Multimedia Analysis and Retrieval System

Sharad Mehrotra, Yong Rui, Kaushik Chakrabarti, Micheal Ortega, Thomas S. Huang
Department of Computer Science
University of Illinois at Urbana-Champaign
Email: sharad,kaushikc,ortega-b{@cs.uiuc.edu} yrui,huang{@ifp.uiuc.edu}

June 15, 1997

1 Introduction

With the advances in storage technology and the advent of the World Wide Web, there has been an explosion in the amount and complexity of digital information being generated, analyzed, stored, accessed and transmitted. Most of this data is multimedia in nature, including digital images, video, audio and simple text data. To make use of this vast amount of multimedia data, we need to develop techniques to efficiently retrieve information over large multimedia repositories based on their content. Due to the difficulty in capturing the content of multimedia objects using textual annotations and the non-scalability of the approach to large data sets (due to a high degree of manual effort required in defining the annotations), the approach based on supporting content-based retrieval over visual features has become a promising research direction. This is evidenced by several prototypes[15, 8, 6, 4] and commercial systems[3, 2] that have been built recently.

Most existing approaches take a *"vision-centric"* view in which objects are represented using automatically extracted visual features and the system allows users to retrieve objects based directly on the extracted features (e.g., retrieve all images containing a lot of "blue"). In contrast, in the *Multimedia Analysis and Retrieval System* (MARS) being built within our group at the University of Illinois, we have adopted a more *"information-centric"* approach. A user expresses his information needs in the form of a higher-level query (e.g., retrieve images similar in content to a given set of images) which is then mapped to lower-level image features by the system possibly with the help of user interaction. Mapping the user query to a set of features extracted from textual documents have been extensively studied in the information retrieval (IR) literature [14]. In developing the MARS system, we have generalized these techniques for content-based retrieval over image features. In the remainder of the paper, we describe some of the problems that arise in such a generalization and the approaches we have taken to address the problems. We conclude with the directions of future research being pursued in the MARS project.

2 Content Based Retrieval in MARS

Before we describe the retrieval approach used in MARS, we briefly review the retrieval process in modern IR systems [14].

2.1 Existing IR Techniques

In an IR system, a text document is represented as a collection of features (also referred to as terms). Examples of features include words in a document, citations, bibliographic references, etc. A user specifies his information needs to the system in the form of a query. Given a representation of the user's information need and a document collection, the IR system estimates the likelihood that a given document matches the users information needs. Such a match is achieved using a *retrieval model*. Term weights are assigned to the keywords extracted from the documents and the queries which estimate the relative importance of the keyword to the document and the query. Simple statistical measures (e.g., term frequency (*tf*), inverse document frequency (*idf*)) are used to estimate the weights of the keywords associated with the document [13]. The most common measure used is the product of *tf* and *idf*. Weights of the keywords associated with queries are typically estimated using the technique of *relevance feedback*. In relevance feedback, the system provides the user with the initial set of documents. Based on the relevance estimates provided by the user, the query is refined until the user's information need is met.

2.2 Challenges

MARS approach to content-based retrieval consists of generalizing the IR techniques to content-based retrieval over feature representations extracted from multimedia objects. Such a generalization poses many significant challenges. Statistical techniques like *tf* and *idf* need to be generalized to estimate the relative importance of the features and the individual components within a feature to a multimedia object. Adapting relevance feedback approach to adjust the user's query in visual databases is also complex. Problems arise since visual features, unlike keywords which are atomic, are composite structures. For example, the color feature, which is represented using a 8×8 histogram, comprises of 64 components. Determining the importance of individual component to a user's query requires generalizing the relevance feedback approach to non-atomic feature representations. Another issue is that of the importance of structural or layout information in the visual domain to retrieval. Typical IR systems ignore the presence of structure in textual documents for retrieval. Finally, existing retrieval models for mapping queries to documents have specifically geared towards textual documents. These models need to be generalized to domains in which visual features are used.

In the following subsections, we provide an overview of the retrieval models, the normalization process, the term-weighting strategy and the relevance feedback approach used in MARS.

2.3 Retrieval Models used in MARS

With a large number of retrieval models in the IR literature, MARS attempts to exploit this research for content-based retrieval over images. In MARS, an image is represented as a collection of low-level image features[1] as well as a manual text descriptions of the image. A user graphically constructs a query by selecting certain images from the collection. A user may choose specific features from the selected images (e.g., retrieve images similar to an image A in color and similar to an image B in texture) or the entire image itself (e.g., retrieve all images similar to image C) User's query is interpreted as a Boolean expression over image features and a Boolean retrieval model (adapted for retrieval over images) is used to retrieve a set of images ranked based on the degree of match. Boolean queries provide a natural interface for the user to formulate and refine *conceptual queries* (e.g, fields with yellow flowers, a sunset by a lake, etc.) to the system using lower-level image features.

To support a ranked retrieval over queries, MARS adopts the following two extensions to the basic Boolean model:

- **Fuzzy Boolean Retrieval:** distance between the image and the query feature is used to compute the degree of membership of the image to the fuzzy set of images that match the query feature. Fuzzy set theory is used to interpret the Boolean query and the images are ranked based on the their degree of membership in the set.
- **Probabilistic Boolean Retrieval:** distance between the image and the query feature is used to compute the probability that the image matches the user's information need. Feature independence is exploited to compute the probability of an image satisfying the query which is used to rank the images.

Unlike the basic Boolean model, both the fuzzy and probabilistic Boolean models provide a ranked retrieval over the image sets. Furthermore, as one expects from Boolean models, both the fuzzy and the probabilistic Boolean models rank images based on queries corresponding to semantically equivalent Boolean expressions in the same order.

2.4 Feature Sequence Normalization in MARS

Since distances between the image and the query based on features are interpreted as the degree of membership in the fuzzy model and the relevance probability in the probabilistic model, these distances need to normalized to the same scale before they can be used. The reason is that the features are over a diverse domain and use different distance metrics, so the actual distance of the image from the query based on a feature may not exactly reflect the degree (or probability) of match of the image to the query based on the feature. Normalization process serves the purpose of putting equal emphasis or importance to each feature. In MARS normalization is done at two levels.

- **Intra-feature Normalization**: It is used to put an equal emphasis on each feature component within a feature vector. To see the importance of this, consider the texture layout representation where the feature

[1]For example, color is represented by a 8×8 HS histogram, texture is represented by CCD (*coarseness, contrast and directionality*) components, shape uses modified Fourier Descriptor (MFD) to represent object boundaries and layout uses a 5×5 matrix of subimages with individual color, texture and shape features.

components may be over totally different physical domains. One feature component can be a mean while the other can be a standard deviation. Thus, their magnitudes can vary drastically, thereby biasing the Euclidean distance measure. To overcome the above problem, MARS uses the Gaussian normalization to map the feature components within the feature vector to the N(0,1) range [7].

- **Inter-feature Normalization**: It is used to map the distance values of the image from the query based on a feature vector into the range [0,1] so that they can be interpreted as the degree of membership in the fuzzy model or relevance probability in the probability model. Similar to intra-feature normalization, Gaussian normalization is also used for inter-feature normalization [7].

2.5 Feature Weighting Techniques used in MARS

While the normalization process achieves the goal of equalizing the contribution of each feature component within the feature vector, and of each feature vector within a multi-feature query in computing the distance of an image from the query, not every feature vector (and a component within the feature vector) is equally important to an image or relevant to the query. MARS uses generalizations of the statistical measures of *tf* and *idf* as feature weights associated with the images in the database and uses a generalization of the relevance feedback process to learn the query weights. The feature weighting techniques used in MARS is discussed below and the relevance feedback approach used is described in the following subsection.

Motivated by the *tf* measure, which estimates the importance of a term to a document, we defined the notion of *ci* (component importance) and *fi* (feature importance) which estimate the importance of a given feature component within the feature vector and the importance of a feature to the image respectively. Similarly, motivated by the *idf* measure, which estimates the discriminatory power of a keyword in a given document collection, we defined the notions of *ici* (inverse component importance) and *ifi* (inverse feature importance) which measure the discriminatory power of a feature component (within a feature vector) and that of a feature vector (among a set of features) in the image collection respectively. These measure are estimated using simple statistical properties. For example, since *ici* measures the discriminatory power of a feature component, standard deviation of the sequence of components values in the entire image collection is used as a measure of *ici*. Smaller the standard deviation, smaller is the discriminating power of the component, and as a result, smaller is its *ici*.

2.6 Relevance Feedback in MARS

In MARS, relevance feedback mechanism is used to allow users to specify their exact information needs [11, 9, 12, 10]. Since different users and the same user at different instances may have different perceptions of the image content, relevance feedback is used to learn the set of features, the representation of the feature (in case multiple representations of a given feature are supported), and the components in the feature vector that are of interest to the user. This is achieved by distributing different weights, which are *dynamically adjusted* based on user's feedback, between the different features, within the different feature representations, and within each component in a representation during the retrieval process. The approach allows a user to specify a coarse initial query and

Texture Feature Representation	Estimate of weights	Precision after ith iteration of Relevance Feedback (RF)			
		i=0 (No RF)	i=1	i=2	i=3
Wavelet	ci	0.7687	0.8213	0.8433	0.8550
Wavelet	$ci \times ici$	0.7727	0.8233	0.8513	0.8553
Co-occurrence Matrix	ci	0.5753	0.6320	0.6513	0.6607
Co-occurrence Matrix	$ci \times ici$	0.5780	0.6347	0.6513	0.6640

Table 1: Experimental Results: The variation of precision with texture feature representation, estimate of feature weights and relevance feedback

continuously interacts with the system until his information need is satisfied.

2.7 Experimental Results

To validate the effectiveness of the IR approach to content based retrieval used in MARS, we have performed experiments over several diverse datasets including a collection of images of ancient African artifacts from the Fowler Museum of Cultural History (online demo at http://jadzia.ifp.uiuc.edu:8001) , texture images from the MIT media Lab (online demo at http://quark.ifp.uiuc.edu:8080/html/texture.html) and terrain databases (online demo at http://jadzia.ifp.uiuc.edu:8002/arldemo.html).

Table 1 illustrates the impact of using feature weights and relevance feedback to retrieval performance over a collection of texture images. Results are shown of experiments over two different representations of texture to establish the applicability of the IR techniques to visual features irrespective of their representation. The experimental results show that:

- The precision is higher with relevance feedback than without relevance feedback.
- $(ci \times ici)$ is a better estimate of weights of the feature components than ci alone [2]. This result is consistent with the results of experiments conducted on text documents which demonstrates $(tf \times idf)$ is a better estimate than tf alone [13].
- The precision increase after the first iteration of relevance feedback is the highest. This is desirable, as it saves the user from giving feedback to the system several times to achieve good retrieval performance. The first iteration of feedback produces much larger improvement in precision than compared to later iterations.

3 Future Work

While our results indicate that extending IR techniques for content-based retrieval over multimedia databases is a promising approach, much further work needs to be done before we can fully validate the effectiveness of IR techniques for multimedia retrieval. So far, our work has been restricted to only content-based retrieval over image databases. Extending the approach to video and other general multimedia databases introduces many significant problems. The problem not only consists of extracting the relevant features (e.g., object movement,

[2]Since only a single feature (i.e. texture) is used in this experiment, fi and ifi measures are not used

motion parameters, activity measures) from video which, in itself, is a big challenge, but even the underlying retrieval models supported need to be extended. While our experiments show that the low-level image features (e.g., color, textures, shapes, etc.) provide a powerful representation for content-based retrieval, such low-level descriptors may not be powerful enough to capture the rich content of a video. Instead, the representation will need to be augmented with a semantic description of video content (as in [1, 5]). The information retrieval techniques will then need to be generalized for content-based retrieval over the semantically augmented representation. Such retrieval techniques are being currently explored in MARS.

References

[1] Sibel Adali, K.S. Candan, Su-Shing Chen, Kutluhan Erol, and V.S. Subrahmanian. Advanced video information system: Data structures and query processing. *To appear in ACM-Springer Multimedia Systems Journal.*

[2] Jeffrey R. Bach, Charles Fuller, Amarnath Gupta, Arun Hampapur, Bradley Horowitz, Rich Humphrey, Ramesh Jain, and Chiao fe Shu. The virage image search engine: An open framework for image management. In *SPIE Storage and Retrieval for Still Image and Video Databases IV.*

[3] M. Flickner, Harpreet Sawhney, Wayne Niblack, and Jonathan Ashley. Query by image and video content: The qbic system. *IEEE Computer*, 1995.

[4] Thomas S. Huang, Sharad Mehrotra, and Kannan Ramchandran. Multimedia analysis and retrieval system (MARS) project. In *Proc of 33rd Annual Clinic on Library Application of Data Processing - Digital Image Access and Retrieval*, 1996.

[5] E. Hwang and V.S. Subrahmanian. Querying video libraries. Technical report, Univ. of Maryland.

[6] B.S. Manjunath and W.Y. Ma. Texture features for browsing and retrieval of image data. Technical report, CIPR TR-95-06, 1995.

[7] Michael Ortega, Yong Rui, Kaushik Chakrabarti, , Sharad Mehratra, and Thomas S. Huang. Supporting similarity queries in MARS. In *(submitted to)ACM Conf. on Multimedia*, 1997.

[8] A. Pentland, R.W.Picard, and S.Sclaroff. Photobook: Tools for content-based manipulation of image databases. In *Proc. Storage and Retrieval for Image and Video Databases II*, volume 2, pages 34–47, Bellingham, Wash, 1994.

[9] Yong Rui, Thomas S. Huang, and Sharad Mehrotra. Content-based image retrieval with relevance feedback in MARS. In *Proc. IEEE Int. Conf. on Image Proc.*, 1997.

[10] Yong Rui, Thomas S. Huang, and Sharad Mehrotra. Human perception subjectivity and relevance feedback in multimedia information retrieval. In *(submitted to) Storage and Retrieval of Images/Video Databases VI, EI'98*, 1998.

[11] Yong Rui, Thomas S. Huang, Sharad Mehrotra, and Michael Ortega. Automatic matching tool selection using relevance feedback in MARS. In *(submitted to) Int. Conf. on Visual Information Retrieval*, 1997.

[12] Yong Rui, Thomas S. Huang, Sharad Mehrotra, and Michael Ortega. A relevance feedback architecutre in content-based multimedia information retrieval systems. In *Proc of IEEE Workshop on Content-based Access of Image and Video Libraries*, 1997.

[13] Gerard Salton and Chris Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing and Management*, 1988.

[14] Gerard Salton and Michael J. McGill. *Introduction to Modern Information Retrieval.* McGraw-Hill Book Company, 1983.

[15] John R. Smith and Shih-Fu Chang. Tools and techniques for color image retrieval. In *IS & T/SPIE proceedings Vol.2670, Storage & Retrieval for Image and Video Databases IV.*