IJICTR International Journal of Information & Communication Technology Research

Volume 3- Number 4- September 2011 (13-25)

An Image Annotation Method Based on CM Similarity Measure and Hybrid Relevance Feedback

Mehran Javani
Dept. of Computer Engineering
Young Researches Club, Behbahan Branch,
Islamic Azad University, Behbahan, Iran
m.javani@qiau.ac.ir

Amir masoud Eftekhari Moghadam
Dept. of Electrical Engineering and computer Science
Islamic Azad University, Qazvin Branch
Qazvin, Iran
eftekhari@qiau.ac.ir

Received: June 6, 2011- Accepted: August 20, 2011

Abstract— **Today, with the advent of digital imagery, the volume of digital images has been growing rapidly in different fields. So, there is an increasing need for providing an effective image retrieval system. In this paper, a semi-supervised k-means clustering method was introduced for image database clustering and image annotation. One of the most important parts of image clustering algorithms is to determine similarity of the images. To compute exact similarity measures, a new CM similarity measure was proposed here to make normalized and weighted features simultaneously so that similarity measure exploits normalized or weighted features in its formula to reach better performance. However, due to semantic gap, some images may be false clustered. A hybrid of three relevance feedback (RF) schemes was used to improve the accuracy of image clustering. (1) The images with the user who knows their irrelevance to a cluster were conducted to correct cluster by a long-term RF. (2) With regard to the images with the user who knows they are relevant to a cluster, feature weight of the clusters was estimated in order to provide a multiple similarity measure using a re-weighting RF. (3) To discover the exact place of the cluster centers, a cluster center movement (CCM) RF was used. Experimental results based on the Corel database including 1000 images and a satellite image database of Tehran city including 2400 images demonstrated the superiority of the proposed method in image database clustering.**

Keywords- ***Image retrieval, Relevance feedback, Image annotation, Image database clustering, Similarity measure.***

I. Introduction

With rapid advances in the software, hardware, cost reducing of data storage and capturing resources, the size of digital picture achievements is increasing rapidly. Therefore, there is a great need for large image database management and effective and efficient tools for image annotation and retrieval techniques. Image retrieval systems are divided to three types [1] of text based approaches, content based image retrieval (CBIR) approaches and automatic image annotation (AIA) approaches. Most existing image retrieval systems such as image search engines in Google and Yahoo are textual based ones in which images are searched using the surrounding keywords, captions etc [2]. In many image databases such as satellite and medical ones, there is no text information to help image retrieval task. With regard to the increasing number of images taken from these areas, manual management, retrieval and searching among the voluminous data are very impractical and challenging tasks. For example, NASA's Terra and Aqua satellites generate more than 3 Tb of images daily [3] or Radiology Hospital University of Geneva took more than 70,000 images per day, just in 2007[4]. So, a system should be developed to be able to do image analysis, interpretation, annotation, management and retrieval automatically based on the image content. Content-based image retrieval systems (CBIR) were introduced in 1990s to automatically annotate images. Nowadays, many techniques have been developed for CBIR. CBIR is the process of retrieving similar images to a specific query from a large image database. It is usually implemented in two phases: 1) annotation the images and 2) retrieval of images from image database. In the annotation phase, each image is annotated with some machine learning techniques according to its visual features. In the retrieval phase, when a user enters a query, a feature vector is extracted from query image content and then, using a similarity measure, this vector is compared with the vectors in the feature database. The images most similar to the query would be returned to the user.

The basic assumption of CBIR systems is that images with similar concepts also have similar visual features. Although these systems are not necessarily able to understand the contents of images, it is necessary to categorize images to similar groups by their visual features. In general, there is no direct link between low-level features and high-level semantic concepts. Low-level features are often incapable of describing the high-level semantic concepts in the user's mind. The difference between the power of low-level descriptors and rich meanings in the user's mind is defined as the semantic gap [5] [37]. There are machine learning techniques and relevance feedback methods for narrowing the 'semantic gap' down. The machine learning techniques used in image annotation can be divided to three categories of supervised, unsupervised and semi-supervised methods [6-10].

Supervised methods need annotation data to work. It should be noted that annotation is a subjective, time consuming and cost ineffective process which is prone to error [11-12]. Semantic concepts are very subjective for human beings so that they are not well-defined. A person at different times may have different opinions about an image. Due to training patterns and training time, supervised methods are used less frequently when the volume of data increases. Today, clustering methods are extremely taken into account since they do not require any training phases.

K-means algorithm is one of the most appropriate clustering algorithms in image retrieval area. The K-means (or hard C-means) algorithm starts with K cluster centers which is initially selected randomly. Each image in the database is then assigned to the closest cluster centre. The centers are updated by means of the associated images. The process is repeated until the criterion of mean square error is (to be) satisfied. These algorithms have two disadvantages. First, they are sensitive to the selection of initial cluster centers so that they might be trapped in a local optimum [35]. Second, the performance of the algorithm depends on the selection of similarity measure [35]. To overcome the former, semi-supervised clustering is used and, to handle the latter, the user's feedback is applied and a new multiple CM (Canberra Mahalanobis) similarity measure is proposed to estimate an individual similarity measure for every cluster.

Relevant feedback is an automatic process to identify user intentions and learn the similarity concept from the user's point of view [34]. Feedback of users is employed to reduce the semantic gap between what queries say (low-level features) and what the user think. By interaction with users, the performance of CBIR systems improves considerably [4]. RF algorithms are usually used to improve the accuracy of supervised CBIR systems. On the other hand, RF algorithms are rarely paid attention in articles for unsupervised methods. In some clustering based image annotation systems, some images may be assigned incorrect labels. In this paper, a new hybrid of three RF including *long-term RF*, *reweighting RF* and *cluster centers movement RF* was introduced to improve the accuracy of image annotation and image database clustering.

II. Related Works

Researchers try to develop RF techniques for achieving higher performance. This is the most important stage in CBIR. For this purpose, there is an attempt to use the user feedback in the best way. In this section, various methods of relevance feedbacks are reviewed and categorized.

A. Learning Method for Relevance Feedback

As shown in Figure 1.a, CBIR system computes the distance between query image and all images of database. Then, according to a similarity measure, it retrieves *t* images with less distance. From the user's point of view, some of the images are relevant and some are irrelevant. Thus, a learning algorithm should be used to learn the subjectivity of the user. In general, RF learning methods can be divided to supervised, semi-supervised or unsupervised methods.

1) Supervised Relevance Feedback

The supervised RF has a paradigm in various disciplines such as pattern recognition and machine learning. These systems are controlled by user feedback and are called Human-Computer Interaction (HCI). HCI tries to simulate the human understanding system using a non-linear separator analysis [8]. Three basic methods of learning user opinion include *query point movement (QPM)* [13-15], *re-weighting* [5][13-14] and *Bayesian inference (BI)* [13-14]. Some researchers have used a hybrid of three methods to achieve better performance [2][14].

- *Query point movement*: this method supposes that there is an ideal point for every query in feature space. The purpose of retrieval process is to discover this point. *QPM* idea is to move the query to positive example (PE) positions and far away from negative example (NE) positions [13]. As shown in Figure 1.b, query point $x_i^{(j)}$ is moved to $x_i^{(j+1)}$ by using Rocchio formula that would be included further relevant images.
- *Re-weighting*: The main purpose of re-weighting is to give more importance to some features

according to feedback samples. The evident feature should have less scattering on the class samples. In MARS [15], any feature is assigned a weight with regard to the reverse variance. Figure 1.c shows much weight dedicated to X feature which causes more relevant images to be included. Deselaers et al. [31] used L_2 as a similarity measure and learned weights of features by gradient descent which simultaneously minimized the distance of relevant images while maximizing distance of irrelevant images. This method has two drawbacks. First, it can evaluate optimal weights for only one query image. Second, this method may converge to local optimum for high dimensional feature space because the number of feedback images is limited.

- *Bayesian inference*: Another method for RF processing is probabilistic Bayesian model [13]. BI approach uses Bayesian framework for deductive probabilistic estimation. Since probabilistic distribution on the whole image database is updated after any feedback, so the whole system can be improved.
- *Hybrid method*: Researchers try to combine different available methods and obtain an optimal one. For instance, in [14], three methods of QPM, re-weighting and BI were combined and reinforcement learning was used to develop an optimal feedback for query. In BALAS [2], a hybrid of QPM and BI methods was proposed. The degrees of relevant/irrelevant images were determined by estimating online probability density functions.

2) Semi-Supervised and Unsupervised Relevance Feedback

Semi-supervised and unsupervised methods are more popular because they are less dependent on user opinion, especially in the contexts such as web based image retrieval that user opinion is very time consuming. Hence, machine-computer interaction (MCI) systems are an effective method. An MCI system was introduced to search image database by training RBF neural network [16]. This system tries to simulate user opinion via an unsupervised RF approach.

3) Short and Long Term Learning

Users usually think differently about a single image at different times [17]. Therefore it is preferred to use multiple user opinions in multiple sessions to achieve more confident results. RF is categorized into *long term* and *short term* learning when we face a couple of query sessions.

- *Short term learning*: The traditional short-term relevance leaning scheme always starts a new query session without any assumption about the query formulations or feature weights of the images and it repeats interaction with user while satisfactory results to be achieved.

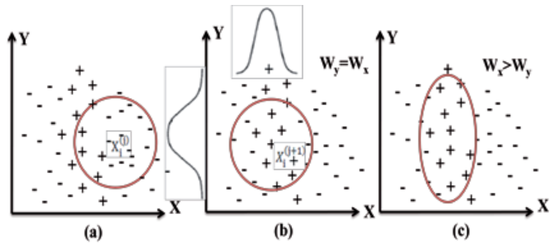


Figure 1. a) initial query b) reweighting is applied to stretch the neighborhood c) movement query with CCM

- *Long term learning*: This learning scheme keeps a global memory for each image database to store the last query formulation and feature weights of the images when the image is previously used as a query. Thus, the relevance feedback learning for a new query session can start from its previous state to speed up the learning process.

Xiaofei et al. [32] used both long term and short term RF to improve the performance of image retrieval. Long term RF first used to construct a semantic space from user interactions, then system used a learning function to distinguish relevant images from irrelevant images. Short term RF is used to reduce the mistakes of choice choosing relevant features and choosing feedback images under control of user.

B. Relevance Feedback for Clustering

In recent years, there has been a growing interest in clustering. In a great number of articles, relevance feedback is used for supervised learning. But, in unsupervised methods similar to the proposed system, RF learning methods are difficult to manage to rear instruction samples. RF is used in unsupervised systems with different purposes to improve the performance of systems. For instance, in CLUE [18], a method was proposed that initially clustered image database based on graph theory and Ncut. In this method, a graph representation of neighborhood target images was defined as G=(V, E), where the nodes V represent the images and the edges E are weights between every pair of nodes. The weights w_{ij} are defined for the edges according to the similarity between the nodes. Graph-partitioning is applied in this graph. For image retrieval, this method is applied in a two-stage hierarchical learning: in the first stage, the top r neighbors of the query are retrieved; in the second stage, the Ncut clustering method is applied to the query image and its nearest neighboring target images; then, the system displays the image clusters and similarity measure is corrected by relevance feedback based on the user's desire. Notwithstanding the great applicability of this method, it cannot be directly mapped between low level feature and high level semantic. Also, computation of this method due to using graph-partitioning and Ncut for large dataset is very complicated.

In SKKC [16], semi-supervised k-means clustering is used to annotate medical images and then relevance feedback is used to improve the performance. However, this method does not use any learning methods in RF stage and only tries to correct clustering images. Thus, this method needs many feedbacks.

Bhanu et al. [33] used a semi-supervised fuzzy clustering (SSFCM) method to learn class distribution by proposing an objective function. The SSFCM method is in conjunction with probabilistic feature relevance feedback to improve the retrieval accuracy. A specific query uses probabilistic feature relevance for weighed features. The SSFCM clustering method needs more computations because objective function must calculate inverse matrix at each iteration. Thus, this method is not suitable in situations which include interaction with users.

III. The Proposed Approach

In this paper, a semi-supervised k-means clustering method was proposed by introducing a new *CM similarity measure* and a *three-stage hybrid relevance feedback scheme*. Figure 2 shows the framework of the proposed system.

K-means algorithm is very suitable in interactive applications because its time complexity is O (n) [35]. But, it has two problems. The main problem of k-means is its sensitivity to the selection of the initial cluster centers which may be converged to local optima. To address this problem, a few annotated images (about five images) were used here as training data in order to find the boundary of the clusters as a semi-supervised approach. Each initial cluster center was determined by the mean values of feature vectors of the annotated images. Second, the efficiency of the clustering results is influenced by similarity measure. A CM similarity measure was proposed to exactly estimate similarity of images. Then, fuzzy membership values of each image to cluster centers were calculated according to the distance of each image from all clusters.

The pseudo code of semi-supervised K-means clustering with CM similarity measure is presented is Figure 3. In many conditions, it is not sufficient to find the boundary of clusters based on the low level features because there is a vast diversity of data. In this way, some images are wrongly assigned to adjacent clusters so that the distance of the false clustered images from the adjacent clusters is less than its distance from the correct cluster. To improve the proposed clustering method and to bridge the low level features and high level semantic concepts, a new hybrid of three-stage relevance feedback was introduced.

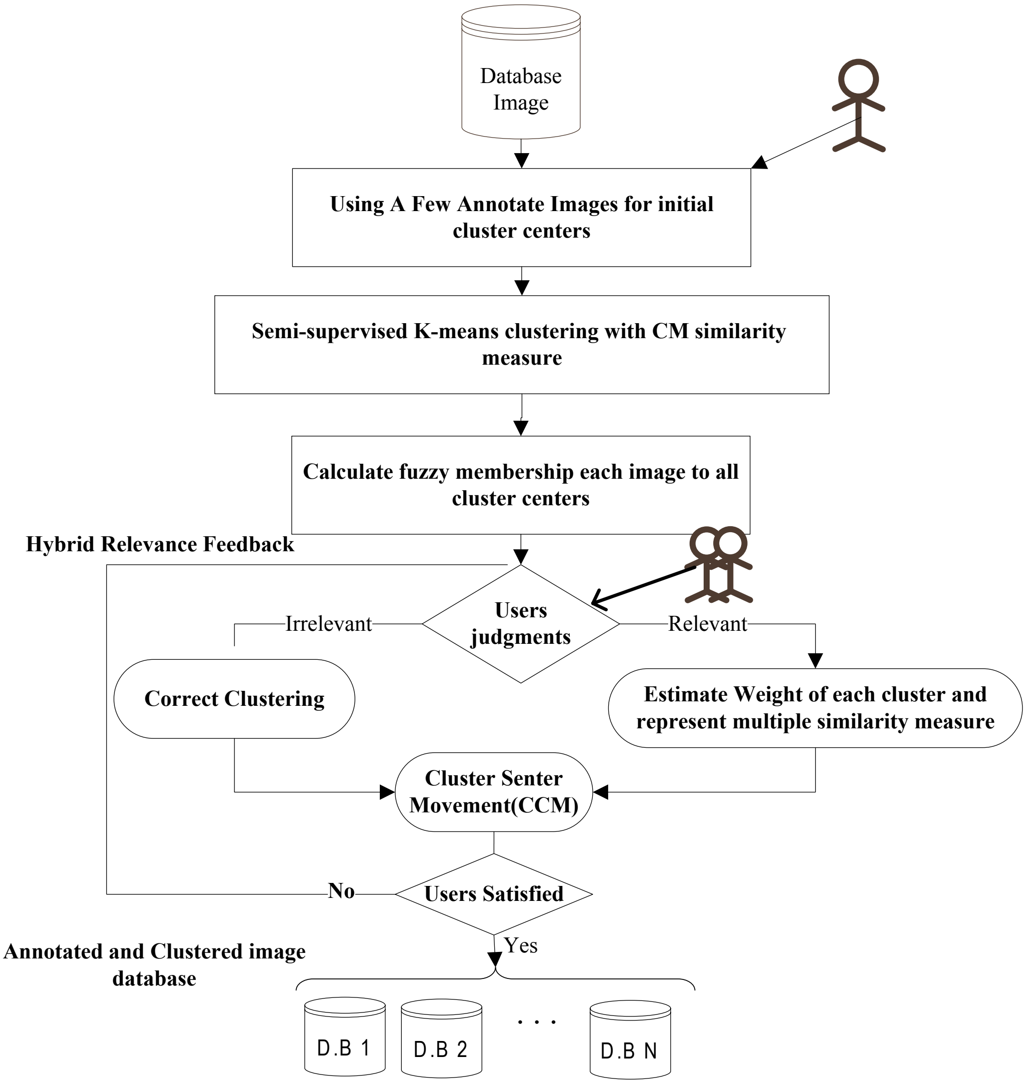


Figure 2. Framework of proposed system

Table 1. Taxonomy of similarity measures

Name	References	Formula	Advantages	Disadvantage
Minkowski	[22-26]	$D=\left(\sum_{i=1}^{d}[x_i-y_i]^{\frac{1}{n}}\right)^n$	Basic-popular;	To dominate features with high value variance on other features;
Euclidean (L2)	[22, 24, 26]	$D=\left(\sum_{i=1}^{d}[x_i-y_i]^{\frac{1}{2}}\right)^2$	Popular; Metric; Special case of Minkowski when n=2; Robust to rotation and transition; Suitable for hyper sphere clusters;	Unsuitable if patterns linear inseparable or complex;
Chebydhev (L∞)	[24, 26]	$D=\max\lvert x_k-y_k\rvert$	Special case of Minkowski when $n=\infty$	
Manhattan(L1)	[22, 24, 26]	$D=\sum_{i=1}^{d}\lvert x_i-y_i\rvert$	Known as the city-block distance; Special case of Minkowski when n=1; Suitable for hyper rectangle clusters;	
Mahalanobis	[23, 26]	$D=(X-Y)^T S^{-1}(X-Y)$ S is covariance matrix between patterns	Suitable for hyper ellipse clusters; Robust to any unique linear transformation;	If features not correlated and squared, Mahalanobis distance equal to Euclidean distance;
Canberra	[22-26]	$D=\sum_{i=1}^{d}\frac{\lvert x_1-y_1\rvert}{\lvert x_1\rvert+\lvert y_1\rvert}$	Can be called comparative (relative) Manhattan;	
Statistic value χ^2	[22, 25,-26]	$D=\sum_{i=1}^{d}\frac{(x_i-y_i)^2}{m_i}$ $m_i=\frac{x_i+y_i}{2}$	It emphasize the elevate discrepancies between the different feature vectors based on the importance of distributing them;	
Jeffrey divergence	[25- 26]	$D=\sum_{i=1}^{d}(x_i\log\frac{x_i}{m_i}+y_i\log\frac{y_i}{m_i})$ $m_i=\frac{x_i+y_i}{2}$	Known as the Kullback-Leibler divergence asymmetric-robust to noise and size of histogram bins;	
Weighted mean variance (WMV)	[26]	$D=\sum_{i=1}^{d}\frac{\lvert x_i-y_i\rvert}{\partial_i}$	It is a kind of Manhattan distance that normalized with variance of features;	
Czekanowski	[24, 26]	$D=1-\frac{2\sum_{i=1}^{d}\min(x_i-y_i)}{\sum_{i=1}^{d}(x_i-y_i)}$		Can be worked only on features with nonnegative value;
Inner product	[23, 26]	$D=\frac{\sum_{i=1}^{d}x_i.y_1}{\lvert x_i\rvert+\lvert y_i\rvert}$	Overcome to difficult Minkowski (dominate features with high value) family by normalized features;	

```
n: is a number of class;
T: is a number of images in each class;
L: is a Labeled of each class
Xi : is a n*T matrix which Contain of Labeled images;
//Initialized Cluster Center
for i=1 : 5
     Xi=L  // Labeled five images of each class
// Semi-supervised K-means Clustering with CM
similarity measure
σ=Diag(1/cov(feature)); // σ is a initial weight of features
Center(1:n)=initial(X(1:n));
//assigned initial cluster center means of labeled imaged
features
for k=1 : iteration
  for i=1 : n*T
    for j=1 : n
       dis(j)=(center(j,:)-feature(i,:))* σ *(center(j,:)-
feature(i,:))'
          /abs(center(j,:))+ abs(feature(j,:));

    End;
    Xi=min(dis);
  end;
  for j=1 : n    // updated cluster center
     Center(1:n)=meanlabeled(Xi);
  end ;
end;
```

Figure 3. Pseudo code of semi-supervised K-means clustering with CM similarity measure.

In this section, the similarity measure is illustrated. Then, image database clustering by k-means algorithm with CM similarity measure are suggested to estimate the similarity of images. In the last part of this section, a new hybrid relevance feedback is introduced to improve the accuracy of the clustering algorithm.

A. Similarity Measure

Similarity measure is the main component of many models and algorithms in machine learning, pattern recognition and computer vision. Determination of similarity between images is an important task in image retrieval systems. Moreover, the role of users in analyzing the response of system in an interactive process is very important for understanding the image similarity model [19]. The basic task of most clustering algorithms is to measure similarity between two patterns in the feature space. A similarity function on a data set X is defined to satisfy the following conditions [20]:

1) *Symmetry:* $S(x_i, x_j)=S(x_j, x_i)$
2) *Positivity:* $0 \leq S(x_j, x_i) \leq 1$ *for all* x_i, x_j
3) *Triangle inequality:* $S(x_i,x_j)S(x_j,x_k) \leq [S(x_i,x_j) + S(x_j,x_k)]\, S(x_i,x_k)$ *for all* x_i, x_j *and* x_k

4) Reflexivity: $S(x_j, x_i)=1$ iff $x_j = x_i$

If conditions 3 and 4 are met, the similarity measure is called metric.

1) Similarity Measures Taxonomy

The performance of many clustering and classification methods such as K-means and RBF depends on the appropriate definition of distance function. Instead, pre- defined distance function based on prior knowledge about the application is very pleasant to be used for learning distance function. This may start with initial choice and continue based on supervisory data about the application. Unfortunately, under unsupervised learning issues, learning of similarity measure is not a well-defined criterion. For example, clustering algorithms such as k-means with different similarity measures produce different clustering results [21]. Table 1 shows some popular similarity measures with their advantages and disadvantages.

2) CM Similarity Measure

As shown in Table 1, eleven similarity measures that are often used in the literature can be classified into three categories: *basic similarity measures*, *weighted similarity measures* and *normalized similarity measures*. Table 2 demonstrates the classification of similarity measures into three categories.

- *Basic similarity measure*: Basic similarity measures are the ones whose formula has the following form:

$$D = |x - y| \quad (1)$$

- *Weighted similarity measure*: Weighted similarity measures have a weighted parameter *W*. The weighted parameter can be obtained in various methods. General form of weighted similarity measure has the following form:

$$D = (X - Y)^T W (X - Y) \quad (2)$$

- *Normalized similarity measure*: Normalized similarity measures are the ones whose formula is divided to combination of their absolute of features. General form of normalized similarity measure is as follows:

$$D = \sum_{i=1}^{d} \frac{|x_i - y_i|}{|x_i| + |y_i|} \quad (3)$$

Figure 4 shows recall-scope curves of the three similarity measure approaches. As we observe, the weighted similarity measures and normalized similarity measures have much better performance than basic similarity measures. According to this principle, we can use from properties of normalization and weighted approaches simultaneously and propose a new similarity measure as in Equation 4.

$$CM = \sum_{i=1}^{d} \frac{(X_i - y_i)^T S_{i,i}^{-1} (x_i - y_i)}{|x_i| + |y_i|} \quad (4)$$

where *S* is a covariance matrix of whole image database that its non-diagonal elements are set to zero. This similarity measure is the combination of Canberra and Mahalanobis, thus we called it relative Mahalanobis or *CM* (Canberra and Mahalanobis).

B. A New Hybrid Relevance Feedback

As shown in Figure 5, after clustering, in previous stages, the fuzzy membership values of each image to cluster centers are calculated according to the distance of each image from each cluster. In the last stage, the performance of system is improved by the proposed hybrid relevance feedback. In the clustering process, some images may be clustered wrongly. Hence, relevance feedback was used to take high level semantic concept into account and to reach more accurate image clustering. All Images in each cluster ranked according to distance from the cluster centers and images with the maximum distance were shown to users for relevance feedback process. In this regard, it is possible to have positive and negative examples. Experimental results showed that the performance of system should be increased. The images that the user sets as relevant were assigned to correct cluster (positive example), so the system recognized them correctly. The images the user sets as irrelevant belonged to other clusters (negative example) but the system assigned them to this cluster. Long term RF was used for correct clustering. Thus, fuzzy membership value would set zero for the cluster that the user knew as irrelevant and image would be allocated to the cluster with the maximum membership value. The experiments showed that irrelevant images after averagely two feedback iterations were assigned to the correct cluster. After each feedback, cluster center movement (CCM) technique based on QPM was used to find the ideal point for every cluster in feature space with Eq. (5):

$$X_i^{(j+1)} = \alpha x_i^{(j)} + \beta \sum_{Y_k R} \frac{Y_k}{|R|} - Y \sum_{Y_k N} \frac{Y_k}{|N|} \quad (5)$$

where $x_i^{(j)}$ and $x_i^{(j+1)}$ are previous and updated cluster centers, respectively, and $x_i^{(j)}$ means that a user submits the *i*th cluster in image database as the query and has *j* RF iterations and $x_i^{(j+1)}$ is updated cluster center in *(j+1)* RF iterations. *α, β* and *γ* are adjustment coefficients that set *α* to 3*|*C-R-N*| and *β* and *γ* are equal to 0.33.

Table 2. Categories of similarity measures

Category	Similarity measure
Basic similarity measure	Minkowski Euclidean Chebydhev Manhattan
Weighted similarity measure	Manhattan Weighted Euclidean Weighted Mahalanobis WMV
Normalized similarity measure	Czekanowski Inner product Jeffrey divergence χ^2 Canberra

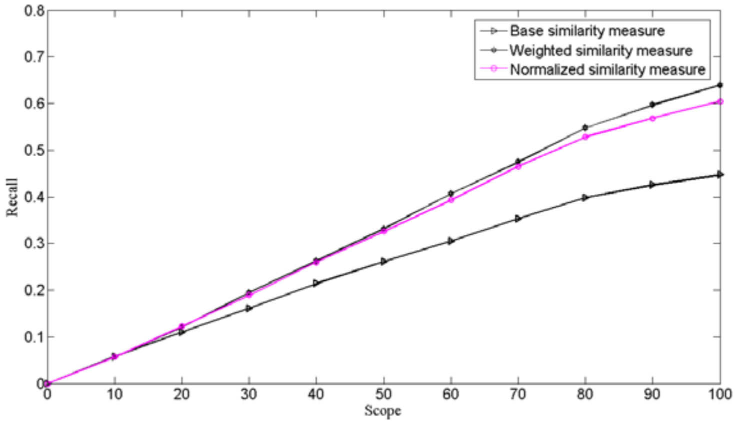


Figure 4. Recall-scope curves for three categoroes of basic, weighted and normalized similarity measures.

```
F=Calculate fuzzy distance of each image from different cluster center;
While (User Satisfied for clustering result)
        Show 20 images near to boundary which max distance from cluster center;
        User labeled relevant or irrelevant to this cluster;
        If (user set relevant)
          Xi=Label of images;
       else  // Correct Clustering
          Fi=0;
          Xi = max index(Fi); // image allocate to cluster with max degree of F
        end;
       Updated cluster center with Equation (5);  // CCM
End;
Updated σ by covariance of current labeled images;  // estimate feature weight of each cluster;
```

Figure 5. Pseudo code of semi-supervised K-means clustering with CM similarity measure.

Also, let the set of relevant images identified at the jth iteration be R and the set of identified irrelevant images be N and C is all images in database. Z_k are images that belong to region R or N.

In this regard, the user preference and his/her point of view toward each feature type is very significant. For example, two objects with similar shapes but different colors like a lemon and an orange are recognized similar based on their shape features by some users and dissimilar based on color features by others [13].

Relevance feedback was repeated until the user's satisfaction with clustering result. Then, the images that the user sets as relevant in several feedbacks to estimate the feature weight of each cluster were used. Since important features in various clusters were different, re-weighting RF was used to obtain suitable feature weight of each cluster according to scattering of features. Multiple similarity measures are presented in Eq. 6. A special weighting vector W was obtained for each cluster via relevance feedback.

$$cluster(I) = \arg\min_{j\in(1..k)} \left(\frac{(F_I - C_J)^T W_J (F_I - C_J)}{|F_I| + |C_J|} \right) \quad (6)$$

In response to the query image I, returned cluster J minimizes Eq. 6. W_j is the weight vector of cluster j that was calculated in the relevance feedback phase.

IV. Experimental Results

In this paper, a clustering technique was used for image annotation and a relevance feedback method was used to improve the accuracy of annotation. The proposed algorithm was implemented by MATLAB 7.6 and the hardware environment of all experiments was a Pentium4 PC, 7.300 CPU, with 1GB memory, Windows XP.

For system assessment, first, the original image was segmented to constitute the required image database. Then, low-level features were extracted from the image database and the proposed algorithm was tested on these features to be measured efficiently.

A. Image Databases

In this work, used two image databases were used. The first one was *Tehran satellite image database* [34], [36]. A satellite image of Tehran with the size of 3243 × 3243, as shown in Figure 6, which was partitioned to 2400 sub-images with the size of 90×90. Adjacent sub-images overlapped with each other by 50 percent. Therefore, Tehran satellite image database included 2400 images with five semantic concepts including *trees*, *road/cars*, *runway/desert*, *urban* and *airplane*. Figure 7 shows five image samples from Tehran satellite image database with their annotated semantic concepts.

The second image database was *Corel database* including 1000 images of 10 semantic classes. Each class in Corel database consisted of 100 images in

384× 256 or 256×384 resolution. Figure 8 shows one sample image from each semantic class.

B. Performance Evaluation

The evaluation of the system was done in two main parts. First, the effect of similarity measure was evaluated. In the second part, the performance of hybrid of three-stage relevance feedback was considered. Recall/Scope was used to evaluate the performance of system. All results presented in this section were achieved from the average of 1000 random queries.

1) Performance Evaluation of Similarity Measure

Determining the similarity of images based on low-level features is one of the major topics in machine vision. There is no general rule or method for selecting appropriate similarity measures. In literature different similarity measures are used which some of these similarity measures are presented in Table 1.

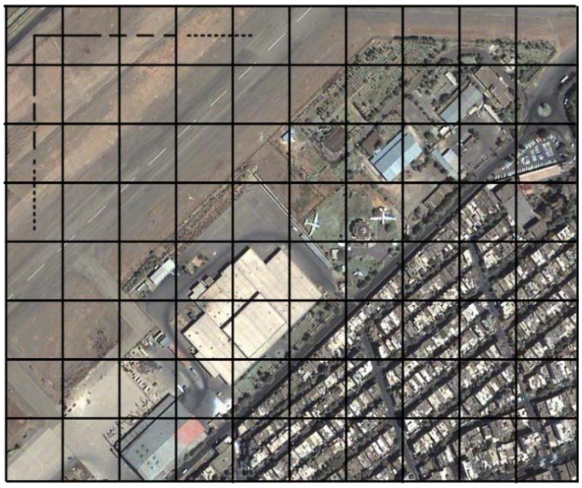
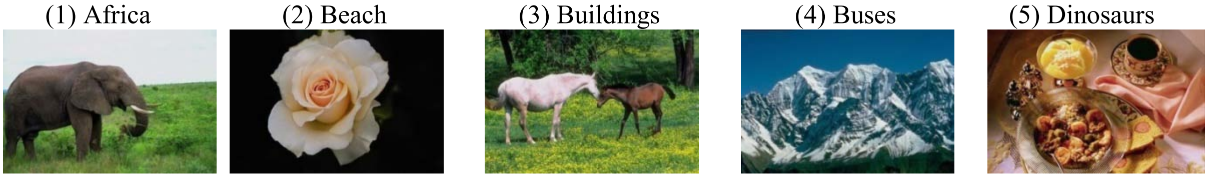


Figure 6. Partial satellite image of Tehran partitioned into overlapped sub-images with size 90x90.

(1) Trees, 520 (2) Road and cars, 712 (3) Runway and desert, 250 (4) Urban, 870 (5) Airplane, 48

Figure 7. Image samples of *Tehran satellite image database* with their annotated semantic concepts. The caption of each image shows its cluster ID (semantic concept) and number of images with this cluster ID in the image database.

(1) Africa (2) Beach (3) Buildings (4) Buses (5) Dinosaurs



(6) Elephants (7) Flowers (8) Horses (9) Mountains (10) Food

Figure 8. Image samples of *Corel database* with their annotated semantic concepts.

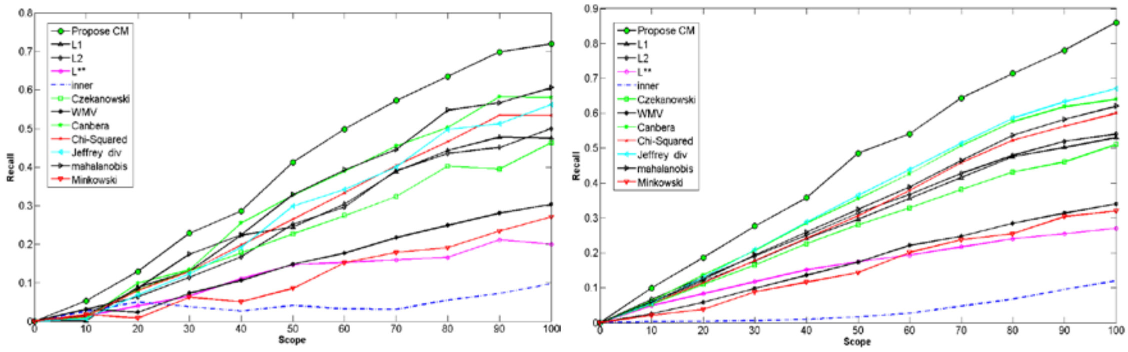
Table 3. Color, texture and shape descriptors

Feature	Advantage	Descriptor	Dimension
Color	Simplest, most efficient and most applicable feature in CBIR and invariant to resize, rotation and transformation but sensitive to illumination and other external factors. Selection a color space is the main step in color feature extraction [27].	mean, standard deviation, and skewness for each of three color channels L, a, b	9
Texture	Texture has a widely used in Remote Sensing (RS) image classification and retrieval. Especially, in area separation with different textures such as tree planting areas and urban areas.	inertia, energy, entropy, contrast, local homogeneity, correlation and difference moment extracted from GLCM matrix [5][30].	7
Shape	Shape has a main role to identify some linear entities (for example, river, and road) and objects with regular shape (for example, building, farmland and hangar). Zernike moments have properties, such as rotation invariance, robustness to noise, expression efficiency.	Zernike moments [28-30]	9

In this section, the performance of the proposed CM similarity measure was evaluated and then it was compared with other similarity methods. Recall/scope curve of 12 similarity measures are shown in Figure 9. As shown in this diagram, the proposed CM similarity measure had the highest accuracy because of making normalization and using weighted features simultaneously 11% and 14% improvement was observed in annotation accuracy for Corel and satellite images databases, respectively, compared with the best similarity measure. After the proposed similarity measure, Canberra, Mahalanobis and Jeffrey divergence had the best results, respectively. Since these similarity measures performed a kind of normalization or weight to the features in their formulas, they had higher performance than other similarity measures. Inner product and L_1 similarity measures had the lowest performance. As shown in Figure 9 in both databases, the performance of similarity measures was not much different because similar features were used in both databases. Moreover, the performance of similarity measure much depended on application and low-level features.

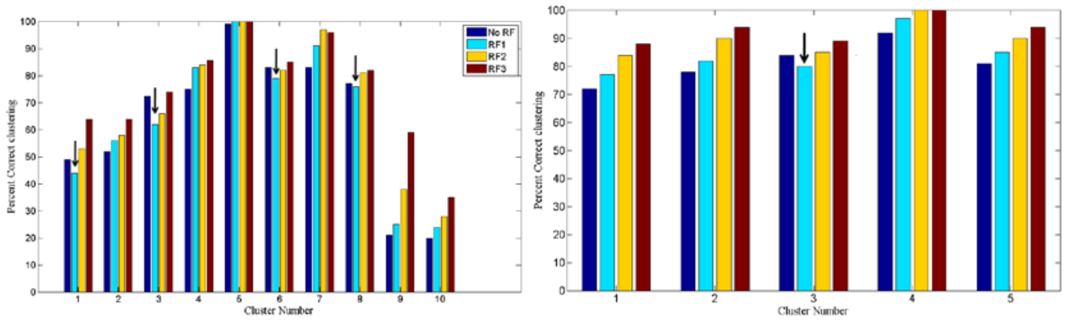
2) Performance Evaluation of Relevance Feedback

To examine the effect of hybrid relevance feedback method, the accuracy of annotation in three-stage feedbacks is presented in Table 4. As shown in Table 4, after three feedbacks, the accuracy increased by 7.3% and 13.6% for the Corel and Tehran satellite image databases, respectively. But, since the accuracy of the proposed clustering system was good, showing the ability of the proposed relevance feedback was not expected.



(a) (b)

Figure 9. Recall of clustering results for: (a) Corel Database; (b) Tehran satellite image database in different similarity measures.



(a) (b)

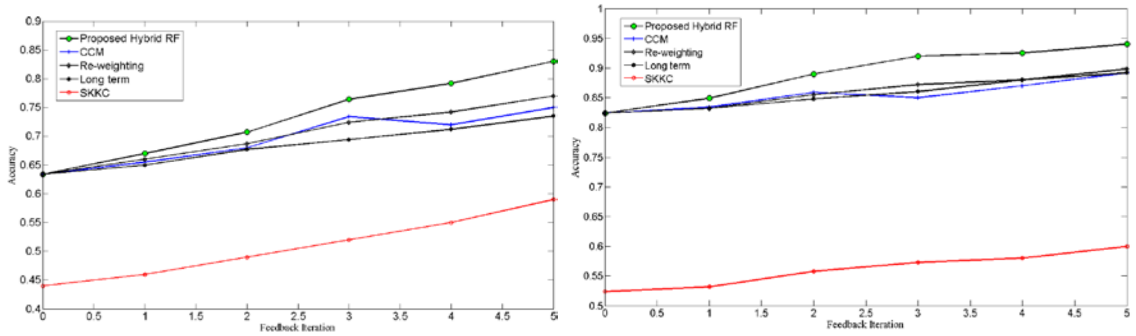
Figure 10. Effective hybrid relevance feedback in clustering accyracy; a) Corel Database b) Tehran satellite image database.

Table 4. Accuracy of annotation hybrid relevance feedback in three feedback for proposed system.

	Without RF	1th RF	2th RF	3th RF
Tehran satellite image database	%86.3	%89.7	%91.2	%94.6
Corel image database	%71.5	%74.6	%80.3	%85.1

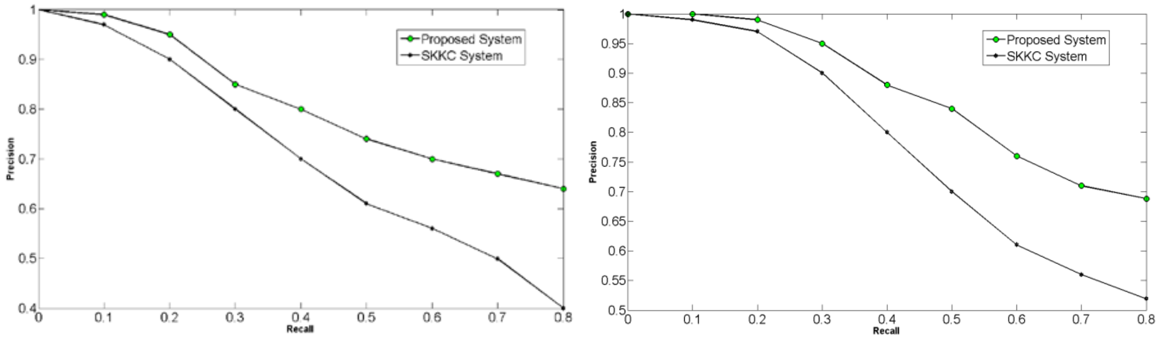
Table 5. Accuracy of annotation hybrid relevance feedback in three feedback for the system with χ^2 similarity measure.

	Without RF	1th RF	2th RF	3th RF
Tehran satellite image database	%58.1	%65.4	%71.2	%79.8
Corel image database	%57.4	%64.9	%72.4	%76.3



(a) (b)

Figure 11. Accuracy / Feedback iteration curve for proposed system and three component CCM , Re-weighting and Correct clustering with SKKC system; a) Corel Database b) Tehran satellite image database.



(a) (b)

Figure 12. Precision/ Recall curve for proposed system and SKKC system; a) Corel Database b) Tehran satellite image database.

Therefore, to evaluate the proposed relevance feedback, the system with χ^2 similarity measure that had lower efficiency was used. As demonstrated in Table 5, the accuracy of the system increased after three feedbacks until 18.9% and 21.7% for the Corel and Tehran satellite image databases, respectively.

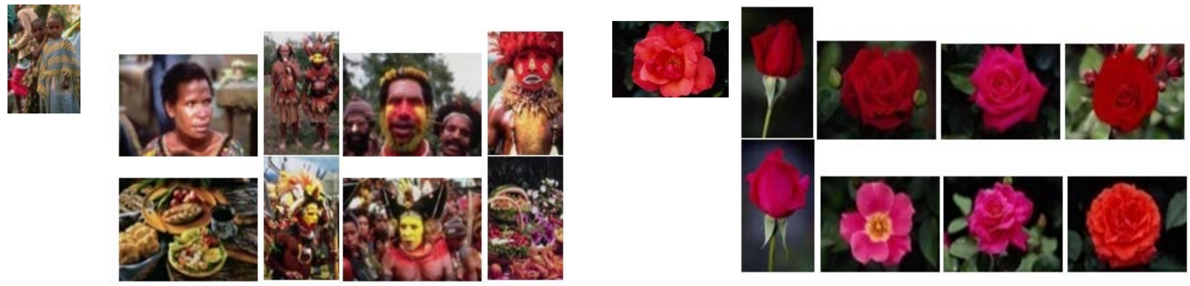
The accuracy of different clusters in three feedbacks is shown in Figure 10. As shown in Figure 10.a, for Corel database, the accuracy of clusters 1, 3, 6 and 8 decreased after performing some feedbacks. According to the experiences by different people after several repetitions, it was concluded that, about 3 to 9 percent of user feedback images were wrong. In general, three reasons can be mentioned for this problem.

- First, human being is sensitive and unpredictable and the user judgment is subjective.
- Second, images in clusters' boundary are very similar and the user may mistakenly mark them as relevant leading to the decrease in cluster accuracy.
- Third, the user may make a mistake in putting check marks or vice versa.

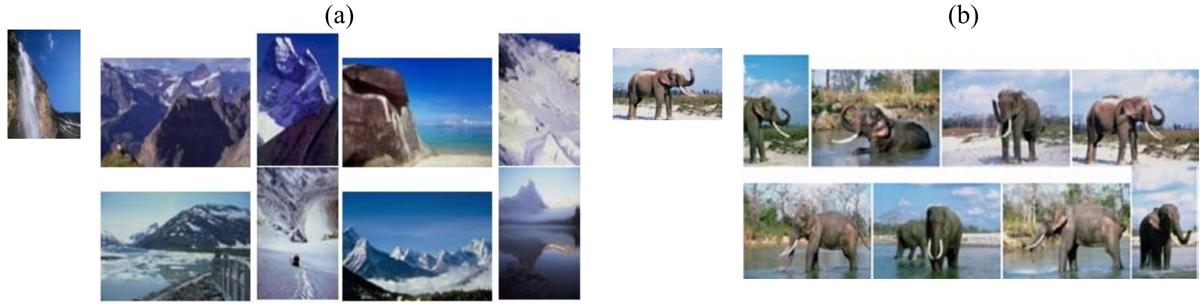
About satellite images, the same kind of mistakes may happen in lower scales, especially in the first and second cases. However, it was more evident in the third case due to the vast variety of categorization in satellite images

Figure 11 illustrates the curve of system accuracy in different relevance feedbacks in comparison with three RF including Long term, Re-weighting, CCM/hybrid RF and SKKC system. The result of two databases showed the accuracy increase in hybrid method by about 18% whereas, for three methods (CCM, Re-weighting and Correct clustering), the accuracy increase was between 9 and 11 percent and, for SKKC, it was approximately 9%.

The Curve precision-recall proposed system and SKKC system are shown in Figure 12 (the precision-recall curves of the proposed system and SKKC system are shown in Figure 12). In both databases, the performance of the proposed system was significantly better.



(a) (b)



(c) (d)

Figure 13: Retrieval example from the Corel dataset for four queries evaluated. The top-left corner image is the query and top 8 retrieval results obtained by proposed scheme algorithm. (a) Africa; 6 matches out of 8. (b) Flowers; 8 matches out of 8. (c) Mountains; 7 matches out of 8. (d) Elephants; 8 matches out of 8.

(a) (b)

(c) (d)

Figure 14: Retrieval example from the Satellite dataset for four queries evaluated. The top-left corner image is the query and top 8 retrieval results obtained by proposed scheme algorithm. (a) Trees; 8 matches out of 8. (b) Airplane; 6 matches out of 8. (c) Urban; 8 matches out of 8. (d) Road and cars; 8 matches out of 8.

To illustrate the performance of the proposed system, several images were selected as query and are shown in Figures 13 and 14. The images were related to semantics such as Africa, flowers, elephants and mountains from Corel database and trees, airplane, building-residential and road-cars from Tehran satellite image database. For each query example, the precision of the query results was examined depending on the relevance of image semantics.

V. CONCLUSION AND FUTURE WORKS

Due to semantic gap between low-level features and high-level semantics, CBIR systems cannot be good at image annotation. To reduce semantic gap, we use semi-supervised k-means clustering by CM similarity measure and hybrid relevance feedback scheme. Due to utilizing normalized and weighted features simultaneously the proposed similarity measure could perform better about 14% rather than the best available similarity measure. In clustering some images may not annotate correctly. To improve performance of system, a hybrid of three relevance feedback is used. (1) *Long-term RF* is used to conduct images to correct clusters, (2) *Re-weighting RF* is used to estimate feature weight of the clusters to provide a multiple similarity measure and (3) *CCM* is used to determine the exact place of cluster centers. After three stages of proposed relevance feedback scheme, accuracy of annotation increased about 18%.

This hybrid relevance feedback is utilizable for every clustering method based on cluster center. Experimental results on Corel and satellite image databases show the effectiveness of proposed method in improving the accuracy of image database clustering and image annotation.

The future work will investigate the extension of this approach to provide a way for predicting feature weight with a small number of sample feedbacks that is used for initial cluster centers. Experiments have shown that, if the feature weight of each cluster can be determined, clustering results without any feedback will be equivalent to the end of the second feedback session. Therefore, the performance can be improved by predicting feature weight.

REFERENCES

[1] D. Zhang, M.M. Islam, and G. Lu, "A review on automatic image annotation techniques, " Pattern Recognition, Vol. 45 , No. 1 , 2012.

[2] R. Zhang and Z. Zhang, "BALAS: Empirical Bayesian learning in the relevance feedback for image retrieval," Image and Vision Computing, Vol. 24, pp. 211–223, 2006.

[3] M.P.S. Silva , G. Camara, R.C.M. Souza, D.M. Valeriano, and M.I.S. Escada, "Mining Patterns of Change in Remote Sensing Image Databases," *Proc. of the Fifth IEEE International Conference on Data Mining (ICDM'05)*, Nov. 2005.

[4] T. Deselaers, H. Müller, and T. Deserno, "Automatic Medical Image Annotation in Image CLEF 2007: Overview, Results, and Discussion," *Pattern Recognition Letters*, Vol. 29, No. 15, pp. 1988–1995, 2008.

[5] Y. Liua, D. Zhanga, G. Lua, and W. Mab, "A survey of content-based image retrieval with high - level semantics," *Pattern Recognition,* Vol. 40, No. 4, pp. 262-282, 2007.

[6] F. Cao, J. Liang, and L. Bai, "A new initialization method for categorical data clustering," Expert Systems with Applications, Vol. 36, pp.10223–10228, 2009.

[7] J. Lee, S. Hwang, Z. Nie, and J. Wen, "Query Result Clustering for Object-level Search," Proc. of *KDD'09*, Paris, France. pp. 1205-1213, Jun. 2009.

[8] P. Muneesawang, L. Guan, "*Multimedia Database Retrieval: A Human-Centered Approach,*" Signals and Communication Technology, 2006 Springer.

[9] K.Y. Yip, D.W. Cheung, and M. K. Ng., "On discovery of extremely low-dimensional clusters using semi-supervised projected clustering," *Proc. of the 21st International Conference on Data Engineering (ICDE)*, 2005.

[10] W. Dai, G. Xue, Q. Yang, and Y. Yu, "Co-clustering based Classification for Out-of-domain Documents," Proc. of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining KDD'07, San Jose, California, USA, pp. 210-219, Aug. 2007.

[11] A.R. Webb, "Statistical Pattern Recognition, Second Edition," 2002 John Wiley & Sons, Ltd.

[12] C. Yen and K.J. Cios, "Image recognition system based on novel measures of image similarity and cluster validity," Neurocomputing, Vol. 72, pp. 401–412, 2008 .

[13] M.L. Kherfi and D. Ziou, "Relevance Feedback for CBIR: A New Approach Based on Probabilistic Feature Weighting With Positive and Negative Examples," IEEE Transactions on Image Processing, Vol. 15, No. 4, APRIL 2006.

[14] P. Yin, B. Bhanu, K. Chang, and A. Dong, "Integrating Relevance Feedback Techniques for Image Retrieval Using Reinforcement Learning," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 10, 2005.

[15] Y. Rui, T.S. Huang, and S. Mehrotra, "Content-based image retrieval with relevance feedback in MARS," in Proc. IEEE Int. Conf. Image Processing, Santa Barbara, CA, pp. 815–818, 1997.

[16] B. QIU, C.S. XU, and Q. TIAN, "Efficient Relevance Feedback Using Semi-supervised Kernel-specified K-means Clustering," *Proc of the 18th International Conference on Pattern Recognition,* Vol. 03, pp. 316 – 319, 2006.

[17] X.S. Zhou and T.S. Huang, "Relevance Feedback in Image Retrieval: A Comprehensive Review," *Multimedia System,* Vol. 8, No. 6, pp. 536–544, 2003.

[18] Y. Chen, J.Z. Wang, and R. Krovetz, "CLUE: Cluster-Based Retrieval of Images by Unsupervised Learning," IEEE Transactions on Image Processing, Vol. 14, No. 8, 2005.

[19] P. Muneesawang and L. Guan, "An Interactive Approach for CBIR Using a Network of Radial Basis Functions," IEEE Transactions on Multimedia, Vol. 6, No. 5, 2004.

[20] R. Xu, S. Wunsch, and D. Wunsch, "Survey of Clustering Algorithms," IEEE Transactions on Neural Networks, Vol. 16, No. 3, 2005.

[21] D. Yeung and H. Chang, "A Kernel Approach for Semisupervised Metric Learning," IEEE Transactions on Neural Networks, Vol. 18, No. 1, 2007.

[22] J.A.M. Zegarra, N.J. Leite, and R. Torres, "Wavelet-based fingerprint image retrieval," Journal of Computational and Applied Mathematics, Vol. 227, pp. 294–307, 2009.

[23] S. Das, A. Abraham, S. Konar, and A. Konar, "Automatic Clustering Using an Improved Differential Evolution Algorithm," IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans, Vol. 38, No. 1, 2008.

[24] S. Antani, R. Kasturi, and R. Jain, "A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video," Pattern Recognition, Vol. 35, pp. 945–965, 2002.

[25] M.X. Ribeiro, P.H. Bugatti, C. Traina Jr, P.M.A. Marques, N.A. Rosa ,and A.J.M. Traina, "Supporting content-based image retrieval and computer-aided diagnosis systems with association rule-based techniques," Data & Knowledge Engineering, Vol. 68 pp. 1370–1382, 2009.

[26] H. Nezamabadi pour and E. Kabir, "Concept learning by fuzzy k-NN classification and relevance feedback for efficient image retrieval," Expert Systems with Applications, Vol. 36, pp. 5948–5954, 2009.

[27] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods," Pattern Recognition, Vol. 40, pp. 1106–1122, 2007.

[28] A. Broumandnia and J. Shanbehzadeh, "Fast Zernike wavelet moments for Farsi character recognition," *Image and Vision Computing*, Vol. 25, pp. 717–726, 2007.

[29] G. Amayeh, S. Kasaei, G Bebis, A. Tavakkoli, and K. Veropoulos, "Improvement of Zernike Moment Descriptors on Affine Transformed," International Symposium on Signal Processing and its Applications (ISSPA07), United Arab Emirates, February 12-15, 2007.

[30] R.S. choras, "Image Feature Extraction Techniques and Their Application for CBIR and Biometrics Systems," *International Symposium on Signal Processing and its Applications (ISSPA07),* UAE, Feb. 2007.

[31] T. Deselaers, R. Paredes, E. Vidal, and H. Ney, "Learning Weighted Distances for Relevance Feedback in Image Retrieval," proc. In 19th International Conference on Pattern Recognition (ICPR), pp.1-4, 2008.

[32] H. Xiaofei, O. king, W.Y. Ma, L. Mingjing, and H.Z. Zhang, "Learning a Semantiv Space From User's Relevance Feedback fr Image Retrieval," IEEE transactions on circuits and systems for video technology, Vol. 13, No. 1. pp. 39-48, 2003.

[33] B. Bhanu and A. Dong, "Concepts Learning with Fuzzy Clustering and Relevance Feedback," *Engineering Applications of Intelligence,* Vol. 12, pp. 123-138, 2002.

[34] M. Javani, A.M. Eftekhari Moghadam, "Satellite Image Annotation Based on KNN Classifier with CM Similarity Measure," Proc. Of the 6th Iranian machine Vision & Image Processing Conference (27, 28 October), pp. 122-129, 2010.

[35] M. Javani, K. Faez, and D. Aghlmandi, "Clustering and feature selection via PSO algorithm," Proc. of International Conf. of Artificial Intelligent and Signal Processing (AISP), Tehran, Iran, pp. 81-86, 2011.

[36] M. Javani, A.M. Eftekhari Moghadam, "Satellite Image Annotation via Co-Clustering Technique," Proc. Of 16th CSI Conference, 2011.

[37] A.M. Tousch, S. Herbin, and J.Y. Audibert, "Semantic hierarchies for image annotation: A survey," Pattern Recognition, Vol. 45, pp. 333–345, 2012.

Mehran Javani was born in Behbahan, Iran, in 1984. He received the Post Diploma in Computer Software from Shiraz University in 2005, B.E. degree in Computer Software engineering from Islamic Azad University, Behbahan, Iran, in 2007 and M.Sc. degree in Artificial Intelligent from Islamic Azad University, Qazvin, Iran, in 2011. In 2009, he was a lecturer of computer engineering at the Islamic Azad University, Behbahan, Iran. His research interests include image processing, content based image retrieval, data-mining and pattern recognition.

Amir Masoud Eftekhari Moghadam was born in Tehran, Iran in 1964. He received B.E. and M.Sc. degree in computer engineering from Iran University of Science and Technology in 1992 and 1995, respectively and Ph.D. degrees in computer engineering in 2002 from Science and Research branch, Islamic Azad University, Tehran, Iran. In 1996 he was a lecturer of computer engineering, and since 2002 he has been assistant professor at the Qazvin branch, Islamic Azad University, Qazvin, Iran. His research interests include image processing, image retrieval, image/video coding, semantic video analysis, data mining and image security.