MPEG7 algorithms implemented for a Multimedia Management System

Liana Stănescu, Dumitru Sorina, Cosmin Stoica Spahiu, Dumitru Dan Burdescu *

** University of Craiova, Faculty of Automation, Computers and Electronics, Bvd. Decebal, No.107, Romania; (e-mail{ stanescu, burdescu, stoica.cosmin}@software.ucv.ro).*

Abstract: The paper presents an original dedicated integrated software system for managing and querying alphanumerical information and images. The system is designed with a modularized architecture which is based on a relational database management server. The system is updated with new algorithms from MPEG 7 for image processing and retrieval. The studies made for the implemented algorithms, have shown that the results obtained by combining the Color Layout, Dominant Color and Texture Edge Histogram descriptors, improved the performance. The visual manner of building this type of query specific for multimedia data and the modified Select command that is sent for execution to the MMDBMS give originality to the software product.

Keywords: Image retrieval, image processing, MPEG7 descriptors.

1. INTRODUCTION

The large scale usage of the digital data in the last years, and the growth of the internet have ensured that huge volumes of high dimensional multimedia data are available all around us. This kind of information is often mixed, the same document containing different data types such as: text, image, audio, speech, hypertext, graphics, and video. All these components are interspersed with each other. The World Wide Web has played an important role in making the data easily accessible to users all over the world, even if they are from geographically distant locations.

All these information must be efficiently managed and retrieved in order to be useful. Different algorithms and systems have been implemented for this goal. The most known set of algorithms is MPEG.

The main objective of the MPEG-7 visual standard is to provide standardized descriptions of streamed or stored images or video-standardized header bits (visual low-level descriptors) that help users or applications to identify, categorize, or filter images or video. These low-level descriptors can be used to compare, filter or browse images or video purely on the basis of non-textual visual descriptions of the content – or in combination with common text-based queries.

The paper presents the development of a multimedia databases management system which combines the functionality of a classical relational system with algorithms for images processing.

A classical DBMS has to provide support for database management like browsing and querying, inserting, updating, storing and deleting information. It has to ensure database integrity and security (Dimitrova, 1999).

In addition, a multimedia database management system must have functions for metadata processing, special functions for images processing and characteristics extractions, and functions for storing all this information.

The paper has the following structure: section 2 presents the related work, section 3 presents the general architecture of the MMDBMS, Section 4 and 5 present in detail the colour descriptors implemented for the SGBD, and Section 6 presents the conclusions and future work related to this paper.

2. RELATED WORK

Many content based retrieval systems have been developed in the last years. Some of them for commercial purposes, others have been developed as results of research projects. Below one presents some of the most known systems:

• QBIC project (Query By Image Content) was one of the first commercial content-based image retrieval system that was developed. It was implemented by IBM and consists in a search engine that sorts database images according to some descriptors: colours, textures, shapes, sizes, and space position. The system has implemented functionality for: query-by-example, user drawings, selected colour and texture patterns, camera and object motion. Any image obtained as a query result can be used as a new query in order to improve the search. (Flickner, M. et. al (1997), Niblack, C. W. et. al, (1993), Royo, C.V., (2010)).

• Virage project (Bach, J. R. et. al, 1996) is also a content-based images retrieval system, developed by Virage Inc. This system is similar to the QBIC project: it has functionality for visual queries based on colour, texture, and shapes. The main difference is that this system allows the users to adjust the weights of each characteristic taken into account (Royo, C.V., 2010).

• RetrievalWare project is also a commercial retrieval engine developed by Excalibur Technologies Corporation.

It was commercially launched in 1992. Its emphasis was in neural nets to image retrieval. Its more recent search engines use features such as colour, shape, texture, brightness, colour layout, and aspect ratio of the image (Excalibur Visual Retrievalware, (1998), Royo, C.V., (2010)).

• Photobook project is a retrieval system which allows the users to browse image databases using both text annotation information added to images, and executing content-based retrieval operations. This project was developed at the MIT Media Lab. It is composed of three sub-modules: shape, texture, and face features extracted. Humans' interaction is needed during the image annotation and retrieval loop (similar to the MIRROR system) (Pentland, A. et. al, (1995), Royo, C.V.,(2010)).

• ImgSeek project is a free open source photo collection manager and viewer which includes content-based retrieval functionality. The query is built by providing either a sketch painted by the user or an image file (Jacobs, C. E. et. al. (1995) , ImgSeek web, (Royo, C.V., (2010)).

• MIRROR project. The acronym "MIRROR" comes from Multimedia Information Retrieval Reducing Information OveRload. It was developed by the University of Twente and it is one of the most complex systems (Wong et. al, 2005, Royo, C.V., (2010)). It was implemented to evaluate MPEG-7 visual descriptors and to design and develop new algorithms for image retrieval. It includes a web-based user interface for query by image example retrieval. The system does not include the DBMS. Figure 1 shows a capture of the MIRROR interface. The MIRROR system has implemented the following color descriptors from MPEG-7: Dominant Color, Color Layout, Scalable Color and Color Structure. The texture characteristic is also a feature considered to be very important in image retrieval. The algorithm used for extraction is Texture Edge Histogram. Various matching tools are defined for different descriptors and description schemes.

3. THE MULTIMEDIA DATABASE MANAGEMENT SYSTEM

The multimedia database management system that was developed is an original application that includes methods for extracting the colour and texture characteristics and for executing content based retrieval queries. It uses both classical data types for storing standard information and it has defined a new data type special for images, called IMAGE (Spahiu, C.S. et. al., (2009a), Spahiu, C.S. et. al, (2009b)). This type is used to store the image file and the extracted metadata.

The multimedia database management system was developed by the Computers and Information Technology Department from the University of Craiova. The basic reason for the designing of this system was that most of the database servers existing on the market nowadays does not offer any support for the images management and processing. Most of them do not offer any particular functionality for images management. It is usually recommended to store the images files outside the database, this way being sensitive to unauthorized changes/deletes.

The biggest advantage of this system is that the client applications' complexity is much lower than that from other similar systems, as the system includes all the necessary functionality for images processing. The clients

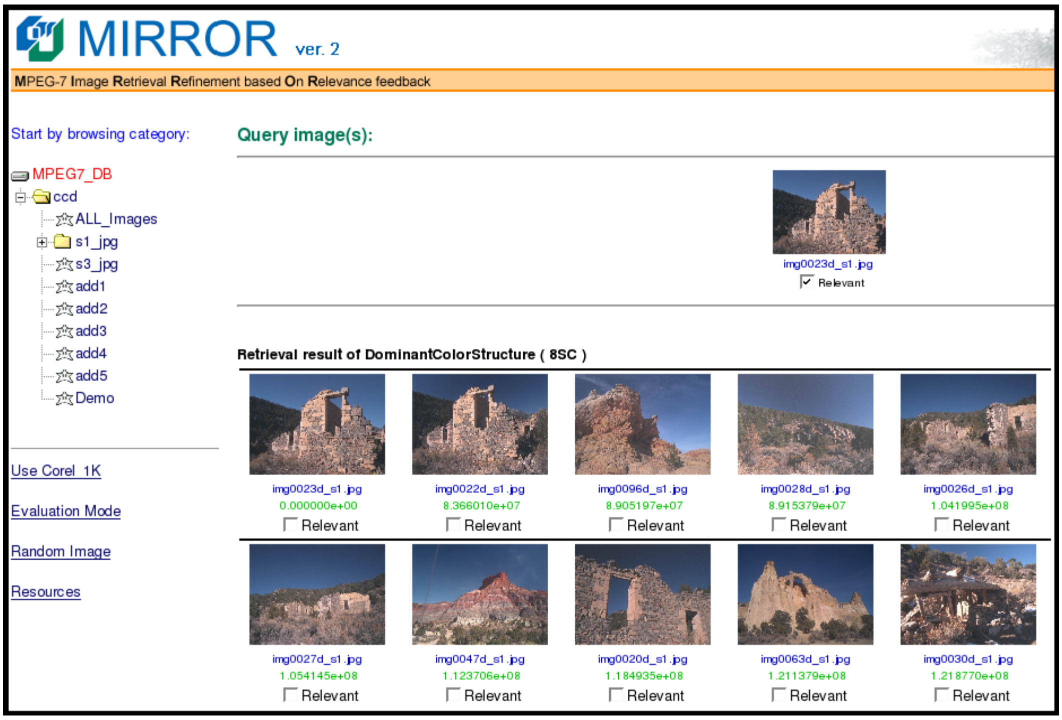


Figure 1. A capture of MIRROR interface.

need only to call the APIs of the system.

The system is a modularly designed, each functionality being managed by a different module. There are implemented modules for: colour characteristics extraction, texture extraction, similarity computation, SQL processing, etc.

If needed, new functionalities can be added very easy by adding new modules to the system and connecting them to the main module.

4. MPEG DESCRIPTORS USED

One of the main objectives of the system is to include all algorithms needed to characterize images and to determine their similarity close to humans' colour perception.

The initial version of the multimedia system used the Gabor Filters and Histograms for images retrieval. The functionality was extended by adding new MPEG7 algorithms for images processing and retrieval. The MPEG-7 standard is a collection of methods used to obtain different colour descriptors from images (MPEG-7).

After studying the advantages and disadvantages of each colour descriptor, the Colour Layout Descriptor (CLD) and the Dominant Colour Descriptor (DCD) were selected for implementation. This decision was carried out to take the advantages offered by DCD and CLD. The DCD makes possible the effective description of the dominant colours of an image and the CLD is a compact, resolution-invariant representation that retains the spatial distribution of an image's colours.

4.1 Dominant Colour Descriptor (DCD)

The Dominant Colour Descriptor (DCD) it is one of the descriptors used in MPEG7. It is used to give a compact description of the salient colours in an image or image region. It gives the possibility to specify a small number of dominant colour values as well as their statistical properties: distribution or variance (Verdaguer, S.,L., (2009)).

The algorithm used to extract the dominant colours from an image needs a clustering method which is used to cluster the pixel colour values.

There are needed three stages for the extraction process: a color space change, a clustering method and a calculation of the percentages of each centroid (Verdaguer, S.L., (2009)).

The first stage is recommended by the MPEG-7 standard, in order to use a perceptually uniform colour space for executing the clustering method. The recommended colour space is CIE LUV. This step must be implemented as most of the images used in this project were defined on the RGB colour space.

The second step can begin once the input image is converted to the CIE LUV colour space. At this step a clustering algorithm is applied in order to extract the dominant colours of the image. Any clustering method can be applied, however the one selected in this project is the K-means algorithm.

The K-means clustering algorithm is a method of cluster analysis which aims to divide n observations into k clusters. Each observation belongs to the cluster with the nearest mean (K-means web a.)(K-means web b.).

The following Figure 2 is an example of a colour clustering using K-means for the DCD extraction of an input image. As the standard recommends the use of three or four dominant colours for the DCD, four colours have been selected in this particular example.

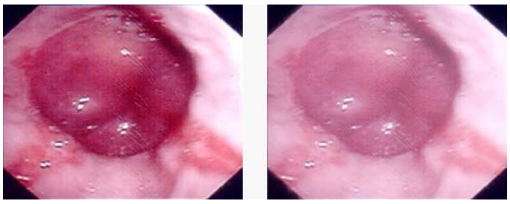


Figure 2. K-means: RGB clustering with 4 clusters.

It can be observed as only four dominant colours of the original image have been selected, which corresponds to the four centroids of the K-means clusters.

The percentage of pixels in the image belonging to each of the clusters is calculated in the last step of the algorithm. The DCD results (four centroids and the percentage) obtained for the previous example, are summarized in the following table:

Table 1. Example of DCD

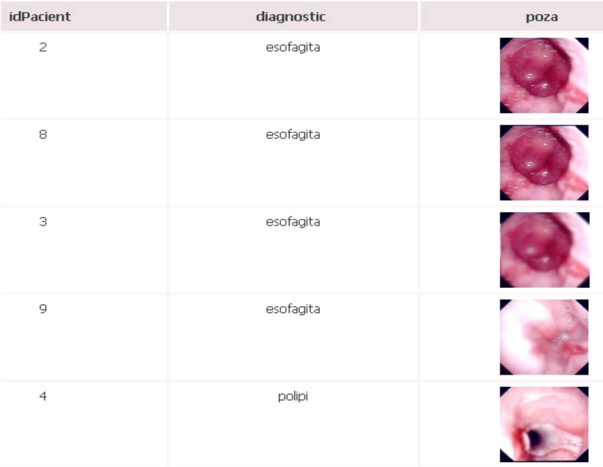
Centroids ()		Percentage ()	
L*	u*	v*	
54.2495	70.0140	12.3690	0.3501 %
80.3203	24.7781	18.8156	0.1321 %
66.4478	49.2724	17.7287	0.2456 %
34.6152	36.4142	-15.7432	0.2722 %

These centroids and its corresponding percentages represent the DCD of the image.

The performance of the system using DCD was evaluated using different images from database. In this case, it is expected to obtain images with similar dominant colours ordered according to increasing distances. An example of the results returned is presented in Figure 3.

In this particular example it can be observed that the image with less distance (distance 0) is the input image (it is detected in the database) and the following ones are the other images with similar dominant colors. Furthermore, the same results have been obtained using, as an input to the system, images with different dominant colors, such as white, red, brown.

The obtained results were satisfactory. However, in this case the computing time when the descriptors were extracted was very high:



idPacient	diagnostic	poza
2	esofagita	
8	esofagita	
3	esofagita	
9	esofagita	
4	polipi	

Figure 3. Results obtained using DCD .

- Analyzing the DCD of one picture: 20 seconds
- Input image DCD extraction and Matching with 100 DCD: 45 seconds (S.L. Verdaguer, S.L., (2009)).

4.2 Color layout descriptor (CLD)

The second descriptor implemented in the system is Colour Layout Descriptor. It captures the spatial layout of the representative colours on a grid superimposed on a region or image. The representation of image is based on coefficients of the Discrete Cosine Transform. It is a very compact descriptor, very efficient, being recommended for fast browsing and search applications. The descriptor can be used with good results both for still images as well as for video files (Verdaguer, S.L., (2009)).

This descriptor is obtained by applying the discrete cosine transform (DCT) transformation on a 2-D array of local representative colours in Y or Cb or Cr colour space.

The functionalities of the CLD is matching both images and video. The CLD is one of the most precise and fast colour descriptor from MPEG.

Four steps are needed for the descriptor's extraction process: image partitioning, representative colour detection, DCT transformation and a zigzag scanning. An additional step is needed to transform colour space from RGB (if necessary) to YCbCr.

The algorithm divides the input picture (represented in RGB color space) in 64 blocks to guarantee the invariance to the image resolution or scale (Fig. 4).

The second step starts after the image partitioning stage. It

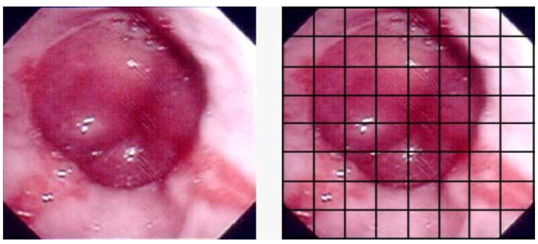


Figure 4. Step 1: Partitioning.

needs the selection of a single representative colour from each block. This selection can be performed using any appropriate method. The standards' recommended method is to use for the corresponding representative colour the average of the pixel colours in a block. This method has a low complexity and the description accuracy is good.

The results of this selection are tiny image icons of size 8x8. Figure 5 shows the process. The resolution of the original image has been maintained only in order to facilitate its representation.

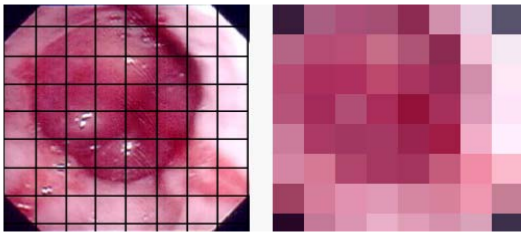


Figure 5. Stage 2: Representative colour selection.

The colour space conversion from RGB to YCbCr is applied once the new icon is obtained. This color space conversion can be done after any step. However, the MPEG-7 standard recommended to proceed with the conversion in this point in order to reduce the computational load needed for this process (Verdaguer, S.L., (2009)).

The following formulas are used to calculate the DCT in a 2D array:

$$B_{pq} = \alpha_p \alpha_q \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A_{mn} \cos\frac{\pi(2m+1)p}{2M} \cos\frac{\pi(2n+1)q}{2N}, \quad \begin{matrix} 0 \le p \le M-1 \\ 0 \le q \le N-1 \end{matrix} \qquad (1)$$

$$\alpha_p = \begin{cases} \frac{1}{\sqrt{M}}, & p = 0 \\ \sqrt{\frac{2}{M}}, & 1 \le p \le M-1 \end{cases} \qquad \alpha_q = \begin{cases} \frac{1}{\sqrt{N}}, & q = 0 \\ \sqrt{\frac{2}{N}}, & 1 \le q \le N-1 \end{cases} \qquad (2)$$

The inputs and outputs for this step are presented by Table 2.

Table 2. Inputs and Outputs Stage 4

Input Stage 4	Output Stage 4
Tiny image icon [8x8] in YCbCr color space	3 [8x8] matrix of 64 coefficients (DCTY, DCTCb, DCTCr)

A zigzag scanning is performed in the next step with these three sets of 64 DCT coefficients. The scanning schema is depicted by Fig. 6.

The purpose of this step is to group the low frequency coefficients of the 8x8 matrix.

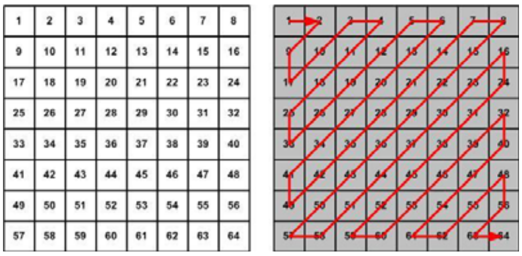


Figure 6. Stage 5: Zigzag scanning

The CLD of the input image are represented by these three set of matrices.

The images similarity between two CLDs obtained from two different images is computed as:

$$D = \sqrt{\Sigma_i w_{yi}\left(DY_i - DY_i'\right)^2} + \sqrt{\Sigma_i w_{bi}\left(DCb_i - DCb_i'\right)^2} + \\ + \sqrt{\Sigma_i w_{ri}\left(DCr_i - DCr_i'\right)^2} \quad (3)$$

where {*DY*, *DCb*, *DCr*}, { *DY'*, *DCb'*, *DCr'* }, represent the CLDs of the images.

The subscript *i* represents the zigzag-scanning order of the coefficients. Furthermore, notice that is possible to weight the coefficients (*w*) in order to adjust the performance of the matching process. These weights let us give to some components of the descriptor more importance than others (Verdaguer, S.L., (2009)).

5. TEXTURE DESCRIPTORS

The texture characteristic is a very important low-level descriptor in images retrieval, similar to the colour characteristic. The MPEG-7 standard uses three texture descriptors (MPEG-7 (a)-(c)):

• texture browsing descriptor (TDB). It uses perceptual attributes such as directionality, regularity, and coarseness of a texture. This algorithm obtains results close to human eye perception. The TBD is a 5-dimensional vector expressed as:
[Regularity Directionality1/ Directionality2/ Scale1/ Scale2], where:

Regularity represents the degree of periodic structure of the texture. The larger the Regularity value is, the more regular the pattern is;

Directionalities are associated to two dominant orientations of the texture;

Scales represents the two dominant scales of the texture.

When this algorithm is used, the similarity is best computed using Euclidean distance

• homogeneous texture descriptor (HTD) . It provides a quantitative characterization of homogeneous texture regions for similarity retrieval. It characterizes the texture characteristic using the energy and energy deviation in a set of frequency channels (ISO/IEC JTC1/SC29/WG11, (2002)). This characteristic is extracted using Gabor filter which partitions the frequency space with equal angle of 30º in angular direction and with octave division in radial direction. According to some previous results, the best numbers of angular and directional parameters are 6 and 5, resulting in 30 channels in total.

• local edge histogram descriptor. This descriptor has good results when the image region that should be processed is not homogeneous. The algorithm divides an image into 4 x 4 non-overlapping sub-images. Each sub-image is divided into an application-specific number of image-blocks. The five types of edge information can be extracted from the image-blocks by edge detection operators. For each sub-image a local edge histogram with 5 bins is generated and the total of 80 histogram bins (16 sub-images multiplying 5 bins) is achieved for the whole image (Xu, F. and Zhang, Y.J., (2005)).

The performances of these algorithms were evaluated from the speed and retrieval quality point of view.

Table 3 shows the average runtime needed for each of the three algorithms to process an image. The results of the HTD recommend it as the most efficient for texture extraction. Although it hasn't the simplest implementation, it has the lowest complexity among the three descriptors. The most complex and time consuming descriptor is TBD.

Table 3. The average runtime of needed for texture extraction

Texture descriptor	HTD	TBD	EHD
Average runtime (ms)	585	814	664

6. CONCLUSIONS AND FUTURE WORK

The paper presented the additional functionality of a multimedia database server using a set of low-level features called visual descriptors that have been defined by MPEG7. The system can be used for managing medium sized collections of images. There is also made a survey of the similar implementations of other well known applications that manage the content based image retrieval.

This MMDBMS is created for managing and querying medium sized personal digital collections that contain both alphanumerical information and digital images. The software tool allows the creating and deleting of databases, the creating and deleting of tables in databases, updating data in tables and querying. The user can use several types of data as integer, char, double and image.

This software can be extended in the following directions:

• Adding new types of traditional and multimedia data types (for example video type or DICOM type - because the main area where this multimedia DBMS is used it is the medical domain and the DICOM type of data is for storing alphanumerical information and images existing in a standard DICOM file provided by a medical device).

• Studying and implementing indexing algorithms for data inserted in the tables.

• Adjusting the weights of the CLD&DCD in order to improve the obtained results.

• Combining the different results obtained by each visual descriptor; for this it will be study the statistical distribution of the distance values obtained for each descriptor in order to normalize each distance before combining them.

• Using Relevance Feedback techniques in order to improve the results. These techniques consist in involving the human factor in the search engine. Thus, once the retrieved images have been shown, the user has to decide which results are considered relevant. The retrieval system takes then the advantage of this information given by the user and tries to improve its results by using all the features extracted from these relevant images.

• Developing a region-based search engine. For this we will need to incorporate shape descriptors, as well as whole set of descriptors used in image retrieval.

REFERENCES

Bach, J. R., Fuller, C., and Gupta, A. (1996). Virage image search engine: an open framework for image management, *Storage and Retrieval for Still Image and Video Databases* IV 2670, no. 1, 76–87.

Del Bimbo, A., (2001). Visual Information Retrieval, *Morgan Kaufmann Publishers*. San Francisco USA

Dimitrova, N. (1999). Multimedia Content Analysis and Indexing for Filtering and Retrieval Applications, *Informing Science Special Issue on Multimedia Informing Technologies*, vol.2 no.4, pp.87-100.

Excalibur Visual Retrievalware, web page (1998). http://www.excalib.com/products/vrw/vrw.html.

Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., and Yanker, P. (1997). Query by image and video content: the QBIC system, *Intelligent multimedia information retrieval*, MIT Press, Cambridge, MA, USA, pp. 7–22.

ImgSeek, web page: http://www.imgseek.net/

ISO/IEC JTC1/SC29/WG11, (2002). MPEG-7 overview, V. 8, Doc. N4980.

Jacobs, C. E., Finkelstein, A., and Salesin, D. H. (1995). Fast multiresolution image querying, SIGGRAPH ’95: *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques* (New York, NY, USA), ACM, pp. 277–286.

K-means: http://people.revoledu.com/kardi/tutorial/kMean/NumericalExample.htm

K-means: http://www.mathworks.com/access/helpdesk/help/toolbox/stats/index.html?/access/helpdesk/help/toolbox/stats/kmeans.html&http://www.mathworks.com/cgibin/texis/webinator/search/?db=MSS&prox=page&rorder=750&rprox=750&rdfreq=500&rwfreq=500&rlead=250&sufs=0&order=r&is_summary_on=1&ResultCount=10&query=kmeans&submitButtonName=Search

Khoshafian S., and Baker, A.B. (1996). Multimedia and Imaging Databases. *Morgan Kaufmann Publishers*, Inc. San Francisco California.

Manjunath, B.S., Salembier, P., and Sikora, T., (2002). Introduction to MPEG-7: Multimedia Content Description Interface, 1st edition.

MPEG: http://www.mpeg.org/

MPEG-7(a): http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm

MPEG-7(b): http://www.m4if.org/resources.php#section40

MPEG-7(c): http://www.multimedia-metadata.info/

Niblack, C. W., Barber, R., Equitz, W., Flickner, M. D., Glasman, E. H., Petkovic, D., Yanker, P., Faloutsos, C., and Taubin, G. (1993). QBIC project: querying images by content, using color, texture, and shape, *Storage and Retrieval for Image and Video Databases*, vol. 1908, SPIE, pp. 173–187.

Pentland, A., Picard, R. W., and Sclaroff, S., (1995). *Photobook: Content-Based Manipulation of Image Databases*.

Royo, C.V. (2010) Image-Based Query by Example Using MPEG-7 Visual Descriptors.

Smith, J.R. (1997). Integrated Spatial and Feature Image Systems: Retrieval, Compression and Analysis. *Ph.D. thesis, Graduate School of Arts and Sciences. Columbia University*

S.C. Spahiu, C., Mihaescu, C., Stanescu, L., Burdescu, D.D., Brezovan, M. (2009 (a)). Database Kernel for Image Retrieval, *The First International Conference on Advances in Multimedia (MMEDIA 2009)*, pp. 169 – 173.

Stoica Spahiu, C. (2009 (b)). A Multimedia Database Server for information storage and querying, *2nd International Symposium on Multimedia – Applications and Processing (MMAP'09)*, Vol. 4, pp. 517 – 522.

Verdaguer, S.L. (2009). Color Based Image Classification and Description

Wong, K. M., Cheung, K. W., and Po, L. M. (2005). MIRROR: an interactive content based image retrieval system, *IEEE International Symposium on Circuits and Systems*, vol. 2, pp. 1541–1544.

Xu, F., Zhang, Y.J. (2005). Evaluation and comparison of texture descriptors proposed in MPEG-7, November 2005, www.sciencedirect.com.