

# An Image Clustering and Feedback-based Retrieval Framework

Chengcui Zhang, University of Alabama at Birmingham, AL, USA

Liping Zhou, University of Alabama at Birmingham, AL, USA

Wen Wan, University of Alabama at Birmingham, AL, USA

Jeffrey Birch, Virginia Polytechnic Institute and State University, VA, USA

Wei-Bang Chen, University of Alabama at Birmingham, AL, USA

---

## ABSTRACT

Most existing object-based image retrieval systems are based on single object matching, with its main limitation being that one individual image region (object) can hardly represent the user's retrieval target, especially when more than one object of interest is involved in the retrieval. Integrated Region Matching (IRM) has been used to improve the retrieval accuracy by evaluating the overall similarity between images and incorporating the properties of all the regions in the images. However, IRM does not take the user's preferred regions into account and has undesirable time complexity. In this article, we present a Feedback-based Image Clustering and Retrieval Framework (FIRM) using a novel image clustering algorithm and integrating it with Integrated Region Matching (IRM) and Relevance Feedback (RF). The performance of the system is evaluated on a large image database, demonstrating the effectiveness of our framework in catching users' retrieval interests in object-based image retrieval.

**Keywords:** Genetic Algorithm, Information Retrieval (IR), Multimedia Information Retrieval (Audio, Speech, Video, Image), Search Personalization, Semi-Structured Information Retrieval, User Modeling for Information Retrieval

---

## INTRODUCTION

Object-based image retrieval has recently become an important research issue in retrieving images on the basis of the underlying semantics of images. Within the context of object-based

image retrieval, the semantic content of an image is represented by one or more of the objects present in that image, where object-based features are used to describe the user's perceived content of that image. However, two critical issues exist in existing object-based image retrieval systems. First, most existing systems retrieve images according to a single

DOI: 10.4018/jmdem.2010111204

object/segment of a user's interest, which cannot meet the requirement of those user queries where more than one object-of-interest is involved. In addition, image segmentation, known as an extremely difficult process, may produce inaccurate segmentation resulting from over-segmentation and/or under-segmentation. Consequently, the inaccurate segments may negatively affect the image retrieval results. Secondly, most existing systems require a complex user interface which is capable of displaying all the segments (regions) of a query image so that users can choose desirable query region(s) from these segments. Such an interface can be cumbersome and confusing, mostly due to the inaccurate segmentation results by over- and under-segmentation, but also partly due to the fact that there are usually 7~8 segments/objects on average in each image.

Integrated Region Matching (IRM) (Carson, Thomas, & Belongie, 1999) has been proposed to alleviate the above two problems to some degree. IRM measures overall similarity between two images with the following two advantages. First, IRM effectively reduces the side effect of inaccurate segmentation by incorporating properties of all the regions in the images into one region matching scheme. Moreover, unlike other existing object-based image retrieval systems, IRM does not require a complex user interface to display all the segments/objects in the query image because IRM adopts an overall image-to-image similarity measure based on the similarity of two sets of objects/segments. Therefore, users only need to specify a query image without having to specify particular objects of interest through a complex user interface. However, several challenges remain in IRM, including: (1) how to efficiently index and search in a large-scale image segment/object database, and (2) how to bridge the "semantic gap" between low level object features and high level perceptions of image content consisting of a set of objects. As an unsupervised similarity measure, the original IRM ranks the retrieved images based on the overall similarity between two sets of image segments without any input of user guidance/

knowledge, where the significance of each image segment is fully determined by low level object features such as region/object size. The significance score is then used to calculate the pair-wise segment similarity score between a segment of the query image and a segment from an image in the database. The overall similarity score is the sum of pair-wise segment similarity scores between two images. However, this matching scheme failed to capture the user's preferences such as the relative importance of certain objects according to the user's subjective perception, and a selected few regions that form the profile of the user's search interest in a query image. In other words, the original IRM scheme does not reflect the relative importance of individual regions/objects in the query image according to the user's own preferences. This article aims to design an object-based image clustering and retrieval framework with feedback-based integrated region matching to address the challenges aforementioned.

In order to support integrated region-based image retrieval, we need to divide each image into several semantic regions. Instead of viewing each image as a whole, we examine integrated region similarity during image retrieval. However, this further increases the search space by a factor of 7~8 when compared with single-region based image retrieval which is already one magnitude more complex than non-object based image retrieval. Clustering is a process of grouping a set of physical or abstract objects into classes based on some similarity criteria. In this study, objects correspond to image regions. Given the huge amount of regions/segments in this problem, we first preprocess image regions by grouping them into clusters. In this way the search space can be reduced to a few clusters that are relevant to the regions/objects in the query image. K-means is a traditional clustering method and has been widely used in image clustering. However, it is incapable of finding non-convex clusters and tends to fall into local optimum especially when the number of data objects is large. In contrast, Genetic Algorithm is known for its robustness and ability to approximate global optimum. In

this article, we propose a new Genetic Algorithm based clustering algorithm for image region clustering which is more robust compared to the state-of-the-art.

To fill the semantic gap between low level region/object features (e.g., color, texture, and shape) and high-level user perceptions, relevance feedback is used as a powerful technique in the field of information retrieval and has been an active research area for the past decade. In this article, we proposed a Feedback-based Integrated Region Matching (FIRM) scheme to address this issue. With the proposed scheme, we consider each image as a set of regions, and the entire image database can be viewed as a large collection of regions. According to the region distance measured based on the similarity of region features, image regions are grouped into clusters. Each cluster represents a specific semantically meaningful concept. In the initial retrieval, the user provides a query image, and the system performs retrieval on the set of image clusters (reduced search space) that include the regions in the query image using the original IRM scheme. After the initial retrieval results are returned to the user, the user is asked to provide feedback (positive/relevant or negative/irrelevant) for each returned image among the top 20-40 returned images. To capture the user's high-level perception on the query image (search target), we integrate the user's relevance feedback with the integrated region-matching scheme, i.e., the so-named FIRM scheme. In brief, we assume that the regions in the query image are not equally important to the user query. Some regions capture more of the user's attention during retrieval than the others. This is modeled in our matching scheme through assigning different weights to individual regions. The region weights are then used in the calculation of the similarity between the query image and an image in the database. A learning component is designed and implemented to automatically calculate and adjust the weights of query regions according to the user's feedback. We take the set of region weights as our hypothesis of the user's search interest and continue with the

feedback-retrieval process using the proposed FIRM scheme. Through several iterations, the user's preferences are captured through gradual refinement of region weights.

The arguments for our framework start with a brief discussion on image clustering and retrieval framework. Next, an overview of our system is presented. Thirdly, the implementation of proposed framework is described, followed by the discussion of the processing, clustering, relevance feedback, and the retrieval components. Finally, the evaluation of system performance with experimental results is presented, followed by the summary and conclusion.

## RELATED WORKS

Several image clustering and retrieval frameworks have been proposed in the literature. Most of the object-based image retrieval systems are based on a single object matching. In order to support region-based image retrieval, systems need to divide each image into several semantic regions and examine region similarity during the image retrieval (Zhang & Chen, 2005; Babu & Nagesh, 2008). In our framework, instead of retrieving images based on a single object matching, we consider all the regions in an image as an entirety and examine the overall similarity between images based on the IRM measure (Wang, Li, & Wiederhold, 2001; Babu & Nagesh, 2008). One well-known system, the SIMPLiCity (Semantics-Sensitive Integrated Matching for Picture Libraries), adopts IRM to measure the overall similarity between images; however, no interactive learning mechanism is used in this retrieval scheme. The drawback is that the retrieval system has no clue about which regions/objects in the query image the user is interested in, and therefore, the retrieval system cannot adjust similarity measures to accommodate individual users' search interest. In order to alleviate this problem, in this study, we propose a Feedback-based Integrated Region Matching (FIRM) scheme by learning from the user's relevance feedback. Using the user feedback, we can refine the retrieval results

with a more sophisticated learning algorithm. The proposed scheme can be categorized as region-based image retrieval with interactive learning capabilities. Several other works in this category can be found in (Ji, Yao, & Liang, 2008; Bradshaw, 2000).

## Genetic Algorithm for Image Data

Genetic Algorithm (GA) (Haupt & Haupt, 2004) is a search technique in providing exact or approximate solutions to optimization and search problems. The concept of this type of algorithms is borrowed from the Darwin's theory – "survival of the fittest". The original GA which was developed by Holland in 1975 simulates an evolutionary process of a living species, using genetic operators such as "selection", "mutation", "crossover", and "reproduction". GAs have been broadly applied to a variety of fields, including ecology, biology, and statistics (Ding & Gasvoda, 2005; Tomassini, 1998). In GA, problems are solved by an evolutionary process, which results in a fittest (optimal) solution, i.e., the survivor. In our research, a Modified GA (MGA) is proposed and serves as a clustering algorithm for grouping image regions according to their object visual features.

In general, the objective of all clustering algorithms is to divide a set of data points into subgroups so that the objects within a subgroup are similar to each other whereas objects in different subgroups have diverse qualities. In this study, the main purpose of clustering is to reduce the search space by grouping a set of similar image regions into clusters based on some similarity criteria, and therefore reduce the time-complexity in the subsequent retrieval.

There are some related works in this research area. Ding and Gasvoda (2005) address the application of GAs for clustering image dataset. It is common that the running time for most GAs will dramatically increase when the number of data points in the input set or the number of clusters desired grows. Ding and Gasvoda (2005) try to solve this problem by reducing the input dataset of GA.

Another problem is that traditional GA tends to fall into local optima especially when clustering discretely represented data such as image data in this study. The CMA-ES (Covariance Matrix Adaptation Evolution Strategy) (Hansen, 2008) is an attractive option for non-linear optimization when general search methods failed due to a discontinuous search landscape or sharp bends. In this article, we introduce an innovative Genetic Algorithm, i.e., Modified Genetic Algorithm (MGA) to cluster image regions. The performance of the proposed MGA has been compared to that of CMA-ES and proved to be more robust when applied to real world problems where discrete data such as image data is involved.

## Relevance Feedback Methods

Relevance feedback is a powerful technique in content-based image retrieval (CBIR). In CBIR systems, there exists inherently "semantic gap" between high level concepts and low level features. Human perception of image similarity is always subjective and task-dependent, and the retrieval systems based on the similarities of pure visual features are not necessarily perceptually or semantically meaningful. Relevance Feedback (RF) is a supervised learning technique used to improve the effectiveness of information retrieval systems. It helps to establish the link between high level concepts and low level features and thus bridges the semantic gap.

Most of the relevance feedback methods can be classified into two categories: query point movement (Su, Zhang, & Ma, 2000) and query re-weighting. In the proposed method, the weights of image regions are updated automatically and therefore it falls into the second category. Several other existing works belong to this category. For example, Ji, Yao, and Liang (2008) proposed a Dynamic Region Matching (DRM) which adopts a probabilistic fuzzy region matching algorithm to retrieve and match images at the object level. In the matching algorithm, the weights of regions in the query image are updated through relevance feedback.

However, the problem with this algorithm is that for all matched regions in the query image, their weights are either increased or decreased by a fixed factor, making the performance gain obtained through RF rather limited.

Some other works also attempt to incorporate Support Vector Machines (SVM) (Su et al., 2000) into the learning process based on relevance feedback. However, SVM is not directly suitable for the relevance feedback because the training data set is too small to be representatives of the true distributions. In (Wang et al., 2001; Zhang & Chen, 2005), One-Class SVM is transformed to model the non-linear distribution of image regions and to separate positive regions from negative ones.

In this article, we propose a relevance feedback algorithm integrated into an innovative region-matching scheme. Each region of the query image is assigned an equal initial weight, and the weight is updated gradually based on its distances to the matched regions in positive images. Therefore, a smaller distance indicates a higher similarity between the matched pair of regions. A new integrated region matching scheme is proposed which calculates the overall similarity between two images as weighted region similarities. The region weights are

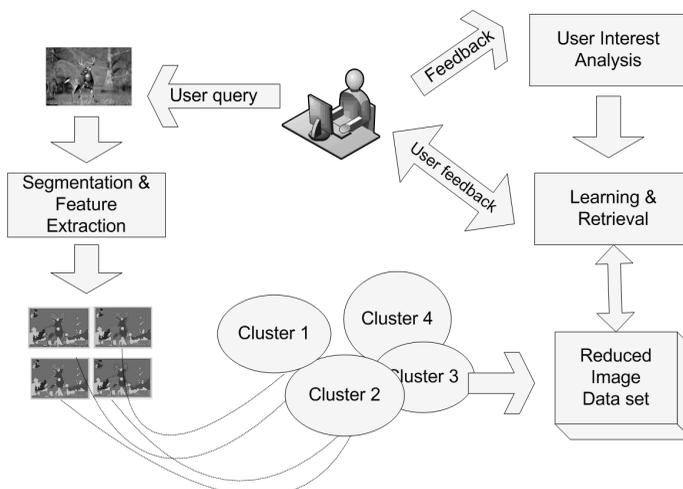
automatically and gradually refined through iterations of the feedback-retrieval process.

## FRAMEWORK OVERVIEW

The system architecture of the proposed framework is illustrated in Figure 1. The proposed framework consists of three major modules, including the preprocessing module, the clustering module, and the relevance feedback and retrieval module. In the preprocessing module, images are segmented into semantic regions. Then, object-level features are extracted for each image region. In our study, 8 features are used – 3 color features, 3 texture features, and 2 shape features. We use the proposed Genetic Algorithm as the clustering method to group image regions into clusters. In this way, an image can belong to more than one cluster if its containing regions are grouped to different clusters. At the time of retrieval, all the image clusters that involve the regions in the query image form the candidate image pool (reduced search space) for subsequent retrieval and user feedback.

In the initial query, since there is no user feedback yet available, the original IRM scheme is used to rank all the images in the candidate

Figure 1. The system architecture



pool. After the initial retrieval, the user can label all the positive images that he thinks relevant to the query image according to his own preferences (e.g., preferred query region(s)). In our current implementation, only the top 30 ranked images are returned to the user for feedback because a larger set may undermine the user experience for feedback-retrieval. Using the proposed matching scheme that integrates relevance feedback, the retrieval system is able to learn automatically the user's preferences from the user's feedback and predict the significance/weights of individual regions in the query image. The similarity between two images is thus evaluated over weighted integrated region similarities. The predicted weights are used in the next round of retrieval, followed by another round of user relevance feedback. It is expected that though multiple iterations, the region weights can be gradually refined so that the user's preferences in the case of region-based retrieval can be captured.

## PREPROCESSING

### Region Segmentation

In our framework, we employ a fast yet effective image segmentation method called WavSeg as proposed in our previous work (Zhang, Chen, Shyu, & Peeta, 2003) to partition images instead of manually dividing each image into a couple of regions (Yang & Lozano-Prez, 2000). In Wavseg, a wavelet analysis in concert with the SPCPE algorithm (Chen, Sista, Shyu, & Kashyap, 2000) is used to segment an image into regions.

By using wavelet transform and choosing proper wavelets (Daubechies wavelets), the high-frequency components will disappear in larger scale sub bands and therefore, the region areas will become more evident. In our implementation, images are pre-processed by Daubechies wavelet transform because it is proven to be suitable for image analysis. The decomposition level is 1. Then by grouping the salient points from each channel, an initial

coarse partition can be obtained and passed as the input to the SPCPE segmentation algorithm. In fact, even the coarse initial partition generated by wavelet transform approximates more closely to some global minima in SPCPE than a random initial partition. In other words, a better initial partition will lead to better segmentation results. Based on our experimental results, the wavelet based SPCPE segmentation framework (WavSeg) outperforms the random initial partition based SPCPE algorithm on average. It is worth pointing out that WavSeg is fast. The processing time for a 240×384 image is only about 0.33 s on average.

### Region Feature Extraction

Region feature extraction is a built-in component in the proposed framework, which is used to extract low-level, object-based visual features that can be used to describe image content. In this article, we extract eight visual features as adopted in Blobworld (Carson, Belongie, & Greenspan, 2002) for each region. The eight visual features include 3 color features, 3 texture features, and 2 shape features for each extracted image region.

## CLUSTERING OF IMAGE REGIONS

In this section, we present an innovative image region clustering algorithm named Modified Genetic Algorithm (MGA). Clustering is especially important when performing image retrieval on a large-scale dataset since the increase of dataset size significantly degrades the searching efficiency. After clustering, the original image dataset is reduced to a few image clusters related to the query image so that the search scope can be narrowed down. Here, we adopt MGA to approximate the global optimal solution, where a solution, within the context of image region clustering, consists of a set of cluster centroids in a non-convex and discontinued search space. In the remaining sections of this article, our experimental results show that

the proposed clustering algorithm alleviates a problem in traditional GAs where they could easily fall into a local optimum.

## Overview of Genetic Algorithms

Genetic Algorithms (Ortiz, Simpson, Pignatiello, & Heredia-Langner, 2004) are iterative optimization procedures that repeatedly apply GA operations (such as selection, crossover, and mutation) to a group of solutions until some criteria of convergence are satisfied. In a GA, a search point, a setting in the search space, is coded into a string which is analogous to a chromosome in biological systems. The string (chromosome) is composed of characters which are analogous to genes. In a statistical application, the chromosome corresponds to a particular setting of  $k$  factors (or regressors), denoted by  $X=[x_1, x_2, \dots, x_k]$  in the design space and  $i^{\text{th}}$  gene in the chromosome corresponds to  $x_{ip}$ , the value of the  $i^{\text{th}}$  regressor. A set of multiple concurrent search points or a set of chromosomes (or individuals) is called a population. Each iterative step where a new population is obtained is called a generation.

In general, the procedure of a GA consists of the following steps:

1. Define an objective/fitness function and its variables. Configure GA operations (such as population size, parent/offspring ratio, selection method, number of crossovers and the mutation rate).
2. Randomly generate the initial population.
3. Evaluate each individual (or chromosome) in the initial population by the objective function.
4. Generate an offspring population by GA operations (such as selection/mating, crossover, and mutation).
5. Evaluate each individual in the offspring population by the objective function.
6. Decide which individuals to include in the next population. This step is referred to as "replacement" in that individuals from the current parent population are "replaced"

by a new population, whose individuals come from the offspring and/or parent population.

7. If a stopping criterion is satisfied, then the procedure stops. Otherwise, go to Step 4.

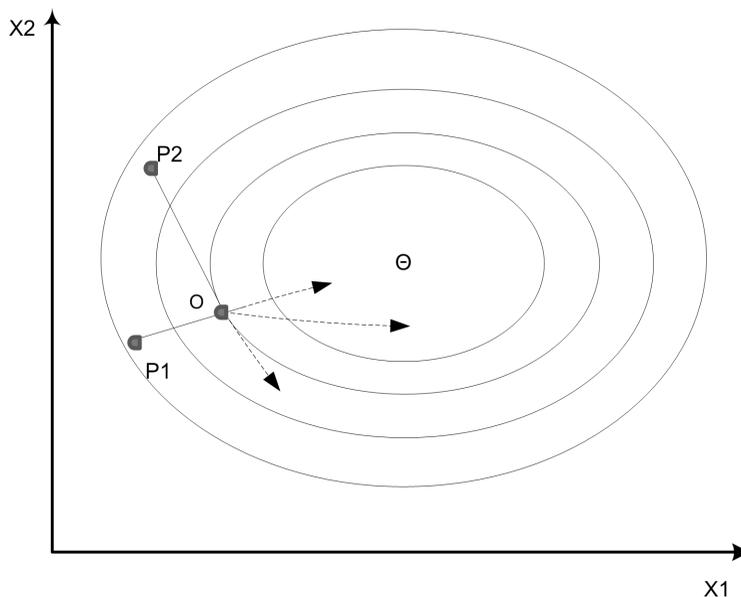
## The Proposed Modified Genetic Algorithm (MGA)

GAs are a large family of algorithms that have the same basic concept but differ from one another with respect to several strategies such as stopping rules and operations which control the search process. Based on previous experiences, in this study, the type of selection we utilize is random pairing. The blending crossover is utilized, and the number of crossover points depends on the number of dimensions of a specific objective function and is set to 2. Random uniform mutation is utilized and the mutation rate is set to 0.04. The type of replacement over both parent and offspring populations is ranking (Hamada, Martz, Reese, & Wilson, 2001; Myers & Montgomery, 2002).

The GA itself does not utilize any directional search explicitly. In order to improve the computational efficiency of the GA, we modify the GA by incorporating local search into the GA process, namely MGA (Wan, 2007). The method of Steepest Descent (SD) (Haupt & Haupt, 2004) and the Newton-Raphson method (NR) are two kinds of well-known local search methods, both of which require the partial derivatives of an objective function  $f$ . It is not expected that SD or NR can always find a proper direction from the current point since an objective function may not be simple or unimodal, but very complicated, locally rough and unsmoothed. Thus, we developed a new local directional search method which is derivative-free and denoted by "DFDS".

The local search approach has the same main idea: utilizing numerical information from a GA process to find some appropriate local directions by only requiring a few extra function evaluations so that the GA process may be guided to further possible improvement. The numerical information we utilize in our study

Figure 2. Three defined directions in a 2-dimension space



is focused on the best offspring among both the current parent and offspring populations.

When the best offspring among both the offspring and the parent populations is found, we can trace back to find its parents. These parents then can be considered as two different starting points. Both of them go to the same point: the current best offspring. Therefore, two directions are established: one is from the first parent to the current best offspring; the other is from the second parent to the same offspring. Both directions have obtained improvement, since the best offspring of interest is an improvement over both of its parents in terms of values of an objective function. We individually project the two directions to  $n$  axis for an  $n$ -dimension space and compare the components of parent directions in  $i^{\text{th}}$  axis, if both parent directions are consistent on  $i^{\text{th}}$  axis (either both positive or both negative), the third direction is the common direction; otherwise, the third direction is zero which means the searching point will not move on  $i^{\text{th}}$  axis for the third direction.

Figure 2 illustrates the three defined directions in a 2-dimension space. The original point

before performing our local search method is represented as “O”, its optimal point is denoted by “ $\Theta$ ”.  $P_1$  and  $P_2$  are its parent’s points from a GA process. It is easy to see the two parents directions  $P_{1O}$  and  $P_{2O}$ , expressed as  $\delta_{P_{1O}} = [\delta_{11}, \delta_{12}]$  and  $\delta_{P_{2O}} = [\delta_{21}, \delta_{22}]$ , respectively. It is obvious that in the third direction  $\delta_3 = [\delta_{31}, \delta_{32}]$ ,  $\delta_{31} > 0$  since both  $\delta_{11} > 0$  and  $\delta_{21} > 0$ . This indicates that the common direction in this case is positive along the  $x_1$  axis. In addition,  $\delta_{32} = 0$  since  $\delta_{12} > 0$  and  $\delta_{22} < 0$ , which indicates that the common direction has no relative movement along the  $x_2$  axis.

### MGA Application on Image Region Clustering

In our application, one image is defined as a set of image regions, and each image region is represented as an 8-dimension data point in the dataset. Therefore, the image region clustering problem is defined as the task of dividing an original image region dataset into a desired number of groups so that the Euclidean distance between each data point and its corresponding cluster centroid is minimized, which is a com-

monly used method for clustering images. The total distance of each point to its cluster centroid, known as the total distance measurement of the clustering, is calculated as in Eq. 1. We use the Euclidean distance to calculate the distance between two images as in Eq. 2.

$$F(C) = \sum_{j=1}^k \sum_{p_i \in C_j} D(p_i, C_j) \quad (1)$$

$$D(I, J) = \left( \sum_t |f_t(I) - f_t(J)|^2 \right)^{1/2} \quad (2)$$

In Eq. 1,  $C_j$  is the  $j^{\text{th}}$  cluster centroid of input dataset (the set of regions);  $p_i$  represents the  $i^{\text{th}}$  data point (image region) in cluster  $j$ ; each data point (region) is described as an 8-feature vector.  $C_j$  is an existing data point (image region) in the dataset chosen as the centroid for cluster  $j$ . Eq. 2 is used to calculate the distance  $D(I, J)$  between region  $I$  and region  $J$ , where  $f_t(I)$  is the  $t^{\text{th}}$  feature of region  $I$  ( $t=1 \dots 8$ ).

It is very common to have as many as 50,000~60,000 fitness evaluations during one single run of a GA. The impact that the fitness function has on the execution time of the GA is obvious. The complexity of the fitness function cannot be too high for any GA that is to be used to perform clustering on a very large dataset. In our research, fitness function is defined as the inverse value of the total distance which is calculated as shown in Eq. 3. Our MGA aims at approximating the maximum of the fitness function.

$$fit = \frac{1}{F(C)} \quad (3)$$

In order to apply the proposed MGA to the clustering of image regions, a chromosome encoding scheme is proposed. In the proposed encoding scheme, the cluster centroids are encoded as genes into the chromosome and the length of each chromosome is the number of cluster centroids and also the number of desired

clusters. In our test image database, there are 8,900 images which are further segmented into 82,556 regions. Therefore, the value of each gene in a chromosome is its index in the set of regions which ranges from 1 to 8,900, and the values of genes in the same chromosome must be unique. One limitation of many clustering algorithms is that they assume the number of clusters is known. However, in practice, the number of clusters may not be known in priori. This problem is called unsupervised clustering. In this study, we use an approach to achieve pseudo-unsupervised clustering which aims at determining an appropriate number of clusters  $|C|$ , without any prior knowledge about it. In particular, we proceed by repeating the clustering for several  $|C|$  values, and choosing the best partition. Since this is done off-line, the high computational cost can be afforded. For scalability, whenever there is a significant growth of the image database, new clusters can be discovered by comparing the distance between each newly added image region and the existing cluster centroids – if the distance value is too large for any existing cluster, the newly added region will form a new cluster of its own, without the need to completely re-cluster all the image regions in the database.

Using this approach, through several runs on the dataset, we finally decide to divide the whole set of image regions into 1,000 clusters since it results in a good balance between good fitness and efficiency. In other words, the original image dataset is represented with a dataset which contains 82,556 points (image regions) centered around 1,000 cluster centroids. Initially, the 1,000 cluster centroids are randomly selected. By several iterative optimization procedures of MGA, the solution chromosome is a set of encoded cluster centroids, which produces the minimum sum of the distances of each point to its cluster centroid (maximum fitness value) through those iterations. Also, the total distance is used to measure the quality of the clustering results.

Table 1. Results of comparison between MGA and CMA-ES

		<i>The total distance</i>
100 Groups	MGA	100,451.0
	CMA-ES	110,950.0
1000 Groups	MGA	67,872.1
	CMA-ES	69,926.4

## Comparison Between MGA and CMA-ES

In order to evaluate the performance of the proposed MGA, we compare it with Covariance Matrix Adaptation Evolution Strategy (CMA-ES) (Hansen, 2008) in solving the image region clustering problem. The quality of clustering is evaluated in terms of the total distance measure as mentioned in previous section, and a smaller total distance value indicates better clustering quality. The CMA-ES is considered as an attractive option for non-linear optimization in a discontinuous search landscape. In our application, the data points (image regions) are scattered in a multi-dimensional space and therefore are discrete. For an input dataset with 82,556 data points (regions), our goal is to cluster these points into 1,000 groups where the total distance of each point to its corresponding cluster centroid is minimized. Table 1 demonstrates the comparison results of MGA and CMA-ES. The comparison is based on 2,400 evaluations of the fitness function. The results presented in the next section are also based on the same number of evaluations. This particular number is chosen because both algorithms start to converge at around 2,000 evaluations for this particular dataset. We can see that MGA has a better performance than CMA-ES based on the total distance measure. We also tested on different numbers of evaluations of the fitness function and the same trend persists.

## RELEVANCE FEEDBACK AND RETRIEVAL

### The Proposed Region-matching Scheme

In our proposed retrieval framework, an innovative region-matching scheme using feedback-based IRM for measuring the similarity of two images is proposed. Compared with the original IRM scheme, our region-matching scheme improves the retrieval accuracy by incorporating the relevance feedback into the IRM scheme. Therefore, the overall similarity of two images is computed as a weighted sum of the region similarity between two images. An advantage of using an integrated region similarity measure is that it is robust against poor segmentation by incorporating the properties of all the regions in the images. In designing the integrated similarity measure, we first attempt to match regions in two images. The principle of matching is that the most similar region pair is matched first. The matching allows one region of an image to be matched to several regions of another image. The distance between two regions can be easily calculated as the Euclidean distance in terms of the features extracted. During the matching process, the similarities of region pairs are calculated, and proper weights are assigned to them. The overall integrated similarity score between two images is computed as a weighted sum of the similarity between region pairs. We define the overall distance of two images

using Eq. 4, where  $I_1$  is the query image and  $I_2$  represents an image in the database;  $i$  and  $j$  denotes the  $i^{\text{th}}$  region of  $I_1$  and the  $j^{\text{th}}$  region of  $I_2$ .  $s_{ij}$  represents the weight assigned to the region pair  $i$  and  $j$  from the two images.  $d_{ij}$  is the Euclidean distance between two regions  $i$  and  $j$ .  $w_i$  is the weight of region  $i$  in the query image ( $I_1$ ). The initial value of  $w_i$  is set to 1, and  $w_i$  is automatically updated through relevance feedback learning process.

$$D(I_1, I_2) = \sum_{i=1}^m w_i \sum_{j=1}^n s_{ij} d_{ij} \quad (4)$$

The matrix  $S$  is referred to as the significance matrix, in which  $s_{ij}$  indicates the significance of the matched region pair  $i$  and  $j$  for determining the similarity between images.  $D$  is referred to as the distance matrix in which  $d_{ij}$  represents the distance of the matched region pairs.

$$D_{1,2} = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{m1} & d_{m2} & \cdots & d_{mn} \end{bmatrix} \quad S_{1,2} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m1} & s_{m2} & \cdots & s_{mn} \end{bmatrix}$$

For each image in the candidate pool, a significance matrix  $S$  and a distance matrix  $D$  will be constructed in order to calculate the integrated region similarity between the query image and that candidate image. Considering that the size of the matrix will increase polynomially with the number of regions in an image, we choose to use the six biggest regions (or less if there are no more than six regions) in the query image for retrieval and ignore the other smaller regions. It is not uncommon that one image may have more than ten regions due to over-segmentation, which leads to significant increase of the time-complexity in calculating the overall distance. Another reason is that some regions are relatively less meaningful due to over-segmentation and cannot carry a concrete semantic concept. For example, some region resulting from poor segmentation covers only a small part of a concrete object and cannot be

used to represent one complete semantic object in that image. In this case, ignoring these smaller regions in the retrieval actually reduces noise in the dataset.

The number of regions ( $\leq 6$ ) in the query image determines the number of rows in  $S$  and  $D$  which is fixed for each  $S$  and  $D$ . For each candidate image, the maximum number of regions used in similarity comparison is also confined to six which determines the number of columns in its corresponding  $S$  and  $D$ . Since the number of regions in each candidate image may not be the same, the number of columns in its corresponding  $S$  and  $D$  varies accordingly. In the initial query, the distance between the query image and a candidate image is calculated using the original IRM scheme, meaning that the initial weights of regions in the query image are all set to 1 (see Eq. 4). The construction of  $S$  and  $D$  matrices is detailed in the next subsection. The top 30 ranked images are returned to the user for feedback. Through the learning via relevance feedback, the weights of regions in the query image are refined automatically and gradually.

### Learning Through Relevance Feedback

In the proposed relevance feedback framework, we collect the user's positive feedback as samples to construct the training dataset. Through feedback, an image labeled as positive indicates that it has regions that match the target regions in the query image according to the user's preferences. This information is used to update the weights of the regions in the query image. The weight of a region in the query image represents the significance of that region for calculating the integrated similarity between the two images, where the significance of a region in the query image can be reflected by the degree of matching between the region and the matched region in the candidate image. The degree of matching between a pair of regions can be determined by the distance between them. The less the distance is, the

better matched they are. Eq. 5 formulates the re-weighting scheme.

$$w_i = w_i' \times \prod_p \frac{1}{\min_k(d_{ik})} \quad (5)$$

where  $w_i'$  is the current weight of  $i^{\text{th}}$  region in the query image;  $d_{ik}$  is the distance between region  $i$  in the query image and region  $k$  in a positive image ( $p$ ) labeled by the user. Assume one query image  $I_1$  has  $m$  regions, represented by a region set  $R_1 = \{r_1^1, r_2^1, \dots, r_m^1\}$ , and a candidate image  $I_2$  has  $n$  regions represented by  $R_2 = \{r_1^2, r_2^2, \dots, r_n^2\}$ . We denote  $d_{ij}$  as the distance of  $r_i^1$  and  $r_j^2$ , and use  $s_{ij}$  to represent the significance value for the pair  $r_i^1$  and  $r_j^2$ . The significance of  $r_i^1$  and  $r_j^2$  in  $I_1$  and  $I_2$  are denoted as  $s_i^1$  and  $s_j^2$ , respectively. We initiate the significance value of a region as the area percentage of that region in a given image, assuming that important objects in an image tend to occupy larger areas. Therefore,  $s_i^1$  is actually the area percentage of region  $i$  in image  $I_1$ .  $s_{ij}$  can be derived from  $s_i^1$  and  $s_j^2$ , subject to the following constraint as given in Eq. 6.

$$\sum_{j=1}^n s_{ij} = s_i^1 \quad \sum_{i=1}^m s_{ij} = s_j^2 \quad (6)$$

To find the first pair of best matched regions, we locate the minimum distance value in  $D$  matrix and obtain the corresponding row and column indices  $i$  and  $j$ . The significance values of regions  $i$  and  $j$  are then updated with Eqs. 7 and 8 which make sure that the best matched region pair has the highest significance value. If  $s_i^1 < s_j^2$ ,  $s_i^1$  becomes 0, and region  $i$  in image  $I_1$  will be removed from the next round of 'finding the best matched pair'. This is because we have already found the best matched region for  $r_i^1$  in  $I_2$ . Similarly, if  $s_j^2 < s_i^1$ ,  $s_j^2$  becomes 0, and region  $j$  in  $I_2$  will be removed from the next round.

$$s_i^1 = s_i^1 - \min(s_i^1, s_j^2) \quad (7)$$

$$s_j^2 = s_j^2 - \min(s_i^1, s_j^2) \quad (8)$$

We detail the construction procedure for  $S$  and  $D$  matrices as follows:

1. Initialize a  $m \times n$  matrix  $S_{1,2}$  for images  $I_1$  (the query image) and  $I_2$  (a candidate image) with its components being set to zeros.
2. Calculate the distance matrix  $D_{1,2}$  for all region pairs in images  $I_1$  and  $I_2$ .
3. Choose the minimum  $d_{ij}$  and obtain the best matched region pair  $(r_i^1, r_j^2)$ .  $\min(s_i^1, s_j^2)$  is assigned to the corresponding element  $s_{ij}$  in the matrix  $S_{1,2}$ .
4. Update the significance values of regions  $s_i^1$  and  $s_j^2$  with Eq. 7 and Eq. 8. Eliminate those distance values from  $D_{1,2}$  matrix that are associated with the region that should be removed according to the discussion aforementioned.
5. If  $\sum_{i=1}^m s_i^1 \neq 0$  and  $\sum_{j=1}^n s_j^2 \neq 0$ , go to Step 3. Otherwise, go to Step 6, which indicates that we have finished the construction of the matrix  $S_{1,2}$ .
6. Calculate the integrated similarity score with Eq. 4, return the top 30 most similar images from the ranked list for user feedback.

The initial weights of regions in the query image  $I_1$  are all set to 1s, and therefore, the initial query results will be the same as that of the original IRM scheme. However, by incorporating users' feedback into the retrieval and integrated region matching, we can dynamically update the significance (weight) of each query region as follows:

For the query image  $I_1$  and its region set  $R_1 = \{r_1^1, r_2^1, \dots, r_m^1\}$ , a positive image  $I_{p1}$  with  $l$  regions is represented by a region set  $R_{p1} = \{r_1^{p1}, r_2^{p1}, \dots, r_l^{p1}\}$ .

1. Construct the distance matrix  $D_{1,p1}$  according to Eq. 2.

2. For each row of  $D_{1,p_1}$ , i.e., for each region in the query image, locate  $\min(d_{ik})_{k=1,\dots,n}$  and calculate the inverse of  $\min(d_{ik})_{k=1,\dots,n}$ . Repeat the above calculation for each positive image labeled by the user and compute the updated weight of region  $i$  with Eq. 5.
3. The updated weights of regions in the query image are recorded and will be used in the next round of retrieval for computing updated integrated similarity scores.
4. This feedback-retrieval process will run for several iterations until the user is satisfied with the returned results.

## PERFORMANCE EVALUATION

In this section, we evaluate the performance of the proposed system by applying the proposed method on a Corel image database consisting of 8,900 images from 100 categories. After segmentation, there are in total 82,556 image segments. The number of clusters is selected to be 1,000 according to the discussion in the previous section. For each region in the query image, the cluster that contains that region is located. All segments in that cluster share similar semantic meaning since the extracted object features from them are similar. A set of such clusters forms the reduced search space for the subsequent retrieval based on IRM. In our case, the size of each region cluster ranges from 14 to 294 (regions). As aforementioned, an image can belong to more than one cluster if its regions are grouped into different clusters. By utilizing the clustering results, we are able to significantly reduce the search space, and the size of which is between 50 and 1,700 (images). The proposed feedback-retrieval framework is based on the candidate images in the reduced image set. Compared with the full image database search, the use of MGA effectively reduces the search space to about 10% of the original search space (8,900 images) on average, and thus, can reduce the time complexity significantly in the subsequent retrieval process.

## Performance Evaluation Measures

In our experiments, we randomly select 50 images from 12 categories as query images. In addition, we adopt the Average Retrieval Rank (AVR), Average Normalized Modified Retrieval Rank (ANMRR) (Cieplinski, 2001), and Accuracy as the standard performance measures for this dataset. The AVR and ANMRR measures are defined in Eqs. 9-10, respectively.

$$AVR(q) = \frac{1}{NG(q)} \sum_{k=1}^{NG(q)} Rank^*(k) \quad (9)$$

In Eq. 9,  $q$  is the current query;  $NG(q)$  is the number of positive images in the top 30 returned images in our case;  $\sum_{k=1}^{NG(q)} Rank^*(k)$

is the retrieval rank capped by the  $Rank^{\max}$  which is defined as the upper limit of the retrieval rank (30 in our case). However, AVR mainly focuses on the quality of ranking in the retrieval and is not very indicative of the total number of positive images returned by the retrieval system. Therefore, we use another measure called ANMRR measure which is the averaged MPEG-7 Normalized Modified Retrieval Rank (NMRR) over the query set, as defined in Eq. 10.

$$NMRR(q) = \frac{AVR(q) - 0.5 * [1 + NG(q)]}{Rank^{\max} - 0.5 * [1 + NG(q)]} \quad (10)$$

In our retrieval system, the proposed system returns the top 30 images with the highest similarity scores as a short check-list to the user. The Accuracy is defined as the percentage of relevant images out of the returned short list, which is commonly and widely used as a criterion for performance measure in content-based image retrieval (Zhang & Chen, 2005). However, as a disadvantage, using Accuracy to measure retrieval performance cannot faithfully reflect the rank of the returned images.

In this article, the rank of a target image is defined as the ordinal position of a relevant/positive image in the retrieved image list. It is obviously that a good retrieval system should return all the relevant images at the top of the list, i.e., a lower value in rank. For example, the lowest (or best) rank is 1, and the highest rank is 30. A true positive image becomes a missed hit or false negative if it does not appear in the short list. Therefore, it is important to consider the rank of the retrieved relevant image since it directly reflects the retrieval performance. We adopt not only the Accuracy measure to demonstrate the effectiveness of the retrieval results, but also the AVR and ANMRR measures to evaluate the rank of the retrieved relevant image.

The AVR measure fairly ranks the retrieval results if the numbers of target images retrieved from different results are the same. However, problem occurs when comparing results with different number of relevant images returned. For instance, the AVR measure is 1 if only one relevant image is retrieved at the top 1 ordinal position. Similarly, the AVR measure is 1.5 if only two target images are retrieved at the top 2 ordinal positions. In the above example, the performance of the latter is better than that of the former; however, the AVR measure does not reflect this fact directly. Therefore, in this article, we evaluate the performance mainly using the ANMRR and Accuracy measures although we list all the three measures in the experimental results.

In our experimental setting, five rounds of relevance feedback are performed for each query image - Initial with no feedback, First, Second, Third, and Fourth. AVR, Accuracy, and ANMRR (Cieplinski, 2001) are individually calculated for the top 30 retrieved images.

### The Effectiveness of MGA in Reducing the Search Scope

In this experiment, we study the performance of the proposed FIRM scheme with and without the use of MGA in reducing the search scope. The motivation of this experiment is to show that without significantly sacrificing the performance of image retrieval, MGA can effectively narrow down the search scope, and therefore, reduce the time complexity. We present the experimental results in Table 2 which compares the performance in terms of AVR, ANMRR, and Accuracy measures.

In Table 2, FIRM indicates that MGA is used in the proposed scheme, while FULL indicates that MGA is not used in the proposed matching scheme, i.e., a full database search is performed. Through this comparison, the lower AVR and ANMRR values indicate better performance. On the contrary, higher accuracy indicates better retrieval results.

In general, full search (FULL) should have a better retrieval performance than that of FIRM which is performed on a much reduced search space, and this is evidenced by ANMRR and Accuracy measures in Table 2. The AVR, as mentioned earlier in the previous subsection,

Table 2. The effectiveness of MGA

#	AVR		ANMRR		Accuracy	
	FIRM	FULL	FIRM	FULL	FIRM	FULL
1	12.64	14.69	0.325	0.255	0.395	0.436
2	12.39	14.56	0.318	0.247	0.399	0.441
3	11.90	17.71	0.320	0.235	0.392	0.451
4	11.98	14.20	0.300	0.230	0.400	0.445
5	12.07	14.16	0.299	0.229	0.408	0.446

#: 1: Initial; 2: First; 3: Second; 4: Third; 5: Fourth

cannot faithfully reflect the retrieval performance when the numbers of target images retrieved are different. However, it is still worth noting that FIRM with MGA demonstrates a better performance than FULL in terms of AVR (see Table 2).

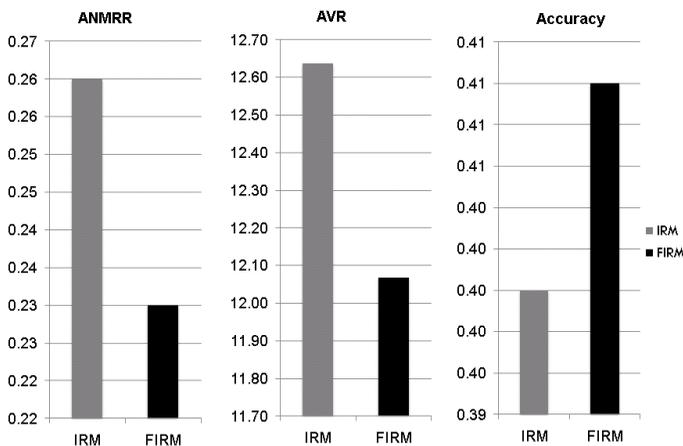
It is worth noting that as a trade off, full-scope search achieves better retrieval accuracy at the cost of significantly higher time complexity, while FIRM trades accuracy for efficiency, which is essential for a practical integrated region-based image retrieval system. In our current implementation, with full search, it takes about 4 minutes to perform one single iteration of query for one query image (there are 5 iterations involved in our experiments), while it only takes about 20 seconds to perform the same query with FIRM and MGA. This indicates that the full search becomes impractical when the size of the image database rapidly grows. Besides, we can observe from Table 2 that the performance of both FIRM and FULL are gradually improved through iterations, owing to the use of relevance feedback. It is also worth noting that at the end of the fourth round of retrieval, the performance of FIRM with MGA is very close to that of FULL.

### The Effectiveness of Relevance Feedback

We further compare the proposed framework (FIRM), i.e., feedback-based integrated region matching, with the traditional IRM (IRM) by using the above three evaluation criteria. We illustrate the comparison results in Figure 3 which shows the AVR, ANMRR, and Accuracy of tradition IRM and that of the proposed framework (FIRM) with four iterations of feedback and retrieval.

In Figure 3, lower AVR and ANMRR scores indicate better performance. On the contrary, higher accuracy indicates better retrieval performance. The AVR, ANMRR, and Accuracy values of the proposed framework (FIRM) are 12.07, 0.23, and 0.41, respectively, while those of the traditional IRM (IRM) are 12.64, 0.26, and 0.39, respectively. The experimental results indicate that the proposed framework (FIRM) outperforms the tradition IRM (IRM) scheme, which also show the effectiveness of integrating IRM with relevance feedback. It is worth pointing out that the retrieval is performed on a much reduced search space with at most 1,700 images in the candidate pool. The noise level is significantly lower than that

Figure 3. Comparison of IRM with FIRM



in a full image database. It is our expectation that the superiority of the proposed method over the original IRM will be more evident if the retrieval is performed on the full image database. However, due to the time consuming nature of full search, the related experiments will be included in our future work.

### Compare the Proposed Framework FIRM with DRM and SVM

In this experiment, we compare the performance of FIRM with two other existing approaches, including Dynamic Region Matching (DRM) and Support Vector Machine (SVM). This experiment aims to evaluate the learning algorithms used in the relevance feedback process. We apply the above three methods on the same test images and compare their performance with the three evaluation measures. Figures 4-6 show the comparison results in terms of the AVR, ANMRR, and Accuracy measures, respectively.

According to our experimental results, the overall results of SVM and DRM are not as good as that of FIRM in terms of the ANMRR and Accuracy measures. In constructing the SVM

for comparison, each positive image is used as a training data sample. And each input dimension of the One-Class SVM corresponds to a region in the query image, and its input value is calcu-

lated as  $\sum_{j=1}^n s_{ij} d_{ij}$  for each region  $i$  in the query

image. Through the learning of SVM using user feedback, a reasonably good combination of all query regions can be gradually discovered and used for future retrieval. In this study, we used a non-linear kernel (Radial Basis Function) for One-Class SVM learning.

The reason that DRM performs worse than FIRM is that the weights of regions in the query image are either increased or decreased by a fixed factor through relevance feedback, and therefore, some regions in the query image may quickly overpower the others through iterations of feedback-retrieval. Consequently, the performance gain obtained through feedback is limited.

In addition, we can observe from these three figures that the proposed framework (FIRM) has the best performance in terms of ANMRR and Accuracy values. However, the AVR measure does not second this observation. The reason

Figure 4. Comparison of FIRM, DRM, and SVM based on AVR measure

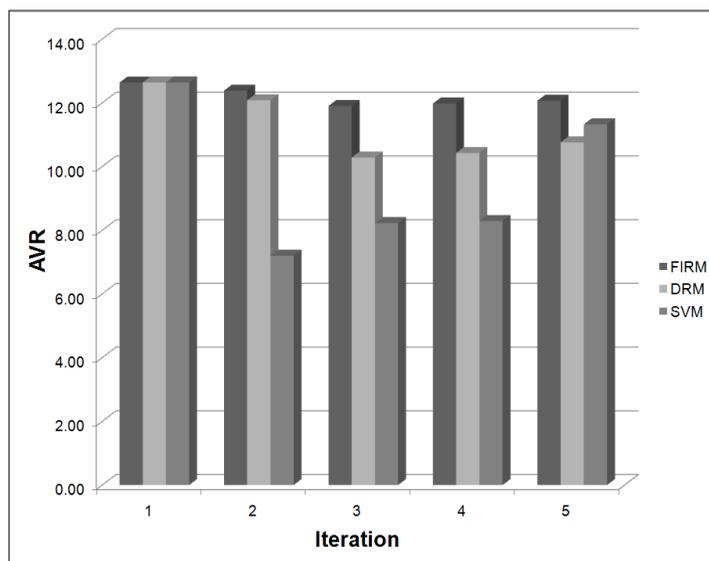


Figure 5. Comparison of FIRM, DRM, and SVM based on ANMRR measure

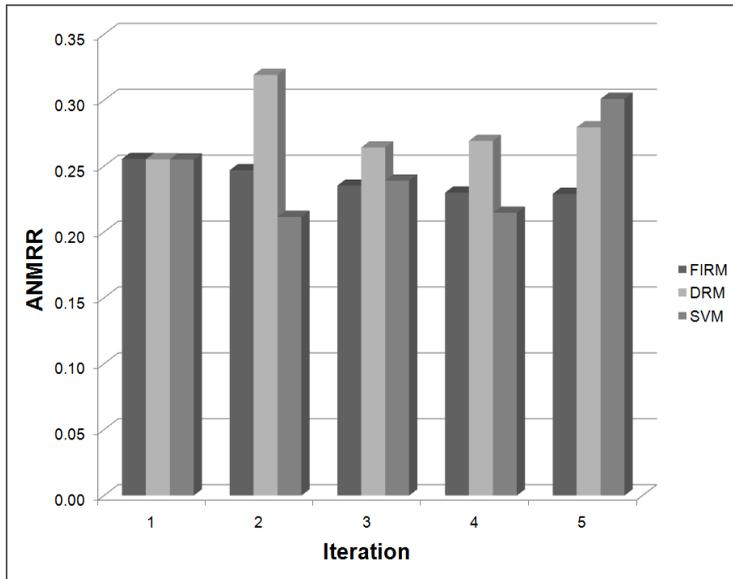
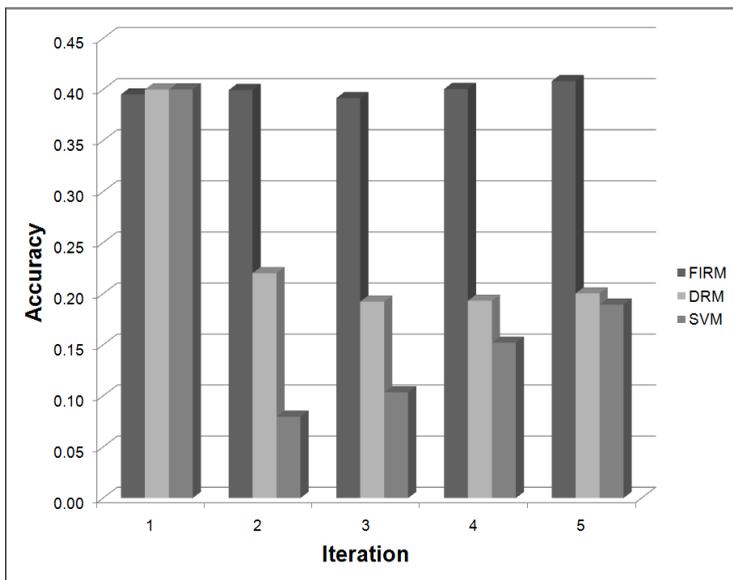


Figure 6. Comparison of FIRM, DRM, and SVM based on Accuracy measure



is due to the problem of AVR as mentioned in the first subsection of this section.

## CONCLUSION

In this article, we present a Feedback-based Image Clustering and Retrieval Framework

(FIRM) which improves the region-based image retrieval accuracy and efficiency by using a novel image clustering algorithm and integrating it with Integrated Region Matching (IRM) and relevance feedback. In our framework, images are first segmented into regions, then the Modified Genetic Algorithm (MGA) is applied to cluster image regions, resulting clusters of images with similar regions. This process effectively narrows down the search scope and, therefore, reduces the time-complexity in the subsequent retrieval step. IRM is then adopted with a new region-matching scheme that is suitable for relevance feedback, which measures the overall similarity between two images based on overall weighted region similarities. In addition, relevance feedback is adopted in this framework to reduce the semantic gap, which helps to progressively learn the user's preferred query regions based on the user selected positive images from the query results.

The performance of the system is evaluated on a large image database containing 8,900 general-purpose images with 82,556 image regions. Our experimental results demonstrate that MGA can effectively reduce the search space by 90%. In addition, the results in Figure 3 show significant improvement in Accuracy, Average Retrieval Rank (AVR), and Normalized Modified Retrieval Rank (NMRR), when compared with IRM without feedback. By comparing the proposed method with dynamic region matching (DRM) and support vector machines (SVM), we also demonstrate that the proposed learning algorithm outperforms both DRM and SVM.

## REFERENCES

- Babu, K. S., & Nagesh, V. (2008, March 9). Optimal solution for image retrieval using integrated region matching. In *Proceedings of the 2<sup>nd</sup> National Conference on Challenges & Opportunities in Information Technology (COIT-2008)*, Mandi Gobindgarh, India (pp. 223-226).
- Bradshaw, B. (2000, October 30-November 3). Semantic based image retrieval: A probabilistic approach. In *Proceedings of ACM Multimedia 2000*, Marina del Rey, CA (pp. 167-176).
- Carson, C., Belongie, S., Greenspan, H., & Malik, J. (2002). Blobworld: image segmentation using expectation-maximization and its application to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8), 1026-1038. doi:10.1109/TPAMI.2002.1023800
- Carson, C., Thomas, M., & Belongie, E. A. S. (1999). Blobworld: A system for region based image indexing and retrieval. In *Proceedings of the 3<sup>rd</sup> International Conference on Visual Information System*, Amsterdam, The Netherlands (pp. 509-516).
- Chen, S.-C., Sista, S., Shyu, M.-L., & Kashyap, R. L. (2000, January). An indexing and searching structure for multimedia database systems. In *Proceedings of the IS&T/SPIE Conference on Storage and Retrieval for Media Databases*, Santa Clara, CA (pp. 262-270).
- Cieplinski, L. (2001). MPEG-7 color descriptors and their applications. In *Computer analysis of images and patterns* (pp. 11-20). Heidelberg, Germany: Springer Berlin.
- Ding, Q., & Gasvoda, J. (2005). A genetic algorithm for clustering on image data. *International Journal of Computational Intelligence*, 1(1), 75-80.
- Hamada, M., Martz, H. F., Reese, C. S., & Wilson, A. G. (2001). Statistical practice: finding near-optimal bayesian experimental designs via genetic algorithms. *The American Statistician*, 55(3), 175-181. doi:10.1198/000313001317098121
- Hansen, N. (2008). *The CMA evolution strategy: A tutorial*. Retrieved April 2008, from <http://lautaro.fb10.tu-berlin.de/user/niko/cmatutorial.pdf>
- Haupt, R. L., & Haupt, S. E. (2004). *Practical genetic algorithms*. New York: John Wiley & Sons.
- Ji, R. R., Yao, H. X., & Liang, D. W. (2008). DRM: dynamic region matching for image retrieval using probabilistic fuzzy matching and boosting feature selection. *Signal, Image and Video Processing*, 2(1), 59-71. doi:10.1007/s11760-007-0037-0
- Myers, R. H., & Montgomery, D. C. (2002). *Response surface methodology: Process and product optimization using designed experiments*. New York: John Wiley & Sons.

- Ortiz, F., Simpson, J. R., Pignatiello, J. J., & Heredia-Langner, A. (2004). A genetic algorithm approach to multiple-response optimization. *Journal of Quality Technology*, 36(4), 432-450.
- Su, Z., Zhang, H. J., & Ma, S. P. (2000, November 13-15). Using bayesian classifier in relevant feedback of image retrieval. In *Proceedings of the IEEE International Conference on Tools with Artificial Intelligence*, Vancouver, BC, Canada (pp. 258-261).
- Tomassini, M. (1998). A survey of genetic algorithm. *Annual Reviews of Computational. Physics*, 3, 87-118.
- Wan, W. (2007). *Semi-parametric techniques for multi-response optimization*. Doctoral dissertation, Virginia Polytechnic Institute and State University.
- Wang, J. Z., Li, J., & Wiederhold, G. (2001). SIM-PLcity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23, 947-963. doi:10.1109/34.955109
- Yang, C., & Lozano-Prez, T. (2000, February 28-March 3). Image database retrieval with multiple-instance learning techniques. In *Proceedings of the 16th International Conference on Data Engineering*, San Diego, CA (pp. 233-243).
- Zhang, C., Chen, S.-C., Shyu, M.-L., & Peeta, S. (2003, December 15-18). Adaptive background learning for vehicle detection and spatio-temporal tracking. In *Proceedings of the 4th IEEE Pacific-Rim Conference on Multimedia*, Singapore (pp. 1-5).
- Zhang, C., & Chen, X. (2005, August 21-24). OCRS: An interactive object-based image clustering and retrieval system. In *Proceedings of the 4th International Workshop on Multimedia Data Mining (MDM/KDD2005), ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Chicago (pp. 71-78).
- Zhang, C., & Chen, X. (2005, July 20-22). Region-based image clustering and retrieval using multiple instance learning. In *Proceedings of the International Conference on Image and Video Retrieval*, Singapore (pp. 194-204).

*Chengcui Zhang is an assistant professor of computer and information sciences at the University of Alabama at Birmingham (UAB) since August, 2004. She received her PhD from the school of computer science at Florida International University, Miami, FL, USA in 2004. She also received her bachelor and master degrees in computer science from Zhejiang University in China. Her research interests include multimedia databases, multimedia data mining, image and video database retrieval, bioinformatics, and GIS data filtering.*

*Liping Zhou is a PhD student in the Department of Computer and Information Sciences at the University of Alabama at Birmingham (UAB) and member of Knowledge Discovery and Data Mining (KDDM) Laboratory. Her research interests include data mining, multimedia database systems, information and image retrieval. She received her BSc in computer science from Qiqihar University, China and MSc in computer science from University of Science and Technology Beijing, China. She is a student member of the ACM.*

*Wen Wan received her PhD in statistics from the Virginia Polytechnic Institute and State University. She is an assistant professor in biostatistics and bioinformatics unit in the Division of Preventive Medicine at the University of Alabama at Birmingham.*

*Jeffrey B. Birch is professor of statistics and director of graduate programs at Virginia Tech, where he has been a faculty member since 1977. Dr. Birch received his MSc (1970) and PhD (1977) degrees in biostatistics from the University of Washington. His area of primary research includes most aspects of regression analysis including robust regression, nonparametric and*

*semiparametric regression, and profile monitoring. Birch is a former associate editor of Biometrics (1989-1993).*

*Wei-Bang Chen is a PhD candidate in the Department of Computer and Information Sciences at the University of Alabama at Birmingham. He received a master's degree in genetics from National Yang-Ming University in Taipei, Taiwan and a master's degree in computer sciences from UAB. His main research area is bioinformatics. His current research involves microarray image and data analysis, biological sequence clustering, and biomedical video and image mining.*