INFORMATION RETRIEVAL BEYOND THE TEXT DOCUMENT

YONG RUI, MICHAEL ORTEGA, THOMAS S. HUANG
BECKMAN INSTITUTE FOR ADVANCED SCIENCE AND TECHNOLOGY
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
URBANA, IL 61801, USA
E-MAIL:{YRUI, ORTEGA-B, HUANG}@IFP.UIUC.EDU,

SHARAD MEHROTRA
DEPARTMENT OF INFORMATION AND COMPUTER SCIENCE
UNIVERSITY OF CALIFORNIA, IRVINE
IRVINE, CA, 92697-3425
E-MAIL: SHARAD@ICS.UCI.EDU

Abstract

With the expansion of the Internet, searching for information goes beyond the boundary of physi libraries. Millions of documents of various media types, such as text, image, video, audio, graphics, a animation, are available around the world and linked by the Internet.

Unfortunately, the state of the art of search engines for media types other than text lags far behi their text counterparts. To address this situation, we have developed the Multimedia Analysis and Retrie System (MARS). This paper reports some of the progress made over the years towards explori Information Retrieval beyond the text domain. In particular, the following aspects of MARS are address in the paper: visual feature extraction, retrieval models, query reformulation techniques, efficient executi

speed performance and user interface considerations. Extensive experimental results are reported to valid the proposed approaches.

1 Introduction

Huge amounts of digital data are being generated every day. Scanners convert the analog/physi data into digital form; digital cameras and camcorders directly generate digital data at the production pha Owing to all these multimedia devices, nowadays information is in all media types, including graphi images, audio, and video, in addition to the conventional text media type. Not only is multime information being generated at an ever increasing rate, it is transmitted all over the world due to expansion of the Internet. Experts say that the Internet is the largest library that ever existed, it is howe also the most disorganized library ever.

Textual document retrieval has achieved considerable progress over the past two decad Unfortunately, the state of the art of search engines for media types other than text lags far behind their t counterparts. Textual indexing of non-textual media, although common practice, has some limitations. T most notable limitations include the human effort required and the difficulty of describing accurately cert properties humans take for granted while having access to the media. Consider how human indexers wou describe the ripples on an ocean; these could be very different under situations such as calm weather o hurricane. To address this situation, we undertook the Multimedia Analysis and Retrieval System (MAR project to provide retrieval capabilities to rich multimedia data. Research in MARS addresses several lev including the multimedia features extracted, the retrieval models used, query reformulation techniqu efficient execution speed performance and user interface considerations.

This paper reports some of the progress made over the years towards exploring Informati Retrieval (IR) beyond the text domain. In particular, this paper will concentrate on Visual Informati Retrieval (VIR) concepts as opposed to implementation issues. MARS explores many different visı feature representations. A review of these features appears in Section 2. These visual features are analogc to keyword features in textual media. Section 3 describes two broad retrieval models we have explored: t Boolean and vector models and the incorporated enhancements to support visual media retrieval such relevance feedback. Experimental results are given in Section 4. Concluding remarks are discussed Section 5.

2 Visual Feature Extraction

The retrieval performance of any IR system is fundamentally limited by the quality of the "feature and the retrieval model it supports. This section sketches the features obtained from visual media. In te based retrieval systems, features can be keywords, phrases or structural elements. There are ma techniques for reliably extracting, for example, keywords from text documents. The *visual counterparts* textual features in visual based systems are visual features such as color, texture, and shape.

For each feature there are several different techniques for representation. The reason for this twofold: a) the field is still under development; and b) more importantly, features are perceived differen by different people and thus different representations cater to different preferences. Image features a generally considered as orthogonal to each other. The idea is that a feature will capture some dimension the content of the image, and different features will effectively capture different aspects of the ima content. In this way two images closely related in one feature could be very different in another feature.

simple example of this are two images, one of a deep blue sky and the other of a blue ocean. These t images could be very similar in terms of just color, however the ripples caused by waves in the ocean ad distinctive pattern that distinguishes the two images in terms of their texture. (Rui et al., 1999) give detailed description of the visual features and the following paragraphs emphasize the important ones.

The *Color* feature is one of the most widely used visual features in VIR. The Color feature captu the color content of images. It is relatively robust to background complication and independent of ima size and orientation. Some representative studies of color perception and color spaces can be found (McCamy et al., 1976; Miyahara, 1988). In VIR, Color Histogram (Swain and Ballard, 1991), Co Moments (Stricker and Orengo, 1995) and Color Sets (Smith and Chang, 1995) are the most us representations.

Texture refers to the visual patterns that have properties of homogeneity that do not result from t presence of only a single color or intensity. It is an innate property of virtually all surfaces, including clou trees, bricks, hair, fabric, etc. It contains important information about the structural arrangement of surfa and their relationship to the surrounding environment (Haralick et al., 1973). Co-occurrence mat (Haralick et al., 1973), Tamura texture (Tamura et al., 1978), and Wavelet texture (Kundu and Chen, 19 are the most puopular texture representations.

In general, the *shape* representations can be divided into two categories, boundary-based and regi based. The former uses only the outer boundary of the shape while the latter uses the entire shape regi (Rui et al., 1996). The most successful representatives for these two categories are Fourier Descriptor a Moment Invariants. Some recent work in shape representation and matching includes the Finite Elem Method (FEM) (Pentland et al., 1996), Turning Function (Arkin et al., 1991), and Wavelet Descrip (Chuang and Kuo, 1996).

3 Retrieval Models used in MARS

With the large number of retrieval models proposed in the IR literature, MARS attempts to expl this research for content-based retrieval over images. The retrieval model comprises the document or obj model (here a collection of feature representations), a set of feature similarity measures, and a query mode

3.1 The Object Model

We first need to formalize how an object is modeled (Rui et al., 1998b). We will use images as example, even though this model can be used for other media types as well. An image object *O* represented as:

$$O = O(D, F, R) \qquad (1)$$

- D is the raw image data, e.g. a JPEG image.

- $F = \{f_i\}$ is a set of low-level visual features associated with the image object, such as color, textu and shape.

- $R = \{r_{ij}\}$ is a set of representations for a given feature f_i, e.g. both color histogram and co moments are representations for the color feature (Swain and Ballard, 1991). Note that, ea representation r_{ij} itself may be a vector consisting of multiple components, i.e.

$$r_{ij} = [r_{ij1}, \ldots r_{ijk}, \ldots r_{ijK}] \tag{2}$$

where K is the length of the vector.

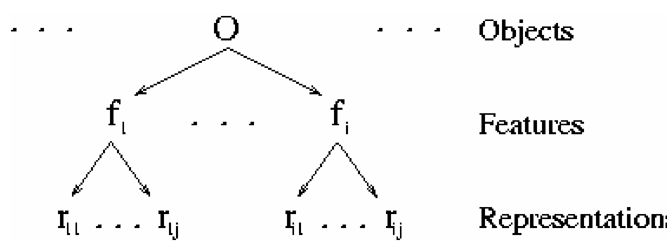


Figure 1: The Object Model

Figure 1 shows a graphic representation of the Object (Image) model. The proposed object mo supports multiple representations to accommodate the rich content in the images. An image is th represented as a collection of low-level image feature representations (Section 2) extracted automatica using computer vision methods, as well as a manual text description of the image.

Each feature representation is associated with some similarity measure (see section 2). All the similarity measures are normalized to lie within [0,1] to denote the degree to which two images are simi in regard to the same feature representation. A value of 1 means they are very similar and a value o means they are very dissimilar. Revisiting our blue sky and ocean example from section 2, the sky a ocean images may have a similarity of 0.9 in the Color Histogram representation of Color and 0.2 in t

Wavelet representation of Texture. Thus the two images are fairly similar in their color content, but ve different in their texture content. This mapping M={<*feature representation*$_i$, *similarity measure*$_i$>, . together with the Object model *O*, forms *(D, F, R, M)*, a foundation on which query models can be built.

3.2 Query Models

Based on the *object model* and the *similarity measures* defined above, Query models that work w these raw features are built. These Query models together with the Object model form complete retrie models used for VIR.

We explore two major models for querying. The first model is an adaptation of the Boolean retrie model to visual retrieval in which selected features are used to build predicates used in a Boole expression. The second model is a vector (weighted summation) model where all the features of the que object play a role in retrieval. Section 3.3 describes the Boolean model and Section 3.4 describes the vec model.

3.3 Boolean Retrieval

A user may not only be interested in a single feature from a single image. It is very likely that t user may choose multiple features from multiple images. For example, using a point-and-click interfac user can specify a query to retrieve images similar to an image *A* in color and similar to an image *B* texture. To cope with composite queries, Boolean retrieval model is used to interpret the query and retrie a set of images ranked based on their similarity to the selected feature.

The basic Boolean retrieval model needs a pre-defined threshold, which has several potent problems [Ortega et al. 1998b]. To overcome these problems, we have adopted the following two extensic to the basic Boolean model to produce a ranked list of answers.

Fuzzy Boolean Retrieval. The similarity between the image and the query feature is interpreted as t degree of membership of the image to the fuzzy set of images that match the query feature. Fuz set theory is used to interpret the Boolean query and the images are ranked based on their degree membership in the set.

Probabilistic Boolean Retrieval. The similarity between the image and the query feature is considered be the probability that the image matches the user's information need. Feature independence exploited to compute the probability of an image satisfying the query which is used to rank t images.

In the discussion below, we will use the following notation. Images in the collection are denoted I_1, I_2, ... I_m. Features over the images are denoted by F_1, F_2, ... F_r, where F_i denotes both the name of t feature as well as the domain of values that the feature can take. The j^{th} instance of feature F_i corresponds image I_j and is denoted by f_{ij}. For example, say F_1 is the color feature which is represented in the databa using a histogram. In that case, F_1 is also used to denote the set of all the color histograms, and $f_{1,5}$ is t color histogram for image 5. Query variables are denoted by v_1, v_2, ... v_n | $v_k \in F_i$ so each v_k refers to instance of a feature F_i (an f_{ij}). Note that $F_i(I_j)=f_{ij}$. During query evaluation, each v_k is used to rank imag in the collection based on the feature domain of f_i (F_i), that is v_k's domain. Thus, v_k can be thought of bei a list of images from the collection ranked based on the similarity of v_k to all instances of F_i. For examp say F_2 is the set of all wavelet texture vectors in the collection, if v_k=$f_{2,5}$, then v_k can be interpreted as bei both, the wavelet texture vector corresponding to image 5 and the ranked list of all $< I, S_{F_2}(F_2(I), f_{2,5})$ with S_{F_2} being the similarity function that applies to two texture values.

A query $Q(v_1, v_2, \ldots v_n)$ is viewed as a query tree whose leaves correspond to single feature varial queries. Internal nodes of the tree correspond to the Boolean operators. Specifically, non-leaf nodes are one of three forms: $\wedge(v_1, v_2, \ldots v_n)$, a conjunction of positive literals; $\wedge(v_1, v_2, \ldots v_p, \neg v_{p+1}, \ldots \neg v_n)$ conjunction consisting of both positive and negative literals; and $\vee(v_1, v_2, \ldots v_n)$, which is a disjunction positive literals. The following is an example of a Boolean query: $Q(v_1, v_2) = (v_1=f_{1,5}) \wedge (v_2=f_{2,6})$ is a que where v_1 has a value equal to the color histogram associated with image I_5 and v_2 has a value of the textu feature associated with I_6. Thus, the query Q represents the desire to retrieve images whose color match that of image I_5 *and* whose texture matches that of image I_6. Figure 2 shows an example query $Q(v_1, v_2, v_4)=((v_1=f_{1,4}) \wedge (v_2=f_{2,8}))\vee((v_3=f_{3,8}) \wedge \neg (v_4=f_{1,9}))$ in its tree representation.

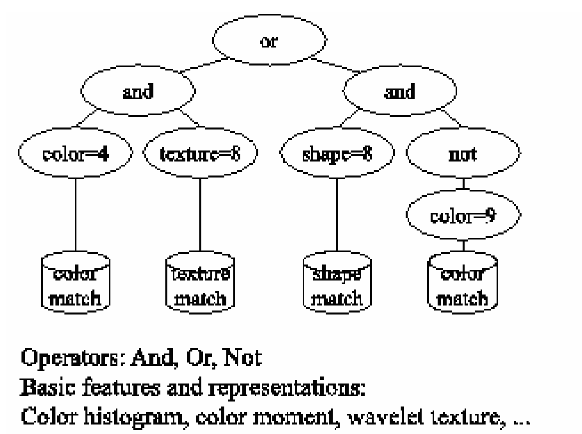


Figure 2 : Sample Query Tree

3.3.1 Weighting in the query tree

In a query, one feature can receive more importance than another according to the user's perceptio The user can assign the desired importance to any feature by a process known as *feature weightin* Traditionally, retrieval systems (Flickner et al., 1995; Bach et al. 1996) use a linear scaling factor as featu weights. Under our Boolean model, this is not desirable. Fagin and Wimmers (1997) noted that such lin weights do not scale to arbitrary functions used to compute the combined similarity of an image. T reason is that the similarity computation for a node in a query tree may be based on operators other tha weighted summation of the similarity of the children. Fagin and Wimmers (1997) present a way to exte linear weighting to the different components for arbitrary scoring functions as long as they satisfy cert properties. We are unable to use their approach since their mapping does not preserve orthogonal properties on which our algorithms rely (Ortega et al. 1998b). Instead, we use a mapping function from [0 $\rightarrow$ [0,1] of the form

$$similarity' = similarity^{\frac{1}{weight}}, 0 < weight < \infty \qquad (3)$$

which preserves the range boundaries [0,1] and boosts or degrades the similarity in a smooth way. Sam mappings are shown in Figure 3. This method preserves most of the properties explained in (Fagin a Wimmers, 1997), except it is undefined for a weight of 0. In (Fagin and Wimmers, 1997), a weight o means the node can be dismissed. Here, $\lim_{weight \rightarrow 0} similarity' = 0$ for similarity $\in$ [0,1). A perfect similar of 1 will remain at 1. This mapping is performed at each link connecting a child to a parent in the query tr

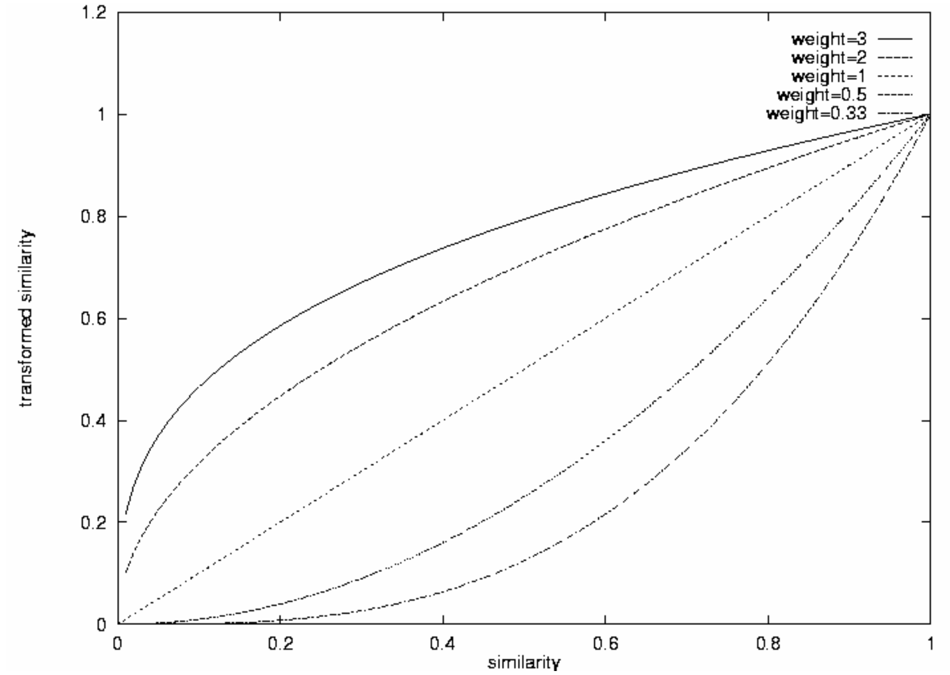


Figure 3 : Various samples for similarity mappings

Figure 4a) shows how the fuzzy model would work with our running example of blue sky and bl ocean images. Figure 4b) shows how the probabilistic model would work with our running example of bl sky and blue ocean images.

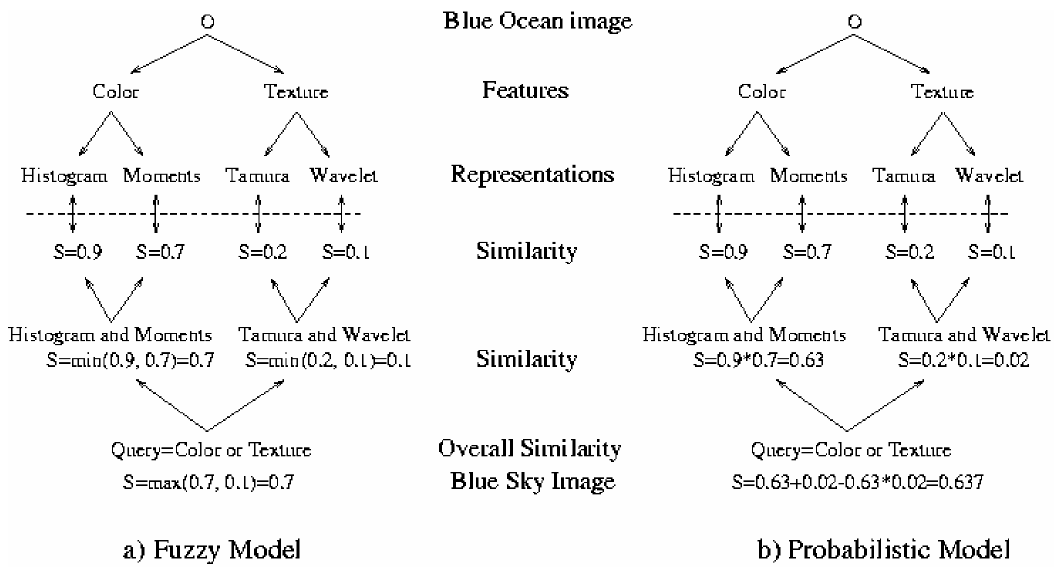


Figure 4 : Various samples for similarity mappings

3.3.2 Computing Boolean Queries

Fagin (1996) proposed an algorithm to return the top k answers for queries with monotonic scori functions that has been adopted by the Garlic multimedia information system under development at the IE Almaden Research Center (Fagin and Wimmers, 1997). A function F is monotonic if $F(x_1, \ldots x_m) \leq F(x'_1,$ $x'_m)$ for $x_i \leq x'_i$ for every i. Note that the scoring functions for both conjunctive and disjunctive queries : both the fuzzy and probabilistic Boolean models satisfy the monotonicity property. This algorithm relies reading a number of objects from each branch in the query tree until it has k objects in the intersecti Then it falls back on probing to enable a definite decision. In contrast, our algorithms (Ortega et al., 1998 are tailored to specific functions that combine object scoring (here called fuzzy and probabilistic models).

Another approach to optimizing query processing over multimedia repositories has been proposed (Chaudhari and Gravano, 1996). It presents a strategy to optimize queries when users specify thresholds the grade of match of acceptable objects as filter conditions. It uses the results in (Fagin, 1996) to conv top-k queries to threshold queries and then process them as filter conditions. It shows that under certa conditions (uniquely graded repository), this approach is expected to access no more objects than t strategy in (Fagin, 1996). Furthermore, while the above approaches have mainly concentrated on the fuz Boolean model, we consider both the fuzzy and probabilistic models in MARS. This is significant since t experimental results illustrate that the probabilistic model outperforms the fuzzy model in terms of retriev performance (discussed in section 4).

3.4 Vector Model

An IR model consists of a document model, a query model, and a model for computing similar between the documents and the queries. One of the most popular IR models is the vector model (Buckl and Salton, 1995; Salton and McGill, 1983; Shaw, 1995). Various effective retrieval techniques have be developed for this model. Among them, *term weighting* and *relevance feedback* are of fundamen importance.

Term weighting is a technique for assigning different weights for different keywords (term according to their relative importance to the document (Shaw, 1995; Salton and McGill, 1983). If we defi w_{ik} to be the weight for term t_k, k=1, ...,N, in document i (D_i), where N is the number of terms. Documen can be represented as a weight vector in the term space:

$$D_i = [w_{i1}, \ldots w_{ik}, \ldots w_{iN}] \tag{4}$$

Experiments have shown that the product of *tf* (term frequency) and *idf* (inverse docum

frequency) is a good estimation of the weights (Buckley and Salton, 1995; Salton and McGill, 1983; Sha 1995).

The query Q has the same model as that of a document D, i.e. it is a weight vector in the term spac

$$Q = [w_{q1}, \ldots w_{qk}, \ldots w_{qN}]. \qquad (5)$$

The similarity between D and Q is defined as the Cosine distance.

$$similarity(D,Q) = \frac{D \times Q}{\|D\| \times \|Q\|} \qquad (6)$$

where || || denotes norm-2.

As we can see from the previous subsection, in the vector model, the specification of w_{qk}'s in Q very critical, since the similarity values (*similarity(D, Q)*'s) are computed based on them. However, it usually difficult for a user to map his information need into a set of terms precisely. To overcome t difficulty, the technique of *relevance feedback* has been proposed (Buckley and Salton, 1995; Salton a McGill, 1983; Shaw, 1995). Relevance feedback is the process of automatically adjusting an existing que using information fed-back by the user about the relevance of previously retrieved documents. Te weighting and relevance feedback are powerful techniques in IR. We next generalize these concepts to VI

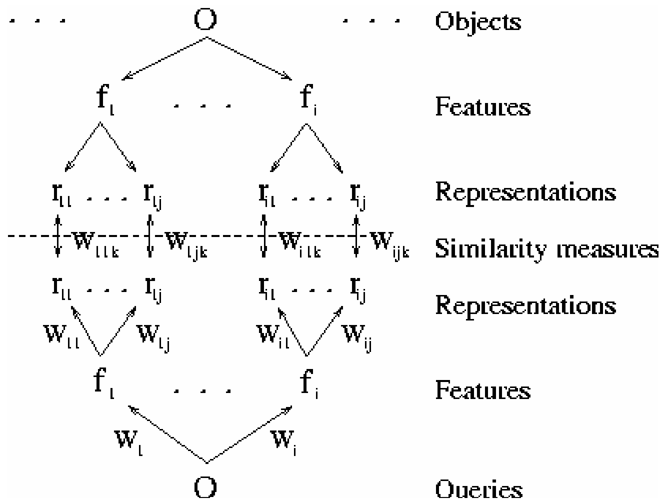


Figure 5 : The retrieval process

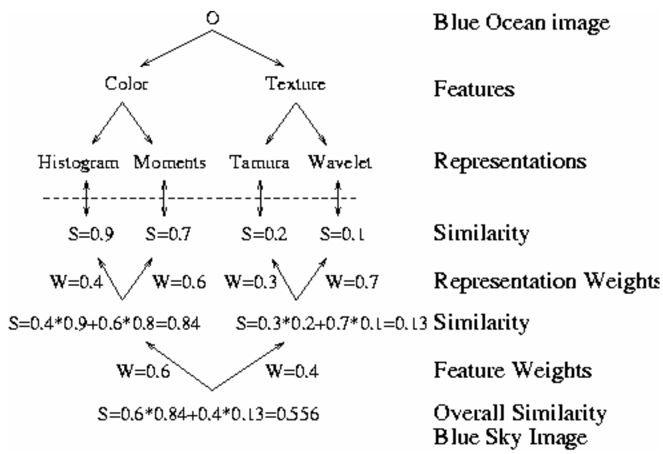


Figure 6 : Example Query calculation pf Blue Sky image against Blue Ocean image

3.4.1 Vector Query Model and Integration of Relevance Feedback to VIR

As discussed in section 3.1, an object model $O(D,F,R)$, together with a set of similarity measu $M=\{m_{ij}\}$, provides the foundation for retrieval (D,F,R,M). The similarity measures are used to determi how similar or dissimilar two objects are. Different similarity measures may be used for different featu representations. For example, Euclidean distance is used for comparing vector-based representations wh Histogram Intersection is used for comparing color histogram representations (see Section 2).

The Query model is shown in Figure 5. The query has the same form as an object, except it h weights at every branch at all levels. W_i, W_{ij}, and W_{ijk}, are associated with features f_i, representations r_{ij}, a components r_{ijk} respectively. The purpose of the weights is to reflect as closely as possible the combinati of feature representations that best represents the users information need. The process of relevance feedba described below aims at updating these weights to form the combination of features that best captures t user's information need.

Intuitively, the similarity between query and object feature representations is computed, and then t feature similarity computed as the weighted sum of the similarity of the individual feature representatio This process is repeated one level higher when the overall similarity of the object is the weighted sum ov all the feature similarities. The weights at the lowest level, the component level, are used by the differe similarity measures internally. Figure 6 traces this process for our familiar example of a blue sky image a query and a blue ocean image in the collection.

Based on the image object model and the set of similarity measures, the retrieval process can described as follows. At the initial query stage, equal weights are associated with the featur representations, and components. Best matches are then displayed back to the user. Depending on his tr information need, the user will mark how good the returned matches are (degree of relevance). Based user's feedback, the retrieval system will automatically update weights to match the user's true informati need. This process is also illustrated in Figure 5. In Figure 5, the information need embedded in Q flows while the content of O's flows down. They meet at the dashed line, where the similarity measures m_{ij} a applied to calculate the similarity values $S(r_{ij})$'s between Q and O's.

Based on the intuition that important representations or components should receive more weigh we have proposed effective algorithms for updating these two levels' weights. Due to page limitation, refer the readers to (Rui et al. 1998b).

4 Experimental Results

In the experiments reported here, we test our approaches over the image collection from the Fow Museum of Cultural History at the University of California-Los Angeles. It contains 286 ancient Afric and Peruvian artifacts and is part of the Museum Educational Site Licensing Project (MESL), sponsored the Getty Information Institute. The size of the MESL test set is relatively small but it allows us to expl all the color, texture, and shape features simultaneously in a meaningful way. More extensive experime with larger collections have been performed and reported in (Ortega et al., 1998b; Rui et al., 1998b).

In the following experiments, the visual features used are color, texture and shape of the objects the image. The representations used are color histogram and color moments (Swain and Ballard, 1991) the color feature; Tamura (Tamura et al., 1978; Equitz and Niblack, 1994) and co-occurrence mat (Haralick et al., 1973; Ohanian and Dubes, 1992) texture representations for the texture feature, and Four descriptor and chamfer shape descriptor (Rui et al., 1997b) for the shape feature.

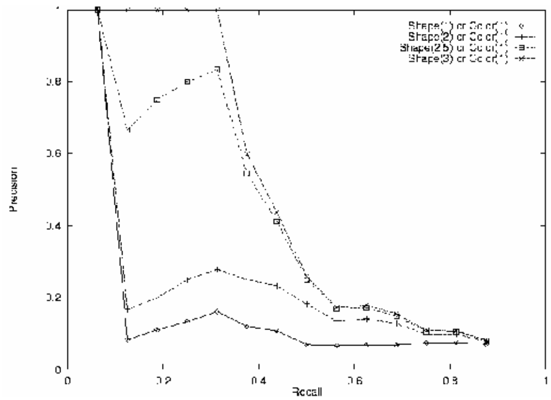
4.1 Boolean Retrieval Model Results

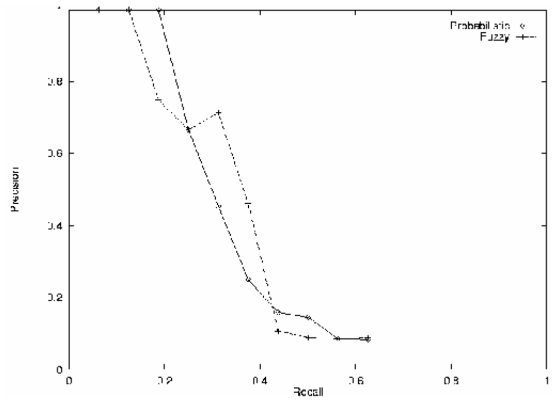
To conduct the experiments we chose several queries and manually determined the relevant set images with help of experts in librarianship as part of a seminar in multimedia retrieval. With the set queries and relevant answers for each of them, we constructed precision-recall curves (Salton and McG 1983). These are based on the well known precision and recall metrics. Precision measures the percentage relevant answers and recall measures the percent of relevant objects returned to the user. The precisi recall graphs are constructed by measuring the precision for various levels of recall.

We conducted experiments to verify the role of feature weighting in retrieval. Figure 7(a) sho results of a *shape or color* query i.e. to retrieve all images having either the same shape or the same color the query image. We obtained four different precision recall curves by varying the feature weights. T retrieval performance improves when the shape feature receives more emphasis.

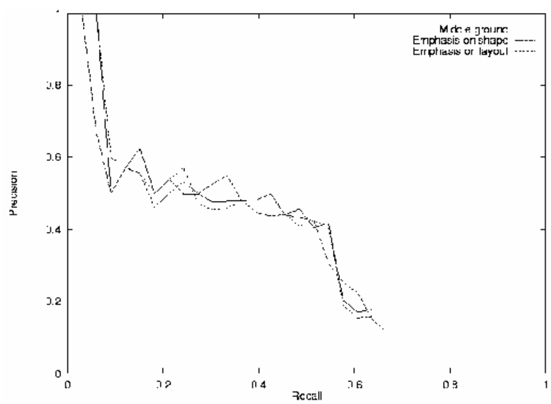
We also conducted experiments to observe the impact of the retrieval model used to evaluate t queries. We observed that the fuzzy and probabilistic interpretation of the same query yields differe results. Figure 7(b) shows the performance of the same query (a *texture or color* query) in the two mode The result shows that neither model is consistently better that the other in terms of retrieval.

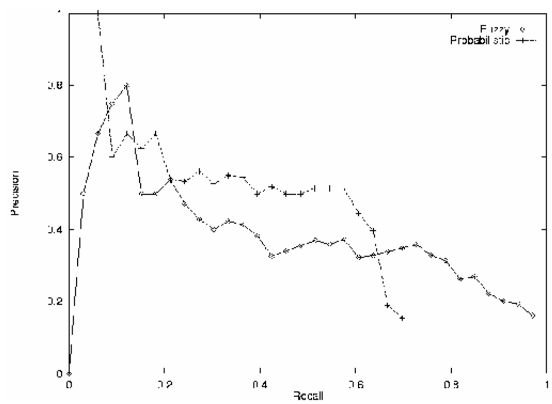
Figure 7(c) shows a complex query (shape(I_i) *and* color(I_i) *or* shape(I_j) *and* layout(I_j)) with differe weightings. The three weightings fared quite similar, which suggests that complex weightings may not ha a significant effect on retrieval performance. We used the same complex query to compare the performar of the retrieval models. The result is shown in Figure 7(d). In general, the probabilistic model outperfor the fuzzy model.





a) Effects of varying the weighting on a quer b) Fuzzy vs. Probabilistic performance for qu





c) Complex query with different weights d) Fuzzy vs. probabilistic for same complex q

Figure 7 : Experimental result graphs

4.2 Vector Retrieval Model with Relevance Feedback Results

There are two sets of experiments reported here. The first set of experiments is on the efficiency the retrieval algorithm, i.e. how fast the retrieval results converge to the true results. The second set experiments is on the effectiveness of the retrieval algorithm, i.e. how good the retrieval results a subjectively.

4.2.1 Efficiency of the Algorithm

As we have discussed in Section 3.1, the image object is modeled by the combinations of representatic with their corresponding weights. If we fix the representations, then a query can be complet characterized by the set of weights embedded in the query object *Q*. Obviously, the retrieval performance affected by the offset of the true weights from the initial weights. We thus classify the test into t categories, i.e. moderate offset, and significant offset, by considering how far away the true weights from the initial weights. The convergence ratio (recall) for these cases is summarized in Figure 8.

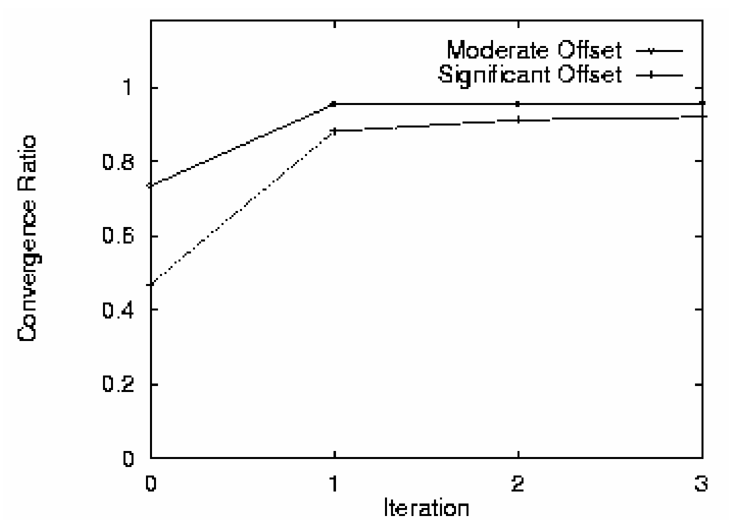


Figure 8 : Convergence Ratio curves

Based on the curves, some observations can be made:

- In all the cases, the convergence ratio (CR) increases the most in the first iteration. Later iteratic only result in minor increases in CR. This is a very desirable property, which ensures that the us gets reasonable results after only one-iteration of feedback.

- CR is affected by the degree of offset. The less the offset, the higher the final absolute C However, the more the offset, the higher the relative increase of CR.

4.2.2 Effectiveness of the Algorithm

Extensive experiments have been carried out. Users from various disciplines, such as Compu Vision, Art, Library Science, etc., as well as users from industry, have been invited to judge the retriev performance of the proposed *interactive* approach. A typical retrieval process on the MESL test set given in Figures 9 and 10.

Figure 9 : The retrieval results before the relevance feedback

Figure 10 : The retrieval results after the relevance feedback

The user can browse through the image database. Once he or she finds an image of interest, tl image is submitted as a query. In Figure 9, the query image is displayed at the upper-left corner and the b 11 retrieved images. The top 11 best matches are displayed in the order from top to bottom and from left right. The retrieved results are obtained based on their overall similarities to the query image, which a computed from all the features and all the representations. Some retrieved images are similar to the que image in terms of the shape feature while others are similar to the query image in terms of color or textu feature.

Assume the user's true information need is to “retrieve similar images based on their shapes”. In t

proposed retrieval approach, the user is no longer required to explicitly map his information need to lo level features, but rather he or she can express his intended information need by marking the relevar scores of the returned images. In this example, images 247, 218, 228 and 164 are marked *highly releva* Images 191, 168, 165, and 78 are marked *highly non-relevant*. Images 154, 152, and 273 are marked *r* *opinion*.

Based on the information fed-back by the user, the system *dynamically* adjusts the weights, putti more emphasis on the *shape feature*, possibly even more emphasis to one of the two shape representatic which better matches the user's subjective perception of shape. The improved retrieval results are display in Figure 10. Note that our shape representations are invariant to translation, rotation, and scalin Therefore, images 164 and 96 are relevant to the query image.

5 Conclusion

This paper discussed techniques to extend information retrieval beyond the textual doma Specifically, it discussed how to extract visual features from images and video; how to adapt a Boole retrieval model (enhanced with Fuzzy and Probabilistic concepts) for VIR systems; and how to general the relevance feedback technique to VIR.

In the past decade, two general approaches to VIR emerged. One is based on text (titles, keywor and annotation) to search for visual information indirectly. This paradigm requires much human labor a suffers from vocabulary inconsistency problems across human indexers. The other paradigm seeks to bu fully automated systems by completely discarding the text information and performing the search on visu information only. Neither paradigm has been very successful. In our view, these two paradigms have bo their advantages and disadvantages; and sometimes are complimentary to each other. For example, in t MESL database, it will be much more meaningful if we first do a text-based search to confine the catego and then use visual feature based search to refine the result. Another promising research direction is t integration of the human user into the retrieval system loop. A fundamental difference between an o Pattern Recognition system and today's VIR system is that the end user of the latter is human. I integrating human knowledge into the retrieval process, we can bypass the unsolved problem of ima understanding. Relevance feedback is one technique designed to deal with this problem.

6 Acknowledgements

This work was supported by NSF CAREER award IIS-9734300; in part by NSF CISE Resear Infrastructure Grant CDA-9624396; in part by the Army Research Laboratory under Cooperati Agreement No. DAAL01-96-0003. Michael Ortega is supported in part by CONACYT Grant 89061. Som example images used in this article are used with permission from the Fowler Museum of Cultural Histo at the University of California-Los Angeles.

7 References

[Arkin et al., 1991] Arkin, E. M., Chew, L., Huttenlocher, D., Kedem, K., and Mitchell, J. (1991). A efficiently computable metric for comparing polygonal shapes. IEEE Trans. Patt. Recog. and Mac Intell., 13(3).

[Bach et al., 1996] Bach, J. R., Fuller, C., Gupta, A., Hampapur, A., Horowitz, B., Humphrey, R., Jain, R., and fe Shu, C. The Virage image search engine: An open framework for image management. Proc. SPIE Storage and Retrieval for Image and Video Databases. February, 1996.

[Beyer et al., 1998] Beyer, K., Goldstein, J., Ramakrishnan, R., and Shaft, U. (1998). When Is "Near Neighbor" Meaningful? Submitted for publication.

[Buckley and Salton, 1995] Buckley, C. and Salton, G. (1995). Optimization of relevance feedback weigh In Proc. of SIGIR'95.

[Chaudhari and Gravano, 1996] Chaudhari, S. and Gravano, L. (1996). Optimizing Queries over Multimed Repositories. Proc. of SIGMOD.

[Chuang and Kuo, 1996] Chuang, G. C.-H. and Kuo, C.-C. J. (1996). Wavelet descriptor of planar curv Theory and applications. IEEE Trans. Image Proc., 5(1):56--70.

[Equitz and Niblack, 1994] Equitz, W. and Niblack, W. (1994). Retrieving images from a database usi texture - algorithms from the QBIC system. Technical Report RJ 9805, Computer Science, IE Research Report.

[Fagin, 1996] Fagin, R. (1996). Combining Fuzzy Information from Multiple Systems. Proc. of the 15 ACM Symp. on PODS.

[Fagin and Wimmers, 1997] Fagin, R. and Wimmers, E. L. (1997). Incorporating user preferences multimedia queries. In Proc of Int. Conf. on Database Theory.

[Flickner et al., 1995] Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorka M., Hafine, J., Lee, D., Petkovic, D., Steele, D., and Yanker, P. (1995). Query by image and vid content: The QBIC system. IEEE Computer.

[Haralick et al., 1973] Haralick, R. M., Shanmugam, K., and Dinstein, I. (1973). Texture features for ima classification. IEEE Trans. on Sys, Man, and Cyb, SMC-3(6).

[Hu, 1962] Hu, M. K. (1962). Visual pattern recognition by moment invariants, computer methods in ima analysis. IRE Transactions on Information Theory, 8.

[Kundu and Chen, 1992] Kundu, A. and Chen, J.-L. (1992). Texture classification using qmf bank-bas subband decomposition. CVGIP: Graphical Models and Image Processing, 54(5):369--384.

[McCamy et al., 1976] McCamy, C. S., Marcus, H., and Davidson, J. G. (1976). A color-rendition cha Journal of Applied Photographic Engineering, 2(3).

[Miyahara, 1988] Miyahara, M. (1988). Mathematical transform of (r,g,b) color data to munsell (h,s,v) co data. In SPIE Visual Communications and Image Processing, volume 1001.

[Ortega et al., 1998a] Ortega, M., Chakrabarti, K., Porkaew, K., and Mehrotra, S. (1998a). Cross me validation in a multimedia retrieval system. ACM Digital Libraries 98 Workshop on Metrics Digital Libraries.

[Ortega et al., 1997] Ortega, M., Rui, Y., Chakrabarti, K., Mehrotra, S., and Huang, T. S. (1997). Supporti similarity queries in MARS. In Proc. of ACM Conf. on Multimedia.

[Ortega et al., 1998b] Ortega, M., Rui, Y., Chakrabarti, K., Porkaew, K., Mehrotra, S., and Huang, T. (1998b). Supporting ranked Boolean similarity queries in mars. IEEE Trans. on Knowledge and D Engineering, 10(6).

[Pentland et al., 1996] Pentland, A., Picard, R. W., and Sclaroff, S. (1996). Photobook: Content-bas manipulation of image databases. International Journal of Computer Vision.

[Rui et al., 1996] Rui, Y., She, A. C., and Huang, T. S. (1996). Modified Fourier descriptors for sha representation - a practical approach. In Proc of First International Workshop on Image Databas and Multi Media Search.

[Rui et al., 1997a] Rui, Y., Huang, T. S., and Mehrotra, S. (1997a). Content-based image retrieval w relevance feedback in MARS. In Proc. IEEE Int. Conf. on Image Proc.

[Rui et al., 1998a] Rui, Y., Huang, T. S., and Mehrotra, S. (1998a). Exploring video structures beyond t shots. In Proc. of IEEE conf. Multimedia Computing and Systems.

Rui et al., 1998b] Rui, Y., Huang, T. S., Ortega, M., and Mehrotra, S. (1998b). Relevance feedback: power tool in interactive content-based image retrieval. IEEE Transactions on Circuits and Syste for Video Technology, 8(5).

[Rui et al., 1999] Rui, Y., Huang, T. S., Chang, S.-F. (1999). Image Retrieval: Past, Present, and Futu Accepted to International Journal on Visual Communiation and Image Representation, 1999.

[Salton and McGill, 1983] Salton, G. and McGill, M. J. (1983). Introduction to Modern Informati Retrieval. McGraw-Hill Book Company.

[Shaw, 1995] Shaw, W. M. Term-relevance computations and perfect retrieval performance. Informati Processing and Management. Vol 31, no 4. Pp491-498.

[Smith and Chang, 1995] Smith, J. R. and Chang, S.-F. (1995b). Tools and techniques for color ima retrieval. In IS & T/SPIE proceedings Vol.2670, Storage & Retrieval for Image and Video Databas IV.

[Stricker and Orengo, 1995] Stricker, M. and Orengo, M. (1995). Similarity of color images. In Proc. SP Storage and Retrieval for Image and Video Databases.

[Swain and Ballard, 1991] Swain, M. and Ballard, D. (1991). Color indexing. International Journal Computer Vision, 7(1).

[Tamura et al., 1978] Tamura, H., Mori, S., and Yamawaki, T. (1978). Texture features corresponding visual perception. IEEE Trans. on Sys, Man, and Cyb, SMC-8(6).

8 Author Biographies

Yong Rui received the B.S. degree from Southeast University, P. R. China in 1991 and the M degree from Tsinghua University, P. R. China in 1994, both in Electrical Engineering. He received Ph.D. degree in Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign 1999. Since March, 1999, he is a researcher at Microsoft Research, Redmond, WA. His research intere include multimedia information retrieval, multimedia signal processing, computer vision and artific intelligence. He has published over 30 technical papers in the above areas. He is a Huitong Univers Fellowship recipient 1989-1990, a Guanghua University Fellowship recipient 1992-1993, and a C Engineering College Fellowship recipient 1996-1998.

Michael Ortega Received his B.E. degree with honors from the Mexican Autonomous Institute Technology in Aug. 1994 with a SEP fellowship for the duration of the studies. Currently he is pursuing graduate studies at the University of Illinois at Urbana Champaign. Michael Ortega received Fulbright/CONACYT/García Robles scholarship to pursue graduate studies as well as the Mavis Award the University of Illinois and is a member of the Phi Kappa Phi honor society, the IEEE computer soci and member of the ACM. His research interests include multimedia databases, database optimization : uncertainty support and content based multimedia information retrieval.

Thomas S. Huang received his B.S. Degree in Electrical Engineering from National Taiw University, Taipei, Taiwan, China; and his M.S. and Sc.D. Degrees in Electrical Engineering from t Massachusetts Institute of Technology, Cambridge, Massachusetts. He was on the Faculty of t Department of Electrical Engineering at MIT from 1963 to 1973; and on the Faculty of the School Electrical Engineering and Director of its Laboratory for Information and Signal Processing at Purc University from 1973 to 1980. In 1980, he joined the University of Illinois at Urbana-Champaign, where is now William L. Everitt Distinguished Professor of Electrical and Computer Engineering, and Resear Professor at the Coordinated Science Laboratory, and Head of the Image Formation and Processing Gro at the Beckman Institute for Advanced Science and Technology.

Dr. Huang's professional interests lie in the broad area of information technology, especially t transmission and processing of multidimensional signals. He has published 12 books, and over 300 pap in Network Theory, Digital Filtering, Image Processing, and Computer Vision. He is a Fellow of t International Association of Pattern Recognition, IEEE, and the Optical Society of American; and h received a Guggenheim Fellowship, an A.V. Humboldt Foundation Senior U.S. Scientist Award, anc Fellowship from the Japan Association for the Promotion of Science. He received the IEEE Acousti Speech, and Signal Processing Society's Technical Achievement Award in 1987, and the Society Award 1991. He is a Founding Editor of the International Journal Computer Vision, Graphics, and Ima Processing; and Editor of the Springer Series in Information Sciences, published by Springer Verlag.

Sharad Mehrotra received his M.S. and PhD at the University of Texas at Austin in 1990 and 19 respectively, both in Computer Science. Subsequently he worked at MITL, Princeton as a scientist fr

1993-1994. He is an assistant professor in the Computer Science department at the University of Illinois Urbana-Champaign since 1994. He specializes in the areas of database management, distributed syster and information retrieval. His current research projects are on multimedia analysis, content-based retriev of multimedia objects, multidimensional indexing, uncertainty management in databases, and concurren and transaction management. Dr. Mehrotra is an author of over 50 research publications in these areas. I Mehrotra is the recipient of the NSF Career Award and the Bill Gear Outstanding junior faculty award 1997.