Multitraining support vector machine for image retrieval

Journal Article

Multitraining Support Vector Machine for Image Retrieval

Jing Li, Nigel Allinson, *Member, IEEE*, Dacheng Tao, and Xuelong Li, *Member, IEEE*

***Abstract*—Relevance feedback (RF) schemes based on support vector machines (SVMs) have been widely used in content-based image retrieval (CBIR). However, the performance of SVM-based RF approaches is often poor when the number of labeled feedback samples is small. This is mainly due to 1) the SVM classifier being unstable for small-size training sets because its optimal hyper plane is too sensitive to the training examples; and 2) the kernel method being ineffective because the feature dimension is much greater than the size of the training samples. In this paper, we develop a new machine learning technique, multitraining SVM (MTSVM), which combines the merits of the cotraining technique and a random sampling method in the feature space. Based on the proposed MTSVM algorithm, the above two problems can be mitigated. Experiments are carried out on a large image set of some 20 000 images, and the preliminary results demonstrate that the developed method consistently improves the performance over conventional SVM-based RFs in terms of precision and standard deviation, which are used to evaluate the effectiveness and robustness of a RF algorithm, respectively.**

***Index Terms*—Content-based image retrieval (CBIR), multitraining SVM (MTSVM), relevance feedback (RF), support vector machine (SVM).**

I. Introduction

Recently, content-based image retrieval (CBIR) [5], [10], [11] has become an important research direction in the multimedia information processing field because of the rapidly increasing requirements in many practical application areas, such as architectural design, museum management, education and fabric design. Due to the semantic gap [10] between low-level visual features and high-level semantic information, the retrieval performances of conventional CBIR schemes are still not good enough for many practical applications. Different users at different times may have different viewpoints on or understanding of the same image. In these situations, interactions between the user and the search engine are introduced to improve the retrieval performance. Therefore, relevance feedback (RF) [10] has been proposed as an effective solution for this kind of interaction. Current RF schemes can exhibit some general limitations of over sensitivity to subjective labeling by users and the inability to accumulate knowledge over different sessions and users [15], [16]. The traditional process of RF is as follows: 1) from the retrieved results, the user labels a number of relevant samples as positive feedbacks, and also labels a number of irrelevant samples as negative feedbacks; 2) to obtain improved retrieval results, the CBIR system then refines its retrieval procedure based on these labeled feedback samples. These two steps can be carried out repeatedly. As a result, the performance of the CBIR system can be improved by means of the system gradually learning the user's preferences.

Several RF methods have been developed in recent years, for example [10] adjusts the weights of various features to adapt to the user's preference, while [3] estimates the density of the positive feedback samples. Discriminant learning has also been employed for feature selection in RFs [18]. However, all of these methods have some limitations; for example, the method in [10] is only heuristically based, the density estimation method in [3] ignores any information contained in the negative feedback samples, and the discriminant learning in [18] often suffers from the so-called matrix singular problem or the curse of dimensionality. More recently, classification-based RFs [5], [7], [17] have become a popular technique in CBIR and the support vector machine (SVM)-based RF approach [2], [14] has shown promising results due to the SVM's good generalization ability for classification. However, SVM-based RF algorithms have their own drawbacks.

1) The SVM classifier is unstable for a small size training set because the optimal hyperplane, determined by the support vectors, can be very sensitive to the training examples. In CBIR RF, the number of feedback samples is usually small, because the user is reluctant to mark many samples. Consequently, the performance of the system may be poor due to insufficient and inexactly labeled samples; and,
2) The kernel method cannot exert its normal capability because the feature dimensions are much higher than the number of training samples. In this circumstance, the kernel machine can achieve a zero training error but poor generalization.

In this paper, which aims to address these issues, we propose a multitraining SVM (MTSVM) mechanism for a CBIR RF. MTSVM is based on the observation that 1) the success of the cotraining model [1] to augment labeled examples with unlabeled examples in information retrieval; and 2) advances in the random subspace method [4] to overcome the small sample size problem. With the incorporation of the SVM and the multitraining model, unlabeled examples can generate new informative training examples for which the predicted labels become more accurate. Ttherefore, the new MTSVM method can work well in practical situations. In the MTSVM learning model, we choose the majority voting rule (MVR) [8] as the similarity measure in combining individual classifiers since every single classifier has its own distinctive ability to classify relevant and irrelevant samples.

This paper is organized as follows. A brief review of the traditional SVM-based RF is given in Section II. The newly developed MTSVM is then described in Section III. In Section IV, we introduce our CBIR platform with embedded RF. Section V describes preliminary experimental results based on a large scale image database, and, finally, Section VI is the conclusion.

II. Support Vector Machine-Based Relevance Feedback

The SVM [2], [14] is a very effective binary classifier. Consider a linearly separable binary classification problem

$$\{(x_i, y_i)\}_{i=1}^{N} \quad \text{and} \quad y_i = \{+1, -1\} \tag{1}$$

where x_i is an n-dimension vector and y_i is the label of this vector. SVM separates these two classes of points by a hyperplane

$$\mathbf{w}^T x + b = 0 \tag{2}$$

where $\mathbf{x}$ is an input vector, $\mathbf{w}$ is an adaptive weight vector, and b is a bias. SVM finds the parameters $\mathbf{w}$ and b for the optimal hyperplane to maximize the geometric margin $2/\|\mathbf{w}\|$, subject to

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq +1. \tag{3}$$

Manuscript received June 14, 2005; revised April 12, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Eli Saber.

J. Li and N. Allinson are with the Department of Electronic and Electrical Engineering, University of Sheffield, Sheffield S1 3JD, U.K. (e-mail: jing.li.2003@gmail.com; n.allinson@sheffield.ac.uk).

D. Tao and X. Li are with the School of Computer Science and Information Systems, Birkbeck, University of London, London WC1E 7HX, U.K. (e-mail: dacheng@dcs.bbk.ac.uk; xuelong@dcs.bbk.ac.uk).

Digital Object Identifier 10.1109/TIP.2006.881938

TABLE I
WEAK CLASSIFIER UPDATE

Input: weak classifier h_i, training samples x_i, training labels L_i, unlabelled training set U, π^+ is the minimum acceptance threshold for the positive label, and π^- is the maximum acceptance threshold for the negative label.

Output: the updated weak classifiers h_i.

```
Use L_i to train a classifier h_i that considers only the x_i portion of x, 1 ≤ i ≤ N_h.
Allow h_i to label p positive and n negative examples from U, 1 ≤ i ≤ N_h.
FOR loop_classifier (i) = 1 TO N_h {
    FOR loop_unlabbled_example (j) = 1 TO |U| {
        IF Σ_{1≤l≤N_h, l≠i} h_l(u_j) ≥ π+ {
            Add u_i to L_i as a positive sample;
        }
        ELSE IF Σ_{1≤l≤N_h, l≠i} h_l(u_j) ≤ π- {
            Add u_i to L_i as a negative sample;
        }
        ELSE {
            Reject the sample;
        }
    }
}
Retrain the weak classifiers h_i according to the updated L_i, 1 ≤ i ≤ N_h.
```

TABLE II
MULTITRAINING SVM (MTSVM)

Input: feature set $\mathbf{F}$, weak SVM classifier h_j, integer N_h (number of generated classifiers), labelled examples set L, unlabelled examples set U (top $|U|$ retrieved images, after the current feedback), integer N_{MTSVM} (rounds of multi-training SVM), and image database D.

Output: a series of SVM classifiers h_j and the strong classifier $C(z) = \text{sign}\left(\sum_{1\le j\le N_h} h_j(z)\right)$.

```
Create L_i = L, 1 ≤ i ≤ N_h.
FOR loop_observation (j) = 1 TO N_h {
    Random sampling without replacement on F to generate the observation
    x_j for weak SVM classifier h_j;
}
FOR loop_MTSVM (i) = 1 TO N_MTSVM {
    Call the routine Weak Classifier Update and obtain the retrained classifiers
    h_j;
    Use majority voting rule to generate the classifier C(z) = sign(Σ_{1≤j≤N_h} h_j(z));
    Using the dissimilarity measure function f to resort images in D|(L,U);
    where D|(L,U) means D without L and U.
    Select top |U| retrieved images as the unlabelled examples to replace the
    existing unlabelled examples;
}
```

The solution can be obtained using the Wolfe dual problem with a Lagrangian-multiplier α_i

$$Q(\alpha) = \sum_{i=1}^{m} \alpha_i - \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) \Big/ 2 \tag{4}$$

subject to $\alpha_i \geq 0$ and $\sum_{i=1}^{m} \alpha_i y_i = 0$.

In the dual format, the data points only appear in the inner product, and to achieve a potentially better representation of the data, the data points are mapped onto the Hilbert Inner Product Space by means of the replacement

$$\mathbf{x}_i \cdot \mathbf{x}_j \rightarrow \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j) \tag{5}$$

where $K(\cdot)$ is a kernel function. We then get the kernel version of the Wolfe dual problem

$$Q(\alpha) = \sum_{i=1}^{m} \alpha_i - \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i \cdot \mathbf{x}_j) \Big/ 2. \tag{6}$$

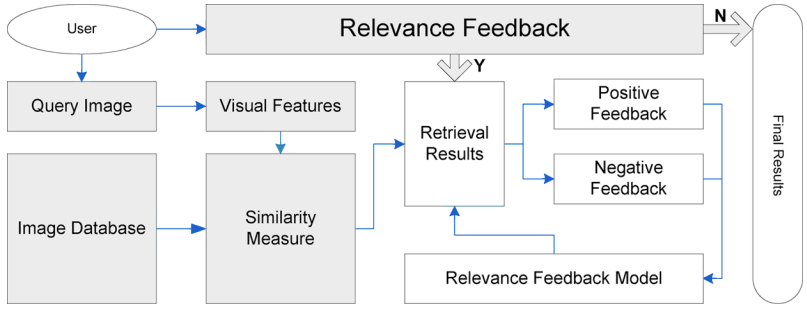

Fig. 1. Flowchart of the image retrieval system. (Color version available online at http://ieeexplore.ieee.org.)

Thus, for a given kernel function, the SVM classifier is given as

$$F(\mathbf{x}) = \mathrm{sgn}(f(\mathbf{x})) \tag{7}$$

where $f(\mathbf{x}) = \sum_{i=1}^{l} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b$ is the output hyperplane decision function of SVM.

In general, when $|f(\mathbf{x})|$ is high for a given pattern, the corresponding prediction confidence will also be high. On the contrary, a low $|f(\mathbf{x})|$ for a given pattern means this pattern is close to the decision boundary and its corresponding prediction confidence is low. Consequently, in our application the output of SVM $f(\mathbf{x})$ is used to measure the dissimilarity [17], [5] between a given pattern and a query image.

III. Multitraining Support Vector Machine

As described in the Introduction, traditional SVM-based RFs often exhibit some intrinsic problems. To address these problems, we propose a new learning scheme to enhance the SVM, which is named the multitraining SVM (MTSVM). The MTSVM has the benefits of the cotraining technique [1] and the random subspace method [4]. Before describing the proposed MTSVM method, we first introduce the cotraining algorithm and the random subspace method as essential background.

A. Cotraining Algorithm

As described in [1], the cotraining algorithm has been proposed as an approach to construct a strong classifier that is trained on a small number of labeled and unlabeled samples. The basis of the cotraining algorithm is that one object (sample) has several different aspects; for example, a biometric classifier could be based on facial, gait, and fingerprint data. For two different aspects of a sample under consideration, two subclassifiers can be trained individually on the given small labeled sample sets. With a well-designed methodology, these initial subclassifiers can gradually gain benefits from the unlabeled samples—each subclassifier independently labels several unlabeled samples and enlarges the training set by including these new samples. After several iterations, the subclassifiers are improved and can be combined as a single one, which can outperform the one obtained by merely directly combining the initial subclassifiers. Therefore, the cotraining algorithm can be very effective for small quantities of labeled data.

Using the cotraining algorithm directly with a SVM is not realistic. This is chiefly because the cotraining algorithm requires that the initial subclassifiers to have been adequately trained, i.e., each subclassifier has already a good generalization ability before the cotraining procedure commences. However, in the CBIR RF procedure, this precondition for subclassifiers cannot be guaranteed. Consequently, we need to enhance the generalization ability of each subclassifier by using classifier committee learning (i.e., the random sampling without replacement in the feature space). With an enhanced ensemble classifier, we can enlarge the training set more accurately by using unlabeled samples.

B. Random Subspace Method

The random subspace method [4], an example of a random sampling algorithm, incorporates the benefits of bootstrapping and aggregation. Multiple classifiers can be generated by training on multiple sets of features that are produced by bootstrapping, i.e., random sampling with replacement on the training features. Aggregation of the generated classifiers can then be implemented by the MVR or other multiple classifiers combination rules. For SVM-based RFs, overfitting is encountered when the training set is relatively small compared to the high dimensionality of the feature vectors. In order to avoid this over-fitting issue, we sample a small subset of features to reduce the discrepancy between the size of the training data size and the length of the feature vector. Exploiting this feature sampling step, we can make the kernel method operate satisfactorily. However, we cannot utilize the random subspace method directly because the cotraining algorithm requires that the different subclassifiers should only be weakly related. Consequently, we randomly select the subset features without replacement in our new algorithm to meet our requirements for multitraining.

In the next part of this section, we develop the multitraining scheme for SVM, which inherits the merits of the existing cotraining technique and the new random subspace method without replacement step.

C. Multitraining Support Vector Machines

The utilitarian cost can be very high in acquiring a large set of labeled examples so we propose a semi-supervised learning method, which can effectively improve the learning performance by introducing the unlabeled data to augment the labeled data. A major concern for our scheme is having multiple views of a sample that are redundant but weakly correlated. During the learning procedure, MTSVM assigns strong labels to the unlabeled examples that are unambiguous. Similar to the cotraining scheme, a key property of MTSVM is that several examples, which may be confidently labeled by one subclassifier, can be misclassified by the other subclassifiers. The subclassifiers can, therefore, train each other by providing additional informative labeled examples. The different elements of the features are almost independent since they describe different parts of an object. So, we can incorporate the cotraining and the random subspace method for the learning procedure to further improve the retrieval or classification performance.

The proposed MTSVM algorithm has two stages: the first is the weak classifier update, which is outlined in Table I, and the second, the main body of MTSVM, is outlined in Table II. The weak subclassifier update stage is modified from the cotraining method described in [1] because cotraining will not reject any samples, even though some samples could be labeled by the weak classifiers with low confidence. In the multitraining scheme, we use the classifier committee learning with a threshold to reject unlabeled samples with low classification confidence. Rejecting the low confidence unlabeled samples can ensure the retrained subclassifiers have a higher confidence without further rejection.

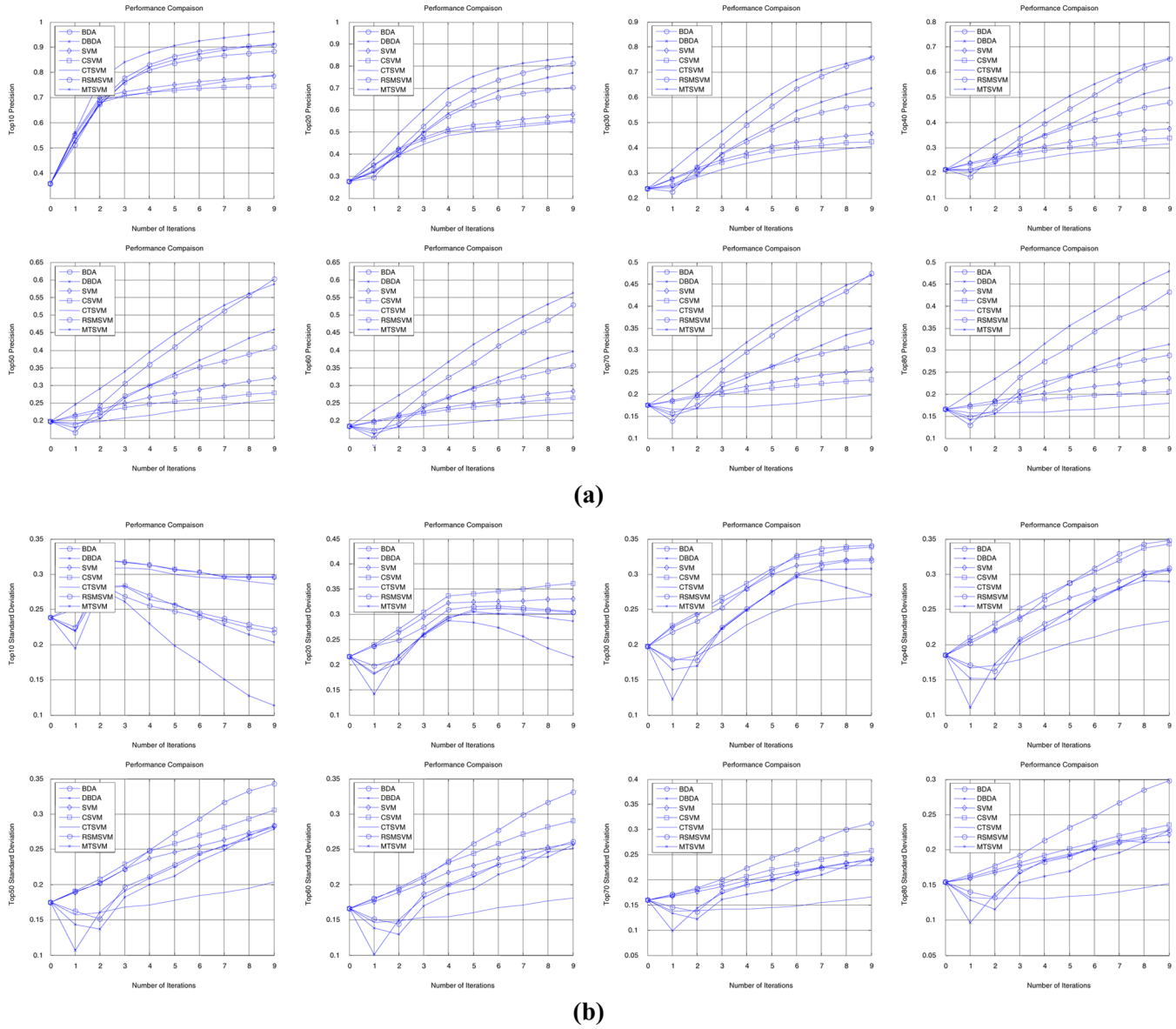

Fig. 2. Relative RF algorithms performance, up to nine iterations, for 500 randomly selected queries. (a) Precision recorded as percentage of returned relevant images in the top-ranked N images ($N = 10, 20, 30, 40, 50, 60, 70, 80$). (b) Corresponding SD curves. The "0" iteration point refers to retrieval based soley on an Euclidean similarity measure and no RF procedure applied. (Color version available online at http://ieeexplore.ieee.org.)

As dissimilarity measure in RF, we first utilize the MVR to identify a given sample as a query relevant or irrelevant sample then we measure the dissimilarity between the sample and the query using an individual SVM classifier, which provides a label with the largest confidence (the absolute value of the corresponding SVM output).

D. RF Requirements and MTSVM Cotraining

The effectiveness of SVM for RF in CBIR systems has been demonstrated previously [3], [5], [7], [17] and the modified approach presented here takes full account of the initially weak classification rules, which are strengthened gradually through repeated user feedback. The different low-level image features are normally independent, e.g., texture and color histograms of an image have no intrinsic linkage and are often of different dimensionality. Our proposed algorithm uses the approach of random sampling without replacement to meet the general requirement of an independent distribution. Finally, in RF, users do not like to label a large number of feedback samples [10], and with these samples, we demonstrate that we can form initial weak classifiers based on different features.

IV. Image Retrieval System

To evaluate the performance of the proposed method, a general-purpose CBIR system was developed, in which any RF algorithm can be embedded. For this system (Fig. 1), when a query image is input, the low-level features are extracted. All images in the database are then sorted based on a specified similarity metric. The user labels some highly ranked images as positive and negative samples. Using these feedback samples, a RF model is trained and the similarity metric is updated based on the output of the RF. All images are resorted using the updated similarity metric, and the procedure is executed repeatedly until the user is satisfied with the outcome.

For our demonstrator system, three main features—color, texture, and shape—are extracted to represent the images. For the color feature, we use the color histogram [12] in the HSV (8:8:4 bits) color

space. Texture is extracted from the Y component of the YCbCr space by the pyramid wavelet transform (PWT) using the Haar wavelet; the mean value and the standard deviation are calculated for each subband at each decomposition level with a resulting feature length of $2 \times 4 \times 3$. For the shape feature, the edge histogram [9] is calculated from the Y component of the YCbCr space; edges are grouped into four categories—horizontal, 45 diagonal, vertical, and 135 diagonal. We combine these features into a single feature vector, and standardize each feature to a normal distribution. The environment of the test system is a Windows XP PC running MATLAB 7.0.

V. Experiments

We have comprehensively compared the proposed MTSVM algorithm with existing popular RF algorithms. Experiments were carried out upon a subset of images from the Corel Photo Gallery. This subset consists of about 20 000 images of very diverse subject matter for which each image was manually labeled with one of 90 concepts. Initially, 500 queries were randomly selected, and the program autonomously performs a RF with the top five most relevant images (i.e., images with the same concept as the query) marked as positive feedback samples within the top 40 images, similarly five negative feedback samples are marked. The procedure is chosen to replicate a common working situation where a user would not label many images for each feedback iteration.

Precision and standard deviation (SD) are used to evaluate the performance for all RF algorithms. Precision is defined as the percentage of relevant images in the top N retrieved images. In our experiments, a precision curve is the averaged precision values over the 500 queries. Similarly, an SD curve is the SD values of the 500 query precision values. The precision curve assesses the effectiveness of a given algorithm and the corresponding SD curve assesses its robustness.

In Fig. 2, we compare the proposed MTSVM with the cotraining SVM (CTSVM) and the random subspace method SVM (RSMSVM) operating on their own. MTSVM is also compared with the original SVM-based RF [17], the constrained similarity measure SVM (CSVM)-based RF [7], the biased discriminant analysis (BDA)-based RF [18], and the direct BDA (DBDA) [13]. The effectiveness of BDA- and SVM-based RF approaches has been demonstrated in [18], [13], and [3], [5], [7], [17], respectively. The results illustrate that 1) MTSVM can consistently outperform the conventional SVM and other SVM-based approaches. Moreover, MTSVM also outperforms BDA- and DBDA-based RF approaches, and 2) from the SD curves, we can also note that MTSVM is generally more robust than the other algorithms.

In our experiments, we chose a Gaussian kernel $K(\mathbf{x}, \mathbf{y}) = e^{-\rho \|\mathbf{x}-\mathbf{y}\|^2}$ with $\rho = 1$ (the default value in the OSU-SVM MatLab toolbox [19]) for all algorithms. The performances of these SVM algorithms are stable over a range of ρ values.

VI. Conclusion

In this paper, we propose a new algorithm for RF in CBIR designed to solve the small sample size problem and improve the capability of the kernel machine compared to traditional SVM-based RFs. The new MTSVM has the advantages of both the cotraining technique and the random sampling method. Extensive experiments using a large subset of the Corel Photo Gallery, some 20 000 manually labeled images, demonstrate that the newly developed MTSVM algorithm outperforms several established CBIR RF methods consistently in terms of effectiveness and robustness.

Acknowledgment

The authors would like to thank the associate editor Dr. E. Saber and all anonymous reviewers for their constructive comments on this paper.

References

[1] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with cotraining," COLT: Proceedings of the Workshop on Computational Learning Theory.

[2] J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, Feb. 1998.

[3] Y. Chen, X. Zhou, and T. S. Huang, "One-class SVM for learning in image retrieval," in *Proc. IEEE Int. Conf. Image Processing*, Thessaloniki, Greece, Oct. 2001, vol. 1, pp. 34–37.

[4] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, Aug. 1998.

[5] P. Hong, Q. Tian, and T. S. Huang, "Incorporate support vector machines to content-based image retrieval with relevant feedback," in *Pro. IEEE Int. Conf. Image Processing*, Vancouver, BC, Canada, Sep. 2000, vol. 1, pp. 750–753.

[6] T. Gevers and A. W. M. Smeulders, "Content-based image retrieval: An overview," in *Emerging Topics in Computer Vision*. Englewood Cliffs, NJ: Prentice-Hall, 2004, pp. 333–384.

[7] G. Guo, A. K. Jain, W. Ma, and H. Zhang, "Learning similarity measure for natural image retrieval with relevance feedback," *IEEE Trans. Neural Netw.*, vol. 12, no. 4, pp. 811–820, Jul. 2002.

[8] J. Kittler, M. Hatef, P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.

[9] B. S. Manjunath, J. Ohm, V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 703–715, Jun. 2001.

[10] Y. Rui, T. S. Huang, and S. Mehrotra, "Content-based image retrieval with relevance feedback in MARS," in *Proc. IEEE Int. Conf. Image Processing*, Washington, DC, Oct. 1997, vol. 2, pp. 815–818.

[11] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000.

[12] M. J. Swain and D. H. Ballard, "Color indexing," *Int. J. Comput. Vis.*, vol. 7, no. 1, pp. 11–32, 1991.

[13] D. Tao, X. Tang, X. Li, and Y. Rui, "Direct kernel biased discriminant analysis: A new content-based image retrieval relevance feedback algorithm," *IEEE Trans. Multimedia*, vol. 8, no. 3, pp. 716–727, Jun. 2006.

[14] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.

[15] Y. Xu, E. Saber, and A. M. Tekalp, "Object segmentation and labeling by learning from examples," *IEEE Trans. Image Process.*, vol. 12, no. 6, pp. 627–637, Jun. 2003.

[16] Y. Xu, E. Saber, and A. M. Tekalp, "Dynamic learning from multiple examples for semantic object segmentation and search," *Comput. Vis. Image Understand.*, vol. 95, no. 3, pp. 334–353, Sep. 2004.

[17] L. Zhang, F. Lin, and B. Zhang, "Support vector machine learning for image retrieval," in *Proc. IEEE Int. Conf. Image Processing*, Thessaloniki, Greece, Oct. 2001, vol. 2, pp. 721–724.

[18] X. Zhou and T. S. Huang, "Small sample learning during multimedia retrieval using biasmap," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, Kauai, HI, Dec. 2001, vol. 1, pp. 11–17.

[19] J. Ma, Y. Zhao, and S. Ahalt, "OSU SVM classifier matlab toolbox (ver. 3.00)," in *Pulsed Neural Networks*. Cambridge, MA: MIT Press, 1999 [Online]. Available: http://eewww.eng.ohiostate.-edu/maj/osu_svm/