Bootstrapping SVM Active Learning by Incorporating Unlabelled Images for Image Retrieval

Lei Wang, Kap Luk Chan, Zhihua Zhang*
School of Electrical and Electronic Engineering
Nanyang Technological University
Nanyang Avenue, Singapore, 639798
{elwang, eklchan}@ntu.edu.sg

Abstract

The performance of image retrieval with SVM active learning is known to be poor when started with few labelled images only. In this paper, the problem is solved by incorporating the unlabelled images into the bootstrapping of the learning process. In this work, the initial SVM classifier is trained with the few labelled images and the unlabelled images randomly selected from the image database. Both theoretical analysis and experimental results show that by incorporating unlabelled images in the bootstrapping, the efficiency of SVM active learning can be improved, and thus improves the overall retrieval performance.

1. Introduction

Content-Based Image Retrieval (CBIR) is known to suffer from the semantic gap, and learning from user's relevance feedback was considered as a way in an attempt to bridge this gap [11, 9, 14, 3, 7]. However, the learning performance is often constrained due to the problem of small sample because a user is unwilling to label too many images in a retrieval. Recently, Support Vector Machines (SVM) based active learning has been proposed to deal with this problem, in which the images shown to query the user are the most uncertain images, e.g., those closest to the optimal separating hyperplane, instead of the most positive images [13]. Though promising improvement is obtained, it is also found that active learning would have difficulties to learn well when it is started with few labelled images, e.g., only one positive and one negative image examples provided by the user at the beginning of retrieval. Currently, this problem is tackled by asking the user to label more randomly selected images in the first round of relevance feedback and starting active learning afterwards. Though labelling more samples can get around this problem, it increases the burden on the user. On the other hand, heuristically labelling a fixed number of images, e.g. twenty images, is not in the spirit of active learning because this number varies with different retrieval tasks and the image databases involved, and may not be known in advance.

In image retrieval, the images labelled by the user are usually very limited in number, especially at the beginning of retrieval. However, the unlabelled images are ample in the image database and are easy to collect. In this paper, a new bootstrapping strategy for SVM active learning is proposed, in which the initial SVM classifier is trained with few labelled images and the unlabelled images randomly selected from the image database. We show, through theoretical analysis and experimental results, that by incorporating unlabelled images, the efficiency of SVM active learning can be improved, leading to an overall better retrieval performance.

2. Background

SVM active learning can be well analyzed by using the version space. In the following, the concept of version space is briefly presented (For the detail, please see [4]).

Let $\mathcal{Z}$ be a set of training samples and $\mathcal{Z} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_m, y_m)\}$, where $\mathbf{x}$ ($\mathbf{x} \in \mathbb{R}^n$) is an n-dimensional input vector and y ($y \in \{+1, -1\}$) is the true label of $\mathbf{x}$. Commonly, the class with $y = +1$ is named as *positive* while the class with $y = -1$ is named as *negative*. Given a kernel $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$, where $\phi(\cdot)$ denotes a mapping and $\langle \cdot, \cdot \rangle$ denotes an inner product, the input vectors in $\mathcal{Z}$ are implicitly mapped into a feature space, $\mathcal{F}$. Let $\mathbf{w}$ be the normal vector of a hyperplane in $\mathcal{F}$, and $f(\mathbf{x}) = \langle \phi(\mathbf{x}), \mathbf{w} \rangle$ represents a hyperplane in $\mathcal{F}$ passing

*The author is now with the Department of Computer Science, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. *Email address*: zhzhang@cs.ust.hk.

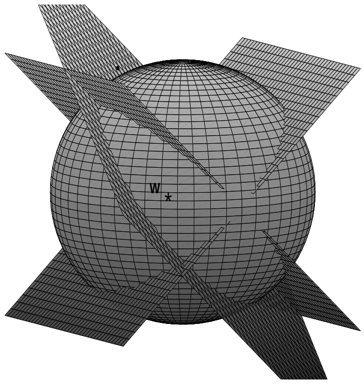

Figure 1. A version space in a 3D space

through the origin[1]. An unseen data, $\mathbf{x}_u$, can be labelled by the sign of $f(\mathbf{x}_u)$. Let $\mathrm{V}(\mathcal{Z})$ denote the version space given $\mathcal{Z}$,

$$\mathrm{V}(\mathcal{Z}) = \{\mathbf{w} \in \mathcal{W} \mid \|\mathbf{w}\| = 1, \forall\, i \in \{1, \cdots, m\} : y_i \langle \phi(\mathbf{x}_i), \mathbf{w} \rangle > 0\}$$

where $\mathcal{W}$ denotes the parameter space. A duality can be found between an $f(\mathbf{x})$ in $\mathcal{F}$ and a $\mathbf{w}$ in $\mathcal{W}$. $y\langle \phi(\mathbf{x}), \mathbf{w} \rangle = 0$ represents a hyperplane in $\mathcal{F}$ with the normal vector $\mathbf{w}$. Also, it represents a hyperplane in $\mathcal{W}$ with the normal vector $y\phi(\mathbf{x})$. Given a training sample, $(\mathbf{x}_i, y_i)$, there is $y_i\langle \phi(\mathbf{x}_i), \mathbf{w} \rangle > 0$ for a hyperplane, $f(\mathbf{x})$, that can correctly classify it. According to the duality, the corresponding hyperplane in $\mathcal{W}$ bisects the surface of the hypersphere of $\|\mathbf{w}\| = 1$, and only the part satisfying $y_i\langle \phi(\mathbf{x}_i), \mathbf{w} \rangle > 0$ is favored. In this way, given the training set $\mathcal{Z}$ with m training samples, the version space is a connected region on the surface of the hypersphere, carved by the corresponding m hyperplanes. Figure 1 illustrates such a version space in a 3D parameter space carved by three hyperplanes. The region closest to the viewer is the version space, and $\mathbf{w}$ represents a normal vector therein. In a version space, an SVM classifier has the following geometrical interpretation. Imagining a largest hypersphere with its center restricted on the version space and not intersecting with any hyperplane, the normal vector of the optimal separating hyperplane, $\mathbf{w}^*$, lies at the center of this largest hypersphere. The training samples are *support vectors* if they correspond to the hyperplanes which are tangent to this largest hypersphere. Note that this interpretation is based on the assumption that $\|\phi(\mathbf{x})\|$ is a constant. It can be well satisfied when a Gaussian RBF kernel is used.

In [13], considering that the normal vector of the optimal separating hyperplane, $\mathbf{w}^*$, can be well captured when the size of version space becomes small enough[2], Tong *et al.* proposed to select the unlabelled sample which can reduce the current version space as much as possible. Assuming that only one unlabelled sample is selected in each learning cycle, they proved that the hyperplane induced by this sample should halve the current version space, and proposed the Simple method in which this sample is approximated by the unlabelled sample closest to the current separating hyperplane. The Simple method is not the best method in identifying the desired sample, and it often leads to poor learning when active learning starts with few labelled data only. However, this method is much less computational intensive, and it is suitable for practical applications, such as image retrieval where multiple samples are selected in each learning cycle. In [2] and [10], selection strategies similar to the Simple method are proposed from different perspectives.

3. SVM active learning incorporating unlabelled data

3.1. The theoretical analysis

Let $\mathcal{D}$ be a given database, and $\mathcal{D}_l^i$ and $\mathcal{D}_u^i$ are the labelled and unlabelled data sets in the i-th active learning cycle, respectively. We have $\mathcal{D} = \mathcal{D}_l^i \cup \mathcal{D}_u^i$ and $\mathcal{D}_l^i \cap \mathcal{D}_u^i = \emptyset$. Let V_i denote the version space given $\mathcal{D}_l^i$, and S_{V_i} denotes the size of V_i. Let $\mathbf{w}_i$ be the normal vector of the separating hyperplane given V_i. Let $\triangle \mathbf{w}_i = \|\mathbf{w}_i - \mathbf{w}^*\|$ denote the error between $\mathbf{w}_i$ and $\mathbf{w}^*$. The expectation of the squared error, $E(\|\triangle \mathbf{w}_i\|^2)$, can be expressed as follows.

$$\begin{array}{ll} & E(\|\triangle \mathbf{w}_i\|^2) \\ = & \int_{\mathrm{V}_i} \|\mathbf{w}_i - \mathbf{w}^*\|^2 p(\mathbf{w}_i, \mathbf{w}^* | \mathrm{V}_i) d\mathbf{w}_i d\mathbf{w}^* \\ = & \int_{\mathrm{V}_i} (\|\mathbf{w}_i\|^2 - 2\mathbf{w}_i^\top \mathbf{w}^* + \|\mathbf{w}^*\|^2) p(\mathbf{w}_i, \mathbf{w}^* | \mathrm{V}_i) d\mathbf{w}_i d\mathbf{w}^* \\ = & 2 - 2\int_{\mathrm{V}_i} \cos\theta\, p(\mathbf{w}_i, \mathbf{w}^* | \mathrm{V}_i) d\mathbf{w}_i d\mathbf{w}^* \\ < & 2 - 2\cos\theta_{max} \int_{\mathrm{V}_i} p(\mathbf{w}_i, \mathbf{w}^* | \mathrm{V}_i) d\mathbf{w}_i d\mathbf{w}^* \\ = & 2(1 - \cos\theta_{max}) \qquad [\theta_{max} \in (0, \pi)] \end{array} \tag{1}$$

where θ_{max} is the maximal angle between $\mathbf{w}_i$ and $\mathbf{w}^*$, and it is positively correlated to S_{V_i}. This equation indicates that the larger the S_{V_i}, the larger the deviation of $\mathbf{w}_i$ from $\mathbf{w}^*$. Hence, SVM active learning seeks the optimal normal vector, $\mathbf{w}^*$, by greedily reducing the size of version space.

In the case of selecting multiple queries in each learning cycle, the optimal selection strategy should also have the version space reduced as much as possible. Hence, the expectation of the size of V_{i+1} in the $(i+1)$-th learning cycle, $E\left(S_{\mathrm{V}_{i+1}}\right)$, should be minimized after the selected k $(k \geq 1)$ unlabelled samples are labelled. Let $\mathcal{D}_k$ be the

[1] The version space introduced requires that the training data are linearly separable in the feature space and that the separating hyperplane passes through the origin. As noted in [13], these two requirements can be well satisfied by modifying the kernel or the input space. Here, it is assumed that they have been satisfied.

[2] Note that it has been assumed that the training data are linearly separable in $\mathcal{F}$. Hence, $\mathbf{w}^*$ will always lie in the series of version spaces induced by the sequence of new samples because the optimal separating hyperplane can always correctly classify all the training data.

set of the k selected unlabelled samples. The optimal $\mathcal{D}_k$ can be described as follows.

$$\mathcal{D}_k^* = \arg \min_{\mathcal{D}_k \subset \mathcal{D}_u^i} E\left(S_{\mathrm{V}_{i+1}}\right) \tag{2}$$

Let n_k denote the number of sub-regions into which the version space, V_i, will be partitioned by the k induced hyperplanes. Let $\mathrm{R}_{i,j}$ $(j = 1, 2, \cdots, n_k)$ be the j-th sub-region, and $S_{\mathrm{R}_{i,j}}$ denotes the size of $\mathrm{R}_{i,j}$. We have $\mathrm{V}_i = \bigcup_{j=1}^{n_k} \mathrm{R}_{i,j}$ and $S_{\mathrm{V}_i} = \sum_{j=1}^{n_k} S_{\mathrm{R}_{i,j}}$. $\mathrm{R}_{i,j}$ will become V_{i+1} if the optimal normal vector, $\mathbf{w}^*$, lies in it. In this way,

$$\begin{aligned} E\left(S_{\mathrm{V}_{i+1}}\right) &= \sum_{j=1}^{n_k} \left[S_{\mathrm{R}_{i,j}} P(\mathbf{w}^* \in \mathrm{R}_{i,j})\right] \\ &= \sum_{j=1}^{n_k} \left[S_{\mathrm{R}_{i,j}} \int_{\mathrm{R}_{i,j}} p(\mathbf{w}^* | \mathrm{V}_i) d\mathbf{w}^*\right] \end{aligned} \tag{3}$$

where $P(\mathbf{w}^* \in \mathrm{R}_{i,j})$ is the probability of $\mathbf{w}^*$ falling into $\mathrm{R}_{i,j}$. Considering that $\mathbf{w}^*$ can lie anywhere in V_i with equal probability, we have $\int_{\mathrm{R}_{i,j}} p(\mathbf{w}^* | \mathrm{V}_i) d\mathbf{w}^* = \frac{1}{S_{\mathrm{V}_i}} \int_{\mathrm{R}_{i,j}} d\mathbf{w}^* = \frac{S_{\mathrm{R}_{i,j}}}{S_{\mathrm{V}_i}}$. Hence, equation (3) becomes

$$E\left(S_{\mathrm{V}_{i+1}}\right) = \frac{1}{S_{\mathrm{V}_i}} \left(\sum_{j=1}^{n_k} S_{\mathrm{R}_{i,j}}^2\right) \geq \frac{1}{S_{\mathrm{V}_i}} \frac{\left(\sum_{j=1}^{n_k} S_{\mathrm{R}_{i,j}}\right)^2}{n_k} = \frac{S_{\mathrm{V}_i}}{n_k} \tag{4}$$

According to Cauchy inequality, the equality can be achieved if and only if $S_{\mathrm{R}_{i,1}} = S_{\mathrm{R}_{i,2}} = \cdots = S_{\mathrm{R}_{i,n_k}} = \frac{S_{\mathrm{V}_i}}{n_k}$. This result indicates that the optimal selection strategy for SVM active learning with multiple queries is to select the unlabelled samples corresponding to the k hyperplanes that can partition the current version space into as many as possible equally-sized sub-regions. The optimal selection strategy for the single query case given in [13] is then a special case, where $k = 1$. Also, equation (4) shows that the Simple method, which selects the k unlabelled samples closest to $\mathbf{w}_i$, is not optimal for the case of multiple queries. However, it is impractical to find the optimal k unlabelled samples according to equation (4) because the parameter space $\mathcal{W}$ often has a very high dimensionality, and the Simple method is a viable option in practice. Hence, a way has to be found to improve the efficiency of the Simple method.

Given a version space, the sum of the squared distances of the hyperplanes induced by the k selected unlabelled samples to $\mathbf{w}^*$ can be used as an indicator of the selection efficiency by the Simple method. The smaller the sum, the more likely the closer of the k hyperplanes to $\mathbf{w}^*$, and the support vectors can be selected with higher probability. Also, a smaller sum means a tighter enclosure to $\mathbf{w}^*$ and a smaller version space is more likely resulted. Figure 2 shows a small version space, V_i, that can be approximated by a region, $ABCD$, on a 2D plane. The k hyperplanes are represented by $h_1, h_2, \cdots, h_k$, respectively. V_i is partitioned into six pieces and the fourth (shadowed region) is the next version space, V_{i+1}. The expectation of the sum of the squared distances can be expressed as

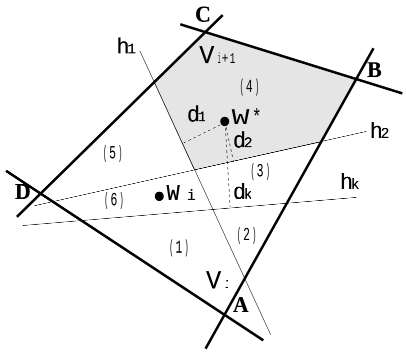

Figure 2. A V_i approximated by a 2D plane

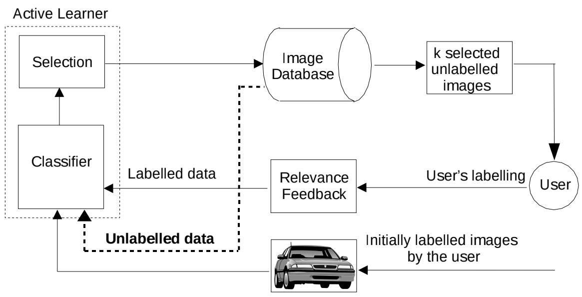

Figure 3. Retrieval model

$$E\left(\sum_{j=1}^{k} d_j^2\right) = k\left[\|\mathbf{w}_i - \mathbf{w}^*\|^2 (r_1 + r_2)^2 + \frac{2}{3}(r_1^2 + r_1 r_2 + r_2^2)\right]$$

where d_j denotes the distance from the i-th hyperplane, h_i, to $\mathbf{w}^*$. r_1 and r_2 are the minimum and maximum of the distances, respectively, and they are regarded as two constants in a selection. This equation shows that a smaller $\|\mathbf{w}_i - \mathbf{w}^*\|$ will incur a smaller $E\left(\sum_{j=1}^{k} d_j^2\right)$. As a result, the selection efficiency by the Simple method will be higher, and a smaller version space can be obtained. It means that reducing the deviation of $\mathbf{w}_0$ from $\mathbf{w}^*$ can make the selection by the Simple method more efficient. The way of labelling more images in [13] was, in fact, to reduce this deviation.

3.2. Bootstrapping SVM active learning by incorporating unlabelled images

Recently, learning with labelled and unlabelled data, also known as semi-supervised learning, has attracted much attention [8, 1, 6]. It aims to achieve good classification performance with the help of unlabelled data in the presence of the small sample problem, and some promising results have been reported. Enlightened by this, instead of asking the user to label more images, we incorporate the unlabelled images before active learning starts. As shown in Figure 3, at the beginning, the user launches a retrieval with few initially labelled images. The labelled images consist of at least one positive and one negative images. Then the unlabelled images in the image database are incorporated to train the initial SVM classifier. By employing a suitable algorithm for learning with labelled and unlabelled data to bootstrap the SVM active learning, the deviation of $\mathbf{w}_0$ from $\mathbf{w}^*$, $\|\triangle \mathbf{w}_0\|$, can then be reduced. In the subsequent learning cycles, the normal SVM active learning is used. When active learning ends, the database images are ranked in descending order of the distances to the final separating hyperplane, and the top images are shown as the retrieval result. The proposed method has the following advantages. First, the initial classifier is improved by incorporating unlabelled images instead of heuristically labelling more images. Hence, it reduces the labelling cost and thus the burden on the user to improve the retrieval. Second, this method is general such that any algorithm for learning from labelled and unlabelled data can be incorporated.

4. Experimental results

The experiments aim to evaluate the effectiveness of the proposed method. Transductive SVM (TSVM) is used as the algorithm learning from labelled and unlabelled data [6]. In the following, the proposed method is named as TSVM-SAL while the method in [13] is named as SAL. An artificial database and a real color image database are used. The artificial database, as shown in [15], includes seven 2D Gaussian-distributed classes. Each class has 100 samples representing 100 "images". The real color image database includes 600 general color images composed from *VisTex* of MIT and *Corel* Stock Photos. Six image classes are defined based on high-level semantics (i.e. defined by a group of human observers), and each class includes 100 image samples. A perceptually uniform color space, $CIE - Lab$, is used to represent general color images, and a feature vector of color moments [12] is defined for each image. The two databases provide the ground truth for evaluation. Besides retrieval precision (PR), classification accuracy (CA) and selection efficiency (SE) are also calculated. Classification accuracy is defined to be the percentage of the database images that are correctly classified as positive and negative. Selection efficiency is defined as the percentage of the true support vectors among the selected unlabelled samples. The true support vectors are found by training an SVM classifier with the ground truth, e.g., the pre-classified databases.

4.1. Experimental procedure

(1) Treat class i $(i = 1, 2, \cdots)$ as positive and the other classes as negative. k_0 images are randomly selected as the initially labelled samples before active learning starts. Among the k_0 image samples, there are, at least, one positive and one negative image samples; (2) Based on the k_0 labelled images, the initial SVM classifier is trained by TSVM-SAL and SAL, respectively. In TSVM-SAL, 10% unlabelled images in the image database are randomly selected in the present work; (3) Select k (k is set to 20 to be consistent with [13]) unlabelled images according to the Simple method. Each image is labelled according to the ground truth to simulate the user's labelling in relevance feedback; (4) Add the k newly labelled samples into the current labelled sample set, and retrain the SVM classifier with labelled images only; (5) Redo steps (3) to (4) with three active learning cycles. In each cycle, after the SVM classifier is trained, the database images are ranked and the retrieval result is determined. The corresponding CA, SE, and PR values are calculated for TSVM-SAL and SAL, respectively; (6) To accumulate statistics, redo steps (1) to (5) fifty times. For each method, the average of CA, SE, and PR are calculated as the values when class i is treated as positive; (7) Redo steps (1) to (6) until each class has been treated as positive once. For each method, the average of CA, SE, and PR on all the classes are calculated as the values for the whole database.

In the experiments, the software for training SVM (and TSVM) is SVM^{light} [5]. Gaussian RBF kernel, $k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}\right)$, is used, where σ is set to the average of the Euclidean distances among the training samples. The regularization parameter, C, is set to 100 for the two databases.

4.2. Results and discussions

Figure 4 shows the comparison between TSVM-SAL and SAL on the artificial database. These sub-figures are arranged, from top-left to bottom-right, according to the order of the three active learning cycles. The horizontal axis of each sub-figure shows the values of k_0 while the vertical axis shows the value of CA, SE, or PR. It can be seen that, by incorporating unlabelled images to train the initial classifier, TSVM-SAL achieves overall better performance on CA, SE, and PR. In sub-figure (a-1) corresponding to the initial classifier, TSVM-SAL achieves higher CA values than SAL for the same k_0. This indicates that the initial classifier is improved by incorporating the unlabelled images. In sub-figure (a-2), the values of SE in the first learning cycle are shown. It can be seen that, benefiting from the improved initial classifier, higher SE values are obtained by TSVM-SAL than SAL for the same k_0. This means that more support vectors are selected in these selections. Hence, better classification and retrieval performance is achieved. These are reflected from the higher values of PR and CA, given by TSVM-SAL after the first learning cycle, shown in sub-figures (a-3) and (b-1), respectively. In the subsequent two leaning cycles (from sub-figure (b-2) to (c-3)), TSVM-SAL still shows better performance because it has a good starting point. Also, it can been seen that, with the increasing values of k_0, the improvement of TSVM-SAL becomes less. This is expected because, when the number of initially labelled samples increases, a good initial classifier can be obtained by using the labelled images only and the help from unlabelled images becomes less. However, it can be seen that, when few labelled images are available only, TSVM-SAL still shows better performance. By cross-referencing the results corresponding to different values of k_0, it can be seen that, in sub-figures (a-3), (b-3), and (c-3), the PR values of TSVM-SAL for a smaller k_0 are comparable to those of SAL for a larger k_0. This result means that, for the case of image retrieval, when few labelled images are available only, incorporating unlabelled images can show the similar effect as asking the user to label more images. Hence, incorporating unlabelled samples is helpful to lighten the burden on the user. Also, these results indicate that TSVM-SAL requires less labelling for achieving a given retrieval precision. Figure 5 shows the comparison on the real color image database, and similar conclusions are drawn. However, the improvement of TSVM-SAL over SAL is less because complex distributions, such as heavily overlapped distributions, of real data can degrade the efficiency of learning from labelled and unlabelled data. Summarily, the above experimental results demonstrate that bootstrapping SVM active learning by incorporating unlabelled data helps achieving efficient active learning using few labelled images and hence a better retrieval performance. This achievement is at the expense of introducing a bootstrapping step which incurs extra computational overhead although it can be reduced by incorporating less unlabelled samples. The benefit of incurring such an overhead is justified in term of the improvement in performance.

5. Conclusion

This paper gives a theoretical analysis on SVM active learning by using the version space. Based on the analysis, a method of bootstrapping SVM active learning is proposed for image retrieval, in which the initial SVM classifier is improved by incorporating unlabelled images instead of asking the user to label more randomly selected images. The proposed method can effectively improve the learning and retrieval efficiency of SVM active learning without increasing the burden on the user. The experimental results on both artificial and real databases demonstrate the better performance achieved by the proposed method.

References

[1] A. Blum and T. Mitchell. Combining labeled and unlabeled data with co-training. *Proceedings of the Workshop on Computational Learning Theory, Morgan Kaufmann Publishers*, 1998.

[2] C.Campbell, N.Cristianini, and A.Smola. Query learning with large margin classifiers. *Proceedings of the 17th International Conference on Machine Learning*, pages 111–118, 2000.

[3] I. J. Cox, M. L. Miller, S. M. Omohundro, and P. N. Yianilos. Pichunter: Bayesian relevance feedback for image retrieval. *Proceedings of the 13th International Conference on Pattern Recognition*, 3:361–369, 1996.

[4] R. Herbrich, T. Graepel, and R. C. Williamson. The structure of version space. *submitted to Journal of Machine Learning Research*, 2001.

[5] T. Joachims. Making large-scale support vector machine learning practical. *B. Scholkopf, C. Burges, and A. Smola. Advances in Kernel Methods: Support Vector Machines, MIT Press, Cambridge, MA, December*, 1998.

[6] T. Joachims. Transductive inference for text classification using support vector machines. *Proceedings of the 16th International Conference on Machine Learning*, pages 200–209, 1999.

[7] T. Minka. An image database browser that learns from user interaction. *Master's Thesis of MIT*, 1996.

[8] K. Nigam, A. K. McCallum, S. Thrun, and T. M. Mitchell. Text classification from labeled and unlabeled documents using EM. *Machine Learning*, 39(2/3):103–134, 2000.

[9] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra. Relevance feedback: A power tool for interactive content-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 8:644–655, 1998.

[10] G. Schohn and D. Cohn. Less is more: Active learning with support vector machines. *Proceedings of the 17th International Conference on Machine Learning*, pages 839–846, 2000.

[11] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, December 2000.

[12] M. Stricker and M. Orengo. Similarity of color images. *Proceedings of SPIE Storage and Retrieval for Image and Video Databases*, 2420:381–392, 1995.

[13] S. Tong and E. Chang. Support vector machine active learning for image retrieval. *Proceedings of ACM International Conference on Multimedia*, pages 107–118, Ottawa, Canada, Oct, 2001.

[14] N. Vasconcelos and A. Lippman. Learning from user feedback in image retrieval systems. *Proceedings of Neural Information Processing Systems (NIPS), Denver, Colorado*, 1999.

[15] X. S. Zhou and T. S. Huang. Small sample learning during multimedia retrieval using biasmap. *Proceedings of IEEE International Conference on Computer Visiona and Pattern Recognition (CVPR), Hawaii*, Dec, 2001.

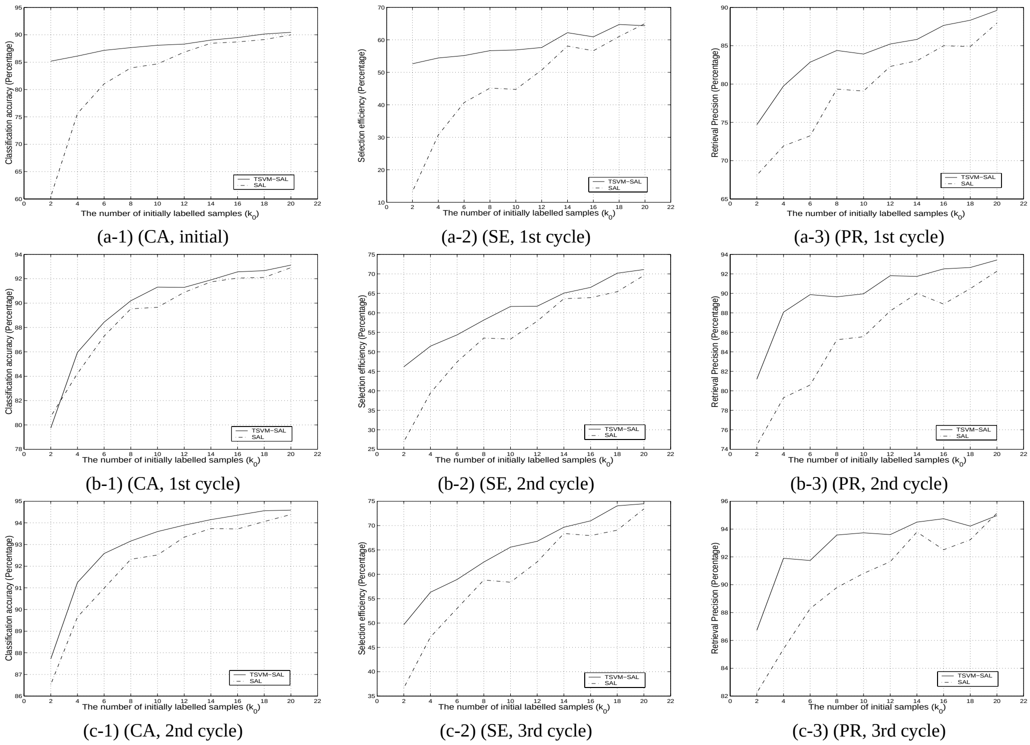

Figure 4. Comparison on the artificial database

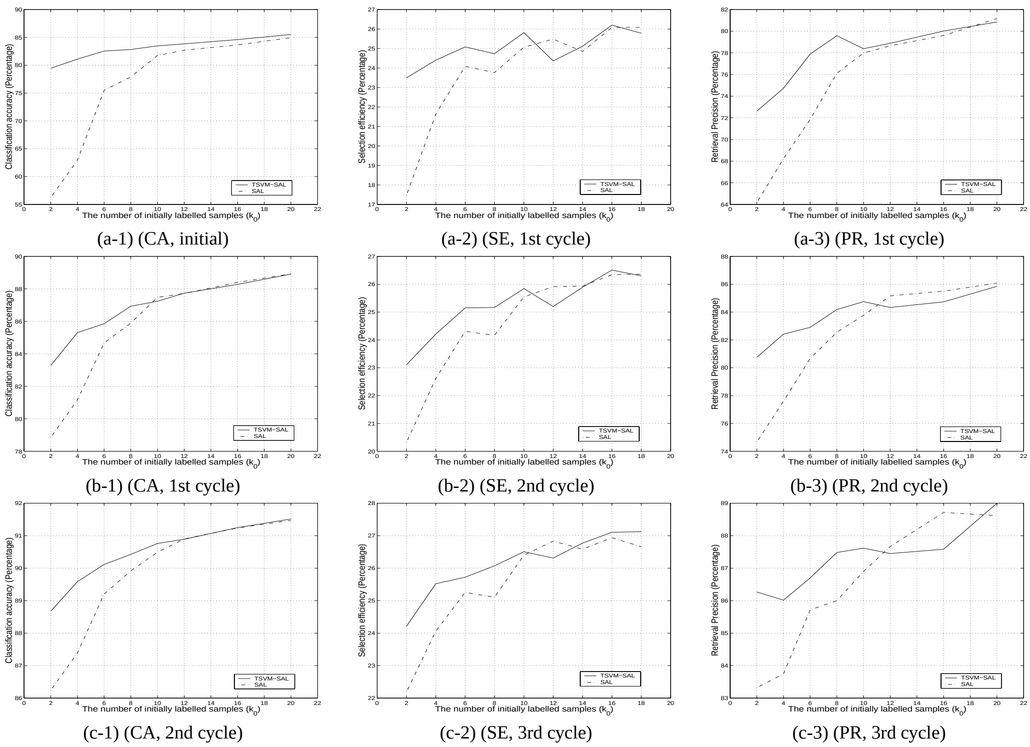

Figure 5. Comparison on the real color image database