OCRS: An Interactive Object-based Image Clustering and Retrieval System

Chengcui Zhang
Department of Computer and Information Sciences,
University of Alabama at Birmingham
Birmingham, AL 35294-1170, U.S.A.
(+1)205-934-8606
zhang@cis.uab.edu

Xin Chen
Department of Computer and Information Sciences,
University of Alabama at Birmingham
Birmingham, AL 35294-1170, U.S.A.
(+1)205-934-8669
chenxin@cis.uab.edu

ABSTRACT
In this paper, we propose an Interactive Object-based Image Clustering and Retrieval System (OCRS). The system incorporates two major modules: Preprocessing and Object-based Image Retrieval. In preprocessing, we use WavSeg to segment images into meaningful semantic regions (image objects). This is an area where a huge number of image regions are involved. Therefore, we propose a Genetic Algorithm based algorithm to cluster these images objects and thus reduce the search space for image retrieval. In learning and retrieval module, Diverse Density is adopted to analyze user's interest and generate the initial hypothesis which provides a prototype for later learning and retrieval. Relevance Feedback technique is incorporated to provide progressive guidance to the learning process. In interacting with user, we propose to use One-Class Support Vector Machine (SVM) to learn user's interest and refine the returned result. Performance is evaluated on a large image database and the effectiveness of our retrieval algorithm is demonstrated through comparative studies.

Categories and Subject Descriptors
H.2.8 [**Database Management**]: Database Applications – *Multimedia Data Mining*.

General Terms
Algorithms, Experimentation.

Keywords
Object-based CBIR.

1. INTRODUCTION
With the rapid increase of various digital image data, image retrieval has drawn attention of many researchers in the computer vision and the database communities. However, the current state-of-art technologies are facing two main problems: 1) semantic gap between low level features and high level concept; 2) curse of dimensionality. This paper aims to build a framework to alleviate the above problems especially for object-based image retrieval.

MDM/KDD 2005 Chicago, August 21, Chicago, Illinois, USA

For "semantic gap" problem, Relevance feedback (RF) technique is widely used to incorporate the user's concept with the learning process [4][6]. As a supervised learning technique, it has been shown to significantly increase the retrieval accuracy. However, most of the existing RF-based approaches consider each image as a whole, which is represented by a vector of N dimensional image features. However, user's query interest is often just one part of the query image i.e. a region in the image that has an obvious semantic meaning. Therefore, rather than viewing each image as a whole, it is more reasonable to view it as a set of semantic regions. In this context, the goal of image retrieval is to find the semantic region(s) of the user's interest. Since each image is composed of several regions and each region can be taken as an instance, region-based CBIR is then transformed into a Multiple Instance Learning (MIL) problem. Maron et al. applied MIL into natural scene image classification [2]. Each image is viewed as a bag of semantic regions (instances). In the scenario of MIL, the labels of individual instances in the training data are not available, instead the bags are labeled. When applied to RF-based image retrieval, this corresponds to the scenario that the user gives feedback on the whole image (bag) although he/she may be only interested in a specific region (instance) of that image. The goal of MIL is to obtain a hypothesis from the training examples that generates labels for unseen bags (images) based on the user's interest in a specific region.

In order to support region-based image retrieval, we need to divide each image into several semantic regions. Instead of viewing each image as a whole, we thus examine region similarity during image retrieval. However, this further increases the search space by a factor of 4~6. Clustering is a process of grouping a set of physical or abstract objects into classes based on some similarity criteria. Given the huge amount of semantic regions in this problem, we first preprocess image regions by dividing them into clusters. In this way the search space can be reduced to a few clusters that are relevant to the query region. K-means is a traditional clustering method and has been widely used in image clustering such as [14] [15]. 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 [9] is known for its robustness and ability to approximate global optimum. In this study, we adapted it to suit our needs of clustering image regions.

After clustering, our proposed system applies Diverse Density (DD) as proposed within the framework of MIL by Maron et al. [2] to learn the region of interest from users' relevance feedback

on the whole image and tells the system to shift its focus of attention to that region. In [1], Zhang et al. further extends Maron's work by incorporating EM (Expectation-Maximization) algorithm in finding the maximum DD point. We adopt Zhang's method because it is less sensitive to the dimension of feature space and scales up well. We take the output of DD as our initial hypothesis of user's interest and continue the relevance feedback with our kernel learning algorithm.

Chen et al. [12] proposed a Support Vector Machine (SVM) based algorithm for Content-based Image Retrieval. In Chen's method, the problem falls into two class classification solved by standard SVM. However, we consider grouping all negative regions into one class somewhat inappropriate. Therefore, after initial analysis of user's interest by DD, our proposed learning algorithm concentrates on those positive images and uses the learned region-of-interest to evaluate all the other images in the image database. The motivation comes from the fact that positive samples are all alike, while negative samples are each bad in their own way. In other words, instead of building models for both positive class and negative class, it makes more sense to assume that all positive regions are in one class while the negative regions are outliers of the positive class. Therefore, we applied One-Class Support Vector Machine (SVM) [3] to solve the MIL problem in CBIR. Chen et al. [7] and Gondra [11] use One-Class SVM in image retrieval but, again, it is applied to the image as a whole. In our approach, One-Class SVM is used to model the non-linear distribution of image regions and to separate positive regions from negative ones. Each region of the test images is given a score by the evaluation function built from the model. The higher the score, the more similar it is to the region of interest. The images with the highest scores are returned to the user as query results. However, the critical issue here is how to transform the traditional SVM learning, in which labeled training instances are readily available, to a MIL learning problem where only the labels of bags (e.g. images with positive/negative feedbacks) are available. In this study, we proposed a method to solve the aforementioned problem and our experiments show that high retrieval accuracy can be achieved usually within 4 iterations.

In Section 2, we present an overview of our OCRS system. The preprocessing module is presented in Section 3 which involves segmentation and clustering. The detailed learning and retrieval approach is discussed in Section 4. In Section 5, system performance evaluation with experimental results is presented. Section 6 concludes the paper.

2. SYSTEM ARCHITECTURE

Figure 1 shows the architecture of our system. In preprocessing module, images are segmented into semantic regions, with each represented by a 19-feature vector. A Genetic Algorithm based clustering method is then implemented to cluster these image segments into clusters so that similar image segments are grouped together.

In initial query, the system first gets the user's query. However, at this point, the system has no clue to the user's interested semantic region. Therefore, a simple Euclidean based similarity comparison is performed to retrieve the initial query results to the user. After the initial query, the user gives feedback to the retrieved images and these feedbacks are examined by Diverse Density trying to analyze user's interest. The output of Diverse Density algorithm is the initial input of One-Class Support Vector Machine (SVM) based algorithm which learns from these feedbacks and starts another round of retrieval. In each round, the refined retrieval result is provided to the user for feedback. One-Class SVM studies these feedbacks and builds a model for future retrieval.

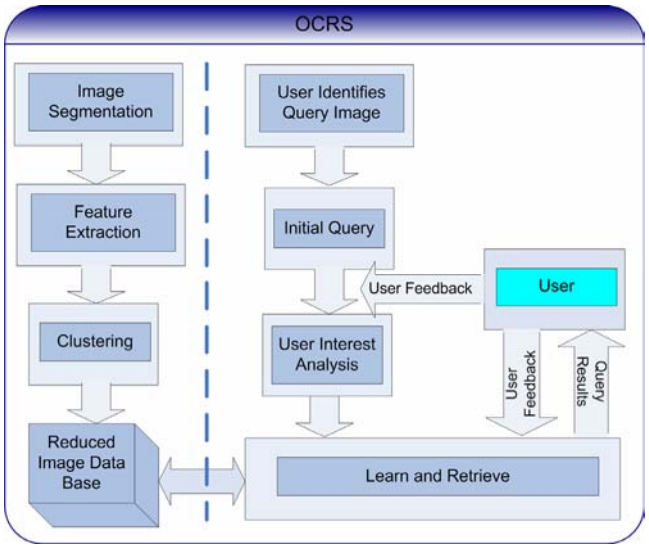

Figure 1. OCRS System Architecture

3. PREPROCESSING

3.1 Segmentation

3.1.1 WavSeg Image Segmentation

Instead of manually dividing each image into many overlapping regions [5], in this study, we propose to use a fast yet effective image segmentation method called WavSeg [13] to partition the images. In Wavseg, a wavelet analysis in concert with the SPCPE algorithm [16] 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 potential regions will become clearly evident. In our experiments, the 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. Actually, even the coarse initial partition generated by wavelet transform is much closer to some global minima in SPCPE than a random initial partition, which means a better initial partition will lead to better segmentation results. In addition, wavelet transform can produce other useful features such as texture features in addition to extracting the region-of-interest within one entry scanning through the image data. Based on our initial testing results, the wavelet based SPCPE segmentation framework (WavSeg) outperforms the random initial partition based SPCPE algorithm in average. It is worth pointing out that WavSeg is fast. The processing time for a 240×384 image is only about 0.33 second in average.

3.1.2 Region Feature Extraction

Both the local color and local texture features are extracted for each image region. For color features, HSV color space and its variants are proven to be particularly amenable to color image analysis. Therefore, we quantize the color space using color categorization based on H S V value ranges. Twelve representative colors are identified. They are black, white, red, red-yellow, yellow, yellow-green, green, green-blue, blue, blue-purple, purple, and purple-red. The Hue is divided into five main color slices and five transition color slices. Each transition color slice such as yellow-green is considered in both adjacent main color slices. We disregard the difference between the bright chromatic colors and the chromatic colors. Each transition color slice is treated as a separate category instead of being combined into both adjacent main color slices. A new category "gray" is added so that there are totally thirteen color features for each image region in our method.

For texture features, one-level wavelet transformation using Daubechies wavelets are used to generate four subbands of the original image. They include the horizontal detail sub-image, the vertical detail sub-image, and the diagonal detail sub-image. For the wavelet coefficients in each of the above three subbands, the mean and variance values are collected respectively. Therefore, totally six texture features are generated for each image region in our method.

The thirteen color features and six texture features of each region are extracted after image segmentation. Thus, for each image, the number of its objects (regions) is equal to the number of regions within that image. Each object has nineteen features.

3.2 Clustering

3.2.1 Overview of Genetic Algorithm

The basic idea of Genetic Algorithm originates from the theory of evolution -- "survival of the fittest". It was formally introduced in the 1970s by John Holland [9]. It is less susceptible to getting "stuck" at local optima. The overview of genetic algorithm is shown in Figure 2.

The algorithm starts with randomly generating the initial population (possible solutions to a real-world problem). In order to be understood in genetic world, the possible solutions to a real world problem are first encoded. Each solution forms a chromosome. A population is a group of chromosomes. From the first generation, these chromosomes are first evaluated. Then they are operated by three genetic operators: *Selection*, *Crossover* and *Mutation* and generate the next generation. The next generation of chromosomes is again evaluated. An objective function is used in evaluation which measures the fitness of each individual solution (chromosome). This accomplishes the evolution of the first generation. Genetic algorithm then starts to run the next generation and goes through the above-mentioned process again until an optimal solution is found.

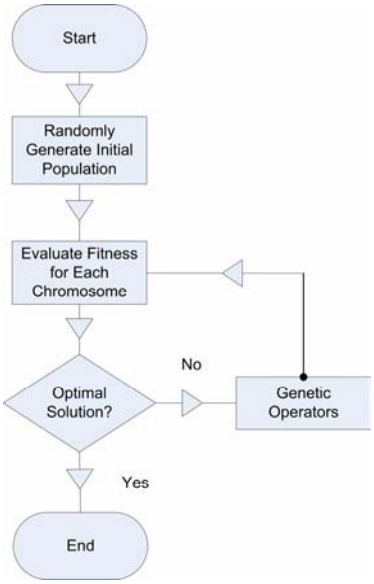

Figure 2. Genetic Algorithm Flow Chart

3.2.2 Genetic Algorithm Design for Image Region Clustering

The objective of image region clustering is to find the optimal combination that minimizes the function below:

$$F(R) = \sum_{j=1}^{k} \sum_{i=1}^{n} d(p_i, rep[p_{i,}, R_j]) \quad (1)$$

p_i is an image region in the cluster R_j which is represented by a representative image region $rep[p_i, R_j]$. n is the total number of image regions and k is the number of clusters. The value of k is determined experimentally as there is no prior knowledge about how many clusters are there. A too large k value would result in over-clustering and increase the number of false negatives, while a too small k value would not help much in reducing the search space. According to our experiment, in which there are 10,000 images with 49,584 regions, we divide the entire set of image regions into 100 clusters since it results in a good balance between accuracy and efficiency. d is some distance measure. In this study, we use the Euclidean distance. Equation 1 is the objective function in our algorithm. The goal is to find its minimum.

In image region clustering, the target is to group semantic image regions into clusters according to their similarities. The above-mentioned representative regions are actually centroids clusters. Therefore, a feasible solution to a clustering problem would be a set of centroids. To encode it, we give each region an ID: 1, 2, …,n (n is an integer). The centroids are represented by their IDs in the chromosome.

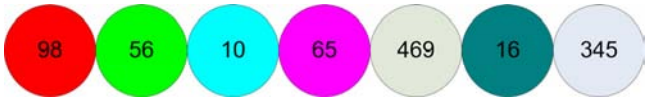

Figure 3. An Example of Chromosome

Figure 3 is an example of a chromosome. In this chromosome, each integer is a gene in genetic world which corresponds to the ID of a centroid image region in the real world.

The initial size of population is set to l which is 50 in this study. For each chromosome we randomly generate k genes, which are actually k integers between 1 and n (the number of image regions). These k genes correspond to the representative image region for each of the k clusters. We then calculate the inverse values of the objective function for these chromosomes: f_1, f_2,…, f_l. The fitness of each individual chromosome is computed according to Equation (2).

$$Fit_i = f_i / \sum_{i=1}^{l} f_i \quad (2)$$

With the first generation, "evolution" begins. In each generation, the whole population goes through three operators: Selection, Recombination and Mutation.

1) *Selection*: There are many kinds of selection operations. We use a Roulette to simulate the selection as shown in Figure 4. For each chromosome we compute its fitness according to Equation (2). Two chromosomes from the population are then randomly selected. The higher the fitness the higher the chance a chromosome is selected. This mechanism is like rotating roulette as shown in Figure 4. C_1, C_2… are chromosomes. The area each chromosome occupies is determined by its fitness. Therefore, chromosomes with higher fitness values would have more chances to be selected in each rotation. We select l pairs of chromosomes and feed them into the next step.
2) *Recombination*: In this step, the recombination operator proposed in [10] is used instead of a simple crossover. Given a pair of chromosomes C_1 and C_2, we use recombination operator to generate their child chromosome C_0 one gene at a time. Each gene in C_0 is either in C_1 or C_2 or both and is not repetitive of other genes in C_0.
3) *Mutation*: In order to obtain high diversity of the genes, a "newly-born" child chromosome may mutate one of its genes to a random integer between 1 and n. However, this mutation is operated at a very low frequency.

At the end of the process, the chromosome with highest fitness in the last population is selected to be a feasible solution. This chromosome is decoded and we have the centroids of clusters as our final output.

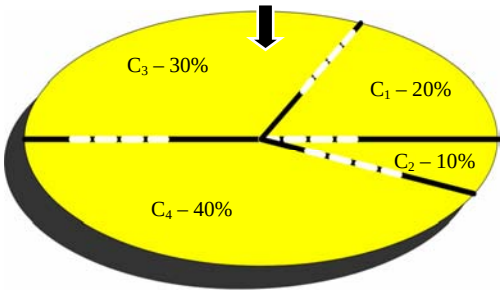

Figure 4. Roulette

4. LEARNING AND RETRIEVAL FRAMEWORK

4.1 User Interest Analysis

In this study, we assume that user is only interested in one semantic region of the query image. The goal is to retrieve those images that contain similar semantic regions. In the initial query, user identifies a query image. At this point, no information is provided as to which specific part of the image is of the user's interest. Therefore, for all the images in the image database, we compute the Euclidean distances between the regions of these images and the regions of the query image. The distance between the query image and an image in the image database is represented by the smallest pair-wise distance of their regions. The "representative" distances are then sorted in an ascending order and the top 30 images are returned to the user for feedback.

The user identifies a returned image as "positive" if it is of his/her interest; otherwise the user labels it "negative". With this information at hand, our next step is to estimate the user's interest i.e. which specific region of the query image user is interested in. We apply Diverse Density (DD) algorithm to accomplish this goal. Diverse Density was first proposed by Maron and Lozano-Pérez in the framework of Multiple Instance Learning [2].

In Multiple Instance Learning (MIL), the label of an individual instance (object) is unknown. Only the label of a set of instances is available, which is called the label of a bag. MIL needs to map an instance to its label according to the information learned from the bag labels. In Content-based Image Retrieval, we have two types of labels – Positive and Negative. Each image is considered a bag of semantic regions (instances). When supplying feedback to the retrieved images, the label of each retrieved image, i.e. bag label, is available. However, the label of each semantic regions in that image bag is still unknown because the user only gives feedback to an image as a whole, not to individual semantic regions in that image. The goal of MIL is to estimate the labels (similarity scores) of the test image regions/instances based on the learned information from the labeled images/bags in the training set.

With Diverse Density approach, an objective function called DD function is defined to measure the co-occurrence of similar instances from different bags (images) with the same label. The target of DD is to find a point which is the closest to all the positive images and farthest from all the negative images. The framework of DD by Maron and Lozano-Pérez [2] is briefly explained below.

We denote the positive bags as B_1^+, B_2^+,... B_n^+ and the negative bags as B_1^-, B_2^-,...B_m^-. The j^{th} instance of bag B_i^+ is represented as B_{ij}^+, while the j^{th} instance of bag B_i^- is written as B_{ij}^-. Each bag may contain any number of instances, but every instance must be represented by a k dimensional vector where k is a constant.

Different semantic concepts may share some common low-level features, but not all k dimensions contribute equally. Therefore, given a semantic concept, greater weights shall be assigned to relevant features as compared to less relevant features. For example, color and texture features shall be assigned greater weights for "grass" compared to shape features. We denote this weight vector $\alpha = \{\alpha_1, \alpha_2, \ldots \alpha_k\}$.

For any point $p = \{p_1, p_2, \ldots p_k\}$ in the feature space, the Diverse Density is defined by the probability of it being our target point, given all the positive and negative bags. So the point we are looking for is the one that maximize the probability below.

$$Arg\max P_r(p| B_1^+, B_2^+, \ldots B_n^+, B_1^-, B_2^-, \ldots B_m^-) \quad (3)$$

Assuming a uniform prior over the concept location $P_r(p)$ and conditional independence of the bags given the target concept p, the above function equals to

$$Arg\max \prod_i P_r(p|B_i^+) \prod_i P_r(p|B_i^-) \quad (4)$$

The following model was introduced by Maron and Lozano-Pérez for estimating hypothesis $h = \{p_1, \ldots p_k, \alpha_1, \ldots \alpha_k\}$.

$$P_r(B_{ij} = p) = exp(-\sum_k \alpha_k (B_{ijk} - p_k)^2) \quad (5)$$

The goal is to find such a hypothesis h such that the above function reaches its maximum. We apply EM (Expectation-Maximization) algorithm as proposed by Zhang et al. [1]. EM starts with an initial hypothesis h, and then repeatedly performs E-step and M-step. In E-step, the current hypothesis h is used to pick one instance from each bag which is most likely to be the one responsible for the label of the bag. In M-step, a two step gradient ascent search (Quasi-Newton search) is performed on DD algorithm to find a new h' that maximizes the above function. In the new iteration h' replaces h. The loop continues until the algorithm converges.

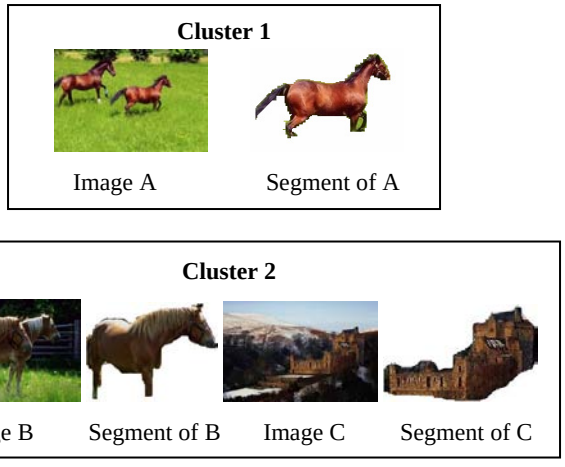

Figure 5. Image Region Clusters

We then use the final result h, the point of the user's interest, to find the cluster this point h belongs to. Hence all the other image regions in this cluster can be located. However, we cannot simply reduce the search space to this cluster alone because it is not a rare case that a particular region is closer to some regions in another cluster than some regions within the same cluster. This situation is illustrated in Figure 5. Suppose the query image region is a "horse" region in Image B. It is in the same cluster with the "castle" region of Image C because these two regions share similar low level features. However, Image B is conceptually closer to Image A whose "horse" region is in another cluster. Therefore, in our system, we choose three clusters whose centroids are the closest to the query region. As an image is composed of several semantic regions, it can fall into any cluster that has at least one of its semantic regions. We then group all the images that have at least one semantic region fall into the three clusters mentioned above and take it as the reduced search space for a given query region. The effectiveness of this reduction is presented in Section 5.

4.2 Learning and Retrieval

As mentioned above, DD algorithm is applied to analyze the user's interest after the initial query. Yet due to the large amount of noise in the image data set, we cannot guarantee that the user's interest is exactly h. Instead, it is taken as our initial hypothesis and the system continues interacting with user to collect more feedbacks. The output of DD is a group of instances, one from each image, that contribute most to the image (bag) label. Specifically, in the output of DD, instances that come from positive bags are positive instances. Because of these instances, their corresponding bags are labeled positive. We then construct the training sample set according to this output of DD. This is then fed into One-Class Support Vector Machine as the initial training sample set, which further learns and models user's interest and refines the retrieval result in the following iterations.

One-Class classification is a kind of unsupervised learning mechanism. It tries to assess whether a test point is likely to belong to the distribution underlying the training data. In our case, the training set is composed of positive samples only. Figure 6 shows how positive image regions are all alike and should be in one class while it is inappropriate to group negative image regions into a single class. In Figure 6, we use images segmented by Blobworld [17] as an example in which image regions are outlined by red lines. Suppose the user's interest is a "horse" object, then ideally positive image regions shall be those "horse" regions. However, negative image regions can be anything other than "horse". As shown in Figure 6, negative image regions can be "flower", "wall", and "sea", etc.

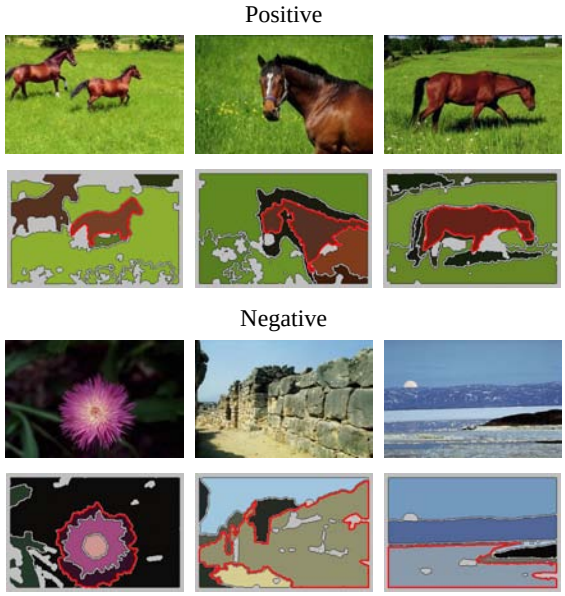

Figure 6. One Class Classifications

One-Class SVM has so far been studied in the context of SVMs [3]. The objective is to create a binary-valued function that is positive in those regions of input space where the data predominantly lies and negative elsewhere. The idea is to model the dense region as a "ball". In our problem, positive instances are inside the "ball" and negative instances are outside. If the origin of the "ball" is $\vec{o}$ and the radius is r, a point $\overrightarrow{p_i}$, in this case an instance (image region) represented by an 19-dimension feature vector, is inside the "ball" *iff* $\left\|\overrightarrow{p_i} - \vec{o}\right\| \le r$. This "ball" is actually a hyper-sphere. The goal is to keep this hyper-sphere as "pure" as possible and include most of the positive objects. Since this involves a non-linear distribution in the original space, the strategy of Schölkopf's One-Class SVM is first to do a mapping θ to transform the data into a feature space F corresponding to the kernel K:

$$\theta(p_1) \cdot \theta(p_2) \equiv K(p_1, p_2) \tag{6}$$

where p_1 and p_2 are two data points. In this study, we choose to use Radial Basis Function (RBF) Machine below.

$$K(p_1, p_2) = \exp\left(\|p_1 - p_2\| / 2\sigma\right) \tag{7}$$

Mathematically, One-Class SVM solves the following quadratic problem:

$$\min_{w,\xi,\rho} \frac{1}{2}\|w\| - u\rho + \frac{1}{n}\sum_{i=1}^{n}\xi_i \tag{8}$$

subject to

$$(w \cdot \theta(p_i)) \ge \rho - \xi_i, \quad \xi_i \ge 0 \text{ and } i = 1,...,n \tag{9}$$

where ξ_i is the slack variable, and $u \in (0,1)$ is a parameter that controls the trade off between maximizing the distance from the origin and containing most of the data in the region created by the hyper-sphere and corresponds to the ratio of "outliers" in the training dataset. If w and ρ are a solution to this problem, then the decision function is $f(x) = sign(w \cdot \theta(p) - \rho)$ and it will be 1 for most examples p_i contained in the training set.

Some images may actually contain more than one positive region. Therefore, we cannot assume that only one region in each image is positive. Suppose the number of positive images is n and the number of all semantic regions in the training set is N. Then the ratio of "outliers" in the training set is set to:

$$u = 1 - (\frac{n}{N} + z) \tag{10}$$

z is a small number used to adjust the u in order to alleviate the above mentioned problem. Our experiment results show that z = 0.01 is a reasonable value.

The training set as well as the parameter u are fed into One-Class SVM to obtain w and ρ, which are used to calculate the value of the decision function for the test data, i.e. all the image regions in the database. Each image region will be assigned a "score" by $w \cdot \theta(p) - \rho$ in the decision function. The higher the score, the more likely this region is in the positive class. The similarity score of each image is then set to the highest score of all its regions.

5. SYSTEM PERFORMANCE EVALUATION

The experiment is conducted on a Corel image database consisting of 10,000 images from 100 categories. After segmentation, there are in total 49,584 image segments. We tested the system performance under different clustering schemes by dividing the entire set of image regions into 40 to 150 clusters. Each time we increase the number of clusters by 10 and find that when the number of clusters k=100, the result is most reasonable in terms of the balance between accuracy and reduction of search space. After the initial query, according to the hypothesis generated by DD, we pull out the three closest clusters as the reduced search space. All the images that have at least one segment fall into these three clusters are identified and fed into the learned One-Class SVM for classification. Sixty images are randomly chosen from 20 categories as the query images. According to our experiment, the search space, in terms of the number of images in the three candidate clusters, is reduced to 13.4% of the original search space (10,000 images) on average.

We compare the performance of our system with two other relevance feedback algorithms: 1) Neural Network based Multiple Instance Learning (MIL) algorithm with relevance feedback [8]; 2) General feature re-weighting algorithm [4] with relevance feedback. For the latter, both Euclidean and Manhattan distances are tested.

Five rounds of relevance feedback are performed for each query image - Initial (no feedback), First, Second, Third, and Fourth. The accuracy rates with different scopes, i.e. the percentage of positive images within the top 6, 12, 18, 24 and 30 retrieved images, are calculated. Figure 7 shows the result after the Fourth Query. "BP" is the Neural Network based MIL which uses both positive and negative examples. "RF_E" is feature re-weighting method with Euclidean Distance while "RF_M" uses Manhattan Distance. "OCRS" is the proposed system.

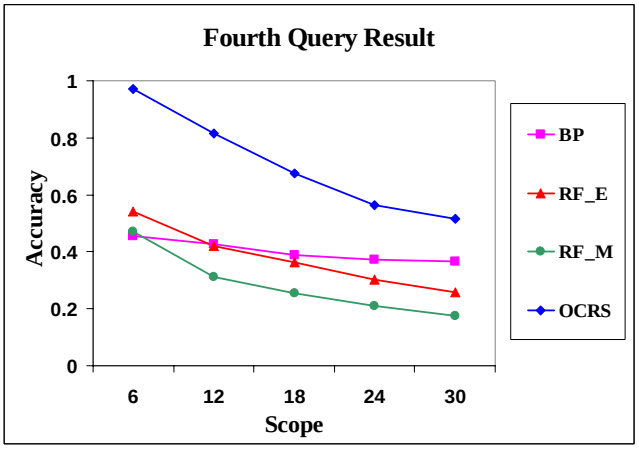

Figure 7. Retrieval Accuracy after the Fourth Query

It can be gleaned from Figure 7 that while the search space is substantially reduced, the accuracy of the proposed framework still outperforms all the other 3 algorithms. It also can be seen that the Neural Network based MIL (BP) shows a better result than

that of general feature re-weighting algorithm after 4 rounds of learning. In addition, the performance of RF_E using Euclidean Distance is slightly better than that of RF_M which uses Manhattan Distance.

Figures 8 and 9 show the first and the fourth Query results of "OCRS", respectively, given the query image on the upper left corner of the interface. In this example, "horse" is the user's interest. It can be seen that only several "horse" images are found correctly initially whereas more "horse" images are found after the fourth round of iteration.

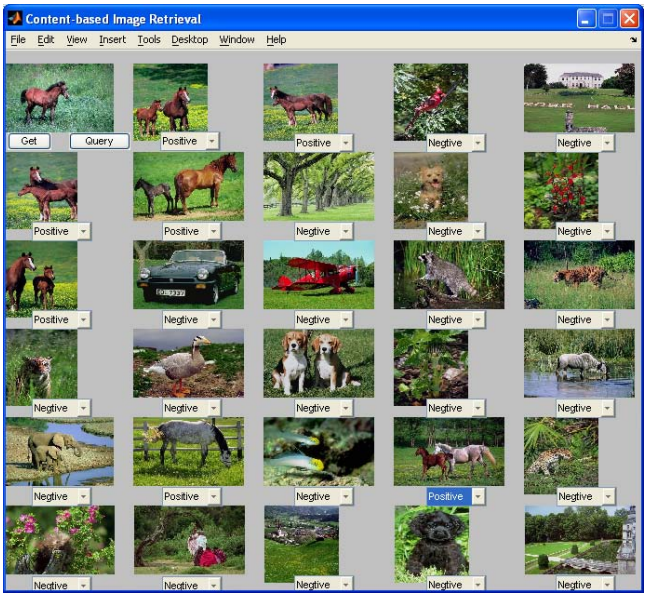

Figure 8. Query Results of OCRS after the First Iteration

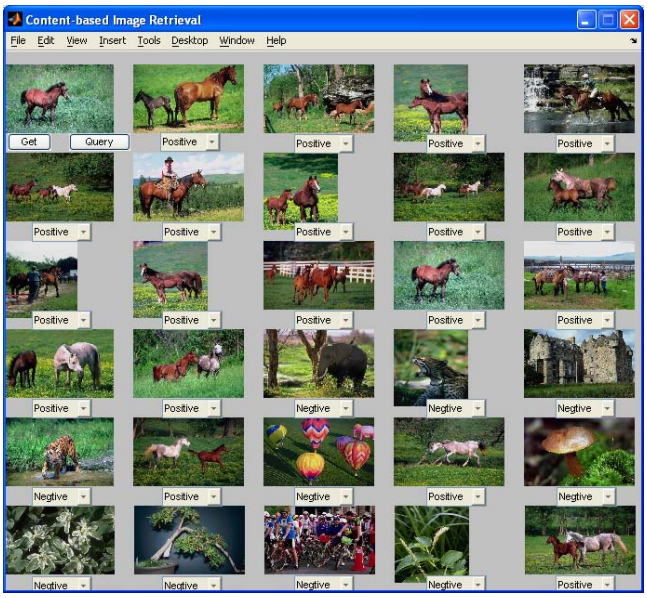

Figure 9. Query Results of OCRS after the Fourth Iteration

It is worth mentioning that, the number of positive images increases steadily through each iteration. Figure 10 gives a concrete view as to how accuracy rates of our algorithm increases across 5 iterations.

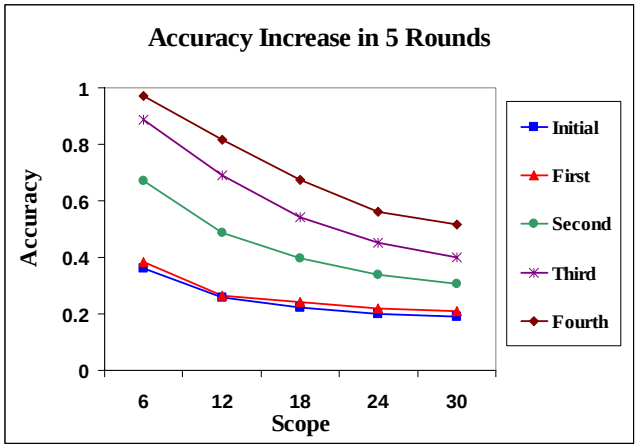

Figure 10. Retrieval Results of OCRS across 5 Iterations

6. CONLUSIONS

In this paper, we proposed a framework, OCRS, for single region based image retrieval. OCRS strives to solve two crucial problems in this area, i.e. time complexity due to the huge amount of high-dimensionality data; semantic gap between low level features and human subjectivity. Specifically in preprocessing, a Genetic Algorithm based clustering mechanism is proposed to reduce the search space. An efficient image segmentation algorithm -- WavSeg is implemented to divide an image into semantic regions. We then adopt Diverse Density to do the initial analysis of user's interest. As initial hypothesis, the output of DD is fed into One-Class SVM in the image retrieval phase. The advantage of our algorithm is that it targets image region retrieval instead of the whole image, which is more reasonable since the user is often interested in only one region in the image. The proposed work also transforms the One-Class SVM learning for region-based image retrieval into a Multiple Instance Learning problem. In addition, due to the robustness of Genetic Algorithm in approximating global optima and the generality of One-Class SVM, the proposed system has proved to be effective in better identifying the user's real need and removing the noise data.

7. ACKNOWLEDGMENTS

The work of Chengcui Zhang was supported in part by SBE-0245090 and the UAB ADVANCE program of the Office for the Advancement of Women in Science and Engineering.

8. REFERENCES

[1] Zhang, Q. and Goldman, S. A. EM-DD: An Improved Multiple-Instance Learning Technique. Advances in Neural Information Processing Systems (NIPS), 2002.

[2] Maron, O. and Lozano-Perez, T.: A Framework for Multiple Instance Learning. Advances in Natural Information Processing System 10. Cambridge, MA, MIT Press, 1998.

[3] Schölkopf, B., Platt, J.C. et al.: Estimating the Support of a High-dimensional Distribution. Microsoft Research Corporation Technical Report MSR-TR-99-87, 1999.

[4] Rui, Y., Huang, T.S., and Mehrotra, S.: Content-based Image Retrieval with Relevance Feedback in MARS. Proceedings of the International Conf. on Image Processing, pp. 815-818, 1997.

[5] Yang, C. and Lozano-Prez, T.: Image Database Retrieval with Multiple-Instance Learning Techniques. Proceedings of the 16th International Conference on Data Engineering, pp. 233-243, 2000.

[6] Su, Z., Zhang, H. J., Li, S., and Ma, S.P.: Relevance Feedback in Content-based Image Retrieval: Bayesian Framework, Feature Subspaces, and Progressing Learning. IEEE Transaction on Image Processing, Vol. 12, No. 8, pp. 924-937, 2003.

[7] Chen, Y., Zhou, X., Tomas, S., and Huang, T.S.: One-Class SVM for Learning in Image Retrieval. Proceedings of IEEE International Conference on Image Processing, 2001.

[8] Huang, X., Chen, S.-C., Shyu, M.-L., and Zhang, C.: User Concept Pattern Discovery Using Relevance Feedback and Multiple Instance Learning for Content-Based Image Retrieval. Proceedings of the 3rd International Workshop on Multimedia Data Mining (MDM/KDD'2002), pp. 100-108, 2002.

[9] Holland, J. H.: Adaptation in Natural and Artificial Systems. University of Michigan Press, 1975.

[10] Vladimir, E. C. and Murray, A. T.: Spatial Clustering for Data Mining with Genetic Algorithms. Technical Report FIT-TR-97-10, Queensland University of Technology, Faculty of Information Management, September 1997.

[11] Gondra, I. and Heisterkamp, D. R.: Adaptive and Efficient Image Retrieval with One-Class Support Vector Machines for Inter-Query Learning. WSEAS Transactions on Circuits and Systems, Vol. 3, No. 2, pp. 324-329, April 2004.

[12] Chen, Y.X., and Wang, J. Z.: Image Categorization by Learning and Reasoning with Regions. Journal of Machine Learning Research, vol. 5, pp. 913-939, 2004.

[13] Zhang, C., Chen, S.-C., Shyu, M.-L., and Peeta, S.: Adaptive Background Learning for Vehicle Detection and Spatio-Temporal Tracking. Proceedings of the 4th IEEE Pacific-Rim Conference on Multimedia, pp. 1-5, 2003.

[14] Kanungo, T., Mount, D., Netanyahu, N., Piatko, C. D., Silverman, R., and Wu, A.Y.: An Efficient K-Means Clustering Algorithm: Analysis and Implementation. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No.7, July 2002.

[15] Wang, L., Liu, L., and Khan, L.: Automatic Image Annotation and Retrieval Using Subspace Clustering Algorithm. Proceedings of the Second ACM International Workshop on Multimedia Databases, pp. 100-108, 2004.

[16] Chen, S.-C., Sista, S., Shyu, M.-L., and Kashyap, R. L.: An Indexing and Searching Structure for Multimedia Database Systems. Proceedings of the IS&T/SPIE Conference on Storage and Retrieval for Media Databases, pp. 262-270, 2000.

[17] Carson, C., Belongie, S., Greenspan, H., and Malik, J.: Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 24, No.8, 2002.