An Intelligent Multimedia System for Efficient Image Browsing and Retrieval

Kostas Karpouzis, George Votsis, Yiannis Xirouhakis, Giorgos Stamou and Stefanos Kollias

National Technical University of Athens, Dept. of Electrical and Computer Engineering,
Computer Science Division, Heroon Polytechniou 9, 157 80 Athens, Greece
{kkarpou, yvotsis, jxiro}@image.ntua.gr

Abstract. In this work, an integrated system is proposed for efficient image browsing and retrieval in multimedia databases. The system incorporates intelligent management and content-based retrieval schemes on top of a fundamental multimedia database structure. Agents are employed in order to limit search into minor subsets of the database. This concept is applied in all modules of the proposed system, minimizing search time and complexity regarding users point of view. Content-based retrieval algorithms are employed facilitating quick browsing in the database contents.

1 Introduction

Multimedia applications involve storage, handling and retrieval of voluminous data, such as images and video sequences. Searching for such kinds of data introduces users to mainly three obstacles. First, the amount of information contained in a big data pool confuses the user as far as the pools subjects in relation to his/her goal are concerned. Besides, the user needs to spend precious time, in order to finally access the information that is of his/her initial exact interest. Finally, looking for multimedia objects often requires their potentially able description through features, which are not effectively described by keywords. Attempting to answer these primary problems, an integrated multimedia system is designed, which facilitates browsing in multimedia (image and video) databases. For this purpose, agents are employed in order to classify a particular user in a predetermined profile and thus, specify the subset of the database that better matches his/her interests. Content-based query mechanisms allow direct access to what the subjects of the database represent, easing their correct retrieval. In this sense, the main intelligent agent framework [1]-[3] is combined with a domain discrimination technique in order to segment the database into distinct virtual subparts. Content-based search schemes [4]-[10] complement the agents functionality by implementing efficient querying within these subparts of the database.

2 System Overview

The proposed system focuses on quick and efficient browsing in a multimedia database that contains images and video clips of various themes. Its main purpose is to ensure that any user finds the information he/she is interested in quickly and without being obliged to answer many questions or browse through irrelevant material. A modular representation of the system architecture is shown in Figure 1. On the top of the diagram, the Client box denotes the web-based environment provided for user navigation into the database contents. A dedicated Web Server provides this web-based environment and collects user data, which in turn are evaluated by the Agent module and the particular user is classified in one of the profiles existing in the database. At the same time, the Web Server module directs image queries to the Search Engine module, which in turn formulates the query and passes it through to the Database . The Database module consists of all information incorporated in the system. The Image Database part corresponds to stored visual material (images and video clips). In addition, it contains all information related to the content-based query and retrieval scheme, such as the employed Feature Extraction Algorithms and Feature Vectors extracted for each image. Finally, all classification information (existing profiles, look-up tables and indexes) are also stored. An initial set of images and video is used to populate the database (Image Database module). Since every image corresponds to one or more categories, a meta-data vector is stored with each image, containing a set of weights, which indicate to which categories the particular image belongs (Look-up Tables module). This vector also accommodates information that depicts the popularity of the image with respect to the different variables that the profile consists of, e.g. age or sex of the user (Indexes module). The system-training scheme is included in Agent module. Besides, it is possible that the database grows enormously large, with hundreds of images corresponding with high

confidence levels to a certain category. For this reason, the Search Engine is related to content-based query and retrieval schemes (Feature Extraction module) so as to implement efficient access. During browsing, users may ask for content-based queries-by-visual-example. The user has also the ability to set the relative importance of each image attribute, e.g. shape, color, etc., by choosing their respective weights in an on-line fashion. The system responds to such queries quickly since the search is limited within the categories characterizing the particular profile.

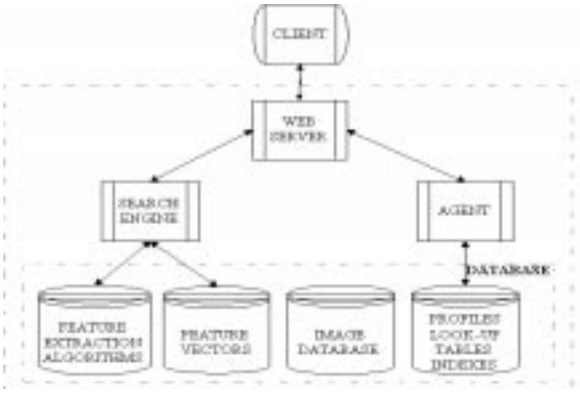

Fig. 1. System Architecture

3 Scenarios

The system consists of three 'abstract' levels (sets): the profiles, the categories and the images set. The former consists of a number of distinct characteristic users (profiles) P_n, $n = 1 \cdots N$, where N denotes the number of profiles stored in the system. In turn, the categories set consists of a number of distinct categories C_m, $m = 1 \cdots M$, where M is the total number of predetermined categories. It is essential that two distinct profiles may correspond to different but overlapping subsets of the categories set. The images set consists of all images I_k, $k = 1 \cdots K$, where K is the number of images stored in the database. It is presumed that every image belongs to all categories (see Figure 2). This presumption is based on the fact that given no relevant information, every image is a candidate for every category. In fact, this relevance cannot be predetermined, since the interpretation of an image may differ for two distinct users of the system. For example, a picture of Acropolis may well be classified in *History*, whereas at the same time be classified in *Sports* for someone involved with car-races. In this sense, no image can be excluded from a certain category.

Every image is saved along with a vector of weights $\mathbf{w}_k^I$ classifying it with higher or lower certainty to all categories. This vector is of length M. For example, if $\mathbf{w}_5^I(1) > \mathbf{w}_5^I(2)$, then image I_5 belongs to category C_1 with higher certainty than to C_2. This definition of weights is also a measure of image 'popularity' in a certain category. For example, if $\mathbf{w}_5^I(2) > \mathbf{w}_6^I(2)$, then image I_5 belongs to category C_2 with higher certainty than image I_6 does. Consequently, an $M \times K$ look-up table $\mathbf{I}$ with its columns corresponding to the vectors of weights, contains sufficient information for images' classification to categories. For example, the profile *Artist* will most likely be associated with *Literature* and *Music* rather that *Sports* and *Technology*. For vectors $\mathbf{w}_n^P$ similar results to that of $\mathbf{w}_k^I$ can be obtained by comparing respective weights. Consequently, a sparse $M \times N$ look-up table $\mathbf{P}$ with its columns corresponding to the vectors of weights, contains sufficient information for categories classification to profiles. For the m-th row the number of non zero entries is equal to the number of profiles associated with the m-th category.

At first, the particular user is assigned a certain profile P_a. This is peformed directly by the user. In the exceptional case that the user wishes to receive a suggestion by the system, he/she gets one through the *'Agent'* module's relevant utility. A sequence of retrieved database entries (images) is available to the user, who is capable of selecting any image in the database starting with the most popular images in profile P_a. In this sense, the following properties of the systembecome clear: (a) every user, no matter what profile he/she is assigned, is capable of choosing any image in the database, (b) depending on the profile chosen, the system will come up with the most popular image in this profile, and (c) the user is discouraged to search for an image irrelevant to the profile he/she has chosen.

Once the user has located an image that bears some resemblance to the one he/she has in mind, in terms of some similarity measures, he/she is given the opportunity to perform a query-by-example using the particular image as input. For every image imported in the database a feature vector is extracted, containing information concerning image attributes with respect to defined similarity metrics. After the user has decided on the relevant attribute similarity, by imposing certain weights for each attribute, a query is performed returning a new set of retrieved data. In this sense, for each query the feature vector of the input image is simply compares to the feature vectors of the rest of the images in the database taking into consideration the weights imposed. The user is able to refine the query until he comes up with the desired output. Every query is somehow limited to the subset of the database corresponding to the selected profile through the selected categories. In fact, all images are retrieved, however feature vector comparison is performed taking them in order of relevant popularity. The results of the query are sorted first by degree of popularity in the profile and then by degree of attribute similarity. In this context, the system will come-up first with the most popular images in the profile which are at most visually similar to the input image. Another three properties of the system become clear at this point: the user is able of: (a) quickly and efficiently browsing in the database contents using content-based queries, (b) choosing which attributes will lead him/her to the desired image, and (c) setting the degree of popularity of the images under comparison, implicitly deciding on the time needed by the system to respond to its query.

Finally, the user finally comes up with the desired image. His choice becomes clear to the system by selecting to view the thumbnail in full size or selecting to download it. This is a valuable information for the user and his/her profile that must not be wasted. In this sense, the look-up tables are updated along the lines of the Section describing the system's adaptation mechanisms.

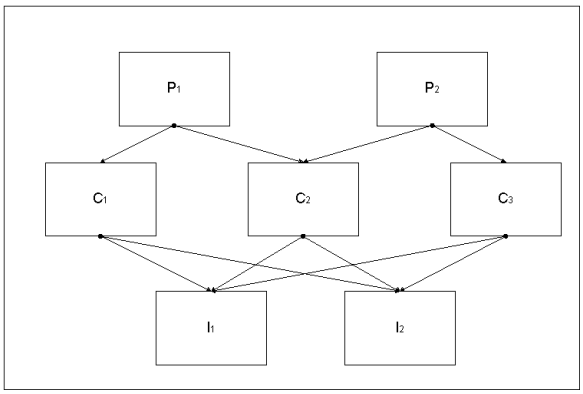

Fig. 2. The three-level schema

4 System Adaptation

Two are the basic problems which are tackled by the 'Agent' module of the system. One is that of evaluating data which has been collected via the 'Web Server' from the user. The outcome of this evaluation is a suggestion to the user concerning which of the supported profiles should be chosen by him/her. The other is that of customizing the retrieved with respect to its relevance: a multimedia database usually contains a large number of images and video which are close to what one is looking for, but usually some of the material is closer to the target. This is the reason why a bi-directional adaptation algorithm has been adopted. Its logic is explained in the following section.

After the profiles are created, they are mapped to all database areas corresponding to different themes. This is essentially a weighted mapping, which means that some of these areas are easier to browse than others, for a user identified to a specific profile. Practically, a vector of length equal to the number of categories is assigned to every profile containing respective weights. This mapping is actually transparent to and thus not perceived by the user of the system, as he/she is not directly asked to choose any of the image and video themes. Queries can sometimes refer to more than one portion of the database; hence, the profiles are mapped to overlapping portions so as to serve such queries. The process, which leads to the actual retrieval of the images, is a part of the internal structure of the image. An image file, especially a photograph, cannot be classified in a single subject or theme, as it usually conveys different meanings

to different users or depicts diversified material or objects. In the proposed system, this is supported by implementing a weighted mapping between the images and the themes.

The system's 'Agent' module assists in directing the user's queries in a specific part of the database, thus decreasing the respective computational load. Unfortunately, the information that is supplied to the user in this stage is completely unstructured, in the sense that it merely represents the datas weights w.r.t. each category. Besides, what characterizes the system is it's ability to adapt its visual features to the actual users' preferences and choices. This is accomplished by altering the weights of each chosen image with regard to the predefined categories of the profile in use, taking under consideration that all weight values should be between 0 and 1. For instance, consider a user that is classified to the Artist profile, which in turn is mapped -to some degree-to Music and Cinematography. When this user chooses to download one of the images presented, this is regarded by the system as an indication that it is popular in these categories and therefore, the 'intelligent agent' should raise the second layer's related weights. The adaptation of these weights is performed through the aid of a learning algorithm that involves the strength of the mapping of the profile to the above categories. Although this bears the risk of refreshing the wrong weights, due to misuse on the part of the user, the 'correct' responses will eventually far surpass this and the users' choices will be accurately reflected. The learning algorithm also changes the first layer's weights, i.e. the fuzzy mapping of the profiles to the predefined categories. The underlying idea of the adaptation has as follows. The 'intelligent agent' raises the weights that connect the user profile to the categories that are strongly associated to the selected image. In this sense, the described system may be viewed as a three-layered fuzzy-neural network (FNN). The nodes of the first layer are the system's profiles, while those of the second layer are the predefined categories. Finally, the nodes of the third layer are the images stored in the database.

5 Content-based Information Retrieval

Once the agent has presented the user with a potential subset of his interest, a content-based retrieval scheme has to be utilized in order to specify his needs in a more detailed manner. The mechanism applied in our system exploits the use of a feature vector, which is composed by the features extracted from an image through the use of processing tools. One may discern mainly two kinds of information related to a digital visual object (either that is an image or a video sequence): the information that describes that object, also called its meta-data, and the information included in the object, known as its visual features [6]. Meta-data is usually alphanumeric and is modeled as some data structure in a database system. On the other hand, visual features need to be extracted from the visual object through the use of image processing tools. As a matter of fact, several of the early image database systems have used pixels as their contents structural and operational cells. However, in most real applications such an approach is not effective due to noise sensitivity and shift, rotation and illumination variance of the images. Let us as well assume that a human restricts the regions of interest in the images, categorizes and characterizes them. According to the above situation, a precise enough content description of the databases visual objects is available. However, in the case that such visual features are not so easily categorized, or even in real time applications, this approach is not realistic. Even in the case where it is, it is also very laborious.

Most visual information retrieval applications are neither satisfied by simple pixel-based queries, nor may they be restricted in a few predefined object categories, whereas annotation mechanisms are proved inadequate. In such generalized applications, visual information is defined as the outcome of image processing transformations applied on the visual objects [6]. The visual features that occur are attributes of the visual objects viewed in the natural two-dimensional space, in the two-dimensional frequency space, through the use of statistical methods and through the utilization of image processing techniques. Hence, for fundamental properties of an image, such as color, shape and texture, a variety of visual features, which are considered to effectively portray content with respect to contemporary image processing tools, are supported by our system [4]-[10].

Color. As a means for content-based retrieval, color is used either straightforwardly, or after it has been submitted to some kind of statistical processing. In our case, the latter scenario is used, since the characteristic feature for color is the image's color histogram. A palette of 256 colors can portray the content of each image, without essentially distorting it. During a query, a metric is used to calculate the distance between the desired chromatic content and the stored histograms, i.e. intersection, euclidean or quadratic distance:

Intersection: $d_I(h,g) = \dfrac{\sum_{m=0}^{255} \min(h[m], g[m])}{\min\left(\sum_{i=0}^{255} h[i], \sum_{j=0}^{255} g[j]\right)}$

Euclidean distance: $d_E(h,g) = \sum_{m=0}^{255} (h[m] - g[m])^2$

Quadratic distance: $d_Q(h,g) = \sum_{i=0}^{255} \sum_{j=0}^{255} (h[i] - g[i]) \cdot a_{ij} \cdot (h[j] - g[j])$

Color composition. This feature contains higher level information, since it involves topological color distribution in a picture. Such a structure is achieved through the combination of quad-trees and histogram computation for each one of the image quadrant. The metric used for comparison is a simple summation of one of the above metrics over all the available quadrants.

Shape. To compare shape similarities between two objects, one could develop shape models for the various objects, which are space- and time-consuming. In our system, a more eloquent form is adopted. Image segmentation along with edge detection techniques give us the opportunity to calculate quantitative amounts, such as area, centrality, elongation and hue, concerning the segmented part under question. Given those characteristics, complicated enough queries may be answered, as long as they are set in a way that permits arithmetic allowances.

Texture. The approach adopted in our system, is the development of a model, which is based on decomposition of normal, static stochastic processes in 2D images. Supposing that a texture image is a homogeneous discrete 2D random field, this decomposition is the sum of three mutually orthogonal components: a harmonic, a generalized-evanescent and a purely non-deterministic field. This description has the advantage that it does not pose limitations as far as existence of predefined forms of texture is concerned [4]. It also depicts perceptual attributes of patterns, allowing simple metrics to be used for similarity comparison. An alternative description is a vector containing energy, entropy, inverse difference moment and correlation coefficients.

Texture composition. As in the case of color, a quadtree structure gives us the topological knowledge of texture synthesis of the image. Employing the above characteristics, the user has the opportunity to drive his query in the feature space and not directly in the image space. Questions of the form: 'if each image is viewed as a point in the n-dimensional space, find the m closest images within distance d from the characteristic vector **v**' are answered by the system [6]. The overall distance metric is a complicated formula due to two reasons; (a) for each feature the meaning of distance cannot be expressed in a unified formalism and (b) each characteristic is weighted, depending on the importance that the user wishes to appoint it during the search.

The main advantage of the used content-based retrieval mechanism, compared to traditional CBR systems, such as QBIC [4], is that of flexibility. Images are compared with respect to several features, both chromatic and morphological, with an interactive selection of comparison metrics. A weighted combination of the user's selections may have a fundamental semantic value; this value is reinforced through the fuzzy partitioning of the database and the use of profiles and categories, which are far more practical than keywords.

6 Extensions of the Basic System

It must be brought to attention that besides image information, video information is also included in the database, in the form of characteristic frames corresponding to video scenes. In the proposed prototype system, this is achieved through scene detection, scene classification and characteristic frame (key-frame) extraction. Characteristic frames are extracted on the basis of minimum similarity and their number depends on the video content and duration. These key-frames are stored in the database and they are treated as images. At the same time, appropriate indexes indicate which images belong to the same video clip. Feature vector extraction for content-based query and retrieval is extended in the case of video clips to include motion information, appropriate for more efficient shape description. The algorithms incorporated in the system for video processing, along with advanced techniques for shape representation developed by our group can be found in [11] and references thereon.

7 Conclusions

In the current work, an integrated system is proposed for efficient image browsing and retrieval in multimedia databases. The system is based on the concepts of intelligent management and content-based retrieval tools. The database is partitioned into categories according to the basic themes of its multimedia contents, through a multiple layer user classification. A simple architecture is presented for simultaneous use and training of the system. Content-based retrieval techniques are subsequently being employed for the search of images and video sequences. The addition of complementary options to the system, involving facial processing and advanced shape extraction algorithms, is currently under investigation.

8 Acknowledgements

The present work is partially supported by the Greek Ministry of Press and Mass Media and by the MODULATES project (Multimedia Organization for Developing the Understanding and Learning of Advanced Technology in European Schools), in the framework of the Educational Multimedia Program of the European Commission 1998-2001.

References

1. C. Guilfoyle and E. Warner, *Intelligent Agents: The New Revolution in Software*, pp. 214, 1994.
2. P. Janca, *Intelligent Agents: Technology and Application*, GiGa Information Group, 1996.
3. C. T. Lin and C. S. G.Lee, *Neural fuzzy systems: a neurofuzzy synergism to Intelligent Systems*, Prentice Hall, 1995.
4. M. Flickner *et al*, "Query by Image and Video Content: The QBIC System," *Computer*, vol. 28, pp. 23–32, Sept. 1995.
5. A. Pentland, R. W. Picard and S. Sclaroff, "Photobook: Content-Based Manipulation of Image Databases," *SPIE Storage and Retrieval Image and Video Databases II*, no. 2185, Feb. 1994.
6. A. Gupta and R. Jain, "Visual Information Retrieval," *Communications of the ACM*, vol. 40, no. 5, May 1997.
7. J.R. Smith and S.F. Chang, "Quad-Tree Segmentation for Texture-Based Image Query," *Multimedia 94*, San Francisco, CA, USA, Oct. 1994.
8. C. E. Jacobs, A. Finkelstein and D. H. Salesin, "Fast Multiresolution Image Querying," in *Proc. SIGGRAPH 95*, Los Angeles, CA, Aug. 1995.
9. M. Abbadi and A. Youssef, "Indexing and Searching of Image Data in Multimedia Databases," George Washington University, 1997.
10. L.D. Bergman, V. Castelli and C.S. Li, "Progressive Content-Based Retrieval from Satellite Image Archives," D-Lib Magazine, October 1997.
11. Y. Xirouhakis, Y. Avrithis, and S. Kollias, "Image Retrieval and Classification using Affine Invariant B-spline Representation and Neural Networks," in *IEE Colloq. on NN in Multimedia Interactive Systems*, London, UK, Oct. 1998.