Suárez02-pfc]. Ambas técnicas conducen al mismo resultado, el reconocimiento de la palabra.

Se van a plantear tres técnicas para la determinación de los parámetros para realizar la discriminación de patrones. Dos de ellas ya se han utilizado con anterioridad, y sólo se introduce una nueva. Se ha tenido en cuenta los resultados con formas manuscritas simples, siendo los mejores resultados los obtenidos con el contorno, y usando las muestras previamente reducidas. Por eso, los métodos presentados se van a centrar en este tipo de técnica, en seguir la secuencia del contorno. Estos sistemas son:

- Detección de la secuencia de las direcciones a partir de la envolvente
- Detección de la secuencia en coordenadas polares, a partir de la envolvente
- Detección de las anchuras y alturas a partir del centro de masas de una palabra.

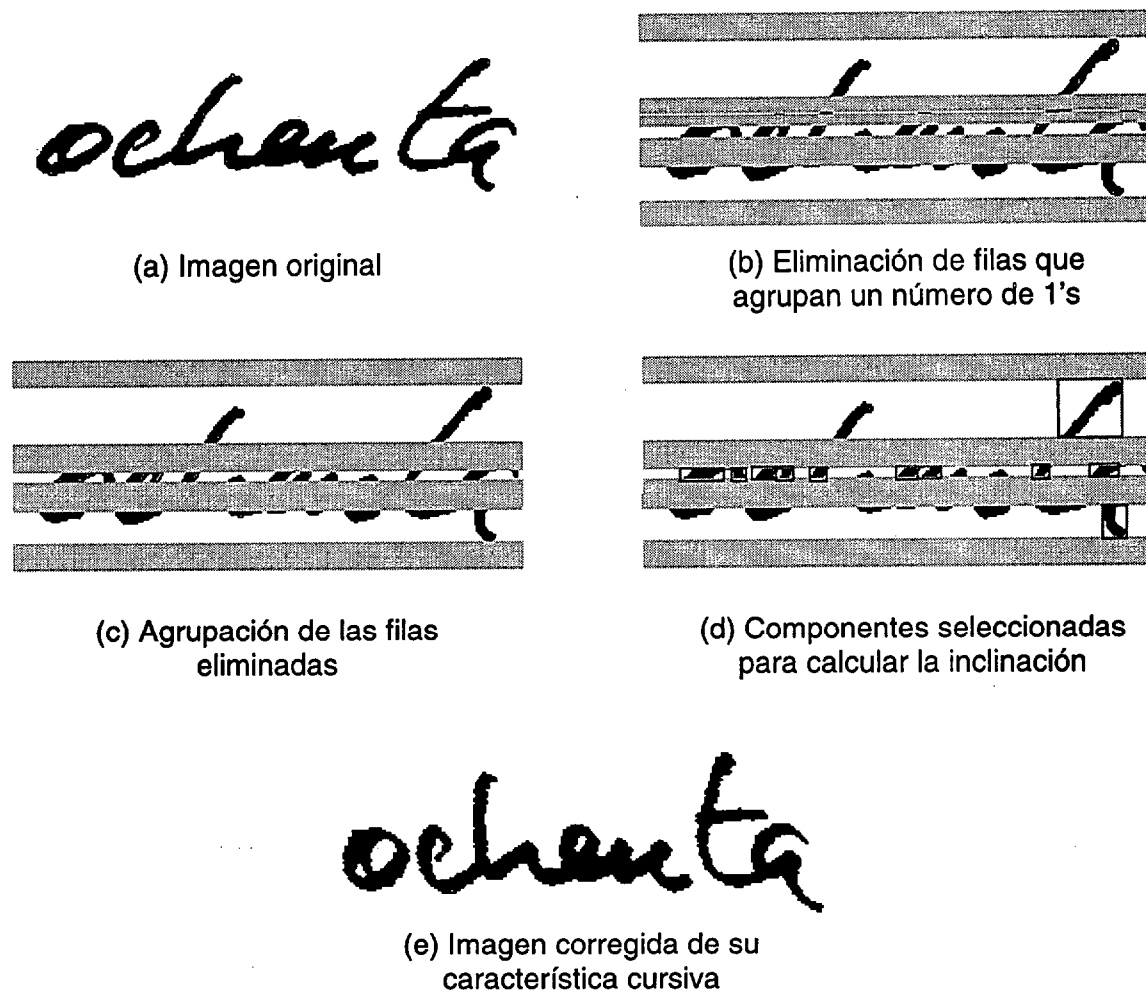


Figura 5.16: Método de corrección de la inclinación.

De estas tres técnicas utilizadas, las dos primeras ya se ha usado para extraer parámetros en las formas manuscritas simples, y la única novedad se introduce con la detección de los altos y los anchos, a partir del centro de masas de la palabra sobre el contorno de las misma.

Entrando más en detalle, la palabra se divide con respecto al centro de masas y trazando una horizontal, se establece el inicio de la medida desde esta referencia hasta el extremo del contorno de la palabra. Básicamente es aplicar el histograma al interior del contorno para su parte superior y lo mismo para su parte inferior, como se puede apreciar en la figura 5.17. Lo mismo se plantea a partir del centro de masa y trazando una recta vertical, para establecer los anchos hacia la derecha y hacia la izquierda. La unión de estos parámetros devuelve un vector que indicará la variación que tiene la palabra en función del número de píxeles introducidos en su escritura.

Otra parámetro utilizado es el número de transiciones que tiene de blanco a negro y viceversa, que se encuentran con respecto a los altos y a los anchos, pero ésta vez se establece para toda la palabra en estas dos vertientes.

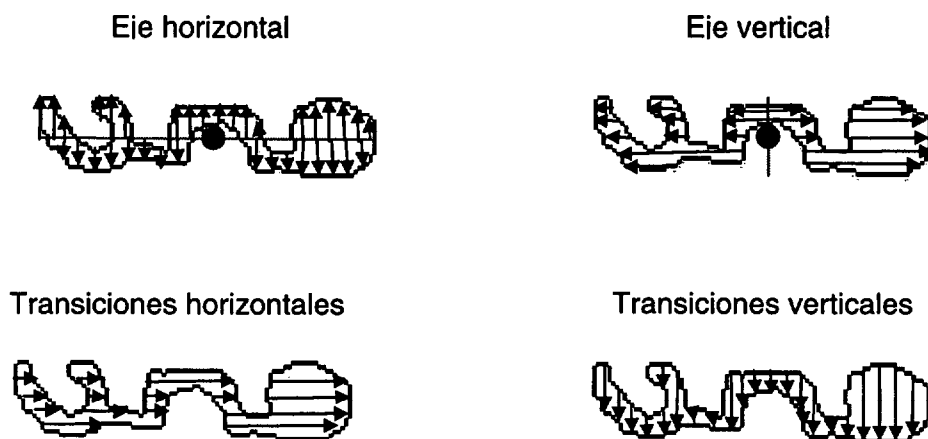


Figura 5.17: Método de parametrizar con altos, anchos y transiciones.

5.4.1 Proceso de la vectorización del contorno

Los procesos de vectorización se corresponden con los empleados a la hora de determinar los parámetros, y por tanto serán tres. Para el seguimiento de la secuencia a partir de la codificación de sus direcciones se ha presentado un nuevo criterio para establecer el punto de inicio de la vectorización. Se parte de la idea que las palabras son generalmente más alargadas que las letras o los dígitos, y con la pretensión de empezar a detectar la parte superior izquierda para el caso de una letra que tenga extremos superiores, se ha partido del sistema de barrer la imagen desde la zona izquierda superior, hacia la zona derecha, descendiendo siempre, con la misma pendiente; hasta detectar un píxel a negro, tal y como se indica en la figura 5.12. En esta figura también se observan dos posibles ejemplos de la detección del punto de inicio.

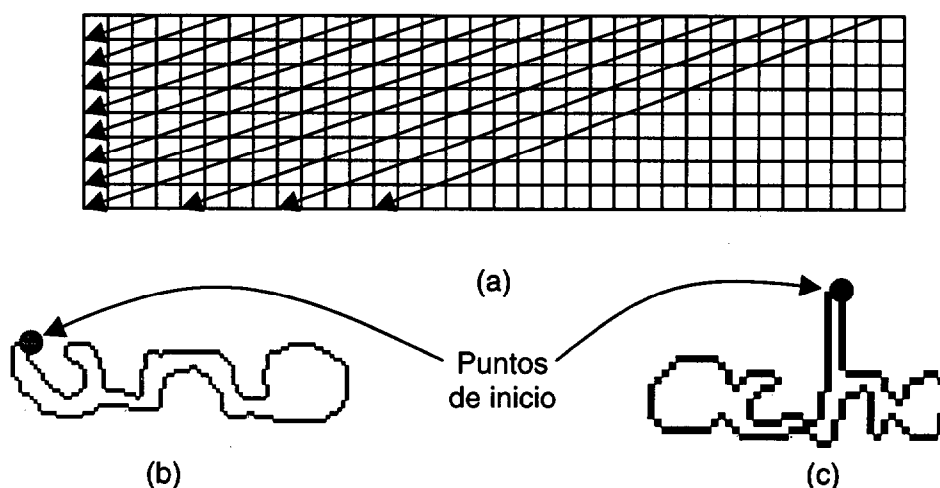


Figura 5.18: Convenio para los patrones de palabras manuscritas.

Una vez definido el punto de inicio, se comienza la vectorización a partir de éste, usando la matriz de saltos. Esta matriz de saltos se utiliza para mostrar cual es el camino que sigue el trazo. Se parte de una matriz de 3×3 , donde el elemento central es el píxel seleccionado, y cada uno de los ocho elementos limítrofes, indican las posibles

variaciones que puede sufrir el trazo (véase figura 5.19), ya que los píxeles del contorno están conectados mediante un solo vecino. Una vez definido el píxel al que ha saltado, éste volverá a ser el nuevo punto central (P), y así, hasta completar el trazo.

3	4	5
2	P	6
1	8	7

Figura 5.19: Matriz de saltos.

Pero los valores que toma el vector, también vienen dados por las matrices de prioridad, para mantener siempre el mismo sentido de giro, y se activan según el valor del salto anterior. Si éste fue codificado con 1, 2, 3, 4 ó se trataba de un punto de inicio, se elige la matriz de prioridad 1. Pero si los valores del salto anterior fueron 5, 6, 7 ó 8, se utiliza la matriz de prioridad 2, consiguiéndose que el giro se realice en el sentido de las agujas del reloj. Estas matrices se encuentran en la figura 5.20 [Morales99-pfc].

4	5	6
3	P	7
2	1	8

Matriz de prioridad 1

8	1	2
7	P	3
6	5	4

Matriz de prioridad 2

Figura 5.20: Matrices de prioridad.

El segundo método para la extracción de parámetros es el seguimiento de la secuencia del contorno en coordenadas polares. Para ello se establece el centro de masas de la palabra, y de seguido se recorre el contorno. Este mismo proceso ya se describe con detalle en el capítulo cuarto, ya que es el mismo método empleado para los patrones manuscritos simples. Por tanto, se recuerda que es independiente del punto de inicio, ya que una ordenación posterior, lo comienza en argumentos de cero grado y lo pone en orden hasta completarlo. Por tanto, los parámetros constarán de dos elementos, del módulo y del argumento para determinar cada punto del contorno. En la siguiente imagen se recoge un ejemplo con este proceso.



Figura 5.21: Vectorización del contorno con coordenadas polares.

Finalmente, quedaría por comentar el último método empleado, por extracción de los altos, anchos y transiciones. De la figura 5.17 se puede deducir cuales van a ser estas características, pero a continuación se va a definir como se agrupan para que puedan ser utilizadas por el HMM. La estructura que se forma con todos estos

parámetros, consta de tres columnas. En la primera columna se coloca el histograma de la parte superior, seguido del histograma de los anchos hacia la derecha. En la segunda columna va el histograma de la parte inferior de los altos, y el del los anchos hacia la izquierda. Y en la tercera columna, van las transiciones de los altos, y a continuación la de los anchos. En la siguiente figura se puede observar esta estructura.

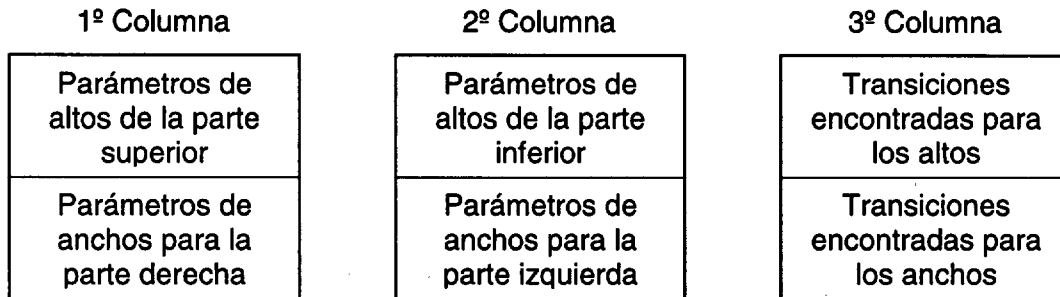


Figura 5.22: Estructura de datos de los parámetros altos, anchos y transiciones.

5.5 Clasificación de las palabras manuscritas

A partir de los vectores de información que se han recabado, se les va a aplicar un clasificador que obtenga buenos resultados para los mismos. Estas pruebas ya fueron realizadas para los dígitos y las letras y por tanto, se va a extrapolar a las palabras ya que se pretende generalizar esta técnica a las aplicaciones manuscritas. Por tanto, el clasificador a utilizar para modelar estas variaciones secuenciales-temporales son los modelos ocultos de Markov (HMM) [Rabiner93].

La clasificación va a costar de dos etapas, el entrenamiento y el test. En la etapa de entrenamiento se van a fijar los parámetros del clasificador tales como el número de estados, el número de símbolos de observaciones por estado, sistema BAKIS y la salida del algoritmo por umbral.

Para el caso de las coordenadas polares y de los histogramas de los altos, anchos y transiciones, el número de símbolos de observaciones por estado se ha fijado en 32, siendo este valor determinado de forma empírica [Gómez01-pfc]. Para la codificación por la matriz de saltos según las direcciones se ha mantenido a 8, al igual que en patrones simples [Morales99].

Una vez que se han fijado estas variables, el modelo devolverá los valores de la matriz de transición de estados (A), la matriz de probabilidad de símbolos (B), y la matriz de la probabilidad de estados iniciales (π). Este modelo $\lambda(A,B,\pi)$, hará uso del procedimiento de Baum-Welch para su ajuste y maximización de las probabilidades de éxito de los patrones.

En la siguiente etapa, se realizará el test, y en función de los resultados obtenidos, se va a determinar si los parámetros del clasificador son los más adecuados para maximizar la discriminación. De esta forma, se puede lograr ajustar los valores de funcionamiento del clasificador para aumentar la tasa de reconocimiento.

5.6 Resultados de la clasificación de palabras manuscritas

En este apartado se va a cuantificar el grado de eficacia que puedan tener los métodos de extracción de parámetros presentados en este capítulo, así como el ajuste de los parámetros del clasificador para su mayor discriminación. Por tanto las variables que se analizarán son:

- El número de estados de la cadena
- El método de extracción de los parámetros para HMM basado en el contorno (por codificación de las direcciones de la secuencia, por seguimiento en coordenadas polares, y por los altos, anchos y transiciones del contorno)
- El número de escritores empleados durante el entrenamiento del clasificador
- Discriminación por la corrección del aspecto cursivo de la palabra

Para la realización de estas pruebas se han fijado unos parámetros ya estudiados para los dígitos y las letras manuscritas, extrapolando de esta manera un único método para reconocer las aplicaciones manuscritas. Por tanto, se hace uso, otra vez, del mismo clasificador (HMM), y además, se establece que todas las pruebas se realicen con las muestras ligeramente reducidas y normalizadas, para aumentar su parecido entre ellas, ya que todas provienen de diferentes escritores. Por otro lado, se van a realizar las pruebas con independencia del escritor, es decir, las muestras usadas para el entrenamiento, no se utilizarán para el test.

Para todas las técnicas de extracción de parámetros se probó el beneficio o no, de corregir la inclinación de la palabra, ya que a priori se puede estimar una mayor correlación entre muestras y una mejor tasa de reconocimiento, y por tanto, se desarrollará experimentalmente.

El primer experimento planteado es la variación del número de estados del HMM en función de las diferentes técnicas empleadas. Para ello, se fijó el número de escritores para entrenar en un 50%, que corresponden a 20 escritores para cada plantilla, y se probó sobre las diversas técnicas y sus variaciones. Estos resultados son los mejores obtenidos para cada técnica, determinando así su número de estados óptimo para maximizar su discriminación, y éstos se pueden observar en la siguiente tabla 5.2.

<i>Técnica de extracción de parámetros</i>	<i>Número de estados</i>	<i>Tasa de reconocimiento</i>
Secuencia	55	72.85 % ± 1.66
Secuencia (corrección de cursiva)	65	76.18 % ± 3.04
Polares	60	56.96 % ± 0.49
Polares (corrección de cursiva)	60	56.08 % ± 2.74
Altos, anchos	55	73.24 % ± 2.05
Altos, anchos (corrección de cursiva)	25	75.10 % ± 0.59
Altos, anchos y transiciones	30	78.73 % ± 2.06
Altos, anchos y transiciones (corrección de cursiva)	30	77.16 % ± 2.45

Tabla 5.2: Resultados de las diversas técnicas planteadas para reconocer palabras, corrigiendo y sin corregir el efecto de cursiva.

Estas pruebas fueron hechas para diez palabras del total del vocabulario planteado. Una vez se optimice el resultado para éstas, se extrapolará al resto de la base

de datos. Estas diez palabras fueron las diez primeras que se pueden encontrar, es decir, del uno al diez.

Sobre los resultados de la tabla 5.2 con respecto a la corrección de la inclinación se observa que para secuencia y altos anchos mejora el resultado en algunos puntos, mientras que en polares y altos, anchos y transiciones el resultado casi se mantiene, sólo descendiendo en un punto. Por tanto, el hecho de aplicarlo es una mejora, en líneas generales.

La técnica que mejor se adapta al HMM para discriminar palabras, es la de altos anchos y transiciones, ya que mantiene las máximas tasas tanto si se corrige como si no su inclinación, y al mismo número de estados (véase la tabla 5.2). A continuación le sigue la técnica de altos anchos y la de secuencia, con unos puntos por debajo. Y a gran distancia se encuentran las coordenadas polares.

Seguidamente se va a estudiar la cantidad de escritores más adecuada para entrenar, para las técnicas de extracción de parámetros de altos, anchos y transiciones y para la de secuencia, teniendo en cuenta la corrección de la inclinación de la palabra. Se va a dividir la base de datos según un porcentaje, que se destinará para el entrenamiento, el resto servirá para el test. Para la tabla anterior, se había fijado el entrenamiento al 50%, en este caso se va a fijar el número de estados (determinado con anterioridad) y se va a variar el porcentaje de entrenamiento.

Para la técnica de seguimiento de la secuencia por codificación de sus direcciones y corregida su inclinación, se ha fijado el número de estados a 50 y a 65, y se han repetido las pruebas en dos ocasiones (véase el anexo C), mostrando en la siguiente tabla los valores promediados:

Número de estados	% de entrenamiento	Tasa de reconocimiento
50	40	71.07 % ± 1.56
50	50	75.69 % ± 0.58
50	60	72.50 % ± 1.75
50	70	71.84 % ± 0.83
65	40	70.91 % ± 0.57
65	50	76.18 % ± 3.04
65	60	72.50 % ± 0.50
65	70	70.84 % ± 2.83

Tabla 5.3: Resultados de los diversos entrenamientos aplicados a la técnica de secuencia, corrigiendo el efecto de cursiva.

De esta tabla 5.3 se puede concluir que el mejor resultado se obtuvo para el 50% de la base de datos para el entrenamiento, por lo que se va desechar el seguir trabajando con la variación del porcentaje de muestras para entrenar el HMM.

Para la técnica de altos, anchos y transiciones corrigiendo su inclinación, se han realizado en este caso, cuatro pruebas diferentes para cada experimento, fijando el número de estados a 30, según la tabla 5.2. Se ha escogido la opción de corrección de inclinación, ya que según las tablas del anexo C, la tendencia es a mejorar, y a la estabilidad, con este efecto. Los resultados promediados que se han obtenido, están registrados en la siguiente tabla:

<i>% de entrenamiento</i>	<i>Tasa de reconocimiento</i>
40	71.72 % ± 1.33
50	77.55 % ± 2.46
55	79.55 % ± 2.33
60	80.13 % ± 0.98
70	76.58 % ± 3.18
80	78.88 % ± 0.96

Tabla 5.4: Resultados de los diversos entrenamientos aplicados a la técnica de altos, anchos y transiciones, corrigiendo el efecto de cursiva.

Por tanto, se va a continuar el estudio con esta técnica de altos, anchos y transiciones, con 30 estados para HMM, y con la corrección de cursiva, ya que si se remite a esta tabla, se puede observar que mejoran los resultados hasta el 80.13%, para un entrenamiento del 60%.

Por tanto, se ha encontrado la mejor técnica que maximiza el reconocimiento, pero aún se le va a aplicar una técnica de multietiquetado, ya que los valores obtenidos en la extracción de parámetros no estaban limitados por el número de símbolos por estado, como en la secuencia, sino que tiene un cierto grado de variación. Por esta razón, se va a estudiar este parámetro propio del clasificador HMM. Estas pruebas se han repetido en dos ocasiones, recogiendo los siguientes resultados promediados:

<i>Multietiquetado (M)</i>	<i>Tasa de reconocimiento</i>
1	78.00 % ± 0.25
2	77.85 % ± 2.75
3	78.63 % ± 4.12
4	82.36 % ± 0.20
5	83.24 % ± 1.47
6	80.49 % ± 4.42
8	81.47 % ± 2.06
10	77.46 % ± 5.30
12	76.28 % ± 1.56

Tabla 5.5: Resultados del multietiquetado aplicado a la técnica de altos, anchos y transiciones, corrigiendo el efecto de cursiva.

De esta tabla se observa, que los mejores resultados se encuentran para un multietiquetado de M=4 y M=5, llegando en media hasta el 83.24% de reconocimiento. Por tanto, se han hallado los parámetros del clasificador más adecuados, para unas muestras totalmente optimizadas en su procesado. Antes de finalizar se va a comprobar con estos mismos parámetros el resultado para los altos, anchos y transiciones con y sin corrección de inclinación, para probar finalmente el beneficio de ésta.

<i>Técnica de extracción de parámetros</i>	<i>Multietiquetado</i>	<i>Tasa de reconocimiento</i>
Altos, anchos y transiciones	4	81.00 % ± 2.89
Altos, anchos y transiciones	5	80.19 % ± 4.34
Altos, anchos y transiciones (corrección de cursiva)	4	82.50 % ± 0.73
Altos, anchos y transiciones (corrección de cursiva)	5	83.13 % ± 3.05

Tabla 5.6: Resultados del multietiquetado aplicado a la técnica de altos, anchos y transiciones, con y sin la corrección del efecto de cursiva.

Se observa, y se puede concluir nuevamente, que el hecho de la corrección de la inclinación proporciona mejores resultados, llegando en este caso hasta 83.13%, muy cercano al anterior de 83.23%. De esta forma, definitivamente se puede decir, que la mejor técnica de las estudiadas para la clasificación de palabras manuscritas es la de altos, anchos y transiciones, corrigiendo el efecto de cursiva, y utilizando un clasificador HMM con 30 estados, siendo entrenado con el 60% de la base de datos, es decir, con 24 escritores de los 40 posibles.

Por tanto, este resultado optimizado se le va aplicar a toda la base de datos. Esta aplicación se le hace agrupando a las palabras por el mismo orden que se tiene de captura, indicado en las figuras 5.4, 5.5, 5.6 y 5.7. Esta prueba se ha realizado en cuatro ocasiones para cada grupo de palabras con la finalidad de validar el reconocimiento, y extraer ningún valor de pico, de esta forma los resultados promediados obtenidos son:

<i>Grupos de palabras</i>	<i>Número de palabras</i>	<i>Tasa de reconocimiento</i>
Del uno al diez	10	83.19 % ± 1.92
Del uno al quince	15	78.54 % ± 2.04
Del uno al veinte	20	75.63 % ± 2.31
Del uno al treinta	30	61.91 % ± 2.92
Del uno al cuatrocientas	40	57.57 % ± 4.68
Del uno al doscientos	50	57.06 % ± 1.26
Del uno a la y	59	52.92 % ± 1.72

Tabla 5.7: Resultados de los diferentes agrupamientos de las palabras para su reconocimiento.

Si se observa del agrupamiento establecido, las caídas más significativas vienen dadas cuando las palabras están más correladas, por ejemplo para 30 palabras, es cuando se incluyen los patrones del veintiuno al veintinueve (parecidos entre ellos) además de los anteriores, produciendo esa bajada significativa de las tasas. Lo mismo ocurre cuando se introducen el trece (parecido al tres), el catorce (parecido al cuatro) y del dieciséis al diecinueve, en el agrupamiento de 20, que tiene otra caída significativa.

Por el contrario, cuando las palabras introducidas son bastantes incorreladas, la tasa de reconocimiento se mantiene, como en el agrupamiento del 40 al 50, cuyas palabras son del doscientas hasta el novecientas entre otras.

5.7 Aplicación: segmentación y clasificación

Como ya se comentaba en los inicios de este capítulo, otra forma de clasificar las cadenas manuscritas es mediante la segmentación de cada uno de los elementos que la componen. Quizás no es directamente el planteamiento de los objetivos de esta tesis, pero si indirectamente, ya que cada forma de la cadena se puede estimar mediante su información secuencial. Desde este punto de vista se ha trabajado, y se está trabajando en la segmentación de aplicaciones manuscritas, como pueden ser los códigos postales y las propias palabras.

5.7.1 Clasificación de códigos postales

Los códigos postales estudiados corresponden al utilizado por el Estado español, cuya composición es totalmente numérica. Si se combinan con repetición cinco dígitos, se pueden obtener todas las posibles formas, aunque no todas tienen asignada una región geográfica.

Esta clasificación es un problema muy cercano al del texto manuscrito, siendo igualmente difícil y muy abierto [Savicc99]. En una aproximación a su resolución, este proceso se ha dividido en varios problemas de análisis de imágenes [Govindaraju95]: detección de las direcciones, separación de las líneas de texto, división de las líneas en palabras, segmentación de las palabras en caracteres, y reconocimiento de letras y símbolos especiales.

Además de estos problemas de procesado digital de imágenes, los desarrollos realizados de reconocimiento de direcciones manuscritas en cartas [Cohen94] [Govindaraju93] se han encontrado como dificultad añadida, la falta de uniformidad en el formato de las direcciones, lo cual dificulta el proceso automático de detección de los números del código postal.

El primer y necesario paso en el desarrollo de los algoritmos para la detección automática de los códigos postales manuscritos en cartas es contar con una base de datos o colección de cartas con las direcciones manuscritas. La base de datos desarrollada consta de 400 cartas con dirección manuscritas todas por diferentes escritores en sobres de diferentes tamaños, colores claros y formatos, con uno o más sellos situados de formas diferentes, y direcciones con distinta colocación y formato (dirigidas a capitales de provincia, poblaciones, pedanías, particulares, empresas, etc.). Como simplificación del problema tan solo se ha tenido en cuenta cartas nacionales. Un ejemplo de carta perteneciente a la base de datos, puede verse en la figura 5.23.

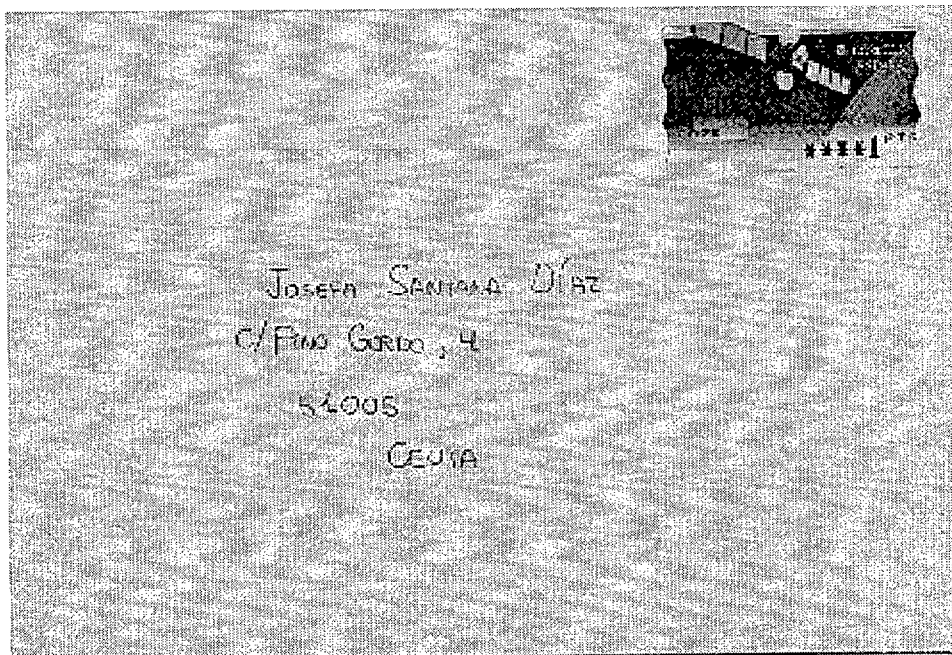


Figura 5.23: Ejemplo de sobre con código postal 51005 digitalizado.

Con el sobre en formato digital, se ha procedido a extraer la parte del sobre que contiene la dirección, y a dividir dicha zona del sobre en líneas. Como simplificación del problema, y tras observar en nuestra base de datos que el código postal siempre se encuentra en la última o penúltima línea, no se han extraído todas las líneas, sino tan solo las dos inferiores. Evidentemente habrá que corregir la inclinación de las líneas para facilitar su posterior tratamiento. La zona de la dirección extraída del sobre del ejemplo de la figura 5.23, así como las líneas separadas, pueden verse en la figura 5.24.

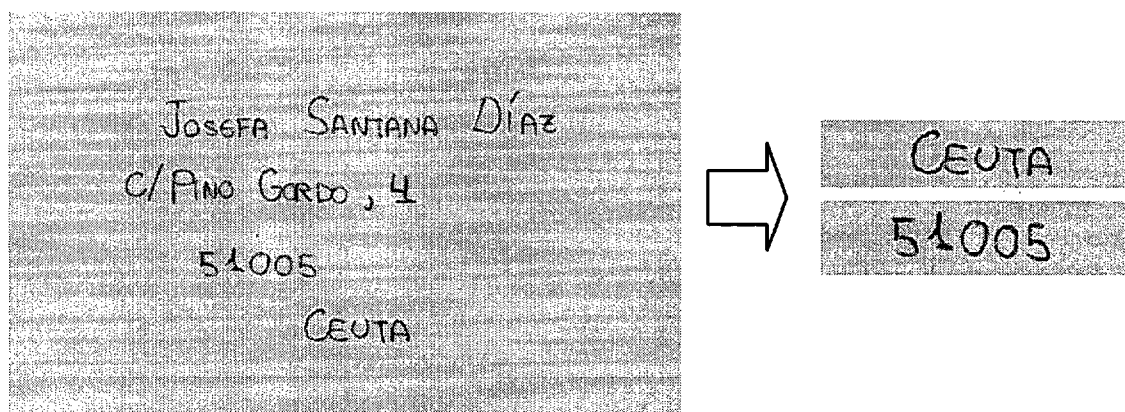


Figura 5.24: Zona de la dirección y líneas separadas extraídas del sobre con código postal 51005.

Una vez obtenidas las líneas, se ha procedido a preprocesarla pasando cada línea a blanco y negro, y eliminando el ruido e invasiones de las líneas vecinas. El proceso de paso a blanco y negro se realiza con la línea, y no con el sobre, para obtener el umbral de paso a blanco y negro que maximice la nitidez del texto. En la línea en blanco y negro se separan las palabras y se coge la primera palabra de cada línea, como en el ejemplo de la figura 5.25. En los casos en que el código postal no está en la primera palabra, éste se encuentra tan pegado a la primera palabra que el algoritmo desarrollado para separar palabras, toma las dos primeras palabras como una, no perdiéndose así el código postal.

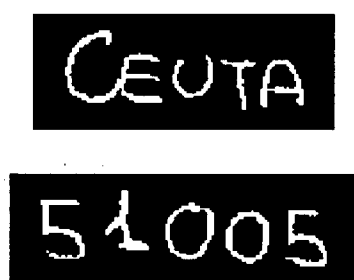


Figura 5.25: Primeras palabras en blanco y negro sin ruido de las dos últimas líneas del sobre con código postal 51005.

Una vez separadas las palabras se segmentan en caracteres, tal y como se ve en la figura 5.26. El algoritmo de segmentación tiene en cuenta que los caracteres de la palabra pueden estar conectados para separarlos (como en el caso de la figura 5.25), o que pueden estar inclinados (escritos en cursiva) para enderezarlos. Evidentemente los algoritmos de segmentación de palabras en caracteres están orientados a segmentar

correctamente cadenas de dígitos, no prestándose atención a errores en la segmentación de palabras.

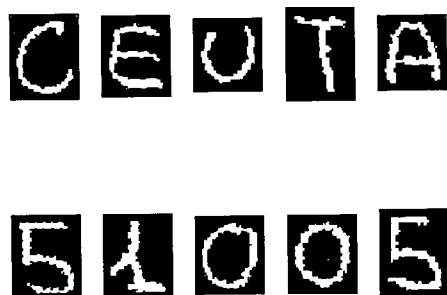


Figura 5.26: Primeras palabras de las dos últimas líneas del sobre con código postal 51005 segmentadas.

La segmentación de una palabra o código postal no siempre será única; en los casos en que los dígitos del código postal se tocan, o el código postal contiene dígitos que presentan columnas con muy pocos píxeles, pueden originarse múltiples posibilidades de segmentación. Ésto se ilustra en el ejemplo de la figura 5.27. En estos casos será el reconocedor de dígitos, el que decidirá cual de las segmentaciones es la buena y el valor numérico del código postal.

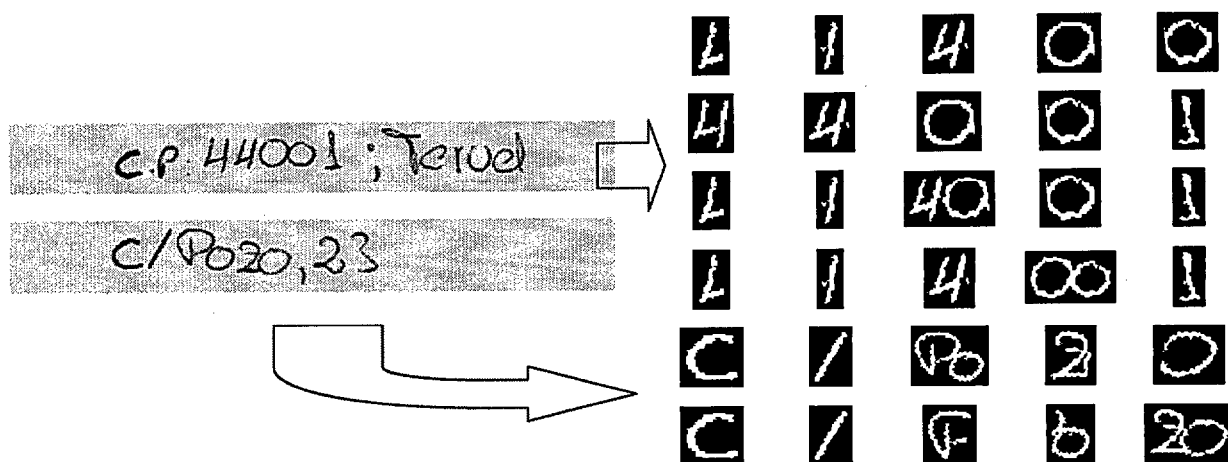


Figura 5.27: Líneas del sobre con código postal 44001 cuyas primeras palabras de las dos últimas líneas presentan múltiples posibilidades de segmentación.

Como reconocedor de dígitos se utilizará el desarrollado por [Travieso99]. Este estudio aportó un reconocedor de dígitos basado en modelos ocultos de Markov, que tras adaptarlo a este proyecto, presenta una tasa de reconocimiento del 92.8% [Exposito01-pfc], lo cual supondrá una tasa teórica de reconocimiento de códigos postales (reconocimiento correcto de cinco dígitos consecutivos) del 68.82%.

El funcionamiento del reconocedor es sencillo: para cada posible segmentación se pasan por el reconocedor sus cinco caracteres, asignando el reconocedor a cada posibilidad un código postal y una probabilidad logarítmica de acierto (véase la tabla 5.8). Las posibilidades cuyos códigos postales asignados no existen en la guía de códigos postales son descartadas. De las posibilidades cuyo código postal existe, en el ejemplo de la tabla 5.8, se tiene el código 21400 de Ayamonte, y el 44001 de León, se

toma como reconocido el código postal de mayor probabilidad logarítmica de acierto, que en el caso de la tabla 5.8, es el código postal correcto.

Posibilidades de códigos postales	Código postal reconocido	¿Existe en correos?	Probabilidad de cada posibilidad	Código postal reconocido
1ª posibilidad	21400	SI	-445	NO
2ª posibilidad	44001	SI	-424	SI
3ª posibilidad	21401	NO	-557	
4ª posibilidad	21401	NO	-498	
5ª posibilidad	00882	NO	-545	
6ª posibilidad	00028	NO	-593	

Tabla 5.8: Ejemplo de reconocimiento del código postal.

En el caso de que no exista ninguna posibilidad con código postal existente en la guía de códigos postales, o que la posibilidad con código postal válido tenga una probabilidad logarítmica muy pequeña por debajo de un umbral predefinido, el reconocedor devuelve código postal desconocido.

Los resultados del reconocedor de códigos postales en cartas con direcciones manuscritas se presentan en la tabla 5.9. Como puede verse, la calidad real del sistema obtenido puede definirse con una tasa de error del 10% y una tasa de acierto del 67%. Estos resultados, que coinciden con los teóricos esperados, están a la par de los presentados en la literatura [Dzuba97][Blumenstein97] sobre reconocimiento de códigos postales manuscritos, si se toman en consideración, tan solo, aquellos resultados que no utilizan la ciudad de destino para verificar el código postal reconocido.

Resultados con la base de datos de 400 sobres		
Número de aciertos	243	67.75%
Número de errores	39	10.75%
Número de desconocidos	73	21.5%

Tabla 5.9: Resultados del reconocedor de códigos postales manuscritos.

Evidentemente, esta tasa no hace comercial la aplicación de este estudio [Exposito01-pfc], pero sí aportará una valiosa experiencia con datos que favorecerán en un futuro una aplicación más cercana a un uso práctico.

5.7.2 Clasificación de palabras por segmentación

En este capítulo se ha desarrollado el reconocimiento de las palabras, como un solo elemento, aunque este constituido por un conjunto de letras. El problema que presenta, como ya se ha planteado, es la baja tasa de reconocimiento a medida que el tamaño del vocabulario de palabras va creciendo, y más acentuado en este caso que se intenta reconocer palabras muy parecidas o correladas entre sí, como son los valores numéricos que se pueden encontrar en un cheque bancario.

La otra posibilidad planteada es el reconocimiento de las palabras a partir de sus componentes o las letras que lo forman. En este sentido se ha desarrollado y se está

trabajando en un sistema de segmentación de palabras incluidas en un texto, para su reconocimiento final. Aunque esta aplicación se sale de lo establecido en los objetivos, representa un sistema paralelo de reconocimiento y una línea de investigación que se deriva del trabajo de esta tesis [Suárez02-pfc].

Este proceso se inicia con la creación de una base de datos de párrafos manuscritos, que posteriormente fueron digitalizados a una imagen. Esta base de datos consta de 66 escritores, donde cada escritor ha desarrollado un formulario con texto. El texto en todos los casos es diferente, para contribuir a la variedad de palabras y por tanto de letras. En la siguiente figura se recoge un formulario tipo.

Recomendaciones

- Escriba normalmente.
- Cuide la caligrafía.
- Intente no unir las palabras entre sí.
- Se recomienda utilizar ROTULADOR de trazo fino.

MUCHAS GRACIAS POR SU COLABORACIÓN

Sexo: Hombre Mujer Escriba entre las líneas

Edad:

Escritura: Zurdo Diestro del texto que se proporciona.

Catástrofe natural más extendida y dañina. Producen daños no solo directos por acción del viento, sino también por las inundaciones. El peor de los últimos años causó la muerte a 500.000 personas en Bangla Desh en 1970, mientras que el desbordamiento del río Yang-tsé en China, tras una serie de tifones, produjo entre 40 y 50 millones de

Figura 5.28: Ejemplo de un formulario de un texto manuscrito.

De esta base de datos inicial, se derivan otras bases de datos surgidas de la inicial. Para poder llevar a cabo el reconocimiento, es necesario entrenar al reconocedor. Por tanto, se necesitan muestras con las cuales el módulo de reconocimiento pueda aprender a discernir entre unas y otras. Por eso, a medida que se va descomponiendo el párrafo, se irán creando bases de datos de las líneas que surjan, de las palabras y de los signos de puntuación contenidos dentro de las líneas, y finalmente de las letras que componen las palabras. Las tres últimas sub-bases de datos, pueden proporcionar información para la ayuda al reconocimiento final de la palabra.

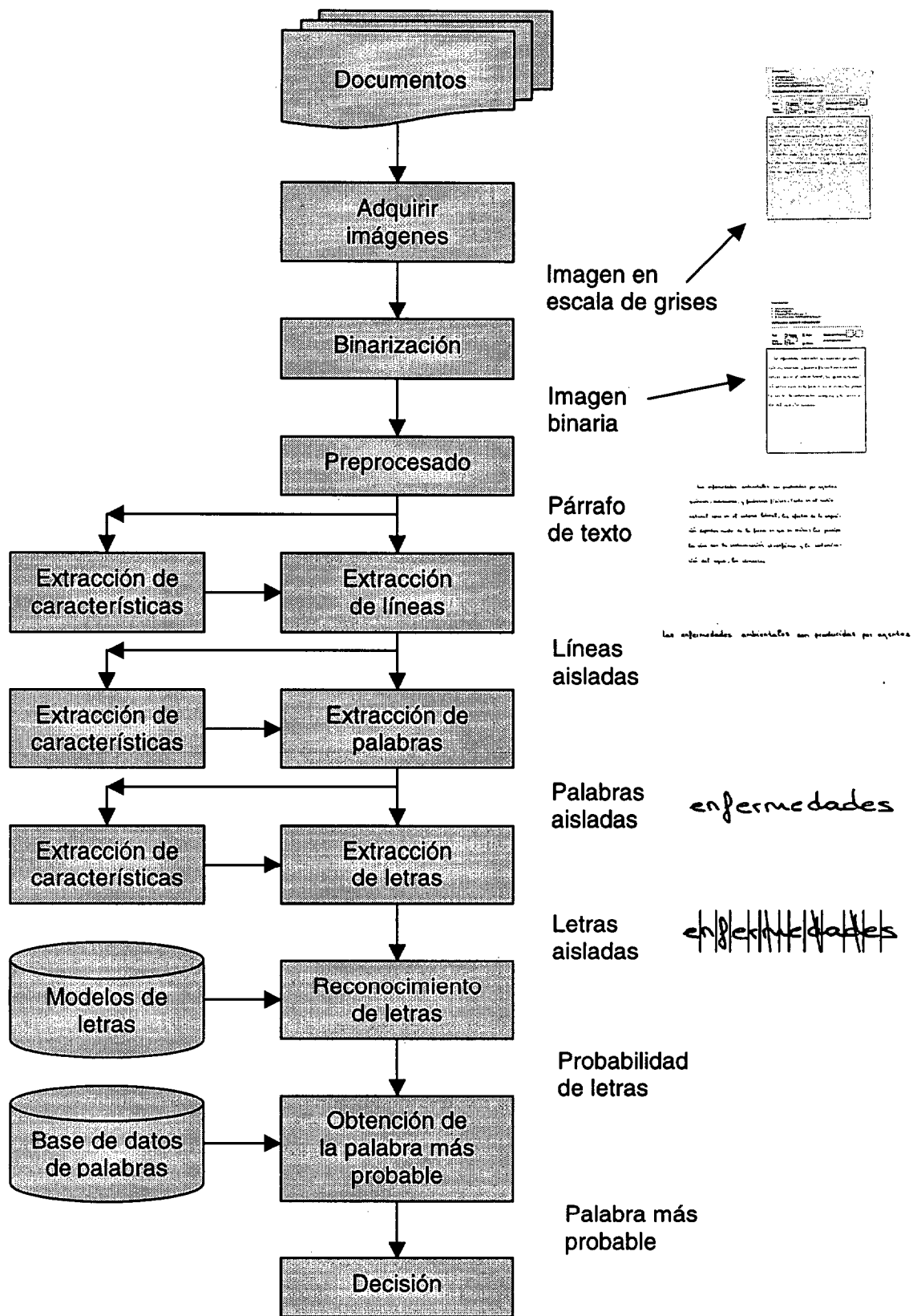


Figura 5.29: Descripción del proceso de reconocimiento.

Si este proceso consiste en la digitalización de la base de datos a posteriori, se establece un sistema *off-line*, el cual, será sometido a un procesado digital de imagen. Estos procesos son los típicos de un sistema de reconocimiento manuscrito: binarización, eliminación del ruido, y normalización, tal y como se recogen a modo de ejemplo en la figura 5.12, pero aplicado al párrafo completo.

Una vez que se tiene la imagen preprocesada, el proceso a seguir se establece con la localización de las líneas de texto, así como su inclinación y si procede la eliminación de dicha inclinación. Las líneas de texto se extraen a la vez que se obtiene información importante, tal como el grosor del trazo, inclinación de la escritura, separación de letras, etc. A partir de toda esta información global junto con la información obtenida localmente se pueden extraer las palabras de las líneas. Finalmente, las palabras hay que segmentarlas para extraer las letras. Todo este proceso puede derivar en el reconocimiento de la palabra a partir de sus componentes más simples. En la figura 5.29 se presenta el organigrama de todo este proceso, pasando a detallar cada parte más en particular.

El primer paso a dar, una vez determinada la imagen del párrafo ya preprocesada, es la estimación y eliminación de la inclinación de las líneas. Se han estudiado los métodos existentes actualmente, tales como el de proyecciones, y la muy conocida transformada Hough. Tanto el uso de esta transformada como de las proyecciones consiguen un 100% en la corrección de las líneas extraídas.

El siguiente paso es detectar las palabras existentes dentro de una línea. La primera labor a realizar es establecer dentro de una línea, los límites para separar las palabras. Para esta estimación se han utilizado tres métodos bien diferenciados, los polígonos convexos, las cajas acotadas, y la distancia horizontal. Los polígonos convexos se basa en cerrar una componente conectada con el menor polígono posible. Las cajas acotadas se definen como el mínimo rectángulo que contiene a la componente conectada. Y por último el método de la mínima distancia horizontal que utiliza la medida de las cajas acotadas si éstas no están solapadas horizontalmente, y en caso de solape, calcula la distancia para cada línea de la imagen, y toma el mínimo de las distancias calculadas.

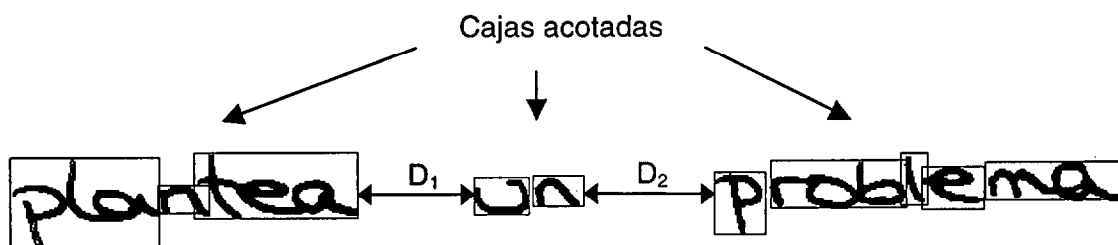


Figura 5.30: Ejemplos de medidas usando cajas acotadas y distancia horizontal.

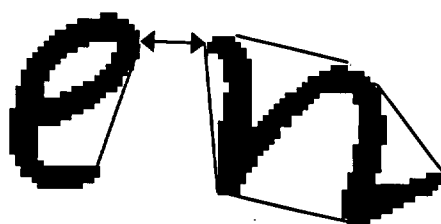


Figura 5.31: Ejemplo de medida basado en polígonos convexos.

En la siguiente tabla se recoge el éxito logrado con estos métodos.

Método utilizado	% Segmentación correcta	% Palabras partidas	% Palabras unidas
Polígonos convexos	94 %	2 %	4 %
Cajas acotadas	92 %	3 %	5 %
Distancia horizontal	85 %	5 %	10 %

Tabla 5.10: Medidas de la separación de las palabras.

En referencia a la estimación y eliminación de la inclinación de las palabras en su escritura, se ha utilizado el mismo algoritmo empleado en las palabras aisladas, encontrándose en [Bozinovic89].

Un paso previo a la separación de las palabras o componentes conectadas es la detección y eliminación de los signos de puntuación, ya que éstos actúan como separadores espaciales, reduciendo la distancia geométrica entre componentes. Como puede observarse en la figura 5.32 si se considera el signo de puntuación como cualquier otra componente, la distancia entre las componentes se reducirá y no se conseguirá separar las palabras. Sin embargo si se detecta el signo de puntuación no sólo se podrá eliminar sino, que esto permitirá separar las palabras (sin necesidad de utilizar otras medidas).

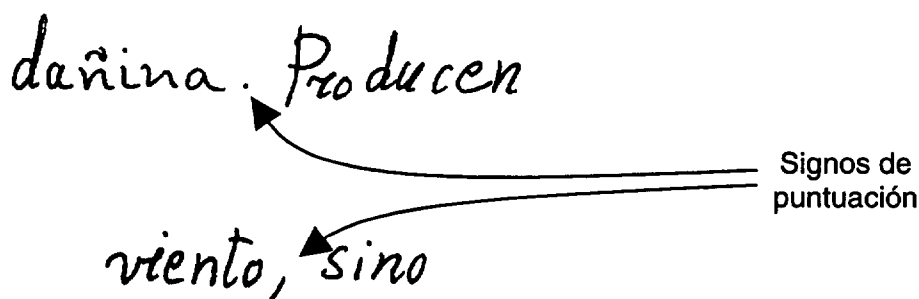


Figura 5.32: Detección de los signos de puntuación.

Dentro del grupo de signo de puntuación, se incluyen los puntos de la “i” y la “j” y los acentos. En principio esta información puede ser útil para un reconocimiento posterior de la letra o de la palabra en su conjunto, pero en esta etapa se va proceder a su

eliminación. En resumen, se van a definir cinco clase en total, tal y como se recoge en la siguiente tabla.

Clase	Descripción
Clase 1	Puntos de la i y j
Clase 2	Acentos y apóstrofe de la ñ
Clase 3	Comas
Clase 4	Punto fin de línea
Clase 5	Otras componentes (letras, ...)

Tabla 5.11: Distintos tipos de clases utilizadas.

La última clase engloba todas las componentes no incluidas en los grupos anteriores, es decir, en ella estarían incluidas las letras y palabras, o dicho de otro modo, lo que no se considera signo de puntuación. De esta manera, una componente conectada siempre deberá estar asociada a una clase.

Para clasificarlo se ha utilizado una red neuronal perceptrón multicapa, con el algoritmo de entrenamiento *Back-propagation*. Los resultados se muestran en la tabla 5.12.

	Clase 1	Clase 2	Clase 3	Clase 4	Clase 5	Total
Clase 1	96	4	0	0	0	100
Clase 2	3.1	96.9	0	0	0	100
Clase 3	0	0	100	0	0	100
Clase 4	0	0	0	100	0	100
Clase 5	0	0	0	1	99	100
Reconocimiento medio						97.3 %

Tabla 5.12: Tasas de reconocimiento de los signos de puntuación.

Si se engloban las clases por sus parecidos, se podría unir la clase 1 y clase 2; y la clase 3 y clase 4. De esta forma se puede volver a recalcular la tasa de reconocimiento, dada en la tabla 5.12, obteniendo una nueva tasa de reconocimiento del 99.8 %.

El siguiente paso, es el de mayor dificultad, ya que consiste en segmentar las palabras en letras. Para ello se han utilizado distintas técnicas de normalización para compensar la variabilidad de escritura entre distintas personas. Para fijar cuales han sido las características utilizadas para establecer los puntos de segmentación, se ha basado en la heurística y en la experimentación para determinarlos. En general, se han utilizado los lazos que pueda tener una palabra; así como del contorno de la misma se han usado los mínimos del contorno superior y los máximos del contorno inferior, zonas de cruces, concavidades y convexidades. Este conjunto de características provocará un corte vertical en el punto de segmentación, para el que se han utilizado seis ángulos fijos entorno a esta vertical para establecer los posibles cortes, que son de -20°, -10°, 0°, 10°, 20° y 30°. Este carácter asimétrico lo justifica el hecho de que la escritura suele sufrir una inclinación positiva o hacia la derecha.

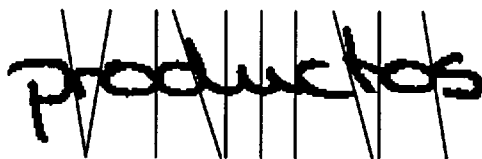


Figura 5.33: Aplicación del método de segmentación de palabras.

Este proceso puede producir una sobre segmentación, que es la primera parte del proceso de segmentación. Como segundo término se determina si la segmentación propuesta es correcta o no. Las separaciones o ligaduras entre caracteres se detectan utilizando histogramas de píxeles en estos seis ángulos mencionados. Con este sistema se obtiene una tasa de éxito en la segmentación del 94.64%, con un 2.94% de sobre segmentación y un 2.42% de letras no segmentadas.

Hay que indicar que la sobre segmentación no implica el error en el reconocimiento de una letra, sino un mayor coste en la posible determinación de la misma. Pero una “no segmentación” si implica un error en el reconocimiento de la letra.

Por último, se ha establecido un proceso de obtención de la palabra más probable, sobre la base de datos de palabras de la que consta el párrafo. Todos estos procesos están investigándose, y mucho más este último, ya que el uso de diccionarios semánticos y léxicos, puede mejorar el reconocimiento.

5.8 Conclusiones

Si se observa el estudio del reconocimiento de palabras manuscritas aisladas no es muy usual en la literatura actual, y más si estas palabras son minúsculas. En este capítulo se ha descrito este estudio, obteniendo una tasa de reconocimiento del 83.2% para diez palabras diferentes, y si se siguen los resultados de la tabla 5.7, se obtienen las diferentes tasas para los diversos agrupamientos, hasta completar para cincuenta y nueve palabras, un 52.9% de reconocimiento.

En particular para este ejemplo, hay que tener presente el gran parecido entre palabras, como al discriminar entre “cuatrocientos” y “cuatrocientas”, y como éste, hay muchos más ejemplos que se pueden observar en las plantillas de captación de muestras de este capítulo. Quizás si la naturaleza de las palabras fuese otra más incorrelada, la tasa de reconocimiento aumentase.

Si se observa la literatura actual sobre este tipo de trabajo, se puede llegar a dos conclusiones. La primera denota la diferencia de naturaleza de los clasificadores utilizados, mientras en esta tesis se utilizan los modelos ocultos de Markov, como clasificador de secuencias extraídas de las palabras, mediante parámetros que siguen el trazo de escritura y las mediciones del contorno; en la literatura se usan clasificadores basados en características más geométricas.

Y en segundo término, se observa que la diferencia entre las tasas de reconocimiento en el caso de no utilizar diccionario, son muy parecidas, algunas son mayores y otras menores, aunque no hay una prueba común o estándar, para poder establecer una comparación, ya que las palabras utilizadas no corresponden a las

palabras provenientes de cheques bancarios. La única excepción viene marcada por [Saon97], que utiliza 26 palabras de cheques bancarios, con la particularidad, de que son palabras de la legua francesa, y las palabras están bastante incorreladas entre sí. Pero en líneas generales, se puede establecer que es un método novedoso y de buenos resultados para estas aplicaciones, tal y como se observa en la siguiente tabla.

<i>Reconocedor de palabras manuscritas</i>	<i>Número de patrones</i>	<i>Tasas de reconocimiento</i>
[Busel94]	12	81%
[Choisy00]	25	83.1%
[Saon97]	26	85.7%
[Garris95]	38	60.5%
Esta tesis en media	10	83.19%
Esta tesis en media	20	75.63%
Esta tesis en media	30	61.91%
Esta tesis en media	40	57.57%
Esta tesis en media	50	57.06%
Esta tesis en media	59	52.92%

Tabla 5.13: Comparación de los sistemas reconocedores actuales para palabras manuscritas.

Para esta tesis se ha aplicado el reconocimiento de palabras aisladas proponiendo como mejor técnica, la extracción de parámetros de altos, anchos y transiciones del contorno de la palabra, habiendo corregido su inclinación. Para esta técnica se propone como clasificador un modelo oculto de Markov de 30 estados y con 32 símbolos por estado, con un multietiquetado de 5. Este modelo debe de entrenarse con el 60% de la base de datos propuesta. Con estos parámetros se obtienen resultados promediados en media del 83.2% para 10 palabras y 52.9% para 59 palabras.

En este capítulo también se proponen otras técnicas para el reconocimiento de cadenas manuscritas, como son las palabras o los códigos postales, a partir de la segmentación de cada uno de sus componentes, y el reconocimiento de los mismos, aplicado a un diccionario. Esta técnica es más usada en la literatura, pero igualmente puede ayudar a la mejora del reconocimiento.

Capítulo 6

Clasificación de manuscritos complejos: firmas

6.1 Introducción

Este capítulo trata sobre la adaptación de los parámetros para que sean clasificados mediante un modelo oculto de Markov, a partir de la información que se pueda extraer de la secuencia de una firma. Ya se han visto en las aplicaciones anteriores que la forma de proceder es común, y para las firmas también será así. Por tanto, se va a crear una base de datos para la extracción de esta información. Lo primero es procesarla y adecuarla para las técnicas que se le aplicarán, para resaltar la secuencia. Y de dicha secuencia se crearán los vectores para atacar a las funciones discriminantes. Finalmente, se pasa al proceso de reconocimiento, donde este clasificador debe ser adecuado a los vectores con esta información secuencial para optimizar la tasa de reconocimiento.

En la literatura, la herramienta más usada en el reconocimiento de las aplicaciones manuscritas, van dirigidas a la extracción de las características geométricas [Neubauer96] [Cho97] [Hwang97] [Seiler96], y en el reconocimiento de formas en general [Kaewkongka99] [Busel94]. En este capítulo de clasificación de firmas, se hará énfasis en la información de la secuencia contenidas en las firmas.

6.2 Base de datos de firmas manuscritas

La base de datos de firmas manuscrita puede estar formada por un universo relativamente ilimitado de patrones, ya que en el mundo hay más de seis mil millones de habitantes, y cada uno tiene su firma característica. Para usar esta aplicación, se ha limitado la base de datos a 60 patrones diferentes, lo que significa, que sólo se va a disponer de 60 escritores para su construcción [Camino99-pfc].

Para el caso de esta base de datos de firmas, se han tenido presente dos aspectos fundamentales en el diseño de la plantilla de recogida de muestras. El primero es la cantidad de muestras recogidas por cada escritor para conseguir un eficaz entrenamiento del sistema, y al mismo tiempo, dejar algunas muestras para la realización del testeo del sistema.

Y el segundo punto ha sido el tamaño seleccionado para que los escritores puedan dejar su firma de forma natural. Tampoco se puede dejar un espacio excesivo para cada firma, ya que se está limitado por una plantilla de dimensión DIN-A4. Si se indaga sobre el tamaño tipo que se suele destinar al espacio de las firmas, se puede llegar a la conclusión, de que en la mayoría de los casos, éste ocupa un tamaño rectangular, como se puede observar en los contratos de trabajo, documentos oficiales, documentos de identificación, o incluso tarjetas de crédito. Una muestra de ello se puede observar en los siguientes ejemplos gráficos:

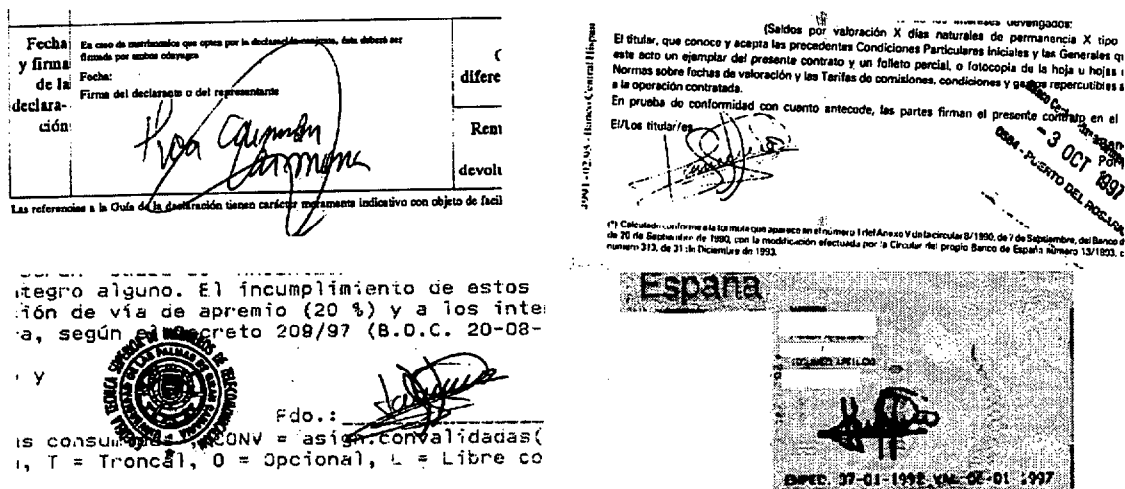


Figura 6.1: Ejemplos de recogida de firmas en documentos oficiales.

Por tanto, el tamaño de la cuadrícula que se ha escogido para depositar cada muestra de firma del escritor, tendrá forma rectangular, y se plantearán dos tamaños diferentes para generalizar la cantidad de tamaños y modelos diferentes que existen en la sociedad.

Como ya se señaló para la construcción de esta base de datos se han utilizado 60 escritores (es decir, 60 patrones), y el número final de muestras que se recogió por escritor fue de 24 repeticiones de firmas, por lo que se obtiene un total de 1440 (60×24) muestras.

6.2.1 Perfil de los escritores

Se puede extraer una serie de características de los escritores que conforman esta base de datos, rasgos recapitulados en la misma hoja de plantilla, y cuyo resumen se muestra en la siguiente tabla:

ESCRITOR	SEXO	MANO ESCRITURA	EDAD	ESCRITOR	SEXO	MANO ESCRITURA	EDAD
1	Hombre	Diestro	26	31	Hombre	Diestro	73
2	Hombre	Diestro	19	32	Mujer	Diestra	22
3	Mujer	Diestra	23	33	Mujer	Diestra	24
4	Hombre	Diestro	22	34	Mujer	Diestra	51
5	Hombre	Diestro	28	35	Mujer	Diestra	39
6	Mujer	Diestra	22	36	Mujer	Diestra	31
7	Mujer	Diestra	41	37	Mujer	Diestra	65
8	Mujer	Diestra	23	38	Mujer	Diestra	37
9	Mujer	Diestra	25	39	Mujer	Diestra	43
10	Hombre	Zurdo	25	40	Mujer	Zurda	14
11	Hombre	Diestro	26	41	Hombre	Diestro	24
12	Hombre	Diestro	22	42	Mujer	Diestra	39
13	Mujer	Diestra	14	43	Hombre	Diestro	43
14	Mujer	Diestra	38	44	Mujer	Diestra	13
15	Mujer	Diestra	25	45	Mujer	Diestra	33
16	Hombre	Diestro	27	46	Hombre	Diestro	24
17	Mujer	Diestra	24	47	Hombre	Diestro	36
18	Hombre	Diestro	22	48	Mujer	Diestra	31
19	Hombre	Diestro	50	49	Hombre	Diestro	37
20	Hombre	Diestro	27	50	Mujer	Diestra	24
21	Hombre	Diestro	28	51	Hombre	Diestro	20
22	Mujer	Diestra	23	52	Hombre	Diestro	22
23	Mujer	Diestra	31	53	Hombre	Diestro	25
24	Mujer	Diestra	33	54	Mujer	Diestra	37
25	Mujer	Diestra	27	55	Mujer	Diestra	24
26	Hombre	Diestro	30	56	Hombre	Diestro	41
27	Hombre	Diestro	29	57	Mujer	Diestra	42
28	Hombre	Zurdo	21	58	Mujer	Diestra	67
29	Hombre	Diestro	59	59	Mujer	Diestra	68
30	Hombre	Zurdo	18	60	Mujer	Diestra	74

Tabla 6.1: Muestrario de los escritores de la base de datos de firmas.

De todos los escritores recogidos en esta base de datos, se puede establecer que:

- las edades de los mismos están comprendidas entre los 13 y los 74 años
- hay ligeramente más escritores femeninos que masculinos
- en general abundan los escritores diestros frente a los zurdos.

Materializando estas observaciones a valores estadísticos, se pueden obtener las siguientes figuras:

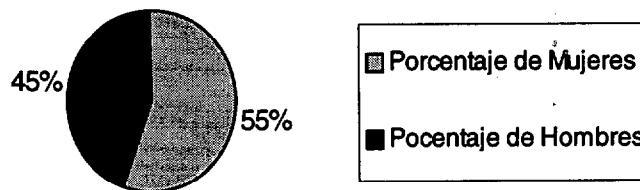


Figura 6.2: Porcentaje de Hombres vs. Mujeres de la base de datos de firmas.

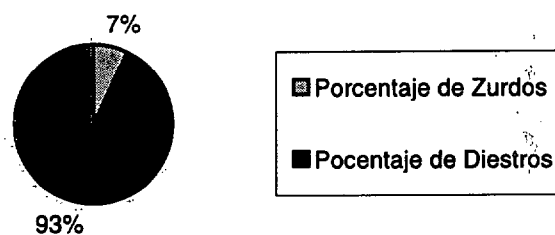


Figura 6.3: Porcentajes de Diestros vs. Zurdos de la base de datos de firmas.

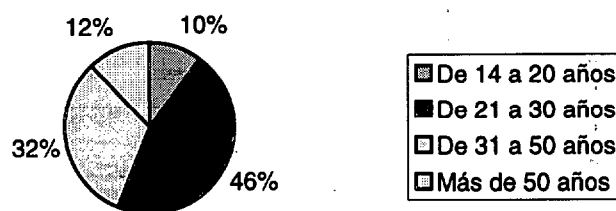


Figura 6.4: Distribución por edades en la base de datos de firmas.

6.2.2 Formato de la plantilla

La plantilla utilizada para la captura de las muestras de los escritores firmantes, consta de las tres partes, al igual que en las aplicaciones de los anteriores capítulos. La primera son las líneas verticales y horizontales para la estimación de la inclinación. La segunda son los tres cajetines que aparecen en la parte inferior de la hoja de la plantilla, que se utilizan para recoger la información estadística de los escritores de la base de datos. Y por último, la rejilla que ocupa casi toda la plantilla y que se sitúa en la zona central de la misma. Esta rejilla toma dos formatos diferentes, y ambos son rectangulares, por los formatos usados más comúnmente en la práctica. Proponiendo esta variedad de tamaños para tener acceso a muestras bajo esta distinta circunstancia, ya que en la práctica no existen formatos estándares de recogida de firmas. Se han dispuesto de las 24 repeticiones, 12 para un tipo de cuadrícula rectangular de 2.5×5.5 centímetros y las 12 restantes en una cuadrícula de 3×5 centímetros. Lógicamente, cumpliendo el requisito

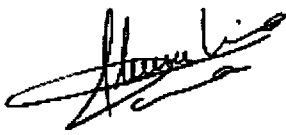


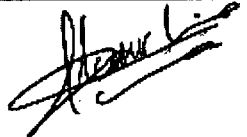

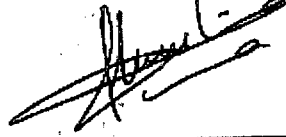


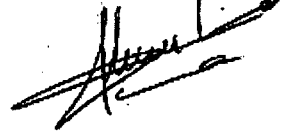
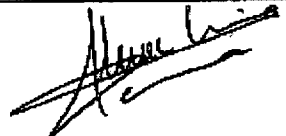

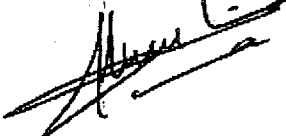
de poder realizar al trazo de la firma de forma natural. En las siguientes figuras se pueden observar la plantilla y un ejemplo de ésta, rellena.

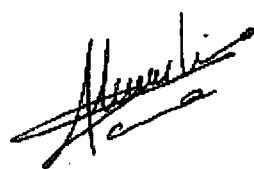
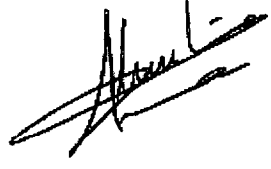
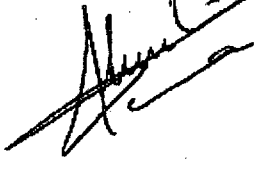
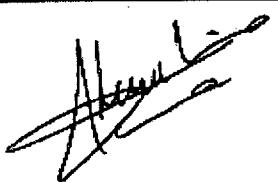
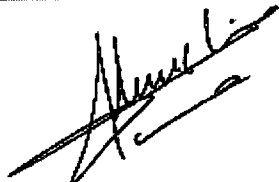
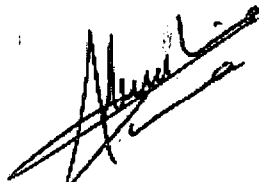
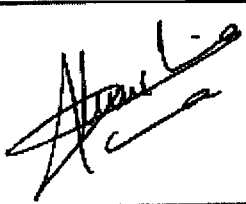
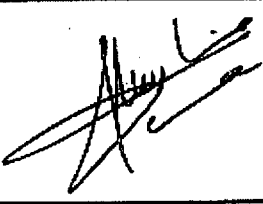
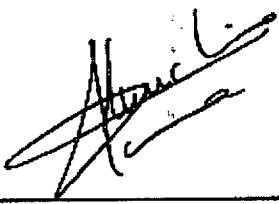

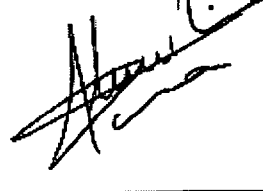
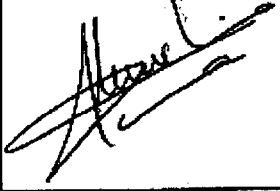
Varón/Mujer	
-------------	--

EDAD:	
-------	--

Zurdo/Diestro	
---------------	--

Figura 6.5: Plantilla de la base de datos de firmas.

Varón / Mujer	<input checked="" type="checkbox"/> Y	EDAD:	<input type="text" value="19"/>	Zurdo / Diestro	<input checked="" type="checkbox"/> D
---------------	---------------------------------------	-------	---------------------------------	-----------------	---------------------------------------

Figura 6.6: Plantilla de la base de datos de firmas rellena.

También se ilustran, en la siguiente figura, el total de las muestras de cada una de las firmas que se han utilizado para construir esta base de datos.





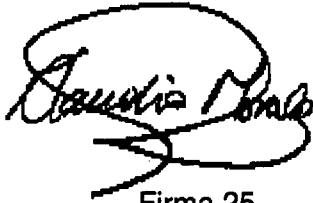
Firma 22



Firma 23



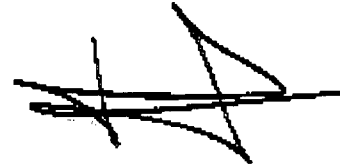
Firma 24



Firma 25



Firma 26



Firma 27



Firma 28



Firma 29



Firma 30



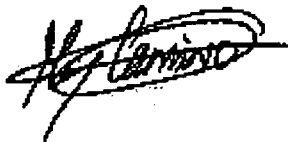
Firma 31



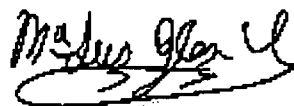
Firma 32



Firma 33



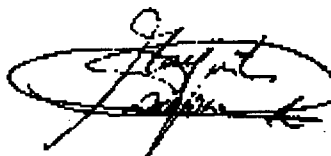
Firma 34



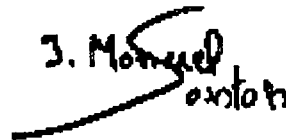
Firma 35



Firma 36



Firma 37



J. Monreal
Santón

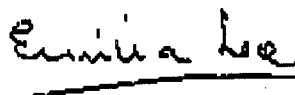
Firma 38



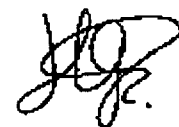
Firma 39



Firma 40



Firma 41



Firma 42

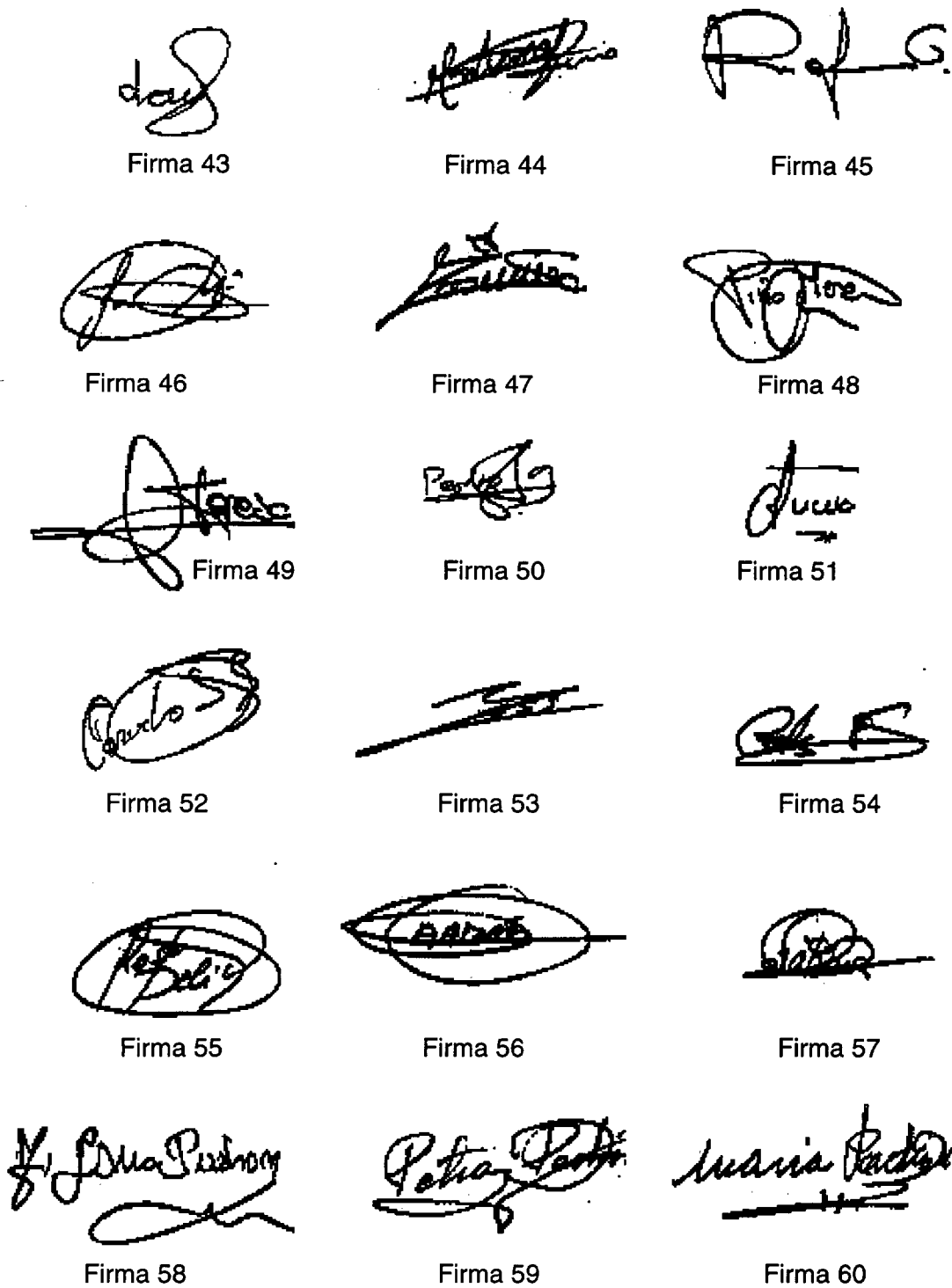


Figura 6.7: Muestrario de las firmas de las consta la base de datos.

6.2.3 Dificultades en la creación de las bases de datos

Las dificultades que se han presentado en la elaboración de esta base de datos, además de las descritas en el capítulo tercero, apartado 3.2.5; se encontró alguna curiosidad más, y fue la desconfianza de algunos escritores en mostrar su firma auténtica, por miedo a falsificaciones, lo que motivó que éstas las realizaran con la escritura de su nombre y primer apellido, aunque fueron casos excepcionales [Camino99-pfc].

Quizás el hecho de no tener una compensación, ha propiciado esta desavenencia, pero finalmente la buena voluntad de todos ha proporcionado la construcción de esta base de datos.

6.3 Preprocesado de imagen de la base de datos

El preprocesado de datos aplicado a las firmas ha sido el mismo que el realizado a las letras, a los dígitos y a las palabras manuscritas, descrito en el apartado 3.3 del tercer capítulo y en el 5.3 del capítulo quinto, respectivamente.

Este proceso consta de los siguientes pasos;

- Extracción de muestras de la plantilla de firmas
- Binarización de las muestras
- Eliminación del ruido
- Normalización y reducción de las firmas

6.4 Parametrización de los patrones manuscritos complejos

Una vez que las firmas ya han pasado por el preprocesado correspondiente, ya se les puede aplicar la extracción de parámetros, estas técnicas que se van a utilizar para tal efecto son:

- Técnica de envolvente vectorizada por agrupación de regiones
- Técnica de envolvente vectorizada por polares
- Técnica de envolvente vectorizada su secuencia
- Técnica de envolvente vectorizada por altos, anchos y transiciones
- Técnica de esqueletización, vectorizada por su secuencia
- Técnica del gradiente para NN

Algunas de ellas ya han sido utilizadas con anterioridad, aplicadas a las anteriores aplicaciones manuscritas, en particular, la única técnica novedosa es la agrupación de regiones a partir de la envolvente, que se describirá a continuación.

Otra peculiaridad añadida, es el uso de la envolvente, que no del contorno, aunque proviene a partir de éste. El cálculo de la envolvente puede tener diversa naturaleza, pero finalmente se ha optado por el uso de la morfología matemática para su obtención.

Si se parte de una imagen, ya se ha visto en múltiples ocasiones en esta tesis como se llega a calcular su contorno. En el caso de las firmas ocurre lo mismo (véase la

figura 6.8), y obsérvese el sencillo proceso que se aplica para obtener su envolvente, según la figura 6.9.

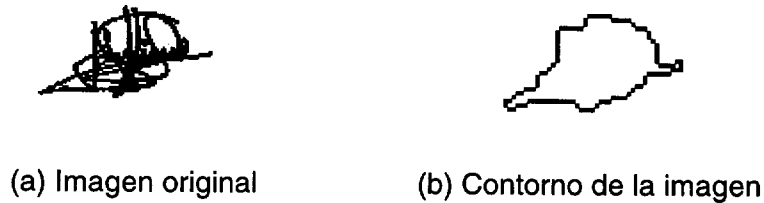


Figura 6.8: Imagen de una firma y su contorno.

Otra variable presente en este estudio es el hecho de disminuir el tamaño de la firma para el cálculo de su envolvente, por un lado se pueden observar ventajas desde el punto de vista del coste computacional, ya que la imagen es menor. Pero por otro lado, como cada escritor sólo realizar un patrón, el tamaño puede ser un elemento discriminante para su posterior reconocimiento. La respuesta final, la tendrá el clasificador que dará los mejores resultados, para firmas reducidas o sin reducir.

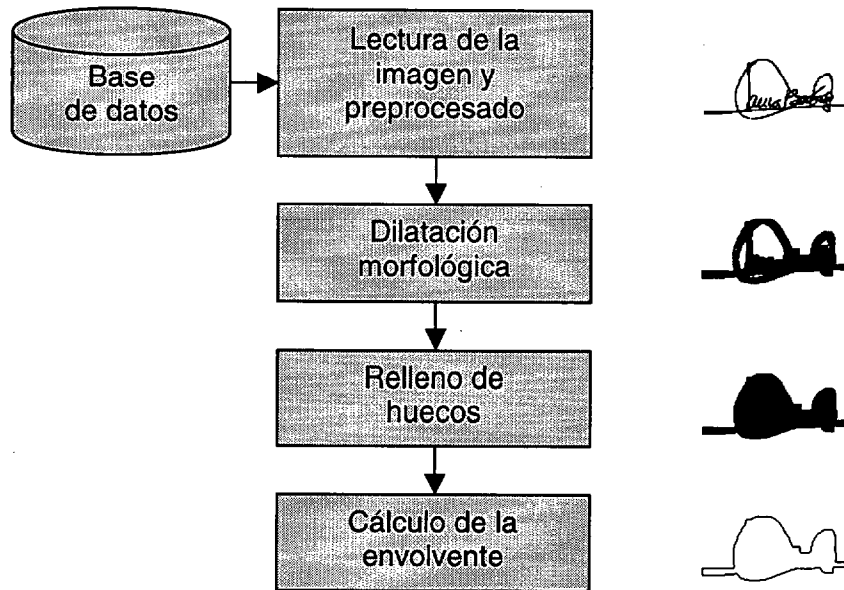


Figura 6.9: Proceso del cálculo de la envolvente.

Básicamente la envolvente, es lo mismo que el contorno exterior, salvo que se ha dilatado. Pero se puede presentar el caso que una firma dilatada, que presente varias componentes o regiones. En ese caso, se procederá a unir dichas regiones, tal y como se observa en la figura 6.10, hasta conseguir una única envolvente. Para ello, se introduce un módulo de dilatación horizontal, y se aplica dicho módulo hasta que la firma es un solo elemento.

Una vez ya se tiene la envolvente, se le puede aplicar las diferentes técnicas para extraer los parámetros, y construir así los vectores. Por último y para la técnica del gradiente [Cho97], se aplican los conceptos ya vistos en capítulos anteriores y explicados en el anexo E, concerniente a la información geométrica en las firmas.

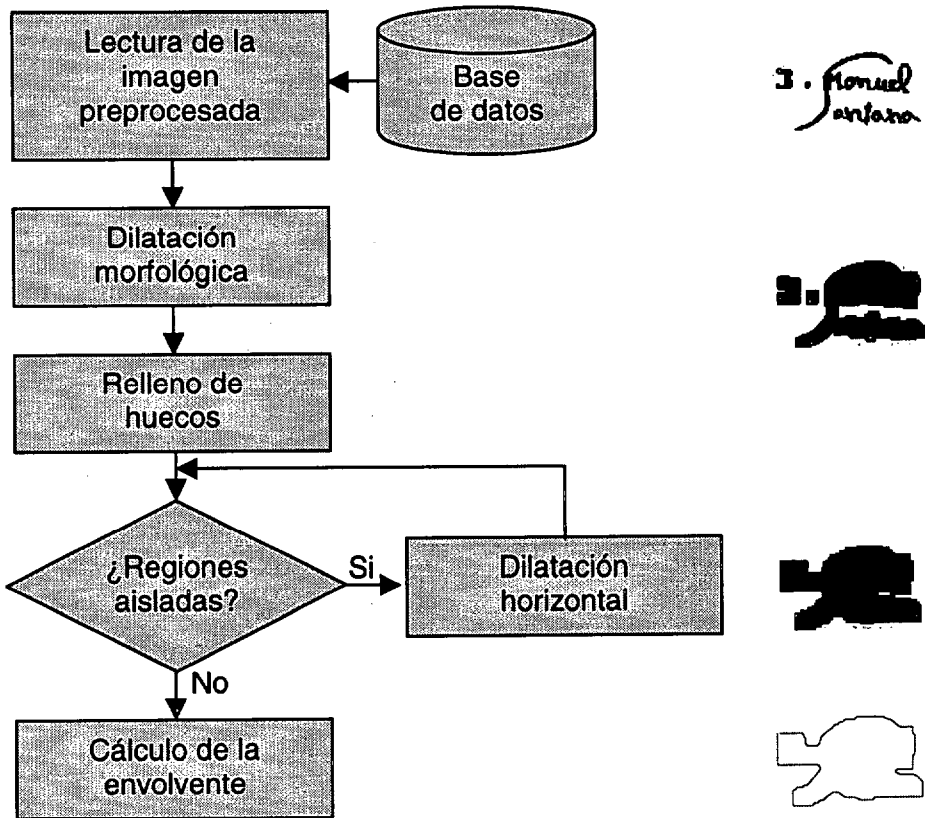


Figura 6.10: Proceso del cálculo de la envolvente con varias regiones aisladas.

6.5 Vectorización de los patrones manuscritos complejos

La primera técnica descrita en el apartado anterior es la vectorización de la envolvente por agrupación de regiones. Esta técnica se basa en el concepto de centroide, y consiste en agrupar la firma en regiones según un criterio de conexiones de los vecinos horizontales y verticales a ambos lados, y posteriormente se calcula un punto característico del interior de la misma, que será dicho centroide. Éste se calcula promediando la localización de los píxeles en la región. Se determinarán tantas regiones como el criterio de conexión refleje, y la extracción de los parámetros a un vector se realiza con las coordenadas cartesianas de los centroides. En la siguiente figura se observa un ejemplo del cálculo de los centroides:



Figura 6.11: Determinación de los centroides a partir de su envolvente.

La segunda técnica para obtener vectores es una ya estudiada, el seguimiento de la envolvente en coordenadas polares. Como se ha visto en las aplicaciones de dígitos, letras y palabras manuscritas, los mejores resultados se encuentran si se utiliza como sistema de referencia el centro de masas de las misma, y por tanto, se desecha el sistema de referencia de inferior izquierda, por extensión de los cálculos de las tasas de reconocimiento recopiladas. Un ejemplo de esta técnica para las firmas, se puede observar en la siguiente figura.

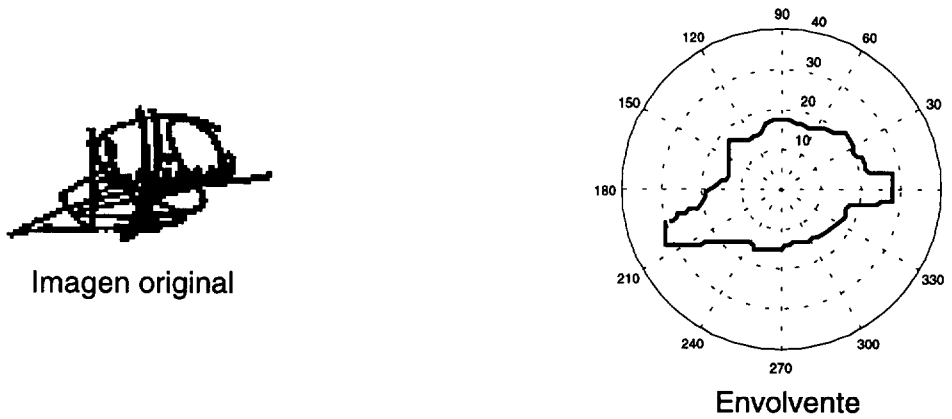


Figura 6.12: Técnica de vectorización por coordenadas polares de la envolvente.

En esta vectorización con coordenadas polares, se ha introducido un nuevo matiz, y es el número de píxeles a negro existentes desde el centro de masas de la firma hasta el extremo de la envolvente. Como ya se estudiará, este nuevo parámetro mejorará los resultados, y se ilustra un ejemplo de éste en la siguiente figura;

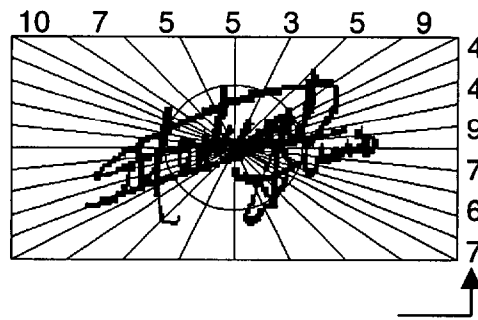


Figura 6.13: Número de píxeles a negros por ángulo para las coordenadas polares.

La siguiente técnica descrita en el cálculo de parámetros para firmas, es la vectorización de la envolvente por su secuencia, al utilizar la matriz de 3x3 de saltos, que registra las variaciones que sufre en su dirección la secuencia.



Figura 6.14: Extracción del vector de la secuencia de la envolvente de una firma.

Otra técnica usada para calcular parámetros de vectores son los altos, anchos y transiciones de una firma. El procedimiento es el mismo al empleado en las palabras manuscritas, y un ejemplo con firmas se puede observar en la siguiente figura;

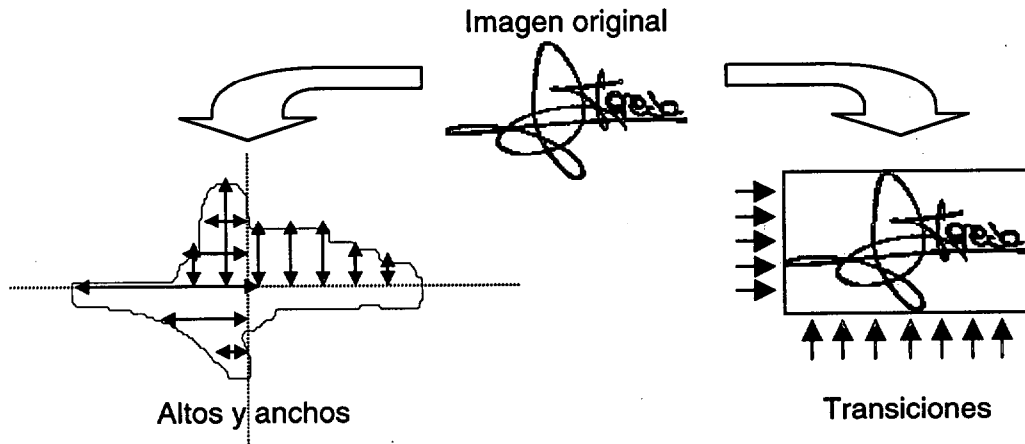


Figura 6.15: Cálculo de los altos, anchos y transiciones de una firma.

Otra técnica estudiada para las firmas, y con antelación para las aplicaciones manuscritas más simples, es la de esqueletización. La forma de obtener el adelgazamiento es la misma, no varía en nada, y los resultados de su aplicación, se ilustran en la siguiente figura;

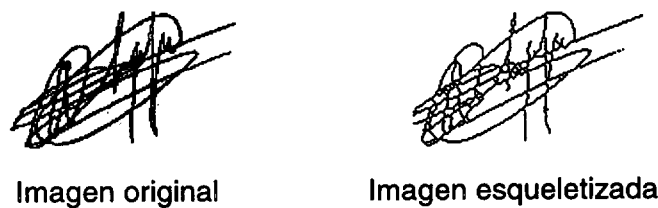


Figura 6.16: Firma original y esqueletizada.

Y por último queda la técnica del gradiente, parece acertado discutir este tipo de técnica ya que en la mayoría de los casos estudiados en el estado del arte del segundo capítulo, sobre firmas en un sistema *off-line*, se utiliza información geométrica y no tanto de la envolvente. Esta discusión puede llegar a establecer alguna buena conclusión en el posterior estudio.

La técnica del gradiente, cuyo algoritmo se ha aplicado en las aplicaciones anteriores, devuelve parámetros geométricos sobre las variaciones de la inclinación de la firma, en la horizontal, la vertical, el lado oblicuo derecho y el lado oblicuo izquierdo. Finalmente, también se incluye una reducción de firma para incluir como parámetro. El desarrollo de este algoritmo se encuentra en el anexo E, y un ejemplo del mismo queda ilustrado en la figura 6.17.

Una vez conocidas todas las técnicas empleadas para la extracción de parámetros de las firmas, ya sólo queda definir el clasificador que se va a utilizar y la descripción de los experimentos planteados, para averiguar cual es la técnica más discriminante.

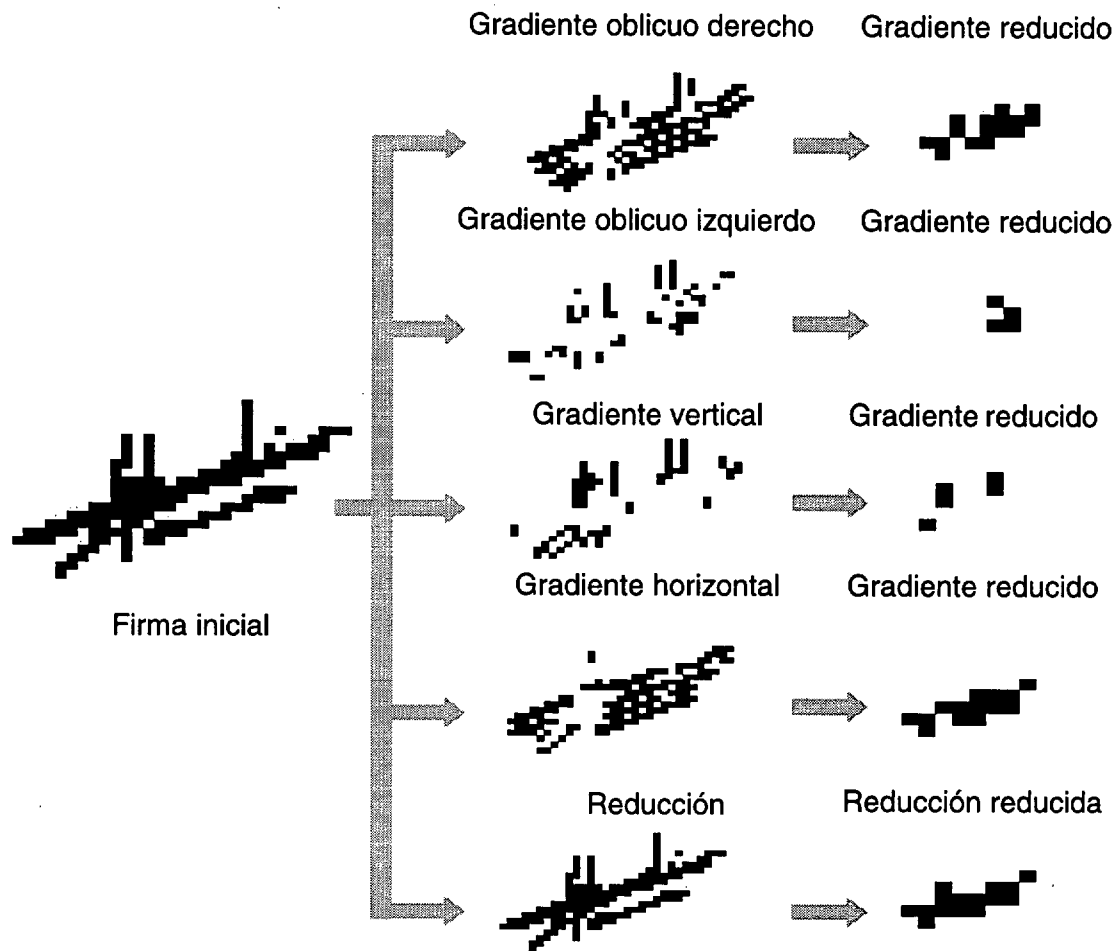


Figura 6.17: Aplicación de la técnica del gradiente en las firmas.

6.6 Clasificación de las firmas manuscritas

Se han planteado dos tipos de vectores de información diferentes, uno que proviene de información de la secuencia y los otros de la geométrica. Estos clasificadores ya fueron usados con las aplicaciones manuscritas anteriores. Para modelar las fuentes secuenciales, se utilizarán los modelos ocultos de Markov (HMM) [Rabiner93], y para las fuentes geométricas redes neuronales (NN) [Cho97][Lee96].

En ambos casos la clasificación será supervisada, y constará de una etapa de entrenamiento y otra de test. En la etapa de entrenamiento se optimizarán los parámetros del clasificador. Para el HMM serán el número de estados, el número de símbolos de observaciones por estado y la salida del algoritmo por umbral, utilizando un sistema BAKIS. Para NN serán el número de neuronas de la capa oculta y el coeficiente de aprendizaje inicial.

Al igual que en la aplicación de letras, para las técnicas de coordenadas polares y de altos, anchos y transiciones; el número de símbolos de observaciones por estado se ha fijado en 32, ya que se ha calculado de forma experimental como un óptimo valor [Gómez01-pfc].

La etapa de test devolverá los mejores resultados para las técnicas empleadas, basados en las probabilidades calculadas por los clasificadores. De esta forma, se averiguará cual de los anteriores métodos es más discriminante.

6.7 Resultados: optimización de los parámetros del clasificador

Las firmas manuscritas forman parte de la identificación que puede hacerse a una persona. La base de datos de firmas manuscritas estaría formada por cada una de las personas que viven en este mundo, pero en esta aplicación se va a ceñir a la descrita en este capítulo. El caso de las firmas tiene la peculiaridad de que siempre deben de crearse los modelos dependientes de escritor, ya que la firma es personal para cada escritor, sino se estaría hablando de una falsificación. Por tanto, estos modelos siempre serán dependientes del propio escritor, unas muestras se usarán para la etapa de entrenamiento y otra parte de las muestras del mismo escritor para la etapa de evaluación.

Otro dato importante a tener presente en la aplicación de las firmas es la doble clasificación que ofrece, el reconocimiento y la verificación. Mientras el reconocimiento identifica a un firmante entre un cierto número de patrones de firmas, la verificación identifica una firma como la de un cierto escritor, siendo conocida la identidad del firmante y su firma. Este capítulo centra su estudio en el reconocimiento, aunque se está investigando la verificación y se han llegado a implementar algunos casos.

Las pruebas efectuadas para las firmas, se realizaron con la base de datos de firmas detallada en el apartado 6.2, con un total de 60 firmantes y 24 repetición por cada escritor, recogiendo todo el estudio de la optimización de los clasificadores para el conjunto de las pruebas planteadas en el anexo D.

En esta aplicación de firmas las pruebas que se han realizado son las mismas planteadas en el apartado 6.4. En estas pruebas se ha realizado la optimización para el HMM con detección de la envolvente. Pero el primer paso es ver si la reducción de la envolvente mantiene mayor discriminación que el tamaño original de la misma, ya que las firmas son únicas por cada escritor, y el tamaño en esta aplicación, puede llegar a ser una forma de diferenciarse.

Para este experimento se ha trabajado una técnica muy utilizada a lo largo de esta tesis, como es la detección por la secuencia de la envolvente, y se ha llevado a la práctica con 20 firmas de la base de datos, dando como resultado:

<i>Tamaño de la envolvente al detectarla por la secuencia</i>	<i>Número de estados</i>	<i>Número de pruebas en el promedio</i>	<i>Tasa de reconocimiento</i>
Tamaño reducido	18	5	79.08% ± 2.26
Tamaño original	35	5	82.25% ± 1.57

Tabla 6.2: Estudio sobre el tamaño de la envolvente para 20 firmas.

Tras promediar el experimento en 5 ocasiones, se puede deducir de esta tabla, que el tamaño reducido puede ofrecer una disminución del coste computacional pero no un mayor éxito que el tamaño original de la envolvente.

Por tanto, a continuación sólo se van a utilizar las firmas sin comprimir su tamaño, y se va a experimentar para las diferentes técnicas mencionadas anteriormente. El siguiente factor a distinguir es la amplia gama de posibilidades de vectorización a partir de la envolvente que se proponen, y se extienden con mayor detalle en el anexo D. Para este experimento se ha utilizado el total de la base de datos, es decir, las 60 firmas de las que consta, para evaluar cual de los métodos es más discriminante. Con respecto a la envolvente los métodos aplicados han devuelto los siguiente resultados, recogidos en estas dos tablas.

<i>Técnica de extracción de parámetros</i>	<i>Número de estados</i>	<i>Pruebas de promediado</i>	<i>Tasa de reconocimiento</i>
Envolvente por agrupación de regiones	25	8	69.93% ± 1.82
Envolvente por coordenadas polares	35	8	85.44% ± 0.70
Envolvente por polares y píxeles negros	35	10	92.52% ± 2.01
Envolvente por secuencia	35	5	74.94% ± 2.06
Envolvente por altos, anchos y transiciones	35	10	92.76% ± 1.67
Envolvente por esqueletización	35	1	67.50% ± 0.00

Tabla 6.3: Tasas de reconocimiento para optimizar el clasificador HMM con la técnica del contorno para firmas manuscritas.

<i>Técnica de extracción de parámetros</i>	<i>Número de neuronas</i>	<i>Pruebas de promediado</i>	<i>Tasa de reconocimiento</i>
Técnica del gradiente por NN	135	1	72.60%

Tabla 6.4: Tasa de reconocimiento para optimizar el clasificador NN con la técnica del gradiente para firmas manuscritas.

Se puede extraer de estas tablas que la diferencia entre las tasas de resultados es significativa, teniendo valores de hasta el 92% de éxito para la parametrización de la envolvente con coordenadas polares en conjunto con los píxeles a negro, y a altos, anchos y transiciones.

Por otro lado, el resto de las técnicas aplicadas a la envolvente, están bastante por debajo de las tasas máximas, la agrupación de regiones en la envolvente [Sánchez01], la secuencia y la esquelización registran entre sí, aproximadamente los mismos órdenes de reconocimiento, empeorando en más de quince puntos.

Respecto a la información geométrica, la técnica del gradiente es inferior a las máximas encontradas con información de la envolvente. El enfrentamiento de estas dos informaciones puedan dar un reflejo de la importancia de la extracción de información de la envolvente para estas aplicaciones manuscritas, tal y como se muestra en la tabla 6.5. Finalmente, la peor de las técnicas queda reflejada en la esqueletización de la firma, con 25 puntos por debajo del máximo valor.

<i>Tipo de Información</i>	<i>Tasa de Reconocimiento</i>
Geométrica	72.60%
Secuencial	92.76%

Tabla 6.5: Tasas de reconocimiento según el tipo de información para firmas.

Por último, se han implementado, algunos sistemas híbridos, compuestos por las técnicas mencionadas con anterioridad, para lograr aumentar la tasa de reconocimiento. Se han planteado de las posibles combinaciones dos casos, ya que experimentalmente, fueron las que mejores resultados dieron. En particular, estos dos experimentos hechos son:

- La combinación de la envolvente por agrupación de regiones con coordenadas polares.
- La combinación de la envolvente por altos, anchos y transiciones con coordenadas polares y píxeles a negro.

En el primer caso, la arquitectura implementada no es más que una combinación lineal de ambas técnicas, como se ilustra en la siguiente figura:

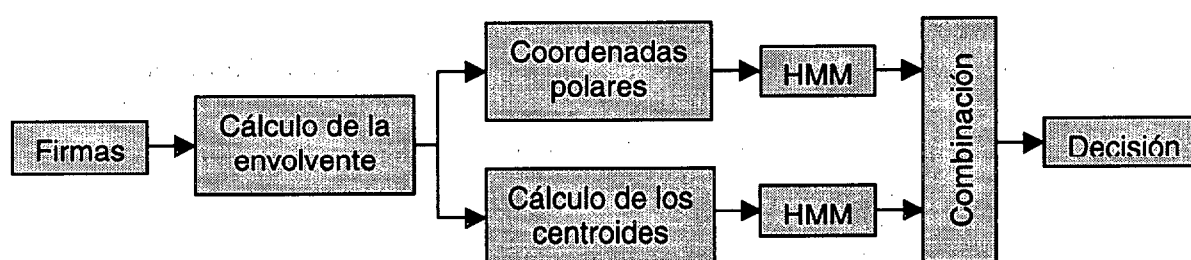


Figura 6.18: Combinación lineal de centroides y coordenadas polares de la envolvente para reconocer firmas manuscritas.

Con este sistema las tasas de reconocimiento han aumentado, tal y como se presentan en la siguiente tabla.

Número de prueba	Envolvente por coordenadas polares	Envolvente por centroides	Sistema lineal combinado
I	85.00%	71.67%	89.07%
II	86.48%	71.67%	91.11%
III	85.74%	71.30%	92.04%
IV	85.93%	67.22%	88.89%
V	85.00%	72.22%	90.19%
VI	85.93%	67.22%	89.07%
VII	85.56%	69.63%	88.52%
VIII	83.89%	68.52%	89.81%
Media ± Varianza	85.44% ± 0.70	69.93% ± 1.82	89.84% ± 1.07

Tabla 6.6: Tasas de reconocimiento del sistema combinado de polares y centroides.

Se ha logrado una tasa del 89.84% de éxito para reconocer firmas, lo cual no llega a superar a los métodos presentados de forma aislada, pero nos da la idea que combinando distintas fuentes de información que provienen de la envolvente, se mejora la clasificación.

El siguiente sistema combinado que se presenta, surge a partir del método de altos, anchos y transiciones para la envolvente y de las coordenadas polares en conjunto

con los píxeles a negro, también para la envolvente. La arquitectura de combinación utilizada es exactamente igual a la reflejada en la figura 6.18, con la salvedad de que está especificada para estos métodos ya mencionados. A continuación se muestra dicho diagrama de combinación en la siguiente figura;

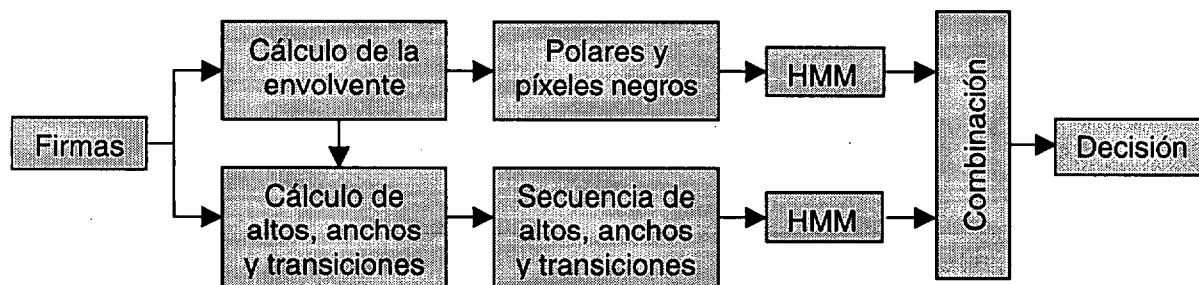


Figura 6.19: Combinación lineal de altos, anchos y transiciones con coordenadas polares y píxeles negros de la envolvente para reconocer firmas manuscritas.

Los datos de esta arquitectura se han recogidos en la siguiente tabla:

Diez pruebas promediadas	Envolvente por coordenadas polares	Envolvente por centroides	Sistema lineal combinado
Media ± Varianza	92.52% ± 2.01	92.76% ± 1.67	95.96% ± 0.60

Tabla 6.7: Tasas de reconocimiento del sistema combinado de altos, anchos y transiciones con coordenadas polares y píxeles negros.

Si se observa este resultado con respecto a la tabla 6.3, se superan todos los resultados de los clasificadores de forma aislada en más de tres puntos. Por tanto, la información de la envolvente al combinarla mejora el éxito y en este caso reduce la varianza, presentando un resultado más estable.

Aún se puede optimizar más los parámetros del clasificador, jugando con el multietiquetado, es decir, el número de etiquetas por símbolo que se puede tener para cada parámetro. Se ha realizado con esta última combinación, ya que presenta el mejor resultado y la mayor estabilidad, se ha variado este parámetro, obteniendo las siguientes tasas de reconocimiento:

Etiquetado	Número de pruebas realizadas	Parámetros del reconocedor		Sistema lineal combinado
		Polares y píxeles negros	Altos, anchos y transiciones	
1	10	92.52% ± 2.01	92.73% ± 1.67	95.96% ± 0.60
2	10	94.52% ± 1.33	95.01% ± 1.95	96.98% ± 0.69
3	10	95.44% ± 0.63	95.46% ± 2.19	97.72% ± 0.41
4	10	95.91% ± 0.61	95.65% ± 0.41	98.28% ± 0.35
5	2	96.94% ± 0.15	95.37% ± 0.62	98.52% ± 0.00
7	6	96.42% ± 0.32	95.03% ± 1.02	98.15% ± 0.20
10	10	96.89% ± 0.07	95.04% ± 1.10	98.75% ± 0.15
32	1	97.04%	96.11%	98.33%

Tabla 6.8: Tasas de reconocimiento del sistema combinado de altos, anchos y transiciones con coordenadas polares y píxeles negros variando el etiquetado.

De esta última tabla, se puede establecer que aún se mejora la tasa de reconocimiento para reconocer firmas, hasta llegar a un 98.5%. Por lo que se puede volver a la misma conclusión dictada con anterioridad, la supremacía del clasificador HMM aplicándole las técnicas de la envolvente.

6.8 Verificación en las firmas manuscritas

Otra tendencia del estudio y la investigación, es la verificación de firmas, un campo que está en proceso de investigación por parte de esta tesis, pero de los que se puede extraer algunos datos. Por ahora, sólo se ha planteado para algunas de las técnicas comentadas con anterioridad, en particular para la técnica de altos, anchos y transiciones, y para la envolvente de la firma con el número de píxeles negros.

En la figura 6.20, se puede apreciar la verificación en régimen de falsa aceptación, es decir, a una firma que no corresponde un escritor se la asigna como tal; y en régimen de falso rechazo, definido como una firma que es de un cierto escritor y no la reconoce como tal.

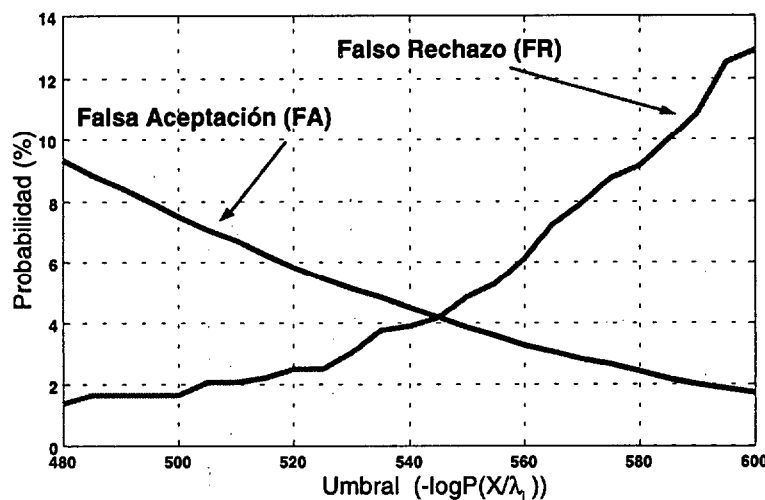


Figura 6.20: Curvas de error en verificación.

De esta figura, se puede concluir que teniendo en cuenta la variación del umbral de la probabilidad asignada aún cierto modelo de firma, si se disminuye el umbral de probabilidad, aumenta el falso rechazo a medida que disminuye la falsa aceptación. Y si se aumenta dicho umbral, justamente ocurre el efecto contrario. En función de cual sea la aplicación se ejecutará de una manera u otra.

6.9 Conclusiones

En referencia a la literatura actual [Deng98][Rigoll98][Jain01], los datos alcanzados llegan a ser comparables y hasta superiores, que los datos que se pueden encontrar en el estado del arte de las firmas manuscritas actual. Con respecto a esta tesis, los datos vuelven a desprender la misma conclusión, que a la llegada en los capítulos anteriores, que la información procedente de la envolvente es útil para reconocer y verificar las firmas manuscritas, y que tiene mejor comportamiento la información secuencial que la geométrica, con estas técnicas estudiadas (véase el anexo D), y comparables a la

literatura actual. En la siguiente tabla se van a establecer los valores recogidos durante el estudio de la revisión del estado de la técnica frente a los alcanzados en esta tesis.

<i>Reconocedor de firmas</i>	<i>Sistema</i>	<i>Número de firmas</i>	<i>Tasas de reconocimiento</i>
[Rigoll98]	Off-line	14	98.1%
[Rigoll98]	On-line	14	99%
[Deng98]	Off-line	25	91.71%
[Jain01]	On-line	59	97%
[Jain01]	On-line	102	98%
[Jain01]	On-line	43	95%
Esta tesis en media	Off-line	60	98.52 %

Tabla 6.9: Comparación de los sistemas reconocedores actuales para firmas.

La conclusión resulta clara, en esta tesis se consiguen mejores resultados de cuanto al reconocimiento, y en cuanto al número de firmas a reconocer. Ambos factores en conjunto hacen que los sistemas con información de la envolvente propuestos en este capítulo sobrepasen a los sistemas estudiados en la literatura actual, algunos de sistema *on-line* y otros de sistema *off-line*, con información geométrica.

Capítulo 7

Conclusiones y líneas futura de trabajo

7.1 Introducción

Este último capítulo resume las conclusiones más importantes obtenidas en esta tesis. Estas conclusiones van a compactar todo el planteamiento realizado en el primer capítulo, analizando los diferentes métodos y técnicas estudiadas sobre los resultados de las tasas de reconocimiento calculadas.

Para finalizar este capítulo, se propondrán cuales son los puntos fundamentales sobre los que se pueden incidir, y cuales son los posibles nuevos puntos a introducir, para usar las ventajas que plantea el HMM y su combinación, ya sea con diferentes técnicas aplicadas al mismo HMM o con otro tipo de adquisición de información.

7.2 Conclusiones

Antes de pasar a describir las conclusiones, se va a repasar el objetivo de la tesis, comprobando así, el cumplimiento de la idea trazada sobre la investigación que se ha desarrollado. Según se puede extraer en cita del primer capítulo;

“El principal objetivo de esta tesis es estudiar el alcance los modelos ocultos de Markov para modelar patrones de formas manuscritas para el reconocimiento de los mismos.”

Si se remite al apartado de resultados de los capítulos cuarto, quinto y sexto de esta tesis, se puede comprobar de forma cuantitativa, que realmente se cumple el objetivo principal de esta tesis, ya que mediante el contorno, y la información secuencial que se encierra en su trazo, se puede reconocer en las aplicaciones manuscritas.

En primer lugar se puede observar de forma clara que la extracción de parámetros a partir del contorno produce muy buenas tasas de reconocimiento en todas las aplicaciones descritas, comparándola con la literatura actual; y por tanto, se debe de considerar a la información proveniente del contorno como muy útil para el reconocimiento de las aplicaciones manuscritas, en general. Como se puede observar en la figura 7.1, donde se comparan los resultados obtenidos en esta tesis, con algunos de la literatura, se podría decir que a medida que se complica la aplicación manuscrita, mejores resultados se obtienen. En los dígitos funciona algo mejor los resultados recabados en la literatura, y para letras y palabras se mantienen, y en el caso de las firmas mejora con respecto a la literatura.

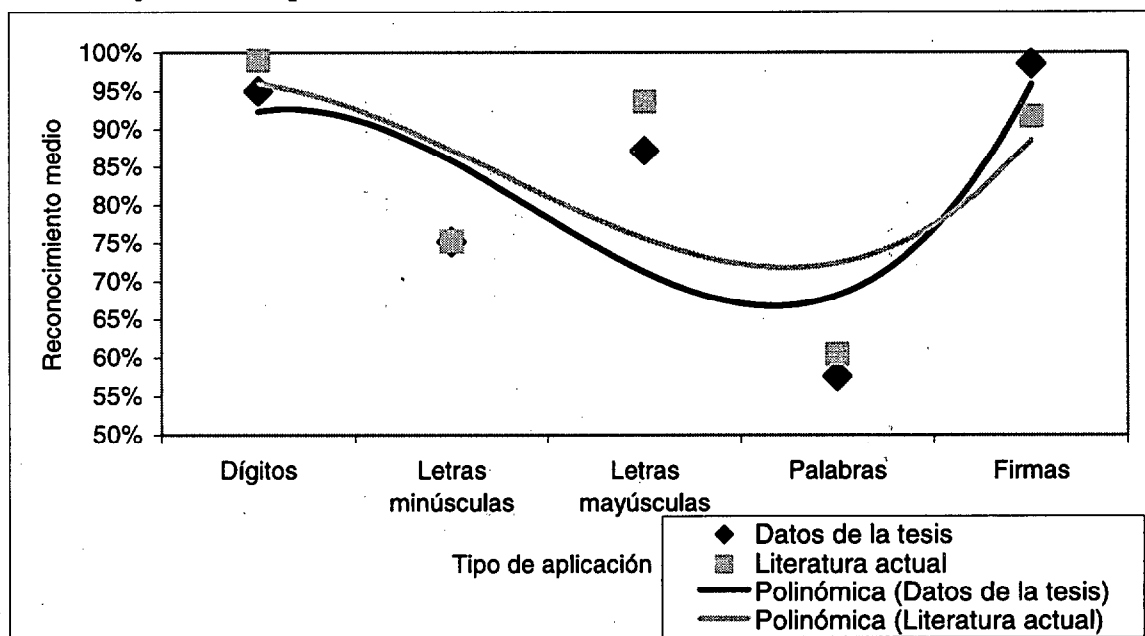


Figura 7.1: Comparativa en las aplicaciones manuscritas según su tasa de reconocimiento.

Por tanto, el uso del contorno para clasificar, es una técnica que se puede extender a muchas aplicaciones manuscritas, y mejor resultado se obtiene, cuanto más compleja es la aplicación a reconocer.

También se puede concluir que la información geométrica es muy utilizada en la práctica y tan efectiva como la información secuencial, pero a medida que se complica la aplicación o se aumenta el número de patrones a discriminar, baja su eficacia. Pero lo importante es que la combinación de los clasificadores a partir de distintos métodos de extracción de parámetros, siempre y en todos los casos estudiados, mejora las tasas de reconocimiento, debido a la incorrelación de los métodos empleados y por tanto de los vectores obtenidos. Por lo que no se plantea la elección de un tipo de información, sino el uso combinado entre ellos.

Y por último, estas técnicas de reconocimiento, y en particular, las referidas a la extracción de parámetros del contorno clasificándolo con HMM, pueden ser extensibles a otras aplicaciones manuscritas, tales como el reconocimiento de código postal de una carta manuscrita, o el reconocimiento de palabras segmentadas, entre otras, como ya se ha venido comentando en el capítulo quinto.

7.3 Líneas futuras de trabajo

A lo largo de esta tesis se han comentado algunas aplicaciones relacionadas con los datos manuscritos y otras técnicas que usan información geométrica, con la finalidad de optimizar los procesos. Estos campos de actuación, resumirán las líneas futuras de trabajo que se quieren proponer, se pueden concretar en los siguientes puntos:

- Mejora en la extracción de parámetros
- Utilización de sistemas clasificadores híbridos
- Expansión del ámbito de las aplicaciones usando las técnicas propuestas

La mejora de la extracción de parámetros va en referencia del uso de sistemas cada vez más eficientes para crear vectores de características más discriminantes entre patrones. Se puede observar como a medida, que la aplicación manuscritas aumentaba de complicación, las técnicas utilizadas eran más numerosas y complejas, pero proporcionaban mayor tasa de clasificación, por tanto esto es una línea que hay que seguir potenciando e investigando.

7.3.1 Sistemas clasificadores híbridos

Los sistemas híbridos o combinados han surgido de la idea de la propia variedad de información, partiendo de que se tienen dos informaciones de distinta naturaleza. Por tanto, resultaría lógico pensar que ambas informaciones están incorreladas, ya que proceden de distinta extracción de datos, a pesar de que la fuente de partida sea la misma, la imagen del patrón manuscrito. Continuando con el proceso, esta información en forma de vector se introduce en el clasificador y finalmente éste devuelve unas probabilidades, siendo la máxima de éstas el valor estimado por el clasificador como cierto. Pero en los casos que ésta sea errónea por parte de una cierta técnica aplicada a un reconocedor, se pueden intentar corregir a partir de las informaciones de éxito de otras técnicas.

Los sistemas híbridos, usan una combinación de informaciones diversas, y esto realmente es un acierto. El hecho de tener varias fuentes de información, con varios

clasificadores diferentes es una mayor riqueza en la aportación de parámetros discriminantes, que faciliten la diferenciación entre patrones.

Se ha venido desarrollando una simple arquitectura, basada en la heurística, pero que puede ser el inicio de sistemas más complejos y con mejor resultado. En la siguiente gráfica se presenta el modelo.

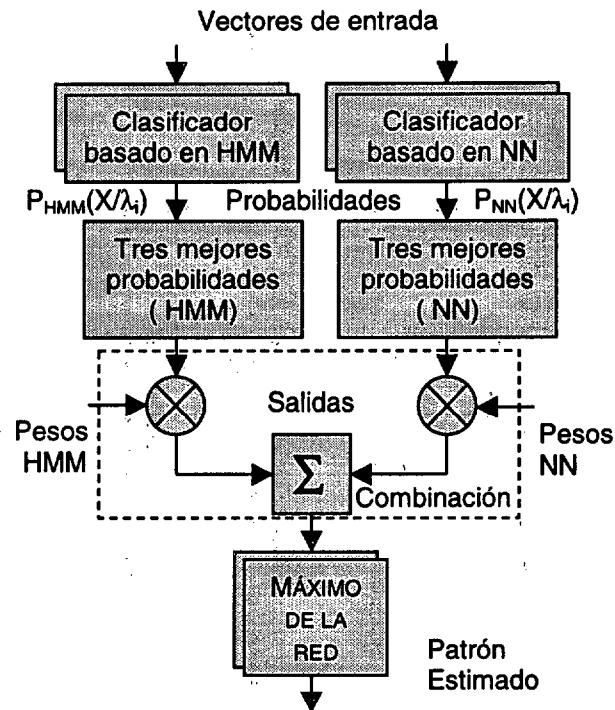


Figura 7.2: Combinación de sistemas de clasificadores de diferente tipo de información.

Lo que se pretende con este combinador es tomar una regla de decisión para que pueda corregir fallos, dados por el clasificador HMM, para que sean corregidos por el clasificador NN y viceversa. Por tanto, se plantea la construcción de esta red combinadora, a partir de los clasificadores de forma aislada, tomando sus tres mejores probabilidades de éxito y combinado estas probabilidades con un cierto peso, para tomar como regla de decisión el mayor valor dado por la suma de las probabilidades por su peso. Este sistema de combinación se puede observar en la anterior figura 7.2.

Una vez definido el sistema de combinación, se va a experimentar para algunas de las aplicaciones manuscritas que se han abordado, en particular, para los dígitos, las letras y las firmas. Como los clasificadores aislados ya están optimizados y se han realizado unas pruebas de promedio para verificar su buen funcionamiento, se va a hacer una prueba extensible a este sistema combinador, para probar así su funcionamiento.

Dadas las tablas 4.4, 4.7, 4.8, 6.5 y 6.6, en las que se reflejan los valores promedios en media y varianza para cada una de las aplicaciones, se va a implementar este sistema para ver la aportación que puede reflejar.

<i>Nº de Test de dígitos</i>	<i>Clasificador HMM</i>	<i>Clasificador NN</i>	<i>Clasificador Combinado</i>
Test 1	94.48%	87.33%	95.37 %
Test 2	93.89%	89.51%	95.34 %
Test 3	94.21%	86.94%	95.19 %
Test 4	93.42%	85.96%	94.65 %
Test 5	94.48%	86.59%	95.48 %
Test 6	94.98%	89.51%	95.46 %
Test 7	93.86%	86.94%	94.77 %
Test 8	94.35%	85.96%	95.57 %
Test 9	94.24%	88.10%	95.66 %
Test 10	94.77%	86.71%	95.76 %
Media ± Varianza	94.27% ± 0.21	87.35% ± 1.23	95.32 % ± 0.13

Tabla 7.1: Sistema combinado para los clasificadores HMM y NN con dígitos.

Para los dígitos, tal y como se observa en la tabla 7.1, se produce un incremento en la tasa media de resultados, y en todos los casos, el clasificador combinado siempre estima mejor resultado que el mejor clasificador de forma aislada. Incluso se mantiene el resultado más estable, ya que su varianza también es menor.

Para las letras, se ha realizado esta prueba tanto para minúsculas como para mayúsculas, y su resultado se recoge en las siguiente tablas.

<i>Nº de Test de letras minúsculas</i>	<i>Clasificador HMM</i>	<i>Clasificador NN</i>	<i>Clasificador Combinado</i>
Test 1	73.63 %	61.06 %	75.54 %
Test 2	73.61 %	59.38 %	75.98 %
Test 3	71.22 %	59.02 %	73.82 %
Test 4	70.85 %	60.14 %	73.89 %
Test 5	74.62 %	60.61%	76.87 %
Media ± Varianza	72.79 % ± 1.29	59.84 % ± 0.80	75.22 % ± 1.16

Tabla 7.2: Sistema combinado para los clasificadores HMM y NN con letras minúsculas.

<i>Nº de Test de letras mayúsculas</i>	<i>Clasificador HMM</i>	<i>Clasificador NN</i>	<i>Clasificador Combinado</i>
Test 1	88.03 %	67.40 %	89.56 %
Test 2	86.54 %	69.86 %	88.00 %
Test 3	83.94 %	68.83 %	85.58 %
Test 4	83.81 %	68.63 %	85.67 %
Test 5	85.27 %	69.14 %	87.13 %
Media ± Varianza	85.52 % ± 1.34	68.77 % ± 0.95	87.19 % ± 1.29

Tabla 7.3: Sistema combinado para los clasificadores HMM y NN con letras mayúsculas.

Para cada una de las dos tablas, 7.2 y 7.3, el resultado que se ha obtenido en media es mayor, a las tasas de los clasificadores de forma aislada, si se observa de forma individual para cada uno de las evaluaciones, el resultado también es siempre superior, al de los clasificadores de forma aislada, y ocurre tanto para las letras minúsculas como para las letras mayúsculas.

Por último, se va a ejecutar este sistema para la aplicación de las firmas, pero se va a realizar con los dos sistemas de detección de contorno propuestos en el capítulo sexto. Se va a realizar con el reconocimiento dado por la clasificación del contorno por agrupamiento de regiones y por la clasificación del contorno a partir de sus coordenadas polares, en primer lugar y de la envolvente expresada por centroides y en coordenadas polares en segundo término.

<i>Nº de Pruebas de firmas</i>	<i>Contorno por agrupación de regiones</i>	<i>Contorno por coordenadas polares</i>	<i>Clasificador Combinado</i>
Test 1	85.00%	71.67%	89.07%
Test 2	86.48%	71.67%	91.11%
Test 3	85.74%	71.30%	92.04%
Test 4	85.93%	67.22%	88.89%
Test 5	85.00%	72.22%	90.19%
Test 6	85.93%	67.22%	89.07%
Test 7	85.56%	69.63%	88.52%
Test 8	83.89%	68.52%	89.81%
Media ± Varianza	85.44% ± 0.70	69.93% ± 1.82	89.84% ± 1.07

Tabla 7.4: Sistema combinado I para el clasificador HMM para la técnica de contorno para firmas manuscritas.

<i>Diez pruebas promediadas</i>	<i>Envolvente por coordenadas polares</i>	<i>Envolvente por centroides</i>	<i>Sistema lineal combinado</i>
Media ± Varianza	92.52% ± 2.01	92.72% ± 1.67	95.96% ± 0.60

Tabla 7.5: Sistema combinado II para el clasificador HMM para la técnica de contorno para firmas manuscritas.

Observando los resultados de estas últimas tablas (7.4 y 7.5), la mejora establecida con respecto a los clasificadores aislados es considerable, tanto en promedio, como en los valores individuales de cada uno de las evaluaciones, y con la correspondiente disminución de varianza, lo que implica un sistema más estable en el reconocimiento.

En conclusión de los sistemas combinados, se podría decir que en todos los casos establece una mejora, en mayor o menor medida, frente a los valores del mejor clasificador de forma aislada. Si se parte de una extracción de parámetros incorrelada, la tasa de reconocimiento final se puede incrementar si se utiliza esta combinación de los clasificadores. Por tanto, hay que profundizar las investigaciones entorno a estos sistemas y las arquitecturas que se puedan utilizar.

7.3.2 Expansión del ámbito de aplicación

Parece lógico pensar que del contorno de las aplicaciones manuscritas se pueden extraer parámetros secuenciales discriminantes para clasificarlos con HMM, llevando esta idea a otras aplicaciones manuscritas y de otra naturaleza, podría conseguirse que el sistema funcionase. Siguiendo esta extrapolación, en esta tesis se está investigando sobre

diferentes ámbitos que responden a resultados preliminales, mientras se indaga en las técnicas más apropiadas. Estas aplicaciones son;

- Aplicaciones manuscritas
- Aplicaciones biométricas
- Aplicaciones biológicas
- Aplicaciones del lenguaje dactilológico

Si se extiende el estudio realizado sobre estas aplicaciones manuscritas, se podría plantear la creación de un OCR (*Object Character Recognition*) para documentos manuscritos, de tal forma que con un escáner se pudiera recuperar los textos manuscritos que contuviera. Este OCR para manuscritos, se puede particularizar para formularios que hay que rellenar a mano, muy típicos de la Administraciones Pública o de la adquisición de datos con distintas finalidades (estadísticas).

También se podría crear una aplicación para reconocer la dirección exacta y la persona a la que va dirigida una carta, mediante la lectura de los datos de un sobre postal. Y por comentar otra aplicación, se podría hacer la verificación de un cheque bancario, tanto de la cifra que contiene, como de la firma de la persona que lo ha extendido.

Dentro de las aplicaciones biométricas se está trabajando en la identificación de personas a partir de algunas de sus características biométricas, en particular, de los labios [Gómez01-pfc] y de las manos.

Si se indaga en la literatura sobre el reconocimiento de labios, se puede encontrar que esta aplicación se venía desarrollando como complemento al reconocimiento automático de voz, y no específicamente al reconocimiento de labios [Zhang97]. La pretensión de este sistema, es la identificación de las personas a partir de sus labios, como parámetro biométrico. Se basa en los principios establecidos para las aplicaciones manuscritas. Se parte de una base de datos a la que se le aplica un preprocesado de imagen antes de calcular su contorno. Una vez establecido, se extraen sus parámetros y se usa un sistema híbrido, en este caso formado por un HMM y una NN, para obtener la decisión final del sistema clasificador. En la siguiente figura, se ilustra este proceso, en un diagrama de bloques.

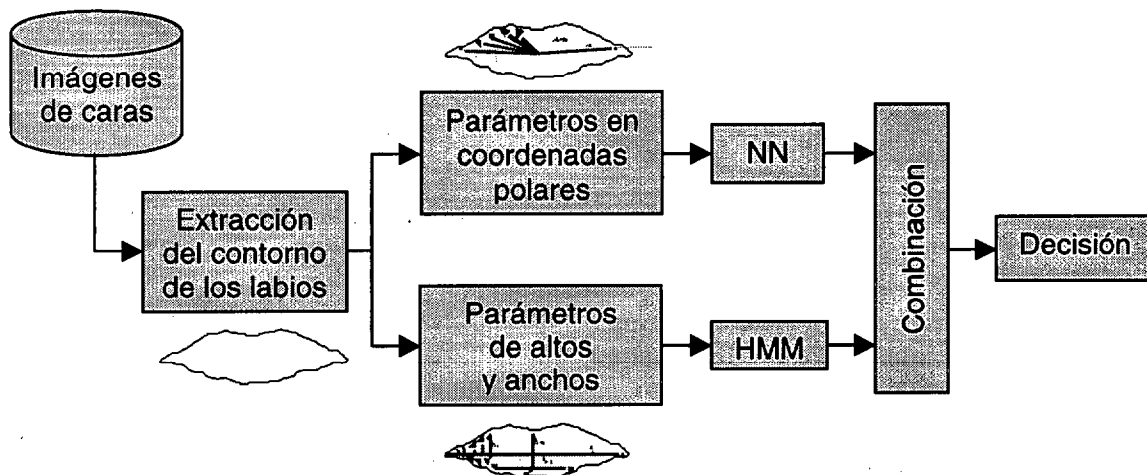


Figura 7.3: Sistema de reconocimiento biométrico mediante labios.

Entrando más en detalle, la base de datos está formada por 50 personas diferentes con 10 imágenes por personas, teniendo un total de 500 muestras. Los labios, en cada imagen, ocupan una posición distinta, ya que las diez imágenes fueron tomadas mientras se leía una frase.

Cada una de estas muestras es una imagen con formato RGB, a la que se le aplica la transformación de Chang, Huang y Novak [Chang94], con la finalidad de conseguir una imagen independiente de la luminosidad, y del bello facial que pueda existir en los rostros, ya que son los principales problemas que se han encontrado. En la siguiente imagen, se puede observar estos ejemplos;

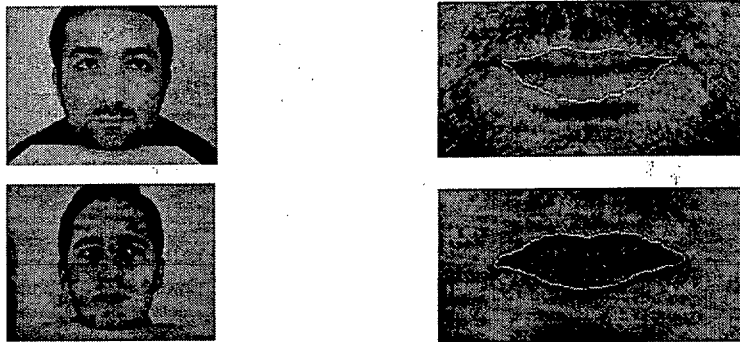


Figura 7.4: Efecto de independencia de luminosidad y bello facial.

Una vez, detectado la zona de los labios, se centra el estudio en el cálculo de la envolvente de los labios, que se realiza en dos pasos. Por un lado, la imagen es binarizada mediante el establecimiento del umbral adaptativo de Ridler and Calvar [Ridler01]. Y en segundo lugar, se aplican algunos operadores morfológicos para producir un suavizado en el contorno del labio, logrando finalmente, obtener la envolvente de los mismos.

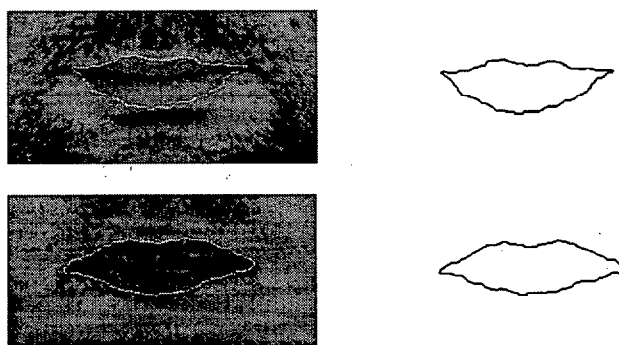


Figura 7.5: Detección de la envolvente de los labios.

Ya finalmente, queda obtener los parámetros a partir del contorno, tal y como se observa en la figura 7.3, con las mismas técnicas ya analizadas en esta tesis. Estos parámetros son clasificados con la arquitectura presentada en dicha figura, usando el 50% de la base de datos para entrenar y el resto para testear; y se logran alcanzar las tasas de reconocimiento que se muestran en la siguiente tabla.

<i>Sistema clasificador</i>	<i>Tasa de reconocimiento</i>
HMM (altos y anchos)	86.1% ± 2.8
NN(coordenadas polares)	96.2% ± 2.5
Sistema combinador (HMM y NN)	96.9% ± 1.2

Tabla 7.6: Tasas de reconocimiento de los diversos sistemas para clasificar labios.

Con esta ejecución, se observa que se llegan a alcanzar un alto nivel de reconocimiento ya que en la literatura no se encuentran referencias, para el reconocimiento de personas, utilizando la información de la secuencia y geométrica que existe en los labios.

Del mismo modo se ha realizado el mismo experimento para las manos. Basándose en la forma de las manos, se sigue la secuencia del contorno de la misma, y se puede llegar a establecer un sistema de identificación de personas según la biometría de sus manos.

Este estudio se está desarrollando para información geométrica, tales como las medidas entre puntos característicos de la mano, y se estudiará para la información extraída del contorno, llegando al uso de un sistema híbrido como en la figura 7.3. Para esta aplicación se parte de una base de datos de 50 manos diferentes con 6 muestras por mano. Una vez que se captura se establece la extracción de parámetros, para crear el vector de características, tal y como se indica en la figura 7.6. Una vez, ya se tenga el vector, se utiliza una red neuronal con funciones radiales (RBF), llegándose a lograr un éxito del 95.81%.

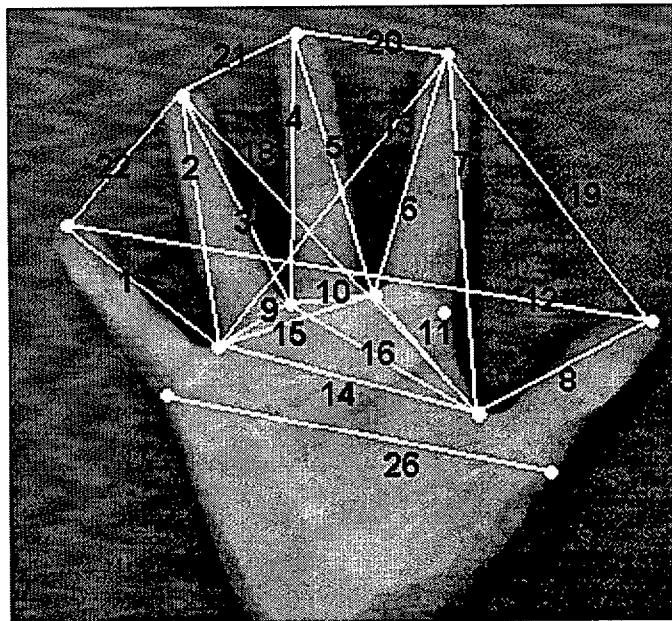


Figura 7.6: Extracción de parámetros en las manos.

La propuesta está en implementar lo mismo visto para las aplicaciones manuscritas en el reconocimiento por la biometría de las manos, usando el contorno de las mismas y sus variaciones modeladas con las diferentes técnicas propuestas.

En las aplicaciones biológicas se está investigando en el reconocimiento de especies endémicas arbóreas de Canarias a partir de las hojas de las mismas. El uso de la información secuencial en esta aplicación que se está desarrollando es básica, para su reconocimiento como se observa en la tabla 7.6. Se está intentando diferenciar entre 16 especies diferentes con 20 muestras por especie, y el esquema que se ha establecido, guarda mucha similitud con los ya estudiados, tal y como se ilustra en la siguiente figura:

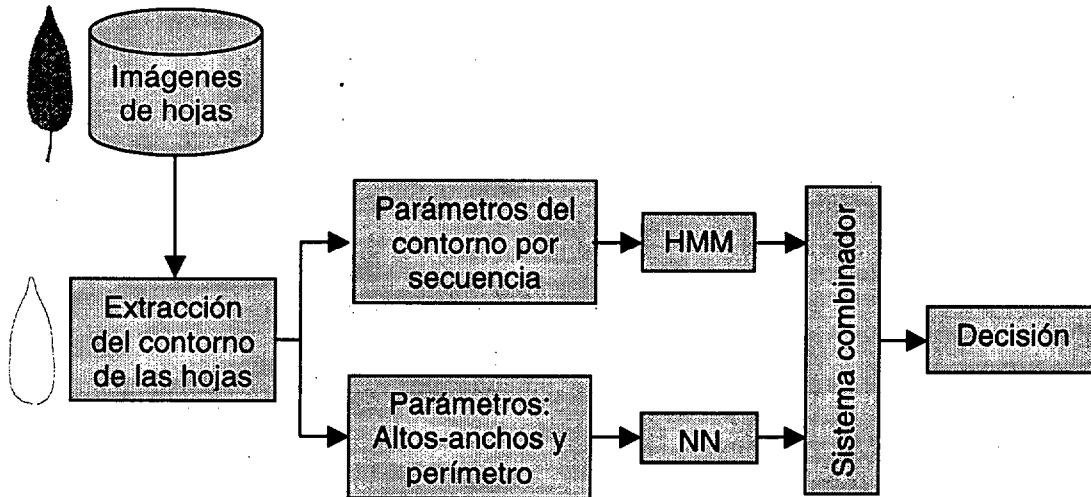


Figura 7.7: Sistema clasificador de especies arbóreas por su hoja.

La extracción de parámetros se basa en el cálculo del contorno a partir de la aplicación de los mismos algoritmos que en las aplicaciones manuscritas. El seguimiento del contorno se ha hecho por la secuencia con la matriz de saltos por la variación de la dirección para clasificarlo con HMM. Y para NN se ha usado la técnica de altos, anchos y el valor del perímetro. Para esta base de datos, estos algoritmos dieron el resultado mostrado en la tabla 7.7.

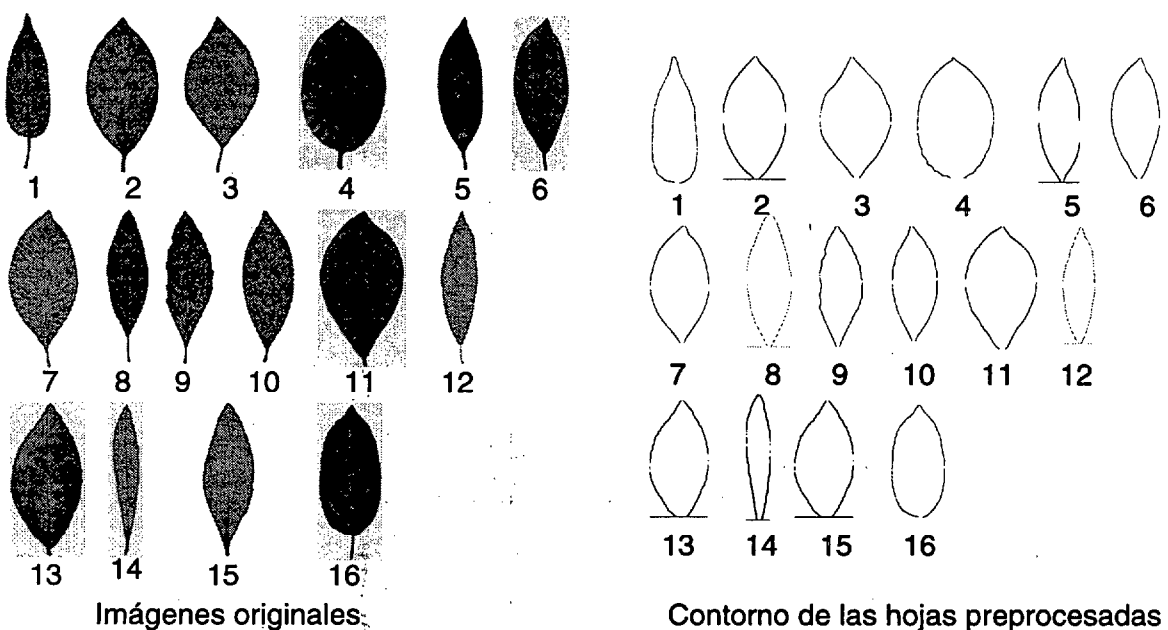


Figura 7.8: Cálculo del contorno de las hojas.

Bajo las condiciones de la figura 7.7, y estableciendo un entrenamiento con el 50% de las muestras por especie, se han logrado las siguientes tasas:

<i>Sistema clasificador</i>	<i>Tasa de reconocimiento</i>
HMM (secuencia)	75.41%
NN(Altos, anchos y perímetro)	43.2%
Sistema combinador (HMM y NN)	78.35%

Tabla 7.7: Tasas de reconocimiento de los diversos sistemas para clasificar árboles endémicos de Canarias.

Estableciendo varias etiquetas a los símbolos contenidos en un estado del HMM, hasta un total de ocho, el sistema mejora, lográndose los siguientes valores:

<i>Sistema clasificador</i>	<i>Tasa de reconocimiento</i>
HMM (secuencia)	78.37%
NN(Altos, anchos y perímetro)	43.2%
Sistema combinador (HMM y NN)	80.2%

Tabla 7.8: Tasas de reconocimiento de los diversos sistemas para clasificar árboles endémicos de Canarias con multietiquetado.

Estos resultados se consideran buenos como primera aproximación, ya que se sigue investigando al respecto, y al no disponer de referencias técnicas en la literatura, ya que sólo se encuentran del tipo biológico [Im98].

Por último se ha llevado esta técnica también al ámbito del alfabeto dactilológico [Ladrón-01-pfc], utilizado por personas con problemas de audición. En esta aplicación, se ha buscado la información del contorno que deja la mano, cuando ésta describe una letra de dicho alfabeto. En particular, hay un total de 25 letras que se representan con 20 muestras por patrón, ya que algunas de ellas se logran con movimiento, y en una imagen estática es muy difícil de registrar.

El procesado aplicado a esta aplicación lleva a la imagen de un formato RGB hasta una imagen binaria, que contiene el vector tal y como se observa en la siguiente figura.

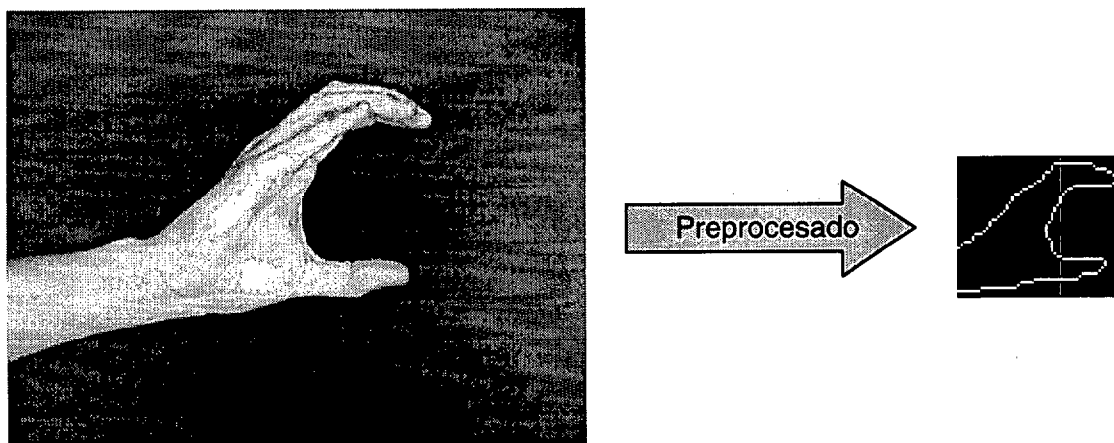


Figura 7.9: Procesado de imagen aplicado a la letra "C".

Una vez se tiene el elemento principal (el contorno), se pasa a extraer los parámetros. Las investigaciones que se están realizando, se han centrado en la técnica de la secuencia para modelar los HMM y la técnica de altos, anchos y transiciones para modelar la NN. Estos resultados, son preliminares, y ni siquiera se ha estudiado el sistema de forma combinada. Los resultados obtenidos hasta ahora se muestran en la siguiente tabla:

<i>Sistema clasificador</i>	<i>Tasa de reconocimiento</i>
HMM (secuencia)	93.6%
NN(Altos, anchos y transiciones)	91.2%

Tabla 7.9: Tasas de reconocimiento sobre el alfabeto dactilológico.

Un primer comentario sobre estos resultados, es el buen nivel que ofrece nuevamente la información secuencial del contorno, al igual que la información modelada por la red neuronal, lo que hace presagiar un aumento del resultado si se investiga sobre las técnicas de parametrización y las de la arquitectura a utilizar. A modo de resumen, se establece un diagrama de bloques que recoge este análisis, según las mencionadas técnicas que se están utilizando para reconocer el alfabeto dactilológico; y podría responder a la siguiente figura:

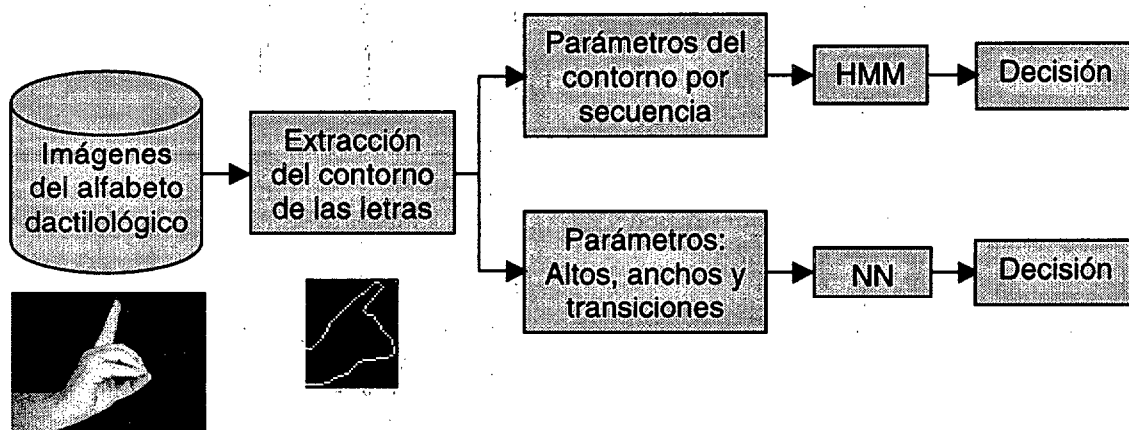


Figura 7.10: Sistema de reconocimiento del alfabeto dactilológico.

Como se refleja en estos estudios, el diagrama de bloques es el mismo, y se amolda con un buen grado de similitud, a estas aplicaciones comentadas, aunque poseen una diversa naturaleza. Todas tienen en común el uso de la clasificación con HMM aplicadas al contorno, como parte fundamental para su reconocimiento. Por tanto, sólo hay que buscar aplicación desde este punto de vista, del uso del HMM, para hacer que este método sea de utilidad en el complejo mundo de la clasificación.

Como conclusión, se ve que la técnica de extraer información del contorno clasificada con HMM, se puede hacer extensible a muchos campos, quizás no radique en buscar campos de actuación sino en buscar donde se encuentra la información secuencial de una aplicación, para aplicar este reconocedor. Al igual que hay que profundizar y seguir investigando sobre las aplicaciones que se han presentado, para buscar la secuencia más discriminante de las formas a clasificar.

Bibliografía

- [Banks90] S. P. Banks, *Signal Processing, Image Processing and Pattern Recognition*, Ed. Prentice Hall International, 1990.
- [Bishop95] C. M. Bishop, *Neural Networks for Pattern Recognition*, Ed. Oxford University Press, 1995.
- [Blumenstein99] M. Blumenstein and B. Verma, "A Neural Network for Real-World Postal Address Recognition", in *European Signal Processing*, pp. 23-27, 1999.
- [Bow92] S. Bow, *Pattern Recognition and Image Preprocessing*, Ed. Marcel Dekker, 1992.
- [Bozinovic89] R.M. Bozinovic and S.N. Srihari, "Offline Cursive Script Word Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, pp. 68-83, enero de 1989.
- [Brakensiek99] A. Brakensiek, A. Kosmala, D. Willett, W. Wrang y C. Rigoll, "Performance evaluation of a new hybrid modelling technique for handwriting recognition using identical On-line and Off-line data", *Proceedings of the 5th International Conference on Document Analysis and Recognition*, pp. 446-449, 1999.
- [Busel94] R. Buse and Z. Liu, "Feature extraction and analysis of handwritten words in grey-scale images using Gabor filters", *Proceedings of IEEE International Conference on Image Processing*, vol. 1, pp. 164-168, 1994.
- [Camino99] José L. Camino, Carlos M Travieso, Ciro R. Morales, Miguel A. Ferrer, "Signature Classification by Hidden Markov Model", in *33rd Annual 1999 International Conference on Security Technology*, pp. 480-484, octubre de 1999.
- [Camino99-pfc] José L. Camino Carmona, *Reconocimiento automático de firmas* Proyecto Final de Carrera de la Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad de Las Palmas de Gran Canaria, mayo de 1999.
- [Chang94] Thomas C.Chang, Thomas S.Huang y Carol Novak. "Facial Feature Extraction from Color Images", *Proceedings of the 12th International Conference on Pattern Recognition*, vol. 2, pp. 39-43, 1994.
- [Chen94] M. Chen, A. Kundu and J. Zhou, "Off-line handwritten word recognition using a hidden Markov model type stochastic network", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, pp. 481-496, mayo de 1994.

- [Chinnasarn98] K. Chinnasarn, "Removing Salt and Pepper noise in text/graphics images", *Proceedings of the Conference on Circuits and Systems*, pp. 459-462, noviembre de 1998.
- [Cho97] S. Cho, "Neural-Network Classifiers for Recognizing Totally Unconstrained Handwritten Numerals", *IEEE Transactions on Neural Network*, vol. 8, n° 1, pp. 43-53, enero de 1997.
- [Choisy00] C. Choisy and A. Belaid, "Analytic word recognition without segmentation based on Markov random fields", *Proceedings of the Seventh International Workshop on Frontiers in Handwriting Recognition*, pp. 487-492, septiembre de 2000.
- [Cohen94] E. Cohen, J. Hull, and S. Srihari, "Control structure for interpreting handwritten addresses", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, n° 10, pp. 1049-1055, 1994.
- [Deng98] P. Deng, H. Liao, C. Ho and H. Tyan, "Wavelet-based Off-line Signature Verification", *National Science Council of Taiwan*, NSC86-2745-E-001-004, 1998.
- [Dzuba97] G. Dzuba, A. Filatov and A. Volgunin, "Handwritten ZIP Code Recognition", in *IEEE International Conference on Image Processing*, pp. 766-770, 1997.
- [El-Yacoubi99] A. El-Yacoubi, M. Gilloux, R. Sabourin and C. Suen, "An HMM-based approach for off-line unconstrained handwritten word modeling and recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 218, pp. 752-760, agosto de 1999.
- [Emiroglu97] I. Emiroglu and M.B. Akhan, "Pre-processing of finger images", *European Conference on Security and Detection*, pp. 147-151, 1997.
- [Expósito01-pfc] Carmen R. Expósito Batista, *Reconocimiento automático del código postal en cartas con dirección manuscrita*, Proyecto Final de Carrera de la Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad de Las Palmas de Gran Canaria, junio de 2001.
- [Fu80] K.S. Fu, "Digital Pattern Recognition" *Communication and Cybernetics*, vol. 10, 1980.
- [Fukunaga90] K. Fukunaga, "Statistical Pattern Recognition", Ed. Academic Press, 1990.
- [Garris95] M. D. Garris, P. J. Grother, S. A. Janet, C. L. Wilson, "Off-line handwriting recognition from forms.", *IEEE International Conference on Systems, Man and Cybernetics*, vol. 3, pp. 2783-2788, 1995.
- [Gómez01-pfc] Enrique Gómez Blanco, *Identificación mediante biometría de los labios*, Proyecto Final de Carrera de la Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad de Las Palmas de Gran Canaria, noviembre de 2001.

- [Gosselim97] B. Gosselim, "Improved Hand-written character recognition thanks to a new geometric distortion method", *Sixth International Conference on Image Processing and its Applications*, vol. 1, pp. 327-331, 1997.
- [Govindaraju93] V. Govindaraju, A. Shekhawat, and S. Srihari, "Interpretation of handwritten addresses in US mail stream", in *Fifth International Conference on Document Analysis and Recognition*, pp. 291-294, 1993.
- [Govindaraju95] Venu Govindaraju and Sagur N. Srihari, "System for reading handwritten documents", *IEEE International Conference on Systems, Man and Cybernetics*, vol. 3, pp. 2347-2352, 1995.
- [Grother92] P.J. Grother, "Karhunen Loéve feature extraction for neural handwritten character recognition", *Proceedings of Applications of Artificial Neural Network III*, vol. 1709, pp. 155-166, Abril de 1992
- [Grother95] P. J. Grother, "NIST Special Database 19: Handprinted Forms and Characters Database", [CD ROM]. National Institute of Standards and Technology, CD ROM SD19, version 1.1, marzo de 1995.
- [Hastie94] T. Hastie, *Handwritten Digit Recognition via Deformable Prototypes*, AT&T Bell Laboratories, marzo de 1994.
- [Hernando97] J. Hernando, C. Nadeu, J.B. Mariño, "Speech recognition in a noisy environment based on LP of the one-sided autocorrelation sequence and robust similarity measuring techniques", *Speech communications*, vol. 21, pp. 17-31, 1997.
- [Hwang97] Y. Hwang y S. Bang, "An Efficient Method to Construct a Radial Basis Function Neural Network Classifier" *IEEE Transactions on Neural Network*, vol. 10, nº 8, pp. 1485-1503, 1997.
- [Im98] C. Im, H. Nishida and T.L. Kunii, "Recognizing plant species by leaf shapes-a case study of the Acer family", *Proceedings of the Fourteenth International Conference on Pattern Recognition*, vol.2, pp. 1171-1173, August 1998.
- [Jain01] A. Jain, R. Bolle and S. Pankanti, *Biometrics. Personal Identification in Networked Society*, Ed. Kluwer Academic Publisher, 2001.
- [Jeong99] C. Jeong and D. Jeong, "Hand-written digit recognition using Fourier Descriptors and Contour information", *Proceedings of the IEEE Region 10 Conference TENCON 99*, vol. 2, pp.1283-1286, 1999.
- [Kaewkongka99] T. Kaewkongka, K. Chamnongthai and B. Thipakom, "Off-line signature recognition using parameterised Hough transform", *Proceedings of the Fifth International Symposium on Signal Processing and Its Applications*, vol. 1, pp. 451-455, 1999.
- [Kawajiri91] H. Kawajiri and N. Suzumura, "Classification of hand-written digits by a large scale neural network 'CombNET-II'", *IEEE International Joint Conference on Neural Network*, vol. 2, pp.1021-1026, 1991.
- [Knerr92] S. Knerr, L. Personnaz and G. Dreyfus, "Handwritten Digit Recognition by Neural Network with Single-Layer Training", *IEEE Transactions on Neural Networks*, vol. 3, pp. 962-968, 1992.

- [Kumar99] S. Kumar, M. Crawford and J. Ghosh, "A versatile framework for labelling imagery with a large number of classes", *International Joint Conference on Neural Network*, Julio de 1999.
- [Ladrón01-pfc] José Ángel Ladrón de Guevara Villar, *Reconocedor del alfabeto dactilológico*, Proyecto Final de Carrera de la Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad de Las Palmas de Gran Canaria, diciembre de 2001.
- [Lam92] L. Lam, S. Lee and C. Suen, "Thinning methodologies: a comprehensive survey", *IEEE Transactions on pattern analysis and machine intelligence*, vol. 4, n° 9 pp. 869-885, septiembre de 1992.
- [LeCun99] Y. LeCun, P. Haffner, L. Bottou and Y. Bengio, *Object Recognition with Gradient-Based Learning*, AT&T Shannon Laboratories, USA, 1999.
- [Lee96] S. Lee, "Off-line recognition of totally unconstrained handwritten numerals using multilayer cluster neural network", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, n° 6, pp. 648-652, 1996.
- [Maravall93] D. Maravall Gómez-Allende, *Reconocimiento de formas y visión artificial*, Ed. Ra-ma, 1993.
- [Morales99] C.R. Morales, C.M. Travieso, I.G. Alonso and M.A. Ferrer, "Algoritmo secuencial de esqueletización par un sistema de reconocimiento de caracteres", XIV Simposium Internacional Científico de Radio, pp. 199-200, 1999.
- [Morales99-pfc] Ciro R. Morales Marrero, *Reconocimiento de dígitos manuscritos con cadenas ocultas de Markov*, Proyecto Final de Carrera de la Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad de Las Palmas de Gran Canaria, abril de 1999.
- [Neubauer96] C. Neubauer, *Recognition of Handwritten Digits and Human Faces by Convolutional Neural Networks*, International Computer Science Institute, diciembre de 1996.
- [O'Gorman92] L. O'Gorman, "Image and document processing techniques for the RightPages Electronic Library System", *Proceedings of the IEEE of 11th International Conference on Pattern Recognition*, vol. 2, pp. 260-263, 1992.
- [O'Gorman95] L. O'Gorman and R. Kasturi, *Document Image Analysis*, IEEE Computer Society Press, 1995.
- [Peinado00-pfc] José D. Peinado Alonso, *Reconocimiento automático de letras manuscritas* Proyecto Final de Carrera de la Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad de Las Palmas de Gran Canaria, septiembre de 2000.
- [Proter98] S. Proter and J. Illingworth, "Combining HMM Classifiers in a Handwritten Text Recognition System", *IEEE International Conference on Image Processing*, vol. 2, octubre de 1998.
- [Provan95] G. Provan and M. Singh, "Learning Bayesian Networks Using Feature Selection", *Fifth International Workshop on Artificial Intelligent and Statistics*, pp. 450-456, 1995.

- [Rabiner89] Lawrence R. Rabiner, "A tutorial on Hidden Markov models and Selected Applications in Speech Recognition", in *Proceedings of the IEEE*, vol. 77, n.º. 2, pp. 257-286, 1989.
- [Rabiner93] L. Rabiner and B. Juang, *Fundamentals of Speech Recognition*, Ed. Prentice Hall Signal, 1993.
- [Randriamahefa93] R. Randriamahefa, J.P. Cocquerez, C. Pepin and S. Philip, "Printed music recognition", *Proceedings of the Second International conference on Document Analysis and Recognition*, pp. 898-901, 1993.
- [Ridler-01] A. Ridler, "Image Processing Fundamentals-Segmentation", [en línea], (1998), Dirección URL:<<http://www.ph.tn.tudelf.nl/Courses/FIP/noframes/fip-Segmenta.html>> [Consulta:13 de junio de 2001]
- [Rigoll98] G. Rigoll, A. Kosnala, "A Systematic Comparison Between On-line and Off-line for Signature Verification with Hidden Markov Model", *International Conference on Pattern Recognition*, pp. 1755-1757, agosto de 1998.
- [Rizki90] M. Rizki, L. Tamburino and M. Zmuda, "A study of morphological feature detector complexity and character recognition rates", *Proceedings of the IEEE National Aerospace and Electronics Conference*, NAECON, 1990.
- [Salvat65] Enciclopedia Salvat-Monitor, Ed. Salvat, 1965.
- [Sánchez01] Jose A. Sánchez, Carlos M. Travieso, Itziar G. Alonso, Miguel A. Ferrer, "Handwritten signature recognizer by its envelope and strokes layout using HMM", *IEEE International Carnahan Conference on Security Technology*, London, pp. 267-271, octubre de 2001
- [Sánchez01-pfc] José A. Sánchez Mena, *Reconocimiento automático de firmas por modelado de la envolvente*, Proyecto Final de Carrera de la Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad de Las Palmas de Gran Canaria, febrero de 2001.
- [Santos99] A.C. Santos, S.S. Furuie and M.A. Gutiérrez, "Estimation of coronary blood flow by contrast propagation using simulated X-ray angiography", *IEEE Computer in Cardiology*, pp. 379-383, 1999.
- [Saon97] G. Saon and A. Belaid, "Off-line handwritten word recognition using a mixed HMM-MRF approach", *Proceedings of the Fourth International Conference on Document Analysis and Recognition*, vol. 1, pp. 118-122, 1997.
- [Savicc99] Milan D.Savicc y Marija Bojovic, "Recognition of handwritten text: basic concepts of a new approach", *IEEE 4th International Conference on Telecommunication in Modern Satellite, Cable and Broadcasting Services*, pp. 468-471, 1999.
- [Schäfer95] J. Schäfer and H. Braun, "Optimizing classifiers for handwritten digits by genetic algorithms", in *Proceedings of the International Conference on Artificial Neural Network and Genetic Algorithms*, pp. 10-13, 1995.
- [Schölkopf95] B. Schölkopf, C. Burges and V. Vapnik, "Extrating Support Data for a Given Task", *Proceedings of first International Conference on Knowledge Discovery and Data Mining*, pp. 252-257, 1995.

- [Seiler96] R. Seiler, M. Schenkel and F. Eggimann, "Off-line cursive handwriting recognition compared with On-line recognition.", *IEEE, Proceedings of the 13th International Conference on Pattern Recognition*, Vol. 3, pp. 505-509, 1996.
- [Serra82] J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press, 1982.
- [Shastri95] L. Shastri and T. Fontaine, "Recognizing Handwritten Digit Strings Using Modular Spatio-temporal Connectionist Networks" *Connection Science*, vol. 7, n° 3, pp. 211-235, 1995.
- [Srihari95] R. Srihari, *Use of lexical and syntactic techniques in recognizing handwritten text*, Center for Document Analysis and Recognition (CEDAR), NY 14228-2567, Buffalo, 1995.
- [Suárez02-pfc] Adolfo Gustavo Suárez Lorenzo, *Segmentación de texto manuscrito*, Proyecto Final de Carrera de la Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad de Las Palmas de Gran Canaria, marzo de 2002.
- [Travieso99] C. Travieso, C. Morales, I. Alonso y M. Ferrer, "Handwritten digits parameterisation for HMM based recognition", *Proceedings of the Image Processing and its Applications*, vol.2, pp. 770-774, julio de 1999.
- [Weber92] M.D. Weber, "A cost effective fingerprint verification algorithm for commercial applications", *Proceedings of the 1992 South African Symposium on Communications and Signal Processing*, pp. 99-104, 1992.
- [Zhang97] L. Zhang. "Estimation of the Mouth Features Using Deformable Templates", *Proceedings of the International Conference on Image Processing*, vol3, pp. 328-331, 1997.

Anexo A

Pruebas realizadas con los dígitos manuscritos

A.1 Introducción

Este anexo contiene el planteamiento de las diferentes pruebas realizadas para optimizar los parámetros del clasificador y de la extracción de características, con el propósito de lograr la mayor tasa de reconocimiento posible. Se ha partido de la base de datos que se ha construido sobre dígitos manuscritos. La extracción de características se realiza principalmente con la información del trazo de la secuencia debido al uso del HMM, aunque también se ha utilizado la geométrica, aplicada a NN. A partir de la interpretación de los resultados, se han buscado los valores óptimos para el mejor funcionamiento de cada clasificador. De esta forma se han realizado un total de seis experimentos, para fijar y concluir con la mejor propuesta.

A.2 Optimización del clasificador para dígitos manuscritos

El uso de los modelos ocultos de Markov marcan de forma clara la tendencia en el planteamiento de la extracción de parámetros, centrando la atención en la información extraída del trazo de las imágenes.

Para esta información del trazo se han utilizado dos técnicas para extraer sus parámetros, la detección de contorno y la esqueletización. Los vectores que se van a obtener por ambos métodos van a pasar por las mismas pruebas, con la finalidad de obtener cual de los dos sistemas devuelve mejores resultados en la tasa de reconocimiento. Como ya se ha comentado anteriormente, se pretende variar el número de estados, las muestras por escritor, y los escritores empleados en el entrenamiento para determinar cuales dan mejores resultados. Una vez determinados estos valores, ya se puede decir que se ha optimizado el uso del clasificador HMM para esta aplicación en particular.

Con respecto a la información geométrica, aunque no es el objetivo de esta tesis se va a realizar una simple prueba. Se expondrá la extracción de parámetros por la técnica del gradiente [Cho97], y el ajuste del clasificador NN se realizará variando, el número de neuronas de la capa oculta, el número de escritores por entrenamiento, así como las muestras que se utilizan por cada uno, y el coeficiente de aprendizaje. De esta forma se puede intentar mejorar la tasa de reconocimiento mediante el planteamiento de un sistema clasificador híbrido, con fuentes de parámetros incorreladas.

Las pruebas que se van a realizar para los dígitos manuscritos serán pruebas dependientes e independientes de escritor. Esto no es más que el uso de las mismas muestras con las que se han entrenado, para realizar el test. De esta forma se puede apreciar si el clasificador es capaz de discriminar las muestras en el caso de que no las haya conocido antes; y también da información de que la base de datos logra tener una buena constitución, ya que con algunas muestras de un escritor se es capaz de reconocer el mismo patrón de diferentes escritores.

La estructura de las pruebas realizadas seguirá la siguiente secuencialización:

1. Optimización realizada para HMM con detección de contornos interior y exterior para dígitos de tamaño original
2. Optimización realizada para HMM con detección de contornos exteriores para dígitos de tamaño original
3. Optimización realizada para HMM con detección de contornos exteriores para dígitos de tamaño reducido
4. Optimización realizada para HMM con detección de contornos exteriores parametrizado con coordenadas polares para dígitos de tamaño reducido.
5. Optimización realizada para HMM con la técnica de esqueletización para dígitos de tamaño reducido
6. Optimización realizada para NN con la técnica del gradiente

A.2.1 Optimización realizada para HMM con detección de contornos interior y exterior para dígitos de tamaño original

La primera prueba ha sido variar el número de estados. Para ello, se han tenido que fijar los demás parámetros, se han escogido 30 escritores para las muestras de entrenamiento y 10 escritores para las muestras de test. A partir de estos datos, se ha escogido de las 12 repeticiones que hay por patrón y escritor, tres para crear el entrenamiento y nueve para el test. Esto se ha invertido, usando nueve repeticiones para entrenamiento y tres para el test, como se muestran en las tablas A.1 y A.2 respectivamente. La extracción de los datos se ha realizado con la técnica del contorno, usando el contorno interior y exterior de los dígitos con tamaño original. Los resultados que se obtuvieron fueron [Morales99-pfc]:

Nº de Estados	Tasa de reconocimiento (test dependiente de escritor)	Tasa de reconocimiento (test independiente de escritor)
10	88.33%	84.77%
20	92.11%	88.44%
30	93.22%	87.77%
40	92.88%	89.44%
50	93.66%	89.33%
60	94.44%	90.88%
70	92.33%	82.88%
80	94.88%	91.22%
90	96.22%	88.11%

Tabla A.1: Tasas de reconocimiento para 3 muestras en el entrenamiento variando el número de estados.

Nº de Estados	Tasa de reconocimiento (test dependiente de escritor)	Tasa de reconocimiento (test independiente de escritor)
10	86.33%	83.66%
20	92%	87.44%
30	87.66%	85.55%
40	96.66%	90.44%
50	95.33%	90.77%
60	96.66%	92.33%
70	96.33%	89.77%
80	97%	91.44%
90	97%	89.33%

Tabla A.2: Tasas de reconocimiento para 9 muestras en el entrenamiento variando el número de estados.

A continuación se exponen las gráficas que permiten relacionar la tasa de reconocimiento con los valores de las variables anteriormente citadas, de manera más intuitiva:

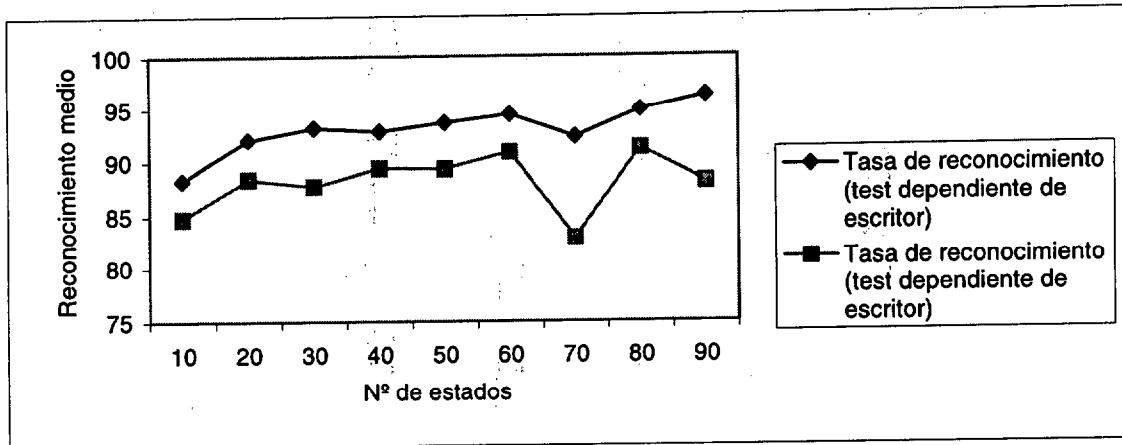


Figura A.1: Tasa de reconocimiento en función del nº de estados (3 muestras por escritor).

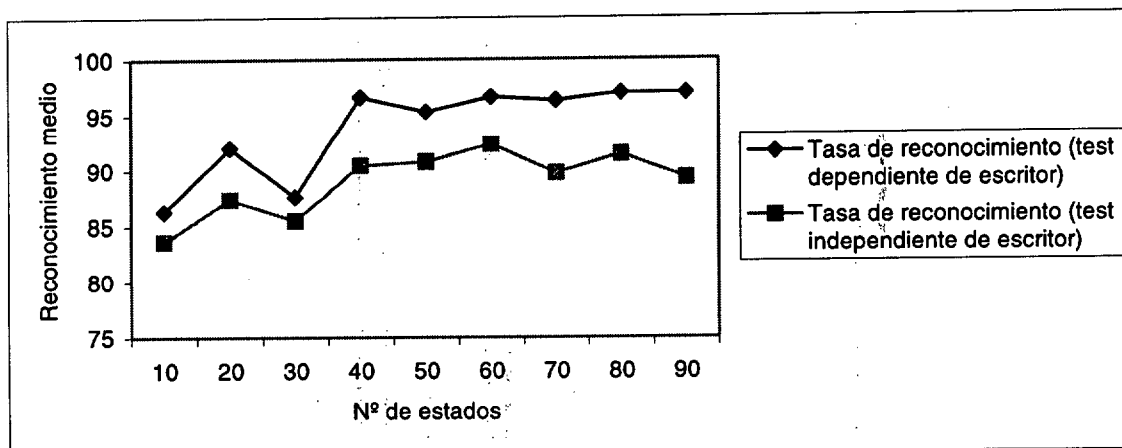


Figura A.2: Tasa de reconocimiento en función del nº de estados (9 muestras por escritor).

En las figuras A.1 y A.2 se puede observar como las tasas de reconocimiento al realizar el test dependiente de escritor son superiores a las obtenidas al realizar el test independiente de escritor. Este resultado es un tanto lógico de esperar, ya que el clasificador funcionará mejor con las propias muestras con las que se entrenó, que con otras diferentes, aunque los resultados siguen más o menos la misma tendencia. Otra característica importante es que los mejores resultados se centran entorno al número de estados situados entre 40 y 60.

Del cruce entre las dos gráficas, se observa que al incrementar las muestras por escritor usadas para entrenar el modelo, las tasas de reconocimiento mejoran. Esto ocurre excepto para los modelos con muy pocos estados, donde el exceso de muestras puede provocar un sobreentrenamiento, provocando la pérdida de generalización y por tanto, la bajada de tasas de éxito.

Otra de las variables a analizar es la influencia del número de escritores usados durante el entrenamiento del modelo. Para ello se ha preparado una prueba utilizando de la base de datos un agrupamiento de 10, otro de 20 y un último de 30 escritores. Y cada agrupamiento utiliza 3 ó 9 muestras por escritor para el entrenamiento. El test realizado se ha hecho con las muestras restantes hasta completar las 12 repeticiones por escritor. Las tasas de reconocimiento obtenidas son:

Nº de Estados	Reconocimiento independiente de escritor		
	Tasa de reconocimiento (10 escritores)	Tasa de reconocimiento (20 escritores)	Tasa de reconocimiento (30 escritores)
10	84.44%	83.88%	84.99%
20	88.44%	79.22%	90%
30	87.77%	88.77%	90%
40	89.44%	91.55%	90.44%
50	89.33%	91.33%	88.6%
60	90.88%	88.33%	92.94%
70	82.88%	93.88%	97.22%
80	91.22%	89.77%	95.55%
90	88.11%	91.33%	93.88

Tabla A.3: Tasas de reconocimiento para 3 muestras en el entrenamiento variando el número de escritores.

Nº de Estados	Reconocimiento independiente de escritor		
	Tasa de reconocimiento (10 escritores)	Tasa de reconocimiento (20 escritores)	Tasa de reconocimiento (30 escritores)
10	83.66%	75.33%	72%
20	87.44%	86.44%	88.33%
30	85.55%	92.77%	90.66%
40	90.44%	89.55%	91.66%
50	90.77%	94.66%	90.55%
60	92.33%	93.66%	91.66%
70	89.77%	92.55%	94.66%
80	91.44%	93.33%	94.83%
90	89.33%	94.11%	94.77%

Tabla A.4: Tasas de reconocimiento para 9 muestras en el entrenamiento variando el número de escritores

Las gráficas que se pueden obtener a partir de estas tasas de reconocimiento se muestran en las siguientes figuras, ya que su interpretación se hace más visible:

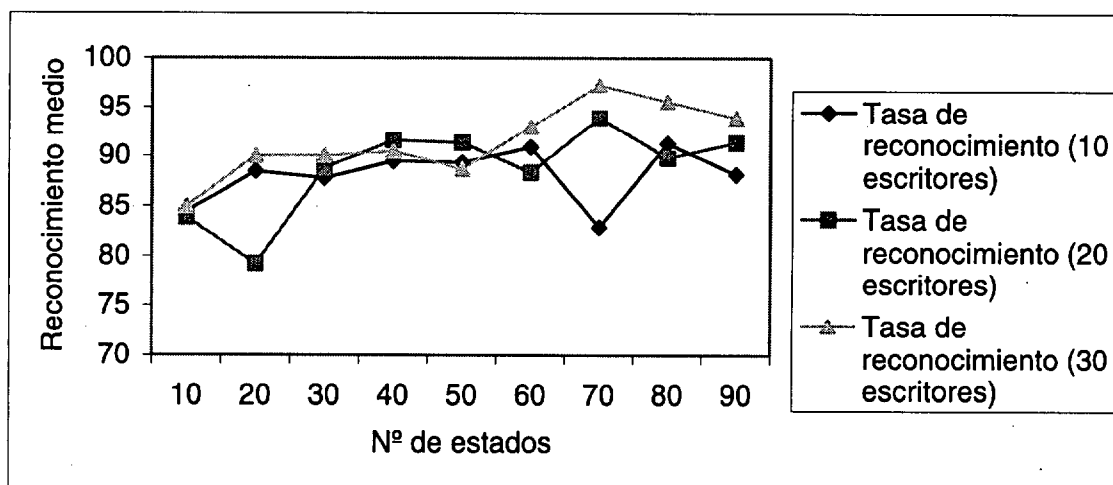


Figura A.3: Tasa de reconocimiento en función del nº de escritores (3 muestras por escritor).

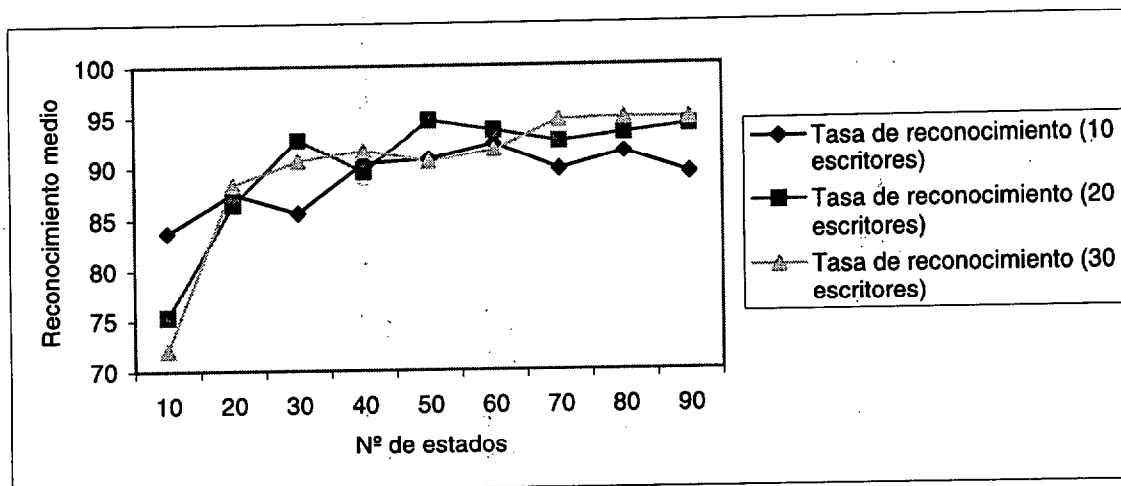


Figura A.4: Tasa de reconocimiento en función del nº de escritores (9 muestras por escritor).

En las últimas dos figuras, se observa como para un número reducido de estados (≤ 30) los modelos entrenados con menos escritores obtienen mejores tasas de reconocimiento. En cambio, para un mayor número de estados la situación es la inversa. Con estas gráficas se demuestra que existe una relación entre el entrenamiento necesario para obtener las mejores tasas y el número de estados de la cadena. Es decir, para un número de estados dado existe un cierto de entrenamiento óptimo, por encima y por debajo del cual caen las tasas de reconocimiento.

Haciendo resumen, la tasa de entrenamiento, al usar los vectores que contienen el contorno interior y exterior de los dígitos originales en las fases de entrenamiento y test del clasificador aumenta al:

- Incrementar el número de estados del modelo, encontrándose el máximo aproximadamente en 80 estados.
- Incrementar el número de muestras por escritor empleadas en el entrenamiento.
- Incrementar el número de escritores empleados en el entrenamiento.

De las gráficas anteriores se puede observar que la tasa de reconocimiento máxima al realizar el test independiente responde a:

Tasa de reconocimiento	97.22%
Nº de estados	70
Número de escritores usados en el entrenamiento	30
Número de muestras por escritor en el entrenamiento	3

Tabla A.5: Mejor resultado para la optimización realizada para HMM con detección de contornos interior y exterior para dígitos de tamaño original.

A.2.2 Optimización realizada para HMM con detección de contorno exterior para dígitos de tamaño original

Las pruebas a realizar serán las mismas que se efectuaron con anterioridad, para poder comparar ambos métodos y decidir cual de las dos técnicas de extracción de parámetros tiene una mejor respuesta frente al HMM.

Dichas pruebas se han realizado para un test dependiente e independiente de escritor, variando el número de estados y las muestras utilizadas para crear los modelos. En las siguientes tablas se muestran los resultados obtenidos:

Nº de Estados	Tasa de reconocimiento (test dependiente de escritor)	Tasa de reconocimiento (test independiente de escritor)
10	76.66%	67.22%
20	83.77%	78.44%
30	91.11%	85.88%
40	90.22%	82.33%
50	83.44%	77.44%
60	89.66%	85.88%
70	95.33%	88.33%
80	90.11%	83.33%
90	93%	85.44%

Tabla A.6: Tasas de reconocimiento para 3 muestras en el entrenamiento variando el número de estados.

Nº de Estados	Tasa de reconocimiento (test dependiente de escritor)	Tasa de reconocimiento (test independiente de escritor)
10	91%	86.55%
20	87.33%	78.66%
30	88%	80.33%
40	90.66%	82.55%
50	96%	91.77%
60	92.33%	86.22%
70	93.66%	86.88%
80	98.66%	90.88%
90	91%	83.22%

Tabla A.7: Tasas de reconocimiento para 9 muestras en el entrenamiento variando el número de estados.

Para estas tablas con las tasas de éxito, se pueden realizar las siguiente gráficas:

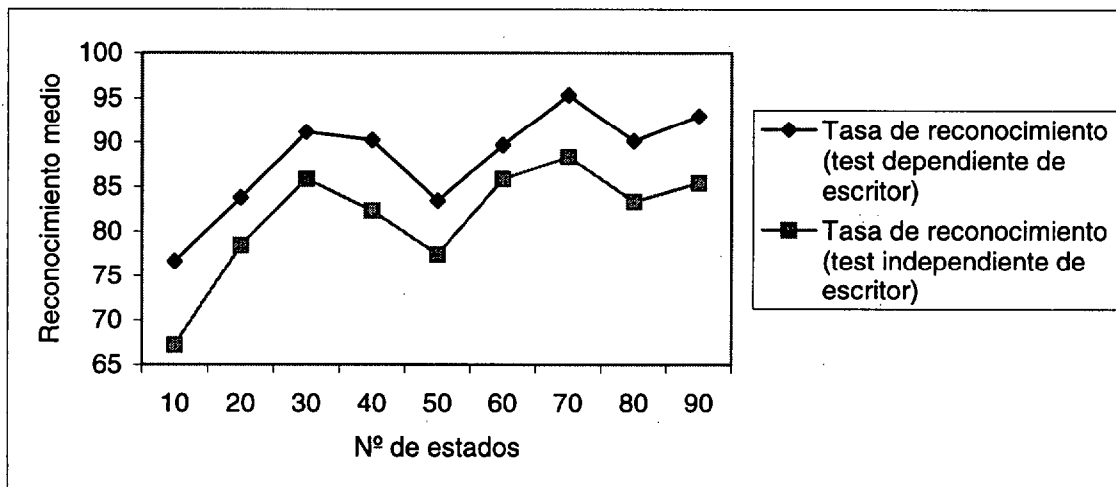


Figura A.5: Tasa de reconocimiento en función del nº de estados (3 muestras por escritor).

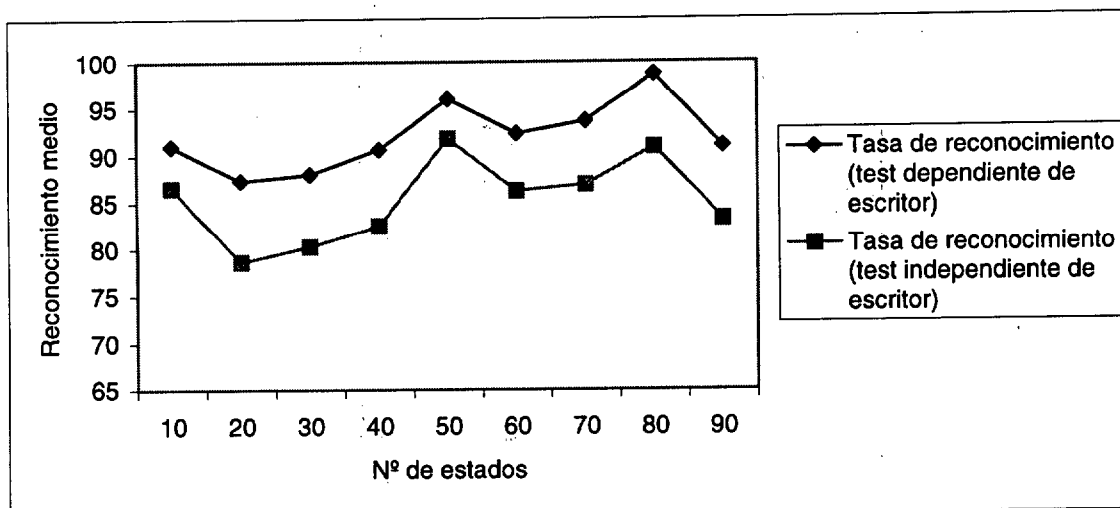


Figura A.6: Tasa de reconocimiento en función del nº de estados (9 muestras por escritor).

De estas gráficas A.5 y A.6, se puede deducir que la tasa de reconocimiento van aumentando a medida que se eleva el número de estados, alcanzando el máximo entre los 70 y 80 estados; tanto para los vectores del test dependiente como independiente. Otra observación encontrada, es la mayor tasa dada en las muestras dependientes, lógico de pensar ya que han sido esos vectores los que han definido el modelo. Y también se puede observar, que entre más muestras se usen para el entrenamiento mejor será la tasa de éxito.

Con la siguiente tabla, se pretende estudiar la evolución de la tasa de reconocimiento en función de la cantidad de escritores usados para entrenar.

Nº de Estados	Reconocimiento independiente de escritor		
	Tasa de reconocimiento (10 escritores)	Tasa de reconocimiento (20 escritores)	Tasa de reconocimiento (30 escritores)
10	67.22%	77.33%	79.66%
20	78.44%	82.11%	90.66%
30	85.88%	86.44%	88.44%
40	82.33%	88.88%	89.88%
50	77.44%	90.55%	86.22%
60	85.88%	83.33%	85.33%
70	88.33%	82.44%	84.55%
80	90.11%	90.22%	85.99%
90	85.44%	85%	87.33%

Tabla A.8: Tasas de reconocimiento para 3 muestras en el entrenamiento variando el número de escritores.

Representando los valores de la tabla A.8, se obtiene esta figura:

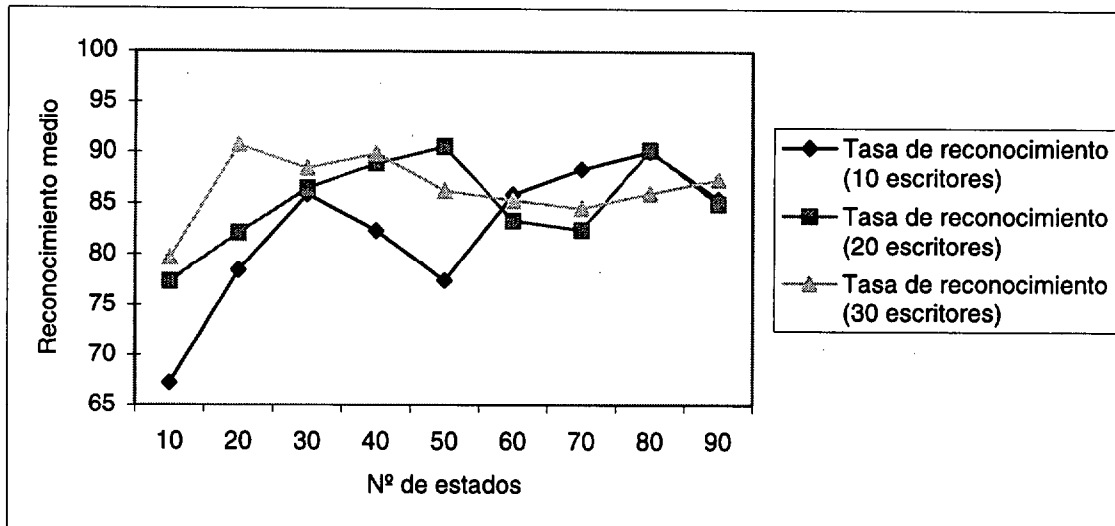


Figura A.7: Tasa de reconocimiento en función del nº de escritores (3 muestras por escritor).

De esta gráfica A.7, se pueden observar dos partes bien diferenciadas, cuando el número de estados es menor a 50, la tendencia es a mejorar la tasa de éxito al aumentar el número de escritores. Pero para un número de estados mayor a 50, la tasa de reconocimiento se mantiene por igual y con indiferencia del aumento de escritores.

Haciendo un resumen de la tasa de reconocimiento usando vectores que contienen el contorno exterior de los dígitos originales, se puede comentar que al:

- Incrementar el número de estados del modelo se localiza los valores máximos de reconocimiento entre 50 y 80 estados.
- Incrementar el número de muestras por escritor y de escritores empleados en el entrenamiento, mejora la tasa de éxito.

De las gráficas anteriores se puede observar que la tasa de reconocimiento máxima al realizar el test independiente responde a:

Tasa de reconocimiento	91.77%
Nº de estados	50
Número de escritores usados en el entrenamiento	10
Número de muestras por escritor en el entrenamiento	9

Tabla A.9: Mejor resultado para la optimización realizada para HMM con detección de contorno exterior para dígitos de tamaño original.

A.2.3 Optimización realizada para HMM con detección de contorno exterior para dígitos de tamaño reducido

Las pruebas que se van a realizar en este caso son las mismas que los otros dos casos anteriores, se pretende optimizar los parámetros del HMM para esta extracción de parámetros con los dígitos reducidos. Ésto se quiere obtener variando el número de

estados, las muestras por escritor usadas para el entrenamiento y el número de escritores utilizados para entrenar.

Las siguientes tablas muestran los valores de la tasa de reconocimiento al variar el número de estados para los casos de usar 3 y 9 muestras por escritor.

Nº de Estados	Tasa de reconocimiento (test dependiente de escritor)	Tasa de reconocimiento (test independiente de escritor)
10	89.22%	84.77%
20	93%	89%
30	95.44%	92.77%
40	94.33%	91.77%
50	92.22%	90.55%
60	94.55%	93.77%
70	91.33%	90.22%
80	84.44%	83.33%

Tabla A.10: Tasas de reconocimiento para 3 muestras en el entrenamiento variando el número de estados.

Nº de Estados	Tasa de reconocimiento (test dependiente de escritor)	Tasa de reconocimiento (test independiente de escritor)
10	91%	82%
20	94.66%	88.33%
30	96%	93.66%
40	95%	94.88%
50	98%	94.11%
60	96%	94.55%
70	93.33%	91.55%
80	76.33%	75.44%

Tabla A.11: Tasas de reconocimiento para 9 muestras en el entrenamiento variando el número de estados.

Representando los valores de las tablas A.10 y A.11 en las siguientes gráficas se obtiene:

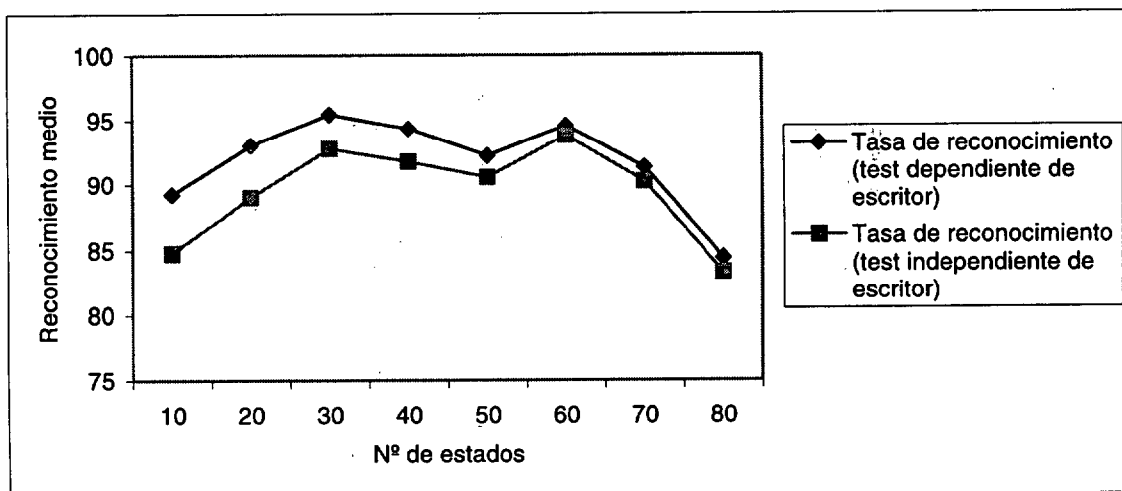


Figura A.8: Tasa de reconocimiento en función del nº de estados (3 muestras por escritor).

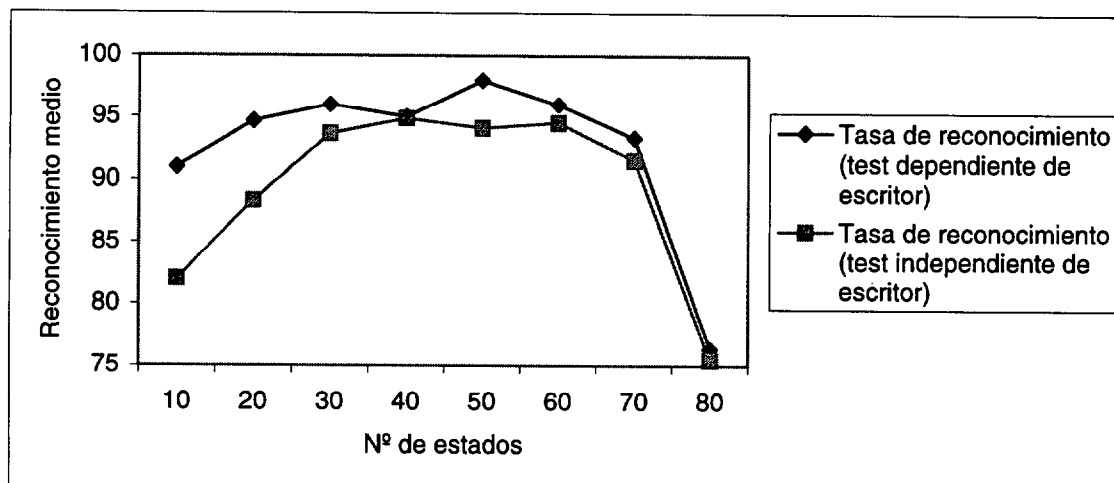


Figura A.9: Tasa de reconocimiento en función del nº de estados (9 muestras por escritor).

De las figuras A.8 y A.9 se obtiene que el número de estados donde se obtiene mejores resultados está entre los 40 y 60 estados, como consecuencia de que los vectores de los parámetros son más pequeños, al usar dígitos más reducidos. Ésto también implica una menor velocidad de procesado y clasificación. También se puede observar como las tasas de reconocimiento a partir de 80 estados caen exageradamente debido a que la longitud de los vectores es demasiado pequeña para entrenar con 80 o más estados, ya que es comparable casi al tamaño del vector. Por último, se muestra como con el aumento de muestras para entrenar, mejora la tasa de reconocimiento.

A continuación, se analiza la tendencia de la tasa de éxito frente al número de escritores y el número de las muestras de éstos utilizadas en el entrenamiento.

Nº de Estados	Reconocimiento independiente de escritor		
	Tasa de reconocimiento (10 escritores)	Tasa de reconocimiento (20 escritores)	Tasa de reconocimiento (30 escritores)
10	84.77%	90.66%	87.66%
20	89%	90.44%	93.77%
30	92.77%	95.33%	93.88%
40	91.77%	94.66%	92.55%
50	90.55%	92.11%	96.22%
60	93.77%	96.22%	95.11%
70	90.22%	92.66%	95.11%
80	83.33%	94.33%	84.66%

Tabla A.12: Tasas de reconocimiento para 3 muestras en el entrenamiento variando el número de escritores.

Nº de Estados	Reconocimiento independiente de escritor		
	Tasa de reconocimiento (10 escritores)	Tasa de reconocimiento (20 escritores)	Tasa de reconocimiento (30 escritores)
10	82%	90.11%	85%
20	88.33%	90.66%	94%
30	93.66%	95.22%	93%
40	94.88%	93.55%	95%
50	94.11%	95.33%	95.11%
60	94.55%	95.22%	95.11%
70	91.55%	96.22%	93.55%
80	75.44%	92.88%	95.88%

Tabla A.13: Tasas de reconocimiento para 9 muestras en el entrenamiento variando el número de escritores.

A continuación se muestran las gráficas A.10 y A.11, que han sido compuestas a partir de estas tablas, las A.12 y A.13.

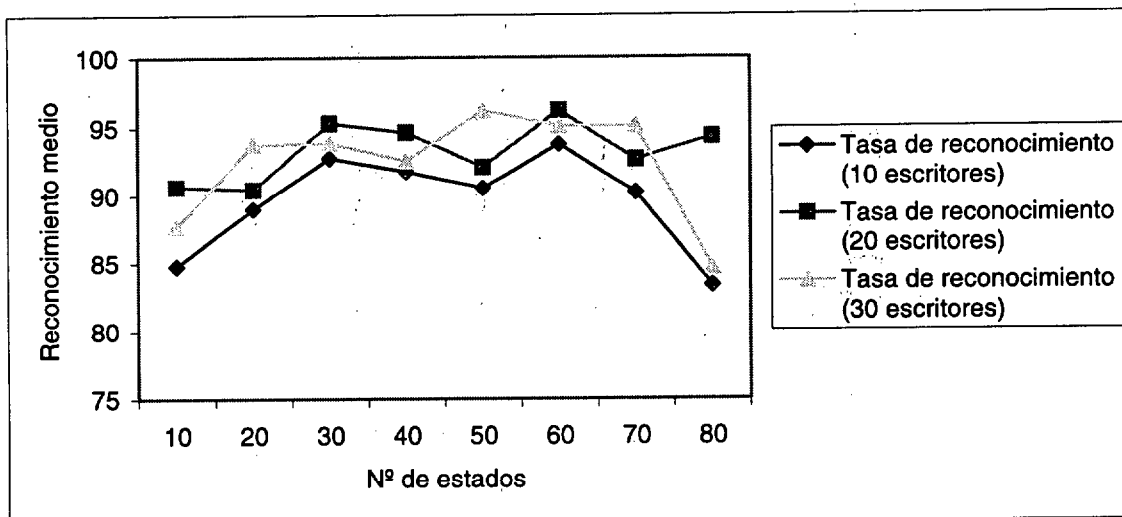


Figura A.10: Tasa de reconocimiento en función del nº de escritores (3 muestras por escritor).

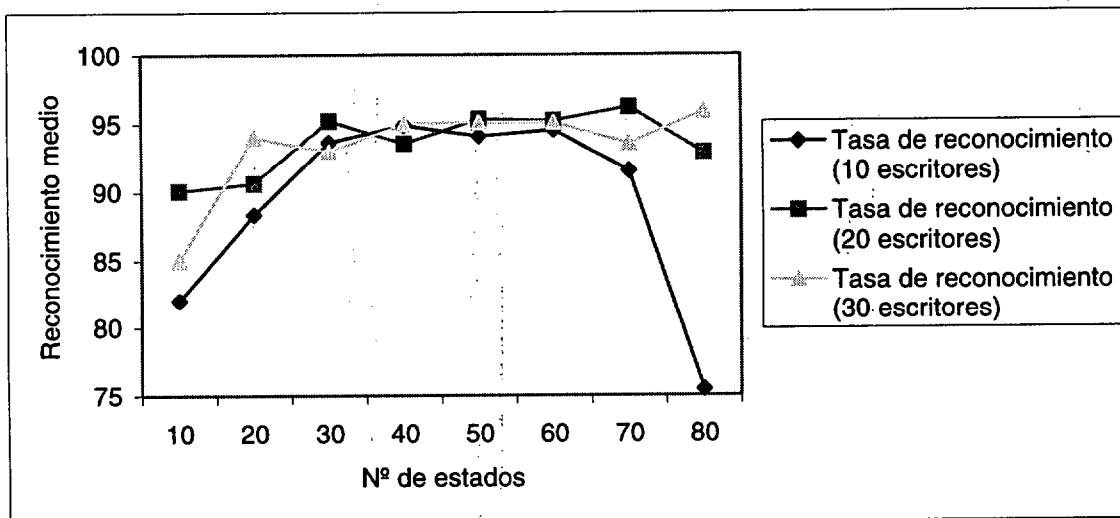


Figura A.11: Tasa de reconocimiento en función del nº de escritores (9 muestras por escritor).

De las figuras A.10 y A.11, se desprende que la tasa de reconocimiento, excepto en algún pico, aumenta a medida que se incrementa el número de escritores y el número de muestras por cada escritor en el entrenamiento. Se percibe la existencia de un rango del número de estados de los modelos, entre los 30 y los 60 estados; en el que la tasa de éxito es bastante estable y de mayor valor que en el resto de los números de estado.

Resumiendo, la tasa de reconocimiento al entrenar el HMM con la extracción de parámetros del contorno exterior de la base de datos de los dígitos reducidos, se aumenta al:

- Establecer el número de estados en la creación de un modelo, en la franja de 40 a 60.
- Aumentar el número de muestras por escritor empleadas en el entrenamiento.
- Aumentar el número de escritores empleados en el entrenamiento.

La mejor tasa de reconocimiento independiente de escritor se presenta en diferentes casos, con el mismo valor del 96.22%, como se muestra en la siguiente tabla;

<i>Tasa de reconocimiento</i>	96.22%		
Nº de estados	60	50	70
Número de escritores usados en el entrenamiento	20	30	20
Número de muestras por escritor en el entrenamiento	3	3	9

Tabla A.14: Mejor resultado para la optimización realizada para HMM con detección de contorno exterior para dígitos de tamaño reducido independiente de escritor.

A.2.4 Optimización realizada para HMM con detección de contornos exteriores parametrizado con coordenadas polares para dígitos de tamaño reducido

Para determinar lo discriminante que va a ser la vectorización con coordenadas polares del contorno exterior de tamaño reducido en los dígitos, se van a determinar dos experimentos. En el primero se va a establecer el número de estados más acertado, y en segundo término, a partir de este número de estados determinado, se fija el valor para variar el entrenamiento. De esta forma el clasificador quedará definido.

Ambas pruebas se van a mostrar gráficamente, y sus valores se recogen en las dos siguientes tablas. En ellas se diferencia entre los distintos orígenes de coordenadas del sistemas en polares, para ver cual de ellos responde mejor a la clasificación, si con sistema de referencia en el centro de masas, o en la parte inferior izquierda.

<i>Nº de Estados</i>	<i>Reconocimiento independiente de escritor</i>		
	<i>Tasa de reconocimiento (40 escritores)</i>	<i>Tasa de reconocimiento (45 escritores)</i>	<i>Tasa de reconocimiento (50 escritores)</i>
30	60.56 %	65.35 %	60.63 %
40	70.28 %	71.89 %	74.28 %
50	63.72 %	67.23 %	69.49 %

Tabla A.15: Tasas de reconocimiento para 12 muestrás en el entrenamiento variando el número de escritores en coordenadas polares con referencia al centro de masas.

Nº de Estados	Reconocimiento independiente de escritor		
	Tasa de reconocimiento (40 escritores)	Tasa de reconocimiento (45 escritores)	Tasa de reconocimiento (50 escritores)
30	53.17 %	53.26 %	57.34 %
40	55.62 %	57.93 %	61.45 %
50	55.64 %	56.42 %	59.39 %

Tabla A.16: Tasas de reconocimiento para 12 muestras en el entrenamiento variando el número de escritores en coordenadas polares con referencia a la parte inferior izquierda.

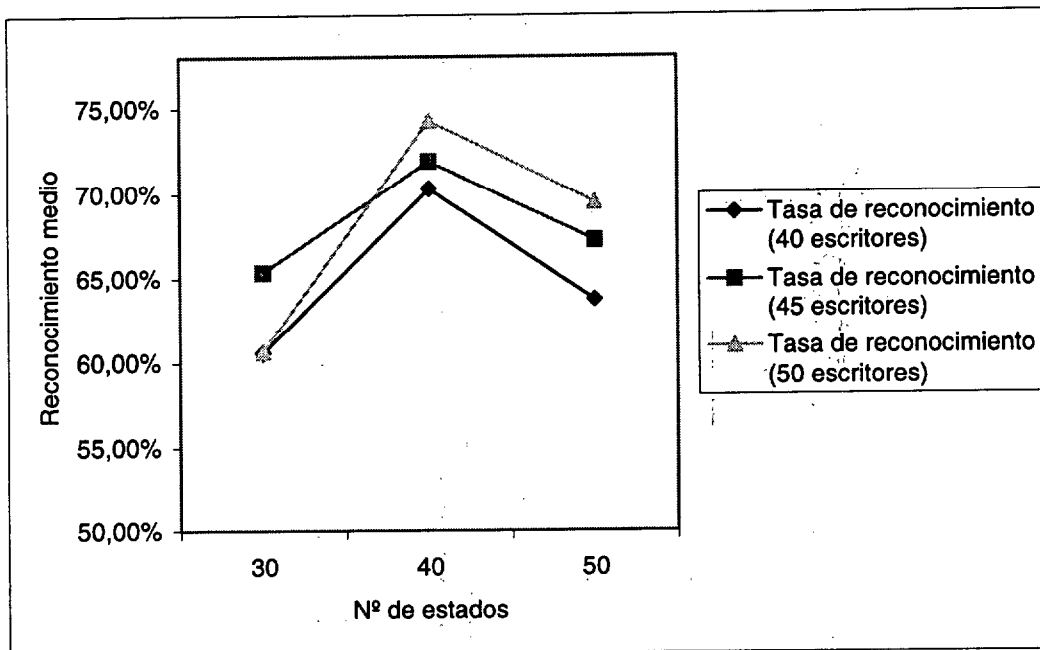


Figura A.12: Tasa de reconocimiento en polares con origen en el centro de masas.

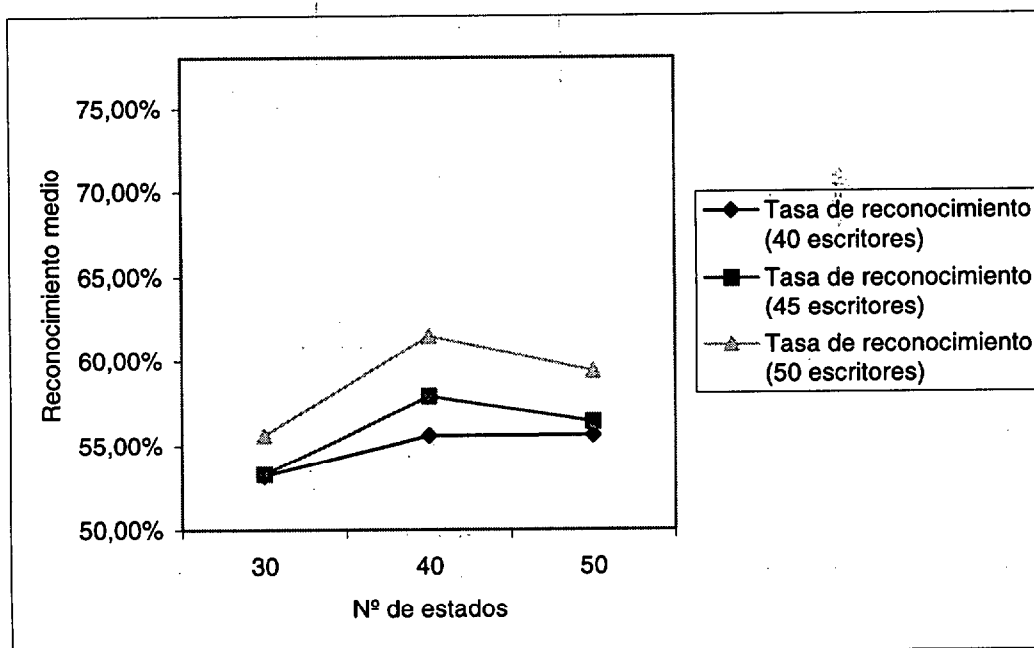


Figura A.13: Tasa de reconocimiento en polares con origen en la parte inferior izquierda.

La conclusión más inmediata de ambas gráficas, es el mejor funcionamiento del clasificador con origen en el centro de masas, frente a la parte inferior izquierda. Y en segundo lugar, es el bajo rendimiento de este tipo de vectorización frente a las estudiadas con antelación.

Por tanto, se va a dejar de simular con este tipo de técnica, debido fundamentalmente a las bajas tasas de reconocimiento encontradas.

A.2.5 Optimización realizada para HMM con la técnica de esqueletización para dígitos de tamaño reducido

De las pruebas realizadas en los casos anteriores, se puede concluir que con los dígitos reducidos se obtiene mejores resultados de reconocimiento, por eso, se va a utilizar en la extracción de parámetros la técnica de esqueletización sólo para la base de datos con los dígitos reducidos, y no para los de tamaño original.

Los valores a optimizar para el HMM son los mismos, por tanto se presentan las tasas de reconocimiento en función del número de estados y las muestras utilizadas por escritor, recogidas en las tablas A.17 y A.18.

<i>Nº de Estados</i>	<i>Tasa de reconocimiento (test dependiente de escritor)</i>	<i>Tasa de reconocimiento (test independiente de escritor)</i>
10	95.11%	89%
20	94.33%	88.88%
30	94.33%	90.22%
40	93.11%	91.33%
50	66.88%	64.22%
60	29.11%	28.36%

Tabla A.17: Tasas de reconocimiento para 3 muestras en el entrenamiento variando el número de estados.

<i>Nº de Estados</i>	<i>Tasa de reconocimiento (test dependiente de escritor)</i>	<i>Tasa de reconocimiento (test independiente de escritor)</i>
10	95.66%	92.11%
20	96.66%	93.44%
30	96%	91.55%
40	96.33%	91.44%
50	87.66%	83.55%
60	67.66%	63.66%

Tabla A.18: Tasas de reconocimiento para 9 muestras en el entrenamiento variando el número de estados.

Representando estas tablas a gráficas, se obtienen las siguientes:

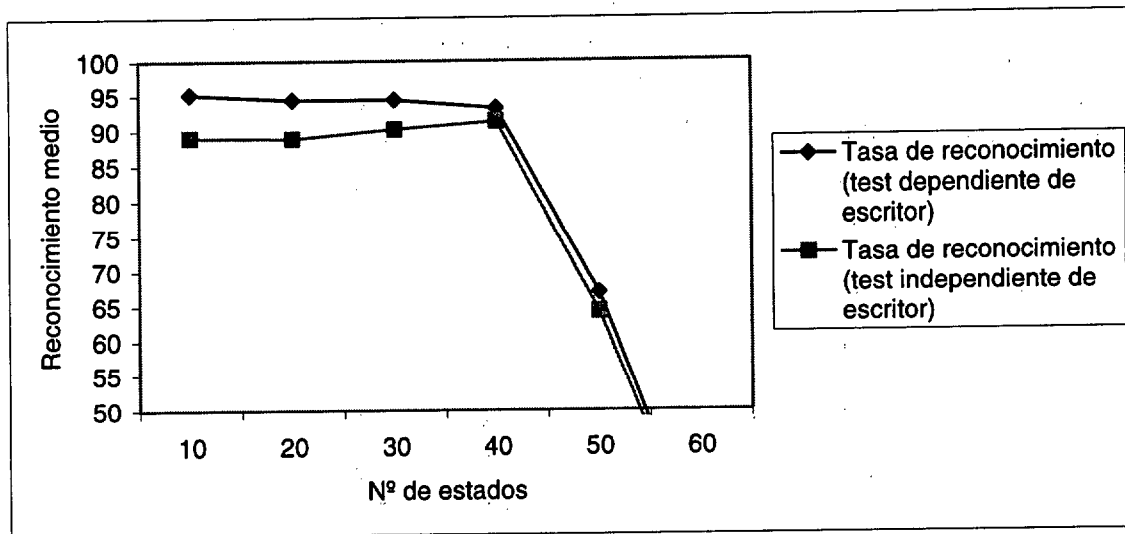


Figura A.14: Tasa de reconocimiento en función del nº de estados (3 muestras por escritor).

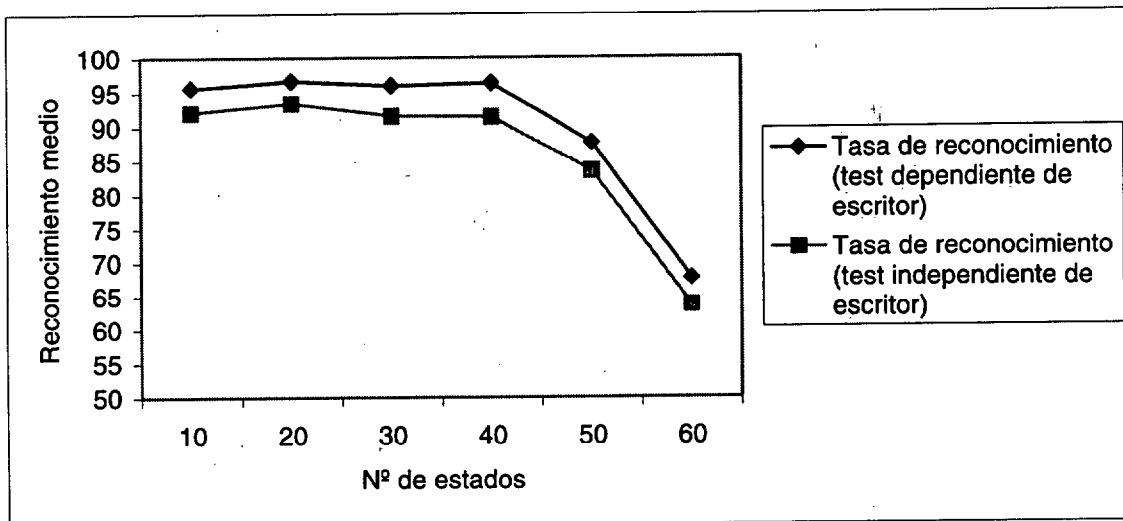


Figura A.15: Tasa de reconocimiento en función del nº de estados (9 muestras por escritor).

Debido al tamaño reducido de los vectores de parámetros para el entrenamiento y el test, se obtiene que para 20 y 30 estados se registran los mejores valores en la tasa de reconocimiento. Mientras que para los modelos creados con 50 y 60 estados, el HMM proporciona tasas de reconocimiento muy bajas, debido a que la longitud de los vectores de parámetros es comparable al número de estados.

También se observa como al aumentar las muestras en los vectores de entrenamiento, la tasa de éxito responde con otro aumento, adaptando mejor los modelos a los vectores a discriminar.

A continuación, se plantea como varía la respuesta del clasificador a partir de la variación del número de escritores y del número de muestras que aporta cada uno al vector de parámetros de entrenamiento, recogido en las tablas A.19 y A.20.

Nº de Estados	Reconocimiento independiente de escritor		
	Tasa de reconocimiento (10 escritores)	Tasa de reconocimiento (20 escritores)	Tasa de reconocimiento (30 escritores)
10	89	88.55	89.33
20	88.88	91.88	89.88
30	90.22	90.11	92.55
40	91.33	91.88	92.77
50	64.22	73	74.55
60	28.36	47.22	46.22

Tabla A.19: Tasas de reconocimiento para 3 muestras en el entrenamiento variando el número de escritores.

Nº de Estados	Reconocimiento independiente de escritor		
	Tasa de reconocimiento (10 escritores)	Tasa de reconocimiento (20 escritores)	Tasa de reconocimiento (30 escritores)
10	92.11	89.11	91.44
20	93.44	93.11	91.44
30	91.55	92.88	90.66
40	91.44	92.88	93.55
50	83.55	83.22	82.44
60	63.55	74	75.88

Tabla A.20: Tasas de reconocimiento para 9 muestras en el entrenamiento variando el número de escritores.

Los valores de estas tablas se representan en las figuras A.16 y A.17.

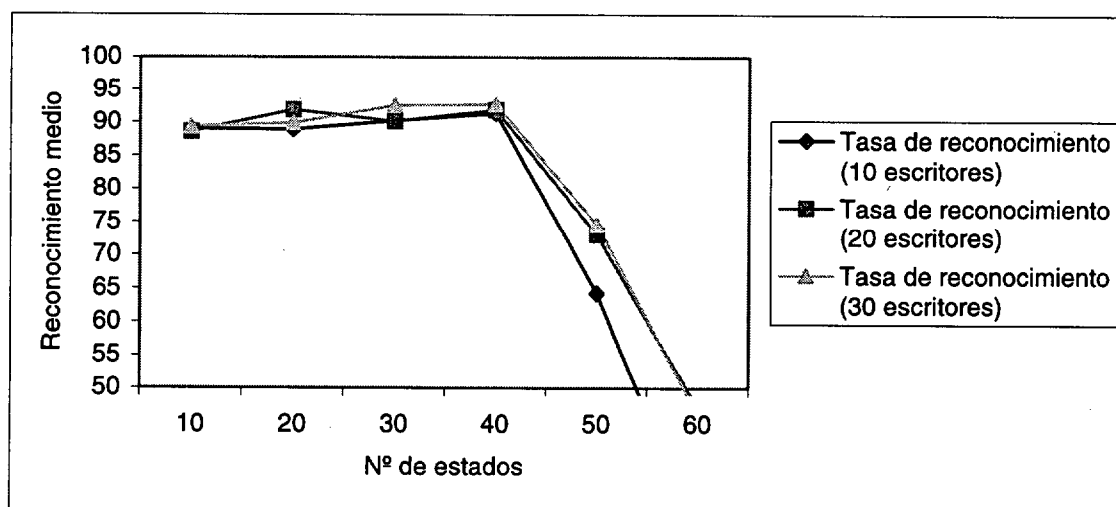


Figura A.16: Tasa de reconocimiento en función del nº de escritores (3 muestras por escritor).

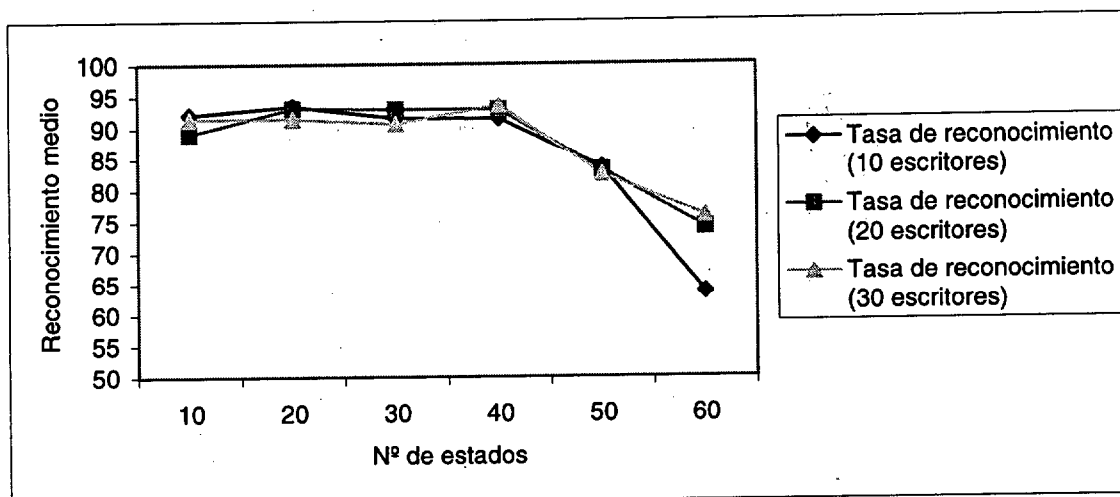


Figura A.17: Tasa de reconocimiento en función del nº de escritores (9 muestras por escritor).

De estas dos gráficas anteriores se puede deducir que al aumentar el número de escritores en el entrenamiento prácticamente no varía la tasa de reconocimiento, desde 10 hasta 40 estados. Este fenómeno puede deberse a que la técnica de esqueletización produce un acercamiento entre las formas de los diversos dígitos, es decir, aumenta la similitud entre ellos, por lo que las tasas de éxito no se verán influidas al aumentar el número de escritores. Pero a partir de 40 estados la tasa disminuye, debido a la similitud del tamaño del vector de parámetros con este número de estados.

Donde se ve un ligero aumento es al incrementar el número de muestras por escritor, debido a un mayor entrenamiento de muestras más parecidas, que disminuyen la dispersión del modelo creado.

La optimización de los parámetros del HMM se obtiene con la extracción de los parámetros, usando esqueletización a los dígitos reducidos cuando;

- Se establece el número de estados del modelo entre el intervalo de 20 y 40.
- Se aumenta el número de muestras por escritor empleadas en el entrenamiento.

La mejor tasa de reconocimiento obtenida en este proceso para un test independiente de escritor, usando esta vectorización de esqueletización de los dígitos reducidos es:

<i>Tasa de reconocimiento</i>	93.55%
Nº de estados	30
Número de escritores usados en el entrenamiento	30
Número de muestras por escritor en el entrenamiento	9

Tabla A.21: Mejor resultado para la optimización realizada para HMM con la técnica de esqueletización para dígitos de tamaño reducidos.

A.2.6 Optimización realizada para NN con la técnica del gradiente

En esta sección se realiza la optimización de la topología de la red neuronal con parámetros geométricos, para obtener la mayor tasa de reconocimiento posible. Este tipo de técnicas no es objetivo de esta tesis, aunque se ha indagado por razones de arquitecturas de clasificadores híbridas, entre HMM y NN, como ya se ha comentado.

Se pretende determinar los valores que realcen la tasa de éxito del coeficiente de aprendizaje, y del número de neuronas óptimo de la capa oculta, ya que se considera fijo el número de neuronas de la capa de entrada (dado por el vector de parámetros) y de la capa de salida (dado por el número de patrones diferentes que se quieren discriminar). El valor óptimo del coeficiente de aprendizaje varía durante la creación del modelo debido al uso de la técnica de coeficientes de aprendizaje adaptativos, por lo que sólo se podrá determinar el valor inicial de éste.

A tal efecto, se muestran las tasas de reconocimiento en función del número de neuronas de la capa oculta para cada valor del coeficiente de aprendizaje inicial, y el efecto del número de muestras usadas por escritor para el entrenamiento.

Nº de Neuronas	Coeficiente de aprendizaje			
	10^{-5}	10^{-6}	10^{-7}	10^{-8}
45	86.44%	87.33%	86.66%	88.77%
75	86.66%	86.66%	87.66%	86.77%
100	86.44%	88.22%	86.44%	87.33%
300	diverge	85.33%	84.22%	85.66%

Tabla A.22: Tasas de reconocimiento con 10 escritores y 3 muestras por escritor para el entrenamiento para dígitos.

Nº de Neuronas	Coeficiente de aprendizaje			
	10^{-5}	10^{-6}	10^{-7}	10^{-8}
45	90%	91%	92%	89.33%
75	91%	90.66%	91%	89.33%
100	89.66%	90.33%	90.33%	91%
300	diverge	90%	89.33%	88.33%

Tabla A.23: Tasas de reconocimiento con 10 escritores y 9 muestras por escritor para el entrenamiento para dígitos.

Pasando estos valores números de las tablas a gráficas, se obtienen las figuras A.18 y A.19.

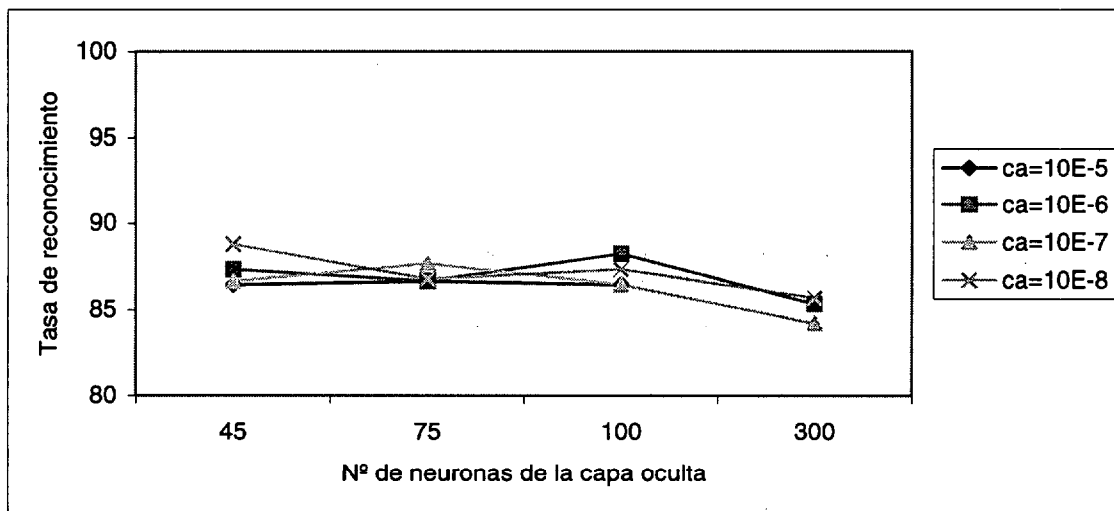


Figura A.18: Tasas de reconocimiento en función del número de neuronas de la capa oculta y el coeficiente de aprendizaje para 10 escritores y 3 muestras por escritor.

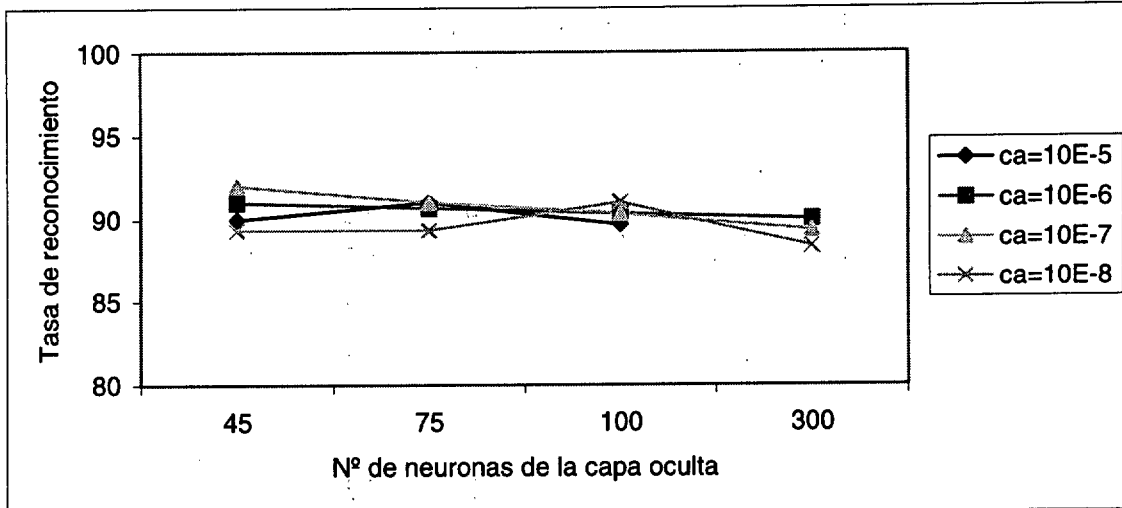


Figura A.19: Tasas de reconocimiento en función del número de neuronas de la capa oculta y el coeficiente de aprendizaje para 10 escritores y 9 muestras por escritor.

Observando las tablas A.22 y A.23, se puede ver que para 300 neuronas en la capa oculta y con un coeficiente de aprendizaje de 10^{-5} la red neuronal diverge, y el algoritmo no encuentra solución. De estas tablas y sus correspondientes figuras se puede sacar la conclusión:

- Las tasas de reconocimiento mejoran al disminuir el número de neuronas de la capa oculta, obteniéndose las mejores tasas para 45 neuronas en dicha capa. Ésto puede deberse al fenómeno denominado sobre-entrenamiento o *overfitting* que se produce cuando hay demasiadas neuronas en la capa oculta.
- Al aumentar el número de muestras usadas en el entrenamiento por escritor, la media de las tasas de reconocimiento mejora sensiblemente debido posiblemente, a que el número de pesos existente en la red neuronal necesita más entrenamiento para ajustarse adecuadamente.
- Se observa como comenzando con valores iniciales de coeficiente de aprendizaje más pequeños, la red obtiene prácticamente las mismas tasas, al realizarse los cambios en los pesos y polarizaciones a saltos menores.

Otra de las variables a optimizar es el efecto del número de escritores utilizados en el entrenamiento, sobre la tasa de reconocimiento. Las pruebas realizadas se basan en variar este número de escritores con la variación del coeficiente de aprendizaje para dos valores fijos de neuronas en la capa oculta.

Nº de Neuronas	Coeficiente de aprendizaje			
	10^{-5}	10^{-6}	10^{-7}	10^{-8}
10	86.44%	88.22%	86.44%	87.33%
20	87.66%	85.83%	86.38%	86.38%
30	88.19%	87.37%	87.41%	87.37%

Tabla A.24: Tasas de reconocimiento variando el número de escritores y con 100 neuronas de capa oculta.



Nº de Neuronas	Coeficiente de aprendizaje		
	10^{-6}	10^{-7}	10^{-8}
10	85.33%	84.22%	85.66%
20	87.22%	85.77%	85.77%
30	87.41%	87.14%	87.14%

Tabla A.25: Tasas de reconocimiento variando el número de escritores y con 300 neuronas de capa oculta.

Representando estos valores numéricos en una gráfica se obtiene;

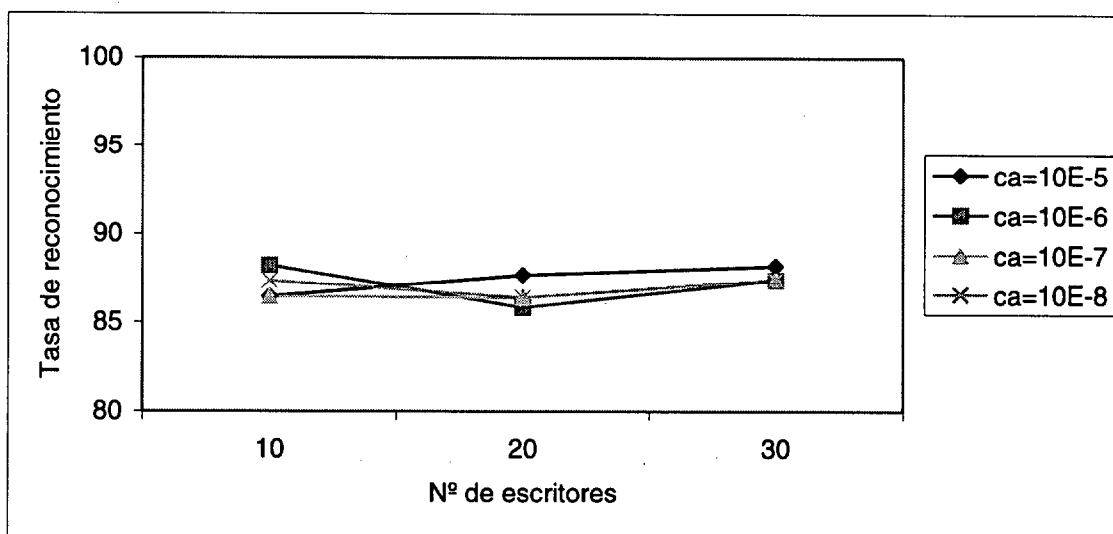


Figura A.20: Tasas de reconocimiento en función del número de escritores y el coeficiente de aprendizaje para 100 neuronas en la capa oculta.

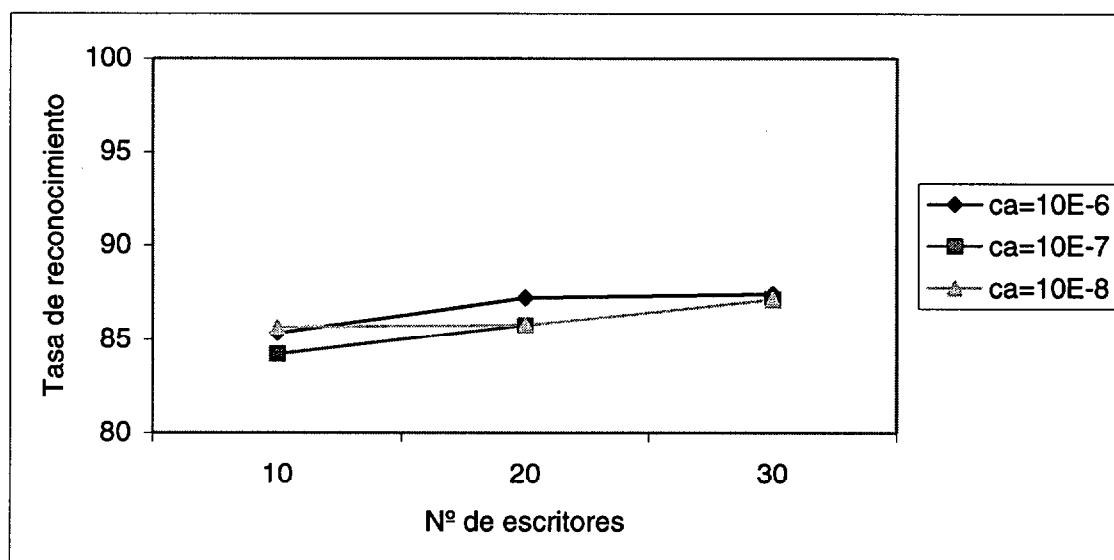


Figura A.21: Tasas de reconocimiento en función del número de escritores y el coeficiente de aprendizaje para 300 neuronas en la capa oculta.

Se puede comentar en primer lugar, que en la tabla A.25 no aparece los resultados para un coeficiente de aprendizaje de 10^{-5} y 300 neuronas en la capa oculta, porque la red divergía. Por otro lado, se descubre de las figuras A.20 y A.21, que las

tasas de reconocimiento aumentan al incrementar el número de escritores usados en el entrenamiento del clasificador, posiblemente por ser entrenado con una mayor diversidad.

Como mejores resultados en la aplicación de las redes neuronales al reconocimiento de dígitos manuscritos, se recoge que:

<i>Tasa de reconocimiento</i>	91.67%
Número de neuronas en la capa oculta	45
Número de escritores usados en el entrenamiento	30
Número de muestras por escritor en el entrenamiento	9
Coefficiente de aprendizaje	10^{-8}

Tabla A.26: Mejor resultado para la optimización realizada para NN con la técnica del gradiente para dígitos manuscritos.

A.3 Conclusiones

En las pruebas con el HMM como clasificador, hay que destacar la mejora en los resultados para los dígitos reducidos, ya que mantienen los niveles en la tasa de éxito con mayor uniformidad en los resultados (véase las figuras A.8, A.9, A.10 y A.11). En contraposición, las pruebas hechas con los dígitos de tamaño original mantiene mayor irregularidad, marcando picos y valles en la tasa de reconocimiento (véase las figuras A.3, A.4, A.5, A.6 y A.7). Con respecto al nivel de la tasa de reconocimiento alcanzado, destaca un pico para los dígitos originales, pero los valores en media de los dígitos reducidos mejora y crea un sistema de reconocimiento más estable en media y varianza. Por lo que se llega a la conclusión de que los vectores de parámetros obtenidos a partir de la base de datos de dígitos reducida, tienen un mejor comportamiento frente al HMM. Esta mejor respuesta en el comportamiento, se debe a la eliminación de un gran número de características particulares de los escritores, que existen en los trazos de la escritura de los dígitos de tamaño original, y que en el proceso de reducción han sido eliminadas. Este proceso de reducción equivale a un filtro paso bajo, ya que se eliminan las características particulares de la escritura, tendiéndose a una generalización de los dígitos de los diferentes escritores.

Otra cuestión a tratar es la técnica de extracción de parámetros que optimiza al HMM. Hay dos técnicas bien diferenciadas en el uso de la información de la secuencia de los trazos. La diferencia es clara, y el mejor comportamiento lo registra la técnica del contorno frente a la de esqueletización, como se puede observar en las tablas A.5, A.9 y A.14 frente a la tabla A.21. Si solamente se centran los resultados con respecto a los dígitos reducidos ya que ofrecen mejores tasas de éxito, la diferencia entre las tablas A.14 y A.21 se hace más marcada. Por tanto, la decisión está clara, una cadena oculta de Markov se puede modelar mejor, con el método de extracción de parámetros de contornos.

Con respecto a las técnicas del contorno, se ha probado con la extracción de la secuencia en coordenadas polares, pero los resultados no son comparables con los encontrados al seguir la secuencia con la matriz de saltos de direcciones (véase las tablas A.15 y A.16). Por tanto, no se van a tener en cuenta a la hora del análisis de la información del trazo de la escritura.

Por otro lado, se utilizó la información geométrica, modelada con NN, como contraste para comparar el estado actual del arte con los valores calculados, aunque hay que indicar que no se ha profundizado mucho en esta técnica, ya que se sale de los objetivos establecidos. También esta información geométrica, va a usarse como complemento a la información secuencial, para el posible uso de arquitecturas de clasificadores híbridas entre informaciones con distinta naturaleza.

Si se fija la observación en la mejor tasa de resultados para HMM (véase la tabla A.14) y para NN (véase la tabla A.26), se puede llegar a la conclusión de que la información secuencial tiene un mejor comportamiento que la información geométrica, de cara a la clasificación de dígitos manuscritos en esta tesis. Pero lo importante es que esta fuente de información, se pueda tener en cuenta a la hora de reconocer este tipo de patrones.

Para demostrar que estas tasas de reconocimiento no responden a picos aislados de éxito, se va a realizar una prueba de promediado de los resultados. Se ha escogido una serie de parámetros de optimización del clasificador HMM y de NN, y a partir de toda la base de datos, se han realizado diez veces la misma prueba, para finalmente obtener un resultado promedio del éxito del reconocimiento.

Para el clasificador de HMM se ha fijado como extractor de parámetros la técnica de contorno exteriores siguiendo la secuencia por la codificación de las direcciones, aplicados a la base de datos de dígitos reducidos. Y para el clasificador NN, se ha escogido los parámetros de la técnica del gradiente que maximizan el reconocimiento, dados por el apartado anterior. Finalmente se han observado los siguientes resultados:

<i>Nº de Test de dígitos</i>	<i>Clasificador HMM</i>	<i>Clasificador NN</i>
Test 1	94.48%	87.33%
Test 2	93.89%	89.51%
Test 3	94.21%	86.94%
Test 4	93.42%	85.96%
Test 5	94.48%	86.59%
Test 6	94.98%	89.51%
Test 7	93.86%	86.94%
Test 8	94.35%	85.96%
Test 9	94.24%	88.10%
Test 10	94.77%	86.71%
Media ± Varianza	94.27% ± 0.21	87.35% ± 1.23

Tabla A.27: Establecimiento de los reconocimientos medios para los clasificadores HMM y NN.

Anexo B

Pruebas realizadas con las letras manuscritas

B.1 Introducción

Este anexo B contiene las pruebas realizadas a la aplicación de las letras manuscritas. Se recuerda que este grupo de patrones esta formado por el conjunto de las letras manuscritas mayúsculas y minúsculas. Las pruebas se han realizado tanto de forma aislada como de forma conjunta para cada uno de los tipos de letras. Las técnicas empleadas para vectorizar son las mismas que se usaron en los dígitos manuscritos, y los parámetros a fijar para cada clasificador, también han coincidido. En este anexo, se ha pretendido plantear una réplica de las pruebas realizadas a los dígitos, pero usando como patrón las letras manuscritas, para finalmente obtener las conclusiones oportunas.

B.2 Optimización del clasificador para letras manuscritas

El objetivo de este anexo es evaluar el clasificador HMM según los tipos de parámetros que se empleen. La primera cuestión es la determinación de los parámetros óptimos de los clasificadores, ya que no hay ninguna fórmula matemática que permita calcularlos directamente. La única forma de determinar sus valores es de manera empírica, se crean los clasificadores dando valores diferentes a sus parámetros y se evalúan, para elegir aquellos que proporcionan la mayor tasa de reconocimiento [Peinado00-pfc].

Las pruebas que se han realizado, han tenido en cuenta que la base de datos está compuesta por letras mayúsculas y letras minúsculas, explicada y detallada en el capítulo tercero. Por tanto los clasificadores se han organizado para cubrir las diferentes posibilidades, que son:

- Clasificador para letras minúsculas
- Clasificador para letras mayúsculas
- Clasificador mixto, para letras minúsculas y mayúsculas

Dependiendo de la aplicación práctica puede tener sentido el uso de unos clasificadores u otros, pero el principal sentido que tiene para esta tesis, es corroborar que para el HMM, la información secuencial se adapta para el reconocimiento de letras. En el clasificador de letras minúsculas y mayúsculas, se encontrarán 27 patrones diferentes, y en el clasificador mixto, se tendrán justo el doble, 54 patrones, el conjunto de las minúsculas y las mayúsculas.

La mayoría de las evaluaciones de los clasificadores son independientes de escritor, es decir, se usan muestras de letras de escritores diferentes para entrenar y para evaluar el clasificador. Para entrenar se utilizarán las 12 repeticiones que hay en la plantilla de cada letra, y se varía el número de escritores entre plantillas diferentes que conforman la base de datos. Para la realización de la etapa del test, se utilizan 40 plantillas de escritores diferentes a los utilizados en la etapa de entrenamiento.

Las pruebas dependientes de escritor se hacen sólo en el modelo mixto. Usando las nueve primeras muestras de un escritor para entrenar y las tres muestras restantes para la etapa de test.

Los parámetros a estimar para cada uno de los clasificadores son los siguientes:

- Para el clasificador basado en HMM
 - Número de estados
 - Número de escritores en la etapa de entrenamiento
- Para el clasificador basado en NN
 - Número de neuronas de la capa oculta
 - Número de escritores en el entrenamiento

Por tanto, esta combinación de pruebas, según cada clasificador, servirá para optimizar sus parámetros con la finalidad de obtener mayor tasa de reconocimiento. Con respecto al clasificador HMM modelado con la información de la secuencia del trazo, hay que introducir otras variables, como son las dos técnicas de vectorización, que ya

han sido vistas en el capítulo 4, la esqueletización y la detección de contornos. Por tanto, para el clasificador basado en HMM, hay que introducir algunas pruebas más, que son las que a continuación se detallan:

- Vectorización del contorno
 - Contorno exterior e interior para letras de tamaño reducido y normalizado
 - Contorno exterior para letras de tamaño reducido y normalizado
- Vectorización del esqueleto

El número total de combinaciones de estas pruebas se ha reducido con respecto al planteamiento inicial. Debido a los siguientes argumentos:

- I. Se utilizarán sólo las letras normalizadas para realizar las pruebas, ya que poseen una mayor correlación entre letras proporcionales en tamaño que entre las originales [Morales99-pfc], ya que en cada escritor difiere el tamaño de su escritura. Por lo que se puede pensar que ofrecerán mayor tasa de reconocimiento, como ya se demostró con la aplicación de los dígitos manuscritos, en el anexo A.
- II. No se realizarán pruebas con la vectorización del contorno exterior e interior, ya que se han obtenido mejores resultados de reconocimiento en la aplicación de dígitos manuscritos con sólo contorno exterior, que con contorno exterior e interior [Morales99-pfc], según se recoge en el anexo A.

Por tanto, después de observar estos argumentos, el global de las pruebas a realizar son las siguientes:

1. Optimización realizada para HMM con detección de contorno exterior para letras de tamaño normalizado vectorizado por la variación de la dirección de la secuencia
2. Optimización realizada para HMM con detección de contorno exterior para letras de tamaño normalizado, vectorizado en coordenadas polares
3. Optimización realizada para HMM con la técnica de esqueletización para letras de tamaño normalizado
4. Optimización realizada para NN con la técnica del gradiente

B.2.1 Optimización realizada para HMM con detección de contorno exterior para letras de tamaño normalizado vectorizado por su secuencia

Las pruebas realizadas intentan definir el valor óptimo de los dos parámetros del clasificador ya comentados, lo que se ha hecho es fijar uno de ellos y variar el otro parámetro hasta definir su valor que maximiza la tasa de reconocimiento. De esta forma, se ha fijado el número de estado a 60 estados, como primera aproximación, que por similitud para el reconocimiento de dígitos manuscritos es un buen resultado, y se ha

variado el número de muestras utilizadas en el entrenamiento del clasificador. Estas pruebas se han realizado independientes de escritor y se han utilizado las doce repeticiones de cada escritor para el entrenamiento. Los resultados aportados en esta prueba se recogen en la siguiente tabla, y ha sido implementado para las letras minúsculas:

<i>Nº de Escritores</i>	<i>Tasa de reconocimiento</i>
20	75.49%
30	72.02%
40	80.09%
50	79.05%
60	80.42%

Tabla B.1: Tasas de reconocimiento con vectorización del contorno para letras minúsculas variando el número de escritores.

Si se registran estos datos en una gráfica se obtiene:

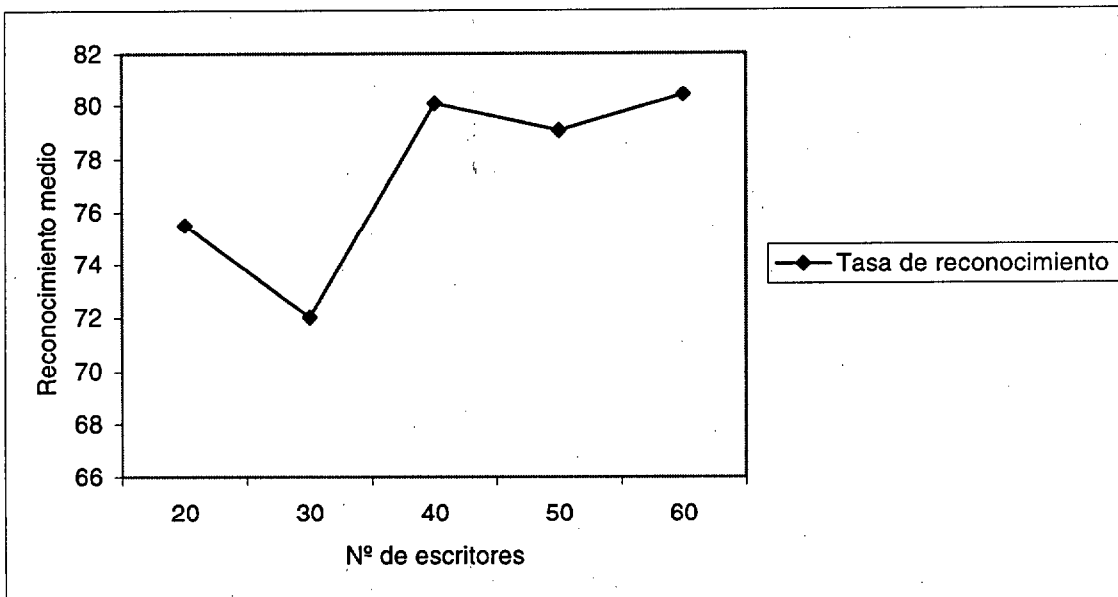


Figura B.1: Tasa de reconocimiento en función del nº de escritores con contornos para letra minúscula.

Se observa de forma clara en la gráfica B.1, para 40, 50 o 60 escritores se obtienen los mejores resultados. Se creó un nuevo modelo para 40 escritores y 60 estados con el fin de verificar el primer resultado, y se obtuvo una tasa de reconocimiento medio del 75%.

El siguiente paso es determinar el valor óptimo del número de estados, para ello se fija el valor del número de escritores a 40, ya que los valores de la tasa de reconocimiento no variaban mucho y la simulación con menos muestras de entrenamiento es más rápida. Y a continuación, se varía el número de estados, resultando las siguientes tasas:

Nº de Estados	Tasa de reconocimiento
40	70.87%
50	73.8%
60	75.04%
70	73.73%
80	74.7%

Tabla B.2: Tasas de reconocimiento con vectorización del contorno para letras minúsculas variando el número de estados.

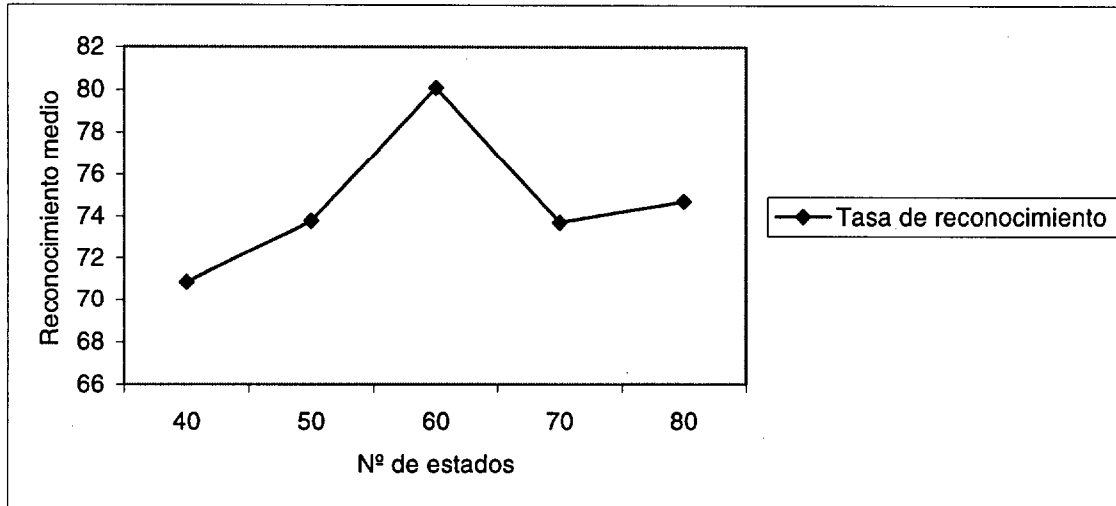


Figura B.2: Tasa de reconocimiento en función del nº de estados con contornos para letra minúscula.

El número de estados que proporciona una tasa de reconocimiento superior es 60, que coincide con el número de estados ideal que se supuso desde el inicio de las pruebas. Por tanto, para la vectorización producida por la técnica del contorno y para letras minúsculas normalizadas, se han obtenido como parámetros óptimos:

Tasa de reconocimiento	75.04%
Nº de estados	60
Número de escritores usados en el entrenamiento	40
Tipo de evaluación de los escritores	Independiente
Tipo de letra	Minúscula

Tabla B.3: Mejor resultado para la optimización realizada para HMM con detección de contornos exterior para letras minúsculas normalizadas.

Las siguientes simulaciones se orientan a medir la tasa de reconocimiento del clasificador de letras mayúsculas y del mixto. Estas simulaciones se hicieron con los parámetros óptimos que se obtuvieron para el clasificador de minúsculas, indicados en la tabla B.3. Con la finalidad de obtener robustez en los resultados se repitieron las pruebas dos veces, variando de forma aleatoria los escritores participantes, y manteniendo su cantidad. Para el clasificador mixto se hizo una prueba dependiente de escritor, entrenando con 9 muestras por escritor y evaluando con las 3 muestras restantes.

Clasificador	Modelo I	Modelo II	Modelo dependiente de escritor
Mixto	69,59%	68,03%	70,45%
Mayúscula	84,9%	86,06%	---
Minúscula	74,39%	75,04%	---

Tabla B.4: Tasas de reconocimiento vectorizando con contornos para las combinaciones de agrupamiento de los patrones de letras manuscritas.

De esta última tabla se deriva, que las mejores tasas de reconocimiento están para letras minúsculas, y es de 75.04%, para las mayúsculas es de 86.06% y para el modelo mixto es de 69.59%.

B.2.2 Optimización realizada para HMM con detección de contorno exterior para letras de tamaño normalizado vectorizado por coordenadas polares

En vistas de los resultados alcanzados con los dígitos manuscritos descrita en el anexo A, se va realizar el estudio para la vectorización del contorno para coordenadas polares, usando como referencia el centro de masas de la letra.

En primer lugar se intentó calcular el número de estados más discriminante para el HMM, partiendo de un mismo entrenamiento con 50 escritores para las letras minúsculas. La tabla y su gráfica se muestran a continuación.

Nº de Estados	Tasa de reconocimiento
40	49.72 %
50	51.51 %
60	48.04 %

Tabla B.5: Tasas de reconocimiento con vectorización del contorno para polares en letra minúscula variando el número de estados.

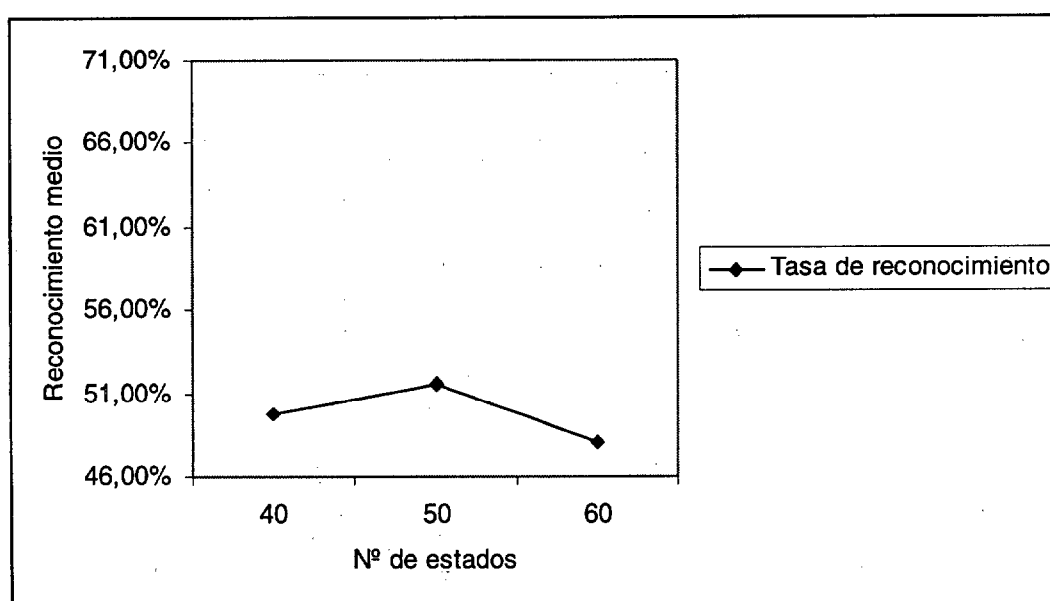


Figura B.3: Tasa de reconocimiento en función del nº de estados con contornos para polares en letra minúscula.

Esta gráfica muestra los valores tan bajos encontrados para este tipo de vectorización, hasta 30 puntos por debajo, por lo que se va a desechar el seguir trabajando con esta técnica.

A pesar de ésto, se probó para las letras mayúsculas, aunque se prevé inicialmente los posibles resultados. La prueba realizada fue exactamente la misma, se fijó el entrenamiento para 50 escritores y se averiguó con que número de éstos, el HMM es más discriminante. Estos valores se recogen en la siguiente tabla:

Nº de Estados	Tasa de reconocimiento
40	56.01 %
50	66.57 %
60	64.52 %

Tabla B.6: Tasas de reconocimiento con vectorización del contorno para polares en letra mayúscula variando el número de estados.

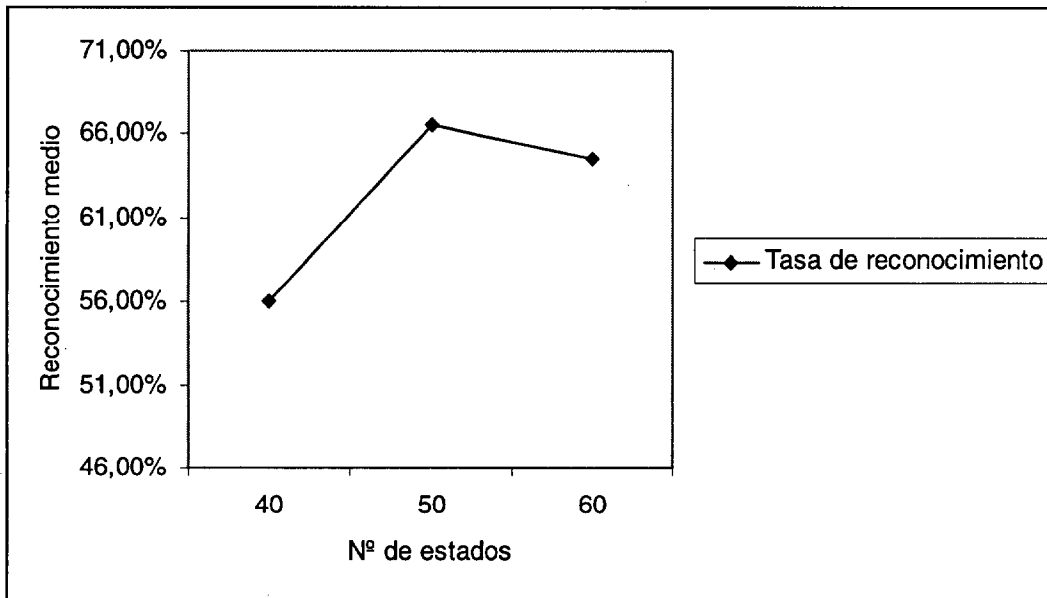


Figura B.4: Tasa de reconocimiento en función del nº de estados con contornos para polares en letra mayúscula.

Para este análisis, también se ha visto que la tasa de reconocimiento para las letras mayúsculas, es también inferior a la obtenida por la secuencia, en aproximadamente 20 puntos, por lo que finalmente se decide dejar este tipo de técnica por su rendimiento tan bajo a la hora de discriminar letras.

B.2.3 Optimización realizada para HMM con la técnica de esqueletización para letras de tamaño normalizado

Como ya se comentó anteriormente se van a utilizar los patrones de letras normalizados y en esta ocasión se van a extraer sus parámetros con la técnica de esqueletización. Las pruebas realizadas siguen el mismo esquema que con la técnica del contorno, se fija el valor del número de estados y se varía el número de escritores para observar cual devuelve la mejor tasa de reconocimiento. El número de estados se ha fijado a 60 y los datos obtenidos se han recogido en la siguiente tabla.

Nº de Escritores	Tasa de reconocimiento
30	62.12 %
40	61.37 %
50	64.78 %
60	66.77 %

Tabla B.7: Tasas de reconocimiento con vectorización por esqueletización para letra minúscula variando el número de escritores.

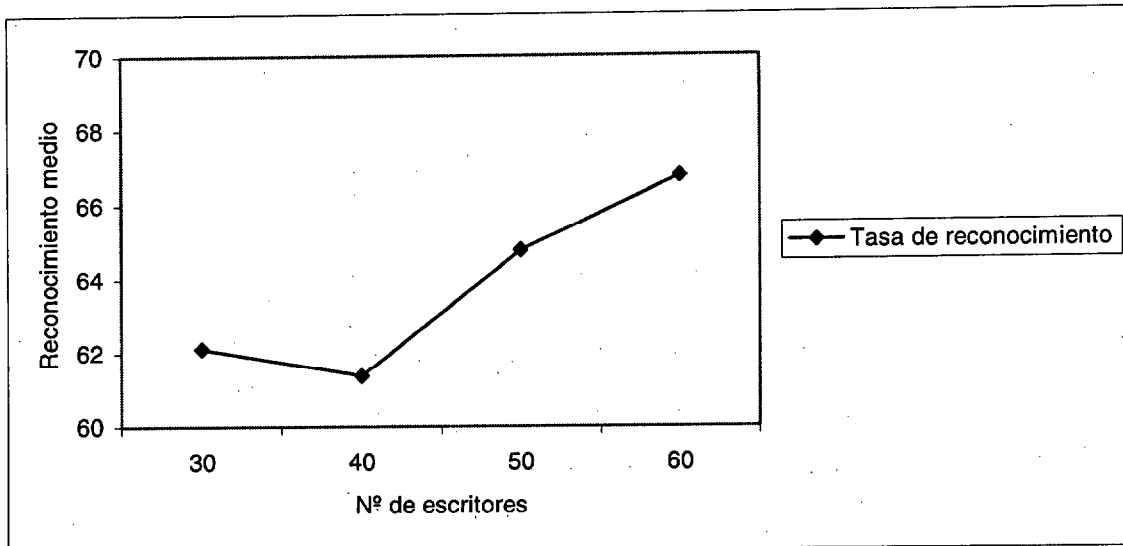


Figura B.5: Tasa de reconocimiento en función del nº de escritores con esqueletización para letra minúscula.

La tasa máxima de éxito fue del 66.77%, por lo que el reconocimiento que ofrece los vectores calculados con esqueletización, queda más de 10 puntos por debajo de las tasas obtenidas con detección del contorno. Ésto despeja dudas sobre la conclusión a tomar, la técnica del contorno clasifica mejor que la técnica de esqueletización usando HMM, por lo que se desecha el realizar más pruebas con el esqueleto de las letras.

B.2.4 Optimización realizada para NN con la técnica del gradiente

En este apartado se trabaja con la información de carácter geométrica, utilizando para ello la técnica del gradiente que extrae este tipo de parámetros, ya explicado en el capítulo cuarto. Como clasificador se utiliza un red neuronal, que hay que optimizar en parámetros para encontrar la mejor tasa de reconocimiento que caracterice a este tipo de vectores. Los parámetros que se tienen que estimar estos patrones, son el número de neuronas de la capa oculta y el número de escritores empleados para entrenar.

La estrategia usada para calcular los parámetros del clasificador es la misma usada en este anexo para otras pruebas. Se realizarán las pruebas para las letras minúsculas, para extrapolar este buen resultado a las letras mayúsculas y el modelo mixto. Y por otro lado, se fijará el valor de las neuronas en la capa oculta y se variará el número de escritores empleados para entrenar. Las tasas de reconocimiento calculadas se incluyen en la siguiente tabla.

Nº de Escritores	Neuronas en la capa oculta						
	40	80	120	160	200	240	280
20	58.44%	61.48%	62.94%	61.21%			
30	61.94%	63.18%	64.4%	64.03%			
40	63.86%	63%	65.18%	66.2%	66.2%	66.62%	66.03%
50	64.66%	65.68%	67.28%	66.47%	66.915	67.03%	67.04%
60	61.69%	66.22%	66.84%	66.4%7	67.4%	66.54%	66.74%

Tabla B.8: Tasas de reconocimiento según las variaciones de número de escritores y neuronas en la capa oculta para letras minúsculas.

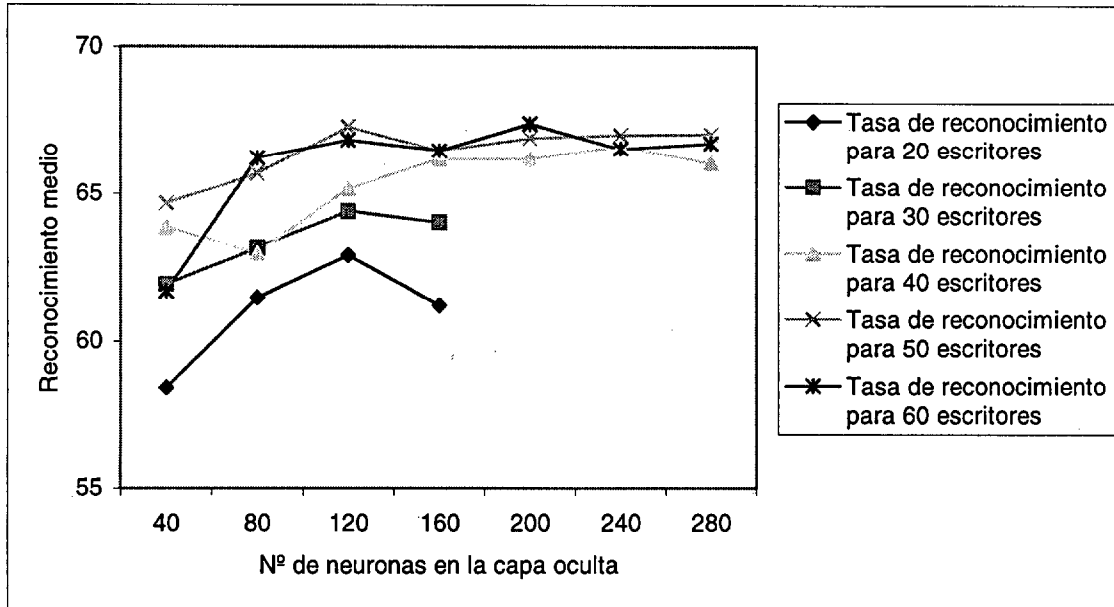


Figura B.6: Tasa de reconocimiento de la técnica del gradiente para letras minúsculas.

Tanto de la tabla B.8 como de la figura B.6, se puede observar como la máxima tasa de reconocimiento se encuentra para un total de 60 escritores y 200 neuronas en la capa oculta, con un resultado del 67.4%. Pero si se presta atención entre el intervalo de 80 y 280 neuronas; y para 40, 50 y 60 escritores, los valores de reconocimiento son bastante parecidos y con muy pequeña varianza. Por lo que cualquier valor dentro de este intervalo podrá servir como válido, en particular, se va a utilizar los parámetros de 50 escritores para entrenamiento y 120 neuronas en la capa oculta, donde se obtuvo una tasa del 67.28%. Esta aproximación se ha hecho para reducir los tiempos de simulación. Para verificar dicha propuesta, se ha repetido el modelo pero con otros escritores diferentes, y se obtuvo una tasa de reconocimiento del 68.22%. En la siguiente tabla se resumen los valores considerados como óptimos.

Tasa de reconocimiento	68.22%
Número de neuronas en la capa oculta	120
Número de escritores usados en el entrenamiento	50
Tipo de evaluación de los escritores	Independiente
Tipo de letra	Minúscula

Tabla B.9: Mejor resultado para la optimización realizada para NN con la técnica del gradiente para letras minúsculas normalizadas.

Una vez definidos los parámetros óptimos se crean los clasificadores de letras mayúsculas y mixto para 120 neuronas en la capa oculta y 50 escritores para el entrenamiento. Para el clasificador mixto se ha realizado la prueba dependiente de escritor, con 9 muestras para entrenar y 3 para evaluar. Los resultados de las simulaciones se presentan en la siguiente tabla.

<i>Clasificador</i>	<i>Modelo I</i>	<i>Modelo II</i>	<i>Modelo III dependiente de escritor</i>	<i>Modelo IV dependiente de escritor</i>
Mixto	68.88%	65.61%	83.16%	85.25%
Mayúscula	77.62%	76.05%	---	---
Minúscula	67.28%	68.22%	---	---

Tabla B.10: Tasas de reconocimiento para NN de las combinaciones de agrupamiento de los patrones de letras manuscritas.

B.3 Conclusiones

Para la base de datos de letras manuscritas descrita en el capítulo tercero, se han realizado cuatro experiencias bien diferentes, y revisando las tablas B.4, B.5, B.7 y B.10, se puede concluir que con gran diferencia, la vectorización de la información secuencial siguiendo la variación de las direcciones por la codificación de la matriz de saltos de 3×3 , a partir del contorno exterior, ofrece mejores resultados que las otras pruebas realizadas.

Ésta es la misma conclusión que se obtuvo para dígitos manuscritos, por tanto, la tendencia es seguir usando este tipo de técnicas para el resto de las aplicaciones manuscritas.

Anexo C

Pruebas realizadas con las palabras manuscritas

C.1 Introducción

En este anexo se va a tratar la discusión planteada al desarrollar las pruebas y experimentos para reconocer las palabras manuscritas aisladas. Se ha partido de la base de datos descrita en el capítulo quinto, así como del clasificador HMM para crear los patrones de las diferentes palabras, a partir de las técnicas de extracción de parámetros de la secuencia, que en dicho capítulo se detallan. Es decir, este anexo va a contener todo el desarrollo para el ajuste de los parámetros del clasificador y de la extracción de parámetros.

C.2 Optimización del clasificador para palabras manuscritas

A continuación se va a desarrollar el estudio de las diferentes técnicas para optimizar el valor del número de estados del HMM, estas técnicas son:

- Seguimiento de la secuencia a través de la codificación de sus direcciones
- Seguimiento de la secuencia en coordenadas polares
- Seguimiento de la secuencia por el valor de los altos, anchos y sus transiciones

El planteamiento que se va a seguir es la optimización de las técnicas de extracción parámetros para un determinado entrenamiento, y su posterior variación para tal efecto. En este estudio se va a incluir la posible mejora que introduzca la corrección de la inclinación de la palabra cuando ésta es escrita de forma cursiva. Es una característica intrínseca de cada escritor, y puede ser que el hecho de su corrección posibilitaría que las palabras fuesen más parecidas entre sí, y podría mejorar el reconocimiento.

Una vez, se determine la mejor técnica para este clasificador HMM, se tratará de averiguar cuando se debe de entrenar para modelar mejor los patrones. Finalmente se tratará de optimizar la técnica seleccionada como la más discriminante.

Estas pruebas se realizaron con una parte de los patrones de palabras, con el fin de ajustar los parámetros y finalmente extrapolarlos al resto. También hay que señalar que para cada experimento se han realizado dos pruebas del mismo (Prueba I y Prueba II), con la finalidad de validarlo y promediario, para evitar picos aislados de las tasas.

C.2.1 Análisis de las técnicas de extracción de parámetros según el HMM

Las variables que se han estudiado en este apartado, intentan descubrir el número de estados para el que mejor funciona el HMM. También se ha barajado la importancia del efecto que produce la letra cursiva y su corrección. En primer lugar se va a realizar esta prueba con la técnica de la secuencia codificada por sus direcciones.

Número de estados	Secuencia			Secuencia-Corrección Cursiva		
	Prueba I	Prueba II	Media \pm Varianza	Prueba I	Prueba II	Media \pm Varianza
25	60.59 %	59.22 %	59.91 % \pm 0.68	40.39 %	58.63 %	49.51 % \pm 9.12
30	67.84 %	64.71 %	66.28 % \pm 1.57	57.65 %	52.75 %	55.20 % \pm 2.45
35	70.78 %	69.61 %	70.20 % \pm 0.58	65.88 %	59.41 %	62.65 % \pm 3.24
40	65.10 %	65.69 %	65.40 % \pm 0.30	69.22 %	73.14 %	71.18 % \pm 1.96
45	68.82 %	74.51 %	71.67 % \pm 2.85	68.82 %	70.39 %	69.61 % \pm 0.78
50	69.22 %	67.84 %	68.53 % \pm 0.69	75.10 %	76.27 %	75.69 % \pm 0.58
55	74.51 %	71.18 %	72.85 % \pm 1.66	69.02 %	72.16 %	70.59 % \pm 1.57
60	74.12 %	69.61 %	71.87 % \pm 2.26	72.94 %	71.37 %	72.16 % \pm 0.79
65	68.63 %	70.78 %	69.71 % \pm 1.08	73.14 %	79.21 %	76.18 % \pm 3.04
70	72.55 %	61.96 %	67.26 % \pm 5.30	72.55 %	69.22 %	70.89 % \pm 1.67

Tabla C.1: Variaciones del número de estados para la secuencia codificada por la matriz de saltos, corrigiendo y sin corregir el efecto de cursiva.

En esta tabla C.1 se ha establecido para la secuencia con y sin la corrección de la inclinación de la palabra, véase en la siguiente figura, como varía el reconocimiento con los números de estados y con la corrección:

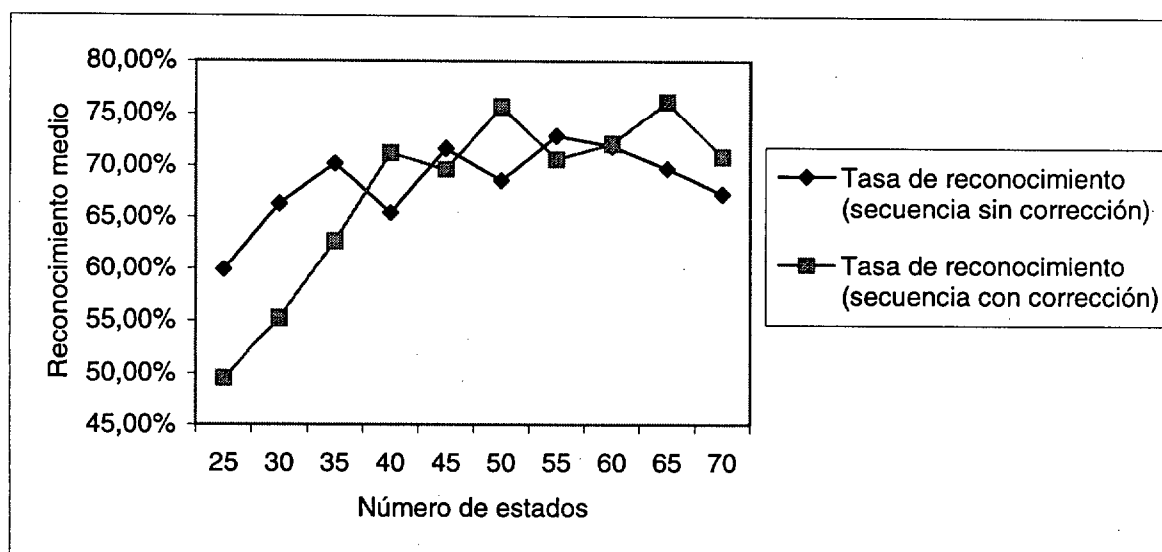


Figura C.1: Tasa de reconocimiento de la secuencia con y sin corrección de la inclinación.

A continuación se realiza el mismo proceso para la secuencia medida en coordenadas polares. En este estudio se han tomado dos criterios como origen de coordenadas, en el centro de masas y el otro en la izquierda inferior. Se presenta en la tabla C.2 y C.3 sus valores y seguidamente, la representación gráfica de estos valores medios.

Número de estados	Polares (con centro de masas)			Polares (con centro de masas) Corrección Cursiva		
	Prueba I	Prueba II	Media ± Varianza	Prueba I	Prueba II	Media ± Varianza
30	49.80 %	50.00 %	49.90 % ± 0.10	49.80 %	47.65 %	48.73 % ± 1.08
40	50.20 %	50.78 %	50.49 % ± 0.29	47.25 %	52.35 %	49.80 % ± 2.55
50	53.73 %	57.06 %	55.40 % ± 1.67	48.24 %	55.10 %	51.67 % ± 3.43
60	56.47 %	57.45 %	56.96 % ± 0.49	58.82 %	53.33 %	56.08 % ± 2.74

Tabla C.2: Variaciones del número de estados para la secuencia en coordenadas polares a partir del centro de masas, corrigiendo y sin corregir el efecto de cursiva.

Número de estados	Polares (sin centro de masas)			Polares (sin centro de masas) Corrección Cursiva		
	Prueba I	Prueba II	Media ± Varianza	Prueba I	Prueba II	Media ± Varianza
30	44.31 %	34.51 %	39.41 % ± 4.90	38.82 %	41.96 %	40.39 % ± 1.57
40	45.49 %	44.51 %	45.00 % ± 0.49	47.45 %	43.14 %	45.30 % ± 2.16
50	46.67 %	50.59 %	48.63 % ± 1.96	47.45 %	49.02 %	48.24 % ± 0.79
60	48.62 %	42.35 %	45.49 % ± 3.14	49.80 %	45.29 %	47.55 % ± 2.25

Tabla C.3: Variaciones del número de estados para la secuencia en coordenadas polares a partir del margen inferior izquierdo, corrigiendo y sin corregir el efecto de cursiva.

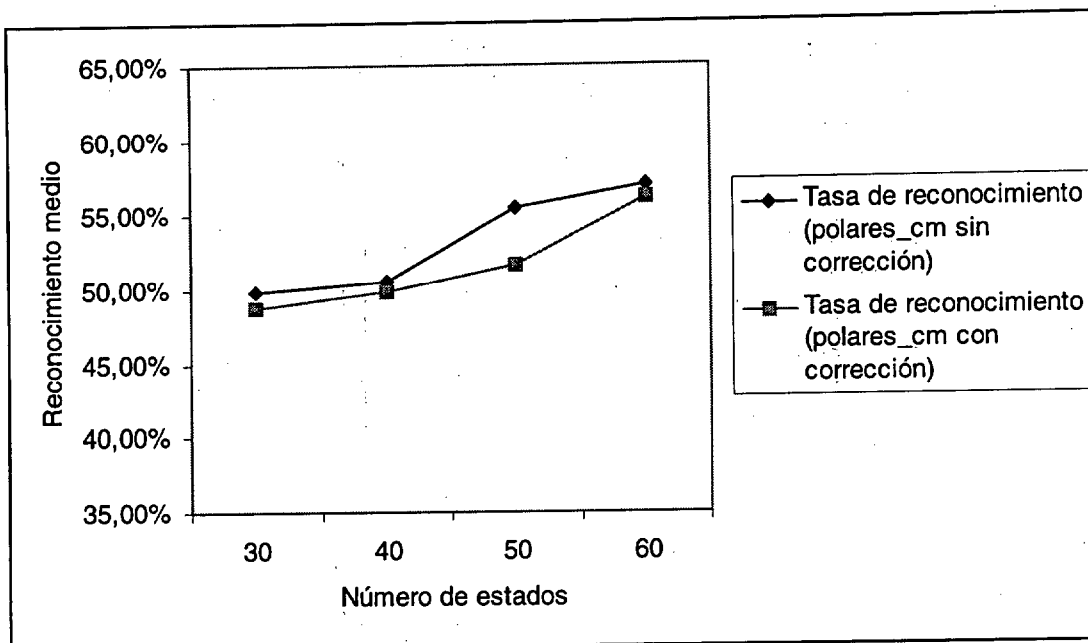


Figura C.2: Tasa de reconocimiento de la secuencia en coordenadas polares a partir del centro de masas, corrigiendo y sin corregir el efecto de cursiva.

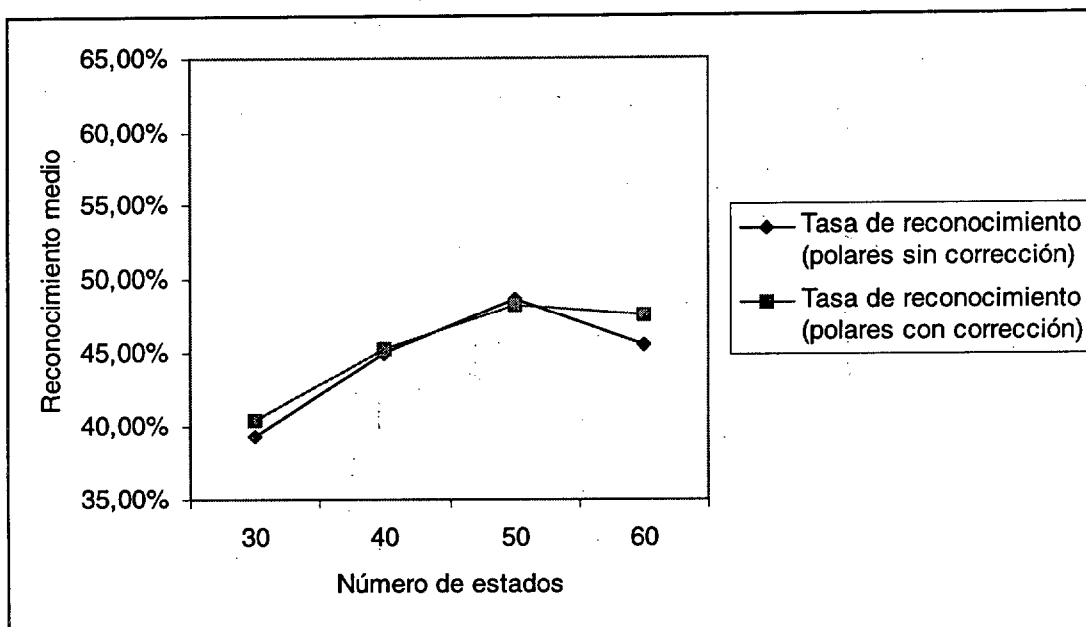


Figura C.3: Tasa de reconocimiento de la secuencia en coordenadas polares a partir del margen inferior izquierdo, corrigiendo y sin corregir el efecto de cursiva.

La técnica que ahora se estudia es la de altos y anchos sin tener en cuenta las transiciones, de esta forma se observa si el hecho de introducir las transiciones es una cuestión discriminante o por el contrario es una mayor carga computacional. Se presentan estos resultados en la tabla C.4, y la representación de la misma en la figura siguiente.

Número de estados	Altos y anchos			Altos y anchos-Corrección Cursiva		
	Prueba I	Prueba II	Media \pm Varianza	Prueba I	Prueba II	Media \pm Varianza
15	69.61 %	71.57 %	70.59 % \pm 0.98	72.94 %	72.75 %	72.85 % \pm 0.10
25	67.84 %	72.35 %	70.10 % \pm 2.26	75.69 %	74.51 %	75.10 % \pm 0.59
35	73.33 %	68.43 %	70.88 % \pm 2.45	72.55 %	68.04 %	70.30 % \pm 2.26
45	73.14 %	70.39 %	71.77 % \pm 1.38	69.80 %	72.35 %	71.08 % \pm 1.28
55	75.29 %	71.18 %	73.24 % \pm 2.05	71.76 %	73.33 %	72.55 % \pm 0.78
65	73.14 %	71.37 %	72.26 % \pm 0.89	73.53 %	70.78 %	72.16 % \pm 1.38

Tabla C.4: Variaciones del número de estados para los altos y anchos del contorno, corrigiendo y sin corregir el efecto de cursiva.

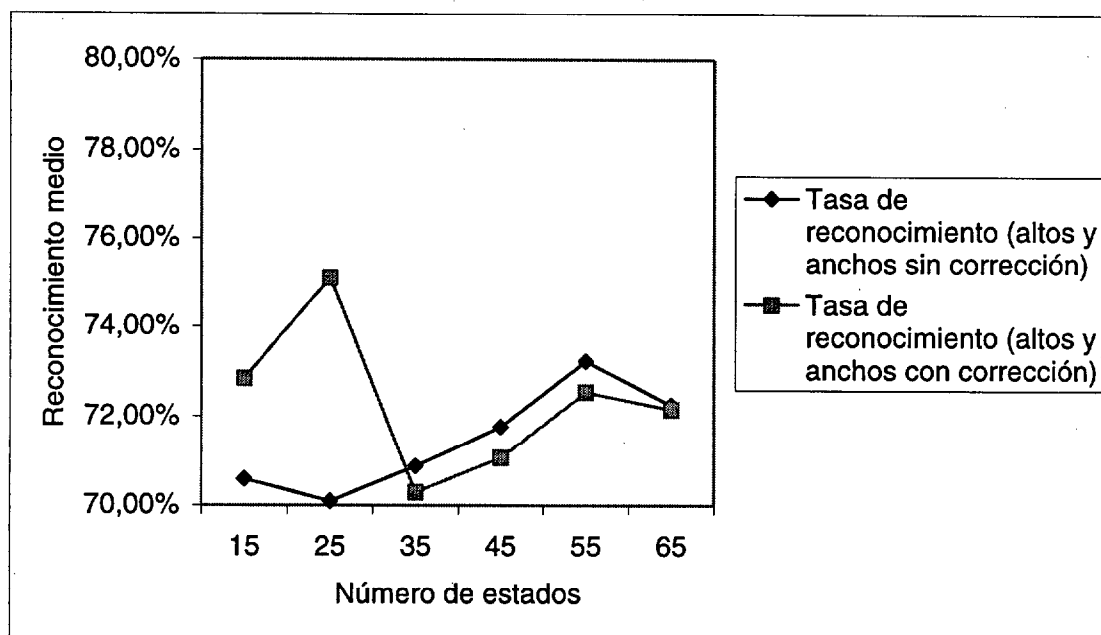


Figura C.4: Tasa de reconocimiento para los altos y anchos del contorno, corrigiendo y sin corregir el efecto de cursiva.

Y finalmente, se representa los resultados de la técnica de altos, anchos y transiciones con su correspondiente gráfica:

Número de estados	Altos, anchos y transiciones			Altos, anchos y transiciones Corrección Cursiva		
	Prueba I	Prueba II	Media \pm Varianza	Prueba I	Prueba II	Media \pm Varianza
15	74.12 %	72.16 %	73.14 % \pm 0.98	72.75 %	74.90 %	73.83 % \pm 1.08
20	73.92 %	77.65 %	75.79 % \pm 1.87	79.80 %	73.53 %	76.67 % \pm 3.14
25	72.16 %	76.86 %	74.51 % \pm 2.35	74.12 %	76.67 %	75.40 % \pm 1.27
30	76.67 %	80.78 %	78.73 % \pm 2.06	74.71 %	79.61 %	77.16 % \pm 2.45
35	75.49 %	74.12 %	74.81 % \pm 0.68	70.78 %	75.88 %	73.33 % \pm 2.55
40	76.86 %	73.33 %	75.10 % \pm 1.77	78.24 %	73.33 %	75.79 % \pm 2.45
45	74.51 %	76.27 %	75.39 % \pm 0.88	75.88 %	74.90 %	75.39 % \pm 0.49
50	73.53 %	74.12 %	73.83 % \pm 0.29	74.31 %	75.10 %	74.71 % \pm 0.40
55	69.61 %	77.45 %	73.53 % \pm 3.92	73.92 %	73.72 %	73.82 % \pm 0.10
60	72.79 %	74.12 %	73.46 % \pm 0.66	72.47 %	70.14 %	71.31 % \pm 1.16
65	74.90 %	75.29 %	75.10 % \pm 0.20	73.73 %	66.47 %	70.10 % \pm 3.63

Tabla C.5: Variaciones del número de estados para los altos, anchos y transiciones del contorno, corrigiendo y sin corregir el efecto de cursiva.

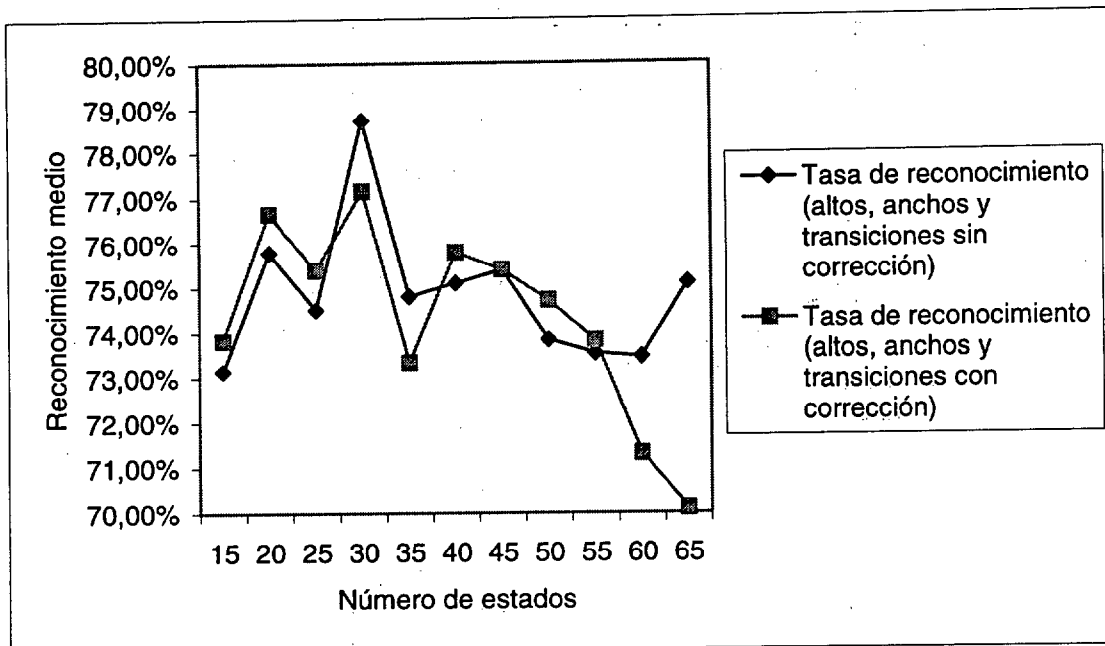


Figura C.5: Tasa de reconocimiento para los altos, anchos y transiciones del contorno, corrigiendo y sin corregir el efecto de cursiva.

De todas estas tablas y sus correspondientes figuras, se puede deducir que el mejor funcionamiento para HMM, los ofrece los altos, anchos y transiciones, tanto si se corrige la inclinación como si no. Por tanto, el hecho de introducir las transiciones aumenta la discriminación, y se seguirá utilizando en los sucesivos experimentos.

A continuación en el mayor éxito alcanzado, le sigue la secuencia, que además mejora con la corrección de la inclinación de la cursiva. Y a más distancia le siguen las polares, que al tener el resultado más bajo, se van a desechar como técnica de obtener parámetros, independientemente de donde se considere el punto de inicio, si con centro de masas o en la parte izquierda inferior.

En líneas generales, y observando las figuras anteriores, se puede decir que la corrección de la inclinación puede ser beneficiosa en el reconocimiento, siempre que se centre bien su rango de actuación, ya que en otras ocasiones funciona peor.

C.2.2 Análisis de las técnicas de extracción de parámetros según el entrenamiento

Para este siguiente experimento, se van a establecer como valores fijos los adquiridos en la prueba anterior, y se va a variar el porcentaje de entrenamiento, para los casos mejores encontrados en el apartado anterior. Estas técnicas son la de altos, anchos y transiciones, que funciona mejor que la de altos y anchos, y la de la secuencia, ya que las polares no han dado buenas tasas de reconocimiento.

El primer caso que se plantea es la variación del entrenamiento para la técnica de altos, anchos y transiciones, usando la corrección de la inclinación de la palabra, y como se determinó anteriormente, para 30 estados en el HMM. El uso de la corrección de cursiva es debido a que en líneas generales, se observa una cierta mejoría con su uso.

Para este experimento en particular, se realizó en cuatro ocasiones, promediando el valor de todas ellas, tal y como se muestra en la tabla C.6 y su representación gráfica.

% de entrenamiento	Altos, anchos y transiciones - Corrección Cursiva				
	Prueba I	Prueba II	Prueba III	Prueba IV	Media \pm Varianza
40	71.15 %	69.84 %	73.28 %	72.62 %	71.72 % \pm 1.33
50	76.67 %	80.78 %	78.63 %	74.12 %	77.55 % \pm 2.46
55	80.44 %	79.33 %	76.00 %	82.44 %	79.55 % \pm 2.33
60	78.50 %	81.00 %	80.25 %	80.75 %	80.13 % \pm 0.98
70	77.00 %	78.33 %	71.33 %	79.67 %	76.58 % \pm 3.18
80	80.50 %	78.00 %	78.50 %	78.50 %	78.88 % \pm 0.96

Tabla C.6: Variaciones del porcentaje de entrenamiento para los altos, anchos y transiciones del contorno, corrigiendo el efecto de cursiva.

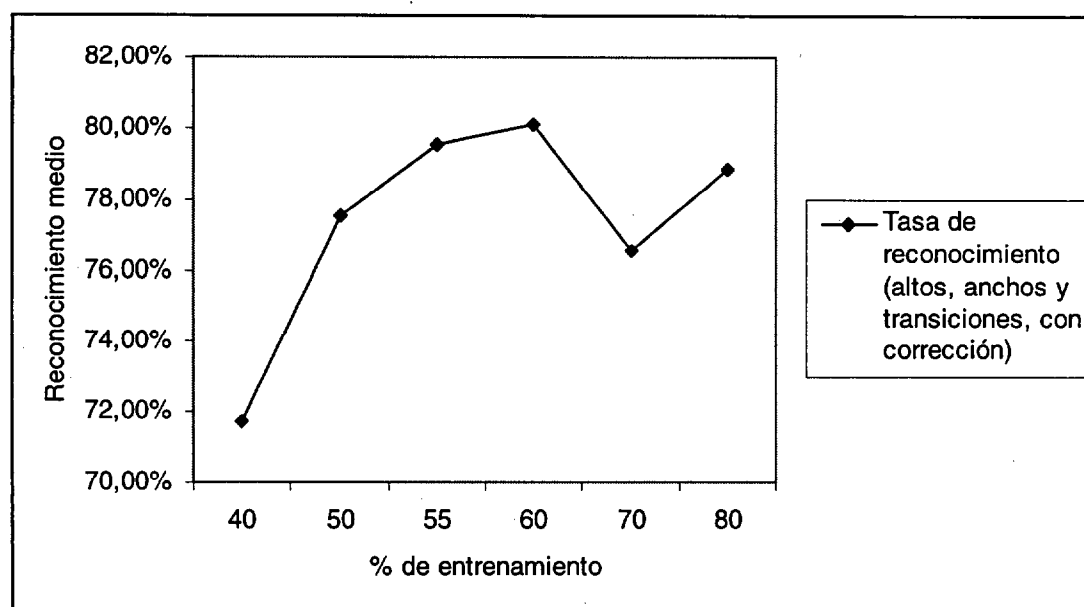


Figura C.6: Tasa de reconocimiento para los altos, anchos y transiciones del contorno, al variar el porcentaje de entrenamiento.

Se puede concluir de esta gráfica, que la mejor tasa de reconocimiento se encuentra para el 60% de la base de datos como entrenamiento, llegando hasta el 80.1%. Para el resto de porcentajes de entrenamiento, se muestra claramente que este valor es inferior.

Por otro lado también se realizó el experimento con la técnica de la secuencia, para los números de estado con mayor éxito, encontrados para 50 y 65 estados (véase la tabla C.1). Los datos calculados se han promediado con dos pruebas, para ambos casos; y la gráfica que describe la variación promediada, se muestra en la figura C.7.

% de entrenamiento	Número de estados = 50			Número de estados = 65		
	Prueba I	Prueba II	Media ± Varianza	Prueba I	Prueba II	Media ± Varianza
40	69.51 %	72.62 %	71.07 % ± 1.5 6	70.33 %	71.48 %	70.91 % ± 0.57
50	75.10 %	76.27 %	75.69 % ± 0.58	73.14 %	79.22 %	76.18 % ± 3.04
60	70.75 %	74.25 %	72.50 % ± 1.75	72.00 %	73.00 %	72.50 % ± 0.50
70	72.67 %	71.00 %	71.84 % ± 0.83	68.00 %	73.67 %	70.84 % ± 2.83

Tabla C.7: Variaciones del porcentaje de entrenamiento para la secuencia del contorno codificando sus direcciones, corrigiendo el efecto de cursiva.

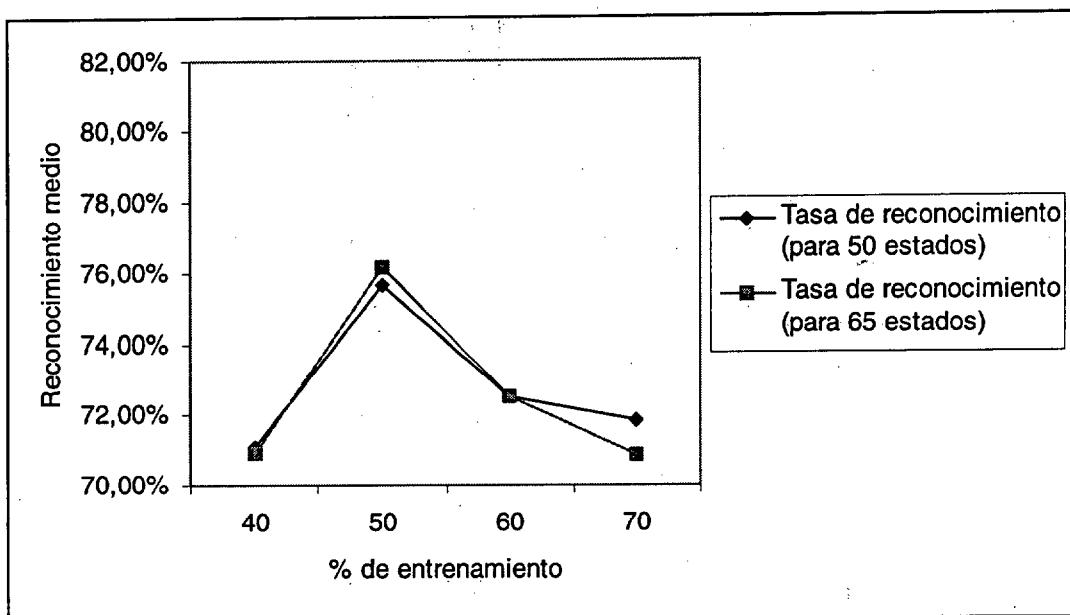


Figura C.7: Tasa de reconocimiento para la secuencia del contorno codificando sus direcciones, al variar el porcentaje de entrenamiento.

Se puede observar que la variación entre ambas es mínima, y mantienen la mejor tasa de éxito para el 50% destinado al entrenamiento. Como su resultado es inferior y no mejora al de altos, anchos y transiciones, se va a desechar esta técnica, siendo ésta última, la más adecuada para el HMM.

C.2.3 Optimización de la extracción de parámetros y su generalización

Una vez determina la técnica que maximiza el resultado, se le va a aplicar a ésta un último experimento. Debido a que los valores numéricos en la extracción de los altos, anchos y las transiciones pueden variar de forma continua en torno a un cierto grado, no se ha fijado en el número de símbolos por estado un valor a este respecto, sino que se delimitó este valor a 32, debido a los buenos resultados alcanzados experimentalmente. Por tanto, la forma de etiquetado a cualquiera de los 32 símbolos de un estado puede ser mayor de 1, como se había fijado. Por tanto, se va a cambiar esta variable por medio de un multietiquetado, y los resultados que se han logrado se describen en la tabla C.8 y su representación en la figura C.8.

Multietiquetado	Altos, anchos y transiciones - Corrección Cursiva con 30 estados y 60% de entrenamiento		
	Prueba I	Prueba II	Media \pm Varianza
1	77.75 %	78.25 %	78.00 % \pm 0.25
2	75.10 %	80.59 %	77.85 % \pm 2.75
3	82.75 %	74.51 %	78.63 % \pm 4.12
4	82.55 %	82.16 %	82.36 % \pm 0.20
5	84.71 %	81.76 %	83.24 % \pm 1.47
6	84.90 %	76.07 %	80.49 % \pm 4.42
8	83.53 %	79.41 %	81.47 % \pm 2.06
10	72.16 %	82.75 %	77.46 % \pm 5.30
12	74.71 %	77.84 %	76.28 % \pm 1.56

Tabla C.8: Variaciones del multietiquetado para altos, anchos y transiciones del contorno, corrigiendo el efecto de cursiva.

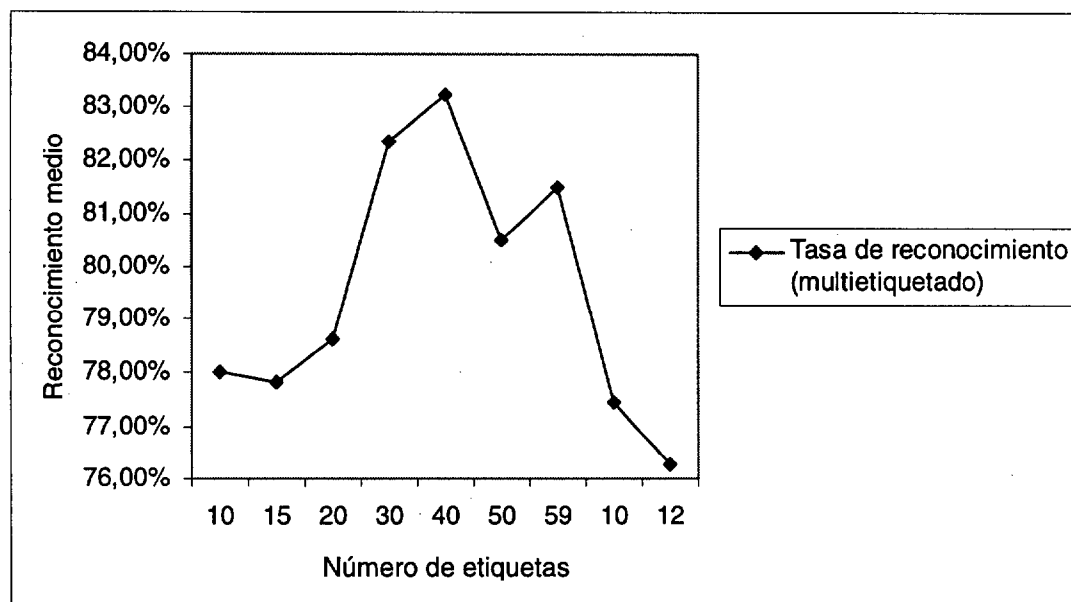


Figura C.8: Tasa de reconocimiento para las variaciones del multietiquetado aplicadas a altos, anchos y transiciones del contorno, corrigiendo el efecto de cursiva.

Se observa de esta gráfica que para $M=4$ y $M=5$ se obtienen los dos mejores valores, maximizando el reconocimiento hasta el 83,2%, con lo que supone una mejora en esta estimación.

Finalmente, y para no dejar dudas, se va realizar este mismo experimento a esta misma técnica, pero sin aplicarle la corrección de la inclinación de las palabras. Las pruebas que se hicieron se muestran en las siguientes tablas, C.9 y C.10. Estas pruebas se repitieron en cuatro ocasiones, promediando las tasas resultantes, y sólo para los dos mejores casos de etiquetado, dado por la tabla C.8.

Multietiquetado	Altos, anchos y transiciones – Sin corrección Cursiva				
	Prueba I	Prueba II	Prueba III	Prueba IV	Media ± Varianza
4	79.75 %	85.75 %	78.00 %	80.50 %	81.00 % ± 2.89
5	86.00 %	82.00 %	78.50 %	74.25 %	80.19 % ± 4.34

Tabla C.9: Resultados del multietiquetado para altos, anchos y transiciones del contorno, sin corregir el efecto de cursiva.

Multietiquetado	Altos, anchos y transiciones – Corrección Cursiva				
	Prueba I	Prueba II	Prueba III	Prueba IV	Media ± Varianza
4	82.75 %	82.25 %	81.50 %	83.50 %	82.50 % ± 0.73
5	84.00 %	78.50 %	87.00 %	83.00 %	83.13 % ± 3.05

Tabla C.10: Resultados del multietiquetado para altos, anchos y transiciones del contorno, corrigiendo el efecto de cursiva.

El resultado de ambas tablas se va a representar en la siguiente gráfica para apreciar las diferencias de las respuestas:

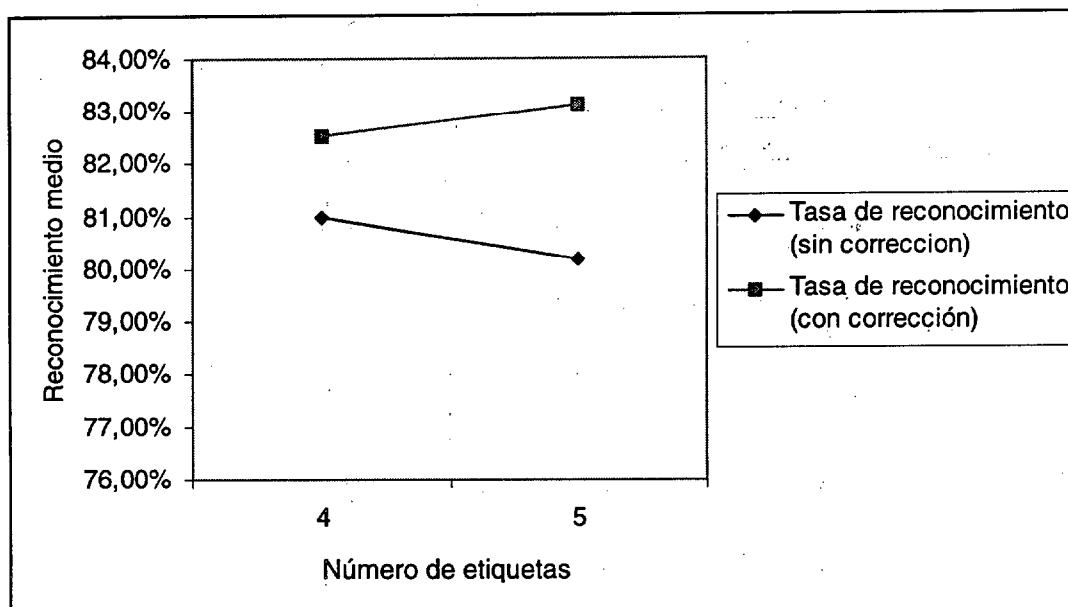


Figura C.9: Tasa de reconocimiento para las variaciones del multietiquetado aplicadas a altos, anchos y transiciones del contorno, con el efecto y sin el efecto de la corrección de la cursiva.

Para finalizar este estudio, sólo queda extrapolar este resultado, logrado con diez palabras, al resto de la base de datos. Para ello, se han agrupado las palabras de diez en diez, hasta completarlas todas. Las pruebas de reconocimiento se han repetido cuatro veces para promediar el resultado, y se recogen en la siguiente tabla.

Número de palabras	Altos, anchos y transiciones - Corrección Cursiva				
	Prueba I	Prueba II	Prueba III	Prueba IV	Media \pm Varianza
10	85.00 %	82.25 %	85.00 %	80.50 %	83.19 % \pm 1.92
15	81.50 %	79.33 %	77.00 %	76.33 %	78.54 % \pm 2.04
20	74.63 %	75.00 %	73.38 %	79.50 %	75.63 % \pm 2.31
30	59.41 %	61.96 %	59.61 %	66.67 %	61.91 % \pm 2.92
40	59.12 %	60.73 %	49.56 %	60.88 %	57.57 % \pm 4.68
50	58.59 %	55.76 %	58.00 %	55.88 %	57.06 % \pm 1.26
59	50.75 %	53.34 %	52.14 %	55.43 %	52.92 % \pm 1.72

Tabla C.11: Resultados de los diferentes agrupamientos de las palabras para su reconocimiento.

La variación del grupo de palabras con respecto a su reconocimiento se observa gráficamente en la siguiente figura:

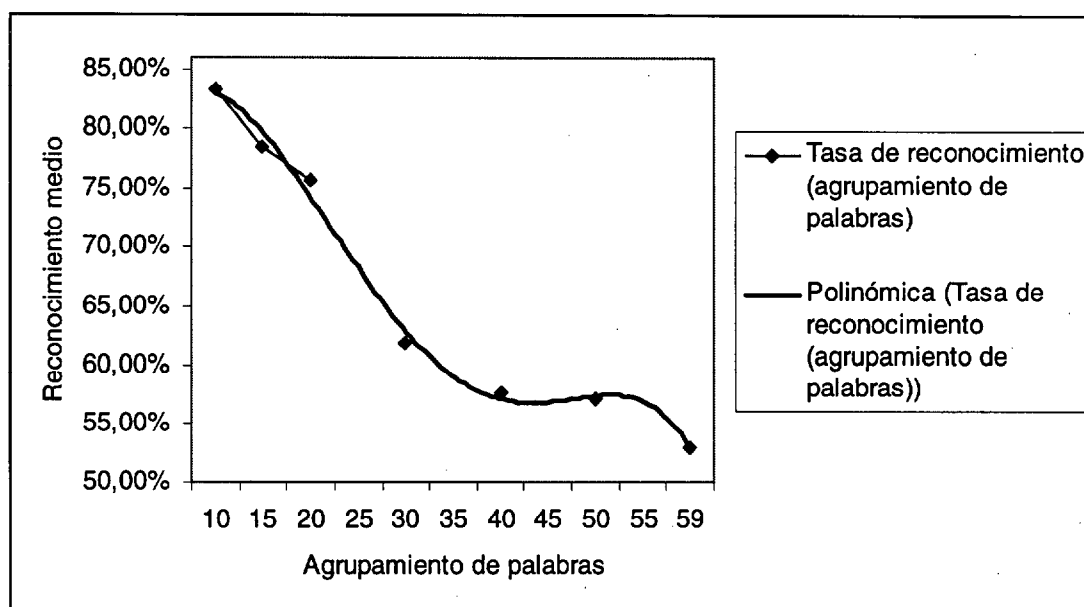


Figura C.10: Tasa de reconocimiento para la agrupación de palabras.

Se observa que la caída en la tasa de reconocimiento va disminuyendo en función del aumento del número de palabras, también hay que tener presente que a medida que las palabras introducidas tienen mayor parecido, la tasa cae en relación con el aumento de la pendiente de la línea polinómica, que cuando las palabras son más diferentes entre sí. Para observar esto, se ha trazado una línea de tendencia polinómica de quinto grado, que se adapta muy bien a las variaciones indicadas por los valores promediados.

Si se observa la figura C.10, al introducir los patrones del veintiuno al veintinueve (muy parecidos entre ellos) además de los anteriores patrones, en el agrupamiento del 30, se produce esa bajada significativa de las tasas. Lo mismo ocurre cuando se introducen el trece (parecido al tres), el catorce (parecido al cuatro) y del dieciséis al diecinueve, en el agrupamiento de 20, que tiene otra caída significativa.

Por el contrario, cuando las palabras introducidas son bastantes incorreladas, la tasa de reconocimiento se mantiene, como en el agrupamiento del 40 al 50, cuyas palabras son del doscientas hasta el novecientas, entre otras. Aquí se observa como la caída del reconocimiento se ve frenada, por este efecto. Por tanto, se puede intuir que para palabras con ciertas diferencias se puede reconocer con mayor éxito.

C.3 Conclusiones

La conclusión de este anexo es bastante clara, la mejor técnica que se adecua al clasificador HMM para 30 estados, un multietiquetado de 5, y entrenado con el 60% de la base de datos, es técnica de altos, anchos y transiciones siempre que se le corrija la inclinación de la palabra, pudiendo llegar en valores promediados al 83,2% de reconocimiento para diez palabras, y un 52,9% para 59 palabras.

Anexo D

Pruebas realizadas con las firmas manuscritas

D.1 Introducción

La aplicación que se quiere reconocer en este anexo son las firmas manuscritas. Para realizar este estudio, se ha partido de la base de datos descrita en el capítulo sexto. El contenido de este anexo consta de las pruebas realizadas con dichos patrones para determinar los parámetros óptimos de los clasificadores a partir de las diferentes formas de obtener la información. Una vez establecidos dichos valores, se podrá observar si mediante el HMM se puede reconocer usando la información del trazo de la secuencia, y comparar la tasa de reconocimiento de todas las pruebas propuestas para ver cual de los sistemas resulta de mayor utilidad en el reconocimiento de las firmas manuscritas.

D.2 Optimización del clasificador para firmas manuscritas

Se persigue continuar con la línea de los anteriores anexos, y por tanto, se sigue clasificando los patrones a partir de la información que éstos contienen. En este caso, son las firmas manuscritas las que están sujetas a estudio.

Se podría indicar que una firma manuscrita caracteriza a una persona, y por tanto, habrán tantas firmas como personas. Para esta aplicación se han creado una base de datos con 60 firmas y 24 repeticiones por firma, descrita en el capítulo sexto. Las firmas se pueden clasificar mediante el reconocimiento o mediante la verificación. En pocas palabras, se puede definir el reconocimiento de firmas como el acto de identificar una firma con respecto a las almacenadas en la base de datos; y la verificación como el hecho de confirmar la identidad de una firma en concreto, siendo conocida la firma y la identidad del firmante, es decir, si la firma es o no de dicho firmante. En el caso de la verificación, se suele trabajar no sólo con la probabilidad dada por el clasificador, sino con la probabilidad de falsa aceptación y la probabilidad de falso rechazo, tal y como se detalla en el capítulo sexto.

En este anexo se ha trabajado para algunas de las técnicas que se plantean el reconocimiento con reservas, es decir, no solo trabajar con la máxima probabilidad dada por el clasificador, sino con los segundos o terceros valores de probabilidad más cercana dada por dicho clasificador. Se suelen coger estos valores de probabilidad cercanos al éxito, como resultados de reserva por si falla el propio reconocimiento. Ésto se realiza, porque si el número de firmas es muy grande, la probabilidad de tener éxito es cada vez menor, y por eso se miran los siguientes casos más probables, como reservas del fallo. El número de reservas va en función del tamaño de la base de datos [Camino99-pfc].

En la clasificación con HMM se van a optimizar los parámetros concernientes al número de estados y el número de repetición por cada patrón de firma usado para el entrenamiento. Este proceso se va a repetir para diferentes grupos de firmantes, desde grupos reducidos hasta el total de la base de datos. Las técnicas empleadas en la extracción de parámetros son las ya usadas, proceso de esqueletización y detección de contorno para obtener esta información secuencial.

En la clasificación con NN, los parámetros a optimizar son el número de neuronas en la capa oculta y el número de repeticiones de cada firma para el entrenamiento, ya que el número de entradas viene dado por el tamaño del vector, en este caso es un valor fijo de 300 parámetros, obtenido a partir de la técnica del gradiente (véase el anexo E), y el número de neuronas de salida.

Para el caso de las firmas no cabe decir si las simulaciones realizadas son dependientes o independientes de escritor, ya que un escritor sólo puede ofrecer una muestra y por tanto, siempre van a ser dependiente de escritor. Es una característica intrínseca de la propia aplicación del reconocimiento de firmas.

Para realizar estas tareas, se va a utilizar información proveniente de la secuencia y la envolvente, además de variaciones en la dirección. Con esta información se elaboran los vectores para discriminar los patrones con los clasificadores HMM para

la información extraída de la envolvente y NN para información de las variaciones de la dirección.

Y por último, el tamaño de las firmas usado para clasificar el HMM, a partir de la información secuencial, se ha tomado reducido y normalizado, ya que si no, el tamaño de los vectores de las firmas es excesivamente grande para su procesado y clasificación.

A partir de todas las consideraciones anteriores, las pruebas que se van a proponer para optimizar los parámetros del clasificador son las siguientes:

1. Optimización realizada para HMM con detección de contorno exterior para firmas
2. Optimización realizada para HMM con la técnica de esqueletización para firmas
3. Optimización realizada para NN con la técnica del gradiente

La primera prueba a ejecutar antes de probar con cada una de las técnicas citadas es la observación de la mayor discriminación de la envolvente reducida o sin reducir. Esto viene inducido por que cada escritor sólo puede realizar una firma, y el tamaño en este caso, al ser una particularidad de cada escritor puede ser una variable a favor del aumento de la discriminación entre patrones. Para comprobarlo, se va a usar el método de la detección de la secuencia a partir de envolvente, realizando cinco pruebas con el mejor número de estados que se ha logrado del HMM. Se han utilizado un total de 20 firmas para esta comprobación, y los datos conseguidos se muestran en la siguiente tabla.

Nº de Prueba	Tamaño de la envolvente	
	Original	Reducida
	Número de estados = 35	Número de estados = 18
I	84.16%	80.83%
II	81.25%	76.26%
III	80.83%	82.50%
IV	84.16%	77.91%
V	80.83%	77.91%
Media ± Varianza	82.25% ± 1.57	79.08% ± 2.26

Tabla D.1: Estudio sobre el tamaño de la envolvente para la secuencia con 20 firmas.

De esta tabla se puede concluir que la envolvente de tamaño original es más discriminante que la reducida. Y por tanto, se va a utilizar este tipo de tamaño, ya que si se reduce en exceso, pierde algo más de información.

Una vez aclarado este paso previo, se va a investigar y averiguar las diferentes tasas de reconocimiento para la técnicas descritas anteriormente.

D.2.1 Optimización realizada para HMM con detección de contorno exterior para firmas

La realización de la optimización del clasificador HMM usando la detección de contornos, ha presentado alguna dificultad, debido al gran tamaño del vector resultante. Por eso se ha optado por calcular la envolvente de la firma, que es una generalización del cálculo del contorno, y se obtiene directamente a través del contorno, para reducir el tamaño que provocan los mayores detalles del contorno, atenuado con la generalización de la envolvente.

Se han empleado diferentes métodos para la vectorización de la envolvente, tanto usando parámetros globales como locales de la misma. Esta vectorización se ha hecho mediante la agrupación de regiones, mediante la parametrización por coordenadas polares [Sánchez01-pfc], usando la misma anterior pero introduciendo el número de píxeles a negro, mediante la secuencia y por último, mediante los altos anchos y transiciones.

La agrupación de regiones se basa en el concepto de centroide, y consiste en agrupar la firma en regiones según un criterio de conexiones de los vecinos horizontales y verticales a ambos lados, y posteriormente se calcula un punto característico del interior de la misma (centro de masa de la región), que en este caso será el centroide. Este centroide se calcula promediando la localización de los píxeles en la región. Se determinarán tantas regiones como el criterio de conexión refleje, y la extracción de los parámetros a un vector se realiza con las coordenadas de los centroides.

Se han realizado una prueba con toda la base de datos, que se ha repetido en ocho ocasiones para determinar el éxito de la misma. Se han utilizado 12 repeticiones por firma para el entrenamiento, y se han utilizado 25 estados en el HMM. Para el test, se utilizaron el resto de las firmas. Las 12 repeticiones elegidas para el entrenamiento se escogieron aleatoriamente, para cada una de las pruebas, y el resto se utilizó para la etapa de evaluación. Bajo estas condiciones las tasas obtenidas fueron;

<i>Nº de prueba</i>	<i>Tasa de reconocimiento</i>
1	71.67%
2	71.67%
3	71.30%
4	67.22%
5	72.22%
6	67.22%
7	69.63%
8	68.52%
Media ± Varianza	69.93% ± 1.82

Tabla D.2: Tasas de reconocimiento con la detección de envolvente por agrupación de regiones.

La siguiente técnica empleada para parametrizar la firma en función de su contorno o envolvente, se basa en el cálculo de las coordenadas polares. A partir de centrar la envolvente y establecer el sistema de referencia como su centro de masas, se vectoriza la envolvente según la disposición de los píxeles respecto a dicho sistema de

referencia, extrayendo de esta manera el recorrido polar de los trazos. Se recorre el trazo para cada posición de píxel, extrayendo el módulo y la fase de las coordenadas polares respecto al sistema de referencia. Por tanto, este sistema de vectorización usa como parámetros del vector, el módulo y la fase a partir de la posición de los trazos respecto a su centro de masas.

Las pruebas realizadas para la mencionada técnica de coordenadas polares, fueron realizadas con 12 repeticiones de las firmas para el entrenamiento, y otras 12 para la etapa de evaluación. El número de estados para el clasificador HMM fue de 35, y se utilizó toda la base de datos (60 firmas), para la ejecución de estas pruebas. El proceso se repitió en ocho ocasiones, cambiando en cada proceso el número de repeticiones para el entrenamiento de forma aleatoria. Las tasas de reconocimiento obtenidas se recogen en la siguiente tabla.

<i>Nº de prueba</i>	<i>Tasa de reconocimiento</i>
1	85.00%
2	86.48%
3	85.74%
4	85.93%
5	85.00%
6	85.93%
7	85.56%
8	83.89%
Media ± Varianza	85.44% ± 0.70

Tabla D.3: Tasas de reconocimiento con la detección de envolvente por coordenadas polares.

Una variación de la técnica anterior, es aumentar el número de parámetros incluyendo el número de píxeles a negro que se encuentran entre el centro de masas y el extremo de la envolvente. Este parámetro junto con el anterior, resulta más discriminativo, ya que bajo las mismas condiciones anteriores, los resultados que se obtuvieron fueron mejores, tal y como se indica en la siguiente tabla.

<i>Nº de prueba</i>	<i>Tasa de reconocimiento</i>
1	90.96%
2	95.30%
3	93.52%
4	92.18%
5	90.94%
6	90.17%
7	94.72%
8	90.24%
9	95.67%
10	91.45%
Media ± Varianza	92.52% ± 2.01

Tabla D.4: Tasas de reconocimiento con la detección de envolvente por coordenadas polares y el número de píxeles a negro.

A continuación sigue la técnica de la vectorización de la envolvente por su secuencia. Este método establece el seguimiento de la envolvente a partir de una matriz

de 3×3, que indica las posibles ocho direcciones a seguir en la secuencia de la firma, de forma que se codifica en un vector, la información de variación en función de dicha matriz y las direcciones que vaya tomando la envolvente.

Las pruebas realizadas se hicieron con las 60 firmas, usando 12 para el entrenamiento y 12 para el test. El clasificador se fijó con 35 estados, para las cinco veces que se hizo la misma prueba, y en todas ellas, variando aleatoriamente las muestras en las dos etapas del clasificador. Los datos que se obtuvieron fueron;

<i>Nº de prueba</i>	<i>Tasa de reconocimiento</i>
1	75.00%
2	76.67%
3	76.80%
4	71.11%
5	75.13%
Media ± Varianza	74.94% ± 2.06

Tabla D.5: Tasa de reconocimiento de la envolvente vectorizada por la secuencia.

Por último, la técnica que queda por analizar es la técnica de altos, anchos y transiciones, al igual que se realizó para las palabras manuscritas. Con esta técnica se captan los altos de la mitad superior y de la mitad inferior, con respecto al centro de masas de la firma. De los anchos se realiza el mismo proceso, pero hacia la izquierda y la derecha. Y finalmente las transiciones que tiene la firma tanto en sus altos como en sus anchos, tal y como se indica en el apartado 6.5 del capítulo sexto.

Las pruebas que se realizaron se repitieron en diez ocasiones, usando la mitad de las muestras de una firma para el entrenamiento y la otra mitad para el test. El número de estados del HMM se determinó en 35 estados, y finalmente se obtuvo;

<i>Nº de prueba</i>	<i>Tasa de reconocimiento</i>
1	92,78%
2	92,96%
3	91,56%
4	94,82%
5	94,39%
6	93,18%
7	90,04%
8	92,76%
9	93,91%
10	90,19%
Media ± Varianza	92,76% ± 1,67

Tabla D.6: Tasas de reconocimiento con la detección de envolvente por altos, anchos y transiciones.

Se puede concluir este análisis que las mejores técnicas, vienen dadas por aquellas tasas mayores, ya que los valores han sido promediados, para evitar valores de pico. En este caso son las técnicas de altos, anchos y transiciones y de coordenadas polares con los píxeles a negro.

En las dos primeras técnicas de reconocimiento para firmas a partir de su envolvente descritas en este apartado, las pruebas realizadas para la optimización de los valores del HMM, se recogen en [Sánchez01-pfc].

D.2.2 Optimización realizada para HMM con la técnica de esqueletización para firmas

Se van a plantear las dos posibilidades de cara al reconocimiento, por un lado el reconocimiento habitual de una firma entre el total de patrones (la más probable), y por otro lado aceptando que la firma a reconocer es una de las más probable.

En primer lugar se va a analizar el caso habitual, que la firma sea la más probable, y para optimizar los parámetros del clasificador en este reconocimiento se va a dividir la base de datos en tres grupos elegidos aleatoriamente. Se van a hacer simulaciones con 10, con 30 y con 60 escritores, y en cada caso se va a variar el número de estados y el número de muestras utilizados en la etapa de entrenamiento.

Para el grupo de 10 firmas se obtuvieron los siguientes resultados;

Nº de Estados	Repeticiones por firmante							
	6	8	10	12	14	16	20	24
25	75.8%	88%	90%	89%	81%	84%	84%	85%
30	73%	84%	90%	89%	88%	80%	88%	91%
35	66%	74%	84%	92%	89%	80%	79%	88%
40	60%	83%	79%	88%	80%	84%	78.3%	80.83%

Tabla D.7: Tasas de reconocimiento de 10 firmantes según el número de estados y el número de repeticiones por firmantes.

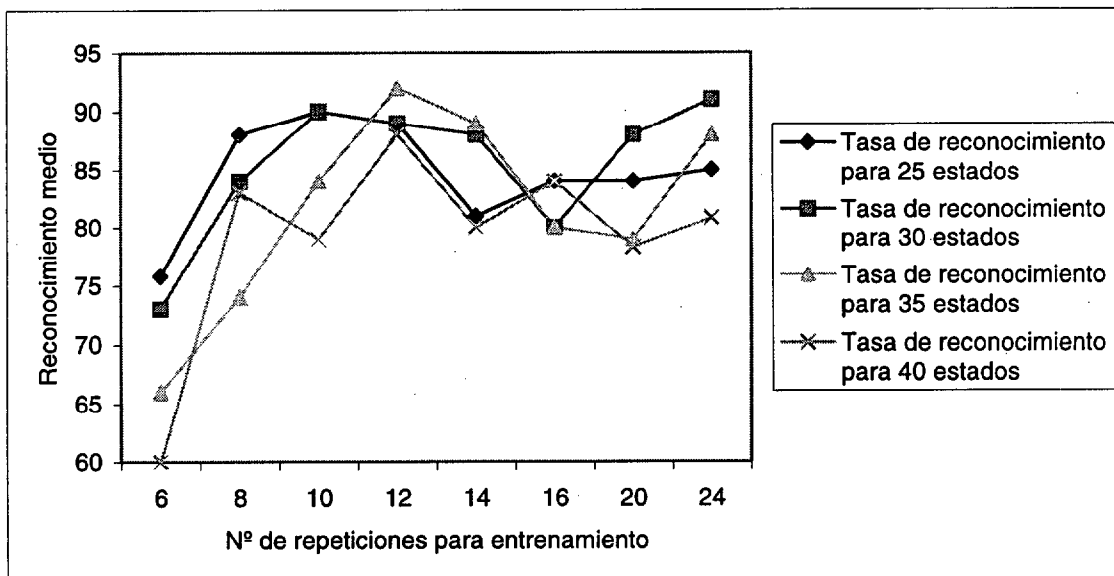


Figura D.1: Tasa de reconocimiento para 10 firmantes con esqueletización para firmas.

Observando estos datos se puede deducir que con pocas repeticiones los resultados de reconocimiento son bajos, pero a medida que estos aumentan la tasa va mejorando. Se puede extraer de la figura D.1, que independientemente del número de estados, con 10 o 12 repeticiones por firmantes, son los que devuelven las mejores tasas de reconocimiento y justamente para 30 y 35 estados se obtienen los picos de mayor éxito. A partir de 14 repeticiones por firmante, la respuesta desciende ligeramente, manteniendo ese valor para el resto de las evaluaciones.

Las siguientes pruebas se han realizado para los mejores valores del clasificador dados por la tabla D.7. Por tanto, para 30 escritores las tasas obtenidas fueron;

Nº de Estados	Repeticiones por firmante	
	10	12
30	68.3%	76.6%
35	70.3%	70.3%

Tabla D.8: Tasas de reconocimiento de 30 firmantes según el número de estados y el número de repeticiones por firmantes.

La tendencia de esta tabla D.8 se ve clara, el número óptimo de repeticiones es de 12, y será el valor utilizado para realizar las pruebas con toda la base de datos, con 60 firmantes.

Nº de Estados	Repeticiones por firmante
	12
30	64%
35	67.5%

Tabla D.9: Tasas de reconocimiento de 60 firmantes según el número de estados.

De todas estas simulaciones, un hecho curioso es la ligera disminución de la efectividad del reconocimiento si se aumenta en demasía el número de repeticiones usadas para la creación de los modelos. La causa que motiva esta disminución de la respuesta es probablemente el sobreentrenamiento o divergencia del clasificador, que hace que las firmas estén excesivamente caracterizadas y ocasiona que las pequeñas variaciones dentro de una firma se tomen como diferencias substanciales, llevando al reconocedor a clasificarla como otra firma diferente.

Del conjunto de las tablas D.7, D.8 y D.9, se puede comentar que a medida que la base de datos va aumentando la tasa de reconocimiento del sistema va disminuyendo. El mejor resultado para la vectorización de esqueletización para toda base de datos se ha obtenido con;

Tasa de reconocimiento	67.5%
Nº de estados	35
Número de repeticiones por firma para el entrenamiento	12

Tabla D.10: Mejor resultado para la optimización realizada para HMM con la técnica esqueletización para reconocimiento de firmas.

El reconocimiento de firmas se va a plantear de otra forma, como los casos más probables que indica el clasificador, no sólo el más probable. Los cánones de la optimización de los parámetros del clasificador, siguen las mismas pautas, se varía el número de estados y las repeticiones por firma. Pero en esta ocasión se han introducido las firmas reservas dentro de las tasas de éxito. Se ha trabajado tanto con una reserva (segunda firma más probable), como con dos reservas (tercera firma más probable). Las tasas calculadas para el grupo de 10 firmantes se encuentran en la siguiente tabla.

Nº de Estados	Repeticiones por firmante			
	8	10	12	14
25 (una reserva)	96%	96%	96%	93%
30 (una reserva)	96%	96%	96%	98%
35 (una reserva)	87%	92%	98%	98%
25 (dos reservas)	99%	96%	98%	97%
30 (dos reservas)	97%	99%	98%	98%
35 (dos reservas)	89%	96%	100%	98%

Tabla D.11: Tasas de reconocimiento con reservas de 10 firmantes según el número de estados y el número de repeticiones por firmantes.

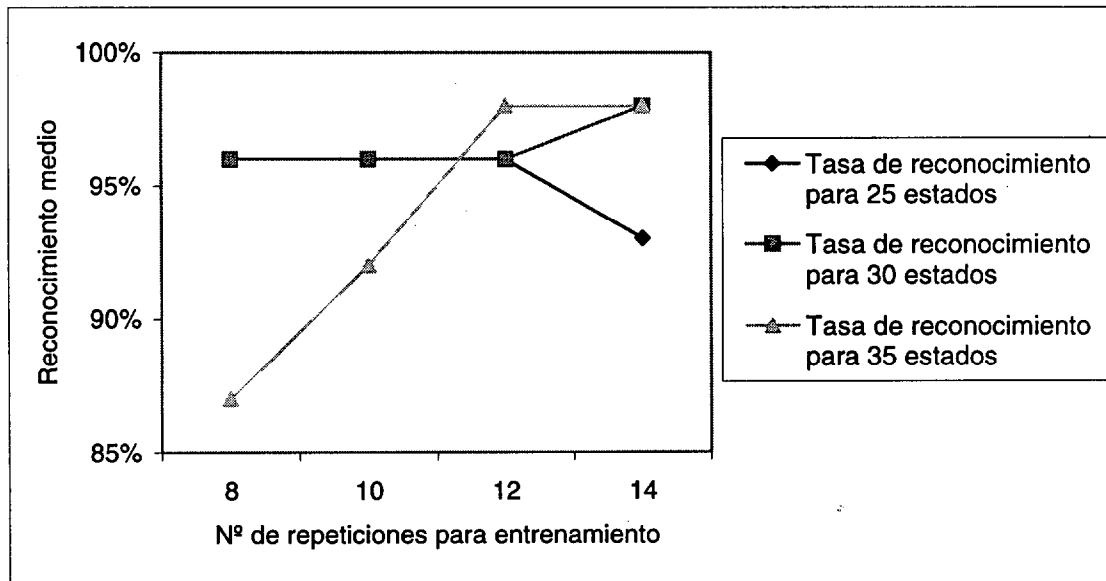


Figura D.2: Tasa de reconocimiento para 10 firmantes con esqueletización para firmas con una reserva.

Se ha indicado en la figura D.2 sólo con una reserva ya que es lo más lógico con solo 10 firmas, y no usar dos reservas aunque en la tabla D.11, se ha indican ambas [Camino99-pfc]. Los resultados están bastante a la par, pero en la figura D.2, queda reflejado como para 35 estados la tendencia con el aumento del número de repeticiones es a incrementar la tasa de éxito, manteniéndose para 12 y 14 repeticiones. Para el resto de número de estados, los valores se mantienen. Por tanto, para el grupo de 30 firmantes con una y dos reservas, se ha escogido 10 y 12 repeticiones por firma, y 30 y 35 estados, resultando;

Nº de Estados	Repeticiones por firmante	
	10	12
30 (una reserva)	81.33%	88%
35 (una reserva)	82.33%	84.33%
30 (dos reservas)	89.66%	92%
35 (dos reservas)	88.66%	88.66%

Tabla D.12: Tasas de reconocimiento con reservas de 30 firmantes según el número de estados y el número de repeticiones por firmantes.

Ya para 60 firmas si es interesante, considerar dos reservas para determinar el reconocimiento de una firma, aunque los resultados para una sola reserva son aceptables. Pero en función de lo reflejado en esta tabla D.12, donde mejor se obtienen resultados es para 12 repeticiones y 30 estados. De esta forma, se deja marcada la simulación con toda la base de datos, aunque se va a incluir las pruebas con 35 estados.

Nº de Estados	Repeticiones por firmante
	12
30 (una reserva)	71.83%
35 (una reserva)	73.5%
30 (dos reservas)	79.83%
35 (dos reservas)	86.16%

Tabla D.13: Tasas de reconocimiento con reservas de 60 firmantes según el número de estados

En esta última prueba, se ha constatado que los mejores resultados para toda la base de datos, se obtiene con 35 estados y 12 repeticiones por escritor, tal y como se recoge en la siguiente tabla.

Tasa de reconocimiento con dos reservas	86.16%
Nº de estados	35
Número de repeticiones por firma para el entrenamiento	12

Tabla D.14: Mejor resultado para la optimización realizada para HMM con la técnica esqueletización para reconocimiento con dos reservas de firmas.

Tanto para reconocimiento de firmas como para reconocimiento con dos reservas, se han obtenido los mismos valores para el mejor funcionamiento del HMM.

D.2.3 Optimización realizada para NN con la técnica del gradiente

Como ya se ha comentado con anterioridad en esta tesis, el sistema de vectorización mediante esta técnica del gradiente, se ha usado con redes neuronales para la clasificación de la información geométrica de las firmas.

Las pruebas efectuadas para optimizar el clasificador se realizaron tanto para reconocimiento como para reconocimiento con reservas. Para el reconocimiento, el

planteamiento de las pruebas realizadas se basa en la variación del número de neuronas de la capa oculta y el número de repeticiones usadas para el entrenamiento de la NN. Al igual que en el apartado anterior, se ha dividido la base de datos en grupos de 10, 30 y 60 escritores, para dar respuesta al reconocimiento. El número de neuronas en la capa oculta se ha variado desde las 130 hasta las 140 neuronas, y las repeticiones se han variado desde las 8 hasta las 24 repeticiones por firma [Camino99-pfc]. Para 10 firmantes elegidos de forma aleatoria, y bajo estas condiciones, se obtuvieron los siguientes resultados;

Nº de neuronas en la capa oculta	Repeticiones por firmante					
	8	10	12	14	16	20
130	82.7%	86%	85%	75%	79%	81%
135	81%	89%	83%	74%	78%	88%
140	84%	87%	86%	74%	78%	82%

Tabla D.15: Tasas de reconocimiento de 10 firmantes según el número de neuronas en la capa oculta y el número de repeticiones por firmantes.

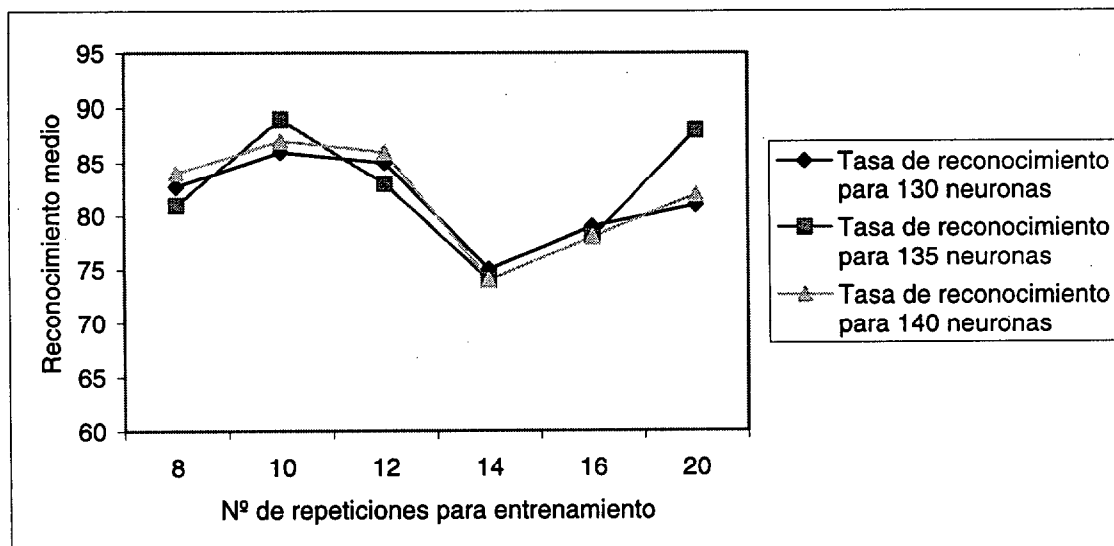


Figura D.3: Tasa de reconocimiento para 10 firmantes con la técnica del gradiente para firmas.

De esta figura D.3, los resultados emanan una tendencia a utilizar pocas repeticiones por firmantes, y la respuesta a estos números de neuronas en la capa oculta es bastante similar, pero de todas destaca la de 135 neuronas, fijando el número de repeticiones entre 10 y 12 por firmante. Con estos parámetros se determina el comportamiento para el grupo de 30 escritores, que también fueron elegidos aleatoriamente de la base de datos.

Nº de neuronas en la capa oculta	Repeticiones por firmante	
	10	12
135	82%	86%
140	81%	85.6%

Tabla D.16: Tasas de reconocimiento de 30 firmantes según el número de neuronas en la capa oculta y el número de repeticiones por firmantes.

De la tabla D.16, se puede extraer que las mejores tasas se obtienen para 12 repeticiones por firmantes, y éste es el valor que se coge para determinar las pruebas a toda la base de datos. Estos resultados para los 60 firmantes se recogen en la siguiente tabla.

<i>Nº de neuronas en la capa oculta</i>	<i>Repeticiones por firmante</i>
	12
135	72.6%
140	72.6%

Tabla D.17: Tasas de reconocimiento de 60 firmantes según el número de neuronas en la capa oculta.

En general, y de las tablas D.15, D.16 y D.17, se puede deducir que a mayor número de firmas a estimar, menor es la tasa de éxito, aunque el descenso no es tan marcado que con la técnica de esqueletización. El mejor dato obtenido para toda la base de datos queda reflejado en la siguiente tabla.

<i>Tasa de reconocimiento</i>	72.6%
<i>Nº de neuronas en la capa oculta</i>	135
<i>Número de repeticiones por firma para el entrenamiento</i>	12

Tabla D.18: Mejor resultado para la optimización realizada para NN con la técnica del gradiente para reconocimiento de firmas.

Para el reconocimiento con reservas, las pruebas planteadas siguen la misma línea que para la técnica de esqueletización y con los mismos parámetros que el reconocimiento de la información geométrica. Por tanto, los datos resultantes para el grupo de 10 escritores tomados aleatoriamente son;

<i>Nº de neuronas en la capa oculta</i>	<i>Repeticiones por firmante</i>		
	8	10	12
130 (una reserva)	99	99	99.5
135 (una reserva)	99.2	99.5	99
140 (una reserva)	99.3	99.7	99.4
130 (dos reservas)	99.4	99.7	99.5
135 (dos reservas)	99.7	99.8	99.5
140 (dos reservas)	99.6	100	99.6

Tabla D.19: Tasas de reconocimiento con reservas de 10 firmantes según el número de neuronas en la capa oculta y el número de repeticiones por firmantes.

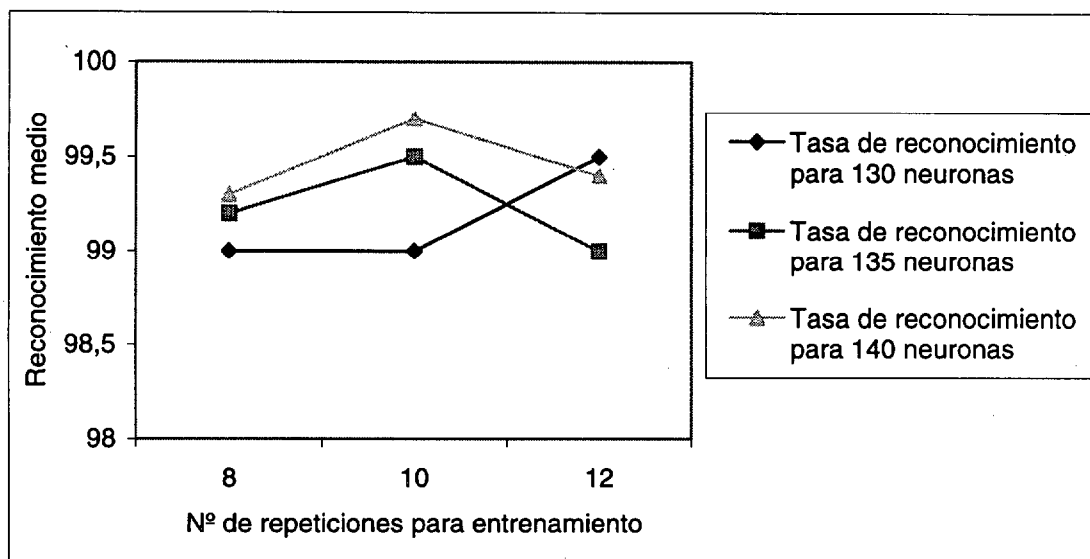


Figura D.4: Tasa de reconocimiento para 10 firmantes con la técnica del gradiente para firmas con una reserva.

De esta última figura, se deduce como, los mejores resultados se centran para más repeticiones por firma, y para más número de neuronas en la capa oculta. Por tanto, se utiliza esta información para realizar las pruebas con el grupo de 30 escritores de firmas.

Nº de neuronas en la capa oculta	Repeticiones por firmante	
	10	12
135 (una reserva)	89.3%	90.6%
140 (una reserva)	89%	92.6%
135 (dos reservas)	91.6%	93.6%
140 (dos reservas)	92%	95%

Tabla D.20: Tasas de reconocimiento con reservas de 30 firmantes según el número de neuronas en la capa oculta y el número de repeticiones por firmantes.

Se plasma de forma clara, como para 12 repeticiones hay una mayor tasa de reconocimiento, y es el dato utilizado para realizar la prueba con toda la base de datos.

Nº de neuronas en la capa oculta	Repeticiones por firmante
	12
135 (una reserva)	82.3%
140 (una reserva)	83.3%
135 (dos reservas)	87%
140 (dos reservas)	86.8%

Tabla D.21: Tasas de reconocimiento con reservas de 60 firmantes según el número de neuronas en la capa oculta.

La conclusión global de las tablas D.19, D.20 y D.21 es que a medida que aumenta el número de patrones a verificar, la dificultad de discriminarlo, aumenta. De

hecho en la siguiente tabla se recoge el mejor dato obtenido con reconocimiento con reservas para 60 firmas para el total de la base de datos.

<i>Tasa de reconocimiento con dos reservas</i>	87%
Nº de neuronas en la capa oculta	135
Número de repeticiones por firma para el entrenamiento	12

Tabla D.22: Mejor resultado para la optimización realizada para NN con la técnica del gradiente para reconocimiento con reservas de firmas.

D.3 Conclusiones

La conclusión de este anexo se centra en dos tendencias. La primera es que la información que parte de la envolvente se puede utilizar para reconocer aplicaciones de firmas manuscritas, y según marcan los resultados con bastante éxito en comparación con la literatura actual. Y la segunda, que parte de esta primera premisa, es que de todas las pruebas propuestas con los diferentes medios de extracción de parámetros y clasificadores, el mejor resultado, y con diferencia, utilizando toda la base de datos, se obtiene para la envolvente de la firma y el uso de su información que procede de ella, sin entrar en detalles de las diferentes técnicas aplicadas.

Anexo E

Técnica del gradiente para clasificación de parámetros geométricos

E.1 Introducción

Este anexo va a tratar la parametrización, la vectorización y la clasificación de la información de carácter geométrico. Este tipo de herramientas queda fuera del ámbito del objetivo de esta tesis, pero parece interesante realizar un somero estudio sobre el mismo, para que se pueda establecer ciertas conclusiones al respecto, en comparación con las técnicas de extracción de parámetros a partir de la secuencia del trazo que se proponen en esta tesis.

Se van a utilizar las bases de datos descritas en el capítulo tercero y sexto, para determinar sus características geométricas. Este tipo de parámetros es la herramienta más usada en el reconocimiento de las aplicaciones manuscritas encontradas en la literatura [Kaewkongka99] [Cho97] [Seiler96] [Busel94], y en el reconocimiento de formas en general [Hwang97] [Neubauer96].

Una vez calculada la imagen con características geométricas, ésta se vectoriza, es decir, se pasa de dos dimensiones (imagen) a una sola dimensión (vector), para situarlo a la entrada del clasificador. Finalmente, se pasa al proceso de reconocimiento, donde este clasificador debe ser adecuado a los vectores con esta información geométrica para optimizar la tasa de reconocimiento, como es la red neuronal (NN).

E.2 Extracción de características: Técnica del gradiente

La elección de un grupo representativo de características es un punto muy importante en el problema de reconocimiento de formas, y de aplicaciones manuscritas. La representación del patrón de una imagen manuscrita es transformada en un juego de datos de alto nivel representativo, pasando a través de una serie de etapa de preprocesado.

Generalmente, una de las técnicas más utilizadas para la extracción de parámetros es a partir de la geometría o la topología de una imagen. En este caso, se ha utilizado un detector de variación de la dirección de los trazos mediante las máscaras de Kirsch [Cho97][Lee96]. Esta técnica aprovecha el hecho de que los patrones manuscritos está compuestos por trazos de líneas, para detectar cada uno de estos segmentos y utilizarlos por separado como características representativas de la imagen. En concreto, el detector por máscara de Kirsch, busca en la imagen las líneas horizontales, verticales, diagonales inclinadas hacia la derecha y diagonales inclinadas hacia la izquierda. El procedimiento que se sigue para la detección de estas líneas es mediante el cálculo del gradiente de los píxeles que ocupan una cierta posición (i,j), utilizando la siguiente expresión;

$$G(i, j) = \max \left\{ 1, \max_{k=0}^7 [|5S_k - 3T_k|] \right\} \quad (\text{E.1})$$

donde,

$$\begin{aligned} S_k &= A_k + A_{k+1} + A_{k+2} \\ T_k &= A_{k+3} + A_{k+4} + A_{k+5} + A_{k+6} + A_{k+7} \end{aligned} \quad (\text{E.2})$$

La función G(i,j) es el gradiente del píxel de la posición de la imagen (i,j), y los subíndices de A son evaluados en módulo 8, y A_k ($k=0,1,\dots,7$) son los vecinos del píxel de la posición (i,j) definidos como se muestra en la siguiente figura;

A_0	A_1	A_2
A_7	P	A_3
A_6	A_5	A_4

Figura E.1: Definición de la vecindad de un píxel.

El siguiente paso será la extracción de las características direccionales de los patrones manuscritos. Estos parámetros de gradiente en la dirección horizontal (H), vertical (V), diagonal derecha (R), y diagonal izquierda (L) serán calculadas con las siguiente ecuaciones;

$$\begin{aligned}
 G(i, j)_H &= \max\{|5S_0 - 3T_0|, |5S_4 - 3T_4|\} \\
 G(i, j)_V &= \max\{|5S_2 - 3T_2|, |5S_6 - 3T_6|\} \\
 G(i, j)_D &= \max\{|5S_1 - 3T_1|, |5S_5 - 3T_5|\} \\
 G(i, j)_I &= \max\{|5S_3 - 3T_3|, |5S_7 - 3T_7|\}
 \end{aligned}
 \tag{E.3}$$

La implementación de la ecuación E.3 a una matriz, es lo que representará las máscaras de Kirsch para realizar el filtrado. No son más que matrices de 3x3, que barren la imagen del patrón manuscrito. Estas matrices quedan de la forma;

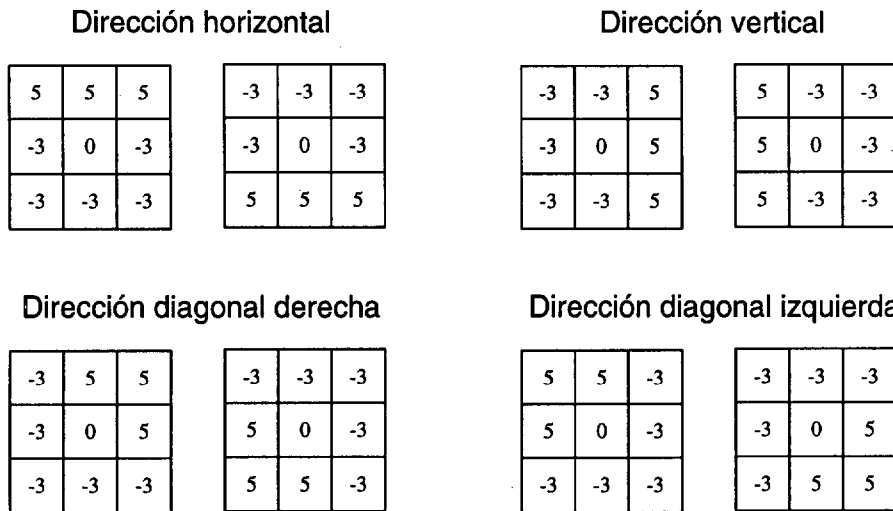


Figura E.2: Máscara de Kirsch usadas para la extracción de parámetros direccionales.

El resultado de este filtrado es otra imagen, una por cada máscara para la detección de las variaciones direccionales de los trazos de la escritura. Según se puede apreciar en la figura E.3, se parte de la imagen inicial, para este caso es una firma, como se ilustra en la figura E.3. Se determina el resultado de la máscara de Kirsch, obteniéndose cuatro matrices con las variaciones del gradiente oblicuo derecho, oblicuo izquierdo, horizontal y vertical. Además se incluye una quinta matriz a los parámetros, que es la reducción de la matriz inicial de la firma.

El paso final es la reducción de las matrices de gradiente, así como una reducción de la matriz reducida, con el propósito de disminuir la cantidad de información a procesar (véase la figura E.3).

E.3 Vectorización de los parámetros

La vectorización se obtiene a partir de las imágenes reducidas, que contiene las variaciones direccionales de los patrones manuscritos. Se traslada de la dimensión del plano, donde se encuentran los trazos manuscritos, a una sola dimensión, para contener en un vector los datos de las variaciones en la dirección. Si se analiza el resultado de las matrices reducidas, se obtienen dos valores, blanco o negro, que se puede binarizar en un vector de "1" y "0", recorriendo estas matrices por filas de izquierda a derecha y por

columnas de arriba a bajo, se consigue como resultado un vector con las características de variaciones de la dirección de los patrones analizados.

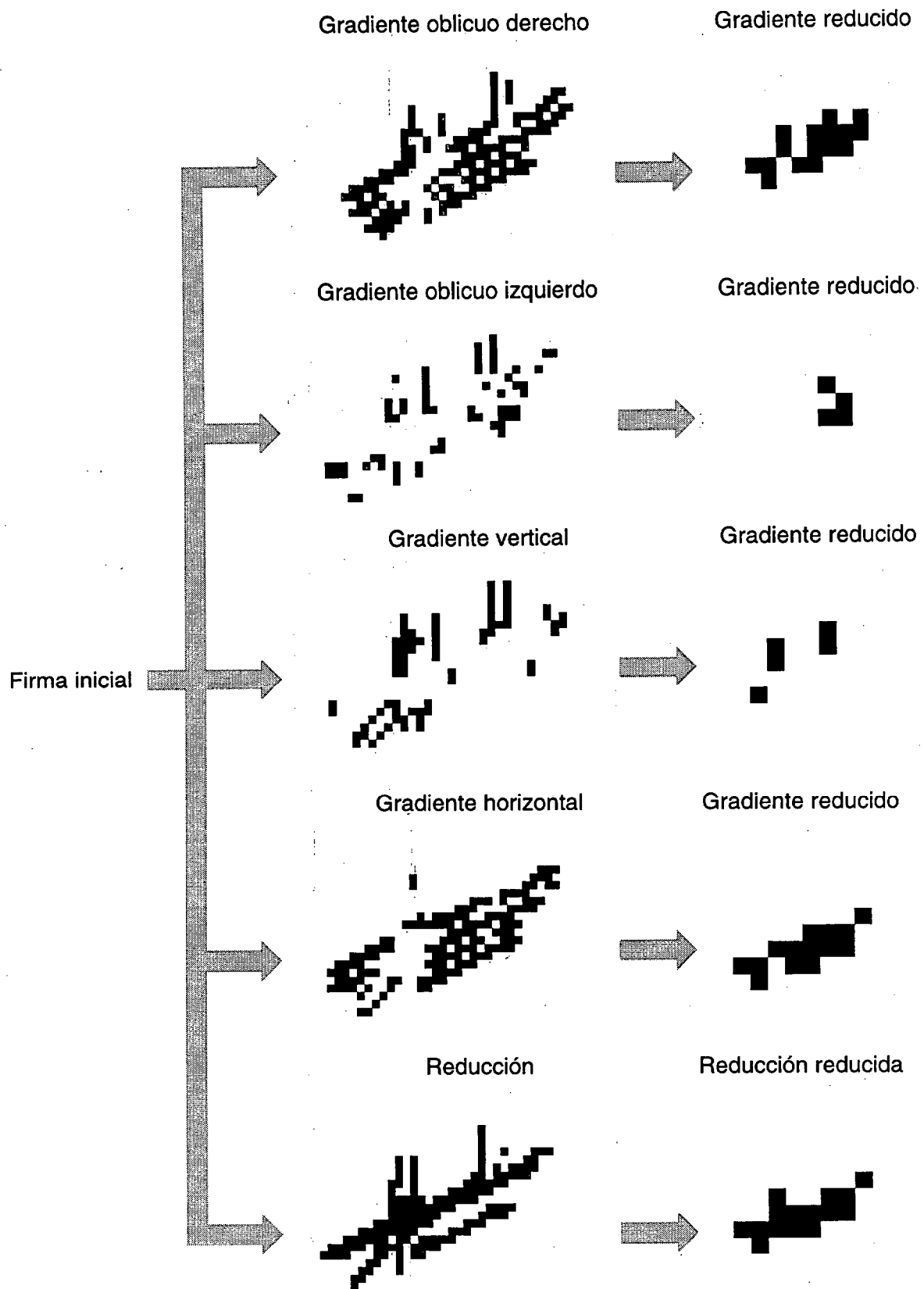


Figura E.3: Proceso de extracción de las características de una firma.

E.4 Clasificador de parámetros geométricos

Para implementar el clasificador que use estos vectores con parámetros basados en las variaciones de la dirección, se va a utilizar una red neuronal del tipo perceptrón multicapas.

Este clasificador tendrá dos capas, una capa oculta y la otra capa de salida, con tantos nodos de salida como número de patrones contenga el conjunto de elementos a reconocer. Se tomará como variable el número de nodos en la capa oculta, para optimizar la tasa de reconocimiento que puede ofrecer el clasificar. La capa de entrada va a depender de la longitud de los vectores de parámetros, que contiene la información de las variaciones en la dirección.

El entrenamiento de la red neuronal se aplica con la finalidad de que los pesos de las distintas ramas se vayan adaptando hasta llegar a un error umbral previamente establecido en la red. Para este propósito el algoritmo utilizado es el *Back-propagation*.

La inicialización de las variables de la red neuronal, ya definidas en el apartado 2.3.3.5, como son los pesos de la red y su valor de offset, se inicializarán también de tal manera, que converja más rápidamente al umbral preestablecido.



