MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan

13-26

SVM-based Relevance Feedback in Image Retrieval using Invariant Feature Histograms

Lokesh Setia, Julia Ick, Hans Burkhardt
Institute of Computer Science
Albert-Ludwigs-University Freiburg
79110 Freiburg im Breisgau, Germany
Email: {setia, ick, burkhardt} @informatik.uni-freiburg.de

Abstract— **Relevance Feedback is an interesting procedure to improve the performance of Content-Based Image Retrieval systems even when using low-level features alone. In this work we compare the efficiency of one class and two class Support Vector Machines in content-based image retrieval using Invariant Feature Histograms. We describe our methodology of performing Relevance Feedback in both cases and report encouraging results on a subset of MPEG-7 content dataset.**

I. INTRODUCTION

Image retrieval is becoming ever more important as the amount of available multimedia data increases. Increasing database sizes also means that manual annotation of image databases becomes prohibitively expensive. Manual annotation has also the drawback that it is very subjective and user dependent even though sometimes it is the only way to retrieve images when semantic similarity is desired. In this work we deal with Content-Based Image Retrieval (CBIR) where the aim of the system is to lead the user to the desired images only through automatic processing of the query images that the user has to offer.

One way to look at Image Retrieval is that it is centered around an idea of a “user query concept”, which stands for the kind of images that the user of the system is looking for. The aim of a CBIR system then is to learn this query concept and deliver appropriate images to the user. The query concept is typically semantic (e.g. images of evening sun setting behind a beach), while the features that can be currently extracted from general databases are mostly visual and lower level. This leads to the so-called *semantic gap* which is the biggest showstopper for the wide-scale adoption of CBIR systems [3].

Fortunately, this semantic gap can be somewhat reduced by different approaches. At one end, one tries to achieve (partial) semantic similarity by pre-segmenting all the images in the database into meaningful regions, possibly singular objects. Similarity of two images then is defined through the similarity of their segmented regions. The success of this approach is of course very dependent on the quality of the segmentation process. Another approach to reduce the semantic gap is through so-called *relevance feedback* where the user provides feedback about the initial results in the hope of getting better results on the basis of this feedback [6].

II. RELEVANCE FEEDBACK USING INVARIANT FEATURE HISTOGRAMS

A. Invariant Features

In many cases during Image Retrieval, the exact position and orientation of objects in an image is only of secondary value. Thus, it is desirable to have features which are invariant to say, translation and rotation. We use invariant image features based on invariant integral which were introduced by Schulz-Mirbach [1]. Fast approximate invariant features were successfully used for image retrieval by Siggelkow et al.[2]. The invariant features are constructed as follows. Let $\mathbf{M} = \{\mathbf{M}(i,j)\}, 0 \leq i < N, 0 \leq j < M$ be an image, with $\mathbf{M}(i,j)$ representing the gray-value at the pixel coordinate (i,j). Let G be the transformation group of translations and rotations with elements $g \in G$ acting on the images, such that the transformed image is $g\mathbf{M}$. An invariant feature must satisfy $F(g\mathbf{M}) = F(\mathbf{M}), \forall g \in G$. Such invariant features can be constructed by integrating $f(g\mathbf{M})$ over the transformation group G.

$$I(\mathbf{M}) = 1/|G| \int_G f(g\mathbf{M})dg$$

which for a discrete image is approximated using summations

$$I(\mathbf{M}) = 1/PNM \sum_{t_0}^{N-1} \sum_{t_1}^{M-1} \sum_{\phi=0,\phi+=2\pi/P}^{2\pi(1-1/P)} f(g\mathbf{M})$$

The summations are replaced by histogramming operation which leads to higher robustness against occlusion or background changes while preserving invariance, although structural information is lost.

We use $f(\mathbf{X}) = (\mathbf{X}(4,0).\mathbf{X}(0,8))^{1/2}$ applied to each color layer of RGB space to yield a 3D histogram of $8*8*8 = 512$ bins.

B. Two-Class SVM

The importance of having good similarity measures for any feature set cannot be overemphasized. Although simple ranking methods based on for e.g. L_1- and L_2- norm have provided good results for single query images, they are not easily adaptable for multiple query images or for performing relevance feedback. Here we use Support Vector Machines

(SVM) which have proved to be very adaptable to various machine learning tasks.

Firstly we use a two-class SVM classifier in which we interpret CBIR as a two class classification problem, the two classes being the relevant (positive) and the not relevant (negative) images. Initially the classifier is trained using a few random images labelled by the user. Two-class SVM solves a classification problem by finding a maximum margin hyperplane that seperates the positive training instances from the negative ones. Each training instance is represented as a vector $\mathbf{x} \in R^n$ and belongs to one of the two classes $L = \{-1, 1\}$. The instances lying closest to the hyperplane are called support vectors and are the only vectors affecting the hyperplane. In many cases the training instances would not be linearly seperable in the original feature space R^n. In this case they can be transformed nonlinearly into a higher dimensional feature space $\mathcal{F}$ with a mapping

$$\phi : R^n \rightarrow \mathcal{F}$$

$$\mathbf{x} \mapsto \phi(\mathbf{x})$$

One obtains then a classification function of the form $f(\mathbf{x}) = sgn(\mathbf{w} \cdot \phi(\mathbf{x}) + b)$. Through the use of a kernel $k(\mathbf{u}, \mathbf{v}) = \phi(\mathbf{u}).\phi(\mathbf{v})$ different boundaries can be obtained. In fact, the kernel function k would lead to classifiers with maximum margin in some mapped feature space even if the mapping ϕ itself is not analytically defined, as long as the kernel satisfies Mercer's condition (Mercer, 1909).

It should be noted that just correct classification is not the goal of a general purpose CBIR system, as the concept of classes does not exist here in the strict sense. More important is an intelligent ordering of the results as the user would most likely see only the top few results. This behaviour is already commonly seen by text-search engines, where for e.g., some query keywords can lead to millions of hits. In a two-class SVM, it makes sense to assume that since the sign of the function $f(\mathbf{x})$ is used as the decision boundary, the images could be ordered on the basis of their decreasing values of $f(\mathbf{x})$. This simple procedure, as we see, provides good results. Furthermore, the user provides feedback not on the most positive images which are shown as intermediate results, but rather the images for which the magnitude of $f(\mathbf{x})$ is as close to zero, i.e. the images closest to the SVM boundary, as is suggested in [8].

C. One-Class SVM

One-Class SVMs were proposed by Schölkopf et al. [5]. One-Class SVMs are binary functions which capture regions in input space where the probability density lives (i.e. its support). Here we are interested only in the distribution of the relevant images. We try to find a hypersphere which contains most of the user-supplied relevant images while being as small as possible. This can be written in primal form as:

$$\min_{R \in \mathcal{R}, \zeta \in \mathcal{R}^l, \mathbf{c} \in \mathcal{F}} \quad R^2 + \frac{1}{\nu l} \sum_i \zeta_i$$

subject to

$$\|\phi(\mathbf{x}_i) - \mathbf{c}\|^2 \leq R^2 + \zeta_i, \;\; \zeta_i \geq 0, \; i = 1, ...l$$

which leads to the dual

$$\min_{\alpha} \sum_{i,j} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) - \sum_i \alpha_i k(\mathbf{x}_i, \mathbf{x}_i)$$

subject to

$$0 \leq \alpha_i \leq \frac{1}{\nu l}, \;\; \sum_i \alpha_i = 1$$

The optimal αs can be computed with the help of QP optimization algorithms. The decision function then is of the form

$$f(\mathbf{x}) = sgn(R^2 - \sum_{i,j} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) + 2 \sum_i \alpha_i k(\mathbf{x}_i, \mathbf{x}) - k(\mathbf{x}, \mathbf{x}))$$

This function returns positive for points inside this hypersphere and negative outside (note that although we use the term hypersphere the actual decision boundary can be varied by choosing different kernel functions). The results are sorted on the basis of their "positiveness". Since the actual value of the function $f(\mathbf{x})$ is not important we can speed up the process by noting that the first two terms in the decision function are constants. Furthermore the last term $k(\mathbf{x}, \mathbf{x})$ is also constant for many kernels. Thus, the images can be ordered simply on the basis of decreasing values of $f'(\mathbf{x}) = \sum_i \alpha_i k(\mathbf{x_i}, \mathbf{x})$

TABLE I

SVM KERNELS FOR IMAGE RETRIEVAL

Kernel	$k(\mathbf{x}, \mathbf{y})$
Linear	$\mathbf{x}.\mathbf{y}$
Polynomial	$(\gamma(\mathbf{x}_i \cdot \mathbf{x}_j) + coef0)^d, \; \gamma > 0$
RBF	$exp(-\gamma \|\mathbf{x} - \mathbf{y}\|^2), \; \gamma > 0$
Histogram Intersection	$\sum_{i=1}^{n} min(x_i, y_i)$

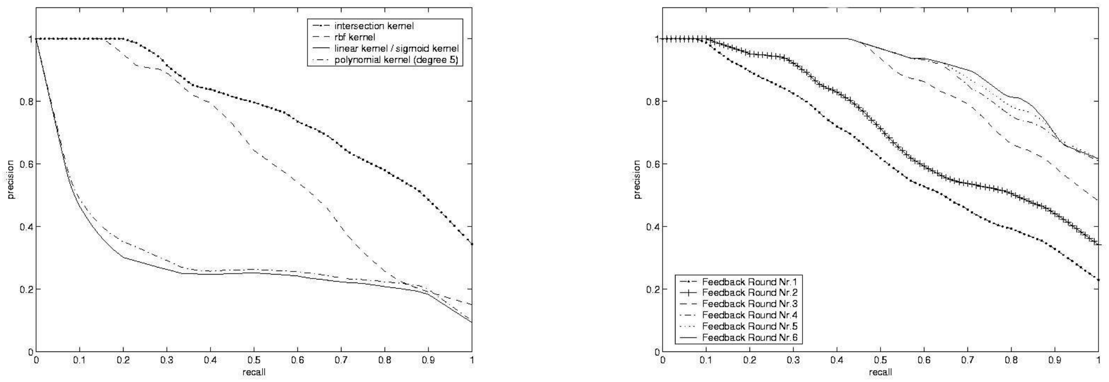
III. EXPERIMENTS AND RESULTS

We analyse results for the following experiments that we conducted among others, on a partially labelled data from the MPEG-7 content set consisting of about 2400 images[1]. The reader is encouraged to try out the web-based demo at http://bart.informatik.uni-freiburg.de/∼setia/svm/svm.php

A. Good kernel functions for invariant feature histograms

Selecting a good kernel function is critical to the performance of an SVM classifier. However, there exists no automatic method to find the optimum kernel function for a particular data set. Moreover, in CBIR a new SVM machine needs to be trained for each new query. Therefore, the best tuned parameters for a particular query image need not work well for all possible queries.

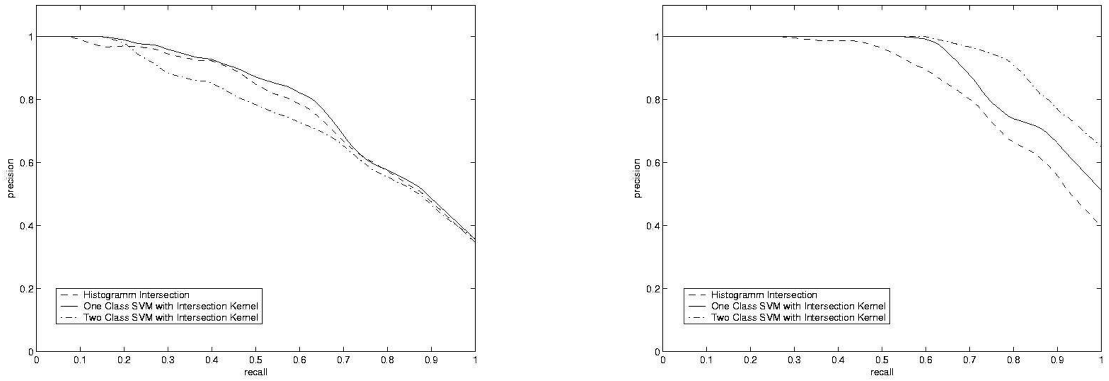
[1] We acknowledge Tristan Savatier, Aljandro Jaimes, and the Department of Water Resources, California, for providing them under the Licensing Agreement for the MPEG-7 Content Set (MPEG 98/N2466).



(a) Effect of different kernel functions

(b) Results after multiple feedback rounds

Fig. 1. Precision Recall plots with two-class SVM



(a) After first round

(b) After six rounds of Relevance Feedback

Fig. 2. Comparison of different retrieval methods

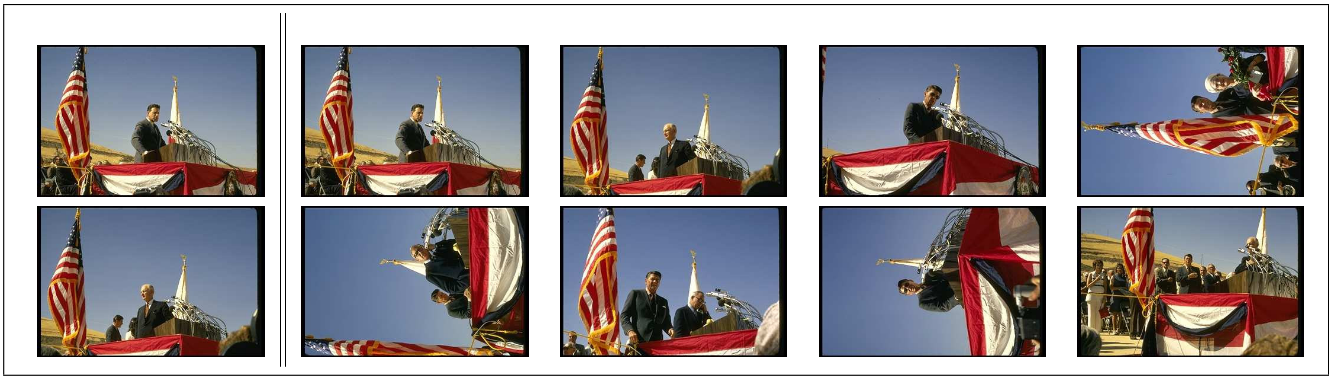


Fig. 3. The 8 most relevant images gained through training the CBIR system with a one-class SVM using the two relevant examples shown left

We tune the parameters for four kernel functions: linear, Gaussian, Polynomial and Histogram Intersection (HI) kernels based on the ground truth we have for this database. The kernels are shown in table I. The first three kernels are common from SVM literature. The HI kernel we try out based on our prior knowledge of our histogram-based features. Indeed, it has also been shown in previous work that L_1 based similarity measures are perceptually closer to human similarity definitions compared to L_2 based measures. In [10], for example interesting results are reported with the Laplacian kernel, which is similar to the HI kernel.

Figure 1(a) shows the results with different kernels using two-class SVM with three relevant and five irrelevant images. As was expected, the Histogram Intersection kernel performs better than the others.

B. Comparison of two-class and one-class SVM

This is a very interesting comparison. On the one hand, one expects a two-class scheme to perform better as it uses all the information that the user provided to the system, i.e. some relevant and some not relevant images. But on the other hand, one can safely assume that although the relevant images might form a cluster in the feature space, the irrelevant images may not, as they can belong to any of the remaining classes in the database. Thus, if these few irrelevant images are say randomly distributed over the feature space then they could possibly be of no help to a classifier which is trying to learn a decision boundary seperating the relevant images from the rest.

Figure 2(a) and 2(b) compare the results of two-class SVM vs. one-class SVM after the first and sixth round of relevance feedback respectively. Also shown for comparison is a simple ranking method based on L_1 similarity measure (Histogram Intersection). As can be seen, a two class SVM does not perform as good as one class SVM after the first round, as there are hardly enough samples (positive and negative) for the classifier. After six rounds, however, a two class SVM outperforms other methods.

C. Improvements over multiple feedback rounds

Figure 1(b) shows the Precision-Recall graph after multiple rounds of Relevance Feedback with two-class SVM have been performed. As can be seen from the graph, Relevance Feedback almost always leads to iterative improvement, but reaches a point of diminishing returns. The improvement in the initial rounds is very encouraging. The fact that the results could not be improved beyond a saturation level could be either due to limitation in the discriminatory performance of the features used or as a learning limitation of the classifier used. Our understanding is that perfect retrieval could not be attained with this combination of classifier and features because some images in our ground truth were only semantically similar while being visually very dissimilar.

IV. Conclusion

We presented Relevance Feedback methods for use with invariant feature histograms. We also compared the performance of one-class SVM and two-class SVM for this purpose. We showed the amount of performance gain that can be achieved after a number of feedback rounds have been performed. We believe that content-based image retrieval can greatly benefit through relevance feedback and future research should strive in improving the performance while demanding the least from the end user of the system.

Acknowledgment

This work is supported by the German Ministry for Education and Research (BMBF) through grant FKZ 01IRB02B.

References

[1] H. Schulz-Mirbach: Invariant features for gray scale images. In G. Sagerer, S. Posch, and F. Kummert, editors, *17. DAGM - Symposium "Mustererkennung"*, pages 1-14, Bielefeld, 1995. Reihe Informatik aktuell, Springer.

[2] S. Siggelkow, M. Schael, and H. Burkhardt. SIMBA - Search IMages By Appearance. In B. Radig and S. Florczyk, editors, Pattern Recognition, *Proc. of 23rd DAGM Symposium, number 2191 in LNCS Pattern Recognition*, pages 9-16. Springer, September 2001.

[3] W. Smeulders et al.; Content-Based Image Retrieval at the End of the Early Years, TPAMI Vol. 22, No. 12, Dec. 2000

[4] Chih-Chung Chang and Chih-Jen Lin, LIBSVM: a library for support vector machines, 2001, Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[5] Schölkopf, B., J.C. Platt, J. Shawe-Taylor, A.J. Smola and R.C. Williamson: Estimating the support of a high-dimensional distribution. Technical report No.(87) Microsoft Research (1999)

[6] Y. Rui,T. Huang, M. Ortega, S. Mehrotra: Relevance feedback: A power tool in interactive content-based image retrival. *IEEE Trans. on Circuits and Systems for Video Technology* 8(5) (Sep. 1998):644-655

[7] Chen, Y., et al, "One-class SVM for Learning in Image Retrival", *IEEE Intl Conf. on Image Proc.* (ICIP 2001), Thessaloniki, Greece, October 7-10, 2001

[8] S. Tong and E- Chang. Support vector machine active learning for image retrival, In *ACM International Conference on Multimedia*, pages 107-118; Otawa, Canada, September 2001

[9] A. Barla, E. Franceschi, F. Odone and A. Verri: Image kernels, In *Proceedings of the International Workshop on Pattern Recognition with Support Vector Machines, satellite event of ICPR 2002*, LNCS 2388, p. 83 ff, 2002.

[10] O. Chapelle, P. Haffner, and V. Vapnik. SVMs for histogram-based image classification. In *IEEE Transactions on Neural Networks*, 1999., special issue on Support Vectors.