Relevance Feedback for Content-Based Image Retrieval Using Support Vector Machines and Feature Selection

Apostolos Marakakis[1], Nikolaos Galatsanos[2], Aristidis Likas[3], and Andreas Stafylopatis[1]

[1] School of Electrical and Computer Engineering, National Technical University of Athens, 15780 Athens, Greece
[2] Department of Electrical and Computer Engineering, University of Patras, 26500 Patras, Greece
[3] Department of Computer Science, University of Ioannina, 45110 Ioannina, Greece
amara@central.ntua.gr, {galatsanos,arly}@cs.uoi.gr, andreas@cs.ntua.gr

Abstract. A relevance feedback (RF) approach for content-based image retrieval (CBIR) is proposed, which is based on Support Vector Machines (SVMs) and uses a feature selection technique to reduce the dimensionality of the image feature space. Specifically, each image is described by a multidimensional vector combining color, texture and shape information. In each RF round, the positive and negative examples provided by the user are used to determine a relatively small number of the most important features for the corresponding classification task, via a feature selection methodology. After the feature selection has been performed, an SVM classifier is trained to distinguish between relevant and irrelevant images according to the preferences of the user, using the restriction of the user examples on the set of selected features. The trained classifier is subsequently used to provide an updated ranking of the database images represented in the space of the selected features. Numerical experiments are presented that demonstrate the merits of the proposed relevance feedback methodology.

Keywords: Content-based image retrieval, relevance feedback, support vector machines, feature selection.

1 Introduction

The target of content-based image retrieval (CBIR) [1] is to retrieve images relevant to a query of a user, which is expressed by example. In CBIR, an image is described by automatically extracted low-level visual features, such as color, texture and shape. After a user has submitted one or more query images as examples of his/her preferences, a criterion based on this image description is used to rank the images of an image database according to their similarity with the examples of the query and, finally, the most similar are returned to the

C. Alippi et al. (Eds.): ICANN 2009, Part I, LNCS 5768, pp. 942–951, 2009.

user as the retrieval results. Nevertheless, there is an intrinsic difficulty for low-level image features to capture the human perception of image similarity. The reason for this is that the user is usually interested in the semantic content of the images. Unfortunately, the semantic content of an image is very difficult to describe using only low-level image features. This is the well-known semantic gap problem.

Relevance feedback (RF) has been proposed as a technique which can be used in order to bridge the semantic gap. RF is an interactive process between the user and the retrieval system. In each RF round, the user assesses the previously retrieved images as relevant or irrelevant to the initial query and provides this assessment as feedback to the system. This feedback is used, subsequently, by the system so that the ranking criterion is updated and a new set of images is retrieved. In this way, the subjective human perception of image similarity is incorporated to the system and the retrieval results are expected to improve, according to the user's viewpoint, with the RF rounds. With regard to RF approaches proposed in the literature, much work has been done during the last years, e.g. [2], [3], [4], [5], [6]. Among the proposed RF methodologies, the most prominent are those which use classifiers to distinguish between the classes of relevant and irrelevant images, e.g. [2], [3], [4]. In this context, the images assessed by the user as relevant or irrelevant up to the current RF round are used as positive and negative examples, respectively, to train a classifier. This classifier is used, subsequently, to update the database image ranking. It must be mentioned here that, among all the learning models proposed for this classification task, Support Vector Machines (SVMs) [12] constitute the most popular one.

The users of a CBIR system are usually not patient enough to provide the system with a large number of examples in each RF round. On the other hand, the description of the images is, generally, of very high dimensionality. From the above, it becomes obvious that, for RF methodologies which are based on training a classifier using the feedback examples, the available training set of the classifier is, almost always, very small compared to the dimensionality of training patterns. This can deteriorate the classifier performance, leading to poor retrieval results. To alleviate this problem, a methodology for selecting a relatively small number of good features for the classification task, based on the properties of the feedback examples, can be used. In this way, a significant dimensionality reduction can be achieved by removing irrelevant or redundant features. Training the classifier on the resulting lower-dimensional feature space can improve its ability to capture the underlying data distribution, thus leading to better classifier performance. Moreover, by reducing feature space dimensionality and, hence, the complexity of data, a decrease in training and image re-ranking time can be achieved. Many feature selection methods for classification have been proposed in the literature, e.g. [7], [8], [9]. As will be shown below, some of these methods can be incorporated straightforwardly in a CBIR system using RF.

The rest of the paper is organized as follows. The SVM methodology in the context of CBIR using RF is described in Section 2. Section 3 presents the feature selection methods used in this work. In Section 4, the details and results of the experiments are provided and, finally, in Section 5, conclusions and directions for future research are presented.

2 Using SVMs for RF in CBIR

Consider the binary classification problem $\{(x_i, y_i)\}_{i=1}^N$, where x_i are the labeled patterns and $y_i \in \{-1, +1\}$ the corresponding labels. Based on this training set, we want to train an SVM classifier. The SVM classifier maps the patterns to a new space, called kernel space, using a transformation $x \mapsto \varphi(x)$, in order to get a potentially better representation of them. This new space can be non-linear and of much higher dimension than the initial one. After the mapping, a linear decision boundary is computed in the kernel space. In the context of SVM methodology, the problem of classification is addressed by maximizing the margin, which is defined as the smallest distance, in the kernel space, between the decision boundary and any of the training patterns. This can be achieved by solving the following quadratic programming problem:

$$\max \left[\sum_{i=1}^{N} a_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j y_i y_j k(x_i, x_j) \right] \quad \text{over} \quad a_i,\ i = 1, \ldots, N \tag{1}$$

$$\text{s.t.} \quad 0 \le a_i \le C \quad \text{and} \quad \sum_{i=1}^{N} a_i y_i = 0 \tag{2}$$

where

$$k(x_i, x_j) = \varphi(x_i)^{\mathrm{T}} \varphi(x_j) \tag{3}$$

is the kernel function and C is a parameter controlling the trade-off between training error and model complexity. The most popular non-linear kernel functions used for SVMs belong to the class of Radial Basis Functions (RBFs). From all RBF functions, the most commonly used is the Gaussian RBF, which is defined by:

$$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \tag{4}$$

After the training of the classifier, the value of the decision function for a new pattern x is computed by:

$$y(x) = \sum_{i=1}^{N} a_i y_i k(x_i, x) + b \tag{5}$$

where b is a bias parameter the value of which can be easily determined after the solution of the optimization problem (see [12]). The value $|y(x)|$ is proportional to the distance of the input pattern x from the decision boundary. Thus, the value $y(x)$ can be regarded as a measure of confidence about the class of x, with large positive values (small negative values) strongly indicating that x belongs to the class denoted by $+1$ (-1). On the contrary, values of $y(x)$ around zero provide little information about the class of x.

In the framework of CBIR with RF, in each round of RF we have to solve a classification problem as the one described above, where a number of images, represented as feature vectors, correspond to the feedback examples provided by the user so far, and each image is labeled by -1 or $+1$ corresponding to irrelevant or relevant, respectively. The initial query is considered to be one of the relevant images and is labeled by $+1$. From the above, it is obvious that we can train an SVM classifier based on the feedback examples and use it to distinguish between the classes of relevant and irrelevant images. Each image in the database will be presented to the trained classifier and the value of the decision function (Eq. (5)) will be used as the ranking criterion. The higher the value of the decision function for an image, the more relevant this image is considered by the system.

3 Feature Selection

Assume, again, we have the binary classification problem presented above, each pattern x_i being a d-dimensional vector of features. Feature selection consists in reducing the dimensionality of the patterns, usually before training the classifier, by removing those features which are irrelevant or redundant for distinguishing between the training set categories, while keeping informative and important features. As far as the problem of re-ranking the database images can be considered as a binary classification problem, feature selection techniques can be applied in each RF round.

Specifically, in this work, we propose an RF scheme for CBIR, which uses SVMs for the RF task along with the methodology introduced in [7] for feature selection. The proposed feature selection methodology is described next. Additionally, another popular and promising methodology for feature selection is considered, which can be incorporated in the RF scheme in exactly the same way. This variation will be used for reasons of comparison with the proposed method.

3.1 Recursive Feature Elimination Using SVMs (SVM-RFE)

The feature selection methodology proposed in [7], called SVM Recursive Feature Elimination (SVM-RFE), is based on a recursive elimination of the less important features, based on the results of classification of the training patterns using SVM classifiers. Thus, the learning model used by this methodology for feature selection is the same as that adopted for the task of RF in this work. This results to the benefit of selecting those features which are the most important for the subsequent training of the SVM classifier used for RF.

Specifically, the SVM-RFE methodology is based on linear-kernel SVMs. Considering a linear SVM kernel:

$$k(x_i, x_j) = x_i^{\mathrm{T}} x_j \tag{6}$$

Eq. (5), for the decision function, takes the form:

$$y(x) = w^{\mathrm{T}} x + b \tag{7}$$

with

$$w = \sum_{i=1}^{N} a_i y_i x_i \tag{8}$$

where the vector w is of the same dimensionality as the training patterns x_i. This form of decision function implies that the higher the value $|w_k|$ or w_k^2 for the k-th coordinate of the vector w, the larger is the influence of this coordinate on the value of the decision function for an unknown pattern x. This notion provides us with a criterion which can be used to rank the image features according to their importance for the classification task.

SVM-RFE is a recursive method. In each repetition, it updates a feature set, S_f, which initially includes all the available features, by eliminating the less important feature of the set. To determine the less important feature, it trains an SVM classifier with a linear kernel, using the training patterns restricted on the features currently included in S_f. After training, the feature with the smaller value w_k^2 is considered the less important one and is eliminated from S_f. This procedure is repeated until a predefined number of features remain in S_f. These are the features selected by the method.

3.2 Minimum Redundancy Maximum Relevance (mRMR)

Another popular feature selection technique, called minimum Redundancy Maximum Relevance (mRMR), is proposed in [8]. This methodology is based on mutual information, which is a measure of relevance between two random variables x, y. The mutual information is defined by:

$$I(x;y) = \int \int p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dxdy \tag{9}$$

where $p(x)$, $p(y)$ and $p(x,y)$ are the probability density function (pdf) of x, the pdf of y and the joint pdf of x, y, respectively. In the framework of mRMR, the d pattern features and the pattern label are considered to be random variables. Under these assumptions, the task consists in the incremental selection of those features which have large relevance with the training labels and, at the same time, have small redundancy among them. This notion is expressed formally by:

$$\max H(f_j) \quad \text{over} \quad f_j \in X - S_f \tag{10}$$

with

$$H(f_j) = D(f_j) - R(f_j) \tag{11}$$

$$D(f_j) = I(f_j; y) \tag{12}$$

$$R(f_j) = \frac{1}{card(S_f)} \sum_{f_i \in S_f} I(f_j; f_i) \tag{13}$$

where f_i (or f_j), y denote the random variables corresponding to the features and the label, respectively. In this case, S_f is the set of features which have been selected until now and X is the set of all features. Initially, S_f is empty. In the first repetition, the feature $f_j \in X$ with the maximum value for $D(f_j)$, i.e. the most relevant with the labels, is selected and inserted in S_f. In each subsequent repetition, a feature f_j from the current set $X - S_f$ is inserted in S_f. This feature is one with a sufficiently large relevance with the labels, $D(f_j)$, and a sufficiently small mean relevance with the already selected features in S_f, $R(f_j)$ (which is used as a measure of redundancy), so as to maximize $H(f_j)$. This procedure is repeated until the set of selected features, S_f, contains the desired number of features.

When the random variables are discrete, the mutual information between them can be computed very easily, as the integrals in Eq. (9) are converted to summations for all the possible values of the random variables. The values of the pattern label are naturally discrete. The values of the pattern features can be easily discretized by computing for each feature the mean (μ) and the standard deviation (σ) of the values it takes for all training examples. Then, a common label is assigned for all values in each one of the intervals $(-\infty, \mu - \sigma]$, $(\mu - \sigma, \mu + \sigma]$ and $(\mu + \sigma, +\infty)$. After discretization, the probabilities needed for the mutual information computation can be determined by simply counting the corresponding instances in the training set.

4 Experiments

In order to assess the performance of the proposed method, an image set containing 3740 images from the image database in [13] is used. These images are manually classified into 17 semantic categories, and this categorization will be the ground truth of the RF simulations.

For each image, color, texture and shape information is extracted. As color features we used a 256-dimensional histogram in HSV color space, with quantization $8 \times 8 \times 4$ for the color coordinates H, S and V, respectively, and a 9-dimensional color moment vector in CIE-Lab color space, containing the first 3 moments (mean, standard deviation and skewness) for each one of the 3 color coordinates L*, a* and b*. As texture features we used the 104-dimensional vector produced by the 3-level tree-structured wavelet transform [10], which includes, for each one of the 4 sub-bands resulted by each signal decomposition, the corresponding mean and standard deviation of the energy, and is based on recursive decomposition of the first 3 sub-bands of lower frequency. Finally, as shape features we used the 80-dimensional edge histogram [11], which is formed by the uniform partitioning of the image into 16 sub-images and the computation, for each one of them, of the frequency of occurrence of 5 types of edges (horizontal, vertical, 45 degrees diagonal, etc.). All the above vectors are merged in a single 449-dimensional one, which is the final representation of the image.

We implemented an RF simulation scheme, using Precision as measure of performance, which is the ratio of relevant images within the top T retrieved images. A retrieved image is considered relevant (irrelevant) if it belongs to the same (different) category as (than) the initial query. In this simulation scheme,

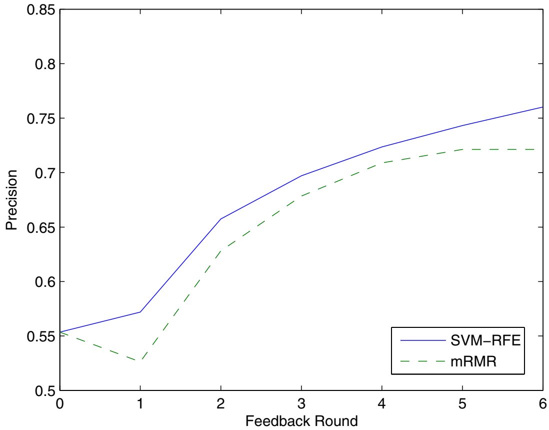

Fig. 1. Average Precision in scope $T = 20$, for $K = 25$ selected features

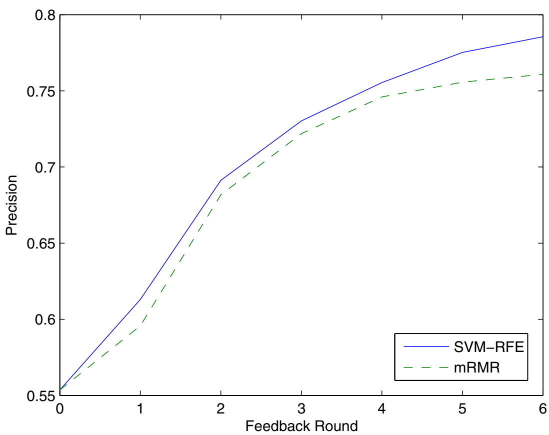

Fig. 2. Average Precision in scope $T = 20$, for $K = 50$ selected features

1000 database images are used once as initial queries. For each initial query, we simulated 6 rounds of RF. In each RF round, at most 3 relevant and 3 irrelevant images are selected randomly from the first 50 images of the ranking. These images are used in combination with the examples provided in the previous RF rounds to select a number, K, of important features and, then, to train a new SVM classifier in the resulting lower-dimensional feature space. Based on this new classifier, the ranking of the database images is updated. For the initial ranking, when no feedback examples have been provided yet and, hence, neither feature selection nor classifier training can be employed, the euclidean distance in the initial feature space is used.

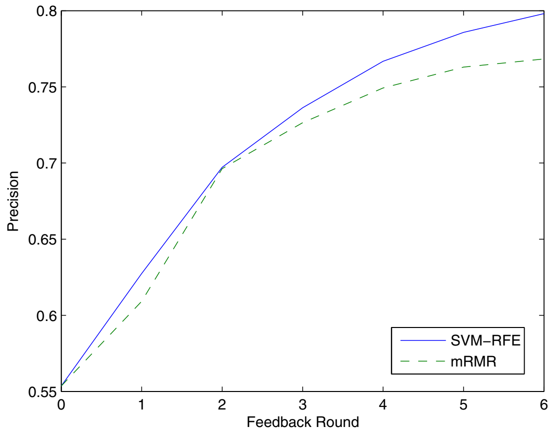

Fig. 3. Average Precision in scope $T = 20$, for $K = 75$ selected features

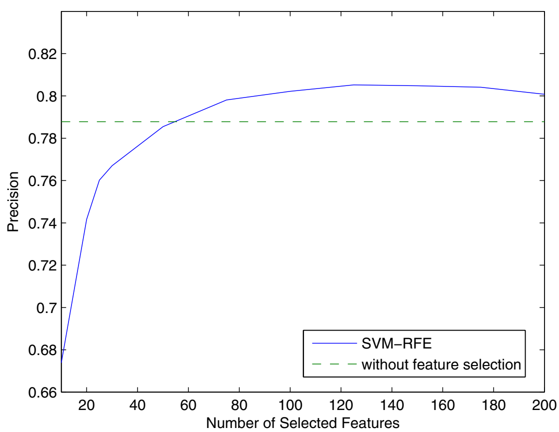

Fig. 4. Average Precision in scope $T = 20$, for the 6th RF round

The first comparison is performed between the two previously described feature selection methodologies, used along with SVMs for the task of CBIR using RF. The number of features to retain is passed as a parameter to the methods. Figures 1, 2 and 3 show the results of the RF simulations, i.e. the average Presicion in scope $T = 20$ during different RF rounds, for selected features $K = 25$, $K = 50$ and $K = 75$, respectively. As can be seen, the proposed method constantly outperforms the mRMR variation. Moreover, regarding the computation time required by the two methods, it must be mentioned that the proposed method is much faster. In particular, when $K = 50$, the average time (on a 3GHz PC) consumed by the proposed method per initial query is 2.5 sec for 6 rounds of RF, whereas the mRMR variation needs 15 sec per initial query.

In a second set of experiments, the performance of the proposed method is compared with that obtained when no feature selection is used. Specifically, Figure 4 displays the average Presicion in scope $T = 20$, for the 6th round of RF, when the proposed method is used with different values of K (e.g., 10, 25, 50, 100, 200 etc.), in comparison with that obtained without feature selection. It can be easily seen that, using a relatively small number of features, we can achieve equal or even better performance with respect to that obtained when using the full feature set.

In the experiments, a Gaussian RBF kernel is adopted for the SVM classifiers (except for the case of SVM-RFE method, which assumes linear kernel SVMs). The values of the SVM parameter C and the Gaussian RBF kernel parameter γ are empirically chosen for each experimental setup, so as to obtain the best performance. We used the SVM implementation provided in [14], and the mRMR implementation in [15].

5 Conclusions – Future Work

A new relevance feedback approach for CBIR is presented in this paper. This approach uses SVM classifiers to distinguish between the classes of relevant and irrelevant images, along with an SVM-based feature selection technique to reduce the feature space dimensionality according to the feedback examples. The experimental results demonstate the superiority of the proposed method compared to an approach based on a very popular feature selection methodology. Furthermore, as indicated by our experiments, even with a very large reduction of the features, a performance equivalent or even better compared to that obtained for the full feature set can be achieved.

In the future, we aim to use more sophisticated image features to represent the image content. Furthermore, we aim to apply techniques for automatic determination of the most appropriate number of features for each round of RF. Finally, we would like to test the scalability of the proposed method using even larger image databases.

Acknowledgement

This work was supported by Public Funds under the PENED 2003 Project co-funded by the European Social Fund (80%) and National Resources (20%) from the Hellenic Ministry of Development - General Secretariat for Research and Technology.

References

1. Datta, R., Li, J., Wang, J.Z.: Content-Based Image Retrieval: Approaches and Trends of the New Age. Multimedia Information Retrieval, 253–262 (2005)
2. Guo, G.D., Jain, A.K., Ma, W.Y., Zhang, H.J.: Learning Similarity Measure for Natural Image Retrieval with Relevance Feedback. IEEE Trans. Neural Netw. 13(4), 811–820 (2002)

3. Jing, F., Li, M., Zhang, H.-J., Zhang, B.: Relevance Feedback in Region-Based Image Retrieval. IEEE Trans. Circuits Syst. Video Technol. 14(5), 672–681 (2004)
4. Tong, S., Chang, E.: Support Vector Machine Active Learning for Image Retrieval. ACM Multimedia, 107–118 (2001)
5. Hsu, C.T., Li, C.Y.: Relevance Feedback Using Generalized Bayesian Framework with Region-Based Optimization Learning. IEEE Trans. Image Process. 14(10), 1617–1631 (2005)
6. Marakakis, A., Galatsanos, N., Likas, A., Stafylopatis, A.: Probabilistic Relevance Feedback Approach for Content-Based Image Retrieval Based on Gaussian Mixture Models. IET Image Process. 3(1), 10–25 (2009)
7. Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene Selection for Cancer Classification Using Support Vector Machines. Machine Learning 46, 389–422 (2002)
8. Peng, H., Long, F., Ding, C.: Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. IEEE Trans. Pattern Anal. Mach. Intell. 27(8), 1226–1238 (2005)
9. Wang, L.: Feature Selection with Kernel Class Separability. IEEE Trans. Pattern Anal. Mach. Intell. 30(9), 1534–1546 (2008)
10. Chang, T., Kuo, C.-C.J.: Texture Analysis and Classification with Tree-Structured Wavelet Transform. IEEE Trans. Image Process. 2(4), 429–441 (1993)
11. Won, C.S., Park, D.K., Park, S.-J.: Efficient Use of MPEG-7 Edge Histogram Descriptor. ETRI Journal 24(1), 23–30 (2002)
12. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, Heidelberg (2006)
13. Microsoft Research Cambridge Object Recognition Image Database, version 1.0, http://research.microsoft.com/research/downloads/Details/b94de342-60dc-45d0-830b-9f6eff91b301/Details.aspx
14. LIBSVM – A Library for Support Vector Machines, http://www.csie.ntu.edu.tw/~cjlin/libsvm/
15. mRMR Feature Selection Site, http://research.janelia.org/peng/proj/mRMR/