

7 Interpretación automática de las imágenes

Una vez que se ha particionado la imagen en regiones de interés (segmentación) y se han corregido los errores de esta etapa (post-procesado), los objetos presentes en el escenario deberán ser cuantificados para tareas de reconocimiento o localización. Se trata de asociar a cada elemento segmentado con un conjunto de valores numéricos o de atributos, al que se le llamará vector de características. Estos valores servirán de entrada al sistema de clasificación de los objetos, dando finalmente una etiqueta cualitativa a cada objeto presente en la imagen, cerrando de esta forma la interpretación automática de las imágenes. Por tanto, en este capítulo se verá que dada la información de partida, la imagen segmentada y post-procesada, se procederá en primer lugar a la etapa de etiquetamiento, donde a cada objeto de interés se le asociará una etiqueta (sección 7.1). Una vez etiquetada la imagen, será posible extraer de forma particularizada las características de cada objeto (sección 7.2). Por último, a cada objeto de la imagen se le asignará una etiqueta cualitativa (sección 7.3), dando por concluida la interpretación de la imagen.

7.1 Etiquetamiento de la imagen binaria

Se parte de una imagen segmentada donde los objetos han sido delimitado y separados del fondo, de manera que los píxeles pertenecientes a las regiones de interés han sido etiquetadas con un uno lógico y el resto con cero lógico. El siguiente paso será etiquetar cada una de los objetos presentes en la imagen, separándolo respecto del fondo y de las otras regiones. Esta etapa se realiza con la operación de etiquetamiento y se

fundamenta en la continuidad de los objetos en el espacio, cuya propiedad se transforma en las imágenes discretas en relaciones de conectividad entre píxeles adyacentes. Partiendo de la esquina superior izquierda de la imagen, se rastrea hacia la derecha y hacia abajo buscando píxeles con etiqueta uno lógico.

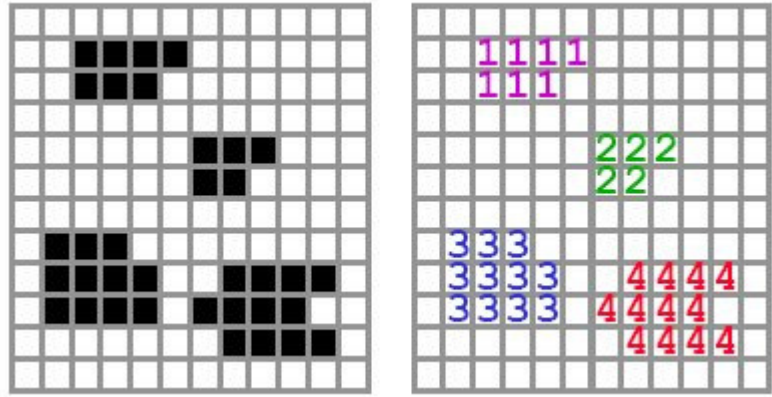


Figura 7. 1 Proceso de etiquetado de una imagen binaria

Cuando se encuentra el primer píxel con dicha etiqueta se le coloca la etiqueta '1', los vecinos que tengan propiedad de conectividad y que posean el nivel lógico activo se les pondrá la misma etiqueta. Al seguir rastreando en la imagen y al encontrarse con un píxel activado sin vecindad con los anteriores se le asociará con la etiqueta 2 y así sucesivamente. Una vez finalizada esta etapa, cada objeto de la imagen tendrá un identificador numérico que le hace ser distinto respecto del fondo y de los otros objetos.

Resolución Matlab

```
>>imgEnt=imread('rice.png');imshow(imgEnt);pause;
>>imgBWMask=im2BW(imgEnt);
>>se = strel('disk',2);
>>imgBWMarcador = imerode(imgBWMask,se);
>>imgReconst=imreconstruct(imgBWMarcador,imgBWMask);
>>imgBWElimBorde = imclearborder(imgReconst);
>>imgEtiq=bwlabel(imgBWElimBorde);
>>subplot(1,2,1);imshow(imgEnt);subplot(1,2,2);imshow(label2rgb(imgEtiq));
```

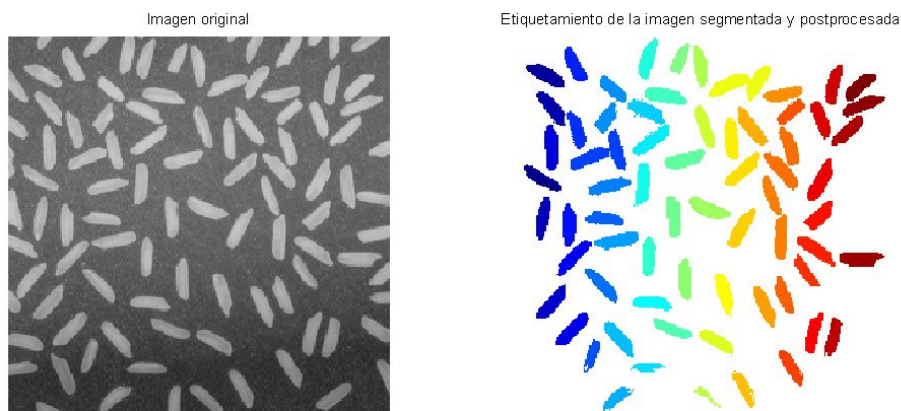


Figura 7. 2 Resultado del etiquetamiento de la imagen de granos de arroz

7.2 Extracción de características

Una vez obtenida la imagen etiquetada, los objetos pueden ser cuantificados, obteniendo así el vector de características. Las propiedades de los objetos se clasifican en dos grandes grupos: aquellas relacionadas con el contorno de los objetos y aquellas características propias del conjunto total del objeto o región de cada una de ellas. En esta sección se hablará de estas medidas.

No hay que olvidar que el objetivo de estos vectores de características servirá de entrada para la clasificación o localización de los objetos. Por tanto, a veces, se requerirá que la métrica empleada sea invariante a efectos de escalado, rotación o posición. También será tratado este aspecto.

7.2.1 Descriptores de regiones

Estas propiedades están relacionadas con el conjunto total de píxeles que constituye el objeto. Se podrían agrupar en tres grandes grupos: métricas, topológicas y texturales.

7.2.1.1 Métricas

Son medidas relacionadas con la distancia euclídea entre píxeles. Las más simples serían el área, el perímetro y el centro de gravedad.

El área, $A(i)$, es el número de píxeles que contiene el objeto, i . Si se pondera cada píxel por su nivel de gris se le define como el peso del objeto, $W(i)$:

$$A(i) = \sum_{x=1}^N \sum_{y=1}^M g_i(x, y) \quad W(i) = \sum_{x=1}^N \sum_{y=1}^M g_i(x, y) \cdot f(x, y) \quad (7.1)$$

Siendo $g_i(x, y)$ una función que es uno si el píxel pertenece al objeto i y nulo en caso contrario. El perímetro, $P(i)$, es el número de píxeles que constituye el borde del objeto, i . También se suele utilizar la relación del perímetro al cuadrado del objeto entre su área, $P^2(i)/A(i)$. Magnitud adimensional, cuyo valor mínimo se tiene cuando el objeto es un círculo, por tanto, describirá la compacidad del objeto. Valores próximos a 4π indicará que se aproxima a un círculo.

El centro de gravedad del objeto, (\hat{x}_i, \hat{y}_i) , indicará su posición puntual en la imagen y estará dado por:

$$\hat{x}_i = \frac{\sum_{x=1}^N \sum_{y=1}^M x \cdot g_i(x, y)}{A(i)} \quad \hat{y}_i = \frac{\sum_{x=1}^N \sum_{y=1}^M y \cdot g_i(x, y)}{A(i)} \quad (7.2)$$

Según el teorema de unicidad de Papoulis dice que, si $g_i(x, y)$ es continua a trozos y tiene valores no nulos sólo en una zona finita del plano x - y , entonces todos sus momentos existen y la aplicación que asigna una secuencia de momentos a la función es biyectiva. En conclusión, tomando un número finito de momentos se obtiene una aproximación del objeto. Para el caso de imágenes discretas y etiquetadas, los momentos de orden $p+q$ se definen como:

$$m_{pq}(i) = \sum_{x=1}^N \sum_{y=1}^M x^p \cdot y^q \cdot g_i(x, y) \quad (7.3)$$

Obsérvese que el área del objeto coincide con el momento de orden cero y que el centro de gravedad está unido con el momento de orden uno ($p=1, q=0$ y $p=0, q=1$). Los momentos se pueden hacer invariantes a traslaciones, si son referidos al centro de gravedad del objeto. A éstos se les llama momentos centrales:

$$mc_{pq}(i) = \sum_{x=1}^N \sum_{y=1}^M (x - \hat{x}_i)^p \cdot (y - \hat{y}_i)^q \cdot g_i(x, y) \quad (7.4)$$

Para ser invariante a escalados, se normaliza los momentos centrales con el área del objeto, esto es, con el momento de orden cero. A esta colección se la define como momentos centrales normalizados:

$$\mu_{pq} = \frac{mc_{pq}}{m_{00}} \quad (7.5)$$

Los ejes mayor y menor de un objeto se definen en términos de sus fronteras y son útiles para indicar su orientación y redondez. Los ejes, dados en píxeles, son obtenidos por la equivalencia del segundo momento central normalizado de la región con una elipse que lo inscribe. Los ejes mayor y menor hacen referencia a la elipse equivalente, mientras su orientación es el ángulo de apertura entre el eje horizontal y el eje mayor del objeto. Otro parámetro empleado es la excentricidad de la elipse



Figura 7.3 Elipse equivalente del objeto etiquetado

equivalente¹. Este parámetro está entre 0 y 1. En el caso de valer 0 indica que es una circunferencia y si es cero es un segmento recto.

Resolución Matlab

```
>>imgEnt=imread('rice.png');
>>imgBWMask=im2BW(imgEnt);
>>se = strel('disk',2);
>>imgBWMarcador = imerode(imgBWMask,se);
>>imgReconst=imreconstruct(imgBWMarcador,imgBWMask);
>>imgBWElimBorde = imclearborder(imgReconst);
>>imgEtiqu=bwlabel(imgBWElimBorde);
>>stat=regionprops(imgEtiqu,'all');
>>stat(20).Eccentricity
```

7.2.2 Descriptores topológicos

Las propiedades topológicas son descripciones globales de los objetos invariantes a rotaciones, escaladas o traslaciones. Por ejemplo, el número de agujeros dentro del objeto. Otro descriptor utilizado es el número de componentes conectados. Se trata de número de elementos separados que forman un objeto. Resulta evidente que estas propiedades no están relacionadas con el concepto de distancia, la cual se ha empleado en las propiedades métricas.

Un descriptor topológico muy empleado es el número de Euler. Éste se calcula con la diferencia del número de componentes conectados de la región menos el número de agujeros de ésta. Por ejemplo, el número de Euler de 'A' será 0, el de 'B' -1 y el 'i' es +2. Los descriptores topológicos se suelen emplear en aplicaciones de reconocimiento de caracteres.



Figura 7. 4 Objeto con número de Euler -1

7.2.3 Texturas

La textura explica la composición de las superficies de los objetos en términos de suavidad, rugosidad, granularidad, regularidad,... Desde el sentido de la visión se observa como repeticiones de patrones locales de luminancia en las imágenes. La sensación que transmiten las imágenes de zonas de tosquedad, rugosidad o suavidad, nacen de las correspondencias de las experiencias humana entre el sentido del tacto y la

¹Excentricidad de la elipse, $e = \sqrt{1 - \left(\frac{b}{a}\right)^2}$, donde a y b son el tamaño de los semiejes mayor y menor respectivamente.

vista. Desde el punto de vista físico, la interacción entre la luz y la materia, en la radiación reflejada, muestra ciertas repeticiones de carácter estadístico.

La imagen resultante no es una repetición local del patrón de forma determinista, sino probabilística y con una distribución en el espacio no exactamente regular. Esta



Figura 7. 5 Imágenes con diferentes texturas

propiedad en la imagen se manifiesta de forma regional, de manera que si se asciende a una visión global pueden aparecer diferentes texturas en la imagen. Por el contrario, al descender y observar la imagen a nivel de píxel o entorno de vecindad del píxel, la propiedad de la textura ha desaparecido. Lo mismo sucedería si se alejase excesivamente del escenario, la textura observada quedaría diluida. Por esto, la idea de textura está asociado a un espacio de escalas. La noción texel o primitiva textural se define como una región que posee ciertas propiedades visuales invariantes a la posición, al nivel de brillo o a cambios de perspectivas.



Figura 7. 6 La textura está relacionada con el espacio de las escalas

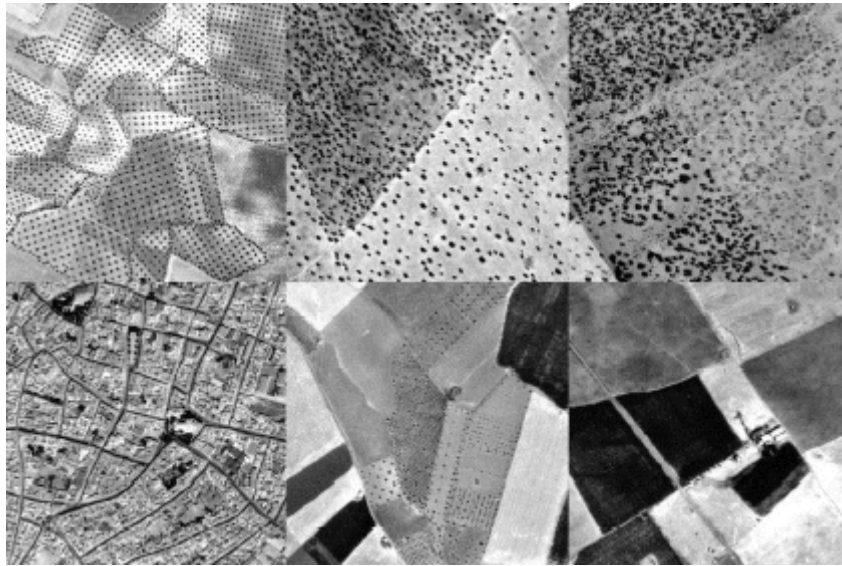


Figura 7. 7 Imágenes de satélite sobre el suelo de la Tierra que presentan diferentes texturas

Las aplicaciones del análisis textural no sólo se dan para la caracterización de los objetos de interés, sino que también son utilizadas en los procesos de segmentación. En la figura 7.8 se observa los procesos de segmentación en la imagen al ir incorporando más información. Se aprecia el resultado de la segmentación desde que sólo se emplea el canal de luminancia hasta la incorporación del color y la textura.

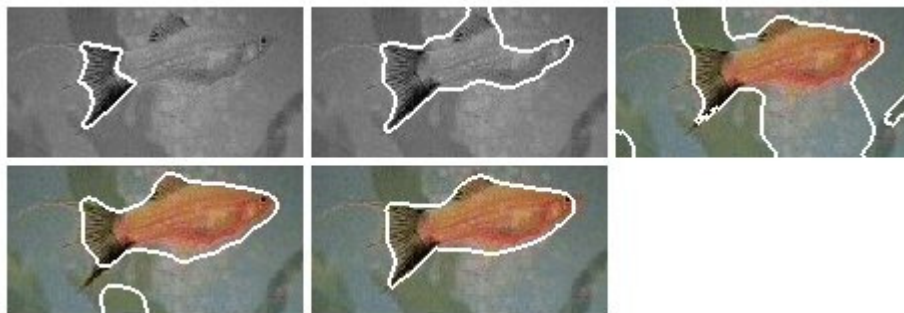


Figura 7. 8 Segmentación del pez con distintos tipo de información. a) niveles de grises, b) nivel de gris + textura, c) color RGB, d) color HSI, e) color + textura

El principal reto al que se enfrentan las técnicas de análisis textural consiste en la descripción de cómo son y cómo se distribuyen los elementos de textura en una imagen. Los descriptores de texturas deben cuantificar ciertas propiedades tales como suavidad, rugosidad y regularidad. Estas características deben ser invariantes a la posición, orientación y nivel de brillo medio. Existen tres enfoques distintos: estructurales, estadísticos y espectrales.

Las técnicas estructurales consideran la textura como compuestas por primitivas que forman un patrón repetitivo y describen este patrón mediante reglas capaces de generarlo o reproducirlo. Formalmente, estas reglas constituyen la gramática de la textura que describe. Estas técnicas resultan aplicables con éxito en el análisis de imágenes que contienen una textura en la que los elementos descriptibles siguen una gran regularidad en su distribución.

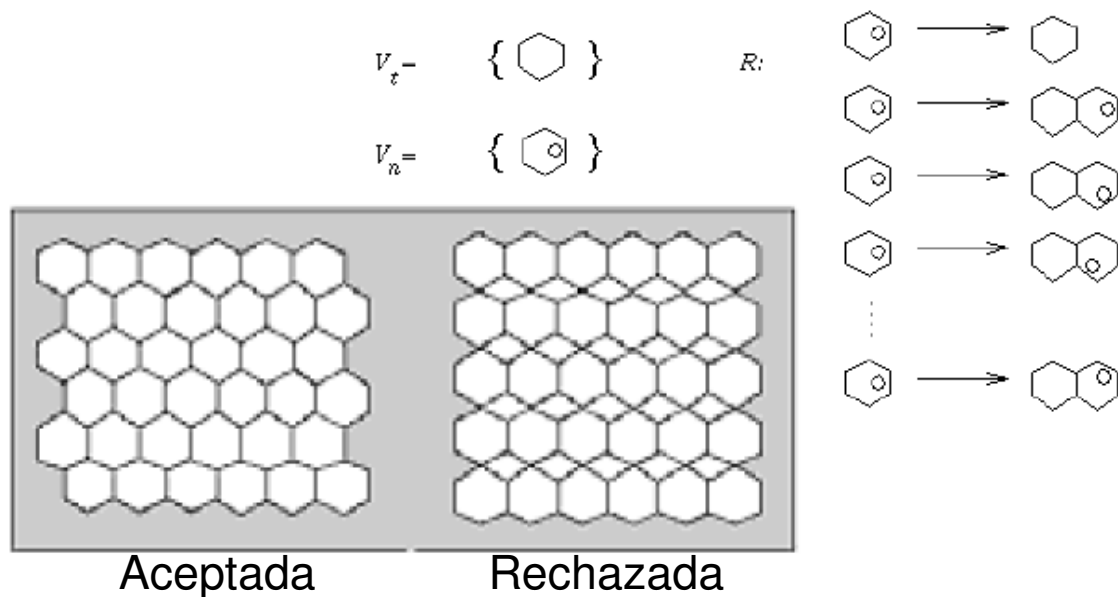


Figura 7. 10. Descripción estructural de la textura. Se define unas primitivas y unas reglas de encadenamiento de las primitivas. En la parte superior se ha representado las primitivas y las reglas de encadenamiento. En la parte inferior un ejemplo de aceptación y rechazo de un tipo de textura

Por el contrario, el análisis espectral busca en las transformadas de Fourier, picos de alta densidad en el módulo o en el argumento para caracterizar los patrones texturales. Se emplea para caracterizar patrones periódicos y de forma global en la imagen. Estos descriptores son invariantes a posición y rotación.

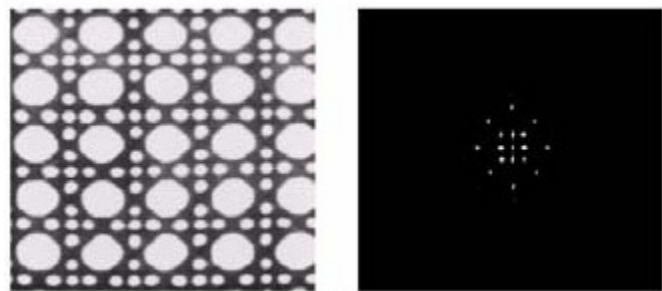


Figura 7. 9 Imagen de una textura sintética y el módulo de su espectro frecuencial

El mayor problema de las transformadas de Fourier es que cada componente del espectro frecuencial depende de la imagen global. Como se muestra en la figura 7.11, puede haber una variación espacial del patrón en la imagen.



Figura 7. 11 Variación de la textura con diferentes escalas

Una mejora de este procedimiento es utilizar los filtros de Gabor. Con estos filtros, el análisis espectral también considera el espacio de escala y una cierta orientación privilegiada. La máscara de convolución de cada filtro se construye con una función gaussiana modulada con un armónico de una determinada orientación y frecuencia:

$$G_{\phi, f}(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \cdot e^{2\pi j f (x \cos \phi + y \sin \phi)} \quad (7.6)$$

La varianza y la frecuencia están relacionadas y definen la escala del patrón:

$$\sigma = 3 \frac{\sqrt{2 \ln 2}}{2\pi} f \quad (7.7)$$

Normalmente, se emplea cuatro orientaciones y tres escalas, teniendo un conjunto de doce máscaras de convolución.

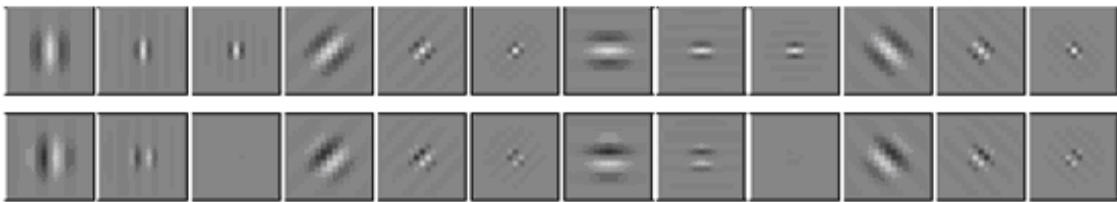


Figura 7. 12 Máscaras de convolución con 4 diferentes orientaciones $\phi = \left\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\right\}$ y tres diferentes frecuencias $f = \{0.2, 0.35, 0.5\}$. La fila superior es la componente real y la inferior la componente imaginaria.

Al convolucionar estos filtros con la imagen, los altos valores indican la presencia del patrón. En el ejemplo de las cebras, se puede observar cómo al procesar la imagen con las máscaras se puede discernir la posición de las cebras respecto del fondo



Figura 7. 13 Imagen original

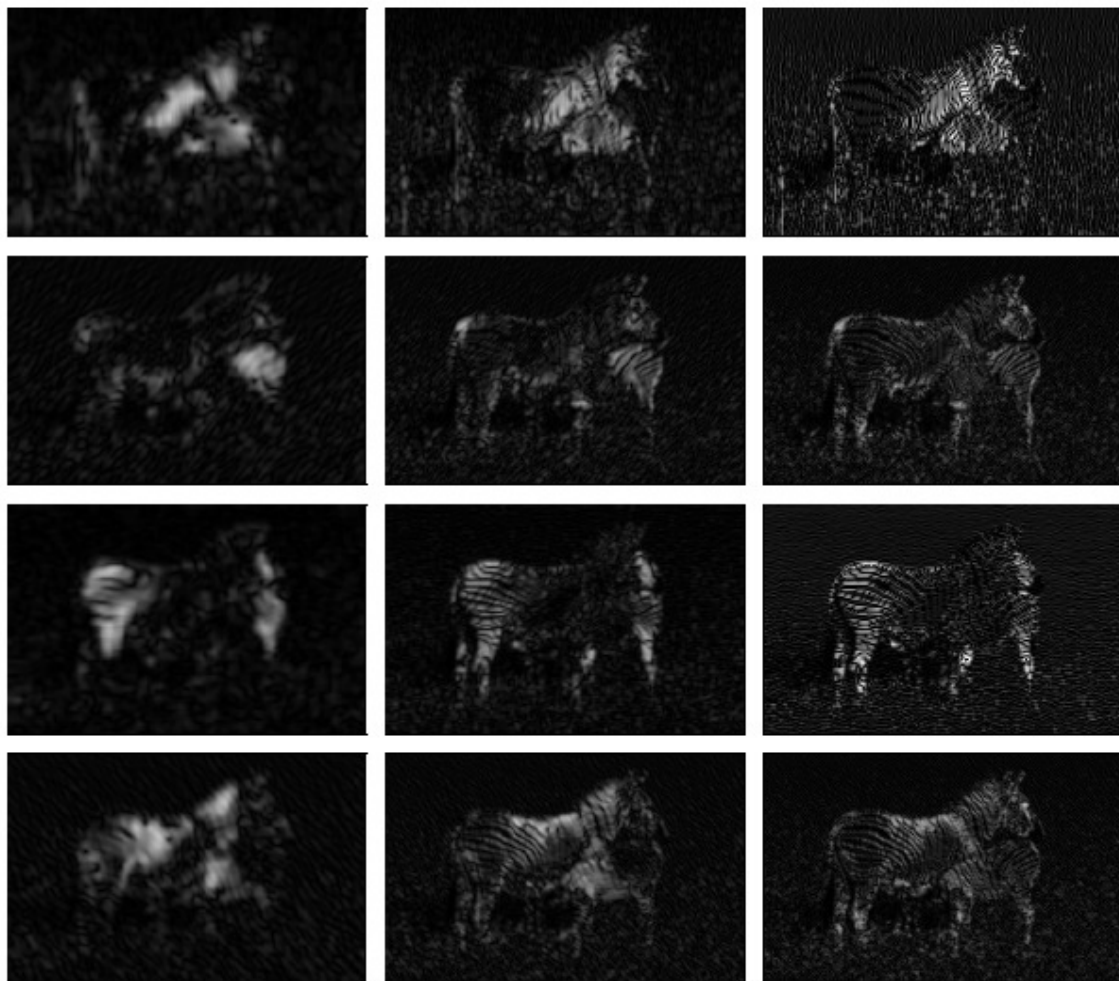


Figura 7. 14 Resultados de la convolución de la imagen con las distintas máscaras

7.2.3.1 Técnicas estadísticas

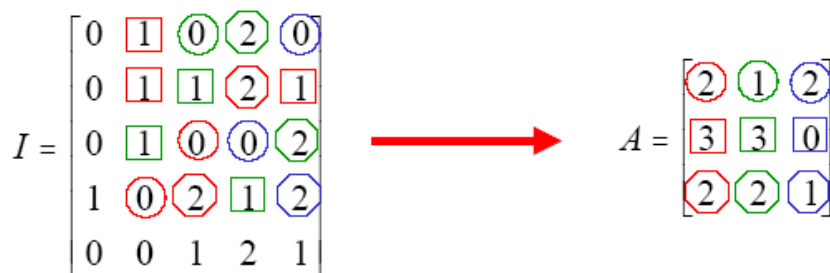
Las técnicas estadísticas describen la textura mediante reglas estadísticas que gobiernan la distribución y la relación espacial de niveles de gris en la imagen. Éstas presentan buen comportamiento en el análisis de texturas naturales o texturas con poca resolución, en donde los elementos de textura resultan difícilmente descriptibles.

Estas se clasifican en: técnicas estadísticas de primer orden, de segundo orden y de orden superior. En el primer grupo, se obtienen medidas considerando el valor de píxel, sin consideraciones de vecindad; mientras en las técnicas de segundo orden tratan de la distribución espacial de parejas de píxeles vecinos y las de orden superior consideran triplete, ternas, ... de píxeles vecinos.

En las estadísticas de primer orden se puede obtener el histograma de la región, cuya normalización proporcionará la función de densidad de probabilidad de la textura. Se pueden comparar los histogramas normalizados entre regiones o utilizar medidas derivadas, tales como la media, la varianza, energía, entropía, etc. (ver sección 4.1.1).

La principal desventaja de las técnicas de primer orden es su falta de sensibilidad ante permutaciones de los píxeles. La vía más apropiada para evitar la limitación anterior es considerar los estadísticos de segundo orden. A partir de ellos se obtienen las matrices de dependencia espacial, generadora de medidas de textura más fiables y robustas. Las matrices de co-ocurrencia del nivel de gris, C , es un ejemplo de esta fuente de medidas de textura. Se define en relación con un desplazamiento de valor h , en una dirección θ , desde una determinada fila y columna de la imagen; el elemento (i,j) de la matriz de co-ocurrencia, c_{ij} , representa el número de veces que un píxel con nivel de gris j , se encuentra a una distancia h , en dirección θ , de un píxel con nivel de gris i (una ocurrencia conjunta). Si N_h representa el número total de ocurrencias, entonces la fracción entre la matriz de co-ocurrencia y N_h representa la matriz de co-ocurrencia normalizada.

Por ejemplo, considerando una imagen con tres niveles de grises (0,1 y 2) y utilizando como



regla de vecindad el píxel superior a la derecha. La

Figura 7. 15 Ejemplo de extracción de la matriz de co-ocurrencias

matriz de co-ocurrencia mostrará por cada celda i,j las ocurrencias de encontrar un píxel de nivel gris i que tenga como vecino superior a la derecha con nivel gris j . En el ejemplo la matriz será de 3x3. La celda (0,0) indica el número de ocurrencias de píxeles con nivel de gris 0 que tenga como superior a la derecha otro píxel con nivel gris 0. Para este caso, aparecen 2 ocurrencias. En cambio el elemento (0,1) indica las ocurrencias de píxeles que tengan nivel de gris 1 y que su vecino superior a la derecha sea nulo.

A partir de la matriz de co-ocurrencia se obtienen diferentes descriptores.

7.2.4 Descriptores de fronteras

7.2.4.1 Códigos encadenados

Partiendo del perímetro del objeto se procede a su codificación a través de segmentos con longitud y orientación determinada. La cadena de códigos describe el objeto mediante la sucesión de estos segmentos conectado píxeles adyacentes pertenecientes al perímetro del objeto. La conexión

se realiza con conectividad a 4 o a 8 píxeles. Partiendo de un origen determinado y en el sentido horario

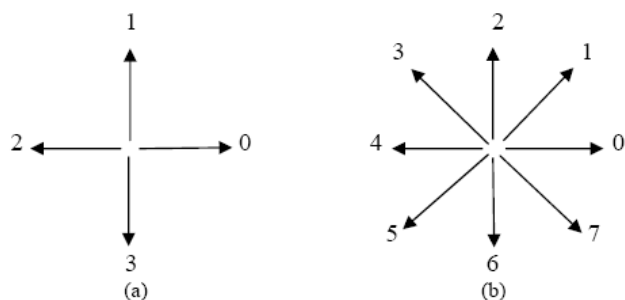


Figura 7. 16 Codificación según vecindad a 4 ó 8

se procede a codificar la frontera del objeto. En la figura 7.17 se muestra un ejemplo de cómo se codificaría a vecindad 4 u 8.

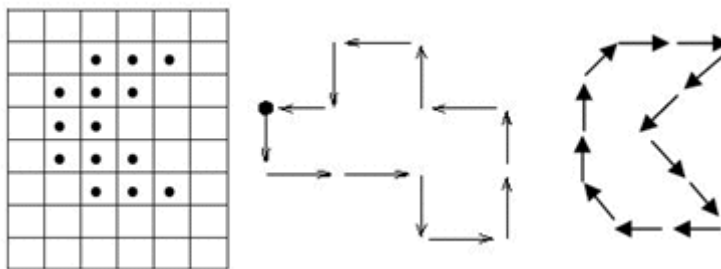


Figura 7. 17 Ejemplo de código encadenado con vecindad a 4 y 8

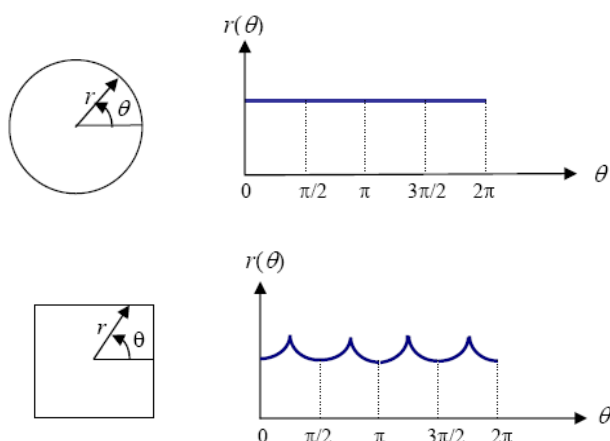
El código, iniciando en el punto y con vecindad a 4 sería 300301121232. El primer elemento debe llevar información de su posición. Si se desea comparar el código con otros se procede a su normalización. Una posibilidad de normalización es iniciar por el píxel de la frontera que genere menor valor en su codificación. En el ejemplo sería 00301121232. Si se emplease vecindad a 8 el código sería 005577443221.

Este descriptor es invariante a traslaciones, permitiendo más fácilmente la comparación entre objetos. Con esta codificación se obtiene de manera más eficiente la medida del perímetro y de los descriptores de Fourier que empleando la imagen etiquetada. El mayor inconveniente es la presencia de ruido.

7.2.4.2 Descriptores geométricos

El perímetro se calcula a partir del código encadenado, sólo debe añadir 1 ó $\sqrt{2}$, dependiendo si hace referencia a un código horizontal/vertical o de carácter diagonal respectivamente.

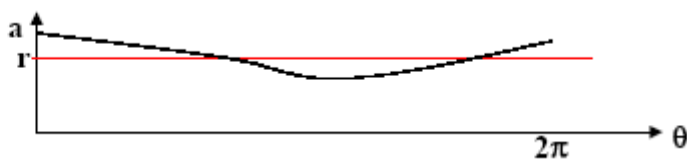
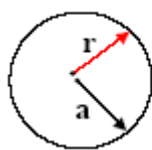
Para facilitar la representación de las curvas se emplean funciones unidimensionales que la describen. La primera sería la signatura. Ésta codifica la distancia de un punto interior del objeto a todos los puntos de la frontera. Normalmente, se suele utilizar el centroide, como punto interior.



Aunque es invariante a traslaciones, depende del tamaño y desde donde se inicializa el ángulo de la representación polar. Para que sea invariante a escala se normaliza la distancia en el rango $[0, 1]$; sólo basta con dividir todas las distancias por la que sea mayor. Respecto a la dependencia al punto inicial, se toma aquel que tengo la máxima distancia.

Figura 7. 18 Ejemplos de signatura

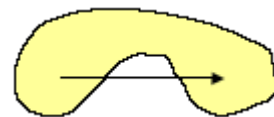
La signatura es muy sensible a la elección del punto interior, cambiando sustancialmente si se desvía del cálculo del centroide.



También tiene el problema de la

Figura 7. 20 Sensibilidad de la signatura con la elección del punto interior

aparición de concavidades en el objeto apareciendo una representación multievaluada para algunos ángulos. Suele emplearse la envolvente convexa².



La curvatura media mide la pendiente (tangente) del contorno. La evolución de la curvatura en la frontera es invariante a traslación, rotación y escalado. Para su cálculo se puede emplear la expresión:

Figura 7. 19 Objeto con concavidades

$$\kappa = \text{div} \left(\frac{\nabla f}{|\nabla f|} \right) \quad (7.8)$$

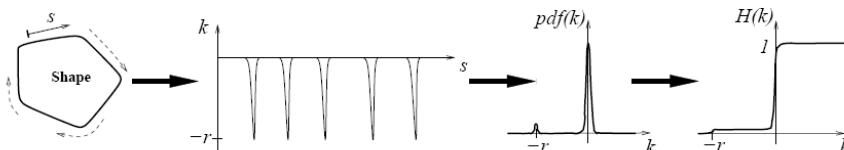


Figura 7. 21 Obtención de la función de distribución de la curvatura media

Partiendo del concepto de curvatura se define la energía de doblado. Se define como la energía necesaria para transformar una varilla en una determinada frontera:

$$ED = \frac{1}{n} \sum_{i=1}^n |\kappa(i)|^2 \quad (7.9)$$

Siendo n el número de puntos pertenecientes a la frontera.

7.2.4.3 Descriptores de Fourier

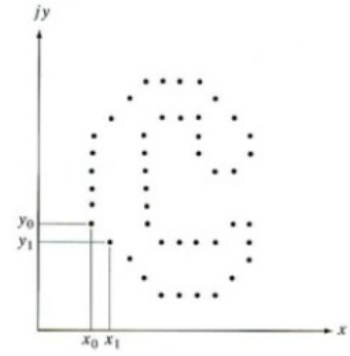
Es otra técnica para pasar la información de bidimensional a otra unidimensional. Dada una curva cerrada, ésta puede ser representada por una

² Una trozo de la curva es convexa si el radio de curvatura está en el interior del objeto y cóncava si está fuera,

transformada discreta de Fourier. La curva viene dada en coordenadas cartesianas, obteniendo una secuencia de n puntos pertenecientes a la frontera:

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i), \dots, (x_n, y_n)\}$$

Cada punto es convertido a variable compleja, $z(i) = x_i + jy_i$. La aplicación sobre esta secuencia compleja de la transformada de Fourier obtendrá una respuesta espectral:



$$Z_k = \sum_{i=1}^n z_i e^{-j\frac{2\pi k}{K}i} \quad k = 0, 1, 2, \dots, K-1 \quad (7.10)$$

Obteniendo K descriptores frecuenciales. La componente de continua representaría el centroide del objeto, las componentes de baja frecuencia estaría relacionados con los aspecto más grueso del objeto y los de alta frecuencia con los detalles.

La secuencia compleja se puede recuperar a partir de la anti-transformada:

$$z_i = \frac{1}{K} \sum_{k=0}^{K-1} Z_k e^{j\frac{2\pi k}{K}i} \quad i = 1, 2, \dots, n \quad (7.11)$$

Si se elige un número de componentes P menor a K se tendrá una descripción sin menos detalles.

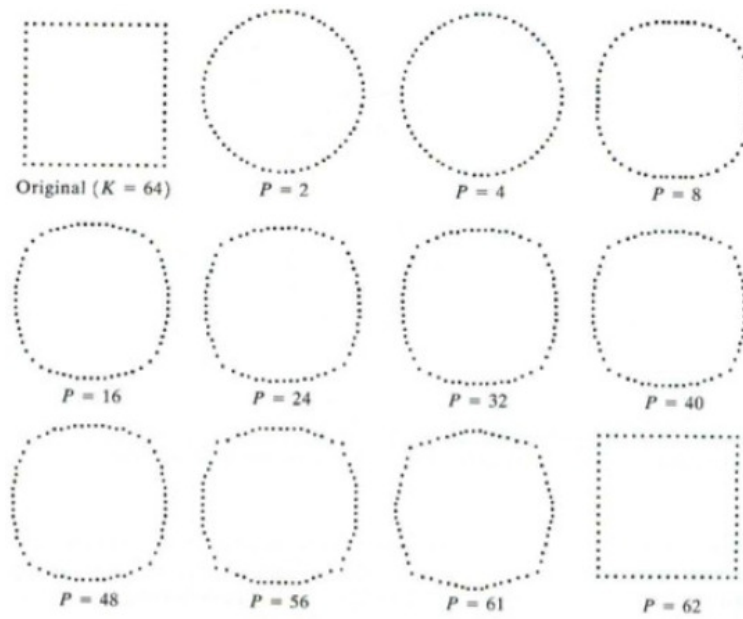
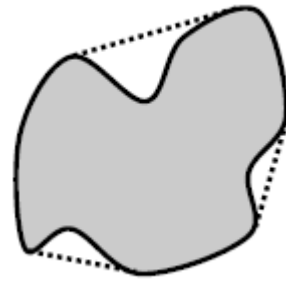


Figura 7.22 Evolución de la forma cuadrada al ir añadiendo P componentes frecuenciales

Estos descriptores son invariantes a traslaciones, rotaciones, escalado y de donde se empiece a tomar la secuencia.

7.2.4.4 Descripción envoltente convexa

El objetivo es descomponer el contorno en una convexidad que simplifique el proceso de descripción. Se trata de eliminar las concavidades del objeto. La envoltente convexa (*convex hull*), EC, de un conjunto C cualquiera se define como el conjunto convexo más pequeño que contiene a C.



7.3 Reconocimiento de patrones

El Reconocimiento de Patrones es una disciplina de la Ciencia encargada de asignar a los objetos una clase determinada. Los objetos no sólo proceden de las imágenes sino también de señales unidimensionales (p.ej. reconocimiento de voz o patrones en el electrocardiograma, ECG) o en cualquier otra medida sobre los objetos.

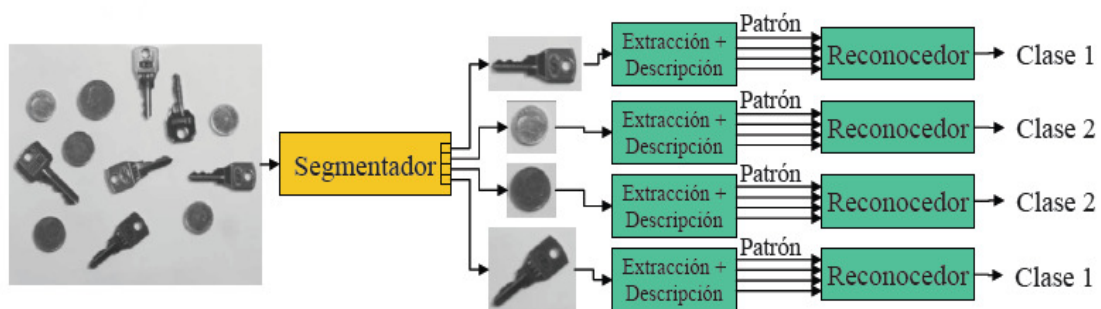


Figura 7. 23 Proceso de clasificación de las llaves/moneda empleando técnicas de Visión Artificial

La tarea de clasificación es una combinación del vector de características de los objetos y del conocimiento a priori que se tenga del universo del problema. Para poder asignar una etiqueta de clasificación a un objeto se debe de tener reglas de pertenencias a cada clase. Estas reglas se consiguen a través del conocimiento que se tenga de cada clase.

Los enfoques que se pueden emplear para la clasificación se basa bien en una función discriminante que divida el espacio de las características (clasificadores estadísticos) o bien empleen relaciones geométricas asociadas a los objetos (clasificadores sintácticos). Existe una tercera vía basada en las redes neuronales artificiales, las cuales se parecen a los clasificadores estadísticos, en cuanto que tratan de particionar el espacio de las características, pero emulando a los sistemas biológicos, empleando etapas previas de aprendizaje.

El esquema general del reconocimiento de objetos, tanto estadístico como neuronal, consiste en obtener funciones discriminantes que divida el espacio de características en tantas clases como estén presentes en el universo del problema. Por

tanto, la selección de las características resulta ser esencial para la buena clasificación de los objetos. Éstas deberán de tener las siguientes propiedades:

1. Capacidad de discriminación: Las características deben tomar valores significativamente distintos para cada clase.
2. Fiabilidad: las características deben tomar valores similares para todos los objetos de la misma clase.
3. Correlación: Las diversas características no deben estar correladas unas con otras, en caso contrario reflejarían la misma propiedad del objeto.
4. Número: la selección de las características debería de elegir el menor número de ellas, ya que permite generalizar más. A más características más datos de entrenamiento se requieren para mantener igual grado de generalidad y no de memorización.

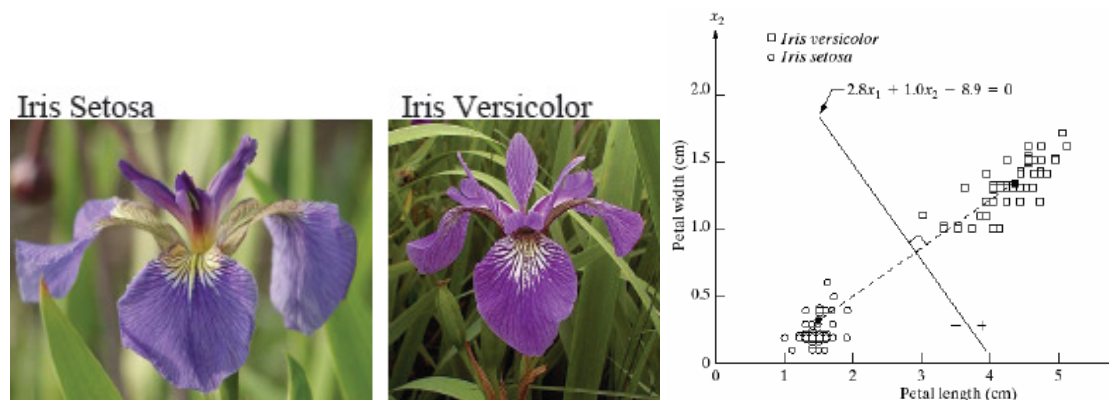


Figura 7. 24 Ejemplo de clasificación entre Iris setoda e Iris Versicolor empleando el ancho y longitud de los pétalos

7.3.1 Alineamiento de plantillas

El objeto a ser reconocido es comparada con un conjunto de plantillas, teniendo en cuenta que pueden haber habido traslaciones, rotaciones o cambio de escala. Muy utilizado en análisis médicas.

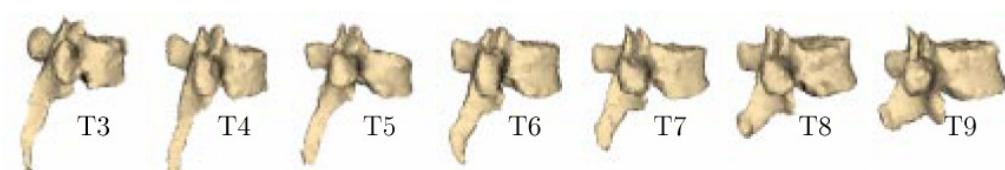


Figura 7. 25 Modelos 3D de vértebras humanas

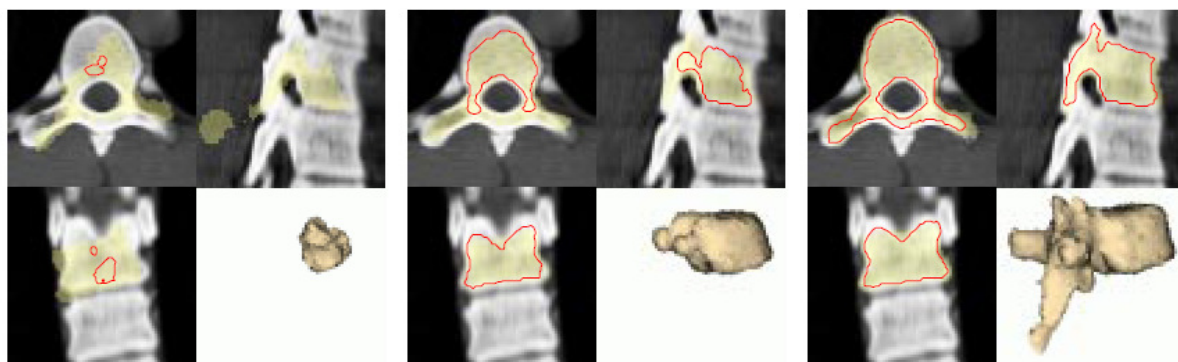


Figura 7. 26 Evolución de un contorno activo para la segmentación y clasificación de una vértebra humana

7.3.2 Métodos sintácticos

En este caso un objeto es visto como una composición de subformas simples que a su vez son construidas de subformas más simples. Las subformas más simple a ser reconocidas son llamadas primitivas y una forma compleja es representada en términos de interrelaciones entre estas primitivas. Este método tiene analogía con la sintaxis de un lenguaje. Hay problemas para separar en sus primitivas una forma que tiene mucho ruido. Requiere bastante esfuerzo computacional.

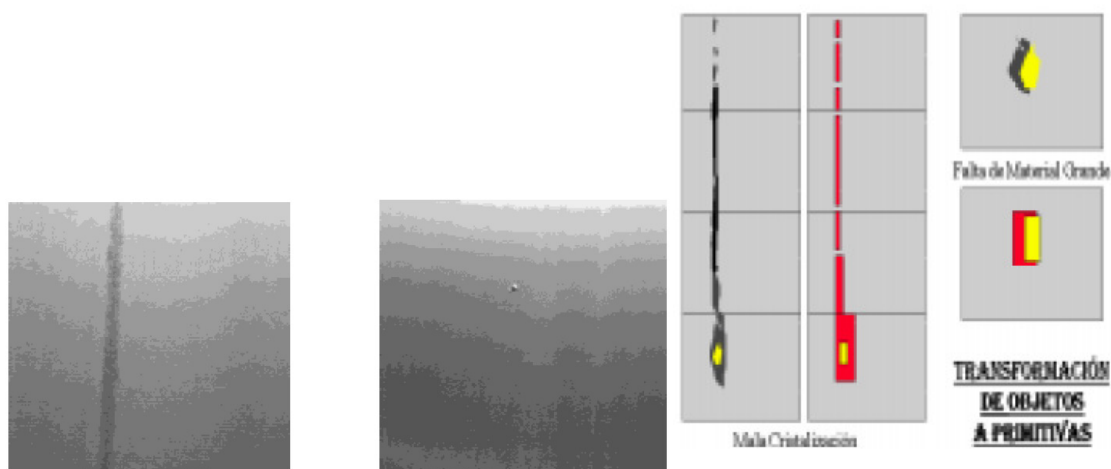


Figura 7. 27 Ejemplos de defectos en el aluminio colado: a) Mala cristalización, b) Desperfecto superficial.

7.3.3 Redes neuronales

Consiste en una red cuyos nodos son neuronas artificiales que se conectan mediante enlaces que tienen distintas ponderaciones. Las redes neuronales tienen la habilidad de aprender complejas relaciones no lineales de entrada-salida usando procedimientos secuenciales de

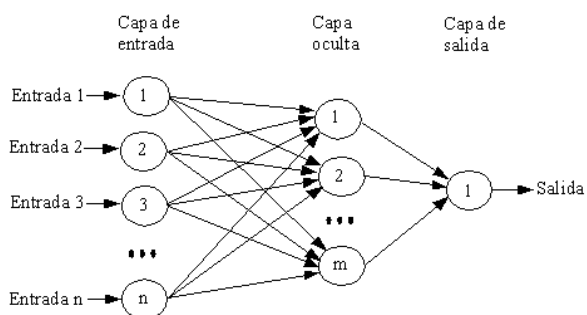


Figura 7. 28 Arquitectura de red neuronal

entrenamiento y además se adaptan ellas mismas a los datos. Las redes neuronales más usadas para efectos de clasificación supervisada son: Feed-forward networks (redes de alimentación hacia adelante) que incluye a MLP (Multilayer perceptron) y las redes de funciones de base radial (RBF). Para clasificación no supervisada se usa las redes de Kohonen conocidas como Self-Organizing Maps (SOM).



Figura 7. 29 Ejemplo de clasificación mediante una red neruronal de los terrenos vistos en imágenes aéreas

7.3.4 Clasificadores estadísticos

Aquí cada forma es representada como un conjunto de mediciones de p características y puede considerada como un punto en el espacio d dimensional. Primero hay que elegir aquellas características que permiten a los vectores de formas de distintas clases ocupar regiones disjuntas en el espacio p -dimensional. Dado un conjunto de formas el objetivo es establecer fronteras en el espacio de características que separen las formas que pertenecen a distintas clases. Hay dos metodologías: la basada en teoría de decisión y la basada en análisis discriminante. En la primera las fronteras son determinadas por las distribuciones de probabilidad de las formas que pertenecen a cada clase. En el segundo caso se especifica una forma paramétrica (lineal, cuadrática, etc) de la frontera y luego se encuentra la mejor frontera basándose en la clasificación de las formas disponibles en la muestra.

Cada una de las N clases, ω_i ($i \in 1..N$), se representa mediante un prototipo o centroide Z_i , el cual es un vector p -dimensional:

$$Z_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \quad (7. 12)$$

Siendo n_i el número de muestras de la clase i y X_{ij} el vector j de p -dimensiones que representa a la clase i . La función distancia euclídea entre una nueva muestra X y cada uno de los centroides Z_i del universo del problema:

$$d(X, \omega_k) = \sqrt{\sum_{i=1}^p (X_i - Z_{k_i})^2} \quad (7.13)$$

Permitirá discriminar a qué clase pertenece la nueva muestra. El objeto será asignado a la clase i que tenga menor distancia a Z_i . La fórmula anterior es equivalente a evaluar la expresión de la *función discriminante* de cada clase $f_i(X)$, siendo: $(i \in 1..k)$, para el patrón X y asignarlo a la clase ω_i para la que $f_i(X)$ sea máximo:

$$f_i(X) = X^T \cdot Z_i - \frac{1}{2} Z_i^T \cdot Z_i$$

A partir de las funciones discriminantes de clases se pueden construir las fronteras de decisión entre clases (hiperplanos), como se ha mostrado en el ejemplo de la familia de las Iris (ver figura 7.24). A veces no se puede conseguir una separación lineal entre clases. Esto se produce debido a que: 1) las características son inadecuadas para distinguir entre clases, 2) las características tienen una alta correlación, 3) las fronteras de decisión no son lineales, 4) hay subclases (dentro de las clases), o 5) el espacio de características es muy complejo. Se puede usar el conocimiento de las *distribuciones de probabilidad* de los valores de las características para clasificar objetos.

En este caso se empleará el teorema de Bayes. Se trata de minimizar el error de clasificación al adjudicar la muestra a la clase con mayor probabilidad a posteriori:

$$p(\omega_i | X) = \frac{p(X | \omega_i) p(\omega_i)}{p(X)} \quad (7.14)$$

Donde $p(\omega_i)$ es la probabilidad de que una muestra pertenezca a la clase ω_i y $p(X | \omega_i)$ es la probabilidad a priori. Ésta refleja que suponiendo que pertenece a la clase ω_i , la probabilidad que tiene el vector X , de hecho, corresponde con la función de densidad de la clase. Por último, $p(X)$ es la probabilidad de que se presente exactamente una muestra con el vector de características X . Cumpliéndose que:

$$p(X) = \sum_{i=1}^N p(X | \omega_i) p(\omega_i) \quad (7.15)$$

Para una nueva muestra X , $p(X)$ será la misma para todas las clases, luego la función discriminante será la probabilidad a posterior sin la probabilidad de la muestra:

$$f_i(X) = p(X | \omega_i) p(\omega_i) \quad (7.16)$$

La estimación de las funciones de densidad de probabilidad $p(X | \omega_i)$ es otra cuestión. Si los patrones vectoriales, X , son de dimensión n , $p(X | \omega_i)$ es una función de n variables, que, si su forma no es conocida, requiere de métodos de la teoría de

probabilidades de varias variables para su estimación. Estos métodos son difíciles de aplicar en la práctica. Por estas razones, el clasificador de Bayes se basa generalmente en la suposición de una expresión analítica para las diversas funciones de densidad y, posteriormente, en una estimación de los parámetros de la expresión para los patrones ejemplo de cada clase. La forma más habitualmente supuesta para $p(X|\omega_i)$ es la distribución gaussiana. Cuanto más se aproxime a la realidad esta suposición, más se aproximará el clasificador de Bayes al mínimo error medio de clasificación.

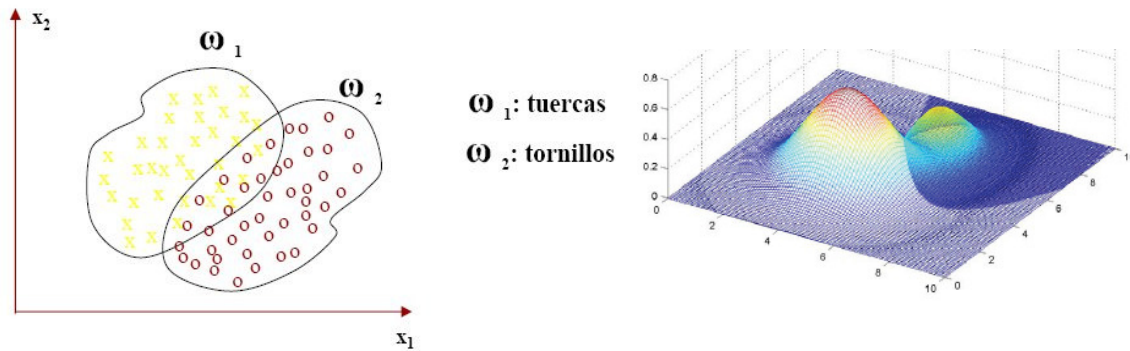


Figura 7. 30 a) nube de puntos de tornillos y tuercas, b) funciones de densidad $p(X|\omega_i)$

El modelo de función de densidad normal está definida por el vector de la media, M_i , y la matriz de covarianza, Σ_i :

$$p(X|\omega_i) = \frac{1}{(2\pi)^{n/2} |\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2}(X - M_i)^T \Sigma_i^{-1} (X - M_i)\right) \quad (7.17)$$

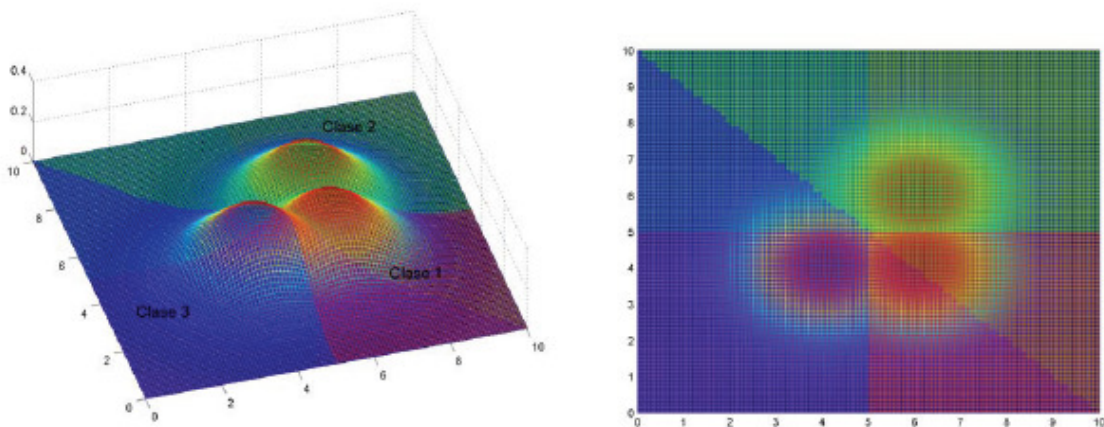


Figura 7. 31 Función de densidad gaussiana 2D para tres clases

Debido a la forma exponencial de la densidad gaussiana, es más conveniente trabajar con el logaritmo neperiano en la función discriminante:

$$f_i(X) = \ln(p(X|\omega_i)p(\omega_i)) = -\frac{n}{2} \left[\ln(2\pi) + \ln\left(|\Sigma_i|^{-\frac{1}{2}}\right) \right] - \frac{1}{2}(X - M_i)^T \Sigma_i^{-1} (X - M_i) + \ln(p(\omega_i)) \quad (7.18)$$

El término $-n/2 \cdot \ln(2\pi)$ es el mismo para todas las clases, por lo que se elimina de la ecuación.

Si todas las matrices de covarianzas de las clases fuesen idénticas y éstas son equiprobables, la función discriminante se define como la distancia de Mahalanobis:

$$f_i(X) = -\frac{1}{2}(X - M_i)^T \Sigma^{-1} (X - M_i)$$

Además para el caso particular de que las componentes del vector X no estuvieran correladas y fuesen de la misma magnitud, coincidiría con la función discriminante lineal o distancia euclídea:

$$f_i(X) = (X - M_i)^T (X - M_i)$$

Si no podemos asumir un modelo preciso para $p(X|\omega_i)$ hay que recurrir a estimadores no paramétricos de la función de densidad, como es el histograma. Sin embargo, el histograma está promediando valores en una región, y por tanto está generando una versión distorsionada de la función de densidad.

Suponiendo que se tiene muestras de una determinada clase, su probabilidad será:

$$p(X) = \frac{\int_R p(X) dX}{\int_R dX} \cong \frac{k/m}{V_R} \quad (7.19)$$

Siendo k el número de muestras del total m que caen en la región R de volumen V_R . Habrá que garantizar que cuando m tiende a infinito, la aproximación discreta coincida con la continua. Una forma de conseguirlo es a través de la ventana de Parzen. Supongamos que en la región R queda definida por una función $\phi(X)$ que encierra un volumen V_R . El número de datos que encierra esa región alrededor de X es:

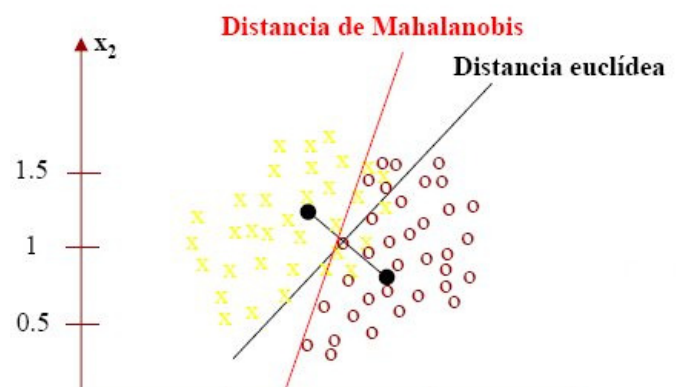


Figura 7. 32 Funciones discriminantes de Mahalanobis y euclídea para el problema de tuercas y tornillos

$$k = \frac{1}{h^n} \sum_{i=1}^m \phi\left(\frac{|x - x_i|}{h}\right) \quad (7.20)$$

Siendo $\phi(X)$ el hipercubo que selecciona la región. La función de densidad quedará como:

$$p(X) \cong \frac{k/m}{V_R} = \frac{1}{m \cdot h^n} \sum_{i=1}^m \phi\left(\frac{|x - x_i|}{h}\right) \quad (7.21)$$

Normalmente $\phi(X)$ emplea una expresión de tipo gaussiano para el suavizado de la función densidad, de manera que h define la apertura de la región y por tanto el carácter de $p(X)$. A medida de que h sea más grande, con más muestras de alrededor interaccionará y más suave será $p(X)$. El papel de h es idéntico a la de varianza.

$$p(X) \cong \frac{1}{m} \sum_{i=1}^m \frac{1}{(2\pi h^2)^{1/2}} \exp\left(-\frac{|X - X_i|^2}{2h^2}\right) \quad (7.22)$$

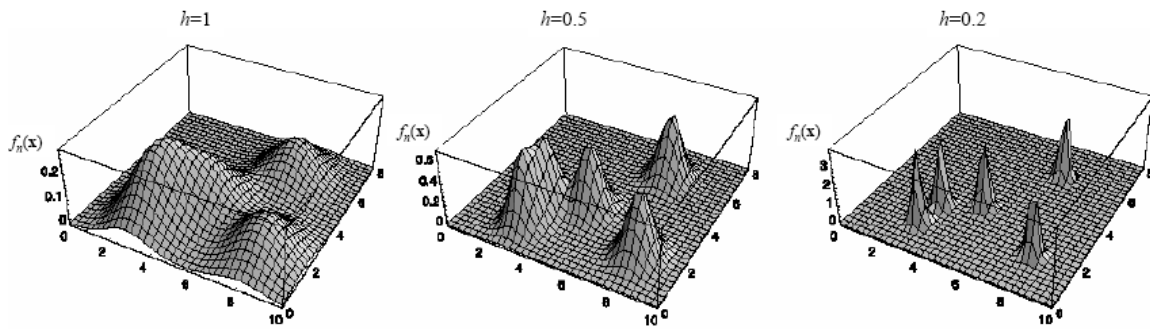


Figura 7.33 Estimación de $p(X)$ con 5 datos con tres diferentes anchos de ventana

Derecho de Autor © 2008 Carlos Platero Dueñas.

Permiso para copiar, distribuir y/o modificar este documento bajo los términos de la Licencia de Documentación Libre GNU, Versión 1.1 o cualquier otra versión posterior publicada por la Free Software Foundation; sin secciones invariantes, sin texto de la Cubierta Frontal, así como el texto de la Cubierta Posterior. Una copia de la licencia es incluida en la sección titulada "Licencia de Documentación Libre GNU".

La Licencia de documentación libre GNU (GNU Free Documentation License) es una licencia con [copyleft](http://www.gnu.org/copyleft/fdl.html) para [contenidos abiertos](http://www.gnu.org/copyleft/fdl.html). Todos los contenidos de estos apuntes están cubiertos por esta licencia. La versión 1.1 se encuentra en <http://www.gnu.org/copyleft/fdl.html>. La traducción (no oficial) al castellano de la versión 1.1 se encuentra en <http://www.es.gnu.org/Licencias/fdles.html>