FIRM: Fuzzily Integrated Region Matching for Content-based Image Retrieval

Yixin Chen
Department of Computer Science and Engineering
Pennsylvania State University
University Park, PA 16801
yixchen@cse.psu.edu

James Z. Wang
School of Information Sciences and Technology
Pennsylvania State University
University Park, PA 16801
jwang@ist.psu.edu

Jia Li
Department of Statistics
Pennsylvania State University
University Park, PA 16801
jiali@stat.psu.edu

ABSTRACT

We propose FIRM (Fuzzily Integrated Region Matching), an efficient and robust similarity measure for region-based image retrieval. Each image in our retrieval system is represented by a set of regions that are characterized by fuzzy sets. The FIRM measure, representing the overall similarity between two images, is defined as the similarity between two families of fuzzy sets. Compared with similarity measures based on individual regions and on all regions with crisp feature representations, our approach greatly reduces the influence of inaccurate segmentation. Experimental results based on a database of about 200,000 general-purpose images demonstrate improved accuracy, robustness, and high speed.

1. INTRODUCTION

Similarity comparison is one of the most important issues in content-based image retrieval (CBIR). In general, the comparison is performed either globally using techniques such as histogram matching and color layout indexing, or locally based on decomposed regions (objects) of the images. A major drawback of the global histogram search lies in its sensitivity to intensity variations, color distortions, and cropping. Color layout indexing is proposed to tackle this problem. However, it is in general sensitive to shifting, cropping, scaling, and rotation. In a human visual system, although color and texture are fundamental aspects of visual perceptions, human discernment of certain visual contents could potentially be associated with interesting classes of objects, or semantic meanings of objects in the image. Motivated by this intrinsic attribute of human visual perception, a region-based retrieval system applies image segmentation to decompose an image into regions, which correspond to objects if the decomposition is ideal. Since the retrieval system has identified objects in the image, it is relatively easy for the system to recognize similar objects at different locations and with different orientations and sizes.

Semantically precise image segmentation by an algorithm is very difficult. However, a single glance is sufficient for human to identify circles, straight lines, and other complex objects in a collection of points and to produce a meaningful assignment between objects and points in the image. Although those points cannot always be assigned unambiguously to objects, human recognition performance is hardly affected. We can often identify the object of interest correctly even when its boundary is very blurry. Based upon the above observations, we believe that, by softening the boundaries between regions, the robustness of a region-based image retrieval system against segmentation-related uncertainties can be improved. In this paper, we present FIRM, a novel similarity measure for region-based image retrieval. To describe the indistinct boundaries between segmented regions, we represent each region as a multidimensional fuzzy set (fuzzy feature) in the feature space. Thus, each image can be characterized by a class of fuzzy features, and, as a result, region matching becomes an issue of finding similarities between fuzzy sets. The FIRM measure is defined as the similarity between two families of fuzzy features.

2. IMAGE SEGMENTATION AND FUZZIFICATION

To segment an image, the system partitions the image into blocks with 4×4 pixels and extracts a feature vector for each block. The k-means algorithm is used to cluster the feature vectors into several classes with every class corresponding to one region in the segmented image. The number of clusters (regions) is selected adaptively. Six features are used for segmentation, as presented in [2]. Three of them are the average color components in a 4×4 block. We use the well-known LUV color space. The other three represent energy in the high frequency bands of the wavelet transforms, that is, the square root of the second order moment of wavelet coefficients in high frequency bands. To obtain them, a Daubechies-4 wavelet transform is applied to the L component of the image. After a one-level wavelet transform, a 4×4 block is decomposed into four frequency bands: the LL, LH, HL, and HH bands. Three features are computed from the HL, LH, and HH bands. Moments of wavelet coefficients in various frequency bands have been shown to be effective for representing texture [3].

Here are some notations. $\mathbb{R}$ and $\mathbb{N}$ denote the sets of real numbers and positive integers, respectively. $\vec{f}_i \in \mathbb{R}^6$ is the feature vector (used for segmentation) of the ith block of an image. $\mathbf{F} = \{\vec{f}_i \in \mathbb{R}^6 : 1 \leq i \leq B, i \in \mathbb{N}\}$ is the set of all block feature vectors of an image, where B is the number of blocks in an image in our database. Feature set $\mathbf{F}_j \subset \mathbf{F}$ contains all feature vectors in the jth cluster, where $1 \leq j \leq C, j \in \mathbb{N}$, $C \geq 2$ is the number of clusters. We also define the center of $\mathbf{F}_j$ as

$$\hat{\vec{f}}_j = \frac{\sum_{\vec{f}:\vec{f}\in\mathbf{F}_j} \vec{f}}{\mathrm{Card}(\mathbf{F}_\mathrm{j})}$$

where $\mathrm{Card}(\mathbf{F}_\mathrm{j})$ is the cardinality (or size) of $\mathbf{F}_j$. $\mathbf{R}_j \subset \mathbb{N}^2$ is the region (set of pixels) corresponding to feature set $\mathbf{F}_j$.

To describe shape properties, three extra features are calculated for each region. They are normalized inertia of order 1 to 3. For a region $\mathbf{R}_j \subset \mathbb{N}^2$ in the image plane, which is a finite set, the normalized inertia of order γ is given as

$$I_{(\mathbf{R}_j,\gamma)} = \frac{\sum_{(x,y):(x,y)\in\mathbf{R}_j} [(x-\hat{x})^2 + (y-\hat{y})^2]^{\frac{\gamma}{2}}}{\mathrm{Card}(\mathbf{R}_\mathrm{j})^{1+\frac{\gamma}{2}}},$$

where $(\hat{x}, \hat{y})$ is the centroid of $\mathbf{R}_j$. The normalized inertia is invariant to scaling and rotation. The minimum normalized inertia is achieved by spheres. Denote the γth order normalized inertia of spheres as $\mathbf{I}_\gamma$. We define shape feature $\vec{h}_j$ of region $\mathbf{R}_j$ as $I_{(\mathbf{R}_j,\gamma)}$ normalized by $\mathbf{I}_\gamma$, i.e.,

$$\vec{h}_j = \left[\frac{I_{(\mathbf{R}_j,1)}}{\mathbf{I}_1}, \frac{I_{(\mathbf{R}_j,2)}}{\mathbf{I}_2}, \frac{I_{(\mathbf{R}_j,3)}}{\mathbf{I}_3}\right]^T.$$

After segmentation, an image can be viewed as a collection of regions. Equivalently, in the feature space, a segmented image is characterized by a collection of feature sets. These feature sets form a partition of $\mathbf{F}$, i.e., $\forall \vec{f} \in \mathbf{F}$, $\vec{f}$ belongs to exactly one feature set. However, segmentation can not be perfect. As a result, for many feature vectors, a unique decision between in and not in the feature set is impossible. Only a degree (between 0 and 1) of membership that it belongs to some feature set should be given, and it could belong to several feature sets with some possibly different degrees. Fuzzy set is a good description for this phenomenon.

To fuzzify feature set $\mathbf{F}_j$, we need to define a membership function $\mathcal{M}_{\mathbf{F}_j} : \mathbf{F} \to [0, 1]$. For any $\vec{f} \in \mathbf{F}$, the value of $\mathcal{M}_{\mathbf{F}_j}(\vec{f})$ is then called the degree of membership of $\vec{f}$ to the fuzzy set $\mathbf{F}_j$ (or, in short, the degree of membership to $\mathbf{F}_j$). The most commonly used prototype membership functions are cone, trapezoidal, B-splines, exponential, Cauchy, and paired sigmoid functions. We have tested these functions on our system. In general, the performance of the exponential and the Cauchy functions is better than that of the rest functions. The exponential and Cauchy functions are comparable in performance. We pick the Cauchy function because it requires much less computations.

So, we define the *membership function for the feature set* $\mathbf{F}_j$, $\mathcal{M}_{\mathbf{F}_j} : \mathbb{R}^6 \to [0, 1]$, as

$$\mathcal{M}_{\mathbf{F}_j}(\vec{f}) = \frac{1}{1 + \left(\frac{\|\vec{f}-\hat{\vec{f}}_j\|}{\sigma_f}\right)^\alpha} \tag{1}$$

where

$$\sigma_f = \frac{2}{C(C-1)} \sum_{i=1}^{C-1} \sum_{k=i+1}^{C} \|\hat{\vec{f}}_i - \hat{\vec{f}}_k\|$$

is the average distance between cluster centers. Similarly, the *membership function for the shape feature* $\vec{h}_j$, $\mathcal{M}_{\vec{h}_j} : \mathbb{R}^3 \to [0, 1]$, is

$$\mathcal{M}_{\vec{h}_j}(\vec{h}) = \frac{1}{1 + \left(\frac{\|\vec{h}-\vec{h}_j\|}{\sigma_h}\right)^\alpha} \tag{2}$$

where

$$\sigma_h = \frac{2}{C(C-1)} \sum_{i=1}^{C-1} \sum_{k=i+1}^{C} \|\vec{h}_i - \vec{h}_k\|$$

is the average distance between shape features. The experiments show that the performance changes insignificantly when α is in the interval $[0.7, 1.5]$, but degrades rapidly outside the interval. So we set $\alpha = 1$ in both (1) and (2) to simplify the computation.

3. THE FIRM SIMILARITY MEASURE

Let $\mathbf{A}$ and $\mathbf{B}$ be fuzzy sets defined on $\mathbb{R}^k$ with corresponding membership functions $\mathcal{M}_\mathbf{A} : \mathbb{R}^k \to [0, 1]$ and $\mathcal{M}_\mathbf{B} : \mathbb{R}^k \to [0, 1]$, respectively. The intersection of $\mathbf{A}$ and $\mathbf{B}$, denoted by $\mathbf{A} \cap \mathbf{B}$, is a fuzzy set on $\mathbb{R}^k$ with membership function, $\mathcal{M}_{\mathbf{A}\cap\mathbf{B}} : \mathbb{R}^k \to [0, 1]$, defined as

$$\mathcal{M}_{\mathbf{A}\cap\mathbf{B}}(\vec{x}) = \min[\mathcal{M}_\mathbf{A}(\vec{x}), \mathcal{M}_\mathbf{B}(\vec{x})]. \tag{3}$$

The union of $\mathbf{A}$ and $\mathbf{B}$, denoted by $\mathbf{A} \cup \mathbf{B}$, is a fuzzy set on $\mathbb{R}^k$ with membership function, $\mathcal{M}_{\mathbf{A}\cup\mathbf{B}} : \mathbb{R}^k \to [0, 1]$, defined as

$$\mathcal{M}_{\mathbf{A}\cup\mathbf{B}}(\vec{x}) = \max[\mathcal{M}_\mathbf{A}(\vec{x}), \mathcal{M}_\mathbf{B}(\vec{x})].$$

The similarity between $\mathbf{A}$ and $\mathbf{B}$, $\mathcal{S}(\mathbf{A}, \mathbf{B})$, is given by

$$\mathcal{S}(\mathbf{A}, \mathbf{B}) = \sup_{\vec{x}:\vec{x}\in\mathbb{R}^k} \mathcal{M}_{\mathbf{A}\cap\mathbf{B}}(\vec{x}). \tag{4}$$

For the fuzzy sets defined by Cauchy functions, calculating similarity according to (4) is relatively simple. This is because Cauchy function is unimodal, and thus the maximum of (3) can only occur on the line segments connecting the center points of two functions. It is not hard to show that for fuzzy sets $\mathbf{A}$ and $\mathbf{B}$ on $\mathbb{R}^k$ with Cauchy membership functions $\mathcal{M}_\mathbf{A}(\vec{x}) = \frac{1}{1+\left(\frac{\|\vec{x}-\vec{u}\|}{\sigma_a}\right)^\alpha}$ and $\mathcal{M}_\mathbf{B}(\vec{x}) = \frac{1}{1+\left(\frac{\|\vec{x}-\vec{v}\|}{\sigma_b}\right)^\alpha}$, the similarity between $\mathbf{A}$ and $\mathbf{B}$ is

$$\mathcal{S}(\mathbf{A}, \mathbf{B}) = \frac{(\sigma_a + \sigma_b)^\alpha}{(\sigma_a + \sigma_b)^\alpha + \|\vec{u} - \vec{v}\|^\alpha}. \tag{5}$$

Let $\mathcal{F}_q = \{\mathbf{F}_j^q : 1 \leq j \leq C_q, j \in \mathbb{N}\}$ denote the collection of fuzzy sets for a query image segmented into C_q regions, and $\mathcal{F}_t = \{\mathbf{F}_j^t : 1 \leq j \leq C_t, j \in \mathbb{N}\}$ be the collection of fuzzy sets for a target image with C_t regions. First, for every $\mathbf{F}_j^q \in \mathcal{F}_q$, we define the similarity between it and $\mathcal{F}_t$ as

$$l_j^{(q,t)} = \mathcal{S}(\mathbf{F}_j^q, \bigcup_{i=1}^{C_t} \mathbf{F}_i^t) = \max_{i=1,\cdots,C_t} \mathcal{S}(\mathbf{F}_j^q, \mathbf{F}_i^t). \tag{6}$$

Combining $l_j^{(q,t)}$'s together, we get a vector

$$\vec{l}^{(q,t)} = [l_1^{(q,t)}, l_2^{(q,t)}, \cdots, l_{C_q}^{(q,t)}]^T.$$

Similarly, the similarity between any $\mathbf{F}_j^t \in \mathcal{F}_t$ and $\mathcal{F}_q$ is

$$l_j^{(t,q)} = \mathcal{S}(\mathbf{F}_j^t, \bigcup_{i=1}^{C_q} \mathbf{F}_i^q) = \max_{i=1,\cdots,C_q} \mathcal{S}(\mathbf{F}_j^t, \mathbf{F}_i^q), \tag{7}$$

and

$$\vec{l}^{(t,q)} = [l_1^{(t,q)}, l_2^{(t,q)}, \cdots, l_{C_t}^{(t,q)}]^T.$$

Finally, we define the similarity vector between $\mathcal{F}_q$ and $\mathcal{F}_t$, denoted by $\vec{L}_{(\mathcal{F}_q,\mathcal{F}_t)}$, as

$$\vec{L}_{(\mathcal{F}_q,\mathcal{F}_t)} = \left[\begin{array}{c} \vec{l}^{(q,t)} \\ \vec{l}^{(t,q)} \end{array} \right].$$

The query image is represented as $\mathcal{F}_q$ and $\mathcal{H}_q$. The target image is represented as $\mathcal{F}_t$ and $\mathcal{H}_t$. $\mathcal{F}_q$ and $\mathcal{F}_t$ are the classes of fuzzy sets which are defined according to (1). $\mathcal{H}_q$ and $\mathcal{H}_t$ are collections of fuzzy sets whose membership functions are consistent with (2). The query image is classified as either a textured or a non-textured image [1]. The similarity measure for two images, $m_{(q,t)}$, is

$$m_{(q,t)} = \frac{1-\delta}{2} \vec{L}^T_{(\mathcal{F}_q,\mathcal{F}_t)}[(1-\lambda)\vec{w} + \lambda\vec{\mu}] + \frac{\delta}{2}\vec{L}^T_{(\mathcal{H}_q,\mathcal{H}_t)}\vec{w}.$$

Here $\vec{w}$ is a vector containing the area percentages of the query and target images, $\vec{\mu}$ contains normalized weights which favor regions near the image boundary. The summation of all entries of $\vec{w}$ or $\vec{\mu}$ equals 2. In our system, $\lambda = 0.1$, if the query image is textured then $\delta = 0$, otherwise $\delta = 0.1$.

4. EXPERIMENTS

The FIRM approach is tested on a general-purpose image database (from COREL) including about 200,000 pictures. For each image, the features, locations, and areas of all its regions are stored. To provide objective comparisons based on numerical results, the system performance is evaluated based on a subset of the COREL database, formed by 10 image categories, each containing 100 pictures. Within this database, it is known whether any two images are of the same category. In particular, a retrieved image is considered a match if and only if it is in the same category as the query. This assumption is reasonable since the 10 categories were chosen so that each depicts a distinct semantic topic.

Every image in the sub-database was tested as a query, and the retrieval ranks of all the rest images were recorded. Three statistics were computed for each query: the precision within the first 100 retrieval images, the mean rank of all the matched images, and the standard deviation of the ranks of matched images. We use entropy to characterize the segmentation-related uncertainties in an image. For an image with C segmented regions, its entropy is defined as

$$E_{\{\mathbf{R}_1,\cdots,\mathbf{R}_C\}} = -\sum_{j=1}^{C} P(\mathbf{R}_j)\log[P(\mathbf{R}_j)],$$

where $P(\mathbf{R}_j)$ is the percentage of the image covered by region $\mathbf{R}_j$. The larger the value of entropy, the higher the uncertainty level. We compare the performance of the FIRM and IRM [2] approaches with respect to the coarseness of segmentation. The entropy is used to measure the segmentation-related uncertainty levels. For different average number of regions, C, the performance in terms of average precision p, average mean rank r, and average standard deviation σ are evaluated for both approaches. The results are given in Figure 1. As we can see, the average entropy E increases when images are, on average, segmented into more regions. In other words, the uncertainty level increases when segmentation becomes finer. At all uncertainty levels, the FIRM scheme performs better than the IRM method in all three statistics. In addition, there is a significant increase in p and a decrease in r for the FIRM scheme as C increases. While for the IRM method, p and r almost remain unchanged for all values of C. This can be explained as follows. When segmentation becomes finer, although the uncertainty level increases, more details (or information) about the original image are also preserved. Compared with the IRM method, the FIRM scheme is more robust to segmentation-related uncertainties and thus benefits more from the increasing of the average amount of information per image.

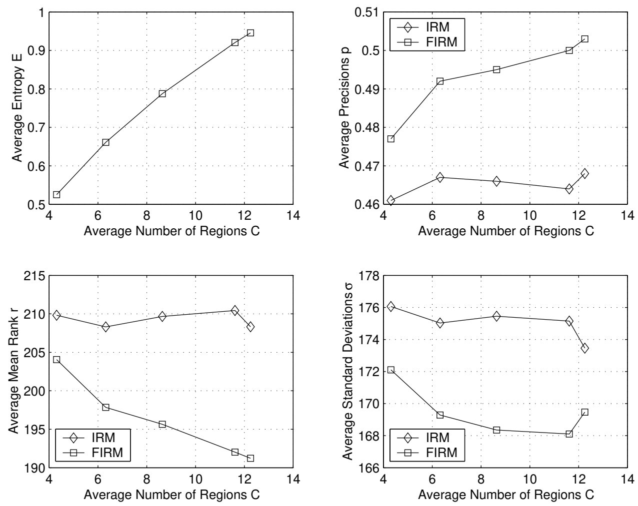

Figure 1: Comparing FIRM and IRM methods on the robustness to image segmentation.

5. CONCLUSIONS

We have developed FIRM for region-based image retrieval. To reduce the adverse impact of imprecise image segmentation, the FIRM uses fuzzy features to represent regions in an image. This naturally depicts the gradual transition of region boundaries, incorporates more information about regions than conventional region representation does, and describes the uncertainties inherent to imprecise image segmentation. The similarity measure of two images is defined as the overall similarity between two classes of fuzzy features. Experiments show that FIRM is more robust than IRM to segmentation-related uncertainties.

6. REFERENCES

[1] J. Li, J. Z. Wang, and G. Wiederhold. Classification of textured and non-textured images using region segmentation. In *Proc. 7th Int. Conf. on Image Processing*, pages 754–757. Vancouver, BC, Canada, September 2000.

[2] J. Li, J. Z. Wang, and G. Wiederhold. IRM: Integrated region matching for image retrieval. In *Proc. 8th ACM Int. Conf. on Multimedia*, pages 147–156. Los Angeles, California, USA, October 2000.

[3] M. Unser. Texture classification and Chansegmentation using wavelet frames. *IEEE Trans. Image Processing*, 4(11):1549–1560, November 1995.