

Aproximación al estudio de los sistemas de recuperación de imágenes “CBIR” desde el ámbito de la documentación

Sara PÉREZ ÁLVAREZ

Departamento de Biblioteconomía y Documentación
Facultad de Ciencias de la Información
Universidad Complutense de Madrid

Recibido: 3-12-2005

Aceptado: 3-2-2006

RESUMEN

Los sistemas CBIR fueron propuestos a comienzos de la década de los '90 desde parcelas relacionadas con la visión automatizada y el procesamiento de la imagen. A pesar del interés que pueden ofrecer los sistemas CBIR para superar ciertas limitaciones del paradigma tradicional de análisis y recuperación de las imágenes basado en texto, lo cierto es que la investigación sobre estos sistemas desde una perspectiva documental ha sido muy escasa. La autora reflexiona sobre las aportaciones que la Documentación puede hacer al estudio de esta tecnología.

Palabras clave: Sistemas de recuperación de imágenes, CBIR, documentación de imagen.

An approach to the Study of CBIR Information Retrieval Systems from the Standpoint of Information Science

ABSTRACT

CBIR systems were put forward in the early 1990's in areas related to automated vision and image process. Despite the possibilities of CBIR in overcoming the traditional text-based paradigm of image analysis and retrieval, research on these systems carried out by Information Scientists is scarce. The author hence reflects on the potential contribution of Information Science to the study of these systems.

Key Words: Image Retrieval Systems, CBIR, Image Documentation.

1. DEFINICIÓN DE LOS SISTEMAS CBIR Y CONSIDERACIONES SOBRE SU SIGNIFICADO

Antes de proceder a una posible definición de los sistemas CBIR, interesa recordar los dos niveles estructurales básicos que la Semiótica establece en el enunciado icónico: el **nivel de la expresión**, también llamado formal o plástico, y el **nivel del contenido**¹.

¹ Estamos recordando el modelo estructural más simple, inspirado en la tríada de Charles Sanders Peirce. Un análisis más profundo de la teoría de este autor en: SANDERS PEIRCE, Charles. *La Ciencia de la Semiótica*, Buenos Aires: Nueva Visión, 1974.

El nivel de la expresión se identifica con la forma de representación de los iconos en la imagen. Por su parte, el nivel del contenido o nivel icónico consiste en la percepción de elementos de la realidad en la imagen. La significación en este nivel deriva de la interpretación de los iconos representados: identificamos categorías, características e identidad de objetos, así como su marco circunstancial –cronológico, geográfico y temático–.

Evidentemente, la separación entre ambos niveles sólo se justifica aquí para explicar los procedimientos analíticos que tienen los sistemas objeto de nuestra reflexión. También para explicar dentro de un momento qué debemos entender cuando hablamos de *sistemas de recuperación de imágenes basados en el contenido*.

John EAKINS y Margaret GRAHAM (1999), dos destacados profesores del *Institute for Image Data Research (IIDR)* de la Universidad de Northumbria (*New-Castle*), consideran que el primero en utilizar la expresión “*content-based image retrieval*” fue Toshikazu KATO², en su trabajo:

Database architecture for content-based image retrieval, En *Image Storage and Retrieval Systems* (Jambardino, A.A. & Niblack, W.R., eds.), Proc. SPIE 2185, 1992, pp.112-123.

Lo hace al describir sus experimentos sobre recuperación automática de imágenes a partir de dos bases de datos distintas, TRADEMARK y ART MUSEUM, que representan sus imágenes mediante los atributos de forma y color, respectivamente. Desde entonces, se ha utilizado ampliamente para referirse a la recuperación de imágenes almacenadas en grandes bases de datos a partir de rasgos como el color, la textura o la forma, que pueden extraerse de manera automática directamente de las imágenes.

Podemos apreciar que la cuestión derivada de la expresión CBIR (*content-based image retrieval*) es el tipo de atributos a partir de los cuales estos sistemas permiten recuperar imágenes. Literalmente la expresión se traduce como “recuperación de imágenes basada en el *contenido*”. Sin embargo, hoy día cuesta entender que esta expresión se refiera al nivel estructural de *contenido* al que aludíamos más arriba, ya que en la actualidad los sistemas CBIR consideran esencialmente los elementos formales intrínsecos que caracterizan a una imagen, es decir, el nivel formal o plástico.

Entre los elementos formales gráficos de carácter intrínseco que pueden ser abstraídos y analizados por estos sistemas se encuentran colores, texturas, figuras y relaciones topológicas entre esos atributos. El análisis de dichos elementos permite inferir **automáticamente** estructuras de composición de la imagen. Debemos tener presente esta apreciación, y quizá atrevernos a entender que el uso del término “*content-based*” alude **en la práctica** a que realmente lo *contenido* en una imagen es una serie de elementos morfológicos, a partir de los cuales el lector puede inferir un significado referente a su realidad, a su mundo visual y cultural. De esta forma tendría

² También otros autores coinciden en atribuir el origen de estos sistemas a los trabajos de Kato (VENTERS y COOPER, 2000b); y, cuando no se afirma directamente, al menos se está de acuerdo en que fue a comienzos de los '90.

sentido hablar de una recuperación que está basada en el contenido, es decir, una recuperación detrás de la cual siempre se esconde un proceso de extracción automática de rasgos visuales, considerados éstos como el verdadero **contenido**³ de una imagen.

Precisamente así lo conciben muchos autores, por ejemplo URCID PLIEGO (2003), para quien la expresión “por contenido” significa que los únicos datos que utilizan estos sistemas para formular las consultas son aquéllos que comprende o encierra la imagen en sí, o sea, su *contenido*⁴. Esta forma de entender la expresión *CBIR* es fruto de la realidad actual de estos sistemas.

Es interesante considerar los dos niveles estructurales, formal y de contenido, en que puede descomponerse una imagen desde otra perspectiva semejante que nos ayude a ver con mayor claridad, si aún cabe, los posibles atributos que sirven de materia prima a los sistemas *CBIR*. El profesor EAKINS (1996) establece un modelo semántico de la imagen orientado a las posibilidades de representación, precisamente, de los planos de significación de la imagen en un sistema de recuperación. Básicamente viene a distinguir entre rasgos primitivos (correspondientes al nivel plástico) y rasgos semánticos⁵. Su distinción tiene en cuenta el tipo de preguntas que el usuario puede hacer en relación con una imagen. A este respecto, establece tres niveles de abstracción: **rasgos primitivos** como colores, formas o texturas; **rasgos lógicos** como la tipología e identidad concreta de los objetos mostrados; y **atributos abstractos**, como las clases de eventos o actividades representadas, o la identificación de elementos de significación emocional. Tanto los rasgos lógicos como los atributos abstractos implican la referencia a un cuerpo de conocimiento ajeno a lo directamente mostrado por la imagen que nos permite identificar una determinada forma como un objeto concreto, o interpretar el significado de una escena; de ahí, que la distinción se resume finalmente en rasgos primitivos y rasgos o atributos semánticos.

Hoy día, tal como hemos dicho, los sistemas *CBIR* recuperan a partir del nivel de abstracción más bajo (color, forma, textura, distribución espacial, posición, bordes, etc.), pero la mayoría de los usuarios demanda un estadio de recuperación más alto. El hecho de que actualmente los sistemas *CBIR* funcionen de manera relativamente exitosa en la recuperación de imágenes desde el punto de vista del plano o nivel formal no significa que su intención no sea la de recuperar también a partir del nivel del contenido. Así, autores como VENTERS, HARTLEY y HEWITT (2004) aseguran expresamente que **el objetivo de la recuperación *CBIR* también es la recuperación semántica de las imágenes**.

Y éste es precisamente el principal reto al que se enfrenta la tecnología *CBIR* en la actualidad: el reto de inferir significados por los que se pueda recuperar partiendo de la extracción automática de atributos visuales. Será cuando quede superado este reto, cuando podamos entender claramente que el término “*content-based*” que

³ Entendido éste como nos indica la segunda entrada del D.R.A.E. (21ª ed., 2001) para este término (2. m. Cosa que se contiene dentro de otra) y no desde la interpretación lingüística (4. *Ling.* Plano del contenido).

⁴ De acuerdo con nota anterior.

⁵ Distinción que, por su parte, ya hicieron con anterioridad GUDIVADA y RAGHAVAN (1995).

se encuentra en su denominación hace plena referencia tanto a la recuperación del contenido formal como a la del contenido icónico de las imágenes.

Finalmente, presentamos a continuación la definición que más completa y precisa nos resulta acerca de la naturaleza de los sistemas CBIR. Se trata de una definición dada por un profesor español proveniente precisamente del ámbito de la Documentación, el profesor Jesús Robledano Arillo, donde se pone de manifiesto no sólo la particularidad que presentan estos sistemas de manejar los atributos visuales de las imágenes, sino también el hecho de que dichos atributos se extraigan automáticamente:

Sistemas CBIR: Modelo de recuperación visual de las imágenes, basado en el uso de características intrínsecas⁶ de los documentos que son **extraídas y representadas automáticamente** a través de estructuras de datos numéricas [ROBLEDANO, 2002].

2. EXPLICACIÓN DEL INTERÉS DE LOS SISTEMAS CBIR DESDE EL PUNTO DE VISTA DOCUMENTAL

El contexto en que se dan a conocer estos sistemas se inserta en un momento de explosión, que arranca fuertemente desde comienzos de la década de los '90, de los sistemas multimedia y de la digitalización de colecciones de imágenes disponibles a través de Internet y de otros medios⁷. Es difícil considerar una sola razón por la cual se originan estos sistemas, pero desde el punto de vista documental se suele señalar como motivo fundamental las limitaciones que, ante la nueva situación de crecimiento de la producción de imágenes digitales, comienzan a advertirse con más firmeza en relación con el modelo lingüístico tradicional aplicado a la recuperación de imágenes [ROBLEDANO, 1999]. Este modelo se basa en la representación textual de las imágenes y el empleo de criterios de recuperación también textuales. A pesar de que las técnicas de indización textual tienen muchas fortalezas, sobresaliendo la capacidad de las palabras para describir casi cualquier aspecto del contenido de una imagen, el proceso de indización manual presenta dos inconvenientes principales:

- Alto coste en términos de tiempo y dinero.
- Problemas de consistencia entre indizadores, e incluso de un mismo indizador, a la hora de determinar los temas.

Así, por ejemplo, entre las razones que aducen los autores AIGRAIN, ZHANG y PETKOVIC (1996) acerca de la necesidad de desarrollar técnicas para el proce-

⁶ En ROBLEDANO (2002) se entiende por características o contenido íntrico de la imagen "el conjunto de atributos derivados de las características materiales de la imagen" (*ob. cit.*, p.182)

⁷ Esta consideración se puede detectar en la práctica totalidad de la bibliografía consultada. Como ejemplo sirva la mención a los trabajos de RUI, HUANG y CHANG (1997; 1999); o LONG, ZHANG y DAGAN (2003).

samiento del contenido visual, se encuentra la de ofrecer acceso a grandes bases de datos de imágenes fijas superando las limitaciones de tiempo y coste que implica la producción humana de información descriptiva sobre las imágenes.

Con respecto al segundo inconveniente que mencionábamos arriba, los problemas de consistencia en la indización manual, a menudo se cita una idea expresada por ENSER (1995), por la que viene a establecer que **si recuperamos una imagen que no es predecible, entonces podemos deducir que su indización temática no es útil**. JAIN (1996) va más allá en este sentido, puesto que para él las palabras clave ofrecen con frecuencia una mejor descripción de la persona que las asigna que de la propia imagen a la que se refieren. A este respecto, VENDRIG (1997) comenta que la recuperación de las imágenes no debería tener que ver con lo que el indizador piensa que se representa en la imagen, sino con lo que el usuario asocia a la imagen. Esa asociación depende de la finalidad del que busca y no sólo varía en virtud del usuario, sino por cada sesión de recuperación en tanto y cuando la motivación del que busca puede cambiar. OGLE y STONEBRAKER (1995) refuerzan esta idea concluyendo que la tarea de indizar textualmente el contenido de las imágenes es muy **subjetiva** y **se enfrenta al cambio diacrónico del vocabulario** empleado como herramienta de trabajo. ROMER (1996) añade también como problemas la **falta de estándares para la denominación de los códigos visuales** y el **desigual entrenamiento visual entre los analistas**⁸.

Una síntesis similar a la ofrecida por los anteriores autores nos la aportan VENTERS y COOPER (2000b), quienes destacan la subjetividad a la hora de interpretar una imagen.

Por último, CORRIDONI, DEL BIMBO y VICARIO (1998) manifiestan un problema distinto a los planteados hasta ahora. Se trata de la **necesidad de superar** la limitación que supone la imposibilidad o gran **dificultad de expresar mediante palabras las cualidades gráficas y las sensaciones estéticas** que proporciona la percepción de una representación visual.

Esta última idea también la plantea ROBLEDANO (1999; 2002) al señalar que **el sistema de significación de la imagen es distinto al del texto lingüístico**, y que, en consecuencia, aspectos muy relevantes de la semántica de una imagen no podrán ser trasladados a descripciones lingüísticas. Como vemos, esto pone en entredicho precisamente una de las ventajas que se consideraba presentaba el modelo lingüístico: la capacidad de las palabras para describir casi cualquier aspecto del contenido de una imagen. Para contrarrestar el efecto de estas afirmaciones, tal vez podamos decir que a pesar de la ambigüedad y subjetividad ligadas a la asignación de texto a las imágenes, lo cierto es que este método de representación permite una recuperación semántica, y en aquellas áreas donde cada objeto tiene una única posible interpretación o descripción, es además una técnica de recuperación muy poderosa [VENDRIG, 1997; PICARD, 1995b].

⁸ Acerca de la falta de estándares para la denominación de los códigos visuales, habríamos de decir que en la actualidad el estándar MPEG-7 resuelve precisamente este problema, pero los sistemas CBIR no excluyen la posibilidad de incluir estos descriptores como complemento. Sirva de ejemplo a este respecto el sistema PICSOM [KOSKELA, LAAKSONEN y OJA, 2001] o el prototipo experimental ERIC7 [GAGNON, FOUCHER y GOUAILLIER, 2004], pues ambos utilizan metadatos de color, textura y forma.

Los sistemas visuales puros, que es como ENSER (1995) y ROBLEDANO ARILO (1999) denominan a los sistemas que permiten tanto la representación automática de los atributos visuales de la imagen como la recuperación a partir de esos criterios, intentan solucionar las debilidades impuestas por el modelo lingüístico. El problema de la inconsistencia en la indización se soluciona siendo el propio sistema quien extrae de forma automática los datos que representan a las imágenes; en consecuencia, no es necesario ningún entrenamiento visual por parte de los analistas, a la vez que permiten expresar cualidades gráficas. Sin embargo, también suponen ciertos problemas que se comentarán más tarde. El fundamental, que ya hemos adelantado, es el de la asignación de significados a los atributos formales extraídos de la imagen para permitir la recuperación a nivel semántico: por objetos, personas concretas y escenas, principalmente. También es cierto que este problema a menudo se resuelve incorporando sobre los sistemas visuales puros funciones de consulta lingüísticas. Es lo que ROBLEDANO ARILLO (1999) considera como un tipo de sistema mixto y una solución bastante acertada, opinión defendida por EAKINS (1996), entre otros muchos autores.

A nuestro juicio, los sistemas CBIR abarcan tanto sistemas visuales puros como sistemas mixtos fruto de la combinación de éstos con sistemas del tipo VL, LV o incluso con sistemas lingüísticos del tipo LL, siguiendo la tipología de sistemas automatizados de recuperación de imágenes establecida por Peter ENSER (1995)⁹, ya que, de forma aislada o en conjunción con herramientas de representación y recuperación textuales, los sistemas resultantes de dichas combinaciones emplearían atributos visuales para representar el contenido de las imágenes al tiempo que siempre permitirían la posibilidad de recuperar por dichos criterios, ajustándose con ello a la definición anteriormente dada de sistemas CBIR.

En todo caso, la integración de estas tipologías aboga por la automatización de los procesos para solucionar los problemas tradicionales e intentar un análisis y una recuperación lo más ajustada posible al contenido total de las imágenes. En consecuencia, y hoy por hoy, creemos que en la medida de lo posible se debe defender el uso conjunto de técnicas de análisis y recuperación visuales y conceptuales como método ideal para considerar en los procesos documentales tanto la naturaleza semántica como la naturaleza morfológica de las imágenes. Cómo adecuar esas combinaciones será un tema que nos corresponda resolver como profesionales de la Documentación en función de las características de los documentos a representar, de las necesidades informativas de los usuarios, de los tiempos de respuesta ante las consultas que el usuario precise y de las formas de difusión requeridas en cada caso; cuestiones todas éstas que, como bien nos recuerdan ROBLEDANO y MOREIRO (2002), han de estar siempre presentes si queremos conseguir sistemas de recuperación factibles.

⁹ Tipología compuesta por cuatro modelos, a saber: LL (representación y recuperación lingüísticas); LV (representación lingüística y recuperación visual); VL (representación visual y recuperación lingüística); VV (representación y recuperación visuales).

2.1. IMPORTANCIA Y FINALIDAD.

Decía JONES (1986) al referirse a la importancia de la indización automatizada frente a la indización humana aplicadas al documento textual, que el valor de la indización automatizada se incrementaría cuando la literatura de forma legible a máquina fuese más importante que la producida por medios tradicionales. Esta afirmación, tal como consideran MÉNDEZ y MOREIRO (1999), quizá se haya hecho presente no tanto en el sentido de que la producción de literatura legible por máquina sea más importante que la impresa, pero sí en cuanto a que **la naturaleza de la información ha cambiado y cada vez se presenta más en formato electrónico**. Ambas reflexiones, tanto la de Jones en 1986 como la perspectiva de Méndez y Moreiro en 1999, son perfectamente extrapolables al ámbito de la imagen. Lo evidente en el terreno documental es que la indización automatizada de imágenes ha venido a centrar la atención de un gran número de estudios cuando la digitalización y los sistemas multimedia, hace ya más de una década, aumentaron la disponibilidad de las colecciones de imágenes y supusieron con ello una solución al problema entre preservación y acceso a la información.

Que el modelo tradicional de indización manual de las imágenes impone limitaciones a la explotación de las mismas como recursos de información es algo ampliamente reconocido, y que la digitalización acentúa los problemas de tiempo invertido en la asignación de texto a las imágenes, es un hecho aún más palpable. Para solventar estos obstáculos, los avances de los últimos años se han centrado en la indización automática y la recuperación de las imágenes por su contenido visual.

Las razones para optar por una u otra solución en el campo de la imagen se han expuesto en el apartado precedente, en el que hemos subrayado las dificultades de la indización manual que han contribuido a una mayor atracción hacia los sistemas CBIR; máxime si tenemos en cuenta que el procesamiento automático del lenguaje natural no resuelve todos los problemas de la indización manual; por ejemplo, es sensible a la ambigüedad que provoca el hecho de asignar una palabra clave a un cierto contexto, obviando otras palabras que pueden ser igualmente aceptables u otros contextos [VENDRIG, 1997]. Precisamente, y como ya hemos comentado antes, en una imagen ha de primar la idea que el usuario asocia a dicha imagen, y no lo que un indizador (automático o humano) ha interpretado que representa. Aún así, hemos dejado entrever que las soluciones más viables en la actualidad pasan por la conjunción de técnicas en función del fondo a consultar.

Para Xiangyu JIN (2002), la necesidad del CBIR se resume en tres ideas que podemos sumar a lo ya comentado y que entroncan directamente con su aplicación a Internet:

- La creciente cantidad de imágenes digitales.
- El hecho de que la Web es un recurso abierto.
- Que los motores de búsqueda actuales se basan en el título de las imágenes, con la limitación que esto supone.

Podemos explicar estas apreciaciones de Xiangyu Jin considerando el fenómeno de desintermediación que ha supuesto Internet al convertirse actualmente en el medio generalizado de acceso a la información gráfica. Las implicaciones que ello conlleva desde el punto de vista del diseño de nuevos sistemas de recuperación de imágenes son diversas [ROBLEDANO y MOREIRO, 2002]: por lo pronto, supone una amplia tipología de usuarios que potencialmente pueden acceder a las distintas colecciones disponibles a través de la Web; el usuario es ahora, además, un usuario activo, que realiza sus propias búsquedas; esta situación conduce al intento de desarrollar sistemas que conjuguen potencia en la recuperación y facilidad de uso; para ello, una posible solución es la de diseñar entornos de descubrimiento sobre sistemas que permitan distintos niveles de descripción de las imágenes y distintas posibilidades de búsqueda que satisfagan la enorme variedad de necesidades manifiestas.

Evidentemente, el interés que suscitan las técnicas CBIR es considerable y dicho interés puede defenderse desde distintas perspectivas. La posibilidad de representar y recuperar visualmente una imagen también plantea interrogantes, como por ejemplo qué campos científicos serán los más beneficiados de su uso. El IIDR señala como principales destinatarios las galerías de arte, museos, periódicos y bibliotecas de diapositivas, así como en general todas aquellas organizaciones cuyo trabajo reside en el tratamiento de colecciones de imágenes. Lo cierto es que las áreas que han recibido ya una atención directa han sido aquéllas para las que el reconocimiento de características formales de la imagen es fundamental: **fondos de patentes, logotipos comerciales, sellos, diseños industriales, imágenes médicas, identificación de huellas digitales, reconocimiento de rostros**, entre otras. En el terreno de la Biblioteconomía y la Documentación interesaría especialmente su integración en **bases de datos de imágenes o multimedia** y su aplicación en **bibliotecas digitales**.

Hemos de remarcar, no obstante, la sugerencia hecha ya por ROBLEDANO ARILLO (1999; 2002) de **aumentar la oferta de sistemas de recuperación automáticos con propósito de aplicación general**. Es una idea formulada también por FROST et al. (1997) en su defensa de desarrollar sistemas para el usuario casual. Dado que nuestra exposición a la información gráfica es más que evidente, es difícil determinar qué contextos científicos o qué personas no necesitan en algún momento la facilidad de acceder a determinadas imágenes digitales, bien sea por una situación profesional o personal. Decía además PEJTERSEN (1995) **que el ser humano es particularmente bueno estableciendo similitudes entre imágenes por sus atributos intrínsecos**; utilizamos millones de conexiones entre imágenes muchas de ellas en virtud del color, textura, forma, orientación, etc. En este sentido, la posibilidad de interrogar por atributos visuales podría estar al alcance de cualquiera, y en consecuencia, se hace necesario establecer de qué forma y en qué momentos ha de ponerse a disposición esta opción de recuperación. Las técnicas CBIR plantean así la importancia de pararnos a reflexionar sobre las necesidades informativas y comportamiento de búsqueda de los usuarios con respecto a las imágenes, aspectos éstos mucho menos estudiados que en lo que a la búsqueda de información textual se refiere. Dicho de otro modo, el advenimiento de sistemas CBIR

ha motivado, y **ha de seguir haciéndolo en mayor medida**, el desarrollo de estudios relacionados con la recuperación y uso de las imágenes, entre otras cuestiones, aportándonos datos de interés al tiempo que generan nuevos planteamientos y líneas de investigación.

En conclusión, podemos decir que la importancia de las técnicas CBIR radica en que no son excluyentes de otros métodos, sino que pueden funcionar en conjunción con las formas habituales de representación y recuperación de imágenes, aportan soluciones a problemas propios del paradigma tradicional de análisis y recuperación, así como también suscitan nuevos intereses en el estudio de la recuperación de imágenes.

3. FORTALEZAS Y DEBILIDADES DE LA INVESTIGACIÓN

Ya EAKINS y GRAHAM (1999) señalaban como temas que cuentan con menos apoyo en el terreno de la investigación sobre CBIR, los de diseño de interfaz, evaluación de la eficacia de estos sistemas y estudios de usuarios. También ROBLE-DANO ARILLO (2002) demandaba más investigación sobre necesidades de usuarios en lo que a las imágenes se refiere, a fin de poder adecuar los sistemas de recuperación a dichas necesidades.

Efectivamente, tanto las necesidades de información como el comportamiento de búsqueda de los usuarios finales son aspectos de los que se sabe muy poco, al menos en comparación con el campo de la recuperación de información textual. Sobre esta materia, EAKINS y GRAHAM (1999) identifican y presentan varios estudios al respecto que se centran en colecciones de imágenes y poblaciones de usuarios específicos. Concluyen que, aunque los resultados ofrecen una valiosa visión, son demasiado limitados en su intento de extrapolarse a un marco teórico global. No obstante, hemos de destacar todos los estudios realizados en esa línea por el Instituto para el que ambos autores trabajan, el ya citado IIDR.

En líneas generales y de forma sintética, tal como manifiestan VENTERS, HARTLEY y HEWITT (2004, p. 323), podemos concretar que las investigaciones relativas al estudio de la recuperación de imágenes por contenido se han dirigido en su inmensa mayoría al desarrollo y avance de algoritmos de recuperación, esfuerzos que, posiblemente, han dado éxitos muy limitados. Sin embargo, se han ignorado en gran medida otros aspectos fundamentales para que esta tecnología pueda dar lugar a herramientas viables de recuperación, aspectos como los arriba mencionados de detección de necesidades y modelos humanos de comportamiento de búsqueda, o la especificación de las interfaces de usuario que, a pesar de haberse considerado reiteradamente como una de las principales áreas de investigación para hacer global y útil la tecnología CBIR, en realidad ha dado aún muy pocos frutos. Quizá una de las llamadas de atención a este respecto más significativas la hacen EAKINS, BRIGGS y BURFORD (2004) al recordarnos que sin un conocimiento real de las necesidades cognitivas y contextuales del usuario final, muchos de los sistemas actuales se arriesgan a fracasar, y por ello han de cobrar importancia los principios de diseño centrados en el usuario.

Para Sagarmay DEB (2004), el punto débil en el intento de desarrollar un modelo universal de sistema CBIR se halla en que sigue sin resolverse la recuperación de significados semánticos o significados de alto nivel. Esta dificultad reside en la importante laguna que aún existe entre la percepción humana y la percepción de un ordenador.

Sin embargo para JAIN, en 1993, el fracaso en el éxito de estos sistemas podía deberse a dos hechos que se daban en el conjunto de los estudios existentes entonces: o se trataban aspectos básicamente teóricos, o aspectos relativos a campos de aplicación demasiado concretos. En este sentido, RUI, HUANG y CHANG (1999) opinan que sólo cuando haya una integración homogénea de las distintas actividades de investigación podrá lograrse el objetivo de una recuperación CBIR efectiva.

En la actualidad, se puede afirmar que la tendencia hacia dicha integración es cada vez mayor, y prueba de ello es la dificultad para aislar los temas que tratan simultáneamente los diferentes estudios sobre CBIR. Se intenta impulsar una mejora de las representaciones de las imágenes, los modelos de datos y los algoritmos de procesamiento, al tiempo que ha aumentado la preocupación por el diseño de interfaces de interrogación inteligentes y la arquitectura de sistemas independientes del dominio [DEB, 2004]. Se observa que crece la interdependencia entre distintos aspectos, y tal vez ello conlleve una disminución en el número de trabajos que se publiquen, pero un mayor cuidado y atención a la diversidad de cuestiones a considerar para superar los retos demandados.

4. LA POSICIÓN DE LA DOCUMENTACIÓN EN EL ESTUDIO DE ESTOS SISTEMAS

En el capítulo V de su tesis doctoral, el profesor Jesús ROBLEDANO (1999¹⁰) señala que es imprescindible para los objetivos de su estudio, el conocimiento del estado de la cuestión respecto a la tecnología aplicable a la recuperación de imágenes. De esta forma, viene un tanto a justificar el hecho de que vaya a tratar de cuestiones tecnológicas, como dirigiéndose a quienes le fueran a responder que no es menester de la Documentación tales asuntos. Va incluso más lejos en su justificación, pues considera que es necesario que exista una integración entre dos perspectivas que han estado aisladas suponiendo así un problema asociado al estudio de la recuperación de imágenes. Se refiere, como también lo hicieran en su momento Enser y Cawkell, a “las técnicas de recuperación conceptual, tradicionalmente situada dentro de la órbita de la teoría y práctica de las Ciencias de la Documentación; y el campo de la recuperación automática de imágenes, abordada desde el ámbito de las tecnologías de tratamiento de la imagen digital” [ROBLEDANO, 1999,

¹⁰ El texto de su tesis publicado en ROBLEDANO (2002). En dicha publicación el capítulo cinco que comentamos aparece como capítulo tres, titulado: “La recuperación de la imagen fija. Perspectiva funcional de los sistemas automatizados de recuperación de imágenes”, pp. 147-200. El objetivo final del trabajo de Robledano es la propuesta de un modelo conceptual basado en un lenguaje orientado a objetos para representar el contenido de las fotografías de prensa.

p.268]. Esta misma idea ENSER (1995) la expresaba diciendo que, ante la ubicuidad de la imagen digital, cabe esperar que los productos informativos que explotan el medio visual provengan de un círculo de profesionales cada vez más amplio que el tradicionalmente asociado con la profesión de la información; y, en consecuencia, los expertos procedentes de estos diversos campos han de intentar comprenderse mutuamente. Con anterioridad, CAWKELL (1993) ya notó esta falta de comunicación entre la comunidad investigadora relacionada con la recuperación de la imagen, identificando dos grupos de investigadores diferenciados en cuanto a su interés por la gestión de la imagen: uno relacionado con el reconocimiento y procesamiento de la imagen; y otro, con la recuperación y clasificación.

MÉNDEZ RODRÍGUEZ (2002) plantea el asunto constatando que la recuperación de la información es inmanente a la disciplina documental, pero también es tarea común compartida por otras actividades/disciplinas, de ahí que los profesionales y las aportaciones implicadas en este desarrollo no vengán únicamente de una sola dirección. Y, no sólo eso, sino que la recuperación de la información ha de apoyarse necesariamente en los avances propuestos desde otras áreas para mejorar la eficiencia de los sistemas automatizados.

En definitiva, aunque el fin último del proceso documental sea la difusión de la información, al estudiar un sistema de recuperación de información no podemos olvidar que éste se constituye de elementos tecnológicos, organizativos y humanos. Con respecto a los elementos tecnológicos, no se ha de pretender que el documentalista se convierta en matemático o informático. El desarrollo tecnológico ha provocado la convergencia de una serie de profesionales en el estudio de una misma entidad, que en este caso son los sistemas automatizados de recuperación de imágenes. Ante esta situación, hemos de asumir el papel que a cada profesional le corresponde, conociendo no obstante las aportaciones que el resto de disciplinas puede suponer a la recuperación efectiva de la información y, en nuestro caso, de las imágenes digitales.

La inmensa mayoría de las aportaciones al estudio de los sistemas CBIR, desde su origen mismo, proviene del ámbito de la visión automatizada y el procesamiento de imágenes. No podemos obviar tales contribuciones, y debemos tener un conocimiento mínimo de las técnicas que habitualmente se emplean para representar los atributos gráficos, pues conociendo las posibilidades que ofrece la técnica está el documentalista en disposición de hacer propuestas, al menos teóricas, acerca de modelos de representación de las imágenes y de definir aspectos funcionales de los sistemas que se adecuen a las necesidades de los usuarios.

El modelo de recuperación visual de las imágenes parece haber despertado un escaso interés en nuestra profesión, pues son mínimas las aportaciones al respecto. Sin embargo, se detecta que las debilidades que sufre la investigación actual en el ámbito de los sistemas CBIR podrían abordarse claramente desde nuestro entorno profesional. El ejemplo en este sentido nos viene, tanto de mano de las actividades del IIDR británico, como de las exposiciones de Robledano en nuestro país. Sin embargo, ¿qué es lo que parece mantener alejados a los departamentos y centros especializados en Biblioteconomía y Documentación del estudio de los sistemas CBIR? El hecho de que, hoy por hoy, los sistemas CBIR no sean utilizados a

escala significativa, no ha de ser un aspecto disuasorio de su estudio, sino todo lo contrario.

Los avances tecnológicos son en la actualidad un foco de estudio muy productivo, pero no así la comprensión que tenemos acerca de las necesidades de los usuarios de imágenes y su comportamiento de búsqueda, ni sobre el diseño de interfaz más adecuado, o cómo integrar más convenientemente la recuperación textual y visual, entre otros aspectos. La Documentación ha de aportar al estudio de los sistemas CBIR la dimensión humana que hasta ahora menos se ha reflejado; hemos de centrar la atención en el usuario, en sus estrategias de recuperación, en su forma de interacción, en definitiva, en sus necesidades de todo tipo para con el sistema.

Podemos sintetizar algunas de las aportaciones con las que la Documentación podría contribuir al diseño de un sistema CBIR, planteando algunas preguntas cuya respuesta se enfocará en función del ámbito de aplicación del sistema:

- ¿Qué rasgos visuales de búsqueda son más útiles?
- ¿Cuáles son las técnicas de búsqueda y las estructuras de indización más adecuadas?
- ¿Cómo se formalizan las necesidades de información sobre las imágenes y cómo se plantean las búsquedas?
- ¿Cómo interactúa el usuario con el sistema?
- ¿Cómo se valoran los resultados obtenidos?
- ¿Qué características individuales influyen en el proceso?
- ¿Cómo debería ser un “lenguaje visual” adecuado?

Los conocimientos y el enfoque documental son imprescindibles para conseguir el diseño de sistemas de recuperación cada vez más amigables. La Documentación ha de estudiar las distintas variables que se ponen en juego a la hora de emplear un sistema de recuperación (características del sistema, características de los usuarios, entorno y proceso de búsqueda, resultados) de forma interconectada, tendiendo a conseguir una sintonía perfecta entre dichos factores.

La Documentación puede y debe aportar avances en temas tan poco tratados como el estudio de las necesidades de información y comportamiento de búsqueda del usuario de imágenes, el diseño de interfaces que deberían desarrollarse para estos sistemas, o incluso acerca de las cuestiones relativas a la evaluación de su eficacia y eficiencia, pues son todos éstos temas determinantes para garantizar el éxito de dichos sistemas de recuperación, y son, al mismo tiempo, aspectos genéricos que los propios planes de estudio de nuestra disciplina reflejan como objeto de nuestra atención bajo el nombre de asignaturas tales como: Planificación y diseño de sistemas de información y documentación; Evaluación de sistemas y usuarios de información y documentación; o Procesamiento automático del conocimiento.

¹¹ Prueba obvia de la necesidad de una mayor integración y relación entre estas dos disciplinas –la Documentación y la Informática– es la creación de estudios combinados a tal efecto, como son los existentes, por ejemplo, en la Universidad Carlos III de Madrid.

El tratamiento de la información es un ámbito interdisciplinar, pero eso no ha de significar que la Documentación tiene menos que aportar en un contexto tecnológico como el que vivimos de continuo, que la Informática¹¹. Cada disciplina tiene un cuerpo de conocimientos y unas técnicas que han de ser readaptadas y aplicadas a su objeto de estudio –en nuestro caso, el tratamiento de la información y, en consecuencia, toda clase de sistemas de recuperación de información–.

Por último, son muchas las aplicaciones relevantes que pueden idearse para la integración de los sistemas CBIR en unidades informativas. En cualquier caso, estos sistemas suponen una importante contribución al reto documental de establecer nuevos niveles de descripción de las imágenes y nuevas posibilidades de búsqueda, permitiendo así satisfacer nuevas demandas informativas. La opción de configurar redes distribuidas de información que combinen técnicas conceptuales y visuales para la recuperación de los documentos fotográficos o gráficos, la puesta en red del patrimonio cultural de una institución, la recuperación de fondos bibliográficos a partir de imágenes que recordemos de ellos, etc., son algunas propuestas interesantes que se han comenzado a explorar y que requieren la participación de nuestros profesionales.

5. RESEÑAS BIBLIOGRÁFICAS¹²

- AIGRAIN, P.; ZHANG, H.; PETKOVIC, D. (1996). **Content-based representation and retrieval of visual media: a state-of-the-art review**, *Multimedia Tools and Applications*, vol. 3 (3), pp. 179-202.
- CAWKELL, A.E. (1993). **Indexing collections of electronic images: a review**, London: British Library Research & Development Department, The British Library Board. [*British Library Research Review*, 15], 43 pp.
- CORRIDONI, J.; DEL BIMBO, A.; VICARIO, E. (1998). **Image Retrieval by Color Semantics with Incomplete Knowledge**, *Journal of the American Society for Information Science*, 49 (3), pp. 267-282.
- DEB, Sagarmay; ZHANG, Yanchun (2004). **An overview of Content-based Image Retrieval Techniques**, *Proc. of the 18th International Conference on Advanced Information Networking and Applications (AINA'04)*, pp. 59-64.
- EAKINS, John P. (1996). **Automatic image content retrieval –are we getting anywhere?** [documento PDF], disponible desde: <http://citeseer.ist.psu.edu/eakins96automatic.html>
- EAKINS, J.P.; GRAHAM, M.E. (1999). **Content-based image retrieval** [documento HTML], *Report to JISC Technology Applications Programme, January*, disponible en <http://www.unn.ac.uk/iidr/report.html> [Consulta: 27/09/04]

¹² Todos los recursos electrónicos que se han manejado, y a los cuales se hace referencia en este listado bibliográfico, se han revisado por última vez el 31 de diciembre de 2005. Por ello, no se reseña la fecha de consulta, como es prescriptivo. Sólo si el recurso ya no está operativo es cuando se ha consignado la fecha original de consulta.

- EAKINS, John P.; BRIGGS, Pam; BURFORD, Bryan (2004). **Image retrieval interfaces: A user perspective**, *CIVR 2004 (Peter Enser et al., eds.)*, pp. 628-637.
- ENSER, P.G.B. (1995). **Progress in Documentation. Pictorial Information Retrieval**, *Journal of Documentation*, 51 (2), June, pp. 126-170.
- FROST, C. et al. (1997). **Research on Human-centered design and evaluation of multimode image retrieval** [documento PDF], *FOURTH DELOS WORKSHOP: Image indexing and retrieval*, San Miniato, Pisa, Italy, 28-30 August, *ERCIM Workshop Proceedings* nº 97-W004. Disponible en: <http://www.ercim.org/publication/ws-proceedings/DELOS4/frost.pdf>
- GAGNON, L.; FOUCHER, S.; GOUAILLIER, V. (2004). **ERIC7: An Experimental Tool for Content-Based Image Encoding and Retrieval under the MPEG-7 Standard**, [documento PDF] *Proceedings of the Winter International Symposium on Information and Communication Technologies (WISICT 2004)*, Hyatt Regency, Cancun, Mexico, January 5-8, disponible en: http://www.crim.ca/rd/publications/VIX_GagLals_WISITC04AbA.pdf [Consulta: 08/09/04]
- GUDIVADA, V.N.; RAGHAVAN, V.V. (1995). **Content-based image retrieval systems**, *IEEE Computer*, vol. 28 (9), pp. 18-22.
- JAIN, R. (1993). **NSF workshop on visual information retrieval management systems**. *Storage and Retrieval for Image and Video Databases: Proceedings of the International Society for Optical Engineering*, San Jose, CA, USA, (February 2-3), pp. 198-218.
- JAIN, R. (1996). **Infosopes: multimedia information systems**, *Multimedia Systems and Techniques*, B. Furht (ed.), Boston: Kluwer Academic Publishers, pp. 217-253.
- JIN, Xiangyu (2002). **Content-based image retrieval: CBIR** [documento PDF], disponible en: <http://www.cs.virginia.edu/~xj3a/publication/CBIR.pdf>
- JONES, Kevin P. (1986). **Getting Started in Computerized Indexing**. *The Indexer*, 15 (April), pp. 9-13.
- KATO, T. (1992). **Database architecture for content-based image retrieval**, en *Image Storage and Retrieval Systems (Jambardino, A.A. & Niblack, W.R., eds.)*, Proc SPIE 2185, pp. 112-123.
- KOSKELA, Markus; LAAKSONEN, Jorma; OJA, Erkki (2001). **Self-organizing image retrieval with MPEG-7 descriptors** [documento PDF], *Proceedings of Infotech Oulu International Conference on Information Retrieval (IR'2001), Oulu (Finland), September*, disponible en: <http://www.cis.hut.fi/picsom/ir2001.pdf>
- LONG, Fuhui; ZHANG, Hongjiang; DAGAN FENG, David (2003). **Fundamentals of content-based image retrieval** [documento PDF], disponible en: http://www.research.microsoft.com/asia/dload_files/group/mcomputing/2003P/ch01_Long_v40-proof.pdf
- MÉNDEZ RODRÍGUEZ, Eva M^a. (2002). **Metadatos y recuperación de información: Estándares, problemas y aplicabilidad en bibliotecas digitales**, Gijón: Trea, 429 pp.
- MÉNDEZ RODRÍGUEZ, Eva M^a; MOREIRO GONZÁLEZ, José Antonio (1999). **Lenguaje natural e indización automatizada**, en *Ciencias de la Información*, vol. 30 (3), septiembre, pp. 11-24.

- OGLE, V.E.; STONEBRAKER, M. (1995). **Chabot: retrieval from a relational database of images**, *IEEE Computer*, vol. 28 (9), pp. 40-48.
- PEJTERSEN, Annelise Mark (1995). **Cognitive engineering in information retrieval domains: Merging paradigms?** [documento PDF], *Bibliothek*, 19 (1), disponible en: http://www.bibliothek-saur.de/1995_1/64-68.pdf
- PICARD, R.W. (1995b) **Toward a visual thesaurus** [documento PDF], *Springer Verlag Workshops in Computing, MIRO 95*, disponible desde: <http://citeseer.ist.psu.edu/picard95toward.html>
- ROBLEDANO ARILLO, Jesús (1999). **La recuperación de la imagen fija. Perspectiva funcional de los sistemas automatizados de recuperación de imágenes**, En *El Análisis de la fotografía de prensa en entornos automatizados* (Tesis Doctoral), pp. 265-310.
- ROBLEDANO ARILLO, Jesús (2002). **El tratamiento documental de la fotografía de prensa: sistemas de análisis y recuperación**. Madrid: Archiviana, 2ª ed., 403 pp.
- ROBLEDANO ARILLO, Jesús; MOREIRO GONZÁLEZ, José Antonio (2002). **La recuperación documental del documento fotográfico: perspectiva tecnológica y documental**. *Primeras Jornadas de Imagen, Cultura y Tecnología* (Universidad Carlos III, Madrid. 1-5 julio)
- ROMER, Donna M. (1996). **Image and Multimedia Retrieval** [documento HTML], *Research Agenda for Cultural Heritage on Information Networks. The Getty Art History Information Program*, disponible en: <http://www.gii.getty.edu/agenda/image.html> [Consulta: 09/09/04]
- RUI, Y.; HUANG, T.S.; GHANG, S-F. (1997). **Image retrieval: Past, present, and future** [documento PDF], *Proc. of Int. Symposium on Multimedia Information Processing*, Dec., disponible desde: <http://citeseer.ist.psu.edu/192987.html>
- RUI, Y.; HUANG, T.S.; GHANG, S-F. (1999). **Image retrieval: Current techniques, promising directions, and open issues**. *Journal of Visual Communication and Image Representation*, 10 (1), pp. 39-62.
- URCID PLIEGO, Paulo (2003). **Búsqueda de imágenes por contenido en bibliotecas digitales (tesis)** [documento PDF], *Universidad de las Américas-Puebla (México), Escuela de Ingeniería*, disponible en: http://www.pue.udlap.mx/~tesis/lis/urcid_p_p/
- VENDRIG, J. (1997). **Filter Image Browsing: A study to image retrieval in large pictorial databases** [documento HTML], *Department of Intelligent Sensoric Information Systems, Faculty WINS, Universiteit van Amsterdam*, disponible en: <http://carol.wins.uva.nl/~vendrig/thesis/>
- VENTERS, Colin C.; COOPER, Matthew (2000b). **Content based image retrieval: A new paradigm for image retrieval?** [documento PDF] *Spectra, Fall Issue November*. <http://mvc.man.ac.uk/staff/venters/pubs/spectra.pdf> [Consulta: 19/09/04]
- VENTERS, C.C.; HARTLEY, R.J.; HEWITT, W.T. (2004). **Mind the Gap: Content-Based Image Retrieval and the User Interface**, en: DEB, S. (ed.) *Multimedia Systems and Content-based Image Retrieval*, Hershey: Idea Publishing, pp. 322-355.