



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

**FACULTAD DE ESTUDIOS SUPERIORES
ARAGÓN**

**RECONOCIMIENTO DE PATRONES EN
IMÁGENES DIGITALES**

T E S I S

**PARA OBTENER EL TÍTULO DE
INGENIERO EN COMPUTACIÓN**

P R E S E N T A :

CLAUDIA ROMERO LUNA

ASESOR: M. EN C. MARCELO PÉREZ MEDEL

MÉXICO 2008





Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

AGRADECIMIENTOS

A la mujer más valiente, noble, sensata, íntegra y amorosa que he tenido la dicha de tener por madre.

A mi padre, quien a pesar de todo siempre me ha apoyado.

Al hombre que inspira mis sueños y mis realidades, mi camino, mi amor Julio Uriel. Te amo.

A mi prima, que más que eso es mi hermana, Wendy. Siempre necesito que cuides de mí.

A mi pequeño Said, el angelito que siempre estará en mi corazón.

A la persona que a pesar de todo creyó que si podía y me brindó todo su apoyo, mi profesor Marcelo.

A todos mis familiares y amigos, que yo sé que siempre están conmigo de corazón, y a quienes yo en verdad quiero.

ÍNDICE

INTRODUCCIÓN	5
1. CONCEPTOS BÁSICOS	7
1.1 Sistema de Reconocimiento Automático	7
1.2 ¿Qué es un Patrón?	8
1.3 Universo de Trabajo (UT)	12
1.4 Clase	13
1.5 Vector de Características	13
1.6 Selección de Variables	15
1.7 Función Discriminante	16
1.8 Segmentación y Extracción de Características	18
1.9 Clasificación	19
1.10 Enfoques y Problemas	20
1.11 Introducción al Análisis de Imágenes	21
1.11.1 Imagen Digital	22
1.11.2 Fundamentos del Color	25
1.11.3 Dispositivos de Captación de Imágenes	27
1.11.4 Formato de Imágenes	29
1.12 Procesamiento Digital de Imágenes (PDI)	31
1.12.1 Conversión de los Niveles de Gris	31
1.12.2 Transformaciones Geométricas	32
1.12.3 Transformaciones del Histograma.....	32
1.12.4 Filtrado	33
1.12.5 Paletas de Colores	37
2. SEGMENTACIÓN Y EXTRACCIÓN DE CARACTERÍSTICAS	39
2.1 Segmentación de Imágenes	39
2.1.1 Umbralización Basada en el Histograma	41
2.1.2 Agrupación de Píxeles.....	44
2.1.3 Cálculo de Bordes	45
2.1.4 <i>Thresholping</i>	45
2.1.5 <i>Watersheed</i>	46
2.1.6 Operaciones Morfológicas	47

2.1.7	Imágenes en Color.....	50
2.2	Extracción de Características	50
2.2.1	Cuantificación	51
2.2.2	Características Geométricas	52
2.2.3	Características Densitométricas.....	58
2.2.4	Características Cromáticas	59
3.	CLASIFICACIÓN	62
3.1	Clasificación Supervisada	63
3.2	Clasificación no Supervisada	67
4.	RECONOCIMIENTO DE ROSTROS BASADOS EN IMÁGENES ..	69
4.1	Tecnología de Reconocimiento de Rostros	70
4.2	<i>Eigenfaces</i> y <i>Eigenfeatures</i>	72
4.3	Generación de <i>Eigenfaces</i>	73
4.4	Reconocimiento de Rostros en 3D	76
4.5	Problemas en el Reconocimiento de Rostros	78
4.6	Enfoques para solucionar los problemas	80
4.7	Aplicaciones	81
	CONCLUSIONES	83
	LISTADO DE FIGURAS	84
	BIBLIOGRAFÍA	86

INTRODUCCIÓN

El **Reconocimiento de Patrones** es un área de la tecnología conocida como aprendizaje de máquina (*machine learning*) o aprendizaje automático. También se le llama lectura de patrones, identificación de figuras o reconocimiento automático de formas. De cualquier manera, se refiere al reconocimiento de patrones en señales, por lo que no sólo es un campo de la informática sino un proceso fundamental que se encuentra en casi todas las acciones humanas.

Jesús Carrasco Ochoa del Instituto Nacional de Astrofísica Óptica y Electrónica opina que el Reconocimiento de Patrones: *“es una ciencia que se ocupa de los procesos sobre ingeniería, computación y matemáticas relacionados con objetos, con el propósito de extraer información que permita establecer propiedades de o entre conjuntos de dichos objetos”*. [11]

El Reconocimiento de Patrones abarca todo un conjunto de técnicas y herramientas matemáticas con aplicación en numerosas disciplinas: ingeniería, medicina, biología, politología, psicología, economía, química, etc. Sin embargo, su desarrollo histórico ha estado estrechamente unido a la informática, de una manera muy especial, suele vincularse a la denominada I.A. - Inteligencia Artificial (A.I., por sus siglas en inglés). Concretamente, constituye la base de la visión artificial o visión por computadora (*computer vision*) una especialidad de la I.A. cuyo propósito es programar una computadora para que “entienda” una escena o las características de una imagen. [1]

El Procesamiento Digital de Imágenes (PDI) es un conjunto de métodos para obtener y modificar información de una imagen. Y precisamente, una de las razones por la cual se ha transformado en una herramienta tan importante para el análisis de imágenes es que los datos son almacenados en forma digital, lo cual permite una manipulación mucho más amplia de los mismos que no son posibles por medios fotográficos. En el análisis de imágenes se toman imágenes digitales como entradas del sistema.

Algunas aplicaciones del Reconocimiento de Patrones son: reconocimiento de voz, clasificación de documentos (por ejemplo *spam/no spam**), reconocimiento de escritura y caracteres ópticos, identificación de sucesiones de ADN, lectores de códigos de barras, lectores de huellas digitales, reconocimiento de rostros humanos, en aplicaciones médicas y científicas, etc.

* Se llama *spam* o correo basura a los mensajes no solicitados, habitualmente de tipo publicitario, enviados en cantidades masivas que perjudican de una u otra manera al receptor. Aunque se puede hacer por distintas vías, la más utilizada entre el público en general es la basada en el correo electrónico.

Para un sistema de visión por computadora, los objetivos incluyen: detección, segmentación, localización y reconocimiento de ciertos objetos en imágenes (por ejemplo, rostros humanos); evaluación de resultados; registro de diferentes imágenes de una misma escena u objeto; seguimiento de un objeto en una secuencia de imágenes; mapeo de una escena para generar un modelo tridimensional de la misma (tal modelo podría ser usado por un robot para navegar por la escena); búsqueda de imágenes digitales por su contenido, etc.

Son los avances en el procesamiento de señales digitales, los sistemas computacionales y la tecnología del circuito integrado los que han hecho posibles sistemas con decisión automatizada y funciones similares. Los algoritmos para el reconocimiento automático de patrones son generalmente implementados en una computadora o sobre una pieza especial de *hardware** que permite llevar a cabo el procedimiento con la menor intervención humana posible o sin ella.

Suena lógico el buscar una forma para diseñar y construir mecanismos que puedan reconocer patrones porque de una forma u otra es uno de los mecanismos naturales que utilizamos a diario y gracias a nuestros sentidos. La automatización o mecanización de estos procesos nos ayuda a realizar labores repetitivas en las que usamos nuestros sentidos cambiando el tacto, el olfato o el oído por sensores que escojan y categoricen objetos quizás en una forma más práctica y con menos errores.

Sin embargo, existen numerosas restricciones en la traducción del mundo humano al de la máquina imponiendo problemas de cuantificación, cualificación y representación en campos tan diversos o tan obvios como por ejemplo en los campos de reconocimiento visual o de voz. Para sobrepasar estas barreras, investigaciones en el área de aprendizaje de máquina ilustran sobre como muchas de estas limitaciones podrían ser resueltas gracias a un conocimiento previo o anticipado del problema y sobre todo de como nuestra propia mente o naturaleza humana analiza, sintetiza y resuelve las incógnitas relevantes en el caso de la vida real. Por lo tanto, es evidente analizar este fenómeno en la naturaleza y traducirlo en forma de algoritmos para poder diseñar sistemas especializados a una tarea de reconocer patrones específicos. [12]

El objetivo de este documento es proporcionar al lector las herramientas básicas del Reconocimiento de Patrones, así como información sobre técnicas actualmente usadas. Debido a que el tema es muy extenso, esta tesina está orientada únicamente al Reconocimiento en Imágenes Digitales.

Quiero hacer hincapié en que este proyecto se llevó a cabo como requisito para obtener el título por la modalidad de Examen General de Conocimientos, esperando que sea de utilidad como documento de consulta.

* *Hardware* es un neologismo proveniente del inglés definido por la Real Academia de la Lengua Española como el conjunto de elementos materiales que conforman una computadora, sin embargo, es usual que sea utilizado en una forma más amplia, generalmente para describir componentes físicos de una tecnología.

1. CONCEPTOS BÁSICOS

El objetivo del Reconocimiento de Patrones es clasificar objetos de interés en una de las categorías o **clases**. Para ello, se extraen características de los objetos que genéricamente son llamadas **patrones**. Los objetos de interés pueden ser de cualquier forma: caracteres, células, ondas, señales eléctricas, “estados” de un sistema, imágenes, movimientos o cualquier otra cosa que uno deseara clasificar; por ejemplo, se pueden clasificar imágenes digitales dependiendo de sus píxeles.

Si existen patrones cuyas clases individuales son conocidas de antemano, entonces se tiene un problema de **Reconocimiento de Patrones Supervisado** o **Clasificación Supervisada**. Si las clases de todos los patrones disponibles son desconocidas, y quizá también es desconocido el número de clases, entonces se tiene un problema de **Reconocimiento de Patrones no Supervisado** o **Clasificación no Supervisada**, también llamada de **Agrupación (*Clustering*)**.

La clasificación utiliza habitualmente uno de los siguientes procedimientos: clasificación estadística (teoría de decisión) o clasificación sintáctica (estructural o determinística).

Como veremos, el Reconocimiento de Patrones gira en torno a varios conceptos claves: patrón, clase, rasgos o características, función discriminante o de decisión, etc. A continuación se revisarán algunos de ellos y se mostrarán las etapas para la elaboración de un Sistema de Reconocimiento Automático. Posteriormente, definiremos algunos conceptos básicos para el análisis de imágenes, que también compete a este documento.

▪ 1.1 Sistema de Reconocimiento Automático

El primer paso para proceder al reconocimiento de los objetos es que estén **segmentados** [aislados] en el espacio de objetos denominado **universo de trabajo**; después se realiza una transformación de ellos antes de ser reconocidos, esto es, se convierten en un vector numérico “X” cuyas componentes se denominan características o rasgos (a esta operación se le denomina **extracción de características**). Cada **vector de características** se compara con un conjunto de vectores preestablecidos o **diccionario**, compuesto por los vectores de todos los objetos del universo de trabajo. Una vez calculado el vector de características “X” asociado a un objeto individual, su reconocimiento automático se basa en determinar su grado de semejanza con los

1. Conceptos Básicos

vectores de características prototipos de cada posible clase de objetos previamente definidos. En diagrama de la figura 1.1 se representa gráficamente el proceso de Reconocimiento Automático de Formas mencionado. [1]

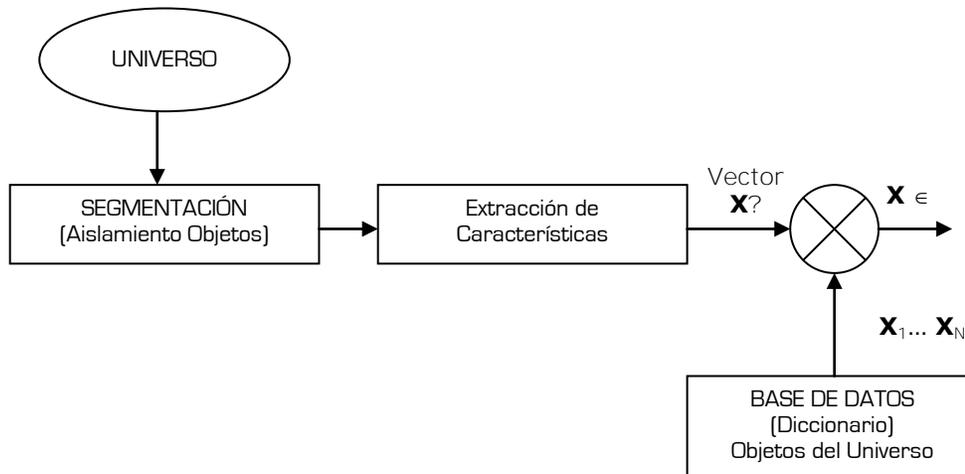
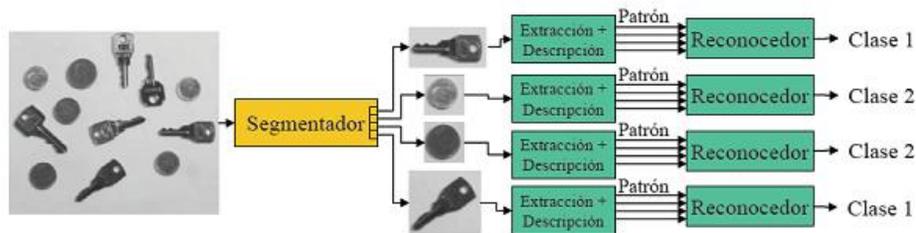


Figura 1.1 Diagrama de bloques de un Sistema de Reconocimiento Automático

En un diagrama más mostrado en la figura 1.2 podemos ver básicamente algunos conceptos a revisar. En él se observan vagamente los pasos de segmentación, extracción de características más descripción y reconocimiento o clasificación; así como los conceptos de patrón y clase.



José Luis Alba - Universidad de Vitor; Jesús Cid - Universidad Carlos III

Figura 1.2 Diagrama de Conceptos Básicos

▪ 1.2 ¿Qué es un Patrón?

Para empezar definiremos un **patrón** como un objeto o sustancia que se emplea como muestra para medir alguna magnitud o para replicarla; se dice que es el elemento ideal o más representativo de una **clase**, un prototipo.

Richard E. Wood y Rafael C. González definen un patrón como: “*una descripción estructural o cuantitativa de un objeto o de alguna otra entidad de interés en una imagen y que está formado por uno o más descriptores*”. [2]

Entonces, un patrón también puede verse como una colección (posiblemente ordenada y estructurada) de descriptores por la cual cada objeto queda representado tras los procesos de segmentación, extracción de características y descripción.

En los problemas de reconocimiento, cada patrón se supone perteneciente a una categoría o clase.

Los tipos de patrones más comunes son:

- **Patrones vectoriales:** $x = [x_1, x_2, \dots, x_n]^T$
- **Patrones estructurados:** cadenas
- **Patrones en árbol**

Para ejemplificar los patrones vectoriales vamos a tomar la clasificación de los tipos de una flor particular, Iris, de acuerdo a: A) Tres categorías: *iris setosa*, *iris versicolor* e *iris virginica* (figura 1.3) y B) Patrones bidimensionales: longitud y ancho del pétalo.



Figura 1.3 Tipos de Iris

Los descriptores utilizados sirven para discriminar *iris setosa* de las otras dos, pero no para discriminar entre *iris virginica* e *iris versicolor*.

Como se puede apreciar en la figura 1.4, los datos muestreados de *iris setosa* quedan agrupados en un área más o menos definida que permite una **discriminación** satisfactoria, ya que se utilizaron las características adecuadas al objetivo.

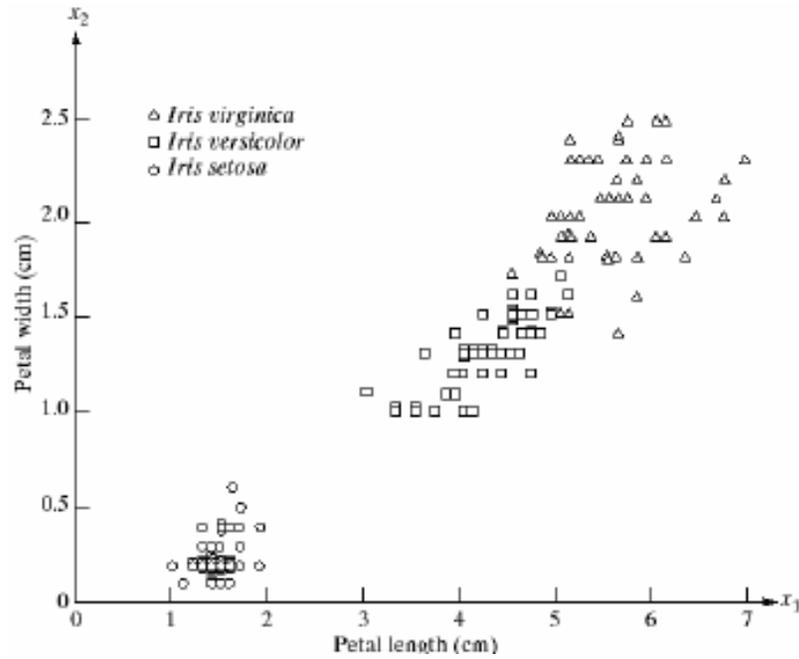


Figura 1.4 Ejemplo de patrones vectoriales. Clasificación de tipos de Iris

Un ejemplo de patrones estructurados (en cadena) es el reconocimiento de huellas dactilares (figura 1.5).



Figura 1.5 Huella dactilar

Este tipo de patrones codifican relaciones (espaciales o de otro tipo) entre componentes del objeto o descriptores. Los algoritmos de reconocimiento suelen basarse en la detección de las minucias [*minutiae*], las cadenas [*ridges*] que forman y su relación entre ellas (figura 1.6). [13]

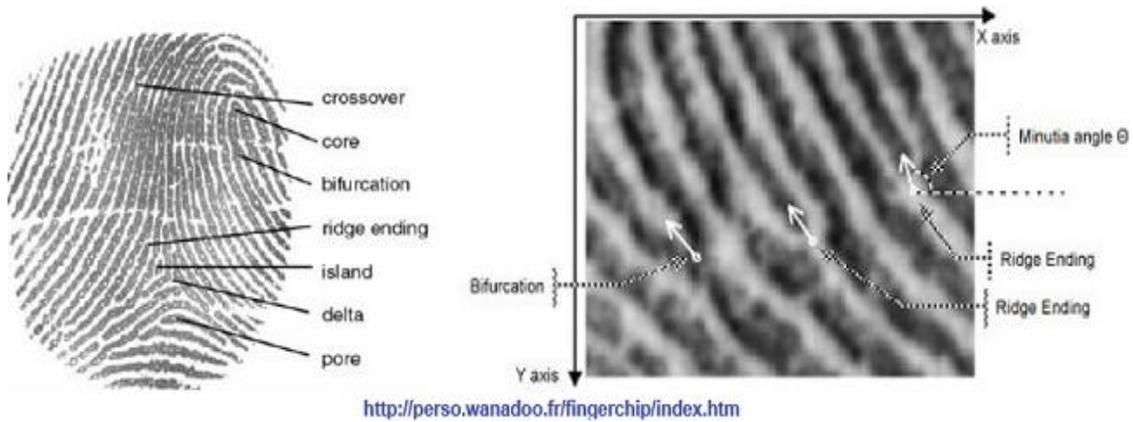


Figura 1.6 Ejemplo de patrones estructurados. Reconocimiento de huellas dactilares

Por último, veremos un ejemplo de patrones en árbol a partir de una imagen de satélite (figura 1.7).

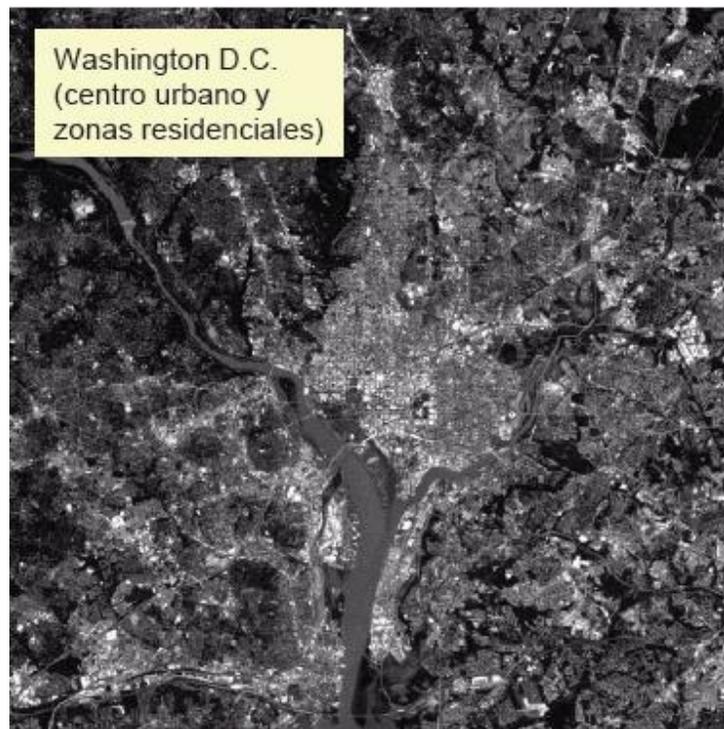


Figura 1.7 Imagen de satélite de una zona de la ciudad de Washington

La descripción en árbol nos muestra que cada rama codifica una relación. En la figura 1.8 observamos el árbol creado a partir de la imagen satelital, en ella pueden separarse los centros urbanos de las zonas residenciales, los edificios de los caminos, las estructuras grandes de las pequeñas, etc.

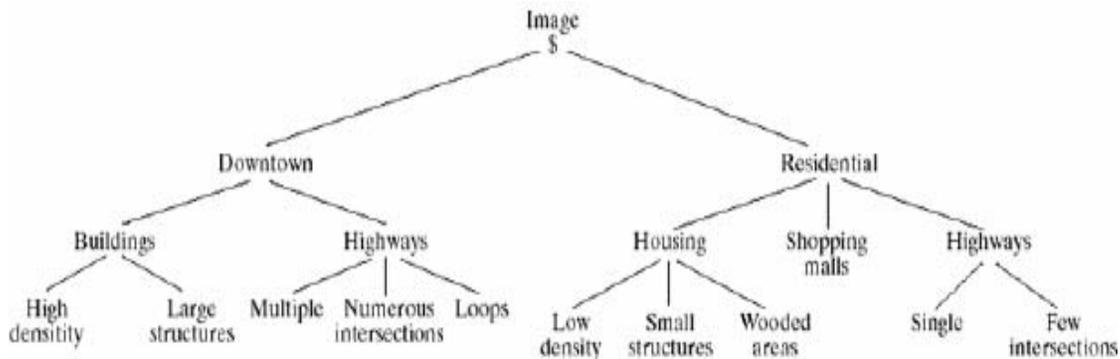


Figura 1.8 Ejemplo de patrones en árbol. Relaciones a partir de una imagen de satélite

▪ 1.3 Universo de Trabajo (UT)

Un Sistema de Reconocimiento Automático opera siempre con un **universo de trabajo (UT)** previamente definido, el cual está formado por todos los posibles objetos individuales que se van a reconocer de manera automática. Estos objetos predefinidos son precisamente los patrones [1]. Como se observó en el diagrama de la figura 1.1 este universo es el principio de todo.

Se puede decir que para una aplicación específica o sistema de reconocimiento concreto el universo de trabajo es el conjunto de todos los patrones que desean reconocerse:

$$UT = \{\text{patrones}\}$$

La primera etapa en el diseño consiste en el establecimiento de las clases, en lo que se denomina **definición del universo de trabajo del sistema**.

En la mayoría de los casos reales esta etapa es directa y trivial, ya que se conocen perfectamente las clases de objetos que han de ser reconocidos. No obstante, puede ocurrir que las clases sean desconocidas *a priori*. En tales situaciones se recurre a las técnicas denominadas de Agrupación (*Clustering*) o Reconocimiento de Patrones no Supervisado.

Para establecer el UT es necesario definir qué objetos se van a reconocer. Por ejemplo, tomando en cuenta el ejemplo de la figura 1.2, es fácil definir el universo de trabajo como sigue:

$$UT = \{\text{llave, moneda}\}$$

▪ 1.4 Clase

Una **clase** es un conjunto de entidades o elementos que comparten alguna característica que las diferencia de otras, es decir que comparten características comunes.

El objetivo del Sistema de Reconocimiento Automático es asignar cada objeto de interés a su clase correspondiente. En el diagrama de la figura 1.2 se observan claramente dos tipos de clases:

Clase 1 = llave Clase 2 = moneda

Para el Reconocimiento de Patrones es importante que:

- Patrones que describen objetos de una misma clase, presenten características similares.
- Patrones que describen objetos de diferentes clases presenten características diferenciadas.

Ahora bien, las clases son representadas por vectores y se dice que una clase es determinística cuando sus elementos se representan por un único vector, que es el prototipo de la clase. [1]

Cuando hablamos de una **clase de rechazo** nos referimos al conjunto de entidades que no se pueden “etiquetar” como ninguna de las clases del problema.

Por ejemplo, tomando en cuenta que se desea hacer el reconocimiento de caracteres que incluya únicamente letras del todo el alfabeto (mayúsculas y minúsculas) y todos los dígitos, las clases estarían definidas como sigue:

- Clase: **C** = {A, B, C, ... Z, a, b, c, ... z, 0, 1, 2, ... 9}
- Clase de rechazo: **Cr** = {¡, !, “, \$, %, &, /, {, }, =, ¿, ?, ...}

▪ 1.5 Vector de Características

Una vez definidas las clases, el siguiente paso consiste en la elección del **vector de características**.

El vector de características, sin duda, constituye el elemento clave en un sistema de reconocimiento automático. Y para la elección y prueba de las características, que es

1. Conceptos Básicos

una etapa compleja en la mayoría de las aplicaciones reales y difícil de formalizar, es necesario que las características tengan rasgos discriminantes.

Se puede decir que no hay reglas para la elección y testeo de las características adecuadas, sin embargo, se sugieren las siguientes propiedades importantes, propuestas por Darío Maravall Gómez-Allende [1]:

1. Las características deben ser discriminantes: valores numéricos diferentes para objetos de clases distintas. Separación lo más nítida posible de las clases.
2. Las características deben ser fiables, es decir, con cambios numéricos pequeños para objetos de una misma clase. Posteriormente veremos que si no se cumple esta propiedad es necesario pasar a un enfoque estadístico.
3. Las características no deben estar correlacionadas, es decir, nunca deberán utilizarse características que dependan fuertemente entre sí, ya que no añaden información discriminante. La independencia de las características equivale al principio de la parsimonia: decir lo mismo con la máxima economía de términos.
4. Las características deben calcularse en un tiempo aceptable. Este es un requisito que puede llegar a ser determinante en ciertas aplicaciones de tiempo real.
5. Las características deben obtenerse con sensores económicos.

Las dos últimas propiedades son las más simples de evaluar, puesto que los parámetros involucrados son fácilmente cuantificables.

En cuanto a la no-correlación, se mide sobre las combinaciones posibles de parejas de dos características, clase a clase, a partir de la correspondiente matriz de covarianza.

Para la fiabilidad se utiliza la matriz de covarianza y se analizan los valores de la diagonal principal, también clase a clase. Los elementos de la diagonal principal del resultado de la matriz son precisamente las varianzas de las correspondientes características. Cuanto mayores sean estos elementos, mayor es la dispersión estadística de las características.

Sin embargo, la selección de las características nunca debe hacerse aplicando el criterio de la fiabilidad individualmente, sino en conexión con la capacidad discriminante, puesto que una característica con una varianza muy elevada puede ser, al mismo tiempo, una magnífica característica discriminante.

La cuantificación de la primera propiedad que debe cumplir toda característica (es decir, la capacidad de discriminación), tal como se indica en el párrafo anterior no debe limitarse a la distancia entre clases, que se mediría como la distancia euclídea entre sus correspondientes medias, sino que es necesario incorporar la dispersión interna de las clases. [1]

La elección de rasgos o características es muy dependiente de la aplicación concreta que se quiera desarrollar: dicen que se trata más de un arte que de una ciencia.

Existen técnicas formales de ayuda en esta etapa, pero sólo pueden aplicarse *a posteriori*; es decir, una vez elegido un conjunto específico de características. Estas técnicas realizan un testeo de las propiedades deseables en toda característica: la capacidad discriminante entre clases y la fiabilidad, fundamentalmente. Estas técnicas son comúnmente estadísticas.

La elección del vector de características se realiza para cada uno de los objetos individuales a reconocer. En el ejemplo de la figura 1.9 las características o rasgos más discriminantes son el perímetro, el número de huecos y el centro de masas (vector de dimensión 3).

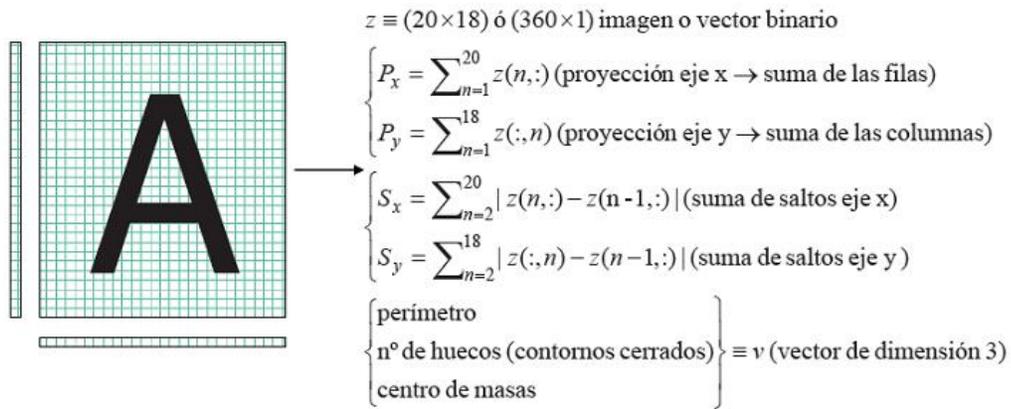


Figura 1.9 Ejemplo de Vector de Características

▪ 1.6 Selección de Variables

Una vez definidos los criterios de evaluación de las características, se puede establecer un proceso operativo para selección de las mismas.

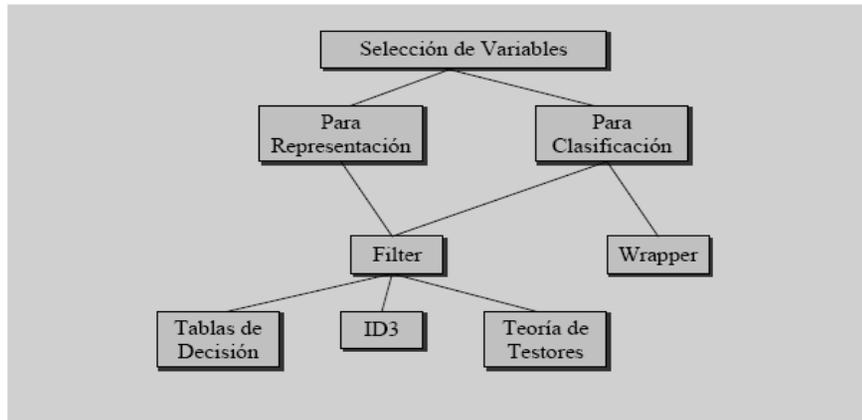


Figura 1.10 Tipos de Selección de Variables

- ✓ **Para la Representación.** Decidir cuáles características representan mejor a cierto tipo de objetos.
- ✓ **Para la Clasificación.** La selección de características relevantes, a partir del conjunto total de características que describen a los objetos, se hace con dos motivos fundamentales: mejorar la clasificación y/o aumentar la velocidad de procesamiento.
- ✓ **Wrapper.** La selección de las características se hace usando información del mecanismo de clasificación.
- ✓ **Filter.** La selección se hace con un criterio independiente del clasificador.
- ✓ **Tablas de Decisión.** Se busca un subconjunto mínimo de variables que no introduzca confusión entre clases.
- ✓ **ID3.** Se crea un árbol de decisión y se selecciona un conjunto de variables que permita discriminar entre clases.
- ✓ **Teoría de Testores.** Se buscan todos los subconjuntos de variables discriminantes minimales, con estos se evalúa la relevancia de cada variable y se seleccionan aquellas con mayor relevancia.

▪ 1.7 Función Discriminante

Después de seleccionado el vector de características del sistema, la etapa que sigue se centra en el cálculo de las correspondientes funciones de decisión o funciones discriminantes. La **función discriminante** mide la relevancia de la clase para el patrón.

Para ilustrar este concepto veremos un ejemplo. Suponiendo que deseamos reconocer de manera automática dos clases de deportistas: jugadores de ajedrez y jugadores de baloncesto.

Para diferenciar estas dos clases de deportistas tomaremos en cuenta aquellos rasgos o características que los unen y que los separan. Cuando hablamos de unir, nos referimos, lógicamente, a los objetos de una misma clase y cuando hablamos de separar nos estamos refiriendo a objetos de clases diferentes.

Escogemos dos características: la altura física y el cociente intelectual normalizado entre 0 y 1 que mida [con 1 como máximo valor] el resultado de un individuo en un test de inteligencia.

Si representamos sobre un plano a jugadores de ajedrez y baloncesto, reducimos éstos a un vector bidimensional de características "X", de manera que se tendría la situación reflejada en el plano de la figura 1.11. Podemos observar que la recta trazada es la **función discriminante (fd)**, divide el plano de características [espacio abstracto de las clases o patrones] en dos semiplanos, correspondiente cada uno de ellos a una clase.

Se dice en este caso que existe separabilidad lineal de las clases. Cuando se eligen acertadamente las características, existirá probablemente separación lineal de clases. Sin embargo, cuando en la distribución de las clases no existe separabilidad lineal, es decir, cuando no exista un hiperplano que separe las clases, es necesario recurrir a funciones discriminantes no lineales, lo cual complica extraordinariamente el diseño del correspondiente clasificador automático y es necesario recurrir a procedimientos más avanzados: discriminadores polinomiales, Aprendizaje Supervisado, Redes Neuronales (RN), etc., o bien, escoger otro conjunto de características discriminantes que proporcionen un nuevo espacio abstracto de patrones con separabilidad lineal de las clases.

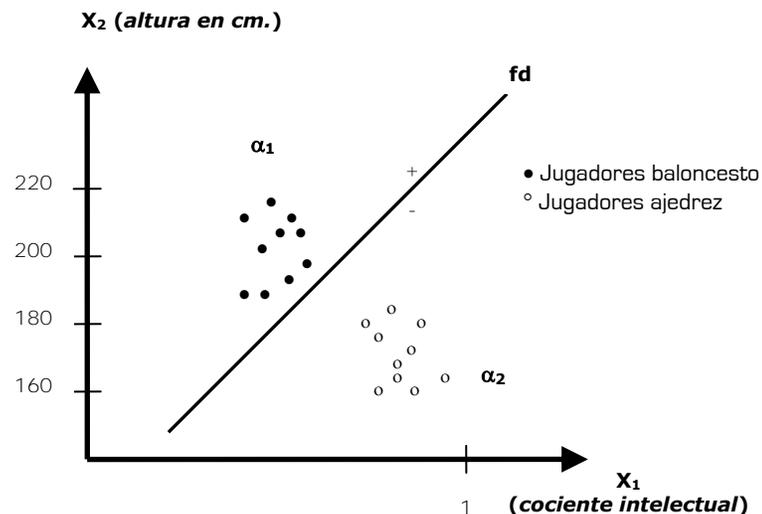


Figura 1.11 Ejemplo de Función Discriminante. Representación en un plano de dos clases de deportistas

La función discriminante de la figura 1.11 es lineal. En un plano se tratará de rectas, en el espacio de tres dimensiones son planos y generalizando, en un espacio n -dimensional las funciones discriminantes lineales son hiperplanos.

El anterior planteamiento del problema conduce a una solución lineal basada en una regionalización del espacio de las clases. Existe otra solución lineal basada en el uso de medidas de distancia, también dentro del espacio abstracto de las clases.

Existen dos familias de funciones discriminantes lineales: las basadas en una regionalización del espacio abstracto de las clases y las basadas en el uso de una función de distancia. Ambas consisten en una combinación lineal de las características. [1]

En general, la solución consiste en encontrar un hiperplano que divida el espacio de las clases en dos regiones, asociadas cada una de ellas a diferentes clases. En el caso general de n características, las funciones discriminantes pasan a ser hiperplanos y obviamente, cuando el número de clases sea superior a dos, es preciso obtener más de un hiperplano. Como podemos imaginar, no es trivial establecer el número mínimo de funciones discriminantes que se necesitarían para dividir el espacio de las clases en n regiones disjuntas, cada una asociada a una sola clase.

✓ **Región de decisión.** El conjunto de todos los puntos del espacio que el reconocedor asigna a la clase.

✓ **Frontera de decisión.** Separa regiones de decisión.

✓ **Decisor.** Típicamente (aunque no siempre), selecciona la clase de mayor o menor valor de la función discriminante.

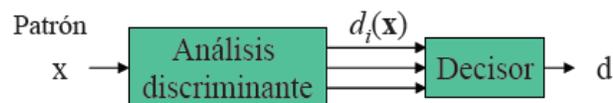


Figura 1.12 Decisor

▪ 1.8 Segmentación y Extracción de Características

Como se ha dicho, se denomina segmentación al aislamiento de los objetos individuales en el universo de trabajo, ya que aparecen normalmente mezclados entre sí; presentando incluso asociaciones e interacciones muy complejas.

Es una operación delicada en la mayoría de los casos prácticos. En realidad, se trata de una problemática ajena, en sentido estricto, al Reconocimiento de Patrones, ya que en éste se parte siempre desde la conversión de un objeto individual en un **vector de observación**. En la visión artificial, la segmentación de las escenas digitales, previas a su análisis y reconocimiento automático, es la etapa más compleja y delicada.

Ahora bien, como pudimos ver en el diagrama de la figura 1.1, el subsistema que extrae la información relevante para la clasificación a partir de las entidades cuantificables es denominado **extractor de características** o **de información** y transforma la información observada en valores numéricos o simbólicos (suele ir acompañado de la **descripción** de dichas características). El resultado de este proceso nos dará el denominado **vector de características** de cada objeto a reconocer.

En este punto es donde podemos notar que el enorme potencial práctico del reconocimiento de patrones está dado por la reducción de la complejidad de los objetos a una información condensada en un vector de datos numéricos.

En el siguiente apartado veremos más a detalle los conceptos de segmentación y extracción de características.

▪ 1.9 Clasificación

La clasificación es propiamente el reconocimiento y es el proceso en el cual se asigna una “etiqueta”, que representa una clase a un patrón concreto. El subsistema encargado de hacer la clasificación es el **clasificador** o **reconocedor**, utiliza un vector de características de la entidad cuantificable y lo asigna a una de m clases, dicho de otra forma, se basa en las características extraídas para clasificar la medición.

En el apartado III de este documento se tratará más a detalle la clasificación.

✓ **Falso rechazo (falso negativo) y falsa aceptación (falso positivo).** Para problemas de 2 clases estas definiciones reflejan la importancia de una decisión contra la opuesta. El sistema de clasificación se puede sintonizar para que trabaje ponderando un tipo de error sobre el otro.

A continuación un diagrama más en la figura 1.13 que ilustra el enfoque del Reconocimiento de Patrones. En él puede observarse el vector de observación “Y” que alimenta el sistema y se transformará en el vector de características “X”. [3]

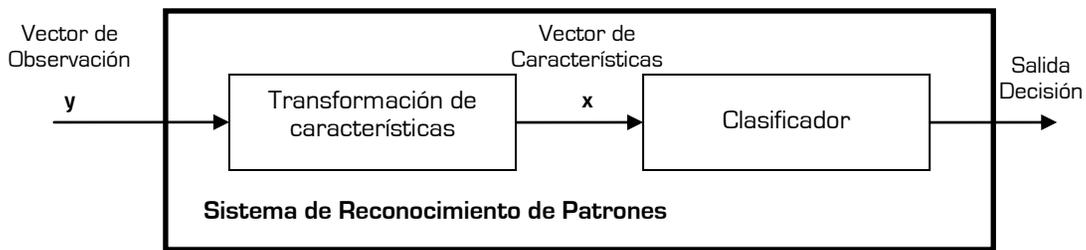


Figura 1.13 El enfoque básico del Reconocimiento de Patrones

1.10 Enfoques y Problemas

Veamos ahora algunos enfoques y problemas del Reconocimiento de Patrones:

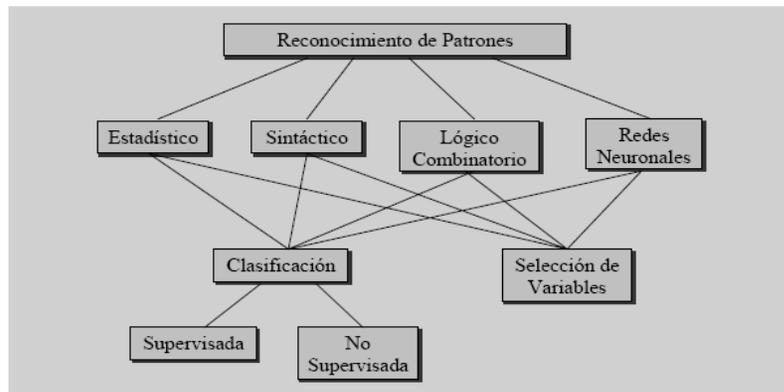


Figura 1.14 Enfoques y Problemas del Reconocimiento de Patrones

- ✓ **Clasificación o reconocimiento estadístico.** Este enfoque se basa en la teoría de probabilidad y estadística y supone que se tiene un conjunto de medidas numéricas con distribuciones de probabilidad conocidas y a partir de ellas se hace el reconocimiento.
- ✓ **Clasificación o reconocimiento sintáctico.** Este enfoque se basa en encontrar las relaciones estructurales que guardan los objetos de estudio, utilizando la teoría de lenguajes formales. El objetivo es construir una gramática que describa la estructura del universo de objetos.
- ✓ **Clasificación o reconocimiento lógico combinatorio.** Este enfoque se basa en la idea de que la modelación del problema debe ser lo más cercana posible a la realidad del mismo, sin hacer suposiciones que no estén fundamentadas. Uno de los aspectos esenciales del enfoque es que las características utilizadas para describir a los objetos de estudio deben ser tratadas cuidadosamente.

- ✓ **Redes Neuronales (RN).** Este enfoque supone que se tiene una estructura de neuronas interconectadas que se estimulan unas a otras, las cuales pueden ser entrenadas para dar una cierta respuesta cuando se le presentan determinados valores.
- ✓ **Selección de Variables.** Consiste en determinar cuál es el conjunto de características más adecuado para describir a los objetos.
- ✓ **Clasificación o reconocimiento supervisado.** Consiste en clasificar nuevos objetos basándose en la información de una muestra ya clasificada.
- ✓ **Clasificación o reconocimiento no supervisado.** Consiste en que dada una muestra no clasificada debe encontrarse una clasificación de la misma.

▪ 1.11 Introducción al Análisis de Imágenes

El “mundo real” es donde vivimos y como lo perciben nuestros sentidos. Datos de una imagen en el mundo real son adquiridos por algún sistema sensorial remoto, como son nuestros ojos. Como sistema de teledetección el ojo humano recibe un haz de energía proveniente de los objetos exteriores, por reflejo de la luz solar o artificial, forma una imagen que se transmite al cerebro que la percibe, e interpreta así, el mundo real que lo rodea. Los datos de sensores remotos pueden ser reproducidos en algo que llamamos **espacio de imágenes (*image space*)** como una función de las características físicas de los objetos y el sensor usado para captarlos.

También se pueden extraer características propias de los datos como reflectancia o textura y a estos podemos ubicarlos en un **espacio de características (*feature space*)**, en este proceso es posible obtener **clases espectrales** que pueden o no coincidir con las clases temáticas definidas por el ojo humano en el mundo real. [16]

La cuantificación de objetos presentes en imágenes es una tarea a la que investigadores de diversos campos se han de enfrentar habitualmente. Muchas veces esta labor de medida se realiza de modo manual, estando sujeta la validez de los resultados a la fatiga del operador que la realiza. Sin embargo, reconocer, contar y medir tamaño, forma, posición, densidad y otro tipo de propiedades similares de determinados objetos presentes en una imagen, es algo que, actualmente, está al alcance de cualquier computadora personal complementado con un *software** específico, realizando estas tareas de un modo mucho más rápido, preciso y reproducible.

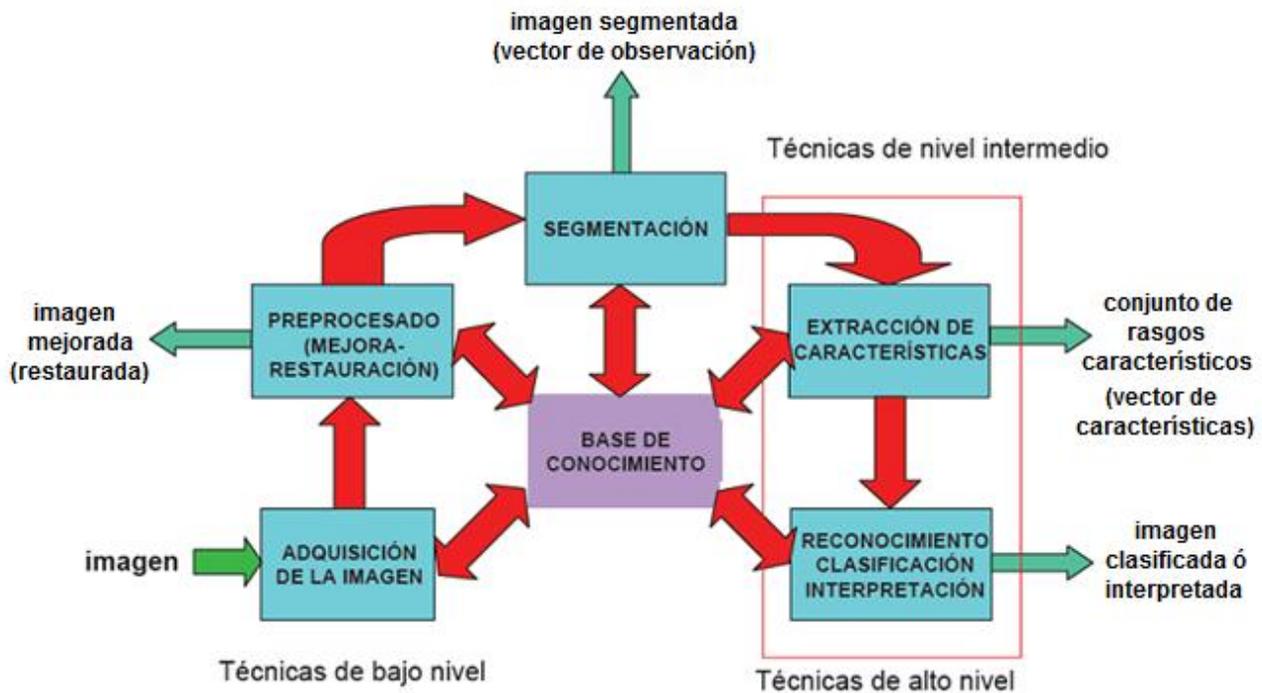
* *Software* (programa, equipamiento lógico o soporte lógico) se le denomina así a todos los componentes intangibles de una computadora, es decir, al conjunto de programas y procedimientos necesarios para hacer posible la realización de una tarea específica, en contraposición a los componentes físicos del sistema (hardware).

1. Conceptos Básicos

El interés por el Procesamiento Digital de Imágenes (PDI) tiene su origen en dos de las áreas existentes de aplicación. El primero de ellos para la mejora de la imagen en la interpretación humana y el último para el procesamiento de imágenes en el área de la visión artificial con la intención de hacer el proceso de percepción de manera autónoma por una máquina.

En visión por computadora, la adquisición de imágenes consiste en digitalizar una señal de vídeo que contiene información luminosa de una escena a interpretar. La señal continua de vídeo debería ser muestreada y digitalizada binariamente para su tratamiento en la computadora.

En la figura 1.15 se muestra el esquema general del análisis de imágenes, se desarrollarán algunos conceptos de las etapas descritas como técnicas de bajo nivel en esta sección del documento.



José Luis Alba - Universidad de Vigo; Jesús Oid - Universidad Carlos III

Figura 1.15 Esquema general del Análisis de Imágenes

1.11.1 Imagen Digital

La imagen digital ha revolucionado el mundo de la fotografía de consumo y de la fotografía científica. Se han incorporado a la práctica una enorme cantidad de equipos que sustituyen rápidamente a las tradicionales cámaras fotográficas.

La principal ventaja de la imagen digital frente a la analógica es su inmediatez. No necesitamos esperar a realizar un complicado proceso de revelado para ver la imagen y en caso de que el resultado no sea de nuestro gusto podemos desecharla y tomar otra sin coste alguno. Además presenta otras muchas ventajas: a una imagen digital se le puede cambiar el contraste, el brillo, el color; se puede combinar con otras imágenes; se puede duplicar, rotar, aumentar o disminuir; puede cuantificarse y puede transmitirse a miles de kilómetros de distancia en pocos segundos.

Desde un punto de vista físico, una imagen puede considerarse como un objeto plano cuya intensidad luminosa y color puede variar de un punto a otro. Si se trata de imágenes monocromas (blanco y negro) podemos verlas como una matriz de números entre 0 y 255, siendo 0 = negro y 255 = blanco; y se pueden representar como una función continua $f(x, y)$ donde (x, y) son sus coordenadas y el valor de f es proporcional a la intensidad luminosa (nivel de gris) en ese punto.

Para obtener una imagen que pueda ser tratada por una computadora es preciso someter la función $f(x, y)$ a un proceso de discretización tanto en las coordenadas como en la intensidad a este proceso se le denomina **digitalización** y se muestra en la figura 1.16.

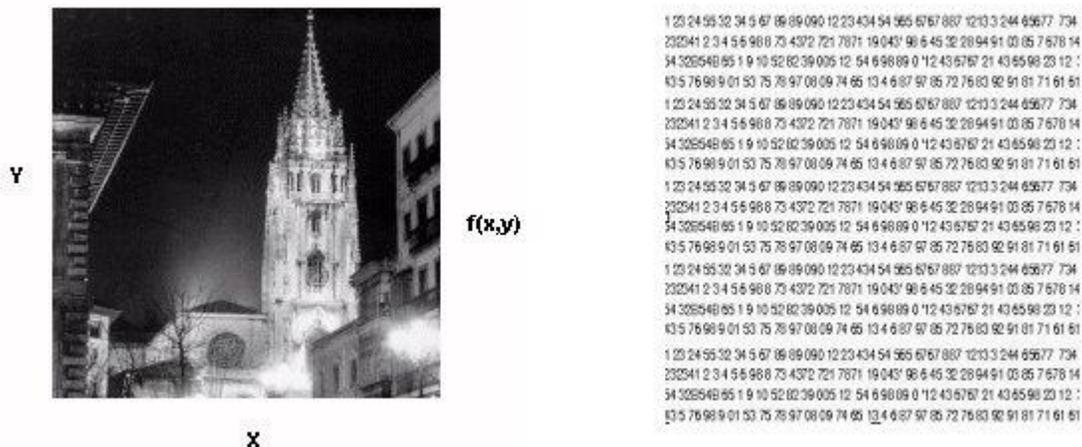


Figura 1.16 Imagen y matriz numérica resultante del proceso de Digitalización

La digitalización consiste en la descomposición de la imagen en una matriz de $m \times m$ puntos, donde cada uno tiene un valor proporcional a su nivel de gris. Dado que este valor puede ser cualquiera dentro de un rango continuo, es preciso dividir dicho rango en una serie de k intervalos, de forma que el nivel de gris de cada punto sea asignado a uno de los valores que representa dicho intervalo. La mayoría de los sistemas de digitalización de imágenes suelen ser capaces de discriminar 256 niveles de gris.

1. Conceptos Básicos

Por consiguiente, una imagen digital es una imagen $f(x, y)$ que ha sido discretizada en sus coordenadas espaciales y en su intensidad luminosa.

Cada elemento en que se divide la imagen recibe el nombre de **píxel** (*picture element*, elementos de imagen). El número de niveles de gris y las dimensiones de la matriz (número de filas por número de columnas) nos condicionan la capacidad de resolución de la imagen digital. La resolución espacial de la imagen viene dada por el número de píxeles (número de filas por número de columnas) que tiene la imagen (figura 1.17). Cuanto mayor sea este número mayor va a ser la resolución. Cada píxel incorpora el nivel de luminosidad del punto correspondiente de la escena digital.

La resolución cromática depende del número de bits que utilicemos para almacenar el valor de un píxel. Si utilizamos un bit podemos tener únicamente dos valores [0, 1] (blanco y negro), si utilizamos 4 bits el número posible de niveles de gris será de 16 (2^4) y si utilizamos 8 bits (1 byte) el número de niveles de gris posibles es de 256 (2^8).

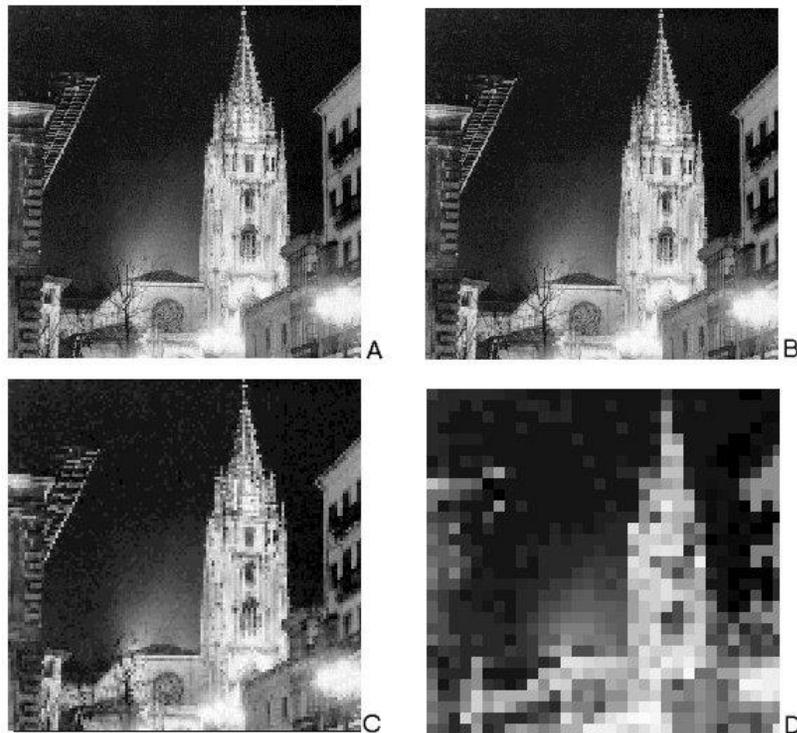


Figura 1.17 Imagen con diferentes resoluciones: A) 256 x 256, B) 128 x 128, C) 64 x 64, D) 32 x 32

Hasta aquí se ha hablado de niveles de gris, esto es, de imágenes monocromáticas. Como se verá después, la visión por computadora se ha desarrollado siempre en torno al principio del reconocimiento de formas. El color no es una forma, sino un atributo de las formas. [1] Sin embargo, se verán *grasso modo*, fundamentos básicos del color.

1.11.2 Fundamentos del Color

- **Colores primarios (modelo aditivo):** Rojo, Verde y Azul.
- **Colores secundarios:** Cian, Magenta y Amarillo. Formados por la suma de colores primarios.
- **Colores primarios de pigmento (modelo sustractivo):** Son aquellos que absorben un color primario de luz y reflejan los restantes. Pigmentos primarios: Amarillo, Magenta y Cian.

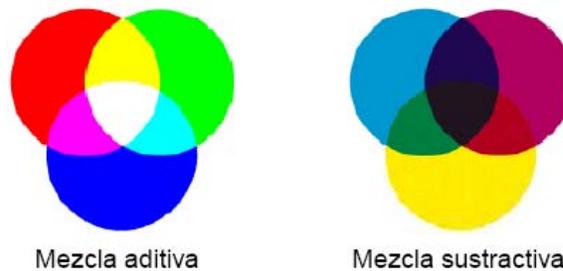


Figura 1.18 Modelos aditivo y sustractivo de color

Características del color:

- **Brillo:** Noción acromática de intensidad.
- **Tono:** Valor dominante en la onda percibida (color del objeto).
- **Saturación:** Pureza relativa del color. Inversamente proporcional a la mezcla de luz blanca. Ejemplo: rojo y verde (completamente saturados); rosa lavanda y verde manzana (menos saturados).
- **Cromaticidad:** Tono y saturación.

Modelos de Color:

Los modelos [o espacios] de color son estándares de especificación de colores. Éstos especifican un sistema de coordenadas 3D y un subespacio en el cual queda definido cualquier color como un punto único. Los más comunes son los siguientes:

Modelo RGB (Red, Green, Blue). Hace referencia a la composición del color en términos de la intensidad de los colores primarios con que se

forma: el rojo, el verde y el azul. Es un modelo de color basado en la síntesis aditiva, con el que es posible representar un color mediante la mezcla por adición de los tres colores luz primarios. Para indicar con qué proporción mezclamos cada color, se asigna un valor a cada uno de los colores primarios, de manera, por ejemplo, el valor 0 significa que no interviene en la mezcla, y a medida que ese valor aumenta, se entiende que aporta más intensidad a la mezcla. Aunque el intervalo de valores podría ser cualquiera, es frecuente que cada color primario se codifique con un byte (8 bits). Así, de una manera estándar, la intensidad de cada una de las componentes se mide según una escala que va del 0 al 255. [14] Ejemplos que lo utilizan: monitores a color, cámaras de video, etc.

- **Modelo CMY (*Cyan, Magenta, Yellow*) y Modelo CMYK (*Cyan, Magenta, Yellow, Black*).** Para impresión, los colores usados son cian, magenta y amarillo. En el modelo CMY, el negro es creado por mezcla de todos los colores, y el blanco es la ausencia de cualquier color (asumiendo que el papel sea blanco). Como la mezcla de los colores es sustractiva, también es llamado modelo de color sustractivo. Una mezcla de cian, magenta y amarillo en realidad resulta en un color negro turbio por lo que normalmente se utiliza tinta negra de verdad. Cuando el negro es añadido, este modelo de color es denominado modelo CMYK. Recientemente, se ha demostrado que el modelo de color CMY es también más preciso para las mezclas de pigmento. [15]
- **Modelo RYB (*Red, Yellow, Blue*).** En el modelo de color RYB, el rojo, el amarillo y el azul son los colores primarios, y en teoría, el resto de colores puros (color materia) puede ser creados mezclando pintura roja, amarilla y azul. Es utilizado en general en conceptos de arte y pintura tradicionales, y en raras ocasiones usado en exteriores en la mezcla de pigmentos de pintura. Aún siendo usado como guía para la mezcla de pigmentos, el modelo RYB no representa con precisión los colores que deberían resultar de mezclar los 3 colores RYB primarios. En el 2004, la ciencia reconoció que este modelo es incorrecto, pero continúa siendo utilizado habitualmente en arte. [15]

En el caso de imágenes en color, la intensidad puede considerarse como un vector tridimensional cuyas componentes son las intensidades en tres bandas espectrales. Hablándose en este caso de una imagen multibanda, puesto que la información de la imagen color se desglosa en tres imágenes correspondientes a cada una de las bandas del espectro visible, por ejemplo RGB (figura 1.19). Para cada una de las bandas se suelen utilizar 8 bits de información, 24 bits en total, lo que nos dan más de 16 millones de posibles combinaciones de colores. Algunos dispositivos pueden llegar a utilizar hasta 16 bits para cada banda, 48 bits en total (más de 281 trillones de colores).

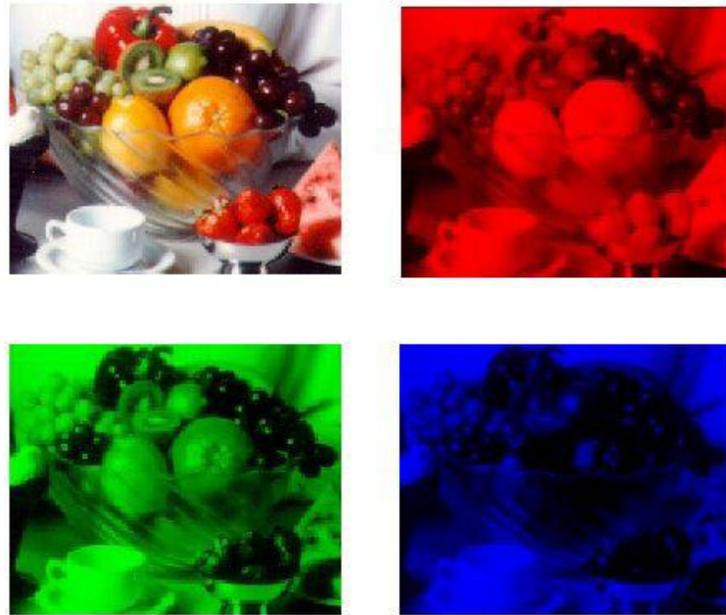


Figura 1.19 Imagen en color y su descomposición en las tres bandas espectrales: rojo, verde y azul (RGB)

1.11.3 Dispositivos de Captación de Imágenes

Para la adquisición de las imágenes se requiere algún dispositivo. La mayoría de los dispositivos para captar imágenes son dispositivos con sensores CCD (*Charge Couple Device*). Un CCD es un dispositivo similar a un microchip con una serie de elementos fotosensibles (píxeles) de polisilicio. Cada uno de estos elementos es capaz de generar una señal eléctrica proporcional al número de fotones (intensidad luminosa) que ha recibido (figura 1.20).

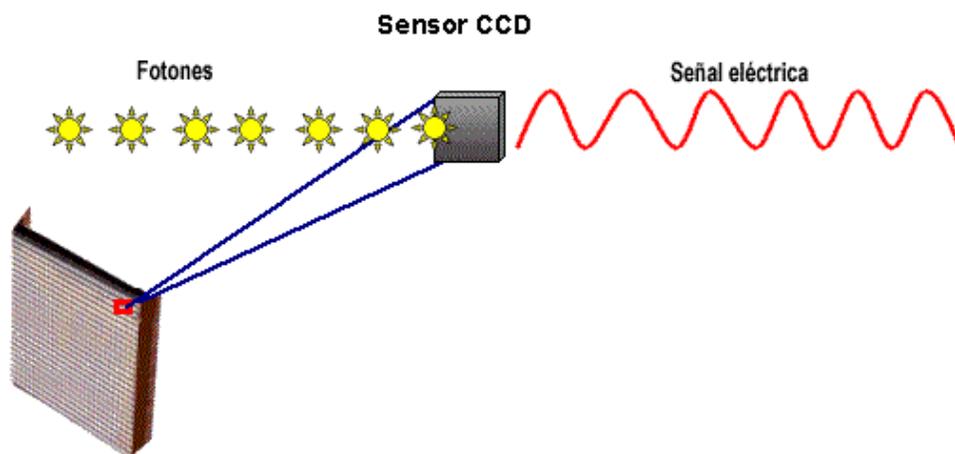


Figura 1.20 Esquema de un CCD de superficie (*array*) y de su funcionamiento.

Existen dos tipos de CCD según la disposición de los elementos fotosensibles: **lineales** y **de superficie** (*array*).

En los primeros los píxeles se ordenan de modo lineal. Para obtener una imagen es preciso desplazar el CCD por la superficie de la imagen. Este tipo de CCD es el que se utiliza en los escáneres.

Los CCD de superficie o *array* son los que se utilizan en las cámaras de vídeo o de fotografía digital. Tienen dispuestos los píxeles en forma de matriz y permiten capturar toda la imagen al mismo tiempo.

Los dispositivos CCD capturan imágenes monocromas. Para poder capturar imágenes en color con un CCD es necesario descomponer previamente la luz en los tres colores básicos del espectro (RGB): rojo verde y azul y recoger la información de cada uno de éstos colores en un píxel. Para conseguir esto se pueden utilizar tres métodos:

- **CCD único.** Un CCD con píxeles sensibles a diferentes colores (rojo, verde y azul).
- **Tres CCD's.** Cada uno sensible a uno de los colores.
- **CCD único con triple exposición.** Se coloca delante del CCD un filtro para cada uno de los colores del espectro y se toman tres imágenes de la misma zona.

Las imágenes de mayor calidad se obtienen con el segundo método. El tercero permite obtener una calidad similar al segundo, pero en objetos estáticos, ya que se necesita el triple de tiempo para obtener la imagen en color.

Los escáneres al contar con un sensor lineal permiten una mayor resolución de imagen, sin embargo el tiempo de digitalización es mucho más lento y en la mayoría de los casos sólo podemos digitalizar imágenes planas (fotografías, transparencias, radiografías, etc.). Por contra los tiempos de digitalización de las cámaras son mucho más rápidos y nos permiten captar cualquier tipo de escenas, aunque la resolución espacial de la imagen sea menor.

Además de por el tipo de sensor CCD (lineales o de superficie) los dispositivos de captación los podemos dividir en **analógicos** y **digitales**. En el primer grupo tenemos las cámaras con sensor CCD que traducen la imagen que capturan en una señal de vídeo (PAL, NTSC o Secam). Esta señal de vídeo será posteriormente digitalizada mediante un *frame grabber** o tarjeta digitalizadora de vídeo. La máxima resolución que podemos

* Un *frame grabber* es un dispositivo electrónico que capta cuadros digitales individuales de una señal de vídeo analógico o un flujo de vídeo digital. Se suele utilizar como un componente de los sistemas de visión por computadora en la que se capturan fotogramas de vídeo en formato digital

obtener viene limitada por el sistema (por ejemplo, PAL 768 x 576 píxeles).

Las cámaras digitales también utilizan CCD's para generar la imagen, pero en este caso es la propia cámara la que la digitaliza. Su principal ventaja es que permiten resoluciones mucho mayores. Hoy en día son relativamente habituales cámaras de 12 Mega píxeles (4080x3072 píxeles). Un problema que se presenta al tomar imágenes de gran resolución es que la transmisión de la imagen a la computadora no es instantánea, siendo necesario unos pocos segundos para su visualización en la pantalla, lo que puede ser un problema cuando trabajamos con imágenes dinámicas.

El criterio más habitual para definir la calidad de una cámara digital es la resolución del sensor, es decir el número de elementos fotosensibles que tiene y que nos dan la máxima resolución de la imagen digital obtenida. Otra serie de factores que hay que tener en cuenta en una cámara son:

- **Sensibilidad espectral o eficiencia cuántica.** Mide la efectividad de un sensor para producir electrones a partir de los fotones incidentes. Varía en función de la longitud de onda. La mayoría de los sensores CCD estándar son más sensibles a las longitudes de onda verde y roja (550-800 nm) y menos sensibles a azul y ultravioleta (375-550 nm).
- **Ruido térmico.** Electrones generados en el sensor debido a que se eleva la temperatura en tiempos de exposición prolongados. Se elimina refrigerando la cámara.
- **Rango dinámico.** Relación entre las zonas de mayor y menor intensidad que se pueden detectar con el sensor. A mayor rango dinámico mayor sensibilidad de la cámara para detectar ligeras variaciones de intensidad luminosa.
- **Binnig.** Consiste en agrupar conjuntos de píxeles como si fuese uno sólo para aumentar la sensibilidad de la cámara. Para intensidades luminosas muy bajas puede ocurrir que la señal que recibamos en un sólo píxel esté por debajo del umbral de detección sin embargo si sumamos las señales recibidas en un conjunto de 4x4 píxeles es posible que superemos el umbral de detección y tengamos señal.

▪ 1.11.4 Formato de Imágenes

A la hora de archivar imágenes el primer factor que se debe de tener en cuenta es el formato. Existen una serie de formatos estándares que son legibles por la mayoría

de los sistemas operativos y programas de tratamiento de imágenes permitiendo compartir imágenes entre personas que utilizan diferentes equipos.

Los ficheros de imágenes digitales pueden tener un tamaño considerable y el tamaño también es un problema a la hora de transferirlas. Cada vez se utiliza más la red Internet como medio para enviar imágenes entre ordenadores distantes. Las velocidades de transmisión de la red nos condicionan el tamaño de nuestras imágenes. Si estas velocidades no son muy altas deberemos utilizar formatos comprimidos.

Los formatos de imágenes más usuales son:

- **TIFF.** Es probablemente el formato de imagen más universal. Soporta imágenes de hasta 48 bits por píxel, pudiendo ser comprimidas o no comprimidas. El formato comprimido utiliza algoritmos de compresión sin pérdida de información con relación al fichero original.
- **PCX.** Es un formato del sistema operativo DOS®. Soporta hasta imágenes de 24 bits y no está comprimido.
- **BMP.** Es el formato de imágenes de Microsoft®, usado frecuentemente en programas de Windows®. Soporta imágenes de 24 bits sin comprimir, por lo que su tamaño suele ser grande.
- **PICT.** Considerado el formato estándar de imágenes para Macintosh®. Soporta imágenes de 24 bits por píxel.
- **JPG.** Es el formato comprimido más popular, compatible con un gran número de plataformas. Los datos son comprimidos para eliminar información no detectable por el ojo humano. La eficiencia de la compresión es excelente pudiendo llegar a 1/20 o 1/30 del original, lo que hace que sea un formato muy aconsejable para almacenar un gran número de imágenes en un espacio de disco limitado, o para cuando se desea transmitir las. Soporta imágenes en color real de 24 bits por píxel. Es el formato más utilizado para imágenes en Internet.
- **GIF.** Es un formato comprimido muy utilizado para guardar iconos, gráficos e imágenes con pocos colores o en niveles de gris. Únicamente soporta imágenes de 8 bits por píxel (256 colores).
- **PNG.** Es un formato de imagen diseñado para Internet y pensado como el sucesor, en código abierto y libre, del formato GIF. Soporta imágenes de hasta 24 bits por píxel y utiliza un algoritmo de compresión sin pérdida de información, sin embargo sus tasas de compresión son menores que las de los formatos GIF y JPG.

▪ **1.12 Procesamiento Digital de Imágenes (PDI)**

En esta sección se verán algunas técnicas que involucran el tratamiento a las imágenes previo a la segmentación. Dicho de otra forma el pre-procesado (mejora-restauración) que se esquematiza en la figura 1.15.

Dentro de lo que se denomina Procesamiento Digital de Imágenes se engloban una serie de técnicas que comprenden operaciones cuyo origen es una imagen y cuyo resultado final es otra imagen.

El valor del píxel en la imagen de salida puede ser función del valor que tenía en la imagen de entrada, de los valores de sus vecinos o del valor de todos los puntos de la imagen de entrada.

Según González, R. C. & Wintz, P. [4], el objetivo de estas técnicas es procesar una imagen de tal modo que la resultante sea más adecuada que la imagen original para una aplicación específica. El término “específico” es importante porque establece que el valor de la imagen resultante esté en función del problema que se trata. Así, un método que es útil para realzar un determinado tipo de imágenes puede no serlo para otras.

Entre las operaciones que se pueden realizar en la fase de pre-procesamiento vamos a comentar brevemente las siguientes: conversión de los niveles de gris, transformaciones geométricas, transformación del histograma y los filtrados espacial y frecuencial.

▪ **1.12.1 Conversión de los Niveles de Gris**

Una transformación adecuada de los niveles de gris puede mejorar sensiblemente la calidad de una imagen cuando ésta es observada por un individuo.

Estas técnicas se han desarrollado empleando por una parte teorías de psicología visual humana y de otra, modelos de los dispositivos físicos de adquisición y presentación de imágenes. Su aplicación, hasta ahora, se ha dirigido hacia el área del procesamiento con observador humano y aunque es evidente que tiene influencia en las fases de interpretación automática, se ha estudiado poco esta problemática.

1.12.2 Transformaciones Geométricas

Tienen dos objetivos básicos: a) corrección de la perspectiva y b) reconstrucción tridimensional de los objetos de una escena.

Dentro de las denominadas **operaciones puntuales** están las **operaciones aritméticas**: suma, resta, multiplicación y división [figura 1.21] así como las combinaciones de varias imágenes para obtener como resultante su máximo, mínimo, media, etc.

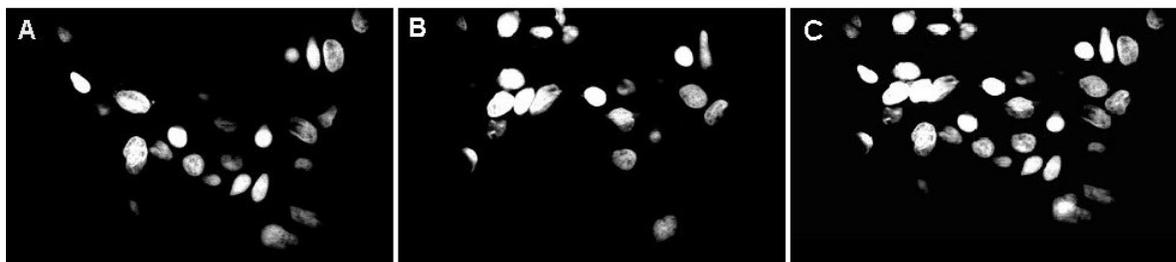


Figura 1.21 Ejemplo de operación aritmética. Suma de imágenes, $A+B=C$

1.12.3 Transformaciones del Histograma

Se definen como aquellas en las que el valor de un píxel en la imagen de salida depende del valor de ese mismo píxel en la original. Dentro de las transformaciones de histograma están: la **normalización**, la **linearización** y el **escalado**.

El histograma de niveles de gris de una imagen representa una distribución de los valores de los píxeles (niveles de gris) que forman la imagen. Normalmente en el eje de las X se representan los valores de gris (entre 0 y 255) y en el eje de las Y el número de píxeles que hay para un valor determinado. Imágenes poco contrastadas, cuyo rango de niveles de gris es escaso, se pueden mejorar realizando una expansión de su histograma de niveles de gris. La normalización consiste en expandir el histograma para que ocupe todo el rango posible de valores. Así en la figura 1.22 tenemos una imagen cuyo histograma ocupa un rango de gris entre 30 y 70, al expandir el histograma entre 0 y 255 la imagen resultante aparece mucho más contrastada.

Una descripción detallada de otro tipo de operaciones de histograma se puede encontrar en Pertusa, J. F. [5].

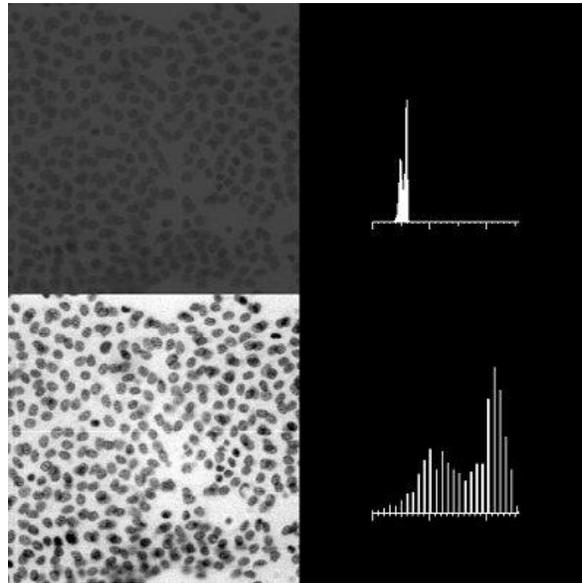


Figura 1.22 Imagen poco contrastada e imagen resultante de la expansión de su histograma

En un histograma, en el eje de abscisas se representan los niveles de gris y en el de ordenadas, el número de píxeles de cada nivel.

En ciertas escenas, como las que suelen aparecer en entornos industriales, el histograma puede ser suficiente para separar objetos dentro de una imagen, con lo que se facilita considerablemente la etapa de interpretación.

Sin embargo, no siempre es factible sacar tanto jugo al histograma. En muchas aplicaciones, su manipulación sólo permite aumentar/disminuir el contraste de la imagen y el margen dinámico de los niveles de gris. Sus efectos son evidentes para un observador humano.

▪ 1.12.4 Filtrado

No es fácil decidir si esta operación, absolutamente vital en un sistema de visión por computadora, pertenece o no a la etapa de simple pre-procesamiento, dada la tremenda modificación que puede suponer respecto a la imagen original.

Entonces consideremos mejor que el filtrado es el procesamiento previo cuando la transformación operada sobre la imagen sea pequeña.

El **filtrado espacial** actúa sobre los niveles de gris de los píxeles de la imagen El **filtrado frecuencial** se basa en trabajar con la transformada discreta de Fourier de la

1. Conceptos Básicos

imagen (una operación muy costosa en tiempo y en memoria de computadora) a la que se aplica un filtro bidimensional.

En las operaciones puntuales el valor del píxel en la imagen de salida es función de sus vecinos más próximos y del suyo propio en la imagen original. Se suelen presentar como operaciones o convoluciones de matrices alrededor del píxel tratado.

La forma de aplicar estos filtros es como sigue:

- Se define un elemento estructural que define una pequeña área de la imagen, por ejemplo una matriz de 3x3 píxeles (figura 1.23).
- Se opera con los valores de la matriz y se asigna al píxel central, en la imagen de salida, el resultado de la operación definida. La operación puede ser obtener el valor de la mediana, la media, el máximo, el mínimo, etc. (figura 1.23).

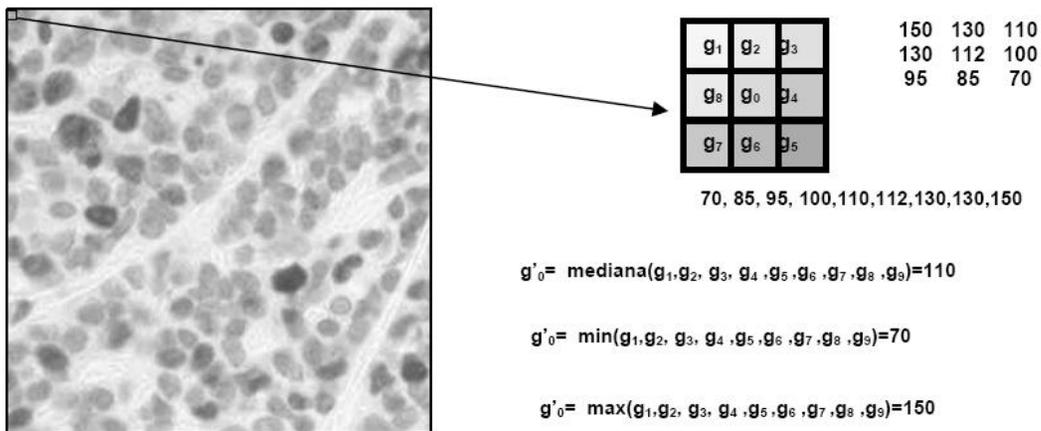


Figura 1.23 Matriz de 3x3 píxeles de una imagen y los valores de mediana, mínimo y máximo

El objetivo de estos procesos es modificar las imágenes para mejorar su calidad o resaltar aspectos de las mismas que nos interesan. El resultado de los filtros viene condicionado por el elemento estructural que utilizemos y el número de iteraciones (repeticiones) que realizamos el filtro (figura 1.24).

Los elementos estructurales circulares o próximos a la forma circular son los que mejor mantienen la forma de los objetos presentes en la imagen.

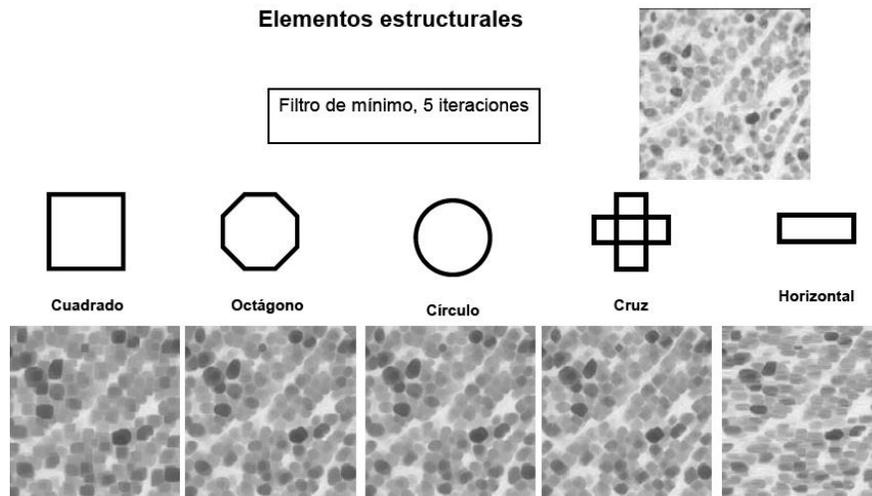


Figura 1.24 5 iteraciones de un filtro de mínimo con diferentes elementos estructurales

En función del resultado obtenido podemos hablar de filtros de suavizamiento que disminuyen las diferencias entre píxeles vecinos produciendo un difuminado de las líneas de contorno de los objetos y una especie de desenfoque general de la imagen. De este tipo es el denominado filtro de promedio que se basa en calcular el valor de la media de los píxeles que forman el elemento estructural (por ejemplo una ventana de 9x9) y asignarlo al píxel central en la imagen de salida (figura 1.25). Este filtro resulta muy útil para eliminar ruido de fondo en imágenes.

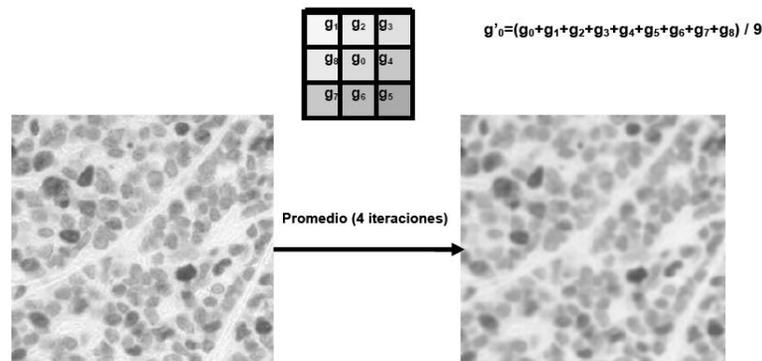


Figura 1.25 Resultado de un filtro de promedio

Cuando lo que nos interesa es enfatizar las zonas en que existen diferencias acusadas de niveles de gris debemos utilizar otro tipo de filtros denominado de realce o enfatizado.

Se basan en definir una matriz en la que el píxel central es positivo, mientras que sus vecinos son negativos o cero y la suma de los valores de la matriz de filtrado es la unidad (figura 1.26).

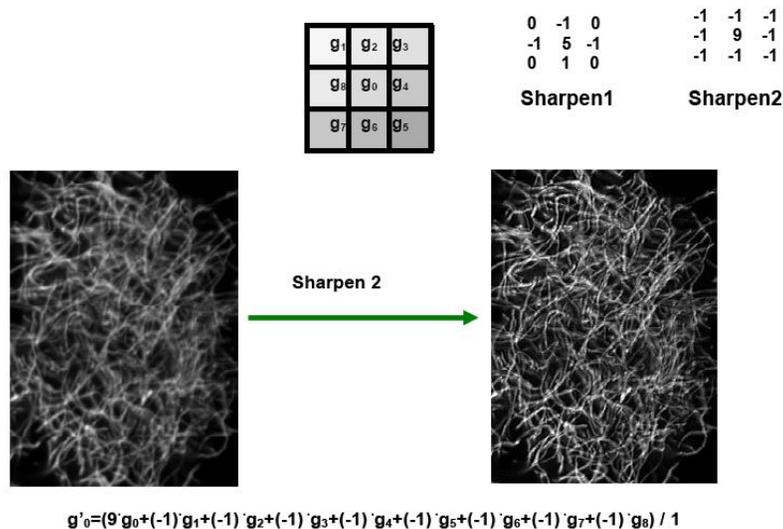


Figura 1.26 Definición de un filtro de enfatizado

Un tipo especial de filtros de realce son los denominados Laplacianos o de extracción de contornos. La matriz utilizada para estos filtros es muy similar a la anterior únicamente varía en que el resultado de la suma de sus valores es igual a cero [figura 1.27].

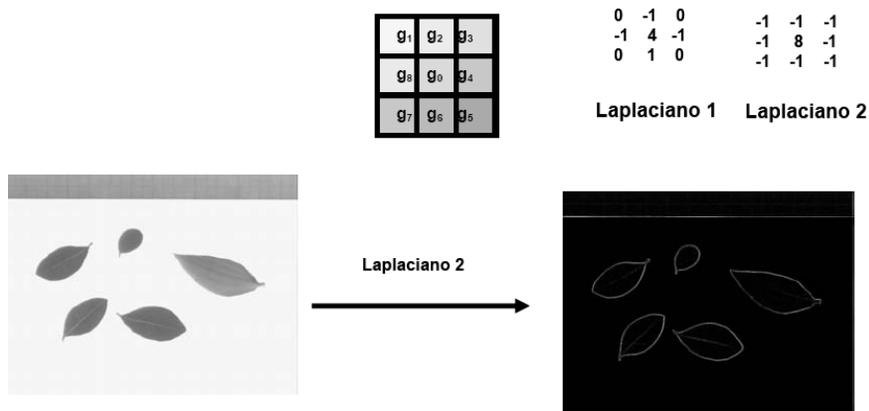


Figura 1.27 Definición de un filtro de extracción de contornos

La mayoría de las veces para obtener el resultado deseado es preciso combinar diferentes tipos de filtros junto con otros tipos de operaciones de imágenes como operaciones aritméticas, manipulación de histogramas, etc. Alguna de estas combinaciones da lugar a nuevos tipos de filtros como los denominados *tophat* que consisten en combinar filtros de máximo y mínimo y a la imagen resultante sustraerle la imagen original o de la imagen original sustraer la resultante. Este tipo de filtros son de gran utilidad para uniformizar el fondo de las imágenes y resaltar objetos de interés.

En la figura 1.28 se pueden ver dos variedades de *tophat* y su utilización para

resaltar objetos de interés del fondo de la imagen en función de que los objetos sean más claros o más oscuros que el fondo.

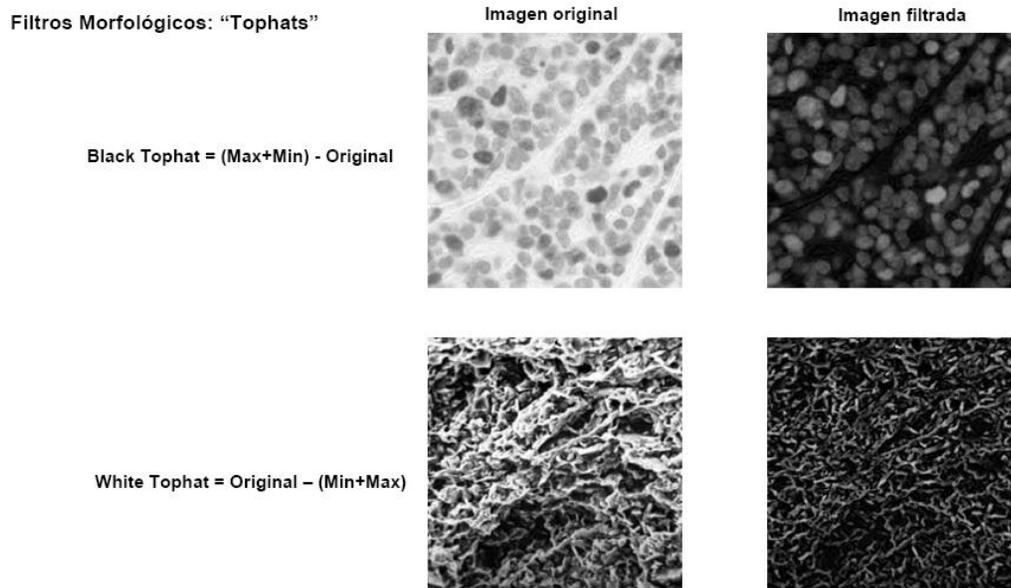


Figura 1.28 Definición de dos tipos de filtro de *tophat*

Otro tipo de operaciones denominadas **lineales** son los ajustes de contraste, brillo y ganancia en los que variando la curva de ajuste podemos modificar a voluntad cualquiera de estos parámetros (figura 1.29).

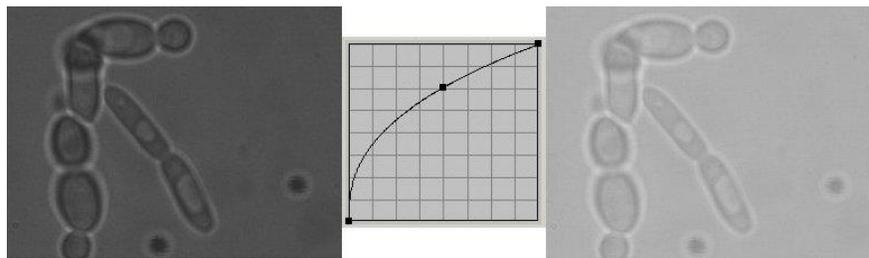


Figura 1.29 Aumento de la ganancia de una imagen manipulando la curva de ajuste

▪ 1.12.5 Paletas de Colores

Las paletas de colores o VLT (*Video Look-up Table*) son tablas que los programas de tratamiento de imágenes utilizan para modificar el aspecto de una imagen digital sin modificar los valores de sus píxeles. Una imagen en niveles de gris se puede visualizar en

falso color utilizando una paleta de colores en la que los valores de los píxeles se traduzcan por diversos colores en vez de por niveles de gris (figura 1.30).

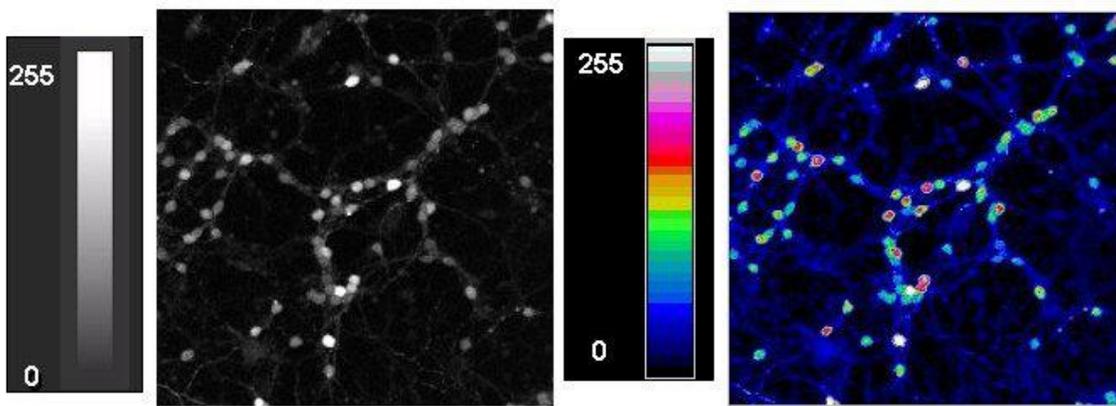


Figura 1.30. Imagen visualizada con una paleta de niveles de gris y en falso color utilizando una paleta de colores

El ojo humano únicamente puede distinguir unos pocos niveles de gris sin embargo es muy sensible a las variaciones de color, por tanto pequeñas variaciones en intensidades luminosas que no son apreciables si visualizamos una imagen en niveles de gris, se pueden apreciar claramente utilizando una paleta de color adecuada.

2. SEGMENTACIÓN Y EXTRACCIÓN DE CARACTERÍSTICAS

Recordemos que la segmentación es el proceso en el cual se aíslan los objetos en el universo de trabajo. Es la etapa más crucial en cualquier sistema de visión artificial, por las dificultades que conlleva y por la importancia de sus resultados.

Después de la segmentación, el resultado que se obtiene es un vector de observación [o vector de medidas] que contiene los valores numéricos a ser evaluados.

Supongamos, hipotéticamente para este ejemplo, que tenemos un lector óptico de caracteres que pone imaginariamente cada caracter en una cuadrícula como la mostrada en la figura 2.1. En esa cuadrícula se tienen 25 cuadrillos, o sea, una matriz de 5x5. La salida eléctrica de cada uno se verá incrementada mientras más aparezca el caracter. Entonces se puede definir un vector X de dimensión 25 compuesto de los valores numéricos de cada cuadrillo, este vector X es precisamente el vector de observación y el espacio multidimensional. En nuestro ejemplo, tenemos el vector de observación X de dimensión 25 de la letra J.

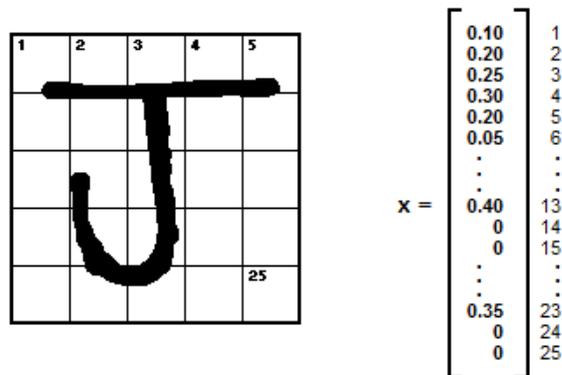


Figura 2.1 Ejemplo del Vector de Observación

2.1 Segmentación de Imágenes

Para el análisis de imágenes la segmentación se define como el proceso orientado a particionar la imagen digital en zonas disjuntas con significado propio. Este significado dependerá de la imagen, así como de los objetivos que se persigan con su interpretación.

Segmentar una imagen digital significa dividirla en zonas disjuntas e individualizadas. En cierta manera, equivale a su reconocimiento, puesto que como consecuencia de la segmentación los diferentes objetos (entendidos como realizaciones físicas de clases o patrones abstractos) se encuentran perfectamente ubicados dentro de la imagen digital.

La segmentación o detección de objetos es el proceso por el cual los píxeles pertenecientes a una imagen son divididos en clases, normalmente dos: objetos de interés y fondo. El fin último de la segmentación es el reconocimiento de objetos y en muchos casos su cuantificación.

En la práctica suelen distinguirse dos segmentaciones: de regiones o zonas y de objetos. La operación es común, lo que varía es aquello que se desea segmentar. En un caso, zonas; en el otro, ya dentro de una región concreta, objetos individuales.

En general, las operaciones de segmentación de una escena, así como el número de clases de regiones y objetos a distinguir, dependerá íntimamente de la propia escena y de la información dentro de la imagen.

Las técnicas de segmentación, ya estén orientadas a objetos o a agrupaciones de objetos, se pueden dividir en tres grupos:

- **Aplicación de umbrales de nivel de gris.**
- **Agrupación por rasgos comunes.**
- **Extracción de bordes.**

La aplicación de umbrales se basa en el empleo del histograma. Básicamente, consiste en distinguir objetos individuales dentro de una escena a partir de sus diferentes niveles de gris.

La segmentación de imágenes mediante la agrupación (*Clustering*) se inspira en las propias técnicas de reconocimiento de formas.

La extracción de bordes es el método más empleado en la segmentación de imágenes (salvo en robótica, donde con frecuencia es suficiente la aplicación de umbrales).

La efectividad, desde el punto de vista del reconocimiento de objetos, de una operación cuyo resultado sea la obtención de los bordes de un objeto, cual es la segmentación, es evidente, puesto que la mayor cantidad de información se encuentra en los bordes. De hecho, la mayoría de los objetos pueden reconocerse a partir de su forma, que está fijada por sus bordes.

Los filtrados espaciales y frecuenciales, que se comentaron en el procesamiento previo, son precisamente operaciones de extracción de bordes, aunque lo hagan con suavidad. De manera análoga, los extractores de bordes empleados en la fase de segmentación pueden clasificarse en espaciales y frecuenciales.

Tras la segmentación de la escena, aparecen dos nuevas fases: extracción de los rasgos característicos de los objetos ya diferenciados y su identificación propiamente dicha.

La segmentación produce una imagen en la que, junto a las coordenadas que definen una posición dentro del plano de la imagen, cada píxel posee una etiqueta distintiva. Una vez segmentada la imagen se puede formar una lista de objetos sin más que agrupar todos los píxeles que posean la misma etiqueta y, al mismo tiempo, que estén conectados espacialmente.

Las técnicas de segmentación que se describirán a continuación funcionan aceptablemente, pero no todas pueden aplicarse con éxito más que a escenas no complejas. Y aunque resulte aventurado realizar una clasificación de este tipo, por escenas no complejas o sencillas entenderemos aquellas que no posean objetos solapados y que dichos objetos tengan niveles de intensidad uniformes o cuasi uniformes.

Es obvio que con las dos restricciones anteriores el universo de trabajo se nos reduce drásticamente en relación al universo físico, pero aún así el rango de posibles aplicaciones de la visión artificial es muy importante y amplio (mucho más de lo que actualmente es en la práctica). Revisaremos algunas técnicas de segmentación.

▪ **2.1.1 Umbralización basada en el Histograma**

Básicamente, un histograma nos permite conocer la frecuencia relativa de aparición de cada uno de los posibles niveles de intensidad dentro de la imagen en cuestión.

Si recordamos que el proceso de segmentar una imagen digital no es otra cosa que agrupar los píxeles (es decir, los elementos básicos y puntuales de la imagen), el histograma nos servirá para agrupar los píxeles en función de su nivel de intensidad luminosa $I(x, y)$, para $x = 0, 1, 2, \dots, N-1$ e $y = 0, 1, 2, \dots, N-1$.

Para entender esta idea vamos a apoyarnos en la figura 2.2, en donde aparece una imagen muy sencilla, formada por un objeto (en este caso, oscuro) y en fondo uniforme. En la misma figura se ha representado el histograma de la imagen.

Este histograma es claramente bimodal (es decir, presenta dos máximos y un único mínimo), ya que los píxeles de la imagen se distribuyen en dos clases o regiones: el objeto y el fondo. En este caso es relativamente sencillo calcular un valor umbral que discrimine de manera óptima las dos clases existentes.

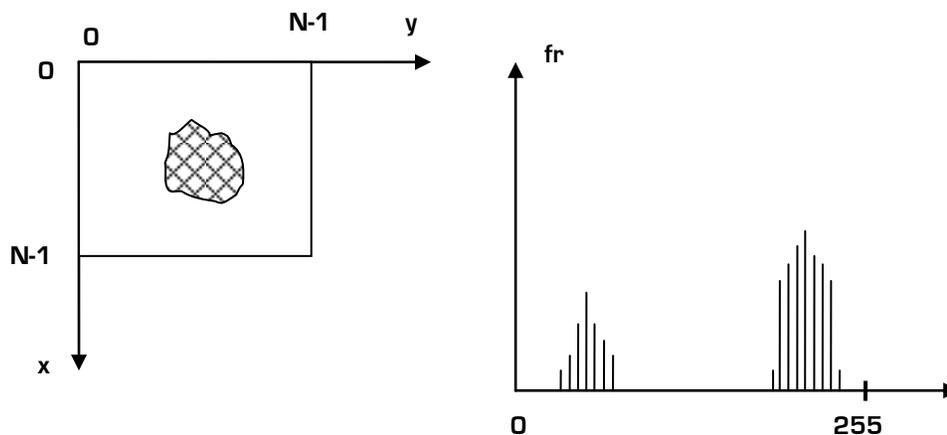


Figura 2.2 Imagen digital de un objeto oscuro sobre un fondo claro y su histograma

La umbralización basada exclusivamente en los niveles de luminosidad o de gris de los píxeles de la imagen se denomina global, en contraposición a una umbralización local en la que aparte de esa propiedad global de un píxel se utiliza información local del mismo; es decir, propiedades de un píxel que dependen de su localización en la imagen.

En la umbralización global existen dos procedimientos prácticos a la hora de la implementación en computadora:

- Umbralización basada en la búsqueda de mínimos.
- Umbralización basada en las técnicas de reconocimiento de patrones o formas.

Umbralización basada en la búsqueda de mínimos. Se trata de un procedimiento heurístico cuyo objetivo es localizar niveles de intensidad luminosa que umbralicen adecuadamente el histograma de una imagen. Ya hemos visto en la figura 2.2 que los píxeles se agrupaban, aproximadamente, en lóbulos individuales dentro del histograma. Cada lóbulo está formado por los niveles de intensidad de los píxeles de objetos físicos equivalentes (es decir, pertenecientes a la misma clase o patrón). Como dos lóbulos consecutivos definen un valle de separación entre ellos, el nivel de intensidad más bajo del valle, en principio, podría aplicarse como umbral de separación entre los dos lóbulos [o clases] adyacentes.

El principal inconveniente de este método proviene de la naturaleza altamente ruidosa del histograma, en donde aparecen multitud de falsos mínimos y falsos máximos,

lo cual redundaría negativamente en la efectividad del procedimiento y más especialmente en el tiempo de cálculo.

Para detectar, y por tanto rechazar, falsos objetos, es necesario realizar un estudio previo, durante la fase del diseño del sistema final, de la distribución de los niveles de intensidad luminosa de los posibles tipos de objetos a reconocer, lo cual limita el alcance y la robustez del sistema ante condiciones dinámicas en la captación de las imágenes a segmentar.

Para evitar estos inconvenientes, es habitual procesar previamente la imagen digital, a fin de disminuir en lo posible el ruido del histograma. En particular, se busca aumentar el contraste de la imagen a reconocer.

La aplicación de un filtro de paso bajo convenientemente sintonizado a una frecuencia de corte adecuada, elimina los componentes de alta frecuencia del histograma, es decir, aquellos producidos por cambios bruscos del histograma en intervalos reducidos.

Umbralización basada en las técnicas de reconocimiento de patrones o formas. Si se analiza la forma que presentan los histogramas, por ejemplo de la figura 2.2, podrá observarse que los lóbulos que en ellas aparecen recuerdan a la campana de Gauss, típica de las distribuciones normales.

Pues bien, este segundo enfoque se basa en considerar que cada lóbulo del histograma corresponde a la distribución en probabilidad de una clase o patrón.

El problema práctico a la hora de segmentar una imagen con este método es calcular las medias y las desviaciones típicas de cada lóbulo. La principal dificultad consiste en delimitar el rango de los lóbulos, es decir, la amplitud de las correspondientes campanas de Gauss. Generalizando al caso de n objetos, aparecerán tantas campanas de Gauss como clases de objetos.

Hasta ahora se ha planteado la umbralización del histograma bajo un enfoque global; es decir, manipulando la imagen completa. Este enfoque puede conducir a una segmentación incorrecta, especialmente si las condiciones de iluminación son defectuosas o cambiantes, de tal manera que los niveles de intensidad luminosa no se distribuyen de forma homogénea en la imagen.

Para evitar los problemas de la umbralización global, se recurre a una umbralización local en la que la clasificación de un píxel se realiza atendiendo a propiedades locales del píxel (es decir, que dependen de su entorno), en vez de la propiedad global que es su nivel de intensidad luminosa.

El procedimiento que se sigue habitualmente a la hora de implementar una

umbralización local es trabajar con el histograma de una subregión de la imagen global, existiendo básicamente dos posibilidades: a) dividir la imagen en un número de subregiones o ventanas cuadradas disjuntas y b) definir una ventana cuadrada centrada en el píxel a clasificar. Este último método implica una elevada carga computacional, por lo que puede ser impracticable en ciertas situaciones.

Una vez establecidas las subregiones de la imagen global (por cualquiera de los dos métodos) la umbralización se realizaría siguiendo las mismas ideas que para la umbralización global.

▪ 2.1.2 Agrupación de Píxeles

Las técnicas de agrupación (*Clustering*) de datos se diferencian del resto de las técnicas de clasificación automática en que no son supervisadas, lo cual es una propiedad muy interesante desde el punto de vista de la segmentación de imágenes digitales.

El precio a pagar es un menor rendimiento de estas técnicas en comparación con las técnicas supervisadas, como es bien sabido.

La idea básica consiste en definir un conjunto de características asociadas a un píxel; es decir, emplear un vector X de características para los píxeles de la imagen a segmentar. En sentido estricto, este es el proceso que se sigue siempre en la segmentación de una imagen, pero en este caso el vector X de características discriminantes asociado a cada píxel ya no es unidimensional, sino de dimensión superior.

La referencia clásica en la segmentación mediante *Clustering* es el trabajo pionero de Coleman y Andrews aparecido en el año 1979. Estos autores plantearon de manera sistemática por vez primera en la literatura especializada un procedimiento de segmentación de imágenes digitales utilizando las técnicas de *Clustering*. En ese mismo trabajo aplicaron sus ideas a dos tipos de imágenes: una en blanco y negro y otra en color.

Para la segmentación de la imagen monocolor emplearon doce características discriminantes: la intensidad luminosa, el módulo y la fase de un filtro especial para la extracción de bordes, las modas de las intensidades luminosas en unas ventanas de 3×3 , 7×7 y 15×15 , respectivamente, centrados en el píxel a agrupar, etc.

Para la segmentación de la escena en color se emplearon quince características, similares a las anteriores, pero teniendo en cuenta los tres canales de información visual básica: rojo, verde y azul.

Sin entrar en detalles, lo realmente notable es el enfoque que estos dos autores propusieron para segmentar una imagen.

▪ **2.1.3 Cálculo de los Bordos**

Esta técnica es una de las más populares y sencillas. Se inspira en un principio muy intuitivo y simple: puesto que el objetivo último de un sistema de reconocimiento visual automático es clasificar los objetos individuales presentes en una imagen, utilizando para ellos atributos o características discriminantes de cada objeto, parece lógico delimitar la frontera (el contorno) de los objetos como paso obligado para el cálculo de los atributos discriminantes.

Pues bien, el cálculo o extracción de los bordes de los objetos se puede plantear como una operación de filtrado de la imagen original, más concretamente, de un filtrado de paso alto.

En efecto, puesto que los píxeles situados en los bordes de un objeto presentan grandes variaciones del nivel de intensidad luminosa en zonas muy pequeñas de la imagen (piénsese en un objeto oscuro situado sobre un fondo claro), es necesario aplicar un filtro de paso alto (es decir, que sólo deje pasar las altas frecuencias de la señal filtrada), pues en tal caso en la imagen resultante aparecerán resaltados los píxeles de los bordes. El principal problema práctico está en determinar la frecuencia de corte adecuada.

Los filtros espaciales, que ya vimos, son aproximaciones a los filtros frecuenciales y tienen como ventaja esencial su sencillez de cálculo.

Para la detección y realce de los bordes, se aplican habitualmente filtros espaciales lineales de dos tipos: a) basados en el gradiente de la función de luminosidad de la imagen y b) basados en su Laplaciana.

▪ **2.1.4 *Thresholping***

Un método sencillo de segmentación es el denominado ***thresholping***. Se utiliza para convertir una imagen de niveles de gris en una imagen binaria. Cada píxel es clasificado como "ON" blanco (valor 255) u "OFF" negro (valor 0) dependiendo de si su nivel de gris excede o no un valor umbral. La selección del umbral se realiza a partir de un estudio del histograma de niveles de gris de la imagen. Si tenemos una imagen que

presenta objetos con valores de gris claro sobre un fondo oscuro y visualizamos el histograma de distribución de niveles de gris de dicha imagen se observa como el histograma es claramente bimodal con dos picos bien diferenciados, el primero corresponde al fondo y el segundo a los objetos (figura 2.3 A). En este caso el proceso de clasificación consistirá en tomar como umbral el valle entre ambos picos y dar el valor 255 a los píxeles con un valor superior al umbral dejando con el valor 0 los restantes (figura 2.3 B).

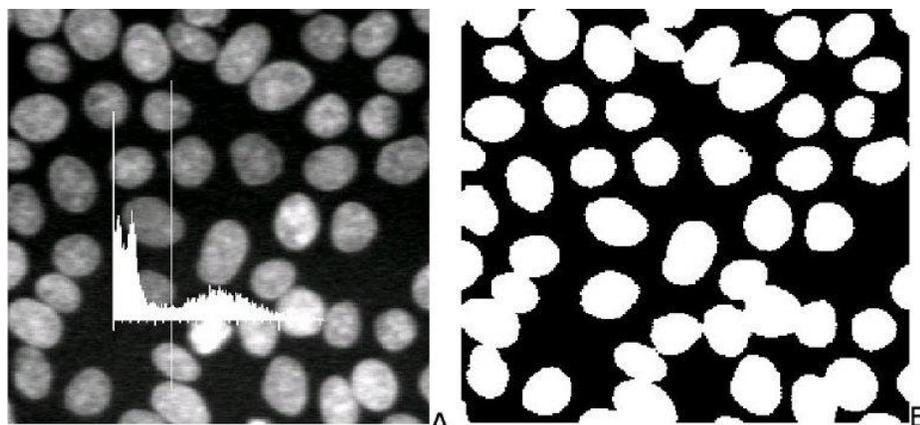


Figura 2.3 A) Imagen de niveles de gris y su histograma, B). Imagen segmentada

La segmentación puede realizarse de modo interactivo visualizando como queda la imagen según varía el valor del umbral, o de modo automático, donde el ordenador determina el valor de los umbrales de segmentación siguiendo diversos criterios estadísticos. [6]

▪ 2.1.5 *Watershed*

Watershed o **segmentación por línea de valle** es una técnica muy usada para separar objetos que se tocan. Supongamos unos objetos que se tocan y que deseamos separar (figura 2.4 A). Para empezar debemos calcular la imagen de distancias, donde cada píxel que estaba en "ON" en la imagen binaria toma el valor de su distancia al borde del objeto, así los píxeles del borde tendrán el valor 1, los siguientes el valor 2 y así sucesivamente hasta los píxeles más internos que tomaran los valores más altos (figura 2.4 B). El resultado será una imagen similar a un mapa topográfico donde los valores de los píxeles se asemejan a curvas de nivel. Si representamos la imagen en tres dimensiones (figura 2.4 C) observaríamos como tres montañas correspondientes cada una a uno de los objetos. Buscando las líneas de valle y poniéndolos a cero en la imagen binaria, tendríamos separados los tres objetos (figura 2.4 D).

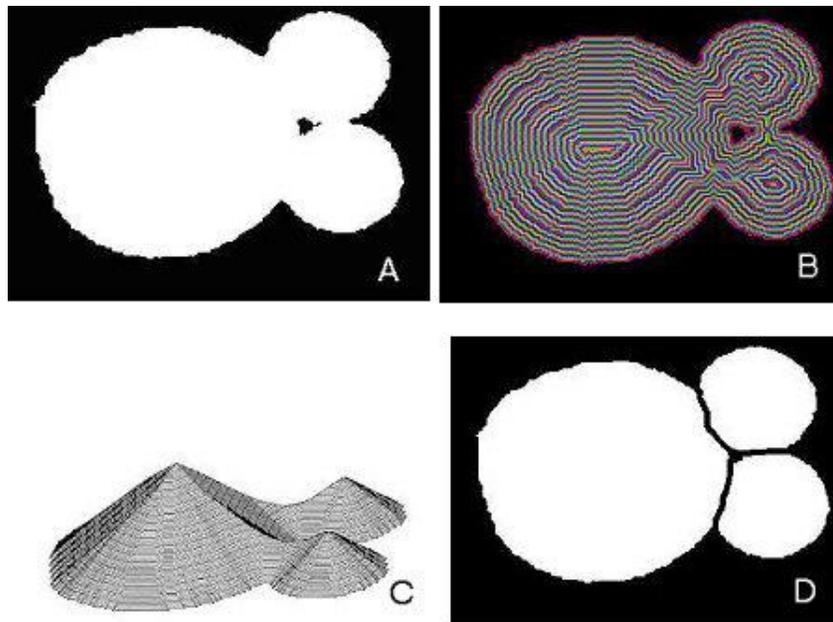


Figura 2.4 Segmentación por *Watershed*. A) Imagen original. B) Mapa de distancias. C) Representación tridimensional. D) Resultado de la Segmentación

▪ 2.1.6 Operaciones Morfológicas

La imagen binaria resultante del proceso de segmentación puede que no represente perfectamente los objetos de interés debido a problemas inherentes en la imagen original, como objetos que se tocan (figura 2.3 B), o problemas del propio proceso de las imágenes, por ejemplo, pequeñas áreas mal clasificadas. Con el fin de solucionar estos problemas se pueden aplicar distintos operadores morfológicos para: separar zonas, cribar áreas de determinados tamaños, erosionar o dilatar regiones, mirar conectividades, extraer bordes, etc.

Dos de las operaciones morfológicas más utilizadas en análisis de imágenes son la **erosión** y la **dilatación**. La primera consiste en examinar cada píxel y cambiarlo de “ON” a “OFF” si alguno de sus vecinos está en “OFF”. Normalmente se utilizan como vecinos los ocho que rodean al píxel examinado, aunque para algunas aplicaciones se pueden utilizar conectividades de 4 vecinos (los 2 verticales y los 2 horizontales) e incluso conectividades de 2 vecinos (los verticales o los horizontales).

La dilatación es el proceso inverso, consiste en cambiar los píxeles de “OFF” a “ON” si alguno de sus vecinos esta “ON”. Al resultado de una erosión más una dilatación se le denomina **apertura** (*opening*). El nombre proviene de la tendencia de esta secuencia de operaciones a separar (abrir) puentes de unión entre objetos próximos o a

abrir cavidades próximas al borde. El proceso puede utilizarse para eliminar pequeños objetos de la imagen. La operación opuesta (dilatación más erosión) es denominada **cierre** (*closing*) y puede usarse para conectar objetos muy próximos o para rellenar pequeños huecos (figura 2.5). En este caso, la imagen se ha invertido, apareciendo los objetos de interés en negro y el fondo en blanco.

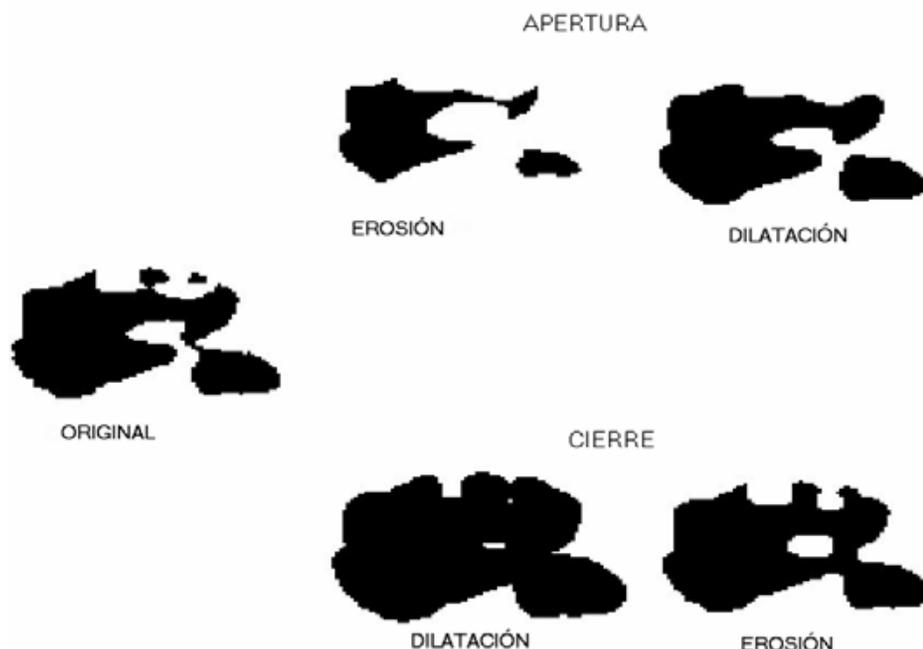


Figura 2.5. Operaciones de apertura (*opening*) y cierre (*closing*).

Otro tipo de operaciones con imágenes binarias son las combinaciones lógicas: **AND**, **OR** y **XOR** (figura 2.6).

- La operación **AND** consiste en poner en "ON" los píxeles que en las dos imágenes estén en "ON" y a "OFF" el resto.
- La operación **OR** pone en "ON" los píxeles que en alguna de las imágenes estén en "ON".
- La operación **XOR** pone en "ON" los píxeles que están en "ON" en alguna de las imágenes pero no en ambas.

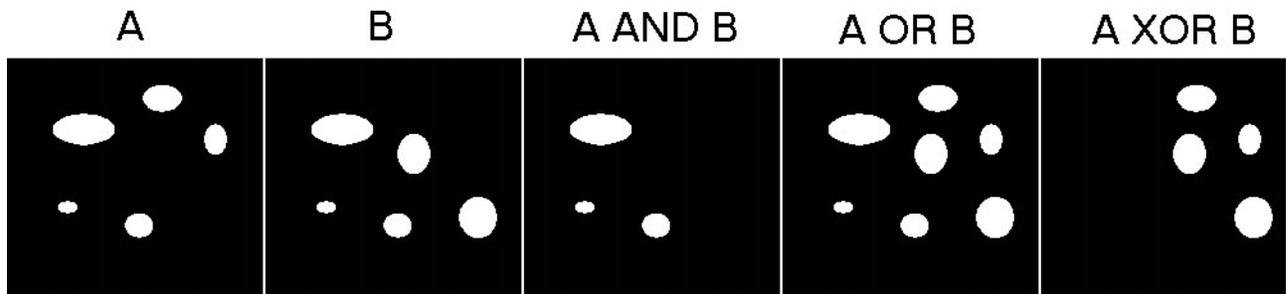


Figura 2.6 Operaciones lógicas: AND, OR y XOR

En muchos casos deberemos de combinar varias operaciones binarias para obtener el resultado esperado.

El *thinning* o **esqueletización** es otro tipo de operación morfológica, consistente en obtener la línea media o eje de un objeto. Este tipo de procedimiento reduce los objetos a un conjunto de líneas sin perderse la conectividad del objeto original (figura 2.7). Resulta de gran utilidad su aplicación al estudio de estructuras alargadas en las que queremos calcular parámetros como: longitudes, orientaciones, conectividades, etc.

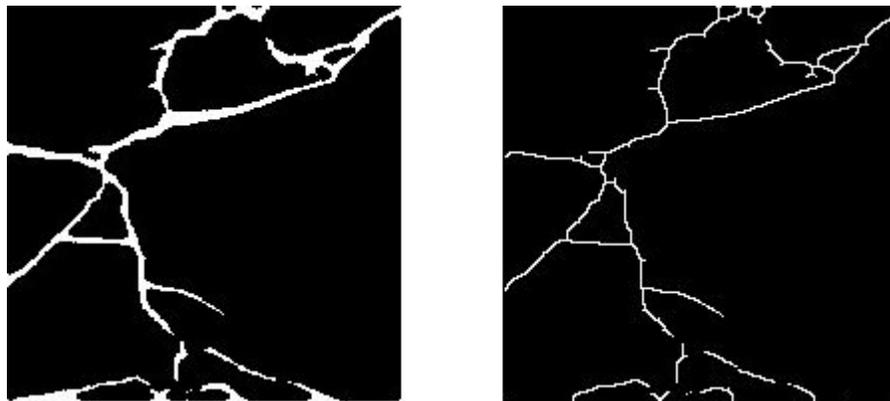


Figura 2.7 Imagen binaria de una red de fisuras e imagen resultante de un proceso de *thinning*

Aunque con un proceso de apertura es posible, en algunos casos, separar objetos que se tocan, la mayoría de las veces es preciso utilizar un procedimiento más complejo para separarlos.

En la bibliografía citada se puede encontrar una completa descripción sobre técnicas de morfología matemática.

2.1.7 Imágenes en Color

En las imágenes en color también es posible aplicar criterios de segmentación para transformarlas en imágenes binarias. Cuando se utilizan marcadores específicos en imágenes histológicas podemos utilizar el color como criterio para segmentar la imagen y obtener los objetos teñidos con ese marcador.

En una imagen en color se trabaja con tres umbrales de segmentación, uno para cada banda del espectro. En la figura 2.8 puede observarse una imagen histológica en color en la que algunos elementos aparecen en color marrón y el resto en color azul. Los umbrales utilizados para segmentar los elementos en marrón han sido 230 para la banda roja, 200 para la verde y 166 para la azul. En la mayoría de los programas de análisis de imágenes los valores de estos umbrales se pueden obtener interactivamente pulsando con el ratón sobre la zona que queremos segmentar.

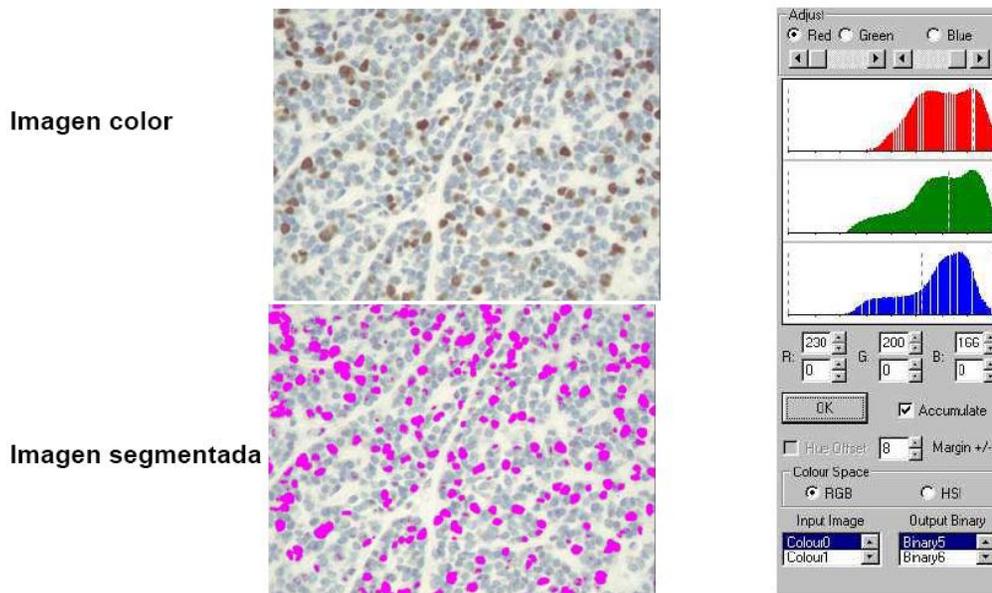


Figura 2.8 Imagen en color de un tejido e imagen segmentada con los elementos en color violeta.
Histogramas de las intensidades

2.2 Extracción de Características

Dado un cierto objeto físico dentro de una imagen, y a partir de su contorno (resultado de la segmentación), puede obtenerse de inmediato un conjunto de parámetros característicos, como perímetro, grado de redondez, momentos, etc.

Estos rasgos hacen posible distinguir objetos mediante el uso de un número reducido de propiedades.

Los rasgos de un objeto forman un vector X (valores numéricos), particularmente, los del objeto a clasificar, determinan un punto en el espacio n -dimensional de las características.

Las propiedades deseables en los rasgos son, principalmente: número no elevado, no estar relacionados entre sí al máximo, discriminación abrupta de objetos, etc., propiedades todas ellas orientadas a realizar una clasificación automática rápida y fiable.

A continuación se describirán algunas características que pueden ser empleadas para describir cuantitativamente regiones que hayan sido segmentadas en una imagen. Se entenderá por región aquel conjunto de píxeles que pertenezcan a una misma zona de la imagen y que esté limitado por bordes. Se asumirá que los bordes no pertenecen a la región.

Para explicar estas características se muestra a continuación un ejemplo de la figura 2.9. En él se presenta una región circular que ha sido segmentada. La región entonces se conforma por los píxeles que pertenecen al círculo (pero no a su perímetro), es decir los píxeles que han sido marcados con color gris. Los bordes de la región definen el límite de la región.

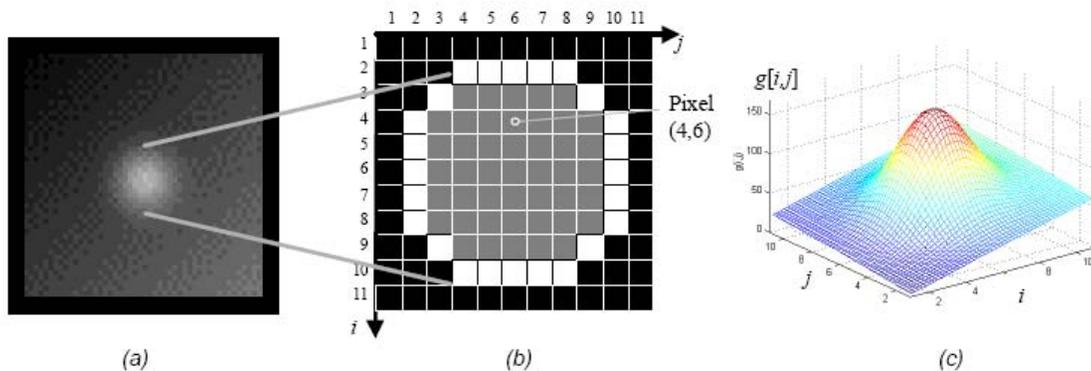


Figura 2.9 Ejemplo de una región. A) Imagen, B) Región segmentada, C) Representación 3D de los valores de gris en la región y su entorno

▪ 2.2.1 Cuantificación

Para llegar a la cuantificación es preciso contar con la imagen segmentada donde tengamos perfectamente identificados los objetos. Una vez que las imágenes

binarias muestran de una forma individualizada los objetos que deseamos cuantificar, se procederá a su identificación y medida [descripción]. Los parámetros o características que se pueden extraer de una región se pueden clasificar en:

- **Características Geométricas.** Contienen información de forma, posición, tamaño, orientación de la región y relaciones espaciales de los objetos.
- **Características Densitométricas.** Miden nivel de gris o parámetros derivados del nivel de gris (transmitancia, densidad óptica, etc.).
- **Características Cromáticas.** Son características relacionadas con el color.

Para que los valores medidos se obtengan en unidades reales (micras, milímetros, metros, etc.) es preciso realizar previamente una calibración de la imagen. La calibración consiste en calcular el tamaño del píxel en unidades reales.

Para ello se digitaliza una imagen en la que se conozca una longitud. Una vez digitalizada la imagen se marca la distancia conocida y se da su valor, como el programa conoce cuantos píxeles hay en la distancia marcada nos puede calcular el tamaño del píxel en unidades reales. A este valor también se le denomina resolución del píxel y se define como la distancia más pequeña que podemos discriminar en nuestra imagen.

▪ 2.2.2 Características Geométricas

Dentro de éstas podemos distinguir entre **características globales**, aquellas que se aplican a una determinada estructura pero sin distinguir detalles de cómo se organizan sus componentes y **características del objeto**, que cuantifican tamaño y forma de cada objeto considerado individualmente. Se enumerarán las más comunes en reconocimiento de patrones.

Características globales:

- **Porcentaje.** Porcentaje de un determinado componente en la imagen. Este porcentaje calculado en el plano de imagen es equivalente al porcentaje en volumen de dicho componente.

$$V_v = \frac{\sum pixels(componente)}{\sum pixels(totales)} \times 100$$

- **Superficie específica.** Superficie de un determinado componente de la imagen por unidad de volumen. Para una sección plana, el valor de la superficie específica viene dado por:

$$S_v = \frac{4 \sum B_i(obj)}{\eta \sum A_i(ref)}$$

donde:

$\sum B_i(obj)$ = longitud de los bordes del componente

$\sum A_i(ref)$ = área de referencia.

- **Número por unidad de área.** Número de objetos individualizados que tenemos por unidad de área.

$$NumArea = \frac{Num}{\sum A_i(ref)}$$

Cuando se trata de contar el número de objetos por unidad de área es preciso tener en cuenta aquellos que aparecen parcialmente en la imagen. Supongamos la imagen de la figura 2.10 A. Si contamos todos los objetos, incluidos los que aparecen parcialmente estaremos sobrevalorando el número de objetos por área. Si sólo contamos los que aparecen completos en la imagen (figura 2.10 B), estaremos minusvalorando el número de objetos por área. La solución correcta es eliminar los que tocan dos de los bordes de la imagen (figura 2.10 C), por ejemplo el borde inferior y el derecho y dejar los que tocan los otros dos.

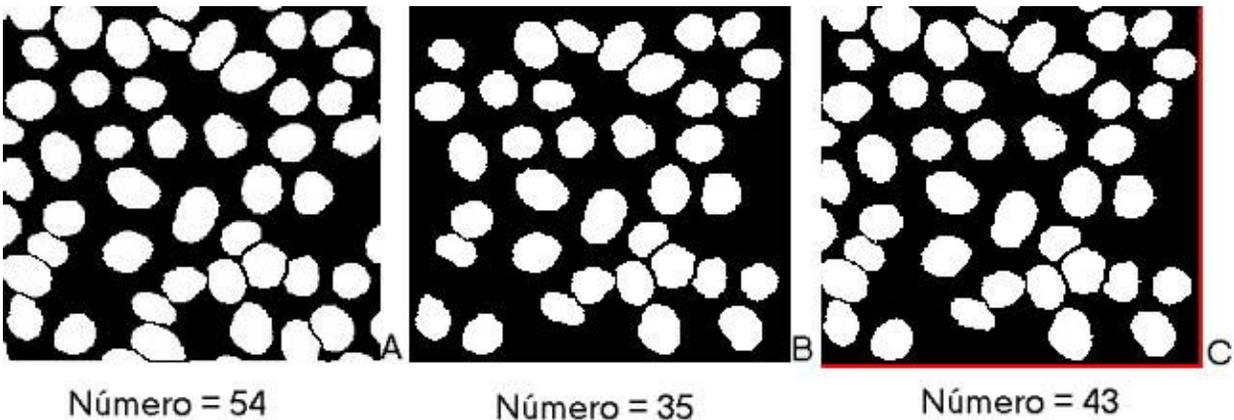


Figura 2.10 A) Distintos valores de número obtenidos según contemos todos los objetos de la imagen.
 B) Eliminando todos los objetos que tocan los bordes. C) Eliminando los que tocan dos de los bordes

Características del objeto:

- **Área.** Se calcula contando el número de píxeles que forman cada objeto y multiplicándolo por la resolución del píxel al cuadrado.



$$Area = NPIX * Resol^2$$

- **Perímetro.** El perímetro de una región puede ser definido de varias maneras. Una definición práctica, más no exacta, es tomar el perímetro como el número de píxeles que pertenecen al borde de la región. O también puede obtenerse de las coordenadas del borde de cada objeto, calculando su longitud y multiplicándola por la resolución del píxel.



$$P = (\sum v + \sum h + \sum d * \sqrt{2}) * Resol$$

donde:

v = segmento vertical

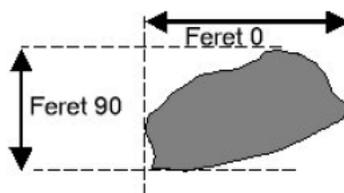
h = segmento horizontal

d = segmento diagonal

- **Perímetro convexo.** Se define como la longitud de una línea convexa que circunscribe el objeto a medir.



- **Feret 0, Feret 90.** Estos parámetros también conocidos como Feret x y Feret y se definen como la distancia entre un par de líneas paralelas, tangentes al objeto en un determinado ángulo.



- **Diámetro máximo y mínimo (longitud y anchura).** Se obtienen calculando los valores máximos y mínimos de los Feret en varias direcciones (normalmente 64). A partir de las coordenadas que definen el diámetro máximo es posible calcular su ángulo con la horizontal, lo que permite conocer la orientación del objeto.



$$D_{\max} = \text{Max}[DP(l)]_{l=1..64}$$

$$D_{\min} = \text{Min}[DP(l)]_{l=1..64}$$

DPI= feret para una dirección determinada

Otra forma para calcular la longitud [h] y anchura [w] es:

$$w = j_{\max} - j_{\min} + 1$$

$$h = i_{\max} - i_{\min} + 1$$

donde:

i_{\max}, j_{\max} = valor máximo que toma i y j en la región

i_{\min}, j_{\min} = valor mínimo que toma i y j en la región

- **Forma Circular (FC).** También se conoce como redondez. Se basa en relacionar la superficie del objeto con la de un círculo de igual área. El valor siempre estará entre 0 y 1. Teóricamente $R = 1$ para un círculo perfecto y $R = 0$ para una región que tenga altura y/o ancho igual a cero. En la práctica, sin embargo, debido al muestreo en el espacio de la región estos valores presentan desviaciones. Entonces decimos que $FC=1$ para objetos circulares.



$$FC = \frac{S_{\text{objeto}}}{S_{\text{circulo}}} = \frac{S}{\eta \frac{P^2}{4\eta^2}} = \frac{4\eta S}{P^2}$$

donde:

S = área del objeto

P = perímetro del objeto

- **Elongación (*Aspect Ratio*)**. Relación entre el diámetro mínimo y el diámetro máximo. Vale 1 para objetos equidimensionales y tiende a 0 para objetos alargados.



$$E = \frac{D_{\min}}{D_{\max}}$$

- **Forma Rugosa (FR)**. Este parámetro también denominado convexidad relaciona el perímetro convexo del objeto con su perímetro total.



$$FR = \frac{P_{\text{convex}}}{P}$$

- **Momentos**. Los momentos estadísticos se definen como:

$$m_{rs} = \sum_{i,j \in \mathbb{R}} i^r j^s \quad \text{para } r, s \in \mathcal{N}$$

donde:

\mathbb{R} = conjunto de píxeles de la región

El parámetro $r+s$ denota el orden del momento. El momento de orden cero m_{00} corresponde al área de la región. El centro de gravedad de una región queda definido por:

$$\bar{i} = \frac{m_{10}}{m_{00}} \quad \bar{j} = \frac{m_{01}}{m_{00}}$$

Con ayuda de las coordenadas del centro de gravedad se definen los momentos centrales que son invariantes al desplazamiento de la región en la imagen:

$$\mu_{rs} = \sum_{i,j \in \mathbb{R}} (i - \bar{i})^r (j - \bar{j})^s \quad \text{para } r, s \in \mathcal{N}$$

Muy conocidos en la teoría de reconocimiento de patrones son las características derivadas de los momentos centrales, denominados momentos de Hu:

$$\begin{aligned}
 \phi_1 &= \eta_{20} + \eta_{02} \\
 \phi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\
 \phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\
 \phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\
 \phi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + \\
 &\quad (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\
 \phi_6 &= (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + \\
 &\quad 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
 \phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - \\
 &\quad (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]
 \end{aligned}$$

con:

$$\eta_{rs} = \frac{\mu_{rs}}{\mu_{00}^t} \quad t = \frac{r+s}{2} + 1$$

Los momentos de Hu son invariantes a la traslación, rotación y escalamiento. Esto quiere decir que dos regiones que tengan la misma forma pero que sean de distinto tamaño y que estén ubicados en posiciones y orientaciones distintas en la imagen tendrán momentos de Hu iguales. A veces, sin embargo, es necesario contar con características que además sean invariantes a las transformadas afines.

Cuando se realiza una cuantificación con un sistema de análisis de imágenes es necesario conocer el procedimiento de cálculo de los diferentes parámetros para poder interpretar correctamente los datos obtenidos.

Supongamos que nos interesa calcular los diámetros máximo y mínimo de los objetos de la figura 2.11. Si utilizamos el procedimiento descrito anteriormente los valores obtenidos para el objeto A son correctos, sin embargo los del objeto B serían erróneos, obteniéndose unos valores de diámetro máximo inferior al real y de diámetro mínimo considerablemente superior a su valor real.

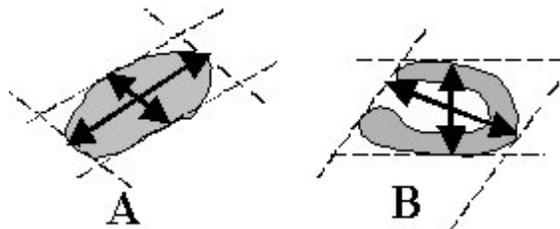


Figura 2.11 Cálculo de los diámetros máximo y mínimo de dos objetos de formas distintas

En el caso de los parámetros derivados (forma circular, elongación, forma

rugosa] el procedimiento de cálculo puede variar de unos sistemas a otros. Por ejemplo la elongación se puede definir como diámetro máximo partido por diámetro mínimo en cuyo caso los valores obtenidos serán iguales o mayores de 1.

▪ 2.2.3 Características Densitométricas

En ocasiones lo que se desea cuantificar no es el tamaño o forma de los objetos sino otro tipo de propiedades como pueden ser su intensidad luminosa (I), su transmitancia (T) o la densidad óptica (DO).

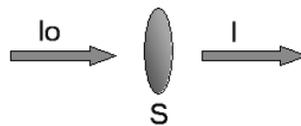


Figura 2.12 Intensidad incidente (Io), Intensidad transmitida (I) y superficie de un objeto (S)

- **Transmitancia.** El valor de la transmitancia viene dado por el cociente entre la intensidad transmitida y la intensidad incidente (Io) (figura 2.12).

$$T = \frac{I}{I_o}$$

- **Intensidad incidente (Io) y transmitida (I).** En una imagen digital con un rango de niveles de gris de 0 a 255 el valor de la intensidad incidente (Io) es igual a 255 y el valor de la intensidad transmitida (I) es el valor medio de gris de todos los píxeles que forman el objeto.
- **Densidad Óptica (DO).** El valor de la densidad óptica (DO) se define como:

$$DO = -\log T = \log \frac{I_o}{I}$$

- **Densidad Óptica Integrada (DOI).** Otro parámetro que se utiliza en las medidas densitométricas es la densidad óptica integrada (DOI) definida como:

$$DOI = DO * S$$

Si previamente hemos calibrado el aparato con una serie de patrones de densidades que queremos medir, podremos tener la equivalencia entre los niveles de gris y nuestras unidades densitométricas.

En el caso de las medidas densitométricas la imagen segmentada se utiliza como máscara para extraer los valores de gris, de las zonas de interés, en la imagen original.

En las medidas densitométricas es muy importante estandarizar las condiciones de tinción e iluminación de la muestra de tal forma que las variaciones en la densidad de la imagen sean debidas a las propias variaciones de densidad de la muestra y no a otros factores externos.

▪ 2.2.4 Características Cromáticas

Antes de definir las características del color es necesario saber si la imagen que se pretende analizar es a color o en blanco y negro. En el primer caso el color se descompone (por convención) en tres componentes RGB (rojo, verde y azul) para cada píxel de la imagen, en el segundo caso se cuenta sólo con el tono de gris en cada píxel, como lo hemos estado manejando.

Las características que se mencionan a continuación son para una sola variable de color. Esta variable puede ser cada una de las componentes del color, una combinación lineal de las tres componentes o bien, simplemente el tono de gris. La información entonces necesaria para calcular estas características es el valor de esta variable de color en cada píxel que es representada como $x[i, j]$ para el píxel (i, j) de la imagen. Es posible que en algunas aplicaciones sea necesario analizar de manera independiente dos variables de color, por ejemplo la componente en rojo y la componente en azul. En este tipo de aplicaciones sería necesario extraer las características de color para cada una de las variables de color requeridas.

- **Color Promedio.** Esta característica es el promedio de la variable de color que se define como:

$$G = \frac{1}{A} \sum_{i,j \in \mathcal{R}} x[i, j]$$

donde:

\mathcal{R} = conjunto de píxeles de la región

$x[i, j]$ = el valor de la variable de color en el píxel (i, j)

El número de píxeles de la región \mathcal{R} es A , el área de la región. Una

representación 3D de la variable de color de una región y su entorno se muestra en la figura 2.9 C. En ese caso se trabaja con el valor de gris, ya que la imagen es blanco y negro. Para ese ejemplo el promedio es $G = 121.90$ ($G = 0$ significa 100% negro y $G = 255$ corresponde a 100% blanco).

- **Gradiente promedio en el borde.** Esta característica toma el valor promedio del gradiente de la variable de color en el borde de la región. Con esta característica se puede medir qué tan abrupto es el cambio en la coloración en la región con respecto a su entorno. El gradiente promedio en el borde se calcula como:

$$C = \frac{1}{L} \sum_{i,j \in \ell} x'[i, j]$$

donde:

$x'[i, j]$ = módulo del gradiente de la variable de color del píxel (i, j)

Los píxeles a evaluar pertenecen exclusivamente al borde. Estos píxeles conforman el conjunto ℓ . El número de píxeles del conjunto ℓ es L , el perímetro de la región. El gradiente puede ser calculado utilizando el operador de gradiente de Gauss, en este caso para el ejemplo de la misma figura 2.9, $C = 35.47$.

- **Contraste.** El contraste de una región es concebido como una medida para la diferencia de color entre la región y su entorno. Región y entorno no tienen píxeles en común, y conforman una zona, que puede ser definida como un rectángulo: la zona entonces queda definida como la ventana

$$g[i, j] = x[i + i_r, j + j_r]$$

para $i = 1, \dots, 2h+1$ y $j = 1, \dots, 2w+1$

donde h y w representan la altura y el ancho de la región respectivamente. Los puntos centrales de estas zonas se definen como i_r y j_r , que se definen con el centro de gravedad de la región.

Entre más pequeña sea la diferencia de la variable de color en la región con respecto a su entorno, más pequeño será el contraste. Para visualizar el contraste se pueden representar la variable de color de una zona como una función 3D, donde el eje x y el eje y representan el eje i y el eje j de la imagen, y el eje z el valor de la variable de color que toma el píxel correspondiente, es decir $g[i, j]$. Nuevamente podemos ver esto en la figura 2.3 C, se reconoce en este ejemplo una región de alto contraste. El

contraste se define matemáticamente de diversas formas. Una definición comúnmente usada es utilizando características de textura.

Una forma de calcular el contraste se obtiene en tres pasos:

1. Extracción del color de los ejes principales de la zona. Se calculan dos funciones de color P1 y P2. La primera función P1 toma los valores de la variable de color en la dirección i y la segunda función P2 en la dirección j. Ambas funciones se centran en los centros de gravedad. En el ejemplo de la figura 2.3B el centro de gravedad está en el píxel (6,6), esto quiere decir que P1 y P2 son los valores del tono de gris de la zona de la columna 6 y de la fila 6 respectivamente, tal como se muestra en la figura 2.13A y 2.13B.
2. Aislamiento de la región. Para aislar la región de su entorno se trata de eliminar el fondo de la región, que se modela como una función lineal de primer orden, es decir una rampa. Se asume entonces que los valores extremos de P1 y P2 pertenecen las rampas R1 y R2, tal como se ilustra en la figura 2.13A y 2.13B. Las rampas serán substraídas de las funciones originales para conformar $Q1 = P1 - R1$ y $Q2 = P2 - R2$ que se fusionan en la nueva función Q como se muestra en la figura 2.13C.
3. Calculo del contraste. A partir de la nueva función Q se definen dos nuevos contrastes:

$$K_{\sigma} = \sigma_Q \quad \text{y} \quad K = \ln(Q_{max} - Q_{min})$$

donde:

σ_Q , Q_{max} y Q_{min} representan la desviación estándar, el máximo y el mínimo de Q respectivamente.

3. CLASIFICACIÓN

Aparte de haber sido segmentado en la escena y de habersele extraído cierto número de características propias, para reconocer un objeto es imprescindible haberlo catalogado previamente; es decir, haberlo definido, dado un nombre.

Esto significa que en un sistema de visión artificial existe un conocimiento almacenado (una memoria), que ha sido elaborado por un equipo de especialistas. Este conocimiento previo es muy variado y depende de la aplicación concreta. Puede ser muy sencillo, cual es el caso de la mayoría de las aplicaciones en robótica, en donde prácticamente sólo hay que distinguir unas pocas figuras simples y, en particular, su orientación en un plano. Y puede ser un conocimiento muy complejo, cual es el caso de la detección automática de tumores (aunque en diagnosis médica siempre existe un observador humano que es el responsable último).

Una vez adquirida la imagen, pre-procesada, segmentada la correspondiente escena, aislados y reconocidos los objetos, se entraría en una fase superior que podría calificarse de cognitiva o de procesamiento del conocimiento, más cercana a lo que ha venido a denominarse Inteligencia Artificial. Aunque, se insiste, todavía no se han resuelto importantes problemas en las fases previas, la segmentación en particular, para imágenes complejas. De hecho, sólo se han encontrado soluciones generales para imágenes bidimensionales sencillas. No hablemos de temas como la percepción de la textura de los objetos, de la visión tridimensional, de las escenas en movimiento, del fenómeno de la estereopsis (visión en profundidad), etc.

Vamos a ver ahora un ejemplo (figura 3.1), donde se observa, primeramente el vector de características v para el reconocimiento de caracteres.

Para llevar a cabo la clasificación se aplica el vector v proveniente del objeto individual al conjunto de funciones discriminantes previamente diseñadas para obtener la clasificación buscada. Obsérvese la figura 3.1, donde el Clasificador asigna el vector “ v ” a la clase que más se “parece”.

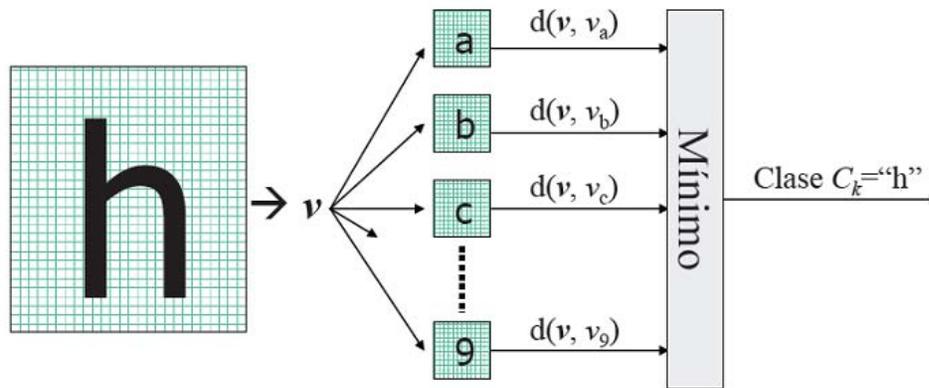


Figura 3.1 Clasificador

Para la clasificación se puede usar un conjunto de aprendizaje, del cual ya se conoce la clasificación de la información *a priori* y se usa para entrenar al sistema (aprendizaje supervisado). Como ya vimos, el aprendizaje puede ser también no supervisado, en el cual el sistema no posee un conjunto de aprendizaje para aprender a clasificar la información *a priori*, sino que se basa en cálculos estadísticos para clasificar los patrones.

▪ 3.1 Clasificación Supervisada

Si existe algún sistema de patrones cuyas clases individuales son conocidas, entonces se tiene un problema de **Reconocimiento de Patrones Supervisado**, también llamado **clasificación** o **aprendizaje supervisado**. Aunque también se denomina clasificación estructural de patrones porque está basado en las relaciones estructurales de las características.

El procedimiento básico a seguir en el diseño de un sistema para este tipo de problemas es el mostrado en la figura 1.14. Una porción del sistema de "patrones etiquetados" se extrae y se utiliza para derivar un **algoritmo de clasificación**. Estos patrones serán parte del **conjunto de entrenamiento**. Los patrones restantes se utilizan para probar el algoritmo de clasificación; estos patrones se refieren colectivamente como el **conjunto de prueba**. Puesto que las clases correctas de los patrones individuales en el sistema de prueba también se saben, se puede evaluar el funcionamiento del algoritmo.

3. Clasificación

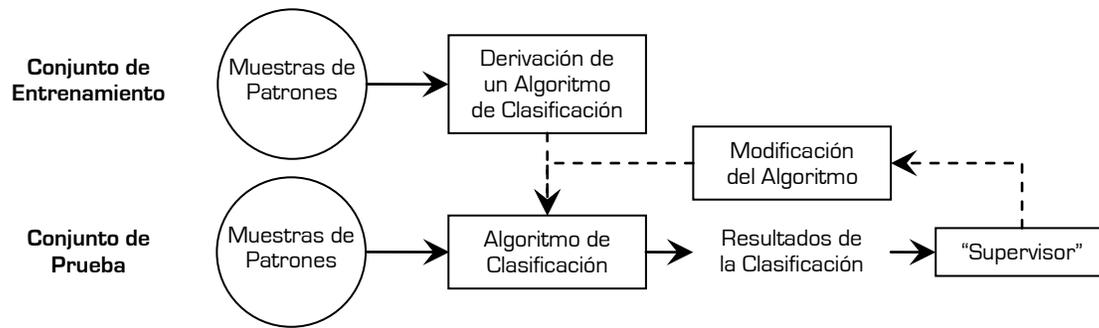


Figura 3.2 Procedimiento para el diseño de un Sistema de Reconocimiento de Patrones Supervisado

En los problemas del reconocimiento de patrones de este tipo, se piensa a menudo en los resultados como si estuvieran siendo evaluados por un supervisor (*teacher*) que dicta las modificaciones convenientes al algoritmo (de ahí que haya tomado ese nombre). Una vez que se haya alcanzado el nivel de funcionamiento deseado, el algoritmo puede ser usado sin patrones etiquetados inicialmente. Y aunque en ese caso el lazo de retroalimentación que involucra al supervisor ya no es de utilidad, generalmente se recomienda tener algún verificador de resultados del sistema en algún punto. [3]

- ✓ **Conjunto o Sistema de Entrenamiento.** Conjunto de patrones etiquetados cuya clase es conocida.
- ✓ **Conjunto o Sistema de Prueba (*Test*).** Conjunto de patrones etiquetados NO utilizados durante el entrenamiento. Sirven para evaluar el rendimiento del clasificador.
- ✓ **Algoritmo de Clasificación.** Es un conjunto de reglas de ajuste de los parámetros de las funciones discriminantes.
- ✓ **Supervisor (*Teacher*).** Hace la evaluación de los resultados de la clasificación y en base a ello modifica el algoritmo hasta obtener los óptimos resultados.
- ✓ **Generalización.** Capacidad para clasificar correctamente patrones no utilizados durante el entrenamiento.

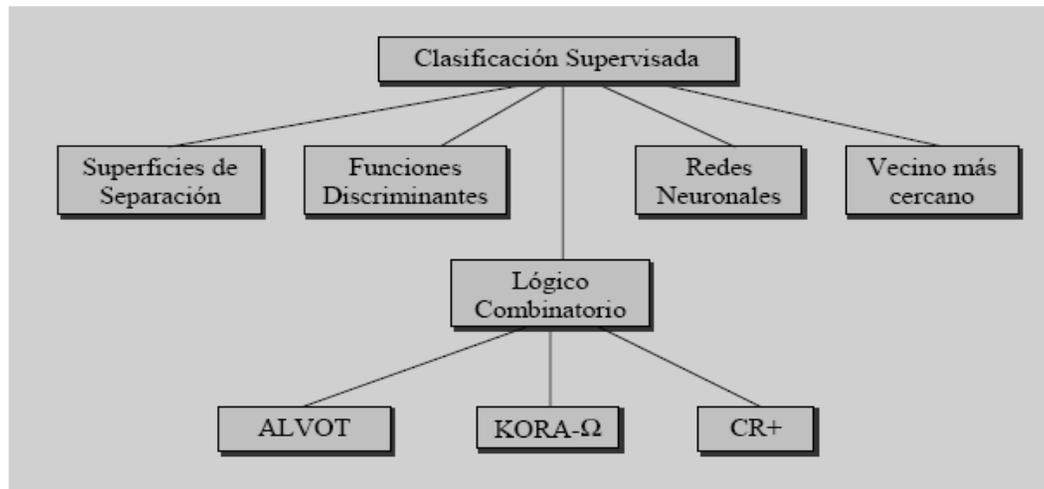


Figura 3.3 Esquema de Clasificación supervisada

- ✓ **Superficies de separación.** Se suponen los objetos en R^n y se busca un conjunto de superficies que separen a las clases.
- ✓ **Funciones discriminantes.** Si son dos clases, se busca obtener una función g tal que para un nuevo objeto P , si $g(P) \geq 0$ se asigna a la clase 1 y en otro caso a la 2. Si son múltiples clases se busca un conjunto de funciones y el nuevo objeto se ubica en la clase donde la función tome el mayor valor.
- ✓ **Vecino más cercano.** Un nuevo objeto se ubica en la clase donde esté el objeto de la muestra original que más se le parece.
- ✓ **Redes Neuronales (RN).** Se entrena una red de neuronas interconectadas, con los ejemplos para cada clase, con esto al presentarle un objeto nuevo en sus entradas la red dará como resultado la clase a la cual pertenece. El término Redes Neuronales engloba a un conjunto de técnicas que proporcionan soluciones flexibles (adaptables a cada problema) para cada una de estas estrategias.
- ✓ **Lógico Combinatorio.** Se definen ciertas condiciones que deben cumplir los agrupamientos y se buscan los grupos que las cumplan. Existen varios modelos: ALVOT, KORA- Ω , CR+ están entre los más conocidos (figura 1.16).

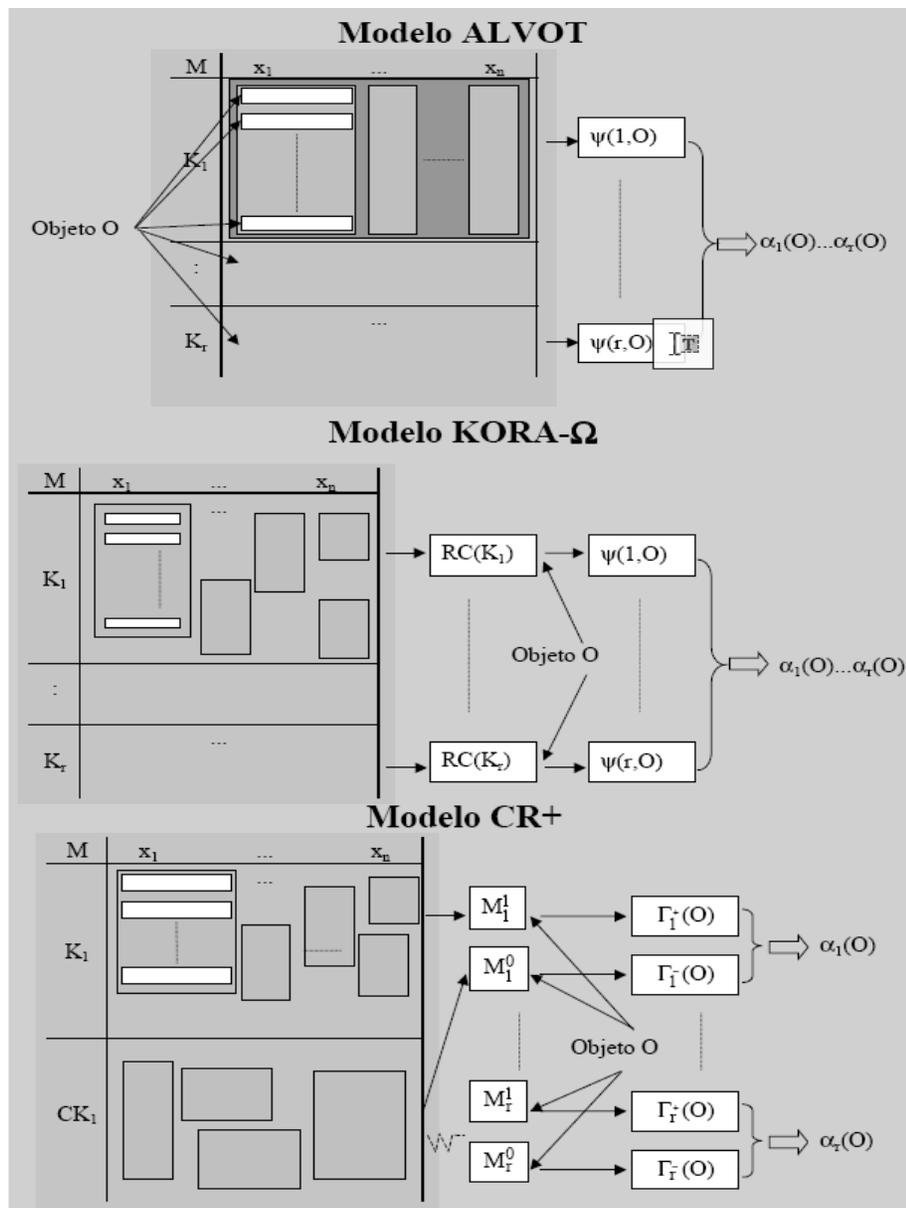


Figura 3.4 Enfoque lógico combinatorio. Modelos ALVOT, KORA- Ω y CR+

Algunas aplicaciones y ejemplos de enfoques serían los siguientes:

- Identificación de rostros (Redes Neuronales, Modelo ALVOT)
- Predicción de magnitudes máximas de terremotos (Modelo ALVOT)
- Búsqueda de petróleo (Modelo CR+)
- Pronóstico postoperatorio en niños con paladar hendido (Modelo ALVOT)
- Clasificación de atmósfera estelares (Vecino más cercano) [11]

3.2 Clasificación no Supervisada

El **Reconocimiento de Patrones no Supervisado** es también conocido como **Agrupación (*Clustering*) o aprendizaje no supervisado**. En este tipo de problemas, se intentan encontrar las clases de patrones con características similares donde algunas veces dichas características pueden ser indefinidas. Esta situación aparece comúnmente en ciertos campos de la medicina, biología y química donde, por ejemplo, se desea categorizar tipos de compuestos o células (y no están clarificados los universos de clases). También pueden encontrarse este tipo de problemas en estudios psicológicos, sociales o económicos, entre otros.

Para resolver este tipo de casos se aplican teorías estadísticas y de decisión, así que el reconocimiento estadístico de patrones está basado en las características estadísticas de los patrones, asumiendo que han sido generados por un sistema probabilístico (figura 1.17).

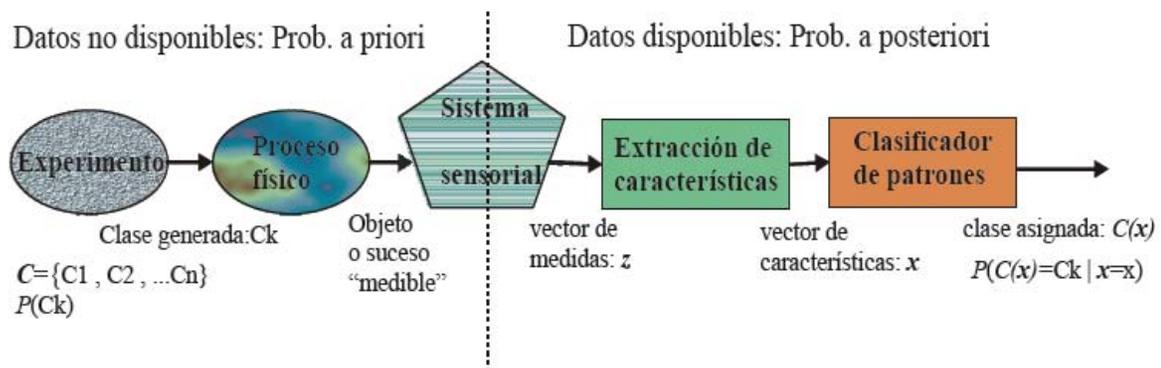


Figura 3.5 Representación de un clasificador estadístico

Sobra decir que el problema de reconocimiento de patrones no supervisado es mucho más complejo que el problema de reconocimiento supervisado. Sin embargo, se han desarrollado diversos algoritmos muy útiles pero su éxito depende del alto grado de capacidad para aprender estructuras de mediciones de patrones en espacios multidimensionales. [3]

Las técnicas de reconocimiento sin supervisión también se pueden emplear como alternativa en situaciones con conocimiento previo de las clases.



Figura 3.6 Esquema de Clasificación no supervisada

- ✓ **Jerárquico.** Puede ser divisiva o aglomerativa.
- ✓ **Reagrupamiento.** Se hace un primer agrupamiento y se va refinando iterativamente.
- ✓ **Lógico Combinatorio.** Se define ciertas condiciones que deben cumplir los agrupamientos y se buscan los grupos que las cumplan.
- ✓ **Simple Link y Complete Link.** Parten de grupos unitarios de objetos y van uniendo los grupos más parecidos en cada etapa, hasta cumplir alguna condición.
- ✓ **ISODATA.** Se van formando grupos que se ajustan iterativamente usando teoría de probabilidades, en alguna versiones se puede hacer la unión o división de algún grupo.
- ✓ **C-means.** Se define un grupo de semillas, se asocia cada objeto al grupo de la semilla más parecida, se toman los centroides de cada grupo como nuevas semillas y se itera hasta que se estabilice.
- ✓ **Criterios Lógico Combinatorios.** Los criterios que se imponen a los grupos son tales como ser conexos, completos maximales, compactos, etc.

4. RECONOCIMIENTO DE ROSTROS BASADOS EN IMÁGENES

Todos los seres humanos tenemos características morfológicas únicas que nos diferencian del resto de los seres humanos, como son la forma de la cara, la geometría de las partes de nuestro cuerpo como las manos, nuestros ojos y las huellas digitales.

El concepto de **biometría** proviene de las palabras *bio* (vida) y *metría* (medida), por lo tanto, con ello se infiere que todo equipo biométrico mide e identifica alguna característica propia de la persona. Biometría es el conjunto de características fisiológicas y de comportamiento que pueden ser utilizadas para verificar la identidad de un individuo, lo cual incluye huellas digitales, reconocimiento de iris, geometría de la mano, reconocimiento de rostros, entre otras técnicas.

La medición biométrica se ha venido estudiando desde tiempo atrás y es considerada en la actualidad como el método ideal de identificación humana, que se utiliza para verificar la identidad de una persona midiendo digitalmente determinados rasgos de alguna característica física y comparando esas medidas con aquéllas de la misma persona guardadas en archivo en una base de datos. En este caso, hablaremos de una de las características físicas utilizadas: los rasgos faciales. [17]

Como una de las aplicaciones más exitosas del análisis y comprensión de imágenes, el Reconocimiento de Rostros ha adquirido mucha atención durante los últimos años. Prueba de ello es la aparición de conferencias específicas sobre el tema como son AFGR^{*} y AVBA[†], y de técnicas sistemáticas de evaluación empírica para el Reconocimiento de Rostros (FRT, por sus siglas en inglés, *Face Recognition Technologies*), incluyendo los protocolos FERET[‡] y XM2VTS[§].

La fuerte demanda de sistemas “amigables” que garanticen la protección de nuestros bienes y nuestra vida privada sin perder nuestra identidad en un mar de números es evidente. En la actualidad, se necesita un NIP para obtener dinero en efectivo de un cajero automático, una contraseña para entrar a una computadora, una docena más para acceder a Internet, etcétera, etcétera. Aunque existen métodos biométricos de

^{*} Proceedings of the International Conferences on Automatic Face and Gesture Recognition, 1995, 1996, 1998, 2000.

[†] Proceedings of the International Conferences on Audio- and Video-Based Person Authentication, 1997, 1999.

[‡] PJ Phillips, H Moon, P Rauss, SA Rizvi. The FERET evaluation methodology for Face-Recognition algorithms. Proceedings of the Conference on Computer Vision and Pattern Recognition, 1997, pp 137–143. SA Rizvi, PJ Phillips, H Moon. The FERET verification testing protocol for Face Recognition algorithms. Proceedings of the International Conference on Automatic Face and Gesture Recognition, 1998.

[§] K Messer, J Matas, J Kittler, J Luettin, G Maire. XM2VTSDB: The extended M2VTS database. Proceedings of the International Conference on Audio- and Video-Based Person Authentication, 1999, pp 72–77.

identificación personal extremadamente fiables que ya se han mencionado, como son el análisis de huellas digitales y escaneos de retina o iris, estos métodos aún no han logrado mayor aceptación por parte de la población en general. Un sistema de identificación personal basado en el análisis de imágenes del rostro (ya sean frontales o de perfil) no es intrusivo y, por lo tanto, amigable al usuario. Lo que es más, la identidad personal puede ser comprobada a menudo sin la asistencia del individuo. Además, la necesidad de aplicar la **Tecnología de Reconocimiento de Rostros (FRT)** ha sido impulsada por los últimos avances en el procesamiento multimedia, junto con otros, tales como las tecnologías de Internet.

Las características de un sistema de identificación de rostros basado en imágenes son: el sistema no es invasivo, permite la identificación de personas en movimiento, tiene la posibilidad de camuflaje, no se necesita la autorización del individuo, al capturar no necesita de una fuente de luz auxiliar, etc. Sin embargo, es susceptible a problemas de iluminación y pose, como ya se describirá y además es vulnerable al reconocimiento de individuos que se han sometido a operaciones de cirugía plástica.

En resumen, existen grandes oportunidades y grandes retos para las FRT. El gran desafío que enfrenta es el buen funcionamiento bajo condiciones severas. Por ejemplo, un sistema de verificación personal puede ser que necesite procesar una imagen facial de baja calidad adquirida mediante una *webcam** de bajo costo y transferida a través de Internet, o que la captura de la imagen ocurra en un ambiente no controlado con mala iluminación, etc.

▪ 4.1 Tecnología de Reconocimiento de Rostros

Un planteamiento general sobre el problema de Reconocimiento de Rostros puede formularse como sigue. Dado un video o una imagen de una escena, identificar o verificar una o más personas en la misma utilizando una base de datos que almacene información de rostros. Se tendrá disponible información colateral como raza, edad, género, expresión facial y lengua, esto puede ayudar a reducir la búsqueda (mejorando el reconocimiento). La solución del problema consiste en la segmentación (detección de los rostros) en la escena originalmente compleja, extracción de características de la región facial y el propio reconocimiento o verificación.

En los **problemas de identificación**, la entrada del sistema es un rostro desconocido, y el sistema se encarga de regresar la identidad del individuo a partir de

* Una *webcam* o cámara web es una pequeña cámara digital conectada a una computadora, la cual puede capturar imágenes y transmitir las a través de Internet en directo, ya sea a una página web o a otra u otras computadoras de forma privada. Las webcams necesitan una computadora para transmitir las imágenes.

una base de datos con la información de personas conocidas; mientras que en los **problemas de verificación**, el sistema necesita confirmar o rechazar la identidad declarada del rostro de entrada.

Existe un rango muy variado de aplicaciones de la FRT que van de fotografías estáticas en condiciones controladas a imágenes de vídeo sin control, lo que plantea una amplia gama de técnicas diferentes y desafíos que requieren al igual diversas técnicas de procesamiento de imágenes, análisis, comprensión y Reconocimiento de Patrones.

En términos generales, se pueden clasificar los problemas y las técnicas en dos grupos: **estáticos** y **dinámicos**. Incluso dentro de cada grupo, existen diferencias significativas, en función de la aplicación que se trate. Éstas diferencias son en términos de calidad de imagen, cantidad de elementos de fondo (lo que hace un reto para los algoritmos de segmentación), disponibilidad de un criterio de concordancia bien definido y de la propia naturaleza y tipo de las imágenes de entrada, entre otras.

Ahora bien, el Reconocimiento Automático de Rostros consiste en diversas subtarefas en forma secuencial: detección del rostro, segmentación y normalización y reconocimiento del rostro o verificación. Muchos métodos de Reconocimiento de Rostros han sido propuestos^{**}, básicamente, se pueden dividir en:

- 1) Sistemas basados en la concordancia de modelos holísticos (método de las *eigenfaces*), que considera las propiedades globales del patrón.
- 2) Sistemas analíticos [*eigenfeatures*], basados en las características geométricas locales de la cara.
- 3) Esquemas híbridos.

Aunque todos estos esquemas se han aplicado con éxito en las tareas de Reconocimiento de Rostros, tienen ciertas ventajas y desventajas. Un sistema apropiado debe ser elegido basado en las necesidades específicas de una determinada tarea. Por ejemplo, el sistema basado en el EBG (Elastic Bunch Graph Matching)^{**} tiene muy buen desempeño en general, sin embargo, se requieren imágenes grandes (por ejemplo, 128 x 128). Esto limita su posible aplicación a los videos de cámaras de seguridad que se utilizan para la vigilancia, en donde el tamaño de la imagen del área facial es muy pequeña. Por otra parte, el subespacio del sistema LDA^{**} funciona bien con imágenes

^{**} W Zhao, R Chellappa, A Rosenfeld, P J Phillips. Face Recognition: A literature survey. Technical Report CS-TR-4167, Center For Automation Research, University of Maryland, 2000.

^{**} K Okada, J Steffans, T Maurer, H Hong, E Elagin, H Neven, C v d Malsburg. The Bochum/USC Face Recognition System and How it fared in the FERET phase III test. In: H Wechsler, P J Phillips, V Bruce, F F Souli, T S Huang, eds. Face Recognition: From Theory to Applications. New York: Springer-Verlag, 1998, pp 186-205.

^{**} W Zhao, R Chellappa, P J Phillips. Subspace Linear Discriminant Analysis for Face Recognition. Technical Report CAR-TR-914, Center for Automation Research, University of Maryland, 1999.

grandes y pequeñas (por ejemplo, 96 x 84 ó 24 x 21). También cuenta con el mejor rendimiento para tareas de verificación de acuerdo a la más reciente prueba FERET.

Todos los productos de Reconocimiento de Rostros almacenan múltiples imágenes de cada individuo, y dependen de un set de reglas para determinar la identidad tomando en cuenta los datos relevantes. Algunos productos usan Inteligencia Artificial y tecnología de Redes Neuronales, en estos la efectividad aumenta con la experiencia y el auto-aprendizaje del propio sistema. Este proceso de aprendizaje permite al sistema disminuir el rango de tipos faciales que debe almacenar en su base de datos a los efectos de la comparación.

Las Redes Neuronales más usadas para el reconocimiento son las redes de “contra-propagación” que básicamente utilizan el método de convertir una imagen en un vector de números y pasarla a la capa de entrada de la red.

Los sistemas de Reconocimiento de Rostros no solo trabajan con imágenes del rostro, algunos incrementan la seguridad almacenando vistas frontales y laterales, esto produce un mapa en 3D, que elimina la posible falla de seguridad utilizando fotos de legítimos usuarios.

Otra aproximación usa imágenes térmicas (denominadas Termogramas Faciales). En esos sistemas se emplean cámaras infrarrojas para capturar los patrones que conforman las arterias y venas bajo la piel. La ventaja de estos sistemas es que son menos susceptibles a cambios en la superficie de la piel y en la posición de la cara, y además pueden operar en la oscuridad.

▪ 4.2 *Eigenfaces* y *Eigenfeatures*

La técnica básica de *eigenfaces* (rostros eigen*) se encarga de encontrar bases de espacios vectoriales compuestas por pocas imágenes que expliquen de forma apropiada el espacio de las imágenes faciales. Como se muestra en la figura 4.1, las *eigenfaces* se obtuvieron de un conjunto de entrenamiento de 600 imágenes.

Un sistema basado en *eigenfaces* considera cada imagen facial como conjunto bidimensional de áreas claras y oscuras dispuestas en un modelo determinado. Éstas áreas son las consideradas *eigenfaces*. El algoritmo de reconocimiento almacena cada imagen como una combinación de *eigenfaces* y después compara las características del *eigenface* del rostro a identificar con las existentes en la base de datos.

* *Eigen* es un término alemán que se refiere a la matemática recursiva usada para analizar características faciales únicas.

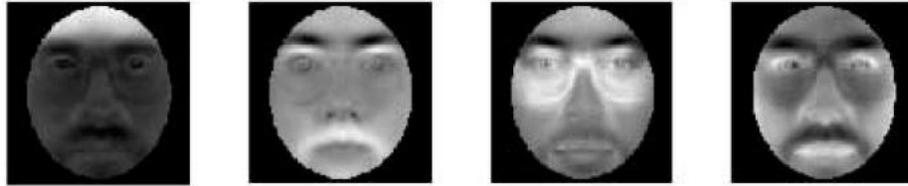


Figura 4.1 *Eigenfaces* de un conjunto de entrenamiento

Por otro lado, un sistema basado en *eigenfeatures* (características eigen) se centra en características faciales específicas tales como la nariz, ojos, boca, cejas y estructuras óseas, y las distancias relativas entre ellas. El sistema analiza la imagen del rostro captado y extrae estos *eigenfeatures* determinados, entonces compara éstos con otros almacenados en la base de datos.

En general, estos sistemas son más exactos en identificar rostros a pesar de variaciones substanciales tales como barbas y anteojos. Sin embargo, los sistemas basados en *eigenfeatures* (métrica facial, técnica analítica) trabajan típicamente en conjunto con los sistemas basados en *eigenfaces* (técnica holística) para producir la identificación más exacta posible.

A partir de la idea original de las *eigenfaces*, han surgido otros enfoques para reducir la dimensión, como el de *fisherfaces* o el Análisis de Componentes Principales (PCA).

▪ 4.3 Generación de *Eigenfaces*

Para generar un conjunto de *eigenfaces*, se deben tomar un gran número de imágenes digitalizadas de rostros humanos tomadas bajo las mismas condiciones de iluminación, dichas imágenes tienen que ser procesadas, es decir, se normalizan en la línea de los ojos y de la boca y luego se reconstruyen a una misma resolución. Las *eigenfaces* pueden extraerse de datos de una imagen a través de una herramienta matemática llamada Análisis de los Componentes Principales (PCA).

A continuación los pasos involucrados en la conversión de una imagen facial en una *eigenface*:

1. Preparar un conjunto de entrenamiento. Las imágenes de los rostros que constituyen el conjunto de entrenamiento T deben estar listos para ser procesados.

4. Reconocimiento de Rostros basados en Imágenes

2. Obtener la media. El promedio de la matriz A tiene que ser calculado y restado de la matriz original en T. El resultado se almacenará en la variable S.
3. Calcular la matriz de covarianza.
4. Calcular los valores y vectores [eigen] de esa matriz de covarianza.
5. Elegir los componentes principales.

Antes de llegar al paso 5 existirá un gran número de *eigenfaces*, pero pocos son necesarios. Se deben seleccionar los que tienen los valores más altos. Por ejemplo, si estamos trabajando con una imagen de 100x100, entonces este sistema creará 10000 vectores *eigen*, dado que la mayoría de los individuos pueden ser identificados usando una base de datos con un tamaño entre 100 y 150, la mayoría de los vectores pueden ser descartados y sólo quedará el más importante.

Los *eigenfaces* creados aparecerán como zonas oscuras y claras agrupadas en un patrón en específico. Este patrón determina como las diferentes características de un rostro se evalúan y califican. Por ejemplo, existirá un patrón para evaluar la simetría, alguno que determine si existe algún tipo de pelo facial o donde empieza el cuero cabelludo o para evaluar el tamaño de la nariz o de la boca. Otros *eigenfaces* tienen patrones que son menos fáciles de identificar, y la imagen de la *eigenface* puede parecerse muy poco a un rostro.

La técnica utilizada en la creación de *eigenfaces* usada para el reconocimiento también son usadas fuera del reconocimiento de rostros. Esta técnica también es usada para el análisis de escritura, reconocimiento de voz, etc. Por lo tanto, algunos no utilizan el término *eigenface* (rostro eigen) y prefieren llamarlo *eigenimage* (imagen eigen).

Básicamente, los *eigenfaces* son un conjunto de "componentes faciales estandarizados", derivados del análisis estadístico de muchas imágenes de rostros. Cualquier rostro humano puede considerarse una combinación de estos rostros estándar. Por ejemplo, un rostro podría estar compuesto del 10% del *eigenface* 1, 55% del *eigenface* 2, y hasta un -3% del *eigenface* 3. Los valores de correspondencia entre un rostro con un *eigenface* pueden ser cualquiera entre 100% y -100%, y entre más algo es el valor, más parecido existe entre el rostro y el *eigenface*.

Sorprendentemente, no se necesitan muchos *eigenfaces* juntos para la evaluación de un rostro en la mayoría de los casos. Así mismo, dado que el rostro de la persona no está registrado como una fotografía digital, sino como una lista de valores (sólo se utiliza un valor por cada *eigenface* en la base de datos – matriz calculada), se utiliza mucho menos espacio en disco para el almacenamiento de la información. [19]

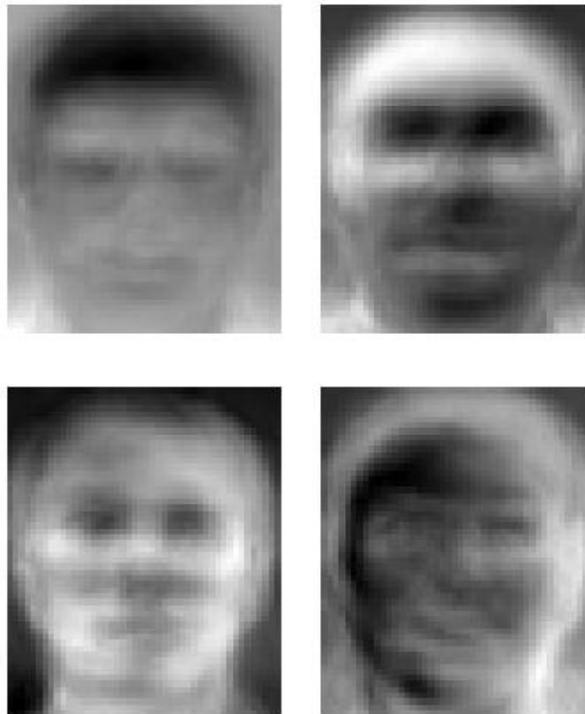


Figura 4.2 *Eigenfaces*

Conocer el contenido de las imágenes faciales puede ser un gran apoyo a la tarea de reconocimiento. Si estructuras complejas, con el grafo que se muestra en la figura 4.3, se pudieran posicionar automáticamente, los sistemas así diseñados serían mucho más eficientes.

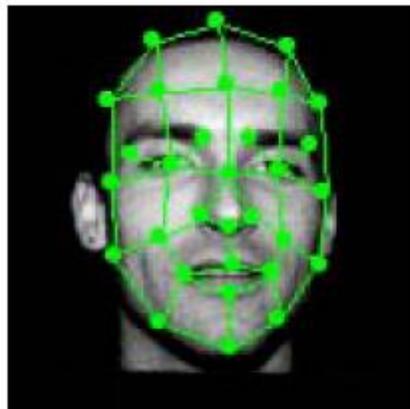


Figura 4.3 Puntos de referencia en un rostro

La localización de algunos puntos de referencia proporciona conocimiento sobre la forma de las caras. Pero, tanto para la localización automática de estos puntos de referencia como para el posterior reconocimiento se necesita incluir propiedades de

textura o de variación de los niveles de gris en torno a estos nodos. Una herramienta habitual para este fin es el análisis de Gabor. Un ejemplo de imágenes filtradas con filtros de Gabor se muestra en la siguiente figura 4.4.

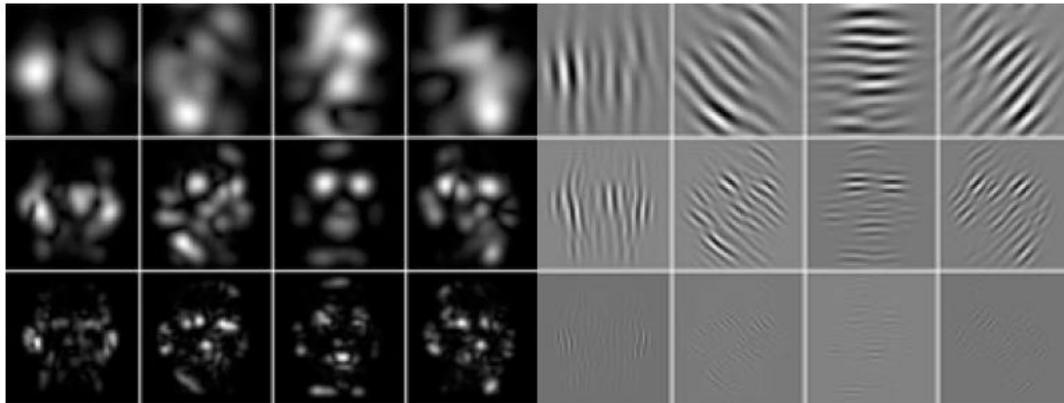


Figura 4.4 Imágenes filtradas con filtros de Gabor

La mayor dificultad para los sistemas de reconocimiento facial es que la cara de la persona cambia a través del tiempo. El sistema debe tener en cuenta esos cambios para hacer una correcta identificación y así poder ir actualizando su base de datos.

▪ 4.4 Reconocimiento de Rostros en 3D

Existe una tendencia en software para el Reconocimiento de Rostros basada en un modelo de imágenes en 3D, que pretende proporcionar más precisión. Capturando en tiempo real una imagen en 3D de la superficie facial de una persona, el sistema utiliza rasgos distintivos del rostro (esto es, rasgos donde los tejidos rígidos y huesos son más evidentes, como las curvas de los globos oculares, la nariz, la barbilla, entre otros) para identificar al individuo, ya que estas áreas son únicas y no cambian con el tiempo.

Usando la profundidad y un eje de medición que no se vea afectado por la luz, el Reconocimiento de Rostros en Imágenes en 3D se puede utilizar incluso en la oscuridad y es capaz de reconocer un rostro con diferentes ángulos de rotación, de hecho tiene el potencial de reconocer un rostro de perfil (que está girado 90 grados).

Ahora bien, el sistema pasa por una serie de pasos para verificar la identidad de un individuo:

- ✓ **Detección (*Detection*).** Esto habla de la adquisición de una imagen, puede lograrse mediante un escaneo digital de una fotografía existente (2D) o mediante el uso de una imagen de video en vivo (3D). Las cámaras utilizadas para ello varían de acuerdo a la aplicación que trate y también varían, obviamente, en cuanto a su calidad y costo.
- ✓ **Alineación (*Alignment*).** Inmediatamente después de que un rostro es detectado en una imagen, el sistema determina la posición de la cabeza, así como el tamaño y la pose. Como se dijo anteriormente, el individuo puede ser reconocido con un ángulo de rotación de la cabeza de hasta 90 grados; mientras que al trabajar con imágenes en 2D, la cabeza sólo puede estar rotada hasta 35 grados frente a la cámara.
- ✓ **Medición (*Measurement*).** A continuación, el sistema mide las curvas del rostro en una escala sub-milimétrica (o incluso menor) y crea una plantilla.
- ✓ **Representación (*Representation*).** El sistema traduce la plantilla en un código único. Este código proporciona a cada plantilla un conjunto de números que representan las características en el rostro del individuo. Entonces, cada individuo, estará representado por un conjunto de números único.
- ✓ **Comparación (*Matching*).** Si la imagen adquirida es en 3D y la base de datos contiene imágenes en 3D, entonces el trabajo de comparación se llevará a cabo sin necesidad de hacer algún cambio a la imagen. Sin embargo, actualmente nos enfrentamos a un reto en las bases de datos que se encuentran aún en 2D. Las imágenes en 3D ofrecen un individuo vivo y variable en cuanto a posición en comparación a una imagen plana y estable. Las nuevas tecnologías hacen frente a este desafío. Cuando una imagen en 3D es tomada, diferentes puntos (usualmente tres) son identificados. Por ejemplo, el exterior del ojo, el interior del ojo y la punta de la nariz se separan y miden. Una vez tomadas estas medidas, un algoritmo se aplicará a la imagen para convertirla a una imagen en 2D. Después de la conversión, el software entonces compara la imagen con las imágenes en 2D existentes en la base de datos para encontrar la que mejor se parezca.
- ✓ **Verificación o Identificación (*Verification or Identification*).** En la verificación, una imagen adquirida es comparada y coincide con solo una imagen en la base de datos (1:1). Es decir, una imagen tomada de un individuo puede ser que coincida con una imagen de una base de datos específica y corrobora que el individuo es quien dice ser. Si el objetivo es la identificación, entonces la imagen se compara con todas las imágenes en la base de datos y se obtiene una puntuación para cada parecido potencial (1:n), la mejor puntuación es la identificación del individuo. En este caso, hablamos de tomar una imagen y compararla en una base de datos para identificar quién es el individuo. [18]

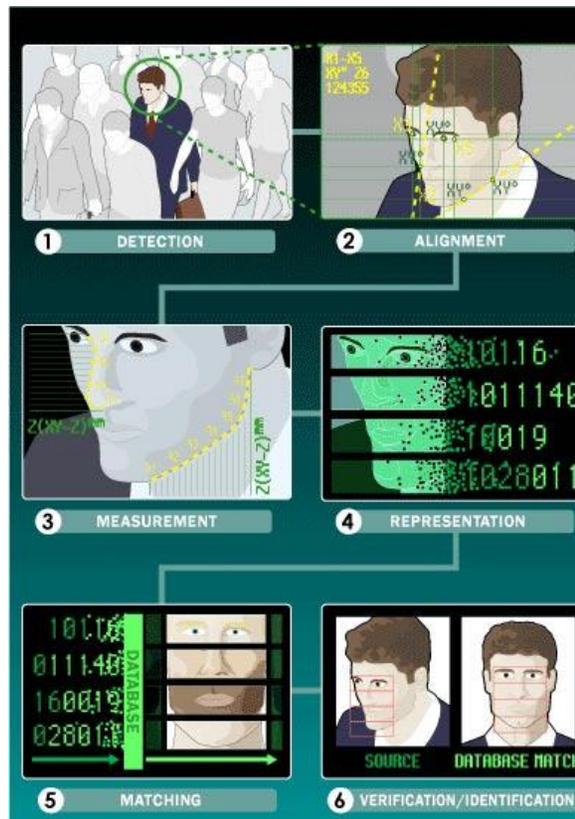


Figura 4.5 Reconocimiento de Rostros en Imágenes 3D

4.5 Problemas en el Reconocimiento de Rostros

A pesar del éxito de muchos sistemas, en base a la prueba FERET, muchas cuestiones siguen sin resolverse. Entre ellas, las dos siguientes son destacadas por la mayoría de los sistemas:

- (1) El problema de iluminación.
- (2) El problema de pose.

El problema de iluminación se ilustra en la figura 4.3, donde el mismo rostro parece diferente debido al cambio en la iluminación. Para entender mejor esto, los cambios inducidos por la iluminación podrían ser mayores que las diferencias entre un individuo y otro, provocando que los sistemas basados en la comparación de imágenes clasifiquen erróneamente la identidad de la imagen de entrada. Esto se ha observado experimentalmente^{ss} con un conjunto de datos de 25 personas.

^{ss} Y Adini, Y Moses, S Ullman. Face Recognition: The problem of compensating for changes in illumination direction.

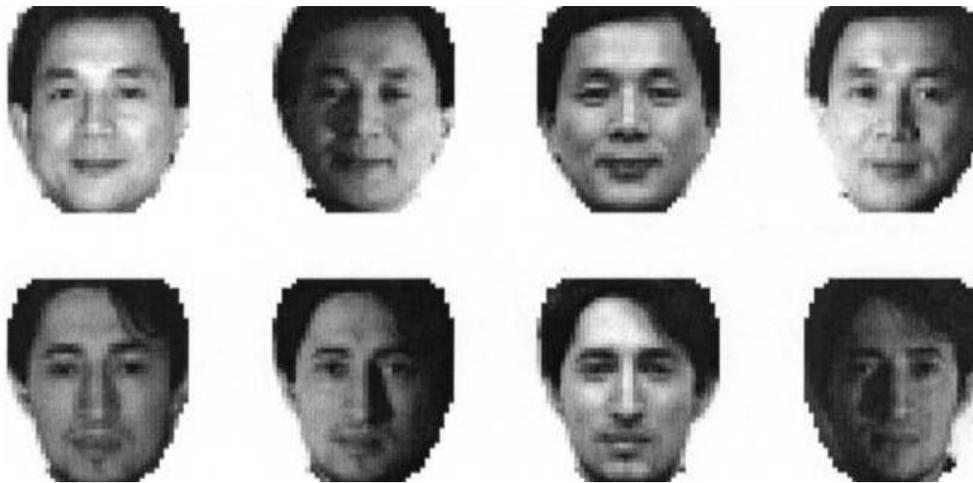


Figura 4.6 El problema de iluminación

El problema de pose se ilustra en la figura 4.7, donde el mismo rostro parece distinto debido a los cambios en las condiciones del ángulo de visión. Además, cuando también existe variación en la iluminación, la tarea de reconocimiento se hace aún más difícil.

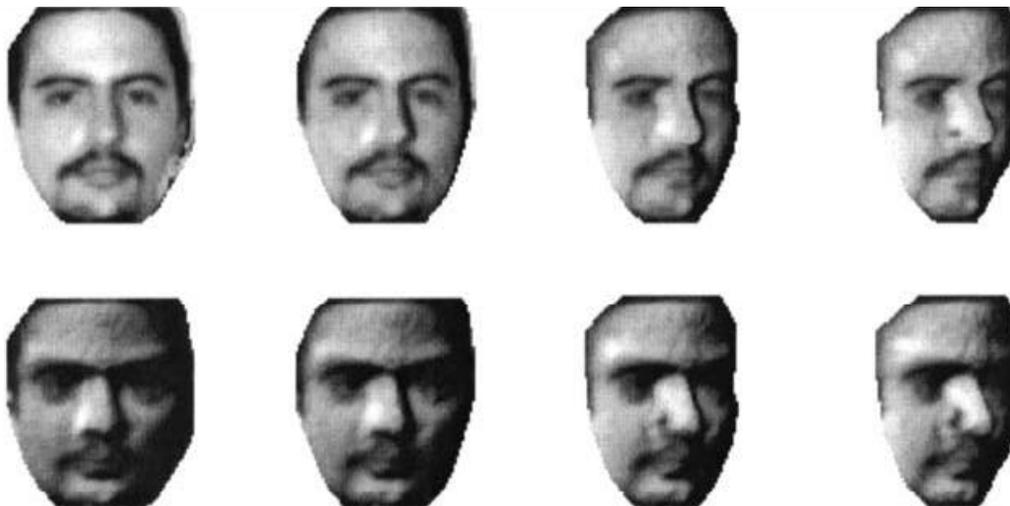


Figura 4.7 El problema de pose (e iluminación). Arriba, sólo el problema de pose. Abajo el problema de pose e iluminación combinados.

Las dificultades debido a las variaciones en la iluminación y pose se han documentado en muchas evaluaciones de los sistemas de Reconocimiento de Rostros. Por desgracia, esto ocurre cuando se adquiere las imágenes en entornos no

controlados, por ejemplo, los videos obtenidos mediante cámaras de seguridad. Las imágenes/patrones en dos dimensiones (2D) de los objetos faciales que están en tres dimensiones (3D) pueden cambiar de forma drástica debido a las variaciones de iluminación y visualización.

Por lo tanto, los problemas de iluminación y pose plantean en la actualidad importantes obstáculos para muchas aplicaciones, lo importante es revisar las ventajas y desventajas de los diferentes métodos a fin de adoptar un enfoque adecuado que pueda ser aplicado a una tarea específica.

Para superar estos problemas, a menudo se utiliza un modelo 3D genérico que básicamente sintetiza el llamado **prototipo*** de una imagen dada adquirida bajo diferentes condiciones de iluminación y visualización. Esta mejora permite a los sistemas existentes manejar tanto los problemas específicos de iluminación como los de pose asumiendo que sólo una imagen por objeto facial estará disponible.

▪ **4.6 Enfoques para solucionar los problemas**

Para el Reconocimiento de Rostros, muchos buenos enfoques que se han propuesto utilizando la idea de que todas las caras pertenecen a una clase, éstos enfoques se pueden dividir en cuatro tipos: (1) métodos heurísticos, (2) métodos de comparación de imágenes, para los cuales se aplican diversas representaciones de la imagen y medidas de la distancia, (3) métodos basados en clases, para los cuales se dispone de varias imágenes de un rostro con poses arregladas pero con diferentes condiciones de luz, y (4) enfoques basados en modelos, en los cuales se emplean modelos en 3D.

Las investigaciones realizadas han propuesto varios métodos para manejar el problema de la rotación. Básicamente, se pueden dividir en tres clases^{***}: (1) métodos basados en múltiples vistas, cuando existen múltiples vistas por cada persona, (2) métodos híbridos, cuando la información de múltiples imágenes están disponibles durante el entrenamiento, pero sólo una imagen por persona está disponible en la base de datos al momento del reconocimiento y (3) métodos basados en imágenes/formas simples, cuando no se realiza el entrenamiento. Hasta ahora, el segundo tipo de enfoque es el más conocido y el tercero no parece haber recibido mucha atención.

*El prototipo de una imagen se define como la vista frontal de un objeto bajo iluminación frontal.

*** W Zhao, R Chellappa. SFS based view synthesis for robust Face Recognition. Proceedings of the Conference on Automatic Face and Gesture Recognition, 2000.

▪ 4.7 Aplicaciones

En cuanto a los usos actuales y futuros de los sistemas de Reconocimiento de Rostros basados en Imágenes, en el pasado, los principales usuarios de éstas técnicas, eran agencias gubernamentales (sobre todo en E.U.A.) que usaban los sistemas para capturar rostros al azar en las multitudes, algunas de estas agencias también lo usaban para sistemas de seguridad y para eliminar el fraude electoral. De hecho, recientemente el gobierno estadounidense comenzó un programa llamado *"US-VISIT" (United States Visitor and Immigrant Status Indicator Technology)*- (Tecnología Indicadora del Estatus de los Visitantes e Inmigrantes a Estados Unidos), que tiene el objetivo de regular el ingreso al país por parte de extranjeros. Cuando un extranjero recibe su Visa, proporciona sus huellas digitales y es fotografiado, esos elementos son comparados en una base de datos de criminales conocidos y sospechosos de terrorismo. Cuando el viajero llega a los E.U.A., en el puerto de entrada, esas mismas huellas y fotografías son usadas para verificar que la persona que recibió la Visa es la misma persona que está entrando al país.

Sin embargo, ahora hay muchas más situaciones en las que el Reconocimiento de Rostros basados en Imágenes se está volviendo más popular, debido a la reducción de costos. Los sistemas ahora son compatibles con cámaras y computadoras que ya se usan en bancos y aeropuertos. Otras aplicaciones potenciales incluyen los famosos ATM (cajeros automáticos). El software es capaz de reconocer rápidamente el rostro del cliente. Aunque previamente se requiere la captura de varias imágenes del rostro del individuo, con esto ya no se requiere el uso de un NIP para verificar la identidad y también podría prevenir situaciones fraudulentas.

CONCLUSIONES

El Reconocimiento de Patrones ofrece una amplia gama de aplicaciones, que involucran reconocer y verificar objetos, fenómenos, individuos con varias características, así como relacionarlos y agruparlos de acuerdo a determinadas reglas, por ejemplo, la agrupación de palabras para obtener frases y oraciones [utilizado por los traductores].

Como puede entenderse con este documento, este tema es muy extenso y requiere dirigirse de acuerdo a la aplicación que se trate. Sin embargo, para un entendimiento óptimo es sugerido familiarizarse en áreas de matemáticas avanzadas como álgebra lineal, probabilidad, teoría de decisiones y estadística, así como con algoritmos, programación, tratamiento de señales, procedimientos para el análisis de imágenes, etc.

El Reconocimiento de Patrones, si bien no es algo trivial es algo con lo que nos enfrentamos día a día, vivimos todos los días reconociendo las cosas a nuestro alrededor, realizando asociaciones, adquiriendo nueva información. Con la asistencia de las herramientas tecnológicas a nuestro alcance podemos hacer maravillas. No sólo se reducen los trabajos tediosos y repetitivos, sino nos proporcionan fiabilidad en algunos casos.

Algunas veces podría considerarse un dilema filosófico el pensar confiar nuestra seguridad a una máquina, confiar más que en una persona; pero la respuesta es simple, un algoritmo bien diseñado implementado en un sistema de reconocimiento para el problema de la identificación y verificación de identidad de las personas es más fiable y rápido que lo que podría hacerlo una persona.

Sólo basta decir, aunque está por demás, que la tecnología avanza exponencialmente y que a este paso, nos enfrentaremos con nuevos problemas en cuanto a reconocimiento de patrones que podrían provocarnos sorpresas interesantes.

LISTADO DE FIGURAS

FIGURA 1.1 DIAGRAMA DE BLOQUES DE UN SISTEMA DE RECONOCIMIENTO AUTOMÁTICO	8
FIGURA 1.2 DIAGRAMA DE CONCEPTOS BÁSICOS	8
FIGURA 1.3 TIPOS DE IRIS	9
FIGURA 1.4 EJEMPLO DE PATRONES VECTORIALES. CLASIFICACIÓN DE TIPOS DE IRIS	10
FIGURA 1.5 HUELLA DACTILAR	10
FIGURA 1.6 EJEMPLO DE PATRONES ESTRUCTURADOS. RECONOCIMIENTO DE HUELLAS DACTILARES	11
FIGURA 1.7 IMAGEN DE SATÉLITE DE UNA ZONA DE LA CIUDAD DE WASHINGTON	11
FIGURA 1.8 EJEMPLO DE PATRONES EN ÁRBOL. RELACIONES A PARTIR DE UNA IMAGEN DE SATÉLITE	12
FIGURA 1.9 EJEMPLO DE VECTOR DE CARACTERÍSTICAS	15
FIGURA 1.10 TIPOS DE SELECCIÓN DE VARIABLES	16
FIGURA 1.11 EJEMPLO DE FUNCIÓN DISCRIMINANTE. REPRESENTACIÓN EN UN PLANO DE DOS CLASES DE DEPORTISTAS	17
FIGURA 1.12 DECISOR	18
FIGURA 1.13 EL ENFOQUE BÁSICO DEL RECONOCIMIENTO DE PATRONES	20
FIGURA 1.14 ENFOQUES Y PROBLEMAS DEL RECONOCIMIENTO DE PATRONES	20
FIGURA 1.15 ESQUEMA GENERAL DEL ANÁLISIS DE IMÁGENES	22
FIGURA 1.16 IMAGEN Y MATRIZ NUMÉRICA RESULTANTE DEL PROCESO DE DIGITALIZACIÓN	23
FIGURA 1.17 IMAGEN CON DIFERENTES RESOLUCIONES: A) 256 X 256, B) 128 X 128, C) 64 X 64, D) 32 X 32	24
FIGURA 1.18 MODELOS ADITIVO Y SUSTRACTIVO DE COLOR	25
FIGURA 1.19 IMAGEN EN COLOR Y SU DESCOMPOSICIÓN EN LAS TRES BANDAS ESPECTRALES: ROJO, VERDE Y AZUL (RGB)	27
FIGURA 1.20 ESQUEMA DE UN CCD DE SUPERFICIE (ARRAY) Y DE SU FUNCIONAMIENTO.	27
FIGURA 1.21 EJEMPLO DE OPERACIÓN ARITMÉTICA. SUMA DE IMÁGENES, $A+B=C$	32
FIGURA 1.22 IMAGEN POCO CONTRASTADA E IMAGEN RESULTANTE DE LA EXPANSIÓN DE SU HISTOGRAMA	33
FIGURA 1.23 MATRIZ DE 3X3 PÍXELES DE UNA IMAGEN Y LOS VALORES DE MEDIANA, MÍNIMO Y MÁXIMO	34
FIGURA 1.24 5 ITERACIONES DE UN FILTRO DE MÍNIMO CON DIFERENTES ELEMENTOS ESTRUCTURALES	35
FIGURA 1.25 RESULTADO DE UN FILTRO DE PROMEDIO	35
FIGURA 1.26 DEFINICIÓN DE UN FILTRO DE ENFATIZADO	36
FIGURA 1.27 DEFINICIÓN DE UN FILTRO DE EXTRACCIÓN DE CONTORNOS	36
FIGURA 1.28 DEFINICIÓN DE DOS TIPOS DE FILTRO DE TOPHAT	37
FIGURA 1.29 AUMENTO DE LA GANANCIA DE UNA IMAGEN MANIPULANDO LA CURVA DE AJUSTE	37
FIGURA 1.30. IMAGEN VISUALIZADA CON UNA PALETA DE NIVELES DE GRIS Y EN FALSO COLOR UTILIZANDO UNA PALETA DE COLORES	38
FIGURA 2.1 EJEMPLO DEL VECTOR DE OBSERVACIÓN	39
FIGURA 2.2 IMAGEN DIGITAL DE UN OBJETO OSCURO SOBRE UN FONDO CLARO Y SU HISTOGRAMA	42
FIGURA 2.3 A) IMAGEN DE NIVELES DE GRIS Y SU HISTOGRAMA, B). IMAGEN SEGMENTADA	46
FIGURA 2.4 SEGMENTACIÓN POR WATERSHEED. A) IMAGEN ORIGINAL. B) MAPA DE DISTANCIAS. C) REPRESENTACIÓN TRIDIMENSIONAL. D) RESULTADO DE LA SEGMENTACIÓN	47
FIGURA 2.5. OPERACIONES DE APERTURA (OPENING) Y CIERRE (CLOSING).	48

FIGURA 2.6 OPERACIONES LÓGICAS: AND, OR Y XOR	49
FIGURA 2.7 IMAGEN BINARIA DE UNA RED DE FISURAS E IMAGEN RESULTANTE DE UN PROCESO DE THINNING	49
FIGURA 2.8 IMAGEN EN COLOR DE UN TEJIDO E IMAGEN SEGMENTADA CON LOS ELEMENTOS EN COLOR VIOLETA. HISTOGRAMAS DE LAS INTENSIDADES	50
FIGURA 2.9 EJEMPLO DE UNA REGIÓN. A) IMAGEN, B) REGIÓN SEGMENTADA, C) REPRESENTACIÓN 3D DE LOS VALORES DE GRIS EN LA REGIÓN Y SU ENTORNO	51
FIGURA 2.10 A) DISTINTOS VALORES DE NÚMERO OBTENIDOS SEGÚN CONTEMOS TODOS LOS OBJETOS DE LA IMAGEN. B) ELIMINANDO TODOS LOS OBJETOS QUE TOCAN LOS BORDES. C) ELIMINANDO LOS QUE TOCAN DOS DE LOS BORDES	53
FIGURA 2.11 CÁLCULO DE LOS DIÁMETROS MÁXIMO Y MÍNIMO DE DOS OBJETOS DE FORMAS DISTINTAS	57
FIGURA 2.12 INTENSIDAD INCIDENTE (IO), INTENSIDAD TRANSMITIDA (I) Y SUPERFICIE DE UN OBJETO (S)	58
FIGURA 3.1 CLASIFICADOR	63
FIGURA 3.2 PROCEDIMIENTO PARA EL DISEÑO DE UN SISTEMA DE RECONOCIMIENTO DE PATRONES SUPERVISADO	64
FIGURA 3.3 ESQUEMA DE CLASIFICACIÓN SUPERVISADA	65
FIGURA 3.4 ENFOQUE LÓGICO COMBINATORIO. MODELOS ALVOT, KORA- σ Y CR+	66
FIGURA 3.5 REPRESENTACIÓN DE UN CLASIFICADOR ESTADÍSTICO	67
FIGURA 3.6 ESQUEMA DE CLASIFICACIÓN NO SUPERVISADA	68
FIGURA 4.1 EIGENFACES DE UN CONJUNTO DE ENTRENAMIENTO	73
FIGURA 4.2 EIGENFACES	75
FIGURA 4.3 PUNTOS DE REFERENCIA EN UN ROSTRO	75
FIGURA 4.4 IMÁGENES FILTRADAS CON FILTROS DE GABOR	76
FIGURA 4.5 RECONOCIMIENTO DE ROSTROS EN IMÁGENES 3D	78
FIGURA 4.6 EL PROBLEMA DE ILUMINACIÓN	79
FIGURA 4.7 EL PROBLEMA DE POSE (E ILUMINACIÓN). ARRIBA, SÓLO EL PROBLEMA DE POSE. ABAJO EL PROBLEMA DE POSE E ILUMINACIÓN COMBINADOS.	79

BIBLIOGRAFÍA

[1] **Maravall Gómez-Allende, Dario.** *Reconocimiento de Formas y Visión Artificial.* (Addison-Wesley Iberoamericana/RA-MA, E. U. A., 1994).

[2] **Wood, Richard E. & González, Rafael C.** *Digital Image Processing.* (Addison-Wesley, E. U. A., 1996).

[3] **Therrien, Charles W.** *Decision, Estimation and Classification. An Introduction to Pattern Recognition and Related Topics.* (John Wiley & Sons, E. U. A., 1989).

[4] **González, R. C. & Wintz, P.** *Digital Image Processing.* (Addison-Wesley, E. U. A., 1977).

[5] **Pertusa, J. F.** *Técnicas de Análisis de Imagen. Aplicaciones en Biología.* (Publicaciones de la Universidad de Valencia, España, 2003).

[6] **Rosenfeld, A & Kak, A. C.** *Digital Picture Processing.* (Academic Press, Inglaterra, 1981).

[7] **Levine, M. D.** *Vision in Man and Machine.* (McGraw-Hill, E. U. A., 1985).

[8] **Castleman, K. R.** *Digital Image Processing.* (Prentice-Hall, E. U. A., 1996).

[9] **Hu, M.-K.** *Visual Pattern Recognition by Moment Invariants.* (IRE Trans. Info. Theory, IT, 1962).

[10] **Javidi, Bahram.** *Image Recognition and Classification. Algorithms, Systems and Applications.* (Marcel Dekker, E. U. A., 2002)

[11] **Instituto Nacional de Astrofísica, Óptica y Electrónica.** INAOE. Ciencias Computacionales.

<http://ccc.inaoep.mx/index.html>.

[12] **Reyes, Juan.** Reconocimiento de Patrones: Una aproximacion para expresion de instrumentos y arte interactiva.

<http://ccrma-www.stanford.edu/~juanig/articles/charlAndes/charlAndes.html>

[13] **Mainguet, Jean-François.** Jean-François Mainguet: page d'accueil.

<http://pagesperso-orange.fr/fingerchip/index.htm>

[14] **Wikipedia.** Wikipedia La enciclopedia libre. *Modelo de color RGB.*
http://es.wikipedia.org/wiki/Modelo_de_color_RGB

[15] **Wikipedia.** Wikipedia La enciclopedia libre. *Teoría del color.*
http://es.wikipedia.org/wiki/Teor%C3%ADa_del_color

[16] **Instituto Latinoamericano de la Comunicación Educativa.** Ciencias de la Tierra.
<http://omega.ilce.edu.mx:3000/sites/ciencia/volumen1/ciencia2/33/htm/percep.htm>

[17] **Departamento de Ingeniería de Sistemas y Automática.** Sistemas Biométricos.
<http://www.disa.bi.ehu.es/>

[18] **How stuff works.** How Facial Recognition Systems Work.
<http://computer.howstuffworks.com/facial-recognition.htm>

[19] **Wikipedia.** Wikipedia La enciclopedia libre. *Eigenface.*
<http://en.wikipedia.org/wiki/Eigenface>