

DORIS: Sistema para la Recuperación de Imágenes de Piezas Mecánicas y de Automoción Utilizando Descriptores de Textura

DORIS: System for Retrieval of Images of Mechanics and Automotive Parts by Using Texture Descriptors

Gloria E. Jaramillo¹, John W. Branch Bedoya²

Escuela de Sistemas, Facultad de Minas, Universidad Nacional de Colombia - sede Medellín, Colombia
gejarami@unalmed.edu.co; jwbranch@unalmed.edu.co

Recibido para revisión 29 de Febrero de 2008, Aceptado 19 de Mayo de 2008, Versión final 28 de MAyo de 2008

Resumen—En este artículo, se propone un sistema para la recuperación de piezas mecánicas y automotrices basándose en su contenido visual, específicamente se aborda un enfoque basado en textura no homogénea como descriptor para el contenido de las imágenes. En la realización de dicho sistema se utiliza una base de datos compuesta por 897 imágenes agrupadas semánticamente en 23 categorías. Se presenta la precisión como medida de desempeño en la recuperación.

Palabras Clave—CBIR, Descriptor de Textura no Homogénea, Indización, Sistemas Gestores de Bases de Datos, SGBD, MPEG-7.

Abstract—In this article, a system for retrieving of mechanics and automotive parts based on image content is proposed, specifically, we adopt an approach based on no-homogeneous texture as descriptor for the content of images. In the development of such a system we use a database composed of 897 images semantically grouped in 23 classes. We present the precision as an evaluation measure in the retrieval.

Keywords—CBIR, no-Homogeneous Texture Descriptor, Indexing, DBMS – Database Management System, MPEG-7.

I. INTRODUCCIÓN

La recuperación de información textual se ha convertido en una de las ramas de investigación más dinámicas desde aproximadamente 1940, impulsado principalmente, por una creciente necesidad de acceso rápido y eficiente a información distribuida. Con la masificación de los dispositivos analógicos y electrónicos de captura y manipulación de contenido audiovisual, se da paso a la creación de las bases de datos multimediales. Es así, como desde finales de 1970 se dispone de una gran cantidad de información visual en diferentes formatos, la cual se hace necesario no sólo almacenar, sino también, contar con métodos que permitan su acceso y recuperación.

En la actualidad, las prácticas utilizadas para la recuperación de información visual se pueden agrupar en tres categorías: libre navegación, basada en texto y basada en el contenido de la imagen – CBIR. La libre navegación o free-browsing es útil en bases de datos o colecciones no orientadas a dominios específicos, debido a que para este tipo de búsqueda el usuario no tiene bien definido un criterio de consulta. El método de recuperación empleado en este enfoque, consiste en “navegar” a través de todo el contenido de la base de datos y detenerse al encontrar el(los) objeto(s) de interés. Un ejemplo de esta técnica de recuperación de información visual es Picasso™ [14] para computadores personales, este programa permite organizar una colección de imágenes, pero para recuperar la información visual que ellas representan es necesario navegar por toda la colección almacenada. En segundo lugar, se encuentra la recuperación de imágenes basada en texto, la cual surge a raíz de los grandes avances logrados en la recuperación textual y en los Sistemas Gestores de Bases de Datos (SGBD). La recuperación basada en texto consiste en la asignación manual del contenido semántico de las imágenes (características de alto nivel) por medio de palabras claves o de la extracción del pie de la imagen, el cual es asignado como índice. Posteriormente, es posible realizar una consulta basada en parámetros textuales, tomando como medida de similitud el índice de la imagen. Aunque este enfoque ha sido ampliamente utilizado en motores de búsqueda tan populares como Google, Alta Vista y Lycos, presenta dos grandes limitaciones:

1. Para al año 2000 se calculaba la cantidad de imágenes de la Web en aproximadamente 489 millones, siendo añadidas nuevas imágenes a una tasa de un millón por día. Para el 2003, Google había indizado alrededor de 425 millones de imágenes [19]; lo cual da una idea del costo humano y económico de una anotación manual para todas las imágenes contenidas en la World Wide Web.

2. Debido a que esta anotación es llevada a cabo de forma manual, la recuperación de imágenes podría ser inexacta debido principalmente a la subjetividad del contenido semántico que puede ser anotado en una descripción. Además, en algunos casos es posible que el pie de la imagen no refleje exactamente el contenido de ésta.

Como una solución a estas limitaciones, se han fusionado técnicas de Procesamiento Digital de Imágenes, Visión por Computador y Bases de Datos para lograr una descripción efectiva de una imagen que dependa de su propio contenido, dando origen a técnicas de Recuperación de Imágenes por Contenido o CBIR (Content-Based Image Retrieval). Este enfoque hace uso de un conjunto de herramientas capaces de detectar automática o semi-automáticamente características visuales de una imagen (características de bajo nivel) como color, textura y forma, evitando la subjetividad al utilizar valores numéricos en las descripciones y haciendo mucho más sencillo la extracción de características por medio de algoritmos no supervisados.

El presente artículo tiene como objetivo proponer un sistema de recuperación de imágenes mecánicas y automotrices, a partir de la detección de discontinuidades como descriptor de textura. El resto del trabajo se estructura como sigue: la sección dos presenta algunos trabajos previos realizados en la recuperación de contenido audiovisual. Posteriormente, se exponen los conceptos relacionados con los descriptores de textura y la manera como fueron aplicados al sistema; luego se presentan los resultados obtenidos con esta propuesta y finalmente las conclusiones y trabajo futuro.

II. TRABAJOS PREVIOS

En el campo de la recuperación de imágenes por contenido han sido numerosas las propuestas y diseños que se han presentado tanto a nivel comercial como académico. En la actualidad se distinguen principalmente cuatro sistemas comerciales: QBIC, VIR, Excalibur e Imatch.

QBIC (Query by Image Content) [2], es tal vez el sistema de recuperación de imágenes más conocido. Este sistema creado por IBM soporta entradas tanto de imágenes estáticas como en movimiento para la extracción de características de color, textura y forma. Aunque algunas aplicaciones que utilizan el sistema QBIC para la búsqueda de contenido visual requieren segmentación manual de las regiones de la imagen, versiones posteriores han implementado algoritmos que permiten la segmentación automática. El sistema Hermitage [20], para la recuperación de imágenes de piezas de arte, es una de las aplicaciones que ha incorporado la búsqueda de color (QBIC Colour Search) y la búsqueda por composición (QBIC Layout Search) de QBIC, permitiendo al usuario especificar el criterio de búsqueda creando proporciones de color desde un espectro o dibujando áreas de color por medio de formas geométricas, aproximando a la descripción visual de una obra de arte.

En el 2000, luego de la fusión de la empresa Virage con el Grupo Autonomy se da la comercialización de nuevas tecnologías para el manejo del contenido de medios ricos

en información y sistemas avanzados de reconocimiento, desarrolladas por estas compañías. Las tecnologías desarrolladas por Virage tienen como objetivo permitir a los usuarios entender e interpretar el potencial de grandes volúmenes de información, como es el caso de imágenes almacenadas en bases de datos. El sistema VIR Image Engine de Virage captura, codifica e indiza el contenido de fuentes de imágenes, video y audio. Las aplicaciones más conocidas de VIR han sido el AV Photo Finder y Virage VideoLogger [22], incorporado al motor de búsqueda de Altavista. Una de las grandes desventajas de este sistema es la imposibilidad de asignar pesos relativos a los criterios de búsqueda (color, textura y forma) [21].

Excalibur Visual RetrievalWare[15], consiste en un conjunto de API's de Java y C++ para la manipulación y construcción de aplicaciones para el procesamiento, indización y recuperación de imágenes. Uno de estos API's es el CST (Color - Shape - Texture), el cual extrae características como orientaciones relativas, histogramas HSV, curvatura y asperezas de los objetos de la imagen [21]. En Excalibur las imágenes son mostradas sin un orden específico, pero a diferencia de Virage, permite asignar niveles de importancia relativa a los criterios de búsqueda. Image Assets es una página web que ha incorporado con éxito la tecnología de Excalibur.

Imatch [5], es tal vez una de las versiones comerciales menos difundidas, esta herramienta permite indizar todos los archivos visuales de un computador personal; esta orientado al público en general y no indiza dominios específicos. Permite recuperación de imágenes por medio de descriptores de color, textura y forma.

Además de los sistemas comerciales, numerosas instituciones académicas se han dedicado a la realización de sistemas CBIR, motivados principalmente por la evaluación y creación de nuevas técnicas de extracción. Algunos de los sistemas más conocidos son Photobook [13] que permite segmentación automática y técnicas sofisticadas para búsqueda por textura, además de incluir técnicas de extracción para forma, y rasgos del rostro. VisualSEEK [16], desarrollado por Columbia University, y MARS [3] (Multimedia Analysis and Retrieval Systems), desarrollado por la Universidad de Illinois, el cual incluye técnicas de retroalimentación que brindan la posibilidad de refinar la búsqueda a partir de una primera recuperación.

En el marco del desarrollo del estándar MPEG-7, para la regularización de esquemas XML que permitieran un lenguaje común de anotación de metadatos de información audiovisual, se creó un software de referencia para la validación del estándar. Dicha aplicación contiene una serie de herramientas desarrolladas en C++ para la recuperación de información visual, entre estas se encuentra una propuesta de extracción de textura por medio de histogramas de bordes [23], dicho enfoque define cinco tipos de direccionalidades: cuatro orientaciones de bordes (horizontal, vertical, diagonal de 45°, diagonal de 135°) y una representación de textura suave. Aunque este enfoque ha servido de pauta para la realización de muchos trabajos de recuperación de información visual, uno de los principales inconvenientes para la utilización de los algoritmos

de extracción propuestos por el MPEG-7 como parte del estándar, es la imposibilidad de operación efectiva de dichas técnicas bajo cualquier contexto, esto debido a que muchas de las etapas (un ejemplo de ello es la etapa de preprocesamiento) esta condicionada a las características particulares de la base de datos. De allí, que muchas de las aplicaciones orientadas a dominios no específicos fallen en la recuperación.

III. SISTEMA DORIS

El sistema DORIS (Domain- ORiented Image Searcher) presenta una propuesta para la recuperación de imágenes de piezas mecánicas y automotrices, definiendo histogramas para dos tipos de discontinuidades: líneas y bordes. DORIS permite al usuario especificar una imagen ejemplo a la cual se le extraen características de textura no homogénea, que luego son comparadas con las características de textura de las imágenes de la base de datos que han sido previamente extraídas.

El descriptor de textura no homogénea describe un histograma por medio de cambios en el brillo de la imagen. En la implementación de extracción de características de textura para el sistema DORIS se abordó un enfoque de detección de discontinuidades, lo cual se lleva a cabo aplicando filtros que refuerzan los píxeles representativos de las discontinuidades, en nuestro caso se pretenden detectar discontinuidades de líneas y bordes, por lo que se utilizan filtros se acentúen las líneas y máscaras representando ocho direcciones (0°, 45°, 90°, 135°, 180°, 225°, 270° y 315°) que refuerzan el píxel central. Las Figuras 1 y 2 muestran los coeficientes de los filtros utilizados en la detección de líneas y bordes (máscaras de Kirsch), respectivamente.

Para la generación del histograma, se procede con un algoritmo iterativo de división de la imagen, en el cual ésta se divide en sub-imágenes de igual tamaño; para cada sub-imagen se genera un histograma de frecuencia que proporciona información relativa de cada tipo de discontinuidad. Al recopilar la información de cada una de las sub-imágenes se genera un solo histograma que sirve como índice de la textura de la imagen, el cual al ser normalizado y cuantizado permite extracción del contenido de la imagen invariante a rotación, traslación y escalado.

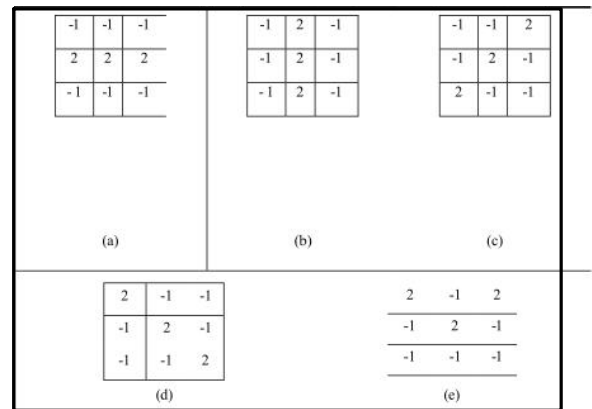


Figura 1. Coeficientes de filtros para discontinuidades de línea (a) horizontal, (b) vertical, (c) 45°, (d)-45°, (e) no direccional.

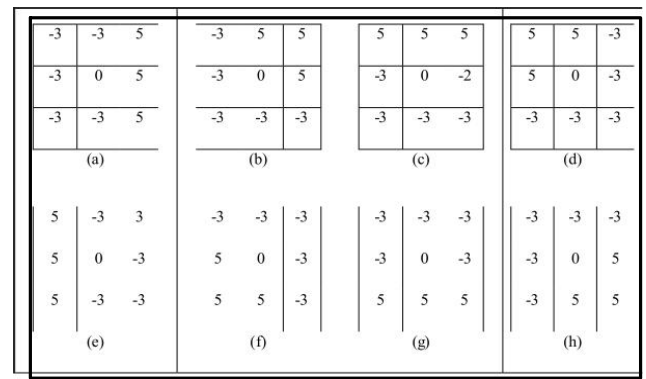


Figura 2. Máscara de Kirsch definiendo ocho orientaciones (a) 0°, (b) 45°, (c) 90°, (d) 135°, (e) 180°, (f) 225°, (g)270°, (h) 315°

Para determinar la información de cada tipo de discontinuidad en las sub-imágenes, se divide cada sub-imagen en bloques, los cuales se dividen, a su vez, en sub-bloques. Para cada sub-bloque, se halla el nivel de gris promedio y se multiplica por el coeficiente correspondiente en cada una de las matrices de filtros digitales; para lo cual, es necesario que el número de sub-bloques y el tamaño de la matriz de los filtros coincidan. Para el sistema propuesto de detección de bordes se utilizan matrices 3x3, por lo que se crean bloques que contengan igualmente 3 sub-bloques por cada fila y columna, formando una matriz de las mismas dimensiones del filtro.

Con estos valores, las magnitudes de los aportes al histograma pueden ser calculadas mediante el siguiente conjunto de ecuaciones, donde (i, j) representa las posiciones de los bloques, (w, z) las posiciones de los subbloques, n el número de subbloques y f(w, z) los coeficientes de los filtros [23]:

$$m_v(i, j) = \left| \sum_{w=0}^n \sum_{z=0}^n a_{wz}(i, j) x f_v(w, z) \right| \tag{1}$$

$$m_h(i, j) = \left| \sum_{w=0}^n \sum_{z=0}^n a_{wz}(i, j) x f_h(w, z) \right| \tag{2}$$

$$m_{45}(i, j) = \left| \sum_{w=0}^n \sum_{z=0}^n a_{wz}(i, j) x f_{45}(w, z) \right| \quad (3)$$

$$m_{-45}(i, j) = \left| \sum_{w=0}^n \sum_{z=0}^n a_{wz}(i, j) x f_{-45}(w, z) \right| \quad (4)$$

$$m_{nd}(i, j) = \left| \sum_{w=0}^n \sum_{z=0}^n a_{wz}(i, j) x f_{nd}(w, z) \right| \quad (5)$$

Una vez se han obtenido los cinco valores, el mayor de estos es comparado con un umbral (asignado empíricamente a 5), si es menor se asigna la dirección correspondiente a ese valor como orientación predominante y su respectivo histograma de frecuencia aumenta en una ocurrencia. De otra manera, se dice que el sub-bloque presenta una textura suave y no aporta información al histograma.

IV. RESULTADOS

Para evaluar el algoritmo en colecciones de imágenes especializadas, se trabajó con una base de datos compuesta por 897 imágenes de piezas mecánicas y automotrices agrupadas en 23 categorías por similitud semántica. Cada una de las imágenes de la base de datos fue ingresada al sistema y para cada categoría se realizaron pruebas modificando el color, rotación y traslación de las imágenes.

Para evaluar la calidad de los resultados arrojados por el sistema se utiliza la precisión, la cual es una métrica de evaluación ampliamente utilizada en los sistemas de recuperación de información y definida como el número de imágenes relevantes recuperadas sobre el número total de imágenes recuperadas. Los resultados obtenidos se muestran en la Tabla 1. Las Figuras 3, 4 y 5 muestran los resultados obtenidos para tres categorías distintas. Para las Figuras 3, 4 y 5 la imagen más a la izquierda representa la imagen ingresada al sistema, en el panel de la derecha se presentan las 20 imágenes con un mayor grado de similitud. Las imágenes resultado se ordenan comenzando desde aquellas con una menor medida de disimilitud.

A partir del análisis de la Tabla 1 es posible observar que algunas de las categorías muestran un bajo desempeño en la recuperación (categoría 2 y 23), lo cual se explica por las características particulares de la base de datos, ya que algunas piezas presentan similitudes en la textura que no corresponden a objetos con el mismo significado semántico como se presenta en la Figura 5.

Tabla 1. Precisión obtenida para cada una de las categorías

Grupo	1	2	3	4	5	6	7	8	9	10	11
Número de imágenes	64	22	56	46	48	23	38	56	48	57	26
Precisión (%) - disc. líneas	0.90	0.25	0.55	0.55	0.8	0.65	0.7	0.9	0.85	0.8	0.8
Precisión (%) - disc. bordes	0.85	0.25	0.65	0.65	0.6	0.65	0.7	0.85	0.65	0.9	0.9

Grupo	12	13	14	15	16	18	19	20	21	22	23
Número de imágenes	33	34	34	40	67	50	37	31	33	31	23
Precisión (%) - disc. líneas	0.5	0.5	0.95	0.9	0.65	0.9	0.85	0.65	0.65	0.65	0.3
Precisión (%) - disc. bordes	0.55	0.45	0.95	0.9	0.8	0.85	0.9	0.8	0.75	0.9	0.35

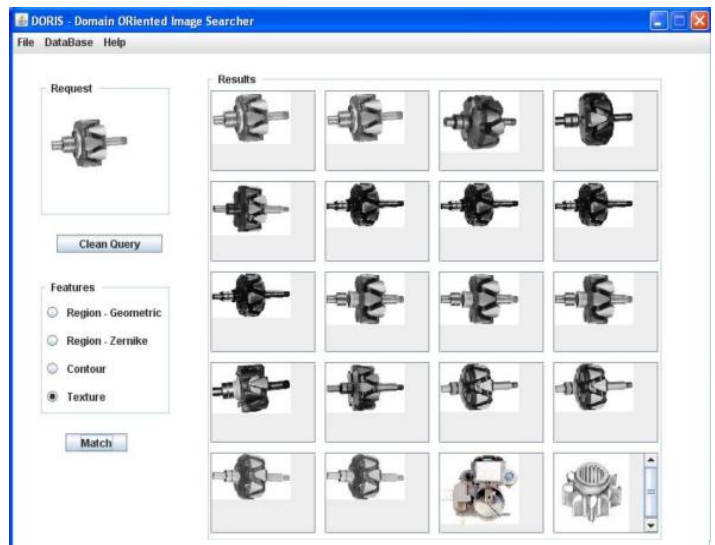


Figura 3. Imágenes recuperadas por el sistema DORIS para rotores

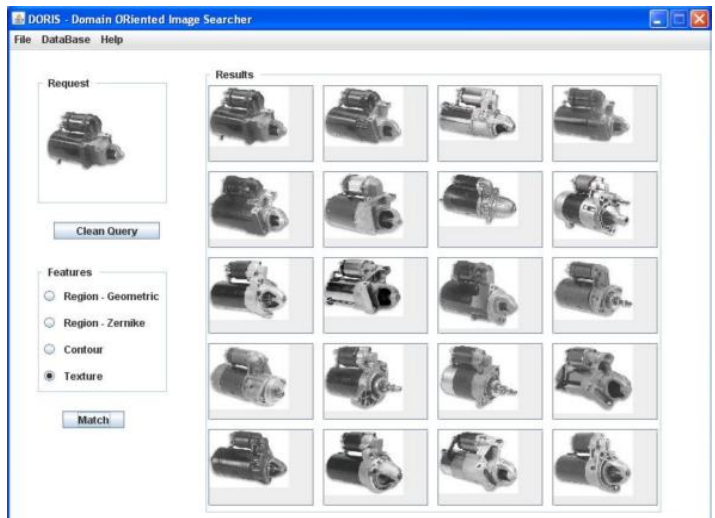


Figura 4. Imágenes recuperadas por el sistema DORIS para arranques

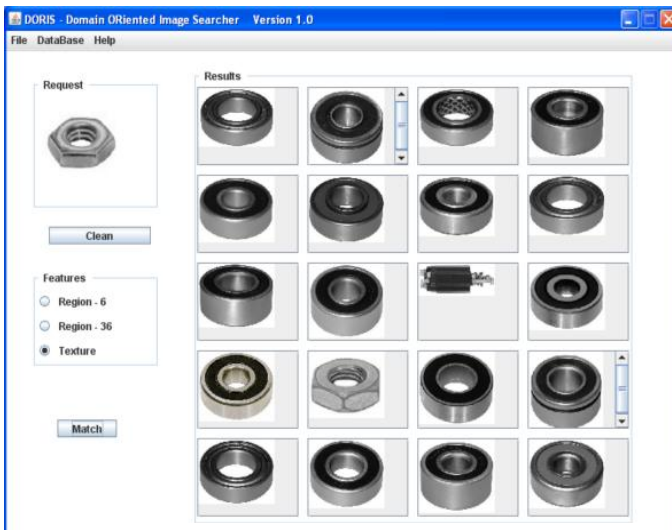


Figura 5. Imágenes recuperadas por el sistema DORIS para tuercas

IV. CONCLUSIONES Y TRABAJO FUTURO

En este artículo, se presentó una propuesta para la recuperación de imágenes de piezas mecánicas y eléctricas, utilizando un algoritmo para la detección de discontinuidades de bordes y líneas. Para tal propósito, se construyó para cada imagen un histograma que representara su distribución de textura no homogénea, el cual fue almacenado en una base de datos y utilizado como índice, para su posterior comparación con las imágenes de prueba ingresadas a la interfaz del sistema DORIS. Los resultados obtenidos muestran una precisión promedio del sistema de alrededor del 70%, demostrando la efectividad de la utilización de la textura como descriptor del contenido visual. La textura no homogénea, muestra una descripción efectiva, tanto para imágenes simples como rodamientos, hasta la recuperación eficaz de imágenes más complejas como arranques de motor.

Los resultados que se muestran en este artículo se derivan de una primera versión del sistema. En el futuro se espera que el sistema DORIS pueda ser llevado a una arquitectura cliente-servidor que tome, ya no imágenes de una base de datos determinada, sino que tenga la capacidad de indizar imágenes de la web provenientes de un dominio específico, para luego recuperar tanto las imágenes como la página web que las contiene, esto con el fin de brindar al usuario un sistema que recupere tanto contenido visual como textual, basado en búsquedas por contenido de imagen.

REFERENCIAS

- [1] Carson C.; Belongie. S.; Greenspan y Jitendra M., 2002. Blobworld: Image segmentation using Expectation-Maximization and its application to image querying. En: IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 24, No. 8, pp 1026-1038.
- [2] Flickner M.; Sawhney H.; Niblack W.; Ashley J.; Huang Q.M.; Dom. B.; Gorkani M.; Hafner J.; Lee D.; Petkovic D.; Steele D. y Yanker. P., 1995. Query by Image and Video Content: The QBIC System. IBM Almaden Research Center. En: IEEE Computer Vol. 28, No. 9, pp 23-32.
- [3] Huang T.; Mehrotra S. y Ramchandran K., 1996. Multimedia Analysis and Retrieval System (MARS) project. En: Proc of 33rd Annual Clinic on Library Application of Data Processing - Digital Image Access and Retrieval.
- [4] Hunter J., 2001. An Overview of the MPEG-7 Description Definition Language (DDL). IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, No. 6.
- [5] Imatch. Digital Image Management Solution. [En línea]: Imatch. <http://mwllabs.de/> [Consulta: 5 de abril de 2008]
- [6] Jaimes A.; Ruiz-del-Solar J.; Verschae R.; Baeza-Yates R.; Castillo C.; Yaksic D. y Davis E., 2004. On the Image Content of a Web Segment: Chile as a Case of Study. Journal of Web Engineering, Vol. 3, No. 2, pp. 153-168.
- [7] Kim S. M.; Park S. J. y Won C. S., 2007 Image Retrieval via Query-by-Layout Using MPEG-7 Visual Descriptors. ETRI Journal, Vol. 29, No. 2, pp. 246-248.
- [8] Ma W. Y.; Deng Y. y Manjunath B. S., 1997. Tools for texture/color based search of images. En: SPIE. Conf. 3106, Human Vision and Electronic Imaging II, pp 496-507
- [9] Marques O. y Furtht B., 2002. Distributed multimedia databases: techniques and applications. Idea Group Publishing. 384 P.
- [10] Martínez J., 2002. MPEG-7. Overview of MPEG-7 Description Tools, Part 2. En: IEEE Multimedia. Vol. 09, No. 3, pp 83-93
- [11] Martínez J. M.; Koenen R. y Pereira F., 2002. MPEG-7: The Generic Multimedia Content Description Standard, Part 1. En: IEEE Multimedia. Vol. 9, No. 2, pp 78-87
- [12] Organisation Internationale de Normalisation ISO/IEC. [En línea]: MPEG-7 Overview. <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7>. [Consulta: 05 de abril de 2008]
- [13] Pertland A.; Picard R. W.; Sclaroff S., 1994. Photobook: Content-Based Manipulation of Image Databases. En: SPIE Storage and Retrieval Image and Video Databases II, No. 2185, pp. 6-10.
- [14] Picassa TM. [En línea]. <http://picasa.google.es/> [Consultada el 10 de septiembre de 2008]
- [15] Rui Y.; Huang T. y Chang S-F., 1999. Image Retrieval: Current Techniques, Promising Directions and Open Issues. En: Journal of Visual Communication and Image Representation, Vol. 10, pp 39-62
- [16] Smith J. R. y Chang S. F., 1996. VisualSEEK: A fully automated Content- Image Query System. En: Proceedings of the fourth ACM International Conference on Multimedia, pp 87-98
- [17] Smith J. R. y Chang S., 1997. An Image and Video Search Engine for the World Wide Web. En: Proceedings SPIE Storage and Retrieval for Image and Video Databases. pp 84-95
- [18] Stanchev P. L.; Amato G.; Falchi F.; Gennaro C.; Rabitti F. y Savino P., 2004. Selection of MPEG-7 Image Features for Improving Image Similarity Search on Specific Data Sets. En: 7th IASTED International Conference on Computer Graphics and Imaging, pp 395-400
- [19] Thao C. y Munson E. V., 2003. A Relevance Model for Web Image Search. En: Proceeding of the Second International Workshop on Web Document Analysis. Pattern Recognition and Image Analysis Group, pp. 57-60.
- [20] The Hermitage Museum, IBM. [En línea]: The State Hermitage Museum <http://www.heritagemuseum.org/>. [Consulta: 05 de abril de 2008]
- [21] Veltkamp R. y Tanase M., 2001. Content Based Image Retrieval Systems: A Survey. Technical Report UU-CS-2000-34
- [22] Virage VideoLogger. [En línea]: Searchable Video: "A Watershed Event for the World Wide Web". <http://publications.autonomy>

- com/pdfs/Virage/Case%20Studies/ [Consultada el 10 de septiembre de 2008].
- [23] Won C. S.; Park K.. y Park S-J., 2002. Efficient Use of MPEG-7 Edge Histogram Descriptor. En: ETRI Journal, Vol. 24, No. 1, pp 23-30