La recuperación automatizada de imágenes: retos y soluciones

Juan Antonio MARTÍNEZ COMECHE
Departamento de Biblioteconomía y Documentación.
Facultad de Ciencias de la Documentación. Universidad Complutense de Madrid
juan.comeche@pdi.ucm.es

Recibido: Agosto 2013
Aceptado: Octubre 2013

Resumen: Análisis de las características peculiares de la imagen como documento y de sus consecuencias de cara al desarrollo de sistemas automatizados de recuperación de este tipo de documentación. Tras desarrollar los conceptos de imagen y recuperación de imagen, se afronta el análisis de los retos que la imagen presenta desde el punto de vista de su recuperación, en especial el vacío semántico, y se describen las principales soluciones encontradas en la literatura sobre el tema principalmente desde 1990 hasta el presente. Se concluye que el enfoque actual (SBVIR) se caracteriza por simultanear el código visual y el código lingüístico en la representación.

Palabras clave: Fotografía; Imagen; Recuperación automatizada de imágenes; Recuperación de imagen; Recuperación de información; Sistemas de Recuperación de Imagen Basada en Contenido (CBIR); Vacío semántico.

Image Retrieval: Challenges and Solutions

Abstract: Analysis of the peculiar characteristics of images as documents and its consequences for the development of automated retrieval of this type of documentation. After developing the concepts of image and image retrieval, the challenges that image presents from the point of view of retrieval are analyzed, especially the semantic gap, and the main solutions found in the literature since 1990 to present are described. The current approach (SBVIR) is characterized by the simultaneous employment of the visual code and the language in representing images.

Keywords: Content-Based Image Retrieval; Image; Image Retrieval; Information Retrieval; Photography.

1 INTRODUCCIÓN

En los últimos años se ha constatado un aumento considerable en el número de elementos audiovisuales que los usuarios manejan e intercambian entre sí. Las fotografías, los vídeos y otras imágenes diversas (desde esquemas y dibujos hasta gráficos) sirven no solo como apoyo o complemento de mensajes de carácter

textual, sino que cada vez en mayor medida constituyen el formato de los mensajes en su integridad.

Desde el ámbito de la representación y recuperación de la información, la transmisión de mensajes informativos exclusivamente icónicos, o multimedia en general, implica un reto de considerable dificultad. La imagen había ocupado inicialmente una posición subsidiaria a la palabra, a la que complementaba dada su incapacidad para transmitir un mensaje de manera autónoma. En cambio, ahora la imagen adopta un papel protagonista, aupado en sus capacidades por la nueva posibilidad de impresionar la luz en un soporte.

En este trabajo analizaremos los problemas que plantea este tipo de documentación y las sucesivas soluciones parciales halladas, principalmente desde 1990 hasta el presente, de cara al desarrollo de sistemas automatizados de recuperación de esta clase de materiales.

Para abordar los problemas planteados por los Sistemas de Recuperación de Imagen y sus posibles soluciones técnicas, se ha adoptado una metodología documental, mediante la recopilación de la literatura existente sobre dicha temática, considerando principalmente las actas de los congresos internacionales dedicados al tema, así como las recopilaciones de los trabajos más destacados y actuales sobre el procesamiento de imágenes. En este trabajo se ha seleccionado la considerada más destacada y al mismo tiempo con información más actualizada.

2 EL CONCEPTO DE IMAGEN

Antes de comenzar esta descripción de los sucesivos métodos de representación y recuperación de imágenes, dadas sus peculiaridades y dificultades en relación a los mensajes textuales, debemos centrar antes los propios conceptos de imagen y de recuperación de imagen, con el fin de fijar claramente los límites que abarca este estudio.

El concepto de imagen que manejaremos depende, en primera instancia, del sistema visual humano. La realidad es observada de manera diversa por distintos seres, conforme a las características morfológicas del ojo (la visión es muy distinta, por ejemplo, si se posee un ojo compuesto o un ojo simple). Si la realidad no es percibida de la misma manera por todos los seres, en nuestro caso nos interesa exclusivamente la imagen tal como es formada en el ojo humano.

El segundo factor que debe ser considerado en la definición de imagen atañe al componente fundamental que es percibido inicialmente por el ojo y transformado finalmente por el cerebro, esto es, la luz. La luz no es más que una radiación electromagnética cuya longitud de onda se halla entre ciertos márgenes o limites (los 400nm y los 750nm) que el ojo humano es capaz de procesar. A su vez, la luz o espectro visible se puede dividir en zonas de frecuencia que corresponden a los diversos colores que distinguimos los seres humanos. Aunque existe un componente subjetivo en la percepción del color, un ojo humano habitualmente

equipara los colores monocromáticos a la luz entre ciertos rangos de frecuencias. Así, el rojo correspondería a la luz de entre 650nm y 750nm (nanometros) de longitud de onda aproximadamente, y así sucesivamente. Si un objeto refleja equilibradamente todas las longitudes de onda del espectro visible, el ojo humano lo percibe como blanco. Finalmente, el negro corresponde a la ausencia total de luz. De lo anterior concluimos que una imagen puede concebirse como una composición de colores.

Un tercer pilar debe, a nuestro juicio, configurar el armazón del concepto de imagen. Dicho componente alude al soporte en que se plasma la imagen. Toda imagen, desde un punto de vista documental, implica el registro de un cierto mensaje icónico[1], manualmente (la pintura o la escultura, por ejemplo) o mediante diversos procedimientos técnicos (es el caso de la fotografía o una película) en un soporte de naturaleza variable (piedra, papel, cristal, madera, tela o cualquier otro material que permita la perdurabilidad del mensaje registrado en él). Este acto de incorporación del mensaje icónico a un registro se realiza, como en cualquier otro documento, con la finalidad primera de garantizar la perdurabilidad del mensaje en el tiempo (Martínez-Comeche, 1995).

La concepción de imagen como colores percibidos por el ojo y registrados en un soporte perdurable es aún extremadamente amplia para nuestros intereses. Acogería toda la arquitectura, la escultura, la pintura, e incluso cualquier objeto visible de la naturaleza. Una posible solución consiste en imponer dos dimensiones al mensaje (Gonzalez; Woods, 2008). Si es cierto que toda imagen debe incorporarse a un soporte que obligatoriamente posee tres dimensiones, lo que por otra parte lo convierte en un objeto físico manejable, es el mensaje icónico el que carece de profundidad, consistiendo esencialmente en una representación bidimensional de un objeto o de una idea. De esta forma mantenemos las pinturas, las fotografías, las películas y vídeos, los dibujos, los mapas, las ilustraciones, los gráficos e incluso los textos (cuando atendemos solo a su valor icónico).

Reuniendo los elementos analizados hasta aquí, podemos concebir imagen como la percepción visual humana de una composición bidimensional de trazos de color registrado en un soporte perdurable.

[1] Empleamos el término icónico desde un enfoque semiótico. La Semiótica concibe el signo como toda entidad o fenómeno que es percibida sensorialmente y que remite a otra realidad que no está presente. A su vez, Ch. S. Peirce clasifica los signos en iconos, indicios y símbolos. Este autor concibe los iconos como signos que presentan una relación de semejanza con la realidad exterior. En este sentido debe entenderse 'mensaje icónico' en el texto; esto es, como un mensaje que alude a otra entidad o fenómeno no presente y que la representa de manera semejante a como es, o al menos tal como su autor concibe dicha entidad o fenómeno (Dubois et al., 1983, s. v. signo e icono).

3 EL CONCEPTO DE RECUPERACIÓN DE IMAGEN

Una vez analizado el concepto de imagen que manejaremos a lo largo del texto, queda por comentar el alcance del término Recuperación de imagen. La Recuperación de información (Information retrieval) constituye un área de conocimiento con una dilatada experiencia desde sus inicios hacia mediados del siglo XX. Entre las múltiples conceptuaciones existentes figura la de Lancaster, quien reconoce que, aunque suele poseer unos amplios límites entre los especialistas, su núcleo consiste en el estudio de los sistemas automáticos que informan al usuario de la existencia de documentos relacionados con su consulta (Lancaster, 1968). Rijsbergen destaca el aspecto automático de la tarea realizada por estos sistemas, en cuanto que llevan a cabo tanto la representación como la extracción de los documentos relevantes sin intervención directa del ser humano (Rijsbergen, 1979). Asimismo, ambos reconocen que se puede describir adecuadamente el tipo de materiales que se manejan en esta especialidad sustituyendo la palabra 'información' por 'documentos'. En consecuencia, la Recuperación de información (Information Retrieval) trata esencialmente documentos en formato digital, de manera que puedan ser representados y recuperados mediante algoritmos automáticos.

Dentro del campo de la Recuperación de información, la consideración de una rama dedicada específicamente al estudio e investigación de la documentación visual no surge hasta la década de 1990. Hasta entonces las imágenes son representadas lingüísticamente, sin diferenciar estos documentos de los textos habituales en otros sistemas de recuperación. La consciencia de rasgos propios de la imagen que no comparten los documentos textuales, como el color y la forma, hace surgir la especialidad de la recuperación de imágenes en cuanto se desarrollan técnicas específicas para representar estas cualidades que no existen en el campo tradicional de la recuperación textual.

El primer autor en utilizar la expresión Recuperación de imagen, denominándola Recuperación de Imagen Basada en Contenido (Content-Based Image Retrieval o CBIR), fue Toshikazu Kato en 1992 (Pérez Álvarez, 2007), aludiendo con el término 'contenido' a los rasgos visuales mencionados de color y forma.

Sin embargo, esta recién creada especialidad equipara imagen con imagen fija, excluyendo el análisis de películas y vídeos. Poco después surge la Recuperación de Vídeo Basada en Contenido (Content-Based Video Retrieval o CBVR), con la intención de aplicar los mismos procedimientos que CBIR a la recuperación de material fílmico. A finales de la década de 1990 ambas disciplinas se combinan en la denominada Recuperación de Información Visual Basada en Contenido (Content-Based Visual Information Retrieval o CBVIR) (Zhang, 2007).

La tendencia actual consiste en limitar la primigenia Recuperación de información (Information Retrieval) a la recuperación de carácter textual, y utilizar la variante Recuperación de información Multimedia o Recuperación Multimedia

(Multimedia Retrieval, MMIR o MIR) cuando se desea considerar la recuperación conjunta de información de cualquier índole, ya sea texto, imágenes, sonido y vídeos en formato digital (Blanken; Vries; Blok; Feng, 2007).

Resumiendo los elementos comentados anteriormente, consideraremos Recuperación de imagen como el proceso automático de representación y búsqueda de imágenes en formato digital.

4 LOS SISTEMAS DE RECUPERACIÓN DE IMAGEN

Para poder cumplir el objetivo planteado de representación y búsqueda, todo Sistema de Recuperación de Imagen incluye los siguientes módulos esenciales:

- Módulo de Descripción. Su misión es representar numéricamente las propiedades o cualidades de las imágenes que ingresan en la colección digital. Estas propiedades pueden ser de dos clases:

  - Propiedades intrínsecas de la imagen: Alude a rasgos visuales que caracterizan toda imagen. Entre ellos destacan el color, la textura, la forma y las relaciones espaciales. Suelen englobarse bajo la denominación de propiedades de bajo nivel.
  - Propiedades extrínsecas de la imagen: Alude a todo elemento no propiamente visual. Dentro de este amplio grupo de las propiedades extrínsecas puede efectuarse la siguiente subdivisión (Müller; Clough; Deselaers; Caputo, 2010):
    - Propiedades de nivel medio: Alude a la detección automática de límites, contornos, objetos (caras, sombreros, edificios, personas...) y de 'conceptos' extraídos de la imagen en su integridad (interior-exterior, día-noche, verano-invierno...).
    - Propiedades de nivel alto: Alude a los elementos objetivos que se incluyen en los metadatos (autor, título, localización geográfica, fecha, formato, género, periodo artístico...) o a propiedades de carácter subjetivo (denominadas semánticas) extraídas a raíz de la contemplación de la imagen (paz, amor, guerra, huelga, soledad...) y que suelen incorporarse en el apartado 'Descripción' o 'notas' en los metadatos.

- Módulo de Consultas. Permite al usuario introducir la consulta o expresar su necesidad informativa. Existen varios métodos que permiten introducir la consulta:

  - Mediante texto. La consulta se expresa lingüísticamente.
  - Mediante ejemplos. El sistema presenta al usuario algunas imágenes de la colección que pueden ser utilizadas por el usuario como ejemplo de lo que busca.

- Mediante navegación por la colección. El sistema permite al usuario recorrer las imágenes de que consta la colección digital para elegir la adecuada a su consulta.

- Módulo de Búsqueda. Incluye el procedimiento automático de extracción de las imágenes más relevantes existentes en la colección en relación a cada consulta y su ordenación por orden de relevancia. Estos procesos de emparejamiento y ordenación requieren la inclusión en el sistema de un algoritmo de similaridad.

5 EL VACÍO SEMÁNTICO

Como hemos visto en el epígrafe anterior, son muy extensas las posibilidades de descripción de una imagen: Desde los atributos visuales más básicos, como el color o su composición y distribución a lo largo de la imagen, hasta los sentimientos subjetivos que puede evocar una imagen.

Los sistemas de recuperación de imagen pueden representar de manera automática y con facilidad las características de bajo nivel o propiedades intrínsecas de las imágenes (color, forma, textura o relaciones espaciales). Todos estos atributos son consustanciales a la imagen, por lo que basta con registrar sus niveles en cada una de las imágenes de la colección.

La cuestión se complica cuando abordamos la representación automática de las propiedades extrínsecas. Describir el concepto de 'cara' en términos de color no es tarea sencilla, y mucho más intrincado sería expresar mediante estos atributos de bajo nivel un concepto subjetivo como el amor o la intranquilidad. En consecuencia, conforme más nos elevamos en el campo de las propiedades de la imagen, más aumenta la complejidad de la 'traducción' o de la equiparación entre los atributos intrínsecos de la imagen (color, forma, textura o disposición espacial), consustanciales a la percepción misma de la imagen, y el concepto o idea que deseamos representar.

Los seres humanos, por su parte, expresan con relativa facilidad los atributos de alto nivel, incluyendo las propiedades semánticas (paz, tristeza...), gracias al empleo de las palabras. En cambio, la dificultad va aumentando conforme trata de expresar con palabras los atributos de más bajo nivel (color, forma, textura...).

Ello provoca finalmente una disfunción en el tercer módulo de todo sistema de recuperación de imagen, el referido a las búsquedas. Si un usuario desea expresar la consulta mediante un código textual, y las descripciones de las imágenes están expresadas mediante un código visual, ¿cómo desarrollar un algoritmo automático que equipare una cierta consulta en palabras con tales descripciones realizadas mediante atributos visuales?

Este fenómeno recibe el nombre de vacío semántico (semantic gap), y supone una dificultad fundamental que debe afrontar necesariamente todo sistema de recuperación de imagen. A la postre, la solución pasa necesariamente por hallar alguna equiparación

plausible entre ambos códigos que pueda replicarse automáticamente. El usuario debe tener allanado en lo posible la expresión de sus intereses, empleando la lengua si así lo desea, mientras que el sistema debe resolver fácilmente la cuestión de la descripción de imágenes limitándose a los atributos primarios.

En relación a las propiedades de nivel medio, relativas a la detección de objetos y de conceptos deducidos de manera directa a partir de los rasgos visuales de nivel bajo (color, forma, textura o disposición espacial), dicha equiparación está siendo investigada con mayor intensidad desde el año 2005 aproximadamente. Se ha comprobado que la tarea es de un nivel de dificultad muy elevado y que los resultados obtenidos deben mejorar aún mucho para ser considerados de factible incorporación en sistemas reales (Müller; Clough; Deselaers; Caputo, 2010: 214-216). Entre las conclusiones más destacadas obtenidas a lo largo de estos pocos años de investigación, destacar los mejores resultados obtenidos empleando programas de aprendizaje máquina que implementan clasificadores a partir de descriptores básicos y el abandono paulatino de herramientas complementarias como taxonomías y ontologías en favor del empleo de datos provenientes del etiquetado social.

En relación a las propiedades de alto nivel, la solución propuesta en los primeros años de investigación consistió esencialmente en hacer convivir ambos códigos (el textual y el visual) y favorecer la equiparación entre ambos mediante técnicas de retroalimentación por relevancia, en las que los usuarios juzgaban la relevancia de las imágenes recuperadas inicialmente, información que era empleada a continuación para mejorar la respuesta del sistema (Zhang, 2007). Actualmente, en cambio, el enfoque adoptado generalmente es el mismo que el comentado en relación a las propiedades de nivel medio, esto es, tratar de hallar una equiparación entre el código visual y las correspondientes propiedades de alto nivel. En definitiva, se pretende hacer corresponder conceptos semánticos con una combinación de rasgos visuales. Los métodos para conseguirlo coinciden con los apuntados en relación a las propiedades de nivel medio: empleo de técnicas de aprendizaje máquina para el desarrollo de algoritmos de clasificación y el empleo del conocimiento humano, ya mediante etiquetado social o retroalimentación por relevancia, y de técnicas estadísticas. Como era de esperar en una tarea de tal dificultad como la de generar automáticamente descriptores que atribuirían las personas subjetivamente al observar la imagen, los resultados aún deben mejorar mucho para poder ser considerados aplicables en un sistema de recuperación operativo.

En resumen, estamos aún muy lejos de solventar el vacío semántico, esto es, de conseguir que los sistemas de recuperación de imagen puedan asignar automáticamente características y propiedades a todos los niveles de significación, desde las puramente visuales hasta las interpretaciones subjetivas realizadas por los seres humanos.

Mientras la investigación en la recuperación semántica logra avances significativos, la Recuperación de imagen no puede prescindir de ninguno de los dos códigos involucrados en la tarea de recuperación, el visual y el lingüístico. En

el siguiente epígrafe analizaremos más detalladamente el grado de imbricación entre ambos códigos desde sus orígenes en la década de 1990.

6 ETAPAS FUNDAMENTALES EN LA RECUPERACIÓN DE IMAGEN

En la evolución de la Recuperación de Imagen pueden distinguirse tres grandes etapas:

- Una primera etapa en la que los Sistemas de Recuperación de Imagen se basan en representaciones textuales de las características de las imágenes.
- Una segunda etapa en la que los Sistemas de Recuperación de Imagen se basan en los rasgos visuales de las imágenes.
- Una tercera etapa en la que los Sistemas de Recuperación de Imagen emplean simultáneamente el código visual y el código textual para representar y recuperar imágenes.

La primera etapa abarca desde los orígenes de la incorporación de imágenes como unidad de descripción documental a las colecciones digitales hasta la década de 1990. Durante estos inicios, los Sistemas de Recuperación de Imagen aplican a estos nuevos documentos las mismas técnicas que se empleaban con los documentos textuales. En consecuencia, importa fundamentalmente todo texto que acompañe a la imagen, pues de ellos se extraerán automáticamente los términos de indización que representarán el contenido de la misma.

Este enfoque todavía se emplea hoy día en sistemas automatizados donde la cantidad ingente de imágenes que se incorporan a la colección de manera constante hace impensable cualquier otro procedimiento. Es el caso, por ejemplo, de los motores de búsqueda generalistas como Google y Bing. Además de la utilización de todos los textos relacionados con la imagen, ya sea en el mismo documento o incluso en documentos ajenos que enlazan con ellas, estos sistemas con contenidos tan amplios se distinguen por el progresivo empleo de caracterizaciones de contenido expresadas por los propios usuarios, lo que ha venido en denominarse anotación social, cada vez más habitual en Internet.

Cuando los sistemas de recuperación de imagen acogen colecciones digitales más restringidas, es habitual la presencia de herramientas documentales clásicas y, en general, el traslado al repositorio digital de los procedimientos habituales en colecciones manuales. Así, en los sistemas de muchas pinacotecas, donde el volumen de pinturas es restringido, es frecuente el empleo de tesauros de historia del arte que permiten normalizar la descripción de movimientos y periodos artísticos.

En los sistemas de esta primera etapa, que incorporan exclusivamente el código lingüístico, la tarea de recuperación se reduce esencialmente a la búsqueda de las palabras clave empleadas al describir la imagen. Los campos más habituales

considerados en la descripción suelen coincidir con un subconjunto de los 15 elementos de que consta el esquema de metadatos Dublin Core básico (Feng; Brussee; Blanken; Veenstra, 2007). Como ejemplo de esta primera etapa puede consultarse el sitio web del Museo del Prado en http://www.museodelprado.es (en concreto, la galería online). Allí podemos comprobar cómo la búsqueda de un cuadro se realiza a través de dos puntos de acceso tradicionales en Biblioteconomía: Autor y Título de la obra.

Este primer enfoque presenta, sin duda, problemas, muchos de ellos puestos de manifiesto tradicionalmente por bibliotecarios y documentalistas sobre las colecciones textuales desarrolladas en bibliotecas. Entre ellos destacan los siguientes:

- La dependencia de analistas humanos. Este enfoque implica la continuidad en relación a la labor documental de descripción esencialmente manual de la colección mediante la adaptación de normas de representación pre-existentes y con la ayuda de vocabularios controlados y otras herramientas documentales. Si el volumen de imágenes es limitado, el problema puede abordarse con relativa facilidad, pero puede suponer un problema irresoluble si el repositorio posee un crecimiento constante (en Internet, por ejemplo).
- La inconsistencia de la descripción entre analistas. Problema conocido de antiguo y ampliamente debatido en el ámbito de la Biblioteconomía y Documentación que se traslada con las mismas características cuando la colección consta de imágenes.
- El volumen de la documentación. Auténtica piedra angular que justifica la necesidad de desarrollar procedimientos automáticos de descripción y recuperación. Es sobradamente conocido el crecimiento exponencial de la información disponible en Internet, en formato multimedia en un porcentaje llamativo. Si la teoría de la Documentación ha demostrado la necesidad de recopilación y gestión de dicha información para garantizar su disponibilidad y uso por parte de los usuarios que la precisen (Desantes Guanter, 1983), se deduce -dada la imposibilidad de realizar estas tareas de manera manual- la obligatoriedad de abordar el problema con medios automáticos.
- Dificultad de descripción de las propiedades perceptuales o de bajo nivel (color, forma...) mediante el código lingüístico. El problema de la limitación terminológica para describir las posibilidades cromáticas de la visión humana ha sido sobradamente puesto de manifiesto por los lingüistas. Sin duda, el código más apropiado para este tipo de descriptores es el propio código visual.

A fin de resolver estos inconvenientes, en la década de 1990 se desarrolla una segunda etapa denominada Recuperación de Imagen Basada en Contenido (Content-Based Image Retrieval o CBIR). La principal novedad de este nuevo enfoque consiste, como ya se ha apuntado, en la adopción de un código

propiamente visual, por completo distinto al lingüístico. La imagen se describe ahora mediante sus propiedades perceptuales, esto es, mediante las características psicofísicas percibidas por el ojo humano, principalmente el color, la textura, la forma o las relaciones espaciales.

En cuanto al color, la Recuperación de imagen parte de un modelo estándar que permite representar cualquier color mediante tres números (basados en los colores primarios rojo (R), verde (G) y azul (B) o RGB). En principio, cada píxel de la imagen posee un color, y puede -por consiguiente- ser descrito mediante una triada de números (González; Woods, 2008: 402-406).

La textura alude a zonas de la imagen que, sin ser estrictamente uniformes en cuanto al color, presentan un patrón que se repite. Por ejemplo, un bosque incluido en una imagen presenta un patrón -tronco del árbol, ramas, hojas...- que no posee un único color, pero que se repite hasta ser percibido por el ojo humano como una propiedad de la imagen. Esta propiedad, al acoger una zona de la imagen, engloba varios píxeles simultáneamente. Se suele representar numéricamente mediante un conjunto de números (uno por característica: energía, contraste, correlación...) por cada matriz de datos correspondiente a una posición espacial considerada dentro de la zona afectada por la textura (la matriz contiene las probabilidades de encontrar parejas de colores en dicha posición) (Mosquera González; Carreira Nouche; González Penedo, 2011).

Una de las maneras más sencillas de representar numéricamente las formas presentes en una imagen consiste en utilizar dos números por cada píxel de la imagen, correspondientes al nivel de variación de la intensidad en dicho píxel con respecto a los ejes de abscisas y ordenadas. Estos dos números se denominan valores de borde, y tienen la propiedad geométrica de señalar la dirección del máximo cambio de intensidad en dicho píxel y la magnitud del cambio en dicha dirección (González; Woods, 2008: 706-712).

Por último, entre los rasgos visuales más habituales empleados en Recuperación de imagen se hallan las relaciones espaciales, consistente en representar las posiciones relativas de ciertas zonas o regiones de la imagen. Las regiones se han detectado previamente, habitualmente seleccionando los píxeles que poseen propiedades (color, intensidad...) semejantes a sus vecinos. De esta manera, por ejemplo, puede detectarse el cielo, el mar, la tierra o las nubes en una determinada imagen (González; Woods, 2008: 763-766). La posición relativa de estas regiones puede aportar información importante sobre la imagen. En consecuencia, una vez detectadas ciertas regiones en la imagen, para representar fácilmente su posición relativa se utilizan varios métodos, entre los que destaca por su sencillez las cadenas 2D. Una primera cadena barre la imagen de arriba abajo, indicando por orden las regiones que se van encontrando; una segunda cadena barre la imagen de izquierda a derecha, indicando igualmente por orden las regiones que se van topando (Mosquera González; Carreira Nouche; González Penedo, 2011: 571-572).

Una vez representadas numéricamente cada una de estas propiedades perceptuales o de bajo nivel (procesado digital de la imagen), el sistema de recuperación de imagen basado en contenido está listo para responder a las necesidades de los usuarios. En un sistema CBIR puro la expresión de la necesidad informativa se realiza mediante una consulta que no emplea texto, sino el propio código visual que ha servido previamente para representar las propiedades perceptuales de la imagen.

En un sistema basado en contenido se pueden realizar consultas en las que el usuario escoge el color o colores más destacados de la imagen que busca, puede también señalar las formas más sobresalientes de la imagen, bien dibujándolas personalmente o a veces mediante iconos que la interfaz pone a su disposición para facilitar esta labor, y de igual forma puede definir regiones completándolas con colores sólidos y en la localización espacial deseada. El lector interesado puede comprobar las posibilidades de este tipo de sistemas de recuperación de imagen basada en contenido en la página web del sistema experimental RETRIEVR (http://labs.systemone.at/retrievr/#sketchName=2013-04-13-21-12-46-480637.8), donde pondrá realizar sus consultas partiendo de una selección de las imágenes de Flickr.

Otro método alternativo al anterior consiste en permitir al usuario visualizar una selección de imágenes de la colección para que escoja una imagen como consulta, sobreentendiendo que el usuario desea imágenes similares a la que sirve de ejemplo. Esta posibilidad es la que posee el sistema PicSOM (http://picsom.ics.aalto.fi/picsom/query/Q:130413:213916:31993:5ae/), donde el lector tiene a su disposición una selección de imágenes con personas, paisajes o aviones entre otras posibilidades.

Introducida la consulta, únicamente resta que el sistema localice en la colección o repositorio aquellas imágenes similares a la consulta. Para ello, como apuntamos en su momento, es preciso que el sistema disponga de un algoritmo de similaridad. Actualmente el enfoque más común consiste en cuantificar numéricamente la similitud entre dos imágenes representando cada imagen de la colección o repositorio como un punto en el espacio de propiedades (color, forma, etc., ya cuantificadas previamente, conforman las coordenadas) y la consulta como otro punto en dicho espacio, y equiparar el grado de similitud entre dos imágenes (puntos en el espacio) a la distancia entre dichos puntos (a mayor distancia entre dos puntos, mayor diferencia entre sus imágenes correspondientes). De esta forma, las imágenes que en el espacio de propiedades estén más cercanas a la consulta serán presentadas al usuario como respuesta a su consulta.

Los Sistemas de Recuperación de Imagen Basada en el Contenido (CBIR) que configuran la segunda etapa también presentan problemas, que podemos subdividir de la siguiente manera:

- Las limitaciones formales. Engloban los problemas al nivel de representación de las propiedades o atributos visuales. Entre ellas, destacan las relativas al color y a las formas:
- En ocasiones las diferencias numéricas entre dos colores (sus representaciones) no se corresponden con la diferencia percibida por el ser humano en relación a tales colores. Además, en dicha percepción pueden estar involucrados aspectos culturales (el color blanco de la nieve para los esquimales, por ejemplo, no es único, distinguiéndose varios términos y colores).
- Formas percibidas como únicas por los seres humanos pueden originar representaciones numéricas muy distintas. Basta, por ejemplo, que dos imágenes capten una misma forma desde ángulos o perspectivas diferentes para que sus representaciones numéricas sean muy dispares (Robledano Arillo, 1999: 291).
- Las limitaciones semánticas. Engloban las dificultades a la hora de efectuar las consultas por parte del usuario, en última instancia motivadas por el vacío semántico. Entre ellas, podemos destacar las siguientes:
- Las necesidades informativas de los usuarios no suelen situarse al nivel de las propiedades perceptuales empleadas en los sistemas CBIR. Es relativamente poco frecuente que un usuario busque expresamente cierto color o una determinada forma en un banco de imágenes. Los usuarios suelen expresar sus necesidades y consultas mediante propiedades de nivel medio y alto (objetos, escenas, conceptos abstractos...). A su vez, encuentra muchas dificultades para traducir estas propiedades extrínsecas mediante propiedades intrínsecas (color, forma, textura o disposición espacial).
- La subjetividad humana inherente a la interpretación de imágenes. Cualquier propiedad extrínseca de alto nivel de una imagen supone la interpretación de dicha imagen, lo que implica su variación entre perceptores distintos, e incluso su posible variación para un mismo observador con el tiempo.

A raíz de los problemas planteados por la Recuperación de Imagen Basada en Contenido (CBIR), surge recientemente una tercera etapa que trata de superarlos. Es la denominada Recuperación de Información Visual Basada en la Semántica (SBVIR). Este enfoque se caracteriza, en líneas generales, por la confluencia de la descripción del contenido visual y la descripción lingüística en una imagen.

Una vez comprobado que la representación de propiedades extrínsecas de alto nivel mediante atributos perceptuales de bajo nivel es una tarea extremadamente compleja, y aunque no se renuncia a este objetivo en ningún momento, parece razonable considerar el texto como un medio eficaz -hoy por hoy- de expresar significados abstractos o de alto nivel presentes en la imagen. En consecuencia, los sistemas de recuperación de imagen de la tercera generación emplean la anotación

de imágenes mediante descriptores textuales para representar propiedades extrínsecas de carácter subjetivo o semántico.

La representación de propiedades extrínsecas de nivel medio mediante rasgos visuales, aunque difícil de alcanzar, supone actualmente uno de los focos de investigación más activos. Mientras tanto, el empleo de descriptores textuales se alterna con la experimentación en el empleo de procedimientos de clasificación automática de objetos y escenas basándose en técnicas estadísticas. Un ejemplo que muestra las posibilidades de este enfoque, aunque de manera todavía inmadura, lo tenemos en el Museo Hermitage (http://www.hermitagemuseum.org/fcgi-bin/db2www/browse.mac/category?selLang=English).

En definitiva, esta tercera generación limita la participación del código lingüístico a aquellos niveles en los que todavía no es posible manejarse con éxito mediante un código puramente visual, empleando el conocimiento humano en dos manifestaciones principales: las etiquetas asignadas por usuarios humanos (anotación social) y la retroalimentación por relevancia durante la recuperación.

7 CONCLUSIONES

Las principales conclusiones que se extraen del presente estudio se pueden resumir como sigue:

1. La solución al reto principal que afronta el desarrollo de Sistemas de Recuperación de Imagen (denominado vacío semántico) pasa necesariamente por hallar alguna equiparación plausible entre los códigos visual y lingüístico que pueda replicarse automáticamente. El usuario debe poder expresar su necesidad informativa empleando palabras si así lo desea, mientras que el sistema debe ser capaz de traducir dichas palabras a los atributos primarios (color, forma, textura...) empleados en la representación de las imágenes.
2. En relación a la equiparación entre las propiedades de nivel medio (objetos y conceptos temporales o geográficos fundamentalmente) o las propiedades de nivel alto (en especial las propiedades semánticas) y los rasgos visuales de nivel bajo o atributos primarios (Recuperación de Imagen Basada en el Contenido), los mejores resultados obtenidos hasta el momento emplean programas de aprendizaje máquina que implementan algoritmos de clasificación. Estamos aún lejos de alcanzar resultados definitivos en esta tarea.
3. Mientras se consiguen avances significativos en la equiparación entre los códigos visual y lingüístico, ha surgido recientemente la denominada Recuperación de Información Visual Basada en la Semántica (SBVIR). Este enfoque se caracteriza por simultanear ambos códigos en la representación de las imágenes, limitando la participación del código lingüístico a aquellos niveles en los que todavía no es posible manejarse con éxito mediante un código puramente visual.
4. La Recuperación de Información Visual Basada en la Semántica (SBVIR)

emplea el conocimiento humano tanto en la representación como en la recuperación mediante dos procedimientos principales: las etiquetas asignadas por usuarios humanos (anotación social) en la representación y la retroalimentación por relevancia en el proceso de recuperación.

8 REFERENCIAS BIBLIOGRÁFICAS

BLANKEN, H.M; VRIES, A.P. De; BLOK, H.E.; FENG, L. (Eds.) (2007). *Multimedia Retrieval*. Berlin: Springer-Verlag.

DESANTES GUANTER, J.M. (1983). "Régimen jurídico de la actividad documentaria modal", en *Documentación de las Ciencias de la Información*, 1983, núm. 7, pp. 11-80.

DUBOIS et al. (1983). *Diccionario de Lingüística*. Madrid: Alianza.

FENG, L.; BRUSSEE, R.; BLANKEN,H.; VEENSTRA, M. (2007). Languages for Metadata, en BLANKEN, H.M; VRIES, A.P. De; BLOK, H.E.; FENG, L. (Eds.) (2007), *Multimedia Retrieval*. Berlin: Springer-Verlag, pp. 23-51.

GONZALEZ, R.C.; WOODS, R.E. (2008). *Digital Image Processing.* Third ed. New Jersey: Pearson Education

LANCASTER, F.W. (1968). *Information Retrieval Systems: Characteristics, Testing and Evaluation*. New York: Wiley.

MARTÍNEZ-COMECHE, J.A. (1995). *Teoría de la información documental y de las instituciones documentales*. Madrid: Síntesis.

MOSQUERA GONZÁLEZ, A.; CARREIRA NOUCHE, M.J.; GONZÁLEZ PENEDO, M.F. (2011). "Recuperación de imagen", en Cacheda Seijo, F. et al. (Eds.). *Recuperación de información: Un enfoque práctico y multidisciplinar*. Madrid: Ra-Ma.

MÜLLER, H.; CLOUGH, P.; DESELAERS, T.; CAPUTO, B. (Eds.) (2010). *ImageCLEF: Experimental Evaluation in Visual Information Retrieval*. Berlin: Springer-Verlag.

PÉREZ ÁLVAREZ, S. (2007). *Sistemas CBIR: Recuperación de imágenes por rasgos visuales*. Gijón: Trea.

RIJSBERGEN, C. J. Van (1979). *Information Retrieval*. 2nd. ed. London: Butterworths.

ROBLEDANO ARILLO, J. (1999). "La recuperación de la imagen fija. Perspectiva funcional de los sistemas automatizados de recuperación de imágenes", en *El análisis documental de la fotografía de prensa en entornos automatizados,* pp. 265-310. <http://hdl.handle.net/10016/499> [Consulta: 03/08/2013]

ZHANG, Y. (2007). *Semantic-Based Visual Information Retrieval*. London: IRM Press.