A UNIFIED FRAMEWORK FOR SIMILARITY CALCULATION BETWEEN IMAGES

Azadeh Kushki, Panagiotis Androutsos, Konstantinos N. Plataniotis, Anastasios N. Venetsanopoulos

Multimedia Laboratory, University of Toronto
10 King's College Road, Toronto, Ontario M5S 3G4, Canada
{azadeh, oracle, kostas, anv}@dsp.toronto.edu

ABSTRACT

Emergence of large digital visual data repositories motivates the need for effective indexing and retrieval tools. In traditional approaches, similarity among two images is a weighted distance between two feature vectors. These techniques, however, do not provide enough flexibility in modelling user queries. This paper proposes a framework where image similarities are obtained through a decision making process using fuzzy theory. The analysis and results of this paper indicate that the aggregation technique presented here provides an effective, general, and flexible tool for similarity calculation based on the combination of individual descriptors and features.

1. INTRODUCTION

Content-Based Image Retrieval (CBIR) has attracted much attention as a tool for locating relevant items within large databases. In CBIR, visual features such as shape, colour, and texture are extracted to characterize images in a machine-understandable format. During the retrieval, users specify queries by providing the system with an example image containing the concepts or features of interest. The visual features of the query are then compared to those of the images in the available database and a ranked set of similar images to the user query is returned.

Generally, a set of complementary visual features are employed during the retrieval to capture information about various properties of the indexed images. The use of multiple features or descriptors gives rise to several possibilities for similarity calculations. One of the most common approaches is to stack the feature description values as a single, large vector and calculate the distance between these "*super-vectors*" in a high dimensional feature space using a measure such as the Euclidean distance and its variants [1, 2]. The use of the Euclidean distance for distance combination, however, does not always provide perceptually meaningful results as the assumption of a Euclidean feature space is generally incorrect [3]. Furthermore, a unilaterally employed distance measures may not necessarily be meaningful for all feature descriptors. In light of this, several systems use distinct distance measures at the descriptor-level before combining resulting distances via an *aggregation operator* to form the overall distance [4, 5]. The weighted average (WA) aggregation operator is very popular among the existing systems. Use of the weighted average, however, imposes a certain structure on the conceptual user queries that can be modelled by a CBIR system. In other words, formulation of queries involving the logical combination of features (i.e. AND, OR, and, NOT) is not possible with this operator. In addressing this issue, systems such as BlobWorld [6], represent the user query as a fuzzy boolean expression over image features using these logical connectives. Yet, decision making which is based on the aggregation of criteria is generally not conjunctive or disjunctive [7], indicating a need for the availability of an operator which has the potential to make such compromises as well as possibly providing disjunctive and conjunctive behaviour.

In light of the aforementioned shortcomings, this work proposes a Unified Framework for Similarity Calculation (UFSC) for the combination of low-level feature similarities. Unlike approaches which directly combine feature distances to arrive at a final, overall similarity value, *decisions* made according to individual distances are *aggregated* to achieve the same effect but with superior flexibility and generality. It is important to mention here that similarity calculations are *not* performed in a high dimensional space, but rather, within a *decision space*. This allows the incorporation of perceptual characteristics of the human decision making process into the design. Furthermore, the proposed system utilizes an entire family of aggregators for feature combination and thus exhibits a high level of flexibility through the use of a wide range of tools for modelling user intentions.

The rest of this paper is organized as follows: Section 2 provides the details of the proposed approach; Experiments and results are given in Section 3; Finally, Section 4 concludes the paper and provides some directions for fu-

This work is partially supported by the National Science and Engineering Council (NSERC).

ture work.

2. PROPOSED FRAMEWORK

2.1. Decision-Space Formulation

We define the conceptual fuzzy set $\mathcal{S}_Q$ as the set of images similar to the user query Q. Each database image I_i belongs to this set to a certain degree or membership grade $\mu_{\mathcal{S}_Q}(I_i)$. Thus,

$$\mathcal{S}_Q = \{\mu_{\mathcal{S}_Q}(I_1)/I_1, \mu_{\mathcal{S}_Q}(I_2)/I_2, \ldots, \mu_{\mathcal{S}_Q}(I_N)//I_N\}. \tag{1}$$

The value $\mu_{\mathcal{S}_Q}(I_i)$ represents the similarity score between the database images and the query; the higher the grade of membership of a database image, I_i, to this set, the more similar it is to the query. Similarity calculation to a given query, thus, translates to determining the membership of each database image to $\mathcal{S}_Q$. The rest of this section is dedicated to an explanation of how this membership values are obtained.

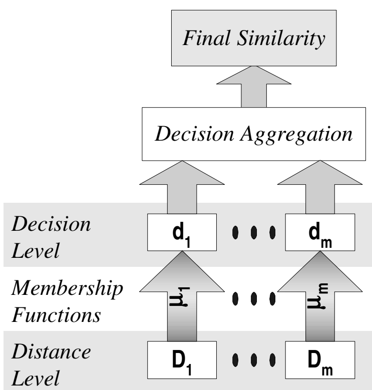

Fig. 1. UFSC overall structure.

In calculating the similarity between two images, each visual feature descriptor is considered individually. Thus, for descriptor j, the distance between the query Q and a given image I_i, denoted as $D_j(Q, I_i)$, is calculated using a distance measure devised with consideration for the nature of the particular descriptor feature space. Fig.1 depicts how $\mu_{\mathcal{S}_Q}(I_i)$ is obtained from the individual feature descriptor distances. From each of these distances, a decision $d_j(Q, I_i)$ is obtained. This decision represents the membership grade of image I_i to $\mathcal{S}_Q$ based only on descriptor j. These decisions are derived from the feature distances through *fuzzy membership functions*. The form of the membership function is highly dependent on the nature of the descriptors and distance measures. The effects of this function on retrieval accuracy along with some example functions are further discussed in Section 3.

As depicted in Fig.1, the final stage in similarity calculation is the aggregation of feature descriptor decisions $d_j(Q, I_i)$. The choice of the aggregation operator $\odot$ is discussed in the next section. For now, let us denote the aggregation operator as $\odot$ to obtain the overall similarity decision or membership value $\mu_{\mathcal{S}_Q}(I_i)$ as follows:

$$\mu_{\mathcal{S}_Q}(I_i) = \bigodot_k \left(\bigodot_{j \in R_k} d_j(Q, I_i) \right), \tag{2}$$

where R_k denotes the k^{th} grouping of the descriptors. The separate grouping of descriptors allows a hierarchical aggregation for taking into account the redundancy between available descriptions. For example, all colour descriptions maybe aggregated together before combination with texture descriptions.

A summary of the UFSC algorithm as discussed in this section is shown in Fig.2.

UFSC Algorithm:

Step 1. Calculate descriptor distances $D_j, 1 \leq j \leq m$.
Step 2. Generate similarity decisions for each descriptor using membership functions.
Step 3. Aggregate descriptor decisions using (2).

Fig. 2. Step-by-step outline of the UFSC operation.

2.2. Aggregation Operator

In order to allow the formulation of various logical query structures, the UFSC moves away from the use of a single aggregator for decision combination and employs a *set* of aggregation operators for addressing various logical conceptual queries.

As previously mentioned, in multi-criteria decision making, the aggregation function is generally neither purely conjunctive nor disjunctive [7] and lies somewhere between the two extremes. The aggregation operator should thus be chosen from an *array* of operators that stretch the full range between AND and OR depending on the nature of the query. To this end, the following family of *compensatory operators* [7] is considered:

$$x_1 \odot \ldots \odot x_n = \gamma \max(x_1, \ldots, x_n) + (1-\gamma) \min(x_1, \ldots, x_n), \tag{3}$$

where, $0 \leq \gamma \leq 1$ is a compensation parameter. Equation (3) represents a family of operators generated by a weighted mean of the logical *AND* and *OR* providing various degrees of compromise depending on the choice of the compensation parameter, γ. Specifically, the choices $\gamma = 0$ and $\gamma = 1$

lead to the logical AND and OR connectives, respectively. For other values of this parameter, a compromise operator is obtained. It must be noted that the behaviour of the NOT operator can be achieved by complementing a decision before aggregation:

$$\overline{d_j(Q, I_i)} = 1 - d_j(Q, I_i). \quad (4)$$

Further desirable properties of this compensative operator and its suitability for similarity decision aggregation are discussed in [8].

As shown in Section 3, the choice of the aggregation operator (by means of the compensation parameter) has a great effect on the performance of the system. Thus, it is important that this parameter is chosen properly for a given query. This selection can be either performed manually or automatically. For the former case, the system user selects the value of the compensation parameter considering the structure of their conceptual query (conjunctive, disjunctive, or compromise). It is also possible that the compensation parameter is adaptively set based on user needs during an interactive process known as relevance feedback [8].

3. EXPERIMENTS & RESULTS

This section describes the experiments conducted to evaluate the retrieval accuracy and flexibility of the proposed UFSC method.

The low-level features selected for image representation are colour and texture. Three of the MPEG-7 descriptors are used for colour; Color Structure, Dominant Color, and Color Layout. The MPEG-7 texture descriptors used are Edge Histogram and Homogeneous Texture. Details regarding both the colour and texture descriptors and the corresponding distance measures are discussed in [9].

The experiments of this section are carried out using a database of 2850 general images from the Corel collection. The performance measure used was the precision-recall measure as in [10]:

$$\begin{array}{rcl} R_k & = & RR_k/N_i \\ P_k & = & RR_k/k \end{array}, \quad (5)$$

where, R_k and P_k are recall and precision after k images are retrieved respectively, RR_k is number of matches after k retrievals, and N_i is the total number of images in category i determined by a human subject. This measure was then averaged over 10 different query categories.

Fig.3 provides a comparison between the UFSC method and two of the most common approaches to similarity calculations namely, the weighted average and Euclidean distance.

The superiority of the proposed approach is contributed to the non-Euclidean nature of the descriptor spaces and inflexibility of the weighted average in modelling user queries.

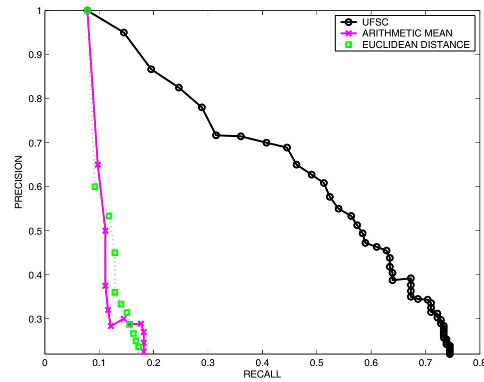

Fig. 3. UFSC vs. Arithmetic Mean and Euclidean distance.

The UFSC framework allows each descriptor to employ a distinct distance measure devised with consideration for the feature space. In addition to the ability to incorporate various descriptor distance measures, the UFSC framework allows descriptors of different nature and size to be utilized in a system. Furthermore, the proposed approach allows for modelling a wider range of conceptual queries than systems using the Euclidean distance, or the weighted average. The advantage of the proposed technique, thus, is seen in its flexibility to easily adapt to different design constraints.

Fig.4 compares the performance of the UFSC framework using various values of the compensation parameter γ. As seen from this plot, the aggregation operator has a great effect on system performance. This is further indication that the use of a fixed aggregator, such as the weighted average, may severely limit the system performance. Although the optimum values of γ have been obtained through an exhaustive search here, this manual optimization is not necessary during actual system functionality. Interactive methods such as Query Feedback [8] are used to determine this parameter adaptively for each query.

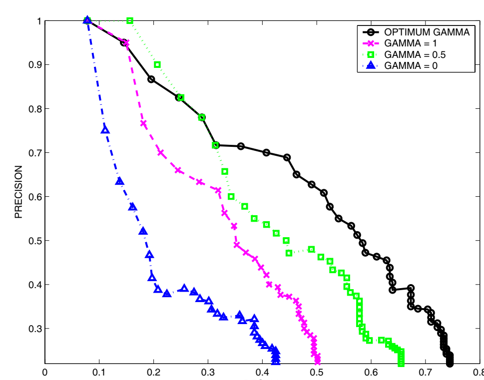

Fig. 4. Performance of the UFSC using different γ values.

Our system offers the functionality of a set of aggregation operators through the adjustment of a compensation parameter. Boolean expression-based systems such as *Blob-World* can in fact be thought of as specific realizations of the UFSC where the compensation parameter is restricted to only taking on binary values of 0 or 1. In this sense, the proposed framework provides a generalization of such systems.

Finally, the performance of the UFSC framework is depicted in Fig.5 for various membership functions. These include simple normalization by the maximum distance, an exponential mapping between distances and decisions, and an S-shaped function of the following form:

$$d(Q, I_i) = \frac{1}{1 + \exp\left(\frac{D(Q,I_i)}{D_{max}}\right)}, \quad (6)$$

where D_{max} is the maximum possible feature descriptor distance.

It can be seen that the exponential and S functions provide reasonable performance. More sophisticated membership functions can also be designed with consideration for the human perceptual characteristics.

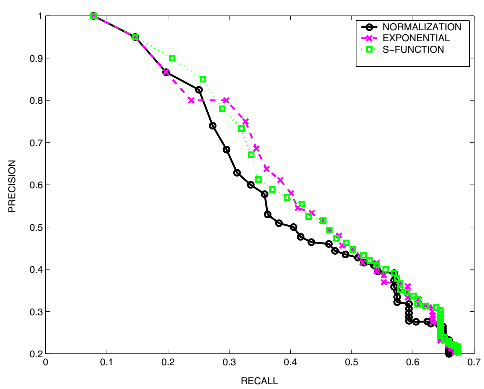

Fig. 5. Retrieval results using different membership functions.

4. CONCLUSIONS

The problem of similarity calculations between images is one of the key challenges in the area of content-based image retrieval. Techniques using the weighted average or the Euclidean distance do not provide enough flexibility for modelling user queries and may not always lead to perceptual meaningful results. In this paper, a similarity calculation framework is proposed based on decision combination. The proposed framework employs membership functions for obtaining similarity decisions. As future research, the design of these functions for incorporation of perceptual elements of human decision making into similarity calculations will be investigated.

Another unique characteristic of the proposed framework is the use of a *compensatory operator* for combining feature similarities. This operator provides the designer and the user with a set of tools for modelling various conceptual queries. Finally, the UFSC framework provides great flexibility in choosing the feature set employed in the measurements in terms of size of descriptor and distance measures.

5. REFERENCES

[1] G. Guo, A. Jain, W. Ma, H. Zhang, “Learning similarity measure for natural image retrieval with relevance feedback”, *IEEE Transactions on Neural Networks*, vol.13, no.4, 2002.

[2] J. Wang, J. Li, G. Wiederhold, “Simplicity: Semantics-sensitive integrated matching for picture libraries”, *IEEE Transactions on Pattern Analysis and Machine Intenlligence*, vol.23, no.9, pp.947-963, 2001.

[3] S. Antani, R. Kasturi, and R. Jain, “A survey on the use of pattern recognition methods for abstraction, indexing, and retrieval of images and video”, *Pattern Recognition*, vol.35, pp.945-965, 2002.

[4] I. Cox, M. Miller, T. Minka, T. Papathomas, P. Yianilos, “The Bayesian image retrieval system, PicHunter: theory, implementation, and psychological experiments”, *IEEE Transactions on Image Processing*, vol.9, no.1, pp.20-37, 2000.

[5] H. Yoo, D. Jang, S. Jung, J. Park, K. Song, “Visual information retrieval via content-based approach”, *Pattern Recognition*, vol.35, pp. 749-769, 2002.

[6] C. Carson, S. Belongie, H. Greenspan, J. Malik, “BlobWorld: image segmentation using expectation-maximization and its applications to image querying”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.24, no.8, pp.1026-1038, 2002.

[7] K.N. Plataniotis, A.N. Venetsanopoulos, *Color image processing and applications*, Springer-Verlag, 2000.

[8] A. Kushki, “An Interactive Framework for Content-Based Image Retrieval”, Master’s Thesis, Unversity of Toronto, 2003.

[9] B.S. Manjunath, P. Salembier, T. Sikora, *Introduction to MPEG-7: multimedia content description interface*, Wiley, 2002.

[10] J. Smith, “Quantitative assement of image retreival effectivess”, *Journal of the American Society for Information Science and Technology*, vol.52, no.11, pp.969-979, 2001.