

## RECUPERACIÓN DE IMAGEN

---

En este capítulo se introduce el concepto de recuperación de imagen (*image retrieval*), campo del ámbito de la recuperación de información en el cual el dominio de datos con el que se trabaja son las imágenes. Además, se realiza una introducción al Procesado Digital de Imágenes, que es el campo de trabajo que sirve de soporte fundamental a su recuperación. Los objetivos concretos del tema son:

- ✕ Conocer el esquema básico de los Sistemas de Recuperación de Imagen Basados en Contenido (SRIBC) y su importante analogía con los sistemas tradicionales de recuperación de texto.
- ✕ Comprender algunos conceptos básicos del Procesado Digital de Imágenes para el tratamiento de los elementos descriptivos fundamentales de una imagen: color, textura y forma.
- ✕ Ser capaz de construir un histograma para propiedades calculadas sobre una imagen y usarlo como descriptor de dicha imagen.
- ✕ Conocer la forma en la que se extraen las propiedades perceptuales y se establecen los rankings de similaridad en los SRIBC.
- ✕ Conocer otros tipos de recuperación de información visual.

## 1 INTRODUCCIÓN

En los últimos años se aprecia un crecimiento exponencial de la información visual disponible en forma de imágenes. Ante este crecimiento, surge la necesidad de desarrollar herramientas que permitan mejorar el acceso a este tipo de información. Esta motivación produce el despegue de una gran línea de investigación, a finales del siglo XX, bajo el nombre genérico de Sistemas de Recuperación de Imagen Basados en Contenido<sup>1</sup> (SRIBC). El objetivo genérico es diseñar sistemas de recuperación eficientes, para bases de datos de imágenes, capaces de ofrecer información relevante ante unos requerimientos de información visual definidos por el usuario. En el mundo actual es bien sabido que el valor real de una información no depende sólo de ella misma sino también de la facilidad con la cual se pueda recuperar cuando se necesite.

En los SRIBC se pueden distinguir dos grandes generaciones: una primera basada casi directamente en los sistemas clásicos de recuperación de texto combinados con un trabajo previo de anotación manual y/o semiautomática y una segunda, cuyo desarrollo comienza en los primeros años de la década de los 90, basada en la extracción automática de propiedades perceptuales sobre la imagen.

En la primera generación de este tipo de sistemas, la descripción de las imágenes se realizaba mediante texto. Estos sumarios de texto, que describían el contenido de la imagen, eran realizados por expertos en la materia de la imagen que indicaban qué estaba representado en la imagen y cuál era su significado. Otra posibilidad era el uso de palabras clave, extraídas manual y/o automáticamente, que hicieran referencia a objetos y/o conceptos que aparecieran en las imágenes. La recuperación efectiva de la información estaba basada en los sistemas de recuperación de texto tradicionales sobre el dominio de los sumarios construidos por los expertos que analizaban las imágenes. Ejemplos típicos de sistemas de recuperación de imagen de primera generación son los buscadores iniciales de imágenes de buscadores Web de referencia como Google o Yahoo<sup>2</sup>. La idea general que utilizan estos sistemas es que cuando el sistema de búsqueda indexa una página Web mediante análisis del texto, y su correspondiente extracción de palabras clave, se pueden asociar estas palabras clave a las imágenes presentes en dicha página web. La hipótesis, que parece razonable ya que la imagen dispone de un contexto, es que la información contenida en las imágenes estará relacionada con la información contenida en el texto y, por tanto, las palabras clave obtenidas se pueden utilizar para indexar el texto y las imágenes de esa página Web.

Esta primera generación de SRIBC tiene importantes problemas que la convierten en una opción no válida para una recuperación de imagen eficiente. El primer problema que aparece está relacionado con la necesidad de contar con un

<sup>1</sup> En inglés, Content-Based Image Retrieval Systems (CBIRS ó sólo CBIR).

<sup>2</sup> [images.google.es](http://images.google.es) y [es.images.search.yahoo.com](http://es.images.search.yahoo.com)

operador humano que describa cada imagen (en el caso de la Web serían los propios creadores de cada página). Si el número de imágenes es pequeño puede ser una tarea abordable pero si es grande la tarea puede convertirse en irrealizable. Tampoco se puede solucionar el problema pensando en la introducción de un conjunto de expertos pues en ese caso se podría tener un problema de consistencia interpretativa. Lo que es importante para un experto en una imagen puede ser irrelevante para otro. De esta forma, la descripción de la misma imagen por diferentes expertos puede producir resultados de recuperación muy diferentes en la fase de consulta. Un ejemplo de este comportamiento es el popular sitio Web Flickr<sup>3</sup> donde los usuarios etiquetan sus propias imágenes sin ningún criterio común. En segundo lugar, resulta muy difícil, por no decir imposible, capturar todos los elementos perceptuales (color, textura, forma,...) presentes en las imágenes con una simple descripción de texto. Como bien dice el refrán, “una imagen vale más que mil palabras”. Por último, el texto no resulta adecuado para modelar la similaridad perceptual relativa de dos imágenes.

Debido a los problemas de los SRIBC de primera generación, en la década de los 90 comienza a desarrollarse una nueva generación. En estos nuevos sistemas, el proceso de recuperación se realiza usando información visual obtenida directamente de las imágenes. Para obtener esta información visual se utilizan técnicas provenientes del procesamiento digital de imágenes, combinadas o no con las técnicas utilizadas por los SRIBC de primera generación. Recientemente, Google ha incorporado este tipo de técnicas en su buscador de imágenes. Cuando se visualizan los resultados iniciales del buscador de imágenes, aparece un enlace debajo de las imágenes con el texto “Similares” que permite acceder a una nueva búsqueda de resultados perceptualmente similares al seleccionado. Este capítulo se centrará en estos SRIBC de segunda generación. El desarrollo de estos sistemas requiere de la contribución de muchas disciplinas científicas: Procesado y Análisis de Imagen, Reconocimiento de Patrones, Visión por Ordenador, Modelado de Datos Multidimensionales, Indexación y Visualización de Datos, etc.

## **2 SISTEMAS DE RECUPERACIÓN DE IMAGEN BASADOS EN CONTENIDO (SRIBC)**

En este apartado se relacionan, brevemente, los principales problemas y soluciones que es necesario considerar en el diseño de un SRIBC. En primer lugar, se presentará la arquitectura general de estos sistemas. Una vez desarrollada la estructura general, se comentan someramente los principales bloques constituyentes. Los dos bloques fundamentales, y distintivos en cuanto a su funcionamiento en este tipo de sistemas respecto a sistemas clásicos de texto, son la representación del contenido y la medición de la similaridad.

<sup>3</sup> [www.flickr.com](http://www.flickr.com)

La arquitectura de los SRIBC (Figura 1) sigue un esquema bastante similar a la de los sistemas convencionales clásicos de recuperación de texto. En este esquema se pueden identificar los siguientes subsistemas:

- ✕ Herramienta gráfica de introducción de consultas. Este subsistema debe permitir la construcción de las consultas sin la necesidad de recurrir a expresar mediante texto las necesidades de información. Algunas opciones utilizadas son la navegación por la base de datos y la introducción externa de imágenes para utilizarlas como ejemplo de consulta y la utilización de aplicaciones de generación de plantillas (*layouts*).
- ✕ Extracción de las propiedades perceptuales sobre los datos visuales. Se usan, fundamentalmente, técnicas de procesamiento digital de imagen y de reconocimiento de patrones para detectar las características visuales más destacadas y calcular alguna medida, descriptor, que las represente.
- ✕ Extracción de las propiedades semánticas sobre los datos visuales. Este tipo de subsistema no siempre existe y, en muchos casos, la semántica se extrae en base a una combinación de las propiedades perceptuales a través de un conjunto de reglas.
- ✕ Anotación manual de metadatos descriptivos. En algunos casos, la extracción automática de propiedades perceptuales de la imagen se acompaña de una etapa de anotación descriptiva en texto del contenido y/o significado de dichas imágenes. Esto sólo es posible cuando se trata de colecciones pequeñas, usualmente con una fuerte especialización temática. Por ejemplo, una base de datos de imágenes de casos médicos de una especialidad dada en la cual los metadatos pueden ser extraídos de las historias clínicas que contienen las imágenes.
- ✕ Una estructura de almacenamiento. Cuando las bases de datos aumentan en tamaño se hace prácticamente imposible realizar comparaciones exhaustivas de la imagen consulta con todas las imágenes almacenadas en la base de datos. Es necesario el empleo de índices que permitan el acceso rápido a las imágenes candidatas a ser relevantes ante una consulta dada.
- ✕ Motor de recuperación. Es el bloque principal del sistema y ejecuta los procedimientos implementados para obtener las imágenes más relevantes frente a las necesidades de información definidas por el usuario. Obtener imágenes que un usuario juzgue como similares y ordenarlas en un ranking de relevancia requiere definir métricas de similaridad que reflejen matemáticamente la percepción humana de la similaridad.

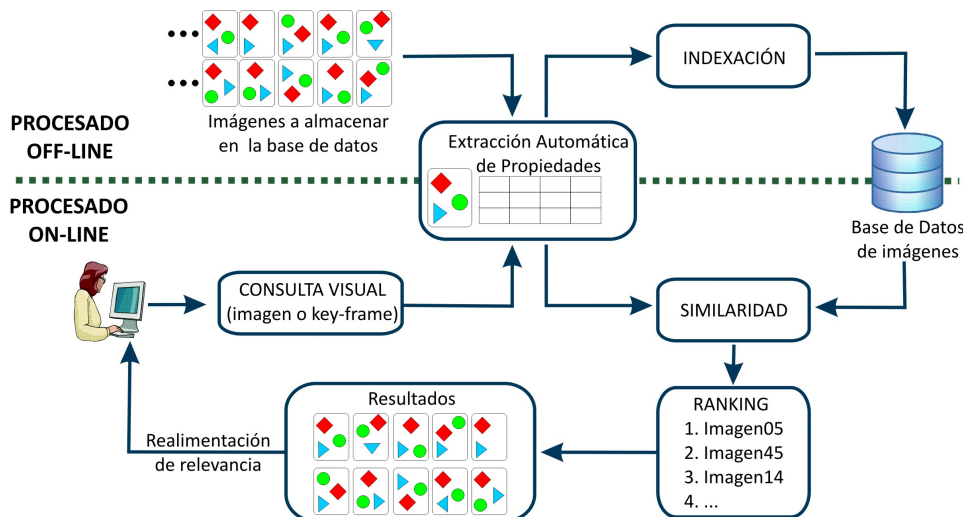


Figura 1: Arquitectura típica de un SRIBC.

- ✗ Herramienta gráfica de visualización de resultados. El usuario que interactúa con el sistema de recuperación de información debe poder inspeccionar los resultados suministrados por el motor de recuperación. Dado que la respuesta de los SRIBC suele contener un gran número de resultados es esencial dotar a los sistemas de herramientas que permitan una fácil navegación por ellos.
- ✗ Mecanismo de realimentación (*feedback*) relevante. En algunos sistemas, el usuario puede introducir realimentación indicando, sobre los resultados obtenidos en una primera consulta, cuales tienen relevancia positiva y/o cuales relevancia negativa. Este tipo de mecanismos, que suponen el establecimiento de una sinergia entre el usuario y el ordenador, permiten refinar los resultados de una consulta de forma progresiva.

## 2.1 Consulta basada en Contenido

En los SRIBC de segunda generación, la introducción de las necesidades de información se debe realizar de forma perceptual, sin recurrir a descripciones mediante texto. El objetivo de un SRIBC es determinar qué imágenes, de la base de datos, son similares a la imagen presentada, o descrita gráficamente, en la consulta. En la práctica, existen tres tipos diferentes de consultas: consulta exacta, que consiste en localizar copias prácticamente iguales a la imagen consulta; consulta de rango, que consiste en establecer un rango en el cual se considere que las imágenes de la base de datos son similares a la imagen consulta; y consulta de primeros k-vecinos, que consiste en obtener las k imágenes más similares a la imagen consulta planteada. Este tipo de consultas se denominan en términos generales consultas

mediante ejemplos visuales y la forma de plantearlas pueden clasificarse en cuatro grandes categorías (Figura 2):

- ✗ Consulta de iconos: El sistema muestra al usuario un área de dibujo dónde este puede situar iconos que representan conceptos. Como resultado final se obtiene un modelo (*layout*) de lo que se va a buscar en la base de datos.
- ✗ Consulta de pintado: El usuario colorea zonas del área de dibujo con colores sólidos. Como resultado final se obtiene una versión esquemática de lo que se va a buscar en la base de datos. La efectividad de esta forma de introducir la consulta tiene una fuerte dependencia con las cualidades de diseño del usuario.
- ✗ Consulta de boceto: El usuario define sobre el área de dibujo las formas de los objetos presentes. Como resultado final se obtiene un esquema de las formas que deben tener las imágenes a buscar en la base de datos. También en este caso, las cualidades de diseño del usuario determinan la efectividad de esta forma de introducir la consulta.
- ✗ Consulta de imagen ejemplo: El usuario introduce en el sistema una imagen con el objetivo de localizar en la base de datos imágenes que sean similares.

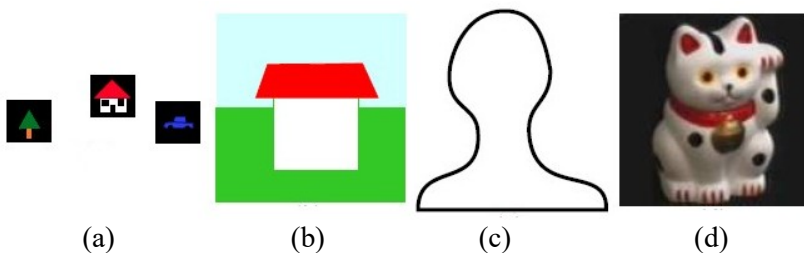


Figura 2: Ejemplos de consulta basada en contenido: (a) Consulta de iconos. (b) Consulta de pintado. (c) Consulta de boceto. (d) Consulta de imagen ejemplo.

## 2.2 Representación del Contenido

Cuando un observador ve una imagen, percibe en su cerebro parámetros psicofísicos como pueden ser el contraste, los bordes, las formas, el color, la textura, la situación espacial, etc. El cerebro humano procesa esos parámetros para tomar la decisión perceptual que corresponda. Dado que los SRIBC tienen como objetivo principal proporcionar resultados similares a la experiencia de percepción humana, deben partir del manejo de esas propiedades perceptuales. Y para poder

manejar esas propiedades en un ordenador es necesario cuantificarlas bajo algún esquema de representación.

La representación de las propiedades perceptuales de las imágenes (como pueden ser el color, la textura, la forma, la estructura, las relaciones espaciales, el movimiento, ...) es el problema fundamental de la recuperación de imagen. Dado que esta representación es la pieza clave de los SRIBC, se dedicará un apartado especial a ella tras introducir los conceptos necesarios de procesado digital de imágenes y de reconocimiento de patrones.

## 2.3 Organización y Acceso

En los SRIBC la información extraída de las imágenes se almacena, habitualmente, en una base de datos. Cuando el tamaño de la información a almacenar (las descripciones cuantitativas de las propiedades perceptuales) es considerable, se hace necesaria la definición de estructuras de índices que eviten un análisis secuencial de los datos. Estas estructuras de índices, idealmente, filtran todas las representaciones irrelevantes mediante la utilización de la descripción de la imagen consulta suministrada por el usuario. Como resultado proporcionan las imágenes relevantes para una consulta sin tener que analizar toda la base de datos completa. Esto permite tener tiempos de respuesta al usuario aceptables.

Entre los sistemas de organización de información de los SRIBC están los ficheros de rejilla, los árboles K-d, los árboles R, R+ y R\*. En general, son adaptaciones específicas a datos multidimensionales de estructuras de datos típicas.

## 2.4 Medición de la Similitud

Una vez que el subsistema de representación del contenido visual ha obtenido la información relevante de las imágenes de la base de datos, y que esta se encuentra almacenada en una estructura de datos adecuada, es necesario comparar estas descripciones con la descripción de la consulta. La base de la obtención de un buen ranking de resultados es la capacidad de medir de forma cuantitativa la similitud entre dos imágenes y que esta medida de similitud tenga algún tipo de correlación con la diferencia perceptual, o con la diferencia semántica, percibida por el usuario entre las dos imágenes.

Una forma muy común de modelar el concepto de percepción humana de la similitud es asimilarla a una distancia métrica en un espacio multidimensional de propiedades (en procesado de imagen, este espacio de propiedades se denomina habitualmente espacio perceptual o espacio psicológico). De esta forma, si un estímulo (imagen) está representado por un punto en un espacio métrico de propiedades concreto de dimensión  $M$ , la similitud entre dos estímulos ( $p_1$  y  $p_2$ )

se puede modelar por una función de distancia que cumpla los cuatro axiomas fundamentales:

- ✗ Autosimilaridad:  $d(p1, p1) = d(p2, p2) = 0$
- ✗ Minimalidad:  $d(p1, p2) \geq d(p1, p1)$
- ✗ Simetría:  $d(p1, p2) = d(p2, p1)$
- ✗ Desigualdad triangular:  $d(p1, p2) \leq d(p1, p3) + d(p3, p2)$

donde  $p3$  es un tercer estímulo. Las funciones de distancia usadas responden, frecuentemente, a distancias tipo Minkowski dadas por la expresión:

$$d_p(p1, p2) = \left( \sum_{i=1}^M (p1_i - p2_i)^p \right)^{1/p}$$

y a distancias cuadráticas ponderadas dadas por la expresión:

$$d_Q(p1, p2) = \sqrt{\sum_{i=1}^M \sum_{j=1}^M (p1_i - p2_j) Q_{ij} (p1_i - p2_j)}$$

donde  $Q_{ij}$  es una matriz definida positiva de pesos entre las  $M$  componentes de los estímulos.

Los casos particulares que se suelen emplear son: la distancia Euclídea (Minkowski con  $p=2$ ), la distancia Manhattan o city-block (Minkowski con  $p=1$ ), la distancia Chebychev o tablero de ajedrez (Minkowski con  $p=\infty$ ) y la distancia Mahalanobis (cuadrática siendo  $Q$  la inversa de la matriz de covarianza entre los estímulos  $p1$  y  $p2$ ).

Este tipo de métricas tiene ventajas que la han convertido en habitual en los sistemas de recuperación de imágenes, aunque estudios sobre psicología han puesto de manifiesto ciertas carencias del modelo métrico a la hora de caracterizar el concepto humano de similaridad. La especificación de la métrica entre imágenes, junto con la representación del contenido visual, conforman los aspectos más significativos de los SRIBC y, por tanto, se comentarán posteriormente con más detalle. Previamente es necesario un repaso a los conceptos básicos del Procesado Digital de Imágenes para poder aplicarlos posteriormente en la creación de descriptores sobre los cuales calcular la medida de similaridad.

### 3 PROCESADO DIGITAL DE IMAGEN

El aspecto más claramente distinguible de los SRIBC de segunda generación es que están basados en la capacidad para representar y cuantificar el contenido visual de las imágenes. Los descriptores necesarios deben manejar



propiedades perceptuales como el color, la textura, la forma, la estructura, las relaciones espaciales, el movimiento, etc. Todas ellas se sitúan en el ámbito de lo que se denomina Procesado Digital de Imágenes. Para poder desarrollar SRIBC es necesario contar con importantes conocimientos en este campo. En este apartado se presentan los aspectos más básicos del Procesado Digital de Imágenes.

### 3.1 Imagen Digital

La luz, desde un punto de vista físico, según el diccionario de la Real Academia Española, es una radiación electromagnética en el espectro visible que supone un agente físico que hace visibles los objetos. Pero desde el punto de vista de los SRIBC es mejor trabajar con un concepto perceptual del proceso de formación de las imágenes. Se puede recurrir a modelos basados en la percepción de los colores por parte del cerebro en los cuales los responsables de las diferencias de color que perciben los humanos son los descriptores perceptuales de la luz:

- ✕ **Brillo.** Dadas dos fuentes de luz con la misma distribución espectral, aquella que tenga mayor intensidad física parece perceptualmente más brillante o clara.



- ✕ **Tono.** El tono es el color puro que define la luz. La longitud de onda, o período espacial de la señal, no es una buena medida del color porque algunos colores no están contenidos en el arco iris que produce un prisma al descomponer luz blanca (por ejemplo, morado = rojo + azul).



- ✕ **Saturación.** La saturación es la responsable de distinguir entre un tono apagado (grisáceo) y un tono claro bien definido.



Una imagen  $I(x,y)$ , en el ámbito del Procesado Digital de Imágenes (denominada habitualmente imagen digital), se puede definir formalmente como un valor particular de la distribución de probabilidad conjunta de un conjunto de  $N \times M$

variables aleatorias ordenadas espacialmente sobre una rejilla rectangular de NxM posiciones. Cada una de esas variables aleatorias representa el valor de la propiedad física que forma la imagen en la posición espacial de dicha variable aleatoria en la rejilla de NxM posiciones. Si la única propiedad que se tiene en cuenta es el brillo, la imagen resultante será una imagen de niveles de gris. Si se tienen en cuenta las tres propiedades perceptuales, o cualquier otra combinación que represente al color existente, la imagen resultante será una imagen en color. Si la propiedad física que se mide no es visual (temperatura, presión, distancia, etc.) la imagen resultante será una representación de la distribución espacial de esa propiedad. El acceso al valor de cada variable aleatoria se realiza a partir de su posición espacial (x,y) en la rejilla de NxM posiciones.

Es habitual referirse a cada una de estas variables aleatorias como píxeles de la imagen, es decir, como elementos de imagen. Así, cuando se dice que una imagen tiene una resolución (formalmente resolución espacial) de 1024x768 píxeles se está diciendo que tiene 786.432 variables aleatorias distribuidas en una rejilla rectangular de 1024x768 posiciones. Los valores que toman las variables aleatorias o píxeles que forman la imagen digital están cuantizados (formalmente resolución radiométrica) entre un valor mínimo que suele ser 0 y un valor máximo que depende del número de bits que se empleen en el ordenador para almacenar dicho valor (1 bit → 2 valores; 2 bits → 4 valores, ..., 4 bits → 16 valores, ..., 8



22,223,249	83,162,192	255,109,39	252,101,0	253,104,2	253,117,0	40	50	69	196	201	217
30,230,255	79,159,184	254,110,37	255,100,0	253,111,1	253,119,0	43	54	71	197	204	215
33,229,253	79,155,179	254,110,37	255,100,0	253,111,1	255,114,8	39	55	60	197	204	215
36,228,251	82,152,177	254,110,37	255,100,0	255,105,6	251,117,0	41	58	50	197	204	215

Figura 3: Imagen Digital con una resolución de 800x300 píxeles y una cuantización de 8 bits (8 bits por componente en la zona en color). En la parte superior se muestra la imagen representando el valor de cada variable aleatoria mediante el color o nivel de gris que define. En la parte inferior se muestran los datos correspondientes a las regiones rectangulares marcadas.

bits  $\rightarrow$  256 valores, 10 bits  $\rightarrow$  1024 valores, ...). En el caso de imágenes en niveles de gris, las variables aleatorias almacenan un único dato y sus valores, habitualmente, varían entre 0 [negro] y 255 [blanco] (número de valores representables con 8 bits = 1 byte de información). En el caso de una imagen en color existirán tres componentes asociadas a cada variable aleatoria. Si los valores de cada una de ellas varían entre 0 [negro] y 255 [máximo de rojo/verde/azul, máximo de luminancia/tono/saturación, etc.] existirán 16.777.216 colores diferentes representables ( $256 \times 256 \times 256$ ).

### 3.2 Color

El color es el efecto principal que se percibe cuando una onda electromagnética incide sobre el sistema de visión humano. Muchos de los modelos desarrollados para explicar la visión en humanos se basan en estudios de la fisiología del ojo como parte fundamental, y más conocida que el cerebro, de la percepción visual humana. No son los modelos más útiles desde el punto de vista de los SRIBC pero es necesario estudiarlos para ver sus debilidades y, por tanto, la necesidad de introducir modelos perceptuales (más complejos matemáticamente).

En 1802, Thomas Young formula el modelo clásico de la visión humana en color. La hipótesis básica es que la retina posee tres tipos de fotorreceptores responsables de la visión del color (los conos) en los cuales los picos de absorción ocurren en diferentes longitudes de onda del espectro. Para un humano promedio, los picos de sensibilidad de esos tres tipos de conos se sitúan en los 430nm, los 560nm y los 610nm. Este posicionamiento espectral ha llevado a que se denominen cono azul, cono verde y cono rojo. En su trayecto al cerebro, estas tres señales básicas se entremezclan proporcionando una señal con la información de luminancia y dos señales con información de crominancia (Figura 4).

En el Procesado Digital de Imágenes, los modelos de color se introducen para dotar a esta propiedad de una representación numérica, necesaria para

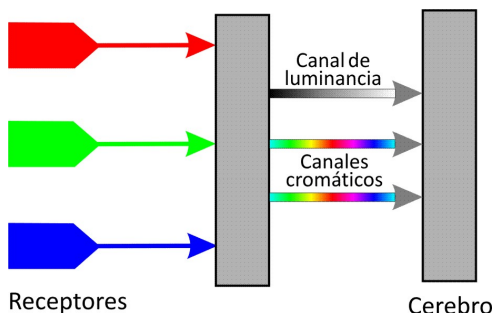


Figura 4: Modelo clásico de la visión humana en color de Thomas Young (Las tres señales recibidas por el ojo se interpretan en el cerebro como una señal de luminancia más dos de crominancia).

posteriores tratamientos. Dado que existen tres tipos de conos, es razonable suponer que los estímulos de color se pueden representar como puntos de un espacio tridimensional (valores de excitación de los conos azul, verde y rojo, respectivamente). En 1931, la Commission Internationale de l'Eclairage (CIE<sup>4</sup>) definió un modelo estándar que permite representar todos los colores mediante síntesis aditiva de tres fuentes primarias. El modelo se denomina CIE 1931<sup>5</sup> (conocido habitualmente como CIE XYZ) y las fuentes primarias propuestas se sitúan en 435.8nm, 546.1nm y 600.0nm (valores basados en los descubrimientos sobre los picos de sensibilidad promedio de los conos de los humanos).

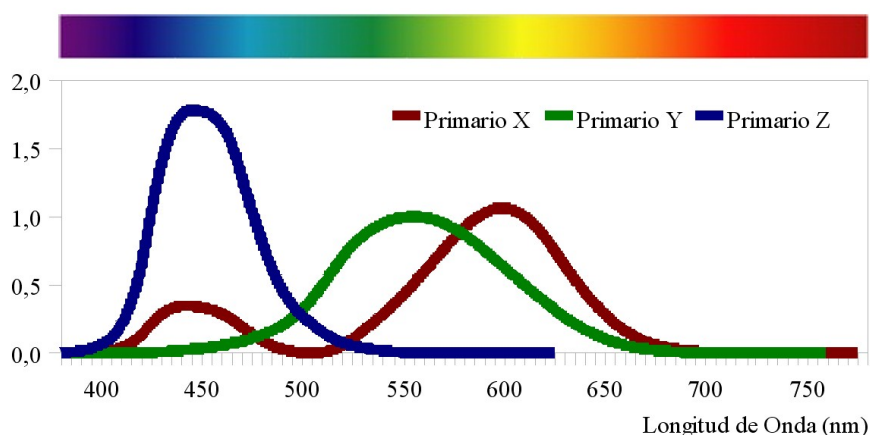


Figura 5: Fuentes primarias del modelo CIE 1931 XYZ.

La introducción del modelo CIE XYZ abrió la puerta a un grupo de modelos denominados de forma genérica modelos de color orientados al hardware (el nombre proviene de su nacimiento como respuesta a la necesidad de mostrar color en dispositivos hardware, fundamentalmente, televisiones y monitores de ordenador). Tienen como característica común que están inspirados en la fisiología de la retina humana y, por tanto, se basan en la definición de unos colores primarios, obteniendo la gama de colores posibles mediante síntesis aditiva de dichos colores primarios. El ejemplo más representativo de este grupo de modelos es el modelo RGB (*Red/Green/Blue*) relacionado con el modelo CIE XYZ mediante la expresión:

$$X = 0,489 R + 0,311 G + 2,201 B$$

$$Y = 0,176 R + 0,813 G + 0,011 B$$

$$Z = 0,000 R + 0,010 G + 0,990 B$$

Se trata de un modelo aditivo que tiene como primarios los colores rojo (R), verde (G) y azul (B). Se emplea en dispositivos de visualización: monitores, televisiones,

<sup>4</sup> <http://www.cie.co.at/>

<sup>5</sup> [http://www.cie.co.at/index.php/Publications/index.php?i\\_ca\\_id=483](http://www.cie.co.at/index.php/Publications/index.php?i_ca_id=483)

etc. Controlando la cantidad de rojo, de verde y de azul que se presenta en posiciones cercanas se puede hacer que el ojo humano perciba un único color mediante integración espacial de las tres fuentes primarias. Otros modelos de color pertenecientes a este grupo son el CMY (modelo sustractivo muy utilizado en dispositivos de impresión) y el YIQ (estándar de las emisiones de televisión). En general, los modelos pertenecientes a este grupo se caracterizan porque sus primarios se pueden obtener a partir de una combinación lineal de los definidos en el modelo CIE XYZ.

Como se había comentado al principio de este apartado, desde el punto de vista de los SRIBC son más interesantes los modelos basados en conceptos perceptuales del sistema de visión humano. Este grupo de modelos, basados en la percepción de los colores, se denomina modelos de color orientados al usuario. Dentro de ellos cabe establecer una subclasificación: los modelos no uniformes y los modelos uniformes. Esta clasificación se establece en términos de la uniformidad de la sensación de distancia de color percibida por un observador humano. Si es igual en todo el espacio de color, es decir, si la diferencia de color percibida por un observador humano se puede aproximar siempre por la distancia métrica entre dos puntos del espacio de color se habla de espacios uniformes. Si no es posible, se habla de espacios no uniformes.

Los modelos de color orientados al usuario no uniformes tienen como característica común que separan la luminancia del resto de componentes cromáticos (típicamente relacionados con los descriptores perceptuales tono y saturación). Sus componentes derivan de combinaciones, lineales o no, del modelo RGB y/o del modelo CIE XYZ. Un ejemplo típico de este tipo de modelos es el modelo HLS (tono, intensidad, saturación) dado por la expresión:

$$H = \arctan \left[ \frac{3^{1/3}(G-B)}{(R-G)+(R-B)} \right]$$

$$L = R + G + B$$

$$S = 1 - \left[ \frac{\min(R, G, B)}{R + g + B} \right]$$

Los modelos de color orientados al usuario uniformes, también denominados esquemas de organización perceptual de colores, han sido estudiados desde hace muchos años. En 1905, Albert Munsell realizó uno de los primeros intentos de organizar perceptualmente los colores. Los ejemplos más representativos de este tipo de modelos son el  $L^*u^*v^*$  y el  $L^*a^*b^*$  dado por la expresión:

$$L^* = 116 \left[ \frac{Y}{Y_0} \right]^{\frac{1}{3}} - 16 \quad \text{para} \quad \frac{Y}{Y_0} > 0,008856$$

$$L^* = 903,3 \left[ \frac{Y}{Y_0} \right] \quad \text{para} \quad \frac{Y}{Y_0} \leq 0,008856$$

$$a^* = 500 \left[ f\left(\frac{X}{X_0}\right) - f\left(\frac{Y}{Y_0}\right) \right]$$

$$b^* = 200 \left[ f\left(\frac{Y}{Y_0}\right) - f\left(\frac{Z}{Z_0}\right) \right]$$

donde:

$$f(t) = \begin{cases} t^{\frac{1}{3}} & \text{para } t > 0,008856 \\ 7,787t + \frac{16}{116} & \text{para } t \leq 0,008856 \end{cases}$$

y  $X_0$ ,  $Y_0$  y  $Z_0$  son los valores de  $X$ ,  $Y$  y  $Z$  del modelo CIE XYZ para la referencia del blanco. El  $L^*u^*v^*$  es el modelo recomendado por la CIE para cuantificar diferencias de color en monitores en condiciones de fuentes de iluminación aditivas. El  $L^*a^*b^*$  es el modelo recomendado por la CIE<sup>6</sup> en 1976 para cuantificar diferencias de color en condiciones de fuentes de iluminación próximas a la luz solar. Este último, muy empleado en sistemas de recuperación de imagen, se basa en los estudios sobre la interpretación del color, iniciados por Ewald Hering, que

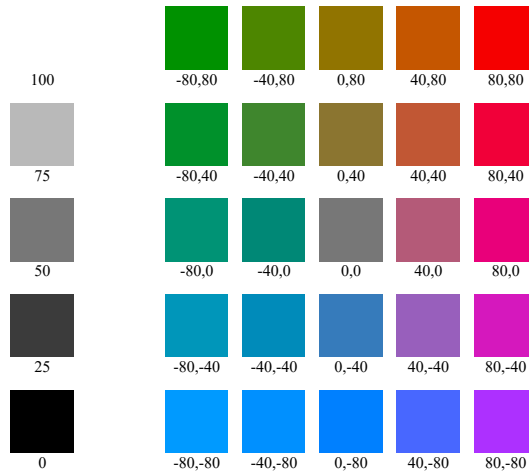


Figura 6: Modelo CIE  $L^*a^*b^*$ . Izquierda: Eje de luminancia  $L^*$  para  $a^*=b^*=0$ . Derecha: Plano cromático  $a^*, b^*$  para  $L^*=50$ .

<sup>6</sup> [http://www.cie.co.at/index.php/Publications/index.php?i\\_ca\\_id=485](http://www.cie.co.at/index.php/Publications/index.php?i_ca_id=485)

sugieren que existen tres tipos de componentes: una componente que mide la luminancia en una escala de gris entre el blanco y el negro, una componente que mide el color en una escala entre el rojo y el verde y una componente que mide el color en una escala entre el amarillo y el azul. Las ecuaciones que relacionan este modelo con el modelo CIE XYZ son complejas porque representan una aproximación de un espacio de Riemann a un espacio Euclídeo para conseguir la propiedad de la uniformidad perceptual (distancias cuantitativas iguales deben corresponder a distancias perceptuales iguales para un humano).

Dado que los SRIBC tratan de reproducir el concepto humano del parecido, es aconsejable desarrollarlos empleando modelos de color orientados al usuario uniformes.

### 3.3 Textura

Las imágenes de objetos reales ofrecen, a menudo, eventos no uniformes en su nivel de gris o en su color, pero que claramente generan regiones diferenciadas del resto del entorno presente en la imagen. Por ejemplo, cuando se observa un paisaje compuesto por árboles, cualquier observador humano reconoce la existencia de troncos, copas, etc. Obviamente, la imagen obtenida de la superficie del tronco de un árbol no es uniforme en niveles de gris o color pero debido a sus continuas variaciones formando un patrón repetitivo puede llegar a

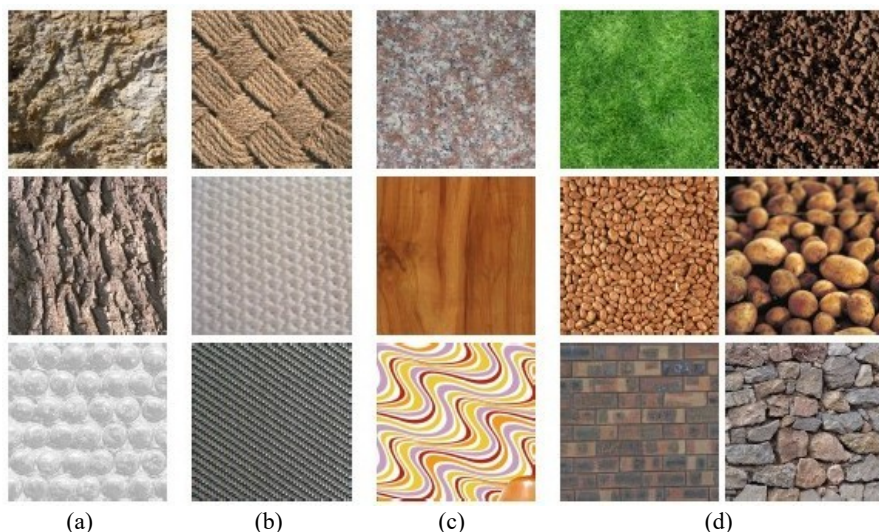


Figura 7: Ejemplos de texturas reales. (a) Generadas por rugosidad de superficies. (b) Generadas por organización estructural del material. (c) Generadas por variación de iluminación. (d) Generadas por agregación de objetos individuales.

considerarse como un evento único caracterizado por alguna propiedad común. Esta propiedad se denomina textura. Es relativamente fácil reconocer una textura cuando se la ve (por ejemplo, una pared de ladrillo, una acera con baldosas, una calle empedrada, una valla, la arena de una playa, un bosque, etc.), sin embargo es difícil definirla cuantitativamente con precisión. Lo que sí se puede afirmar con rotundidad es que la textura no es una propiedad asociada a cada píxel individual de la imagen, como era el color, sino que es una propiedad asociada a una zona de la imagen. Por ello, en el análisis de textura es muy habitual hablar del concepto de relación de un píxel con sus píxeles vecinos (en realidad, de las propiedades de un píxel con las propiedades de sus píxeles vecinos).

Estudios sobre la visión humana, como los realizados por Pollack y Julesz, han puesto en evidencia que el cerebro humano realiza un análisis de frecuencias de aparición de motivos en las imágenes que percibe. Los métodos de análisis de texturas desarrollan dos grandes vías que explotan desde diferentes perspectivas el concepto de frecuencia: métodos estadísticos y métodos estructurales. En las aproximaciones estadísticas se caracteriza la textura a partir de medidas sobre aspectos cualitativos tales como uniformidad, tosquedad, granulado, etc. En las aproximaciones estructurales se trabaja con la regularidad de la distribución de motivos. Para ello, en los métodos que desarrollan las aproximaciones estadísticas se estudia la frecuencia de aparición de niveles de gris o colores de la imagen y en los métodos que desarrollan las aproximaciones estructurales se estudia la frecuencia de aparición de estructuras organizadas (texeles) sobre la imagen, más complejas que el nivel de gris o color de un simple píxel.

### 3.4 Forma

Las personas identifican los objetos presentes en una escena por la distribución de sus bordes, siempre y cuando puedan distinguirlos del fondo sin demasiado esfuerzo como ponen de manifiesto las ilusiones ópticas. Así pues, el concepto básico para trabajar con la forma será el concepto de borde. Al igual que la textura, el borde es un concepto asociado a una región de la imagen.

En un píxel de una imagen existirá un borde si hay una diferencia significativa entre el valor de alguna propiedad (usualmente, color o nivel de gris) y los valores de esa propiedad para píxeles vecinos. Esta medición de la diferencia de propiedad se realiza mediante el empleo de derivadas, especialmente el cálculo del gradiente. Dado que las imágenes digitales son discretas, las derivadas se aproximan por diferencias finitas cuyo valor se calcula mediante una medida de la diferencia entre el valor de la propiedad del píxel que se está analizando y el valor de la propiedad en píxeles situados en el entorno próximo (típicamente  $n=1$ ) a través de las expresiones:



$$\begin{aligned}\nabla_x I(x, y) &= I(x, y) - I(x - n, y) \\ \nabla_y I(x, y) &= I(x, y) - I(x, y - n)\end{aligned}$$

Un gradiente tiene dos componentes: su magnitud y su dirección. La magnitud del gradiente en un punto proporcionará información sobre la fuerza de borde que existe en ese punto y la dirección proporcionará información sobre la dirección del borde en ese punto.

Sin embargo, un objeto es algo más que una serie de puntos de borde distribuidos sobre una imagen. En procesamiento digital de imagen se habla de contorno cuando se tiene una serie de puntos bordes (píxeles con valor alto en la propiedad de borde, es decir, con valores alto de primera derivada) unidos espacialmente y se habla de objeto, o región, cuando ese contorno es cerrado.

### 3.5 Descriptores

Dado que los SRIBC basan su trabajo en la comparación entre las propiedades perceptuales de las imágenes, es necesario disponer de herramientas capaces de cuantificar dichas propiedades. Estas herramientas reciben el nombre genérico de descriptores de imagen. El proceso de recuperación en los SRIBC se realiza a partir de la comparación de las representaciones de las propiedades perceptuales realizadas mediante descriptores de imagen.

El histograma es la forma más sencilla de representar en un solo dato global (un vector) información sobre los valores de alguna propiedad que toman los diferentes píxeles de una imagen. Se trata de una estimación discreta de la distribución de probabilidad de la propiedad en la imagen asumiendo independencia estadística entre los valores de la propiedad en los píxeles de dicha imagen. Esta asunción es usualmente incorrecta pues los valores de una propiedad de un píxel suelen estar muy correlacionados con los valores de esa propiedad en píxeles vecinos, pero es una solución estadística sencilla que se ha mostrado suficientemente válida en la mayor parte de los problemas de procesamiento de imagen. Se pueden obtener histogramas de los niveles de gris, del color, de la textura, de la magnitud de los puntos borde, de la dirección de los puntos borde, etc., aplicando la misma estrategia a los valores obtenidos para las diferentes propiedades.

Su construcción es muy sencilla. Simplemente, se calcula el porcentaje de píxeles que tienen un determinado valor o un rango de valores, frente al número total de píxeles que tiene la imagen completa mediante la expresión:

$$H(i) = \frac{\# I(x, y) = i}{\#(x, y)}$$

donde # es un operador de conteo. La concatenación de diferentes rangos  $i$  de propiedad analizados, denominados bins, proporciona los valores necesarios para la construcción del histograma.

En el caso particular de la forma, es posible trabajar no sólo con estadísticas de valores de los puntos borde sino que se pueden definir descriptores más complejos si se pueden obtener los contornos cerrados que delimitan los objetos presentes en la imagen. Se abre así todo un abanico de posibilidades desde sencillos descriptores geométricos (como el área de la región abarcada, la compacidad del contorno, la elongación, etc.) hasta complejos procesos de comparación de forma basados en el cálculo de la energía necesaria para transformar un contorno concreto en otro contorno dado. También se pueden utilizar formulaciones matemáticas más estrictas como son el cálculo de momentos invariantes o los coeficientes de la transformada de Fourier (representación del contorno en el dominio de la frecuencia).

## 4 RECUPERACIÓN DE IMAGEN

La idea fundamental de la segunda generación de los SRIBC es eliminar la necesidad de descripciones textuales del contenido de las imágenes. En su lugar, se usa un conjunto de descriptores que permitirán al usuario la obtención de imágenes visualmente similares a la presentada como consulta. En los siguientes apartados se repasan algunas de las propiedades usadas en los SRIBC y la forma en la que se puede calcular la similaridad entre dos imágenes en base a dichas propiedades.

### 4.1 Extracción y Representación de Propiedades

Para la extracción y representación de las propiedades perceptuales que manejan los SRIBC se utilizan diferentes tipos de descriptores sobre dichas propiedades perceptuales. En los siguientes subapartados se describen algunas de las representaciones más utilizadas.

#### 4.1.1 COLOR

La representación más usada del color de una imagen (también aplicable a la representación de los niveles de gris) es la utilización de histogramas para describir la distribución de probabilidad de los diferentes colores que aparecen en la imagen caracterizados cuantitativamente mediante algún modelo de color (típicamente  $L^*a^*b^*$ ). La principal decisión en este caso es elegir el número de colores diferentes que tendrá el histograma. Una representación completa de todos los colores es inviable (y, además, se ha probado que no aporta nada a la obtención de resultados de recuperación) pues, por ejemplo, si se considera una

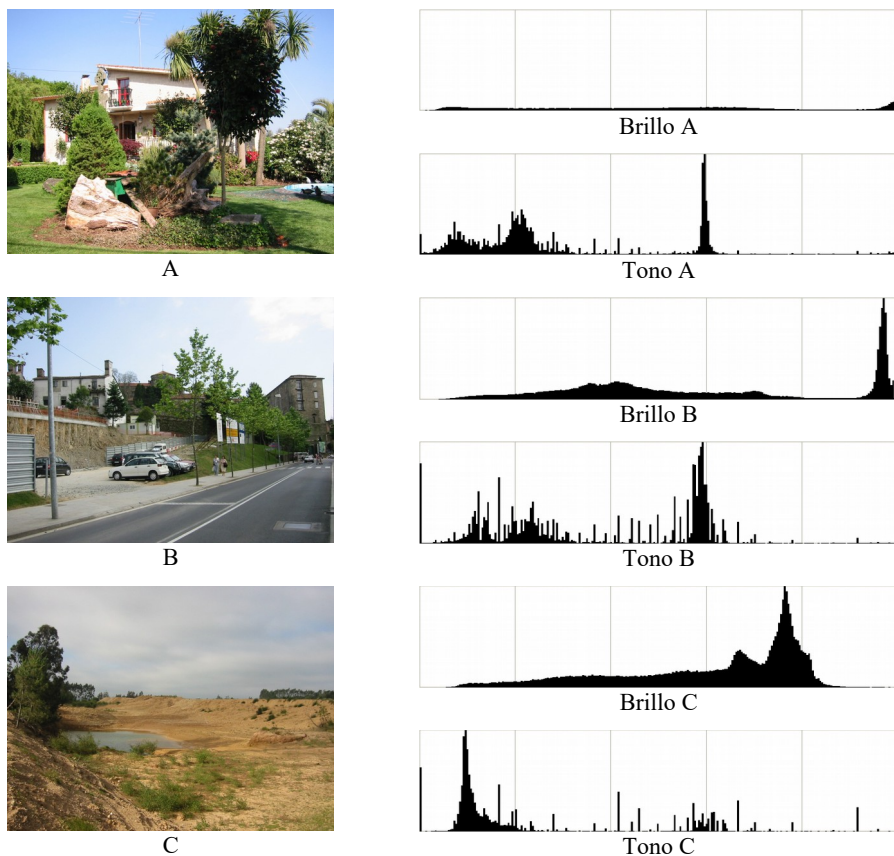


Figura 8: Ejemplos de imágenes y sus correspondientes histogramas de 256 bins para el brillo y el tono de sus píxeles. Como se puede apreciar en los diferentes histogramas de brillo, las imágenes A y B presentan más luminosidad que la imagen C (más puntos en la parte derecha del histograma). En cuanto a los histogramas de tono o color, se puede ver cómo informan sobre los colores predominantes. Las imágenes A y B tienen un pico muy destacado en la zona de los azules correspondiente al cielo (pico de máximo de histograma). Todas las imágenes tienen una contribución significativa en la zona de los verdes (parte izquierda del histograma), aunque la A más que la B y la B más que la C. El pico más significativo de la imagen C, así como una parte importante de la imagen A, están en la zona de los naranjas y los ocres. En general, se puede comprobar cómo imágenes diferentes presentan histogramas significativamente diferentes.

representación de cada color en 8 bits (256 valores diferentes) la representación completa del espacio de color tendría 16.777.216 colores diferentes.

Usualmente, se emplean histogramas que van desde los 16 a los 256 valores diferentes (denominados bins). Para obtener la región del espacio de color que representa cada bin es necesario cuantizar dicho espacio mediante técnicas de clusterización, o bien hacerlo de forma heurística. La principal ventaja del empleo de histogramas es que su cálculo es trivial. Una vez definida la región de influencia en el espacio de color asociada a un bin del histograma, se obtiene su valor contando el número de píxeles que tienen un color situado en el interior de esa región y dividiéndolo entre el número total de píxeles de la imagen.

El histograma es insensible a rotaciones y cambios de resolución de la imagen, aunque es sensible a cambios en la iluminación. En algunos casos se puede obtener un histograma independiente de la iluminación trabajando exclusivamente con las dos componentes cromáticas de los modelos de color orientados al usuario.

Otra forma sencilla de describir el contenido de color de una imagen es emplear conjuntos de colores. Se trata de una representación simplificada del histograma de colores en la cual el valor de los bins que lo forman no se calculan como un porcentaje del total sino que son binarios (valor 0 ó valor 1). Los diferentes bins toman valor 1 si existe una presencia significativa del color que representan (por ejemplo, más de un cierto porcentaje del total de la imagen). Los colores que se incluyen en el conjunto se toman de un diccionario de nombres asociados a los colores básicos. La aplicación de estos colores básicos no debe estar restringida a clases de objetos muy específicas (por ejemplo, color oliva no sería un color válido). El significado de estos colores básicos no debe ser predecible a partir del significado de una parte de un objeto (por ejemplo, color de las hojas no sería un color válido) y no puede estar incluido en otro color (por ejemplo, color verdoso no sería un color válido). En resumen, deben destacar psicológicamente. En un trabajo clásico basado en los modelos de Munsell, Berlin y Kay proponen el siguiente conjunto: Blanco, Negro, Gris, Rojo, Verde, Azul, Amarillo, Marrón, Rosa, Morado y Naranja.

Hasta ahora siempre se ha supuesto que se estaba trabajando con descriptores globales del color, es decir, descriptores sobre el color de toda la imagen. Otra posibilidad interesante es realizar descripciones locales del color sobre regiones de la imagen. Para obtener estas regiones se puede recurrir a particiones uniformes o a particiones adaptativas (segmentación) de la imagen. Una vez separada la imagen en varias regiones se aplican los descriptores a cada región de la misma forma que se hacía para la imagen global.

### 4.1.2 TEXTURA

La textura se pueden caracterizar por la dependencia espacial que aparece en la distribución del color, o de los niveles de gris, de la imagen. Por tanto, el uso de propiedades estadísticas sobre ella es, quizás, el método más simple para analizarlas. Las matrices de coocurrencia dadas por la expresión:

$$P_{xy}(i_1, i_2) = \frac{\#[(x_1, y_1), (x_2, y_2)]}{\#I} \quad \begin{array}{ll} |x_1 - x_2| = x & |y_1 - y_2| = y \\ I(x_1, y_1) = i_1 & I(x_2, y_2) = i_2 \end{array}$$

donde # es un operador de conteo, son un ejemplo típico de estos métodos. Son histogramas de segundo orden en los cuales se caracteriza la textura mediante la dependencia espacial ( $|x_1 - x_2| = x$ ;  $|y_1 - y_2| = y$ ) que aparece en la distribución de los colores o niveles de gris de la imagen. Formalmente se trata de un conjunto de matrices en las que cada elemento representa la probabilidad de encontrar dos colores ( $i_1$  e  $i_2$ ), o dos niveles de gris, en una posición espacial relativa,  $xy$ , dada. Existirán tantas matrices como posiciones espaciales relativas se consideren. Finalmente, la textura se caracteriza midiendo una serie de características sobre cada una de las matrices de coocurrencia: energía (Eg), entropía (En), máximo contraste (MC), contraste (Ct), correlación (Co), clusterización (Cl), etc. Estas características se calculan mediante las expresiones:

$$\begin{aligned} Eg &= \sum_{i_1, i_2} P_{xy}^2(i_1, i_2) & En &= \sum_{i_1, i_2} P_{xy}(i_1, i_2) \log P_{xy}(i_1, i_2) \\ MC &= \max P_{xy}(i_1, i_2) & Ct &= \sum_{i_1, i_2} (i_1 - i_2)^k P_{xy}^l(i_1, i_2) \\ Co &= \sum_{i_1, i_2} i_1 i_2 P_{xy}^2(i_1, i_2) & Cl &= \sum_{i_1, i_2} (i_1 + i_2 - 2\mu)^k P_{xy}(i_1, i_2) \end{aligned}$$

donde  $\mu$  es la media de los colores o niveles de gris  $i$ .

Otro método con sustento similar es el uso de funciones de autocorrelación, es decir, la medición de la correlación de una imagen consigo misma desplazada. Si el tamaño del patrón de textura es grande, la función de autocorrelación variará lentamente frente a la distancia. Sin embargo, si el tamaño del patrón de textura es pequeño, la función de autocorrelación variará rápidamente frente a la distancia.

Otra posibilidad, dentro de las soluciones estadísticas, es el empleo de métodos basados en modelos estadísticos. Estos métodos de análisis de texturas buscan procesos matemáticos subyacentes en las imágenes de texturas que no sólo sean capaces de describirlas sino que además puedan emplearse para sintetizarlas artificialmente. Estos métodos basados en modelos capturan las cualidades esenciales que pueden percibirse en una textura. Un ejemplo típico de este tipo de aproximaciones son los modelos de campo aleatorio de Markov dados por expresiones como:

$$I(x, y) = \sum_{(i, j) \in N} h(i, j) I(x-i, y-j) + u(x, y)$$

donde  $u(x, y)$  es una función de ruido y  $h(i, j)$  es una matriz de coeficientes que caracteriza la dependencia espacial entre los niveles de gris. Estos modelos asumen que la textura está producida por un modelo probabilístico estocástico estacionario espacialmente que satisface la asunción de independencia condicional. De esta forma, el nivel de gris o color de un píxel,  $I(x, y)$ , depende de los niveles de gris o colores de píxeles situados en una vecindad próxima,  $N$ , a través de un conjunto de coeficientes  $h(i, j)$ . El ajuste de este conjunto de coeficientes dará lugar a los diferentes tipos de textura. Otros ejemplos de métodos basados en modelos estadísticos son los modelos autorregresivos y los fractales. En los modelos autorregresivos, la textura se caracteriza mediante los coeficientes de un modelo que relaciona el color o nivel de gris de un píxel con el color o nivel de gris de píxeles situados en una ventana centrada en el píxel bajo estudio (con factores de corrección dados por la media de color o nivel de gris de la imagen y una componente de ruido modelada como un ruido gaussiano de media nula y una determinada varianza). El coeficiente de Hurst, utilizado para medir la dimensión fractal de una imagen, se define como la pendiente de la mayor diferencia de color o nivel de gris en píxeles situados a lo largo de una línea en escala logarítmica frente a la distancia.

En el campo de las aproximaciones estructurales, los métodos más sencillos son los que emplean modelos geométricos o estructuras. Se basan, esencialmente, en interpretar la imagen de texturas como una estructura compuesta por elementos básicos denominados primitivas y en calcular las propiedades geométricas de las relaciones entre esas primitivas. Un ejemplo de este tipo de métodos es el empleo de los mosaicos de Voronoi que representan la partición del espacio en zonas de vecindad. La existencia de primitivas similares en zonas de vecindad contiguas es un indicio de la presencia de una textura. Otros ejemplos de métodos de este tipo son los mosaicos aleatorios (se parte el espacio en celdas con una relación espacial), los métodos basados en estructuras (buscan el posicionamiento de determinadas primitivas tipo) y el empleo de morfología matemática (para localizar el posicionamiento de primitivas de interés).

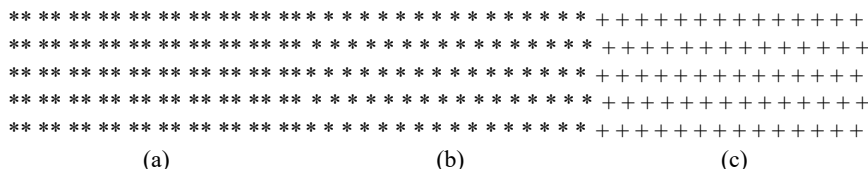


Figura 9: Ejemplos de texturas construidas artificialmente mediante repetición periódica de primitivas básicas. (a) y (b) Misma primitiva con diferentes relaciones espaciales. (b) y (c) Diferentes primitivas con igual relación espacial.

Como se comentaba en el apartado 3.3, el estudio de la visión humana ha puesto de manifiesto que el cerebro realiza un análisis de las frecuencias de la imagen. La textura es una propiedad cuyas características la hacen especialmente susceptible de ser analizada de esta forma. Los métodos basados en técnicas de procesamiento de señal utilizan diversas técnicas de este tipo para analizar la textura. En particular, muchos de los métodos utilizan algún tipo de filtro. El método más característico de este grupo son los filtros de Gabor dados por la siguiente expresión:

$$G_{x_0, y_0, f_0, \Theta_0}(x, y) = \exp \left\{ i \left[ 2\pi f_0 (x \cos \Theta_0 + y \sin \Theta_0) + \Phi \right] \right\} \text{gauss}(x - x_0, y - y_0)$$

donde  $f$  y  $\Theta$  son la frecuencia y la orientación del filtro, el subíndice 0 indica los valores de referencia y gauss una distribución gaussiana. Se trata de filtros sintonizados en frecuencia y orientación que permiten caracterizar la frecuencia de aparición de las primitivas. Otros ejemplos de métodos de este grupo son filtros en el dominio espacial (se aplican plantillas sobre las imágenes para que destaque la presencia de determinadas situaciones), transformadas digitales (con diferentes kernels orientados a destacar características concretas) y cálculos de energía y momentos (operadores matemáticos que extraen determinadas características de las imágenes).

#### 4.1.3 FORMA

Las descripciones de forma se pueden clasificar en dos grandes tipos: descriptores de forma global y descriptores de forma local. Los descriptores de forma global sólo necesitan los valores de borde (magnitud y dirección) de los píxeles que forman la imagen. Los descriptores de forma local asumen conocido el contorno de los objetos presentes en la imagen a describir. Esta última hipótesis es muy restrictiva ya que disponer de curvas cerradas para cada objeto de la imagen implica que se puedan separar automáticamente todos los objetos presentes en la imagen del fondo.

Una manera sencilla de describir la forma global es realizar un histograma de orientaciones a partir de la dirección de los valores de borde de los píxeles de la imagen. En este caso, el número de bins del histograma es una cuantización de las posibles direcciones en el espacio. Es típico utilizar entre 4 y 16 direcciones diferentes.

Otra manera de describir la forma global es utilizar la magnitud de los valores de borde de los píxeles para controlar la evolución de una curva cerrada que localice formas presentes en la imagen. Se trata de una técnica conocida con el nombre de contornos activos (*snakes*). El proceso de deformación elástica de un contorno activo corresponde a la minimización de un funcional de energía

compuesto por dos términos: el primero, relacionado con la forma del propio contorno activo, que controla las deformaciones admisibles y el segundo, relacionado con su posición respecto a los valores altos de borde presentes en la imagen, que controla que el contorno activo se sitúe sobre posiciones significativas. En realidad, se trata de un método híbrido entre forma global y forma local pues comienza utilizando los valores de borde de todos los píxeles de la imagen (forma global) para terminar localizando el contorno de los objetos presentes en la imagen (forma local).

Para la descripción de la forma local existen bastantes alternativas. Las más simples están basadas en la geometría: calcular el área encerrada por la curva, su compacidad, su elongación, la elipse inscrita, etc. Se trata, en general de descriptores muy simples pero muy eficaces en determinados contextos. También se puede optar por calcular el baricentro del objeto y caracterizar los puntos que forman el contorno mediante sus coordenadas polares respecto a dicho baricentro. En este caso, el descriptor se podría construir como una función radio frente a ángulo para los puntos del contorno. Dentro de estas aproximaciones sencillas también se puede encuadrar la caracterización mediante códigos de cadena. Se subdivide la curva en segmentos de una longitud fija y se caracteriza cada segmento mediante su orientación.

Otra posibilidad de descripción de la forma local es utilizar herramientas con base matemática. En general, estas herramientas proporcionan una serie de coeficientes que permiten elegir el nivel de detalle con el que se quiere trabajar. Los primeros coeficientes de la serie proporcionan una información general de la forma del contorno. Conforme se considera un número mayor de coeficientes de la serie, se aumenta el número de detalles de la forma del contorno que se incorporan a la descripción. Entre las herramientas más usadas de este tipo destacan el cálculo de momentos invariantes dados por la expresión:

$$M_{ij} = \sum_{x,y} I(x, y) x^i y^j$$

donde  $ij$  indica el orden del momento y la utilización de la transformada de Fourier dada por la expresión:

$$F(m, n) = \sum_{x=0}^N \sum_{y=0}^M f(x, y) e^{-2\pi i \left( \frac{mx}{N} + \frac{ny}{M} \right)}$$

donde  $(x,y)$  son coordenadas en el dominio espacial y  $(m,n)$  son coordenadas en el dominio de la frecuencia. Los momentos invariantes son, por su definición, invariantes a traslaciones, rotaciones y escalados del objeto. Son muy utilizados en la descripción de formas sencillas. Si las formas son muy complejas, es necesario utilizar muchos coeficientes de la serie lo cual supone problemas en el tiempo de computación necesario para su cálculo. La transformada de Fourier es muy



conocida pues proporciona información sobre el dominio de la frecuencia de una señal. Su uso no resulta sencillo pues no existe una relación inmediata entre las componentes de frecuencia que proporciona y su significado en las variaciones en la forma del objeto que describe.

#### 4.1.4 RELACIONES ESPACIALES

La descripción escueta del color, de la textura y/o de la forma de las regiones presentes en una imagen ofrecen una información descriptiva de bajo nivel del contenido de una imagen. Por ejemplo, si en una imagen se detecta una zona azul sin textura y una zona verde con una textura homogénea, esta información de color y textura describe sintácticamente la imagen. Sin embargo, las posiciones espaciales relativas de ambas zonas pueden tener información semántica sobre el contenido. Si la zona azul está debajo de la zona verde la imagen podría ser, por ejemplo, un paisaje formado por un lago y un bosque. Si la zona azul está encima de la zona verde la imagen podría estar formada, por ejemplo, por el bosque con un cielo azul. Pero si ambas zonas están una al lado de otra es difícil que sea una de las dos suposiciones previas (podría serlo, pero parece que este esquema vertical recuerda menos esa sensación). Así pues, la relación espacial entre las diferentes zonas de la imagen es también una información interesante.

Una forma sencilla de especificar las relaciones espaciales que aparecen en una imagen es segmentar la imagen en sus regiones constituyentes y definir descriptores que midan las posiciones relativas de dichas regiones constituyentes. Esta tarea no es sencilla pues la segmentación es, sin duda, el proceso más complejo del procesamiento digital de imágenes y es prácticamente imposible que contemple todos los casos salvo que se trabaje en un dominio reducido y muy controlado. Un descriptor simple es, por ejemplo, las cadenas 2D que son una representación simbólica de la disposición de los objetos. Se usan exclusivamente dos cadenas, una que representa las posiciones en la horizontal y otra que representa las posiciones en la vertical, en las cuales el orden de cada símbolo de cada región se obtiene proyectando la posición media de cada región sobre los ejes cartesianos.

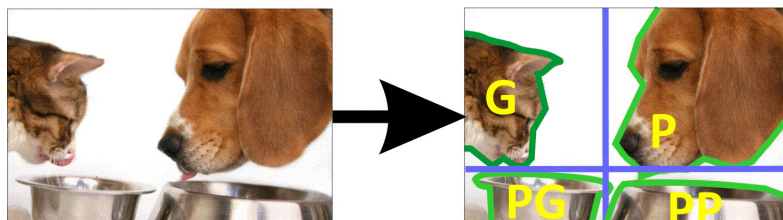


Figura 10: Representación de una imagen mediante cadenas 2D. En horizontal  $G, PG < P, PP$  y en vertical  $PG, PP < G, P$ .

Ejemplos de descriptores de mayor complejidad pueden ser los grafos de adyacencia y el cálculo de los walkthrough.

En los grafos de adyacencia, los nodos son las regiones resultantes del proceso de segmentación de la imagen y los arcos entre nodos toman valores binarios. Estos arcos toman valor 1 si las regiones que representan los nodos que conecta son adyacentes entre sí y toman valor 0 si no se cumple dicho criterio de adyacencia.

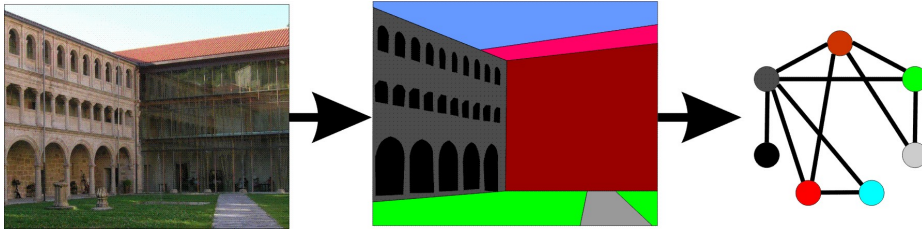


Figura 11: Descripción de una imagen mediante su grafo de adyacencias.

El walkthrough de dos regiones A y B es una clase de equivalencia sobre el camino entre cada punto de la región A y cada punto de la región B. Para calcular el camino sólo se tienen en cuenta dos tipos de desplazamientos, el horizontal y el vertical. Y para cada tipo de desplazamiento sólo se tienen en cuenta tres tipos de situaciones, desplazamiento positivo, desplazamiento cero y desplazamiento negativo (según el eje de coordenadas correspondiente en cada caso). La combinación de los dos tipos de desplazamiento y de las tres situaciones posibles en cada uno proporciona las nueve componentes de la matriz del walkthrough de un par de regiones A y B. Cada componente de la matriz dará cuenta del número de parejas de puntos, uno perteneciente a la región A y otro perteneciente a la región B, que tienen una pareja de situaciones concretas en los dos tipos de desplazamiento posibles.

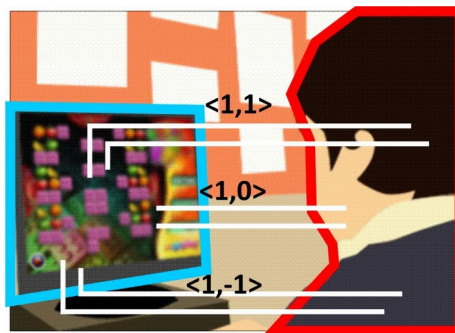


Figura 12: Descripción de una imagen mediante Walkthrough.

## 4.2 Medidas de Similitud

Una vez que se dispone de los descriptores de las propiedades perceptuales, una de las tareas más importantes de un SRIBC es definir una medida de similitud entre las representaciones obtenidas de las diferentes imágenes. Esto permitirá comparar la descripción de la consulta (valores de los descriptores utilizados calculados sobre la imagen pregunta) con la descripción de los elementos almacenados en la base de datos (valores de los descriptores utilizados calculados sobre cada una de las imágenes almacenadas en la base de datos). Esta medida de similitud permitirá la construcción del ranking de resultados que se le presentará al usuario. En los siguientes subapartados se describen algunas de las medidas de similitud más utilizadas sobre los descriptores comentados en el apartado anterior.

### 4.2.1 COLOR

Si el color está descrito mediante histogramas, se pueden emplear medidas de distancia tipo Minkowski sobre histograma dadas por la expresión:

$$d_h(I_q, I_d) = \left[ \sum_{i=1}^n (H(I_q, i) - H(I_d, i))^p \right]^{1/p}$$

donde  $H(I_q, i)$  y  $H(I_d, i)$  son respectivamente el bin  $i$ -ésimo de los histogramas de  $n$  bins de una imagen consulta  $I_q$  y de una imagen de la base de datos  $I_d$ . De esta forma se puede medir la similitud entre el histograma proveniente de la imagen consulta y los histogramas provenientes de cada una de las imágenes de la base de datos. Estas técnicas acumulan las diferencias medibles entre bins equivalentes de ambos histogramas. También es posible utilizar una forma cuadrática de distancia introduciendo una matriz de coeficientes de similitud  $a_{ij}$  entre los diferentes bins del histograma mediante la expresión:

$$d_h(I_q, I_d) = \sqrt{\sum_{i=1}^M \sum_{j=1}^M (H(I_q, i) - H(I_d, j))^t a_{ij} (H(I_q, i) - H(I_d, j))}$$

El ranking de resultados se obtiene ordenando estos valores de distancia de histograma obtenidos.

Un ejemplo puede ilustrar las posibilidades y limitaciones de esta aproximación. Dada la base de datos de imágenes mostrada a continuación, para calcular la similitud entre una imagen seleccionada como pregunta y el resto de imágenes de la base de datos es necesario calcular el histograma de tono o color de cada imagen (Figura 8) y calcular la distancia entre cada par de histogramas (de la imagen pregunta con cada una de las otras imágenes) mediante, por ejemplo, distancias tipo Minkowski.



1



2



3



4



5



6



7



8



9



10



11



12



13



14



15



16

Suponer que se eligen las imágenes 1, 5, 9 y 13 como imágenes pregunta de cuatro consultas de recuperación. Y suponer que para calcular la medida de similaridad en cada consulta, entre la imagen pregunta y cada imagen de la base de datos, se usa una distancia tipo Minkowski con  $p=1$  sobre histogramas de 32 bins de cada imagen de la base de datos dada por:

$$d_h(I_q, I_d) = \left[ \sum_{i=1}^{32} |H(I_q, i) - H(I_d, i)| \right]$$

Los resultados de distancia obtenidos en cada consulta son los siguientes:

Consulta Im1	$d_h(I_1, I_7)=40,5$	$d_h(I_1, I_{14})=109,7$	$d_h(I_5, I_3)=138,8$
$d_h(I_1, I_1)=0$	$d_h(I_1, I_8)=62,6$	$d_h(I_1, I_{15})=116,3$	$d_h(I_5, I_4)=115,8$
$d_h(I_1, I_2)=41,4$	$d_h(I_1, I_9)=123,9$	$d_h(I_1, I_{16})=104,7$	$d_h(I_5, I_5)=0$
$d_h(I_1, I_3)=51,5$	$d_h(I_1, I_{10})=120,2$		$d_h(I_5, I_6)=101,0$
$d_h(I_1, I_4)=54,1$	$d_h(I_1, I_{11})=135,5$	Consulta Im5	$d_h(I_5, I_7)=101,0$
$d_h(I_1, I_5)=107,9$	$d_h(I_1, I_{12})=104,3$	$d_h(I_5, I_1)=107,9$	$d_h(I_5, I_8)=95,7$
$d_h(I_1, I_6)=104,9$	$d_h(I_1, I_{13})=98,3$	$d_h(I_5, I_2)=120,7$	$d_h(I_5, I_9)=132,8$

$d_h(I_5, I_{10})=134,1$	$d_h(I_9, I_3)=125,2$	$d_h(I_9, I_{14})=124,2$	$d_h(I_{13}, I_7)=88,2$
$d_h(I_5, I_{11})=133,6$	$d_h(I_9, I_4)=142,8$	$d_h(I_9, I_{15})=108,4$	$d_h(I_{13}, I_8)=91,7$
$d_h(I_5, I_{12})=119,4$	$d_h(I_9, I_5)=132,8$	$d_h(I_9, I_{16})=155,2$	$d_h(I_{13}, I_9)=116,8$
$d_h(I_5, I_{13})=52,2$	$d_h(I_9, I_6)=145,8$		$d_h(I_{13}, I_{10})=122,9$
$d_h(I_5, I_{14})=96,8$	$d_h(I_9, I_7)=102,4$	Consulta Im13	$d_h(I_{13}, I_{11})=120,8$
$d_h(I_5, I_{15})=95,8$	$d_h(I_9, I_8)=106,9$	$d_h(I_{13}, I_1)=98,3$	$d_h(I_{13}, I_{12})=101,6$
$d_h(I_5, I_{16})=128,0$	$d_h(I_9, I_9)=0$	$d_h(I_{13}, I_2)=111,7$	$d_h(I_{13}, I_{13})=0$
	$d_h(I_9, I_{10})=16,3$	$d_h(I_{13}, I_3)=123,0$	$d_h(I_{13}, I_{14})=62,4$
Consulta Im9	$d_h(I_9, I_{11})=17,4$	$d_h(I_{13}, I_4)=107,2$	$d_h(I_{13}, I_{15})=59,4$
$d_h(I_9, I_1)=123,9$	$d_h(I_9, I_{12})=26,0$	$d_h(I_{13}, I_5)=52,2$	$d_h(I_{13}, I_{16})=119,6$
$d_h(I_9, I_2)=110,3$	$d_h(I_9, I_{13})=116,8$	$d_h(I_{13}, I_6)=79,0$	

Así, para la imagen 1 las imágenes más similares de la base de datos son:



Puede parecer sorprendente que la imagen 7 resulte más similar a la imagen 1 que las imágenes 2, 3 y 4. Hay que recordar que la comparación no es semántica sino sólo en función de los colores presentes. El azul del cielo es muy parecido en ambas imágenes, la imagen 7 tiene algo de verde y la imagen 1 tiene algo de gris en la casa que aparece detrás del motivo principal. A partir de la imagen 8, las distancias son bastante considerables y por tanto su similaridad es muy pequeña.

Para la imagen 5, las imágenes más similares de la base de datos son:



En este caso, en realidad, la única imagen que tiene una similaridad aceptable respecto a la imagen 5 es la imagen 13. La imagen 8 tiene una distancia de similaridad a la imagen 5 que es casi el doble que la imagen 13. De nuevo es necesario considerar que la recuperación no es semántica sino exclusivamente por colores. No es lo mismo un cielo azul que un cielo grisáceo.

Para la imagen 9, las imágenes más similares de la base de datos son:



El grupo de imágenes 9, 10, 11 y 12 son muy similares entre ellas y muy diferentes del resto de imágenes de la base de datos. La siguiente en la lista de imágenes similares, la imagen 7, tiene una diferencia que cuadriplica las diferencias de similaridad de las imágenes 10, 11 y 12. El hecho diferencial es que este grupo de

30 RECUPERACIÓN DE INFORMACIÓN. UN ENFOQUE PRÁCTICO Y MULTIDISCIPLINAR © RA-MA  
 imágenes no tienen cielo, ni azul ni grisáceo. Sólo el verde de la hierba y el gris de las piedras.

Y para la imagen 13, las imágenes más similares de la base de datos son:



Si la imagen 13 era la más similar a la imagen 5, no debe extrañar que la imagen 5 sea la más similar a la imagen 13 (aunque no siempre tiene porqué ser así; véase donde está la imagen 8 en ambas consultas). Lo que más puede sorprender es la ausencia de la imagen 16 (de hecho, sólo tres imágenes son menos similares y por muy poco). Aunque semánticamente pueda resultar muy parecida, a nivel de colores es muy diferente pues se trata de tierra labrada bajo un cielo azul (y la imagen 13 es arena bajo un cielo grisáceo).

Si, por contra, el color está descrito mediante conjuntos de colores, se puede calcular el ranking mediante un producto cartesiano de los vectores binarios de colores mediante la expresión:

$$d = \sqrt{(c_q - c_d)^t a_{ij} (c_q - c_d)}$$

donde  $a_{ij}$  es una matriz de coeficientes de similaridad entre los colores empleados para dar más peso al producto de colores parecidos. Resulta un cálculo más sencillo que las métricas de distancia de los histogramas pero, a cambio, se obtiene un ranking menos preciso.

## 4.2.2 TEXTURA

Casi todas las técnicas de descripción de texturas tienen, al final, un aspecto en común. Este aspecto común es que permiten obtener un conjunto, usualmente bastante grande, de  $N$  valores que describen la textura. Cuando se trabaja con texturas, lo usual es interpretar este conjunto de valores como puntos de un espacio métrico  $N$ -dimensional. Así, se pueden utilizar las funciones típicas de distancia tipo Minkowski sobre ese espacio  $N$ -dimensional. Esos valores de distancia servirán para la creación del ranking de resultados.

## 4.2.3 FORMA

Existen evidencias experimentales que indican que la percepción humana de la similaridad entre formas no sigue completamente los axiomas de las métricas. A pesar de ello, la mayoría de los sistemas de recuperación usan estos modelos de similaridad.

Si la forma se ha descrito mediante histogramas de orientaciones, se puede calcular la similaridad entre imágenes mediante cualquiera de las medidas de



distancia de histogramas disponibles. Estas distancias de histogramas permiten construir directamente el ranking de resultados.

Si la forma se describe mediante contornos activos, los valores del funcional de energía se pueden usar para calcular directamente la similaridad entre objetos presentes en la imagen. También se puede utilizar, para calcular el ranking, la energía necesaria para transformar el contorno de un objeto (objeto pregunta) en el contorno de otro objeto (objetos en la base de datos).

Si la forma se ha descrito mediante medidas geométricas sencillas, se pueden utilizar las diferencias de sus valores para evaluar la similaridad. En este caso hay que realizar un estudio para decidir la importancia relativa que tienen entre sí las diferentes medidas disponibles. En el caso de emplear coordenadas polares respecto al baricentro, se puede estimar la similaridad de la forma de dos objetos a partir de la distancia entre sus respectivas curvas. En este caso es necesario realizar ajustes de normalización para que las comparaciones sean independientes a rotaciones y a escalados (a traslación serán siempre invariantes pues la medida es relativa al baricentro del objeto). Para formas aproximadas mediante curvas poligonales, como es el caso de los códigos de cadena, la similitud se puede estimar midiendo el número de cambios necesarios para transformar una forma en otra.

Los métodos que emplean representaciones de forma basadas en propiedades, como los momentos invariantes o los coeficientes de la transformada de Fourier tratan la forma como vectores de propiedades y calculan la similaridad como una métrica en el espacio de propiedades. Esta aproximación supone que todos los coeficientes de la serie tienen la misma importancia en la descripción de la forma, situación que no tiene por qué ser necesariamente cierta pues los primeros coeficientes capturan los aspectos generales y añadir coeficientes supone añadir detalles cada vez más finos y, por tanto, menos importantes.

#### 4.2.4 RELACIONES ESPACIALES

Cuando se trata de modelar la similaridad para descripciones de alto nivel que intenten aproximar la semántica de la imagen, es imposible establecer un patrón común de estrategia, pues cada descripción lleva aparejada medidas de similaridad específicas e interpretaciones de su aporte a la información semántica diferentes.

Por ejemplo, para el caso de una descripción mediante cadenas 2D, la similaridad se establece comparando, obviamente, las cadenas de símbolos que representan a las imágenes. Sin embargo, esta comparación no puede ser estricta pues se estarían desechando situaciones similares en las que faltara alguna región (y, además, no se puede garantizar una segmentación perfecta de la imagen). En

general, se considera que dos imágenes son similares si la cadena que define a una de ellas está incluida en la cadena que define a la otra.

Si se emplea el modelado de las relaciones espaciales mediante grafos de adyacencia, una medida usual de similaridad suele ser la comparación de dichos grafos mediante la estimación de su isomorfismo. En el caso del cálculo de *walkthrough*, la similaridad se obtiene comparando directamente los nueve coeficientes que forman las correspondientes matrices.

## 5 OTRAS DIMENSIONES

La imagen digital bidimensional no es la única información multimedia sobre la cual se puede implantar un sistema de recuperación de información. También es posible realizarlo sobre imágenes tridimensionales, sobre vídeos e incluso sobre imágenes tridimensionales en movimiento.

Una imagen tridimensional se puede definir como un valor particular de la distribución de probabilidad conjunta de un conjunto de  $N_1 \times N_2 \times N_3$  variables aleatorias ordenadas espacialmente sobre una rejilla rectangular de  $N_1 \times N_2 \times N_3$  posiciones. En este caso el concepto de pixel, bidimensional, se reemplaza por el concepto de voxel, tridimensional. El procesamiento de estas imágenes tridimensionales es similar al de las imágenes bidimensionales extendiendo las dependencias espaciales a la tercera dimensión.

Un vídeo puede considerarse como una sucesión de imágenes. Cada una de estas imágenes constituyentes del vídeo recibe el nombre de *frame*. Una forma sencilla de construir un sistema de recuperación de vídeos es utilizar un sistema de recuperación de imágenes sobre los *key-frames*, es decir, sobre los frames representativos de las diferentes secuencias del vídeo.

Ambas extensiones se pueden combinar para diseñar sistemas de recuperación de información sobre imágenes tridimensionales en movimiento.

## 6 EL DOMINIO DE LA IMAGEN Y EL GAP SEMÁNTICO

En la reflexión sobre la colección de imágenes que gestiona un SRIBC, denominada usualmente el dominio imagen, hay que hacer una distinción entre los casos de dominio limitado y dominio amplio. En los dominios limitados existe una variabilidad predecible en las características relevantes que aparecen en las imágenes (condiciones similares de captura como en el caso de las radiografías o las huellas dactilares). En los dominios amplios, la variabilidad es ilimitada e impredecible, incluso para un mismo significado semántico. Nunca se puede



asegurar que el algoritmo de extracción no se encontrará con contenido desconocido. En este caso, la semántica de las imágenes sólo puede ser descrita de forma parcial (por ejemplo, archivos de gran tamaño de fotografías o las imágenes contenidas en la Web).

En el caso de dominios amplios, el gap entre los descriptores de propiedades y la interpretación semántica es, en general, importante. Por contra, para dominios limitados, el gap entre sus características y su interpretación semántica suele ser más reducido porque las características se diseñan expresamente para resaltar elementos semánticos muy concretos.

En los dominios amplios aparece también un gap sensorial. La información que se puede extraer de una imagen depende directamente de cómo se captura (por ejemplo, si la escena no está bien iluminada pueden, o no, ser visibles determinadas características). En los dominios limitados no suele aparecer este problema porque la captura de las imágenes suele estar muy controlada.

Otro problema mostrado por los SRIBC, especialmente los primeros, es que no se tomó en consideración la existencia de un gap semántico. Esto es, una falta de coincidencia entre la información que se puede extraer de los datos visuales y la interpretación de esos datos por un usuario concreto en una situación dada. En los sistemas más recientes se ha tratado de reducir este gap cambiando la forma de interacción con el sistema. A la hora de sugerir interpretaciones, se examina la base de datos de imágenes y se establecen relaciones entre las imágenes. De estas relaciones surge el posible significado de los resultados proporcionados al usuario.

## 7 EJEMPLOS DE SRIBC

Existe una buena cantidad de ejemplos de implementación de SRIBC de segunda generación. No hay más que acceder a cualquier buscador web e introducir consultas como “image retrieval”, “content-based image retrieval”, “search for image”, etc. Algunos son publicaciones en revistas especializadas y otros demos en páginas web. A continuación se presentan algunos ejemplos, aunque no debe tomarse como una lista exhaustiva. No son más que una fotografía parcial del estado de la cuestión tomada a comienzos del mes de abril de 2010.

- ✕ SIMBA (Search Images by Appearance)<sup>7</sup>. Aproximación basada en propiedades invariantes frente a traslaciones y rotaciones. Utiliza un histograma de propiedades similar al histograma de color pero calculado de forma local sobre la vecindad de cada píxel. Además, incorpora información sobre textura.

<sup>7</sup> [simba.informatik.uni-freiburg.de/](http://simba.informatik.uni-freiburg.de/)

- x CIREs (Content Based Image Retrieval System)<sup>8</sup>. Combina información de alto y bajo nivel. Para el análisis de información de alto nivel emplea organización perceptual y técnicas de inferencia y agrupamiento para extraer información semántica que describa el contenido estructural de las imágenes. En el análisis de información a bajo nivel emplea histograma de color y filtros de Gabor con diferentes canales espaciofrecuenciales para describir el contenido de textura.
- x SIMPLIcity (Semantics-sensitive Integrated Matching for Picture Libraries)<sup>9</sup>. Realiza la comparación de similaridad sobre regiones resultantes de un proceso de segmentación. De esta forma, una imagen está representada por un conjunto de regiones, que deberían corresponder aproximadamente a los objetos presentes en la imagen, que se caracterizan mediante su color, textura (aproximación basada en wavelet), forma y localización. El sistema clasifica las imágenes en diferentes categorías semánticas con el fin de mejorar los resultados de clasificación permitiendo una búsqueda adaptada a esta clasificación semántica.
- x PicSOM (Content-based Information Browsing and Retrieval System based on the Self-Organizing Map)<sup>10</sup>. Utiliza mapas autoorganizativos sobre color descrito en RGB, textura y forma descrita mediante transformadas de Fourier para recuperar imágenes similares. La premisa básica de la recuperación es que imágenes con descripciones similares estarán próximas en el mapa autoorganizado. Este mismo motivo les permite también incorporar de forma sencilla estrategias de navegación sobre la base de datos de imágenes.

Por último, terminar citando una aplicación real de un SRIBC de segunda generación. The State Hermitage Museum<sup>11</sup> en San Petersburgo, Rusia, dispone de dos aplicaciones para realizar consultas sobre sus fondos. Ambas aplicaciones están basadas en QBIC, Query By Image Content, desarrollado por IBM<sup>12</sup>. En QBIC Colour Search<sup>13</sup> la búsqueda se realiza especificando los porcentajes de los diferentes colores que deben aparecer en los cuadros resultantes y en QBIC Layout Search<sup>14</sup> no sólo se especifican los colores que deben aparecer, sino también su posición relativa, o layout, en los cuadros resultantes.

<sup>8</sup> [amazon.ece.utexas.edu/~qasim/research.htm](http://amazon.ece.utexas.edu/~qasim/research.htm)

<sup>9</sup> [wang14.ist.psu.edu/cgi-bin/zwang/regionsearch\\_show.cgi](http://wang14.ist.psu.edu/cgi-bin/zwang/regionsearch_show.cgi)

<sup>10</sup> [www.cis.hut.fi/picsom/](http://www.cis.hut.fi/picsom/)

<sup>11</sup> [www.hermitagemuseum.org/](http://www.hermitagemuseum.org/)

<sup>12</sup> [www.qbic.almaden.ibm.com/](http://www.qbic.almaden.ibm.com/)

<sup>13</sup> [www.hermitagemuseum.org/fcgi-bin/db2www/qbicColor.mac/qbic?selLang=English/](http://www.hermitagemuseum.org/fcgi-bin/db2www/qbicColor.mac/qbic?selLang=English/)

<sup>14</sup> [www.hermitagemuseum.org/fcgi-bin/db2www/qbicLayout.mac/qbic?selLang=English](http://www.hermitagemuseum.org/fcgi-bin/db2www/qbicLayout.mac/qbic?selLang=English)

## 8 EJERCICIOS PROPUESTOS

1. Suponer que la siguiente tabla representa los datos de una imagen digital en niveles de gris con una resolución de 16x8 píxeles y 32 niveles de gris. Y suponer que los valores representativos de cada bin de un histograma de 8 bins son: 3, 6, 9, 12, 17, 20, 23 y 26. Calcular el histograma de la imagen.

17	18	21	22	22	21	20	17	15	15	15	15	14	13	14	13
19	19	21	24	23	22	19	15	13	14	15	16	15	14	15	14
18	19	20	23	21	19	16	13	12	12	13	13	14	12	13	12
16	18	19	17	17	17	15	12	11	11	10	10	9	11	12	10
15	17	18	16	16	16	16	13	10	8	9	8	9	10	11	9
14	15	16	15	14	15	14	13	9	7	6	7	9	9	10	9
13	14	13	14	13	13	11	10	8	8	8	7	9	10	11	13
12	13	13	14	13	12	10	9	8	8	7	7	10	11	13	15

2. Suponer que la siguiente tabla representa los datos de una imagen digital en niveles de gris con una resolución de 16x8 píxeles y 4 niveles de gris. Calcular la matriz de coocurrencia para la relación espacial  $x=1$  e  $y=0$ .

2	2	3	3	3	3	3	2	1	1	1	1	1	1	1	1
2	2	3	3	3	3	2	1	1	1	1	2	1	1	1	1
2	2	3	3	3	2	2	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	1	1	1	1	0	0	0	1	1	0
1	2	2	2	2	2	2	1	0	0	0	0	0	0	1	0
1	1	2	1	1	1	1	1	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1
1	1	1	1	1	1	0	0	0	0	0	0	0	1	1	2

## 9 EJERCICIOS DE PROGRAMACIÓN

1. Desarrollar un sistema de recuperación de imágenes basado en contenido de color que utilice el espacio de color  $L^*a^*b^*$  para caracterizar el color de los píxeles de las imágenes, el histograma como descriptor del color global de las imágenes y una métrica euclídea para calcular la similaridad de los histogramas.

## 10 BIBLIOGRAFÍA BÁSICA COMENTADA

Entre los muchos libros de referencia en el ámbito del Procesado Digital de Imágenes destacan los libros clásicos escritos por González y diversos colaboradores. Repasan todos los temas importantes del campo y son una lectura obligada para cualquiera que se introduzca en este campo. A las versiones tradicionales ha añadido en los últimos años una versión del libro especialmente dedicada a la librería de imágenes de MATLAB. Además existen buenas traducciones al castellano.

- ✕ Rafael C. González, Richard E. Woods. Digital image processing. Pearson-Prentice Hall, cop. 2008, ISBN: 9780135052679
- ✕ Rafael C. Gonzalez, Richard E. Woods, Steven L. Eddins. Digital image processing using MATLAB. Gatesmark publishing, cop. 2009. ISBN: 9780982085400

El libro escrito por Sonka puede considerarse como otra referencia clásica en el ámbito. Su nivel es un poco más avanzado que los libros de Gonzalez y es, por tanto, conveniente leerlo cuando ya se tienen algunos conocimientos del campo.

- ✕ Milan Sonka, Vaclav Hlavac, and Roger Boyle. Image processing, analysis, and machine vision. Thompson Learning, cop. 2008. ISBN: 9780495082521

Este libro se ha incorporado porque es la referencia utilizada para la definición de los conjuntos de colores. Es un buen ejemplo de un libro dedicado a los principios fundamentales del color.

- ✕ Brent Berlin and Paul Kay. Basic Color Terms: Their Universality and Evolution, 1969, ISBN: 1-57586-162-3

El siguiente grupo de libros suponen una selección de libros de lectura obligatoria para cualquiera que se introduzca en el campo de la recuperación de imágenes.

- ✕ Del Bimbo, Alberto, Visual Information Retrieval, Morgan Kaufmann Publishers, 1999, ISBN: 1-55860-624-6
- ✕ Marques, Oge y Furht, Borko, Content-Based Image and Video Retrieval, Kluwer Academic Publishers, 2002. ISBN: 1-4020-7004-7

- x Jörgensen, Corinne, Image Retrieval: Theory and Research, Scarecrow Press, 2003. ISBN: 0-8108-4734-5, 2003

En este libro se comentan las últimas líneas de investigación en el campo de la recuperación de imágenes.

- x V.V.A.A., Artificial Intelligence for Maximizing Content Based Image Retrieval, Zongmin Ma, editor, IGI Global, 2009. ISBN: 978-1-60566-174-2

## 11 PARA SABER MÁS

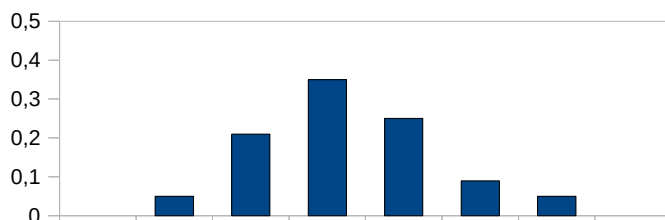
No se incluye nada por el carácter introductorio de este capítulo .

## 12 REFERENCIAS EN LA WEB

Las referencias en la Web se han introducido dentro del texto en el apartado 7.

## 13 SOLUCIONES A ALGUNOS EJERCICIOS PROPUESTOS

1. Para calcular el histograma se toma el valor representativo del primer bin (3) y se cuenta el número de píxeles que tienen su valor más próximo al valor de este bin que a los valores del resto de los bins (0 píxeles). A continuación, se toma el valor representativo del segundo bin (6) y se cuenta, de nuevo, el número de píxeles que tienen su valor más próximo al valor de este bin que a los valores del resto de los bins (6 píxeles; los de valor 6 y 7). Este proceso se repite con todos los valores representativos de los diferentes bins que forman el histograma. Por último, para obtener el valor del histograma normalizado en cada bin se divide el número de píxeles hallado entre el número total de píxeles de la imagen (128).



2. Para calcular una matriz de coocurrencia de esta imagen se debe comenzar fijando las relaciones espaciales,  $x$  e  $y$ , que se van a caracterizar. En este caso se estudiarán parejas de píxeles en las cuales el segundo píxel está colocado a la derecha del primero a una distancia unidad. Cada elemento de la matriz de coocurrencia  $P_{10}$  es el número de parejas de píxeles (con la relación espacial adecuada  $x=1$  e  $y=0$ ) en las cuales el primer píxel tiene el nivel de gris del primer índice de la matriz y el segundo píxel tiene el nivel de gris del segundo índice de la matriz. Finalmente, cada valor se divide por el número de píxeles totales de la imagen (128).

$$P_{10} = \begin{bmatrix} 0,20 & 0,05 & 0,00 & 0,00 \\ 0,05 & 0,34 & 0,05 & 0,00 \\ 0,01 & 0,04 & 0,11 & 0,02 \\ 0,00 & 0,00 & 0,02 & 0,07 \end{bmatrix}$$

## 14 CÓDIGO FUENTE

No hay debido al carácter introductorio del capítulo.