Region-based Image Clustering and Retrieval using Fuzzy Similarity and Relevance Feedback

Reza Fakouri[1], Behzad Zamani[1,2], Mahmood Fathy[1], Behrouz Minaei[1]
[1]Iran University of Science and Technology
fakouri@comp.iust.ac.ir, {mahfathi, bminaei}@iust.ac.ir
[2]Shahrekord Islamic Azad University
bzamani@iust.ac.ir

Abstract

This paper proposes an interactive approach for region-based image clustering and retrieval. By performing clustering before image retrieval, the search space can be reduced to those clusters that are close to the query target. First, the image is segmented to regions by using an unsupervised segmentation method. This is an area where a vast number of regions are involved. To reduce search space for region-based image retrieval, we use clustering based on genetic algorithm. Fuzzy similarity is used in order to compute the similarity of two images. Moreover, a two-class SVM is trained based on user interests to improve image retrieval. Experiments were performed on COREL image database and show the effectiveness of the proposed approach.

1. Introduction

Content-based image retrieval, a technique which uses visual contents to search images from large scale image databases according to users' interests, has been an active and fast advancing research area since the 1990s. During the past decade, impressive progress has been made in both theoretical research and system development. However, there remain many challenging research problems that continue to attract researchers from multiple disciplines.

With the rapid increase of digital image data, image retrieval has drawn attention of researchers in computer vision and database communities. However, the current state-of-art technologies are facing two main problems: (1) “semantic gap” between low level features and high level concept; (2) high dimensionality.

For “semantic gap” problem, Relevance Feedback (RF) technique is widely used to incorporate the user's concept with the learning process [1, 2]. For reducing dimensionality, firstly we preprocess image regions by dividing them into clusters In this way, the search space can be reduced to a few clusters that are relevant to the query region. We use genetic algorithm [3] because of its robustness and ability to be approximate global optimum.

The rest of the paper is organized as follow, in section 2 the image segmentation method and feature extraction from regions are described. In section 3, GA clustering of images regions is presented. Section 4 describes the manner of computation the fuzzy similarity of two images. RF using two-class SVM is represented in section 5. Finally, in section 6 experiments and results are presented.

2. Image Segmentation

For the purpose of image segmentation, the first images should be partitioned into blocks and in lieu of each block a feature vector would be extracted. K-means algorithm classifies the feature vector in some classes in order to each class represents an image region. Size of each block would be selected according to the two effective factors which are texture and computation time. Small size of block causes the preservation details of textures; on the contrary it would increase the time of computation. Mutually, the increase in the block size would decrease the time of calculation, but it would deduct the texture data [4]. In this system, it is uses 4×4 pixel blocks.

Size of Images in the data bases are 256×384 or 384×256. So, each image has 6144 feature vectors. Each feature vector of $\vec{f_i}$ has six features ($\vec{f_i} \in \mathbb{R}^6$). In each feature vector, three features are computed by an averaging on color components in 4×4 blocks. Here LUV color space has been used, L stands for luminance, and U and V are for chrominance. The other three features represent energy in the high

978-0-7695-3504-3/08 $25.00 © 2008 IEEE
DOI 10.1109/ICCEE.2008.57

frequency bands of the wavelet transforms, which is the square root of the second order moment of wavelet coefficients in high frequency bands. Using a one-level wavelet transform, a 4x4 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 [5]. The intuition behind this is that coefficients in different frequency bands show variations in different directions. For instance, the HL band shows activities in the horizontal direction. An image with vertical strips thus has high energy in the HL band and low energy in the LH band.

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. For an image with the set of feature vectors $\mathbf{F}=\left\{\vec{f}_i \in \mathbb{R}^6 : 1 \le i \le 6144\right\}$, **F** is partitions into the C group $\{F_1,\ldots..F_C\}$ that as result, the image is segmented into the C regions $\{R_1,\ldots,R_C\}$.. Clustering is performed in the feature space; blocks in each cluster do not necessarily form a connected region in the images. This way, we preserve the natural clustering of objects in textured images and allow classification of textured images. K-means algorithm would not distinguish the amount of categories. For this reason we add number of categories with the gradual increase of C until the stop condition gratifies. Average number of categories in data bases for all images would be changed according to the stop condition regulation.

To describe shape properties, three other features are calculated for each region. They are normalized inertia of order 1 to 3. For a region R_j in the image plane, the normalized inertia of order γ is given as

$$I_{(R_j,\gamma)} = \frac{\sum_{(x,y):(x,y)\in R_j}\left[(x-\hat{x})^2+(y-\hat{y})^2\right]^{\frac{\gamma}{2}}}{V(R_j)^{1+\frac{\gamma}{2}}}$$

where $(\hat{x},\hat{y})$ is the centroid of R_j and $V(R_j)$ is the volume of R_j. The normalized inertia is invariant to scaling and rotation. The minimum normalized inertia is achieved by spheres. The γth order normalized inertia of spheres denotes as I_γ. We define shape feature $\vec{h}_j$ of region R_j as $I_{(R_j,\gamma)}$ normalized by I_γ,

$$\vec{h}_j = \left[\frac{I_{(R_j,1)}}{I_1}, \frac{I_{(R_j,2)}}{I_2}, \frac{I_{(R_j,3)}}{I_3}\right]^T .$$

3. GA Clustering of Image Regions

Genetic algorithms are methods that according to the randomly space search try to find the best existing answer of problems. But these random searches go toward the best answers. So these algorithms can not be known as totally random processes. According to these specifications, genetic algorithms have high capability in searching in complex spaces [3]. Genetic algorithms start with an initial random population of individuals which are the solutions of the problem. Certainly first these solutions should be coded. Each solution or individual is represented by a chromosome and so, the population is a set of these chromosomes. From the first generation, these chromosomes are first evaluated. Then they are operated by three genetic operators: *Selection*, *Crossover* and *Mutation* and generate the next generation. The next generation of chromosomes is again evaluated. This process continues until termination condition has been met. The overview of genetic algorithm is shown in Figure 1.

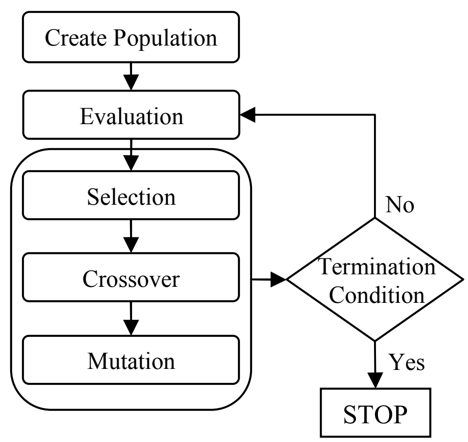

Figure 1. Genetic algorithm flowchart

For clustering the image regions with the use of genetic algorithm, at first we should introduce a chromosome structure. We use a unique number for each region (integers between 1 to n). Each chromosome has a constant length of gene (k) and each gene shows the center of each cluster. The above-mentioned representative regions are actually centroids of clusters. Figure 2 represents a sample of chromosome.

109	234	8	873	4	10	5987

Figure 2. Example of a chromosome

The aim of images clustering is categorization of images into clusters so that images in each cluster have the most similarity to each other. To do so, the following function should be optimized.

$$F(C)=\sum_{i=1}^{n}\min_{j=1}^{k}(d(R_i,C_j)) \qquad (1)$$

which R_i is the image region, C_j is the center of the cluster, that is one of the regions in the cluster and it is not a virtual region, n is the total number of image regions, k is number of clusters and d is a distance measure.

At the beginning we initialize k genes of each chromosome randomly and none repeated integers between 1 to n. Then we calculate the inverse values of the objective function for these chromosomes as fitness: $f_1, f_2, ..., f_l$. The fitness of each individual chromosome is computed as follows:

$$Fitness_i = f_i \Big/ \sum_{i=1}^{l} f_i \qquad (2)$$

l is the size of population. With the first generation, "evolution" begins. In each generation, the whole population goes through three operators: Selection, Crossover and Mutation.

- *Selection*: The selection mechanism is 2-fold tournament selection. From two randomly chosen chromosomes, the fitter is chosen to be one parent. This process is repeated with the two new competitors chosen from the entire population to find the second parent. The two chromosomes selected are then used in the crossover operator. All parental pairs for crossover are selected in this way.
- *Crossover*: We used a simple crossover operator presented in [6]. This operator would generate child chromosome C_0 from parent chromosomes C_1 and C_2. In each iteration, one of C_0 genes would be designating that can be C_1 or C_2 or both randomly. This action would be repeated k times in order to achieve C_0.
- *Mutation*: In order to increase the population diversity, we use a mutation operator. For this reason we select one of the genes and give it amount of 1 to n which this amount should not be repeated.

At the end of the process, the population member with the highest fitness value through all populations is selected as a feasible solution. This member is then decoded to obtain the centroids of clusters as the final output.

4. Fuzzy Similarity of Images

Image partitioning can be shown as the set of regions $\{R_1,\ldots.R_c\}$. Similarly, in feature space, an image can be distinguished with set of features $\{F_1,\ldots.F_c\}$ that each of them is a partition of **F**. For description of region R_j, we can use the related set of F_j and the similarity between two images which can be computed according to the F_j, [7]. Each region can be defined as

$$\hat{\vec{f}}_j = \frac{\sum_{\vec{f}\in \mathbf{F}_j}\vec{f}}{V(\mathbf{F}_j)}$$

that is the center of all F_j and can be none of the members of F_j.

To fuzzify the feature set F_j, we need to define a membership function: $\mu_{\tilde{\mathbf{F}}}:\mathbb{R}^6 \rightarrow [0,1]$. For any $\vec{f}\in \mathrm{F}$, the value of $\mu_{\tilde{\mathbf{F}}}(\vec{f})$ is then called the degree of membership of $\vec{f}$ to the fuzzy set F_j. choosing a proper membership function depends on the problem. We use the Cauchy function due to its high calculative efficiency and good meaningfulness. Therefore, the membership function for the feature set F_j is defined as

$$\mu_{\tilde{\mathbf{F}}_j}(\vec{f}) = \frac{1}{1+\left(\frac{\|\vec{f}-\hat{\vec{f}}_j\|}{d_f}\right)^{\alpha}}$$

that

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

is the average distance between cluster centers.

For finding out the similarity of two images, first we should have fuzzy similarity measure. The fuzzy similarity measure for fuzzy sets $\tilde{A}$ and $\tilde{B}$, $S(\tilde{A},\tilde{B})$, is defining as follows [7]:

$$S(\tilde{A},\tilde{B}) = \sup_{\vec{x}\in\mathbb{R}^k}\mu_{\tilde{A}\cap\tilde{B}}(\vec{x}) \qquad (3)$$

It is obvious that $S(\tilde{A},\tilde{B})$ is in the interval [0, 1] and a high value shows more similarity between $\tilde{A}$ and $\tilde{B}$. It is not hard to show that for fuzzy sets $\tilde{A}$ and $\tilde{B}$ with Cauchy membership functions

$$\mu_{\tilde{\mathbf{A}}}(\vec{x}) = \frac{1}{1+\left(\frac{\|\vec{x}-\vec{u}\|}{d_a}\right)^{\alpha}}$$

and

$$\mu_{\tilde{B}}(\vec{x}) = \frac{1}{1+\left(\frac{\|\vec{x}-\vec{v}\|}{d_b}\right)^{\alpha}},$$

the similarity between $\tilde{A}$ and $\tilde{B}$ is

$$S(\tilde{A},\tilde{B}) = \frac{(d_a+d_b)^{\alpha}}{(d_a+d_b)^{\alpha}+\|\vec{u}-\vec{v}\|^{\alpha}}.$$

Let $\mathcal{A} = \{\tilde{A}_i : 1 \le i \le C_a, i \in \mathbb{N}\}$ and $\mathcal{B} = \{\tilde{B}_j : 1 \le j \le C_b, j \in \mathbb{N}\}$ denote two collections of fuzzy sets. First, for every $\tilde{A}_i \in \mathcal{A}$, the similarity measure for it and $\mathcal{B}$ define as

$$l_i^{\mathcal{B}} = S(\tilde{A}_i, \bigcup_{j=1}^{C_b} \tilde{B}_j)$$

combining $l_i^{\mathcal{B}}$'s together, we get a vector

$$\vec{l}^{\,\mathcal{B}} = [l_1^{\mathcal{B}}, l_2^{\mathcal{B}}, \ldots, l_{C_a}^{\mathcal{B}}]^T .$$

Similarly for every $\tilde{B}_i \in \mathcal{B}$, the similarity measure between it and $\mathcal{A}$ define as

$$l_j^{\mathcal{A}} = S(\tilde{B}_j, \bigcup_{i=1}^{C_a} \tilde{A}_i)$$

combining $l_j^{\mathcal{A}}$'s together, we get a vector

$$\vec{l}^{\,\mathcal{A}} = [l_1^{\mathcal{A}}, l_2^{\mathcal{A}}, \ldots, l_{C_b}^{\mathcal{A}}]^T .$$

Finally, the similarity vector between $\mathcal{A}$ and $\mathcal{B}$, denoted by $\vec{\mathrm{L}}^{(\mathcal{A},\mathcal{B})}$, define as

$$\vec{\mathrm{L}}^{(\mathcal{A},\mathcal{B})} = \begin{bmatrix} \vec{l}^{\,\mathcal{B}} \\ \vec{l}^{\,\mathcal{A}} \end{bmatrix}.$$

Let $(\mathcal{F}_q, \mathcal{H}_q)$ and $(\mathcal{F}_t, \mathcal{H}_t)$ be fuzzy feature representations, for query image (q) and target image (t), respectively. Similarity between question image and target image would achieve through vectors of $(\vec{L}^{(\mathcal{F}_q,\mathcal{F}_t)}, \vec{L}^{(\mathcal{H}_q,\mathcal{H}_t)})$ that $\vec{L}^{(\mathcal{F}_q,\mathcal{F}_t)}$ is the similarity between color and texture and $\vec{L}^{(\mathcal{H}_q,\mathcal{H}_t)}$ shows the regions similarity. The similarity measure for two images, $m_{(q,t)}$, is

$$m_{(q,t)} = [(1-\lambda)\vec{w}_a + \lambda \vec{w}_b]^T \vec{L}^{(\mathcal{F}_q,\mathcal{F}_t)} + \vec{w}_a^T \vec{L}^{(\mathcal{H}_q,\mathcal{H}_t)} \quad (4)$$

Here $\vec{w}_a$ is a vector containing the normalized area percentages of the query and target images and $\vec{w}_b$ contains normalized weights which favor regions near the image boundary.

5. Relevance Feedback

Relevance Feedback (RF) is an interesting procedure to improve the performance of Content-Based Image Retrieval (CBIR) systems even when using low-level features alone. Here for RF, we use two-class Support Vector Machines (SVMs) classifier in which we interpret CBIR as a two class classification problem [8]; these classes are the relevant (positive) and the irrelevant (negative) images. There are several reasons for selecting SVMs:

- SVMs are very flexible. For example, prior knowledge regarding the problem can be used to tune the kernel.
- SVMs allow fast learning (with the rather limited number of examples provided by feedback) and relatively fast evaluation for medium-sized databases.
- By relying only on support vectors, SVMs are usually less sensitive than density-based learners to the imbalance between positive and negative examples in the training data.

Initially, classifier is trained using returned images labeled by the user. Two-class SVM solves a classification problem by finding a maximum margin hyperplane that separates the positive training instances from the negative ones.

6. Experiments

The system is tested on a general-purpose image database named COREL including about 60'000 pictures from 600 categories, which are stored in JPEG format with size 384x256 or 256x384. This database contains a textured and non-textured photograph. We removed 58 categories which represent texture photograph. Totally, in our system, there are about 54200 pictures from 542 categories. After segmentation, there are in total 289158 image regions and average number of regions per image for all images in the sub-database is 5.33. In order to obtain the number of clusters in GA clustering, we tested the system performance under different clustering schemes by dividing the entire set of image regions into 450 to 600 clusters; this range is chosen according to the number of categories. Each time we increase the number of clusters by 10 and then find that the number of clusters is *k=500*.

Now, the query image segments to regions. Consequently, for each region, we choose the five closest clusters as the reduced search space. All the images that have at least one region into these five clusters are identified. Therefore, fuzzy similarity of these images and the query image is computed according to (4) and m top images are retrieved. To evaluate the GA clustering phase, five hundreds images are randomly chosen from 100 categories as the query images. According to our experiment, the search space, in terms of the number of images in the five candidate clusters, on average is reduced to 26.82% of the original search space. Hence, speed of our system is four-fold more than the speed of a system without clustering similar to [7].

To qualitatively evaluate the performance of the system over the 54200-images COREL database, we randomly pick twenty image categories with different semantics, namely*, barnyard, beach, buildings, scene3, firearms, ...* . From each selected category, we randomly pick 5 images as query images. Totally, we used 100 query images. Five rounds of RF are performed for each query image: Initial (without feedback), First, Second, Third, and Fourth. The precision within the top 6, 12, 18, 24 and 30 retrieved images is calculated. It is important to mention that the accuracy increases after each iteration. Figure 3 present the accuracy rates of our algorithm.

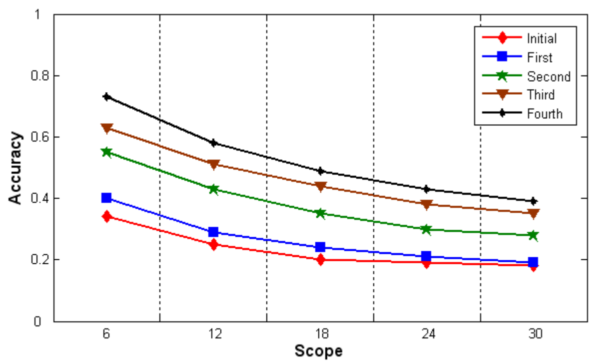

Figure 3: The accuracy within the top 6, 12, 18, 24 and 30 retrieved images after each feedback (initial without feedback).

7. Conclusion

In this paper, we proposed a region-based image retrieval approach using GA clustering based on Genetic Algorithm in order to reduce search space and Relevance Feedback strives to solve a semantic gap between low level features and human perception. Firstly, an image was segmented into regions. For each region, feature vector were extracted and stored in the database. Therefore, image regions clustered with GA clustering to reduce the search space. This phase is offline. After this phase, each region is then represented by a fuzzy feature that is determined by center location. The similarity measure of two images was defined as the overall similarity between two regions of fuzzy features. One of the most important advantages of our approach is its high-speed retrieval. Speed of our approach is four-fold more than the speed of a system without clustering. Furthermore, due to the generality of two-class SVM and robustness of GA in approximating global optima, the proposed system has proven to be effective in better identifying the user's real need.

References

[1] Y. Rui, T.S. Huang and S. Mehrotra, “Content-based image retrieval with relevance feedback in MARS,” *Proceedings of the international conf. on image processing*, Oct. 1997, vol 2(26-29), pp. 815–818.

[2] Zhong Su, Hongjiang Zhang, S. Li and Ma. Shaoping, “Relevance feedback in content-based image retrieval: bayesian framework, feature subspaces, and progressing learning,” *Trans. on Image Processing*, IEEE, vol. 12(8), Aug. 2003, pp. 924–937.

[3] John H. Holland, *Adaptation in natural and artificial systems*, University of Michigan Press, 1975.

[4] Jia Li, James Z. Wang and Gio Wiederhold, “IRM: Integrated Region Matching for Image Retrieval,” *Proceedings of Multimedia Conf.*, ACM, Oct. 2000, pp. 147-156.

[5] M. Unser, “Texture classification and Chansegmentation using wavelet frames,” *Trans. Image Processing*, IEEE, Nov. 1995, vol. 4, no. 11, pp. 1549-1560.

[6] V. Estivill-Castro, Alan T. Murray, “Spatial clustering for data mining with genetic algorithms,” *Technical report FIT-TR-97-10*, Queensland University of Technology, Faculty of Information Management, September

[7] Yixin Chen and James Z. Wang, “A Region-Based Fuzzy Feature Matching Approach to Content-Based Image Retrieval,” *Trans. on Pattern Analysis and Machine Intelligence*, IEEE, 2002, vol. 24, no. 9, pp. 1252-1267.

[8] P.S. Karthik and C.V. Jawahar, “Analysis of Relevance Feedback in Content Based Image Retrieval,” *9th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, IEEE, Dec. 2006, pp. 1-6.