Support Vector Machine Learning for Region-Based Image Retrieval with Relevance Feedback

Deok-Hwan Kim, Jae-Won Song, Ju-Hong Lee, and Bum-Ghi Choi

ABSTRACT—We present a relevance feedback approach based on multi-class support vector machine (SVM) learning and cluster-merging which can significantly improve the retrieval performance in region-based image retrieval. Semantically relevant images may exhibit various visual characteristics and may be scattered in several classes in the feature space due to the semantic gap between low-level features and high-level semantics in the user's mind. To find the semantic classes through relevance feedback, the proposed method reduces the burden of completely re-clustering the classes at iterations and classifies multiple classes. Experimental results show that the proposed method is more effective and efficient than the two-class SVM and multi-class relevance feedback methods.

Keywords—Support vector machine, cluster-merging, relevance feedback, region-based image retrieval.

I. Introduction

In most early approaches to content-based image retrieval (CBIR), images were represented by a set of global features, and retrieval was performed based on similarities in the feature space. However, the retrieval accuracy remains far from users' expectations. This is primarily due to the large gap between high-level semantic concepts and low-level visual features. To reduce this gap, region-based features and relevance feedback technologies have been widely used.

Region-based image retrieval performs more meaningful searches which are closer to the user's perception of an image's content. Instead of looking at an image as a whole, the objects in the image and their relationships are viewed. The similarity between the images perceived by humans does not necessarily correlate with the similarity between them in the feature space. That is, semantically relevant images may exhibit very different visual characteristics, and may be scattered across several classes rather than existing in one class. From this viewpoint, the problem of reducing the semantic gap becomes one of finding semantically related classes.

Recently, relevance feedback (RF)-based classification has become popular among various CBIR techniques, and due to its good generalization ability, RF schemes based on support vector machine (SVM) learning methods have been applied to improve the retrieval performance in CBIR systems using feature representation. Zhang and others proposed a classifier based on SVM learned from the training data of relevant images and irrelevant images marked by users. This model can be used to find more relevant images from the entire database [1]. Jing and others applied a two-class SVM classifier using a generalization of the Gaussian kernel to RF for RBIR using variable length representations [2].

Two-class RF is often inadequate to provide sufficient information rapidly enough to improve retrieval performance. When the instances of positive feedback are few, the performance becomes poor. This is primarily due to two reasons. First, the SVM classifier is unstable when the size of the training set is small. Second, there are usually many more negative feedback samples than positive ones in the RF process.

In order to overcome these problems, we apply a one-versus-one SVM to the multi-classes learning of regions in the relevant set. The learning results are further used to help automatically classify multiple relevant image classes for relevance feedback. Although the number of relevant images is

small, the SVM classifier is stable since it uses a large number of regions in the relevant set.

Alternatively, Peng proposed an adaptive multi-class relevance feedback (MURF) approach using a Chi-square analysis for image retrieval [3]. It uses a metric minimizing the Bayes error criterion. In comparison with the MURF algorithm, the multi-class SVM approach is considered a good alternative because it has a high generalization performance without the need to add *a priori* knowledge, even when the feature dimension is very large. To achieve good generalization, it minimizes the combination of the empirical risk and the Vapnik-Cheronenkis (VC) dimension, while MURF only minimizes the empirical risk.

The proposed method has two main advantages. First, it reduces the burden of completely re-clustering the classes by using a multi-class SVM model for each iteration. Second, it can model the non-linear distributions of image regions, even though the regions of relevant images are scattered in several non-linearly separated classes.

II. Proposed Approach Using Multi-class SVM

Figure 1 shows the proposed RBIR with the RF approach using multi-class SVM learning.

1. Retrieval Phase

An example image submitted by the user is parsed to generate an initial query $Q=(q, d, w, k)$. The earth mover's distance (EMD) function d is used to measure the distance between two images. According to d, the result set consisting of the k images closest to the query point q is returned to the user.

2. User Feedback Phase

The user evaluates the relevance of images in $Result(Q)$ by distinguishing each of them as relevant or irrelevant. Based on these evaluations, a set of relevant images is obtained. The set includes newly added relevant images and relevant images from previous iterations.

3. Multi-class SVM Learning Phase

In the initial iteration, the initial classes of the regions in the relevant set form the basis of the hierarchy and the hierarchical clustering algorithm is adopted to group the regions into a few classes, each of which corresponds to a new pseudo-region for the next query. Level g at the hierarchy of classes corresponds to g classes. Hotelling's T^2 tests whether the locations of the two classes are equal. For C_i and C_j classes with $i \neq j$, it is defined by

$$T^2 = \frac{n_i n_j}{n_i + n_j}(\overline{x}_i - \overline{x}_j)'S_{pooled}^{-1}(\overline{x}_i - \overline{x}_j),$$

where $S_{pooled}=(S_i+S_j)/(n_i+n_j-2)$, and the two classes are characterized by the mean vector $\overline{x}, \overline{y} \in \Re^n$, covariance matrix S_i, S_j, and the number of elements in the class, n_i, n_j, respectively. To estimate the number of optimal classes, it is necessary to decide which level is more optimal than the others. At the g-th classifying level, $\binom{g}{2}$ of Hotelling's T^2s are used to decide which pair of classes is to be merged. If no significant merging occurs, then g is closer to optimal than $(g-1)$; otherwise, $(g-1)$ is closer to optimal.

Once the initial classes are obtained, the multi-class SVMs are trained with all the examples in these classes. In our experiments, the one-versus-one strategy was applied to the multi-class SVM learning. Let D denote a set of given m training data and $D=\{(x_1, y_1), \cdots, (x_m, y_m)\}$, where $x_i \in \Re^n$, $i=1, \cdots, m$ and $y_i \in \{1, \cdots, g\}$ is the class of x_i. It trains $g(g-1)/2$ pairwise binary SVM classifiers. For training data from the i-th and j-th classes, an (i, j)th pairwise decision function is obtained by

minimizing $\quad \frac{1}{2}(w_{i,j})^T w_{i,j} + C\sum_{l=1}^{m}\xi_{i,j,l}$

subject to $\quad (w_{i,j})^T\phi(x_l) + b_{i,j} \geq 1 - \xi_{i,j,l}, \qquad$ for $\quad y_l = i,$
$\quad (w_{i,j})^T\phi(x_l) + b_{i,j} \leq \xi_{i,j,l} - 1 \qquad$ for $\quad y_l = j,$

where $\xi_{i,j,l}$ is the slack variable, C is the penalty parameter, $\phi(\cdot)$ is the nonlinear mapping from the input space $\Re^n$ to a feature space F, and $b_{i,j}$ is a bias.

4. SVM Modeling Phase

For the remaining iteration, the SVM modeling phase is performed and consists of the region-based classification process and the region-merging process. As more relevant images become available, the number of regions in the query increases rapidly. The time required to calculate the EMD distance value between the query and an image is proportional
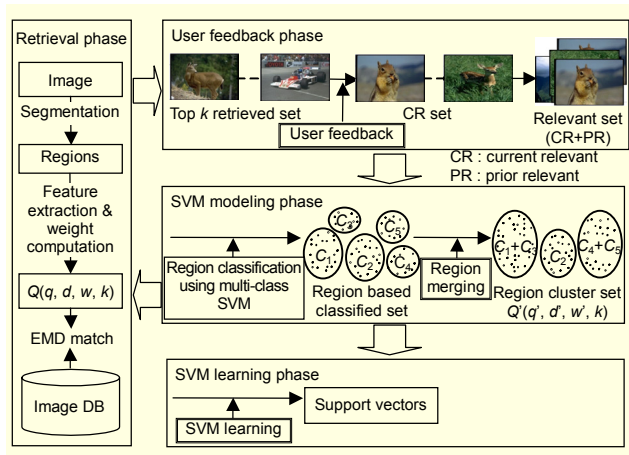

Fig. 1. RBIR with RF using multi-class SVM learning.

to the number of regions in the query. To reduce the retrieval time of the system, multi-class SVM classifiers are used to assign each region of the relevant images to one of the g classes as shown in Fig. 2. If the (i,j)th decision function $sign\,(w^T_{i,j}\,\phi\,(x_l) + b_{i,j})$ says x_l is in the i-th class, then the vote for the i-th class is increased by one; otherwise, the j-th class is increased by one. It is predicted that x_l is in the class with the largest vote. The classes that remain after the classification stage can be merged further into bigger clusters.

Given g classes, the region-merging process determines the number of classes and merges certain classes at the same level to reduce the number of query points in the next iteration. Finally, a new query $Q'=(q', d', w', k)$ is computed and used as an input for the second round.
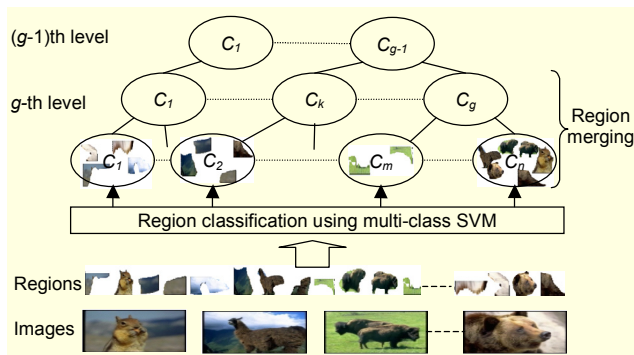

Fig. 2. Multi-class SVM modeling.

III. Experiments and Evaluations

In the experiments, the performance between RBIR using the multi-class SVM, RBIR using hierarchical clustering [4], RBIR using the two-class SVM [2], and MURF [3] were evaluated.

The presented algorithm was tested with approximately 10,000 general purpose color images from the COREL image database. One hundred random initial query images were generated from ten selected categories like sunsets, shores, animals, airplanes, birds, trees, flowers, cars, people, and fruit. High level category information was used as the ground truth to obtain the relevance feedback.

As shown in Fig. 3, RBIR using multi-class SVM yields better performance after one iteration, and its average recall after five iterations is approximately 22.7% higher than that of RBIR using hierarchical clustering, 12.2% higher than that of the two-class SVM, and 3.28% higher than that of MURF. The retrieval time of the multi-class SVM method is half the time of the hierarchical clustering method since the hierarchical clustering requires excessive time to determine the optimal class level. The retrieval time of MURF is similar to that of multi-class SVM, while that of two-class SVM is eight times longer than that of multi-class SVM.
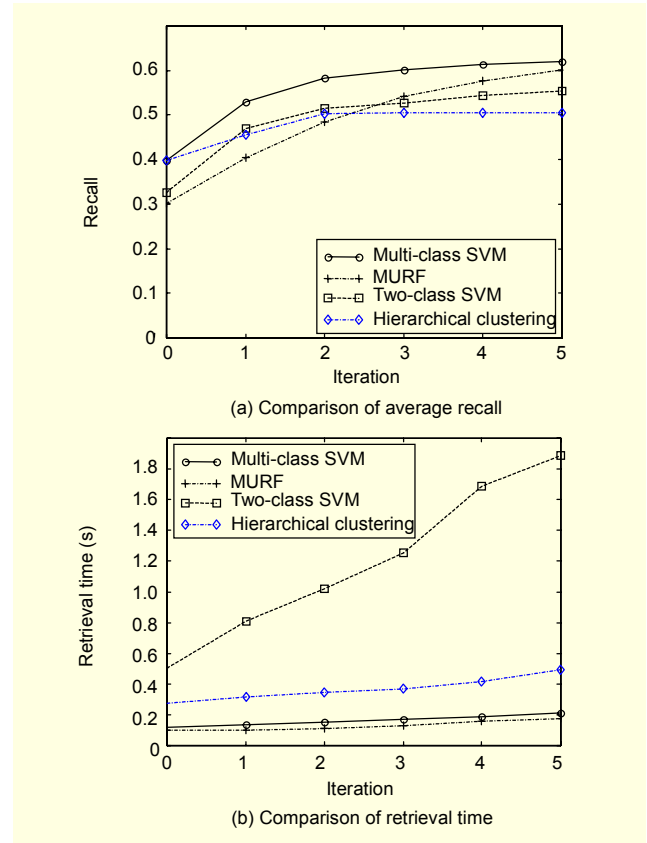

(a) Comparison of average recall

(b) Comparison of retrieval time

Fig. 3. Performance evaluation of relevance feedback with the multi-class SVM method.

IV. Conclusion

This study contributes to the integration of the multi-class SVM learning methods into relevance feedback with adaptive clustering. The SVM classification reduces the burden of completely re-clustering the classes, which exhibit the visual characteristics of semantically relevant images. It can also be incorporated into any RBIR system.

References

[1] L. Zhang, F. Lin, and B. Zhang, "Support Vector Machine Learning for Image Retrieval," *Proc. IEEE ICIP*, 2001.

[2] F. Jing, M. Li, H.-J. Zhang, and B. Zhang, "Support Vector Machines for Region-Based Image Retrieval," *Proc. Int'l Conf. on Multimedia and Expo*, vol. 1, 2003, pp. 21-24.

[3] J. Peng, "Multi-Class Relevance Feedback Content-Based Image Retrieval," *Computer Vision and Image Understanding*, vol. 90, no. 1, 2003, pp. 42-67.

[4] D.-H. Kim and S.-L.Lee, "Relevance Feedback Using Adaptive Clustering for Region Based Image Similarity Retrieval," *LNAI 4099*, 2006, pp. 641-650.