EFFICIENT BROWSING IN MULTIMEDIA DATABASES USING INTELLIGENT AGENTS AND CONTENT-BASED RETRIEVAL SCHEMES

Y.Xirouhakis, G.Votsis, K.Karpouzis, and S.Kollias
Computer Science Division, Dept. of Electrical Eng.,
National Technical University of Athens,
Athens GR-15773, GREECE

Abstract - In this work, an intelligent multimedia system is developed for efficient retrieval and processing of information stored in multimedia databases. Intelligent agents are employed in order to limit the search into minor subsets of the database; a training strategy is introduced for quick browsing in the database contents. Content-based query and retrieval schemes are also employed to increase the effectiveness of the browsing procedure.

INTRODUCTION

Due to recent growth of interest in multimedia applications, an increasing demand has emerged for efficient storage, management and browsing in multimedia databases. The latter has been given considerable attention after the recent guidelines of the Moving Pictures Expert Group regarding the MPEG-4 and MPEG-7 standards. In this work, a prototype system is introduced, facilitating browsing in multimedia databases. For this purpose, intelligent agents are employed in order to specify the subset of the database that matches the interests of the particular user. At the same time, content-based query and retrieval schemes give the opportunity of quick and efficient access to database contents. As described in the next Section, the main intelligent agent framework [1, 2] is combined with a domain discrimination technique in order to segment the database into distinct virtual subparts. In addition, content-based query and retrieval schemes [6, 4] are developed for efficient querying in these subparts of the database.

SYSTEM OVERVIEW

The proposed system focuses on the quick and efficient browsing in a multimedia database containing mainly images of various themes. The main purpose of this system is to ensure that any user finds the information he/she is interested in quickly and without being obliged to answer many questions.

A modular schema of the system overview is depicted in Figure (1). The first module of the proposed system stands for the database structure. It is assumed that the database contains **images** that may correspond to one or more pre-defined **categories**, w.r.t the specific application. Let us for example consider *History*, *Engineering*, *Music*, *Cinematography*, and *Sports*. In this way, a knowledge domain discrimination is performed on a higher level of the database.

In order to exploit this discrimination, the second module of the system consists of a scheme that decides on the specific user's interests. This scheme is developed through intelligent agents. In this sense, a number of characteristic users or **profiles** is defined. Each profile is related to one or more of the above categories. Let's for example assume that the profiles *Scientist*, *Teacher*, *Artist* and *Student* are defined. Every characteristic user is interested in a subset of the categories' set; however, these subsets may overlap, for example $Scientist \equiv \{Engineering, History\}$, $Teacher \equiv \{History, Music\}$, $Artist \equiv \{Music, Cinematography\}$ and $Student \equiv \{Music, Sports, Cinematography\}$. The profiles and categories are defined before the system is set on-line. In addition, a set of images is introduced as initial input in the database. Since every image corresponds to one or more categories, a vector of length equal to the number of existing categories is stored with each image. This vector contains a set of weights, which indicate, in a probabilistic sense, to which categories the particular image belongs.

The system training scheme is included in the third module. The system is trained by the preferences of each individual user, every time one enters it, following his having chosen a specific profile. Explicitly, the user is capable of browsing through all the images in the database; however, the system comes up first with thumbnails of the most popular images of this profile. For example, if the user chooses the profile *Artist*, the system will come up first with the most popular images from *Music* and *Cinematography*, according to the images' weights. When the user views (or downloads) a particular image, the system automatically increases the weights indicating that it belongs to either *Music* or *Cinematography*.

On the other hand, it is possible that the database grows enormously large, with hundreds of images corresponding with high possibility to a certain category. For this reason, in the fourth module, content-based query and retrieval schemes are employed for efficient image retrieval. During browsing in the database under a certain profile and after having selected an image that bears some relevance with the one the user has in mind, he/she may ask for a content-based query-by-example using the particular image as input. The user has also the ability to set the relative importance of each image attribute by choosing respective weights on-line. The system responds to such queries quickly since the search is limited within the categories characterizing the particular profile. Moreover, the search is performed starting from the most popular images of the categories, while the user may execute successive queries giving as input any of the retrieved images.

INTELLIGENT AGENTS

It is usually very time-consuming to browse through a multimedia database, mainly because it contains very large data objects, such as images and video. Besides, querying multi-dimensional objects can be computationally complex, because of the nature of the related algorithms. For that reason, an intelligent classification and searching scheme is implemented in order to serve such queries in an efficient way as indicated in [1, 2]. At first, a set of dinstinct profiles P is created, each characterized by different interests and preferences. After these profiles are created, they are mapped to one or more areas A of the database 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 $\mathbf{w}_P$ of length equal to the number of categories is assigned to every profile containing respective weights. The mapping is transparent to the user of the system, as he/she is not directly asked to choose any of the themes.

Queries can sometimes refer to more than one portion of the database; hence, the profiles that are created can be mapped to overlapping portions in order 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 in most cases be catalogued in a single subject as it can convey different meanings to different users or can depict different material or objects. In the proposed system, this is supported by implementing a weighted mapping between the images and the subjects; this means that an image of an ancient amphorae depicting wrestlers can be classified in a category named *History*, with a relatively high confidence coefficient, as well as in a category named *Sports*, but this time with a substantially lower coefficient. When a user browses the database by means of a specific profile, say *Teacher*, he/she will have more efficient access to the categories this profile is mapped to (e.g. *History*, *Music*). As a result, the image in question, which has a high coefficient w.r.t. this subject will be one of the first to be acquired. On the contrary, if a user categorized to the profile *Artist* browses the database, the system expects him/her to be interested in *Music* and *Cinematography*, a fact that will delay the retrieval of the amphorae image, because of its lower relevancy. An image I^j is inserted in the database with an arbitrary vector of coefficients $\mathbf{w}_I^j$ w.r.t each subject or category. This vector is of length equal to the number of categories and its initial values are specified by the administrator of the system. When a user selects a specific image for viewing or downloading, the image's coefficients, concerning the themes this user's profile is mapped to, are raised; in this way, after due time, the 'correct' coefficients are much higher than the remaining and the image will show up early after relevant queries, but much later when a query concerns a different category.

It can be seen that the system's structure discourages the user from choosing an image that has not become popular in a category, although the latter is possible. This fact supports the stability of the system. It must be also pointed out that overlapping categories help in classifying images in their 'correct' categories. For example, when *Teachers* enter the system, they vote for images concerning *History* and *Music*. This helps *Scientists* finding images concerning *History*, since the latter is their common category and vice versa.

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. A feature vector for each image is extracted before it enters the database and represents it. The definition of content, as far as our system is concerned, involves the following forms of information:

Color. As a means for content-based retrieval, color may be 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.

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 quadtrees and histogram computation for each one of the image quadrant.

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.

Form. To determine the similarity between two classes of objects, as far as their generic form is concerned, one should find a set of characteristics which is efficient for their description. A classical way of facing such a problem is the well known KLT transform. The eigenvectors of the covariance matrix of a set of characteristic images may be viewed as a set of parametric variations from the mean form of the described objects. By keeping the eigenvalue vector for each image, one manages to maintain robust information that discriminates the various objects of the same class.

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].

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 $\mathbf{v}$' are answered by the system [6]. The 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) the use of weights for each characteristic, depending on the importance that the user wishes to appoint it during the search.

EXTENSIONS OF THE BASIC SYSTEM

Presently, we investigate the possibility that besides the profession or occupation, the profiles are characterized by a number of other user attributes such as age, sex and preferences. In this sense, the user can be assigned a more intelligent profile by submitting a simple text form to the system. These more generic profiles can be created beforehand, by training a neural network in a neurofuzzy framework, using the queries of actual users as input. Neurofuzzy techniques are adopted, since they match better the nature of input data, which include non-numerical, fuzzy responses of the user. It must be brought to attention that besides image information, video information could be also included in the database, in the form of characteristic frames corresponding to video scenes. For that purpose, another algorithm is developed taking into consideration that images belonging to the same video clip may have dinstinct themes. At the same time, a number of additional features can be incorporated in the system. These involve face extraction and 3D-modelling techniques [5].

CONCLUSIONS

In the current work, a multimedia database architecture is being designed, based on the concepts of intelligent agents and content-based retrieval. The database is partitioned into categories according to the basic themes of its multimedia contents. The search area of the database for each user is being restricted through the utilization of intelligent agents that classify the user's interests. A simple and quick algorithm is proposed for simultaneous use and training of the system. Content-based retrieval techniques are subsequently being employed for the search of images and video sequences.

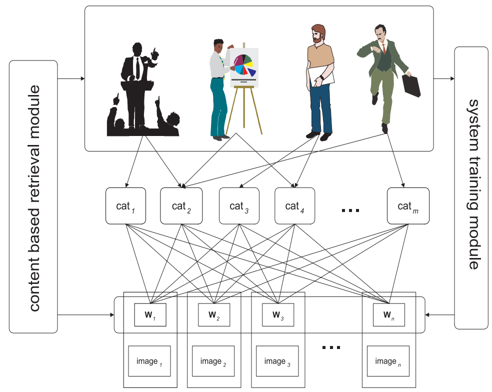

Figure 1: System Overview

ACKNOWLEDGMENTS

The present work is partially supported by the Greek Ministry of Press and Mass Media and by the *MODULATES* project (Multimedia Organisation for Developing the Understanding and Learning of Advanced Technology in European Schools), in the framework of the Educational Multimedia Program of the European Commisssion 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] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker "Query by Image and Video Content: The QBIC System," *Computer,* vol. 28, pp. 23–32, Sept. 1995.

[4] 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, San Jose, Feb. 1994.

[5] G. Votsis, Y. Xirouhakis, S. Kollias, and A. Stafylopatis, "Feature Extraction and 3D Recovery of Human Faces in Video Sequences," *Int. Conf. on Inf. Syst. Analysis and Synthesis,* Orlando, USA, July 1998.

[6] A. Gupta and R. Jain, "Visual Information Retrieval," *Communications of the ACM,* vol. 40, no. 5, May 1997.