Combining Gaussian Mixture Models and Support Vector Machines for Relevance Feedback in Content Based Image Retrieval

Apostolos Marakakis[1], Nikolaos Galatsanos[2], Aristidis Likas[3], 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@cs.uoi.gr, arly@cs.uoi.gr, andreas@cs.ntua.gr

Abstract A relevance feedback (RF) approach for content based image retrieval (CBIR) is proposed, which combines Support Vector Machines (SVMs) with Gaussian Mixture (GM) models. Specifically, it constructs GM models of the image features distribution to describe the image content and trains an SVM classifier to distinguish between the relevant and irrelevant images according to the preferences of the user. The method is based on distance measures between probability density functions (pdfs), which can be computed in closed form for GM models. In particular, these distance measures are used to define a new SVM kernel function expressing the similarity between the corresponding images modeled as GMs. Using this kernel function and the user provided feedback examples, an SVM classifier is trained in each RF round, resulting in an updated ranking of the database images. Numerical experiments are presented that demonstrate the merits of the proposed relevance feedback methodology and the advantages of using GMs for image modeling in the RF framework.

1 Introduction

Content based image retrieval (CBIR) assumes an image description of automatically extracted low-level visual features, such as color, texture and shape. Using this image description, and after a user has submitted one or more query images as examples of his/her preferences, the images of an image database are ranked according to their similarity with the queries and the most similar are returned to the user as the retrieval results, e.g. [1]-[3]. Nevertheless, low-level image features

have an intrinsic difficulty in capturing the human perception of image similarity. In other words, it is very difficult to describe the semantic content of an image using only low-level image features. This is the well known in the CBIR community *semantic gap* problem.

In order to alleviate the aforementioned problem, relevance feedback (RF) has been proposed. RF is an interactive process. During a round of RF, users are required to assess the retrieved images as relevant or irrelevant to the initial query. Then, the retrieval system takes into account the user's feedback to update the ranking criterion. In recent years much work has been devoted to the RF problem for CBIR, e.g. [4]-[8], [10], [13]-[15]. The most promising approaches to this problem are based on training a classifier in each RF round, using the user provided feedback examples, e.g. [4], [7], [14], [15]. The most popular learning models used for this classification task are the Support Vector Machines (SVMs) [17].

In the proposed method, before the database images are used for CBIR, they are appropriately modeled using GMs, which are a well-established methodology to model probability density functions (pdfs), e.g. [2], [6], [9], [10], [13], [16]. This methodology is proven to have significant advantages, such as adaptability to the data, modeling flexibility and robustness. The main challenge when using GMs for CBIR is to define a distance measure between GMs which, in addition to quantifying well the difference of GM models, can be computed efficiently. The traditionally used distance measure between pdfs is the Kullback-Leibler (KL) divergence that cannot be computed in closed form for GM models. Thus, one has to resort to time consuming random sampling Monte-Carlo methods to compute this measure for GMs, which makes its use impractical for CBIR. In [18], a new distance measure was introduced, based on the KL divergence, which can be computed in closed form for GMs. Moreover, in [12], the Asymptotic Likelihood Approximation (ALA) was proposed as a measure which, under certain assumptions, approximates the KL divergence and can also be computed in closed form for GMs.

The rest of this paper is organized as follows. In Section 2, we describe GMs in the context of image modeling for CBIR. In Section 3, we present the approximations of the KL divergence which are used in this work. In Section 4, we describe the SVM methodology for binary classification. In Section 5, we present the proposed SVM kernel functions for classification of GMs. In Section 6, we provide the details and the results of the experiments. Finally, in Section 7, we present conclusions and directions for future research.

2 Using GM Models for CBIR

GM models have been used extensively in many data modeling applications. Furthermore, they have already been used in CBIR as probability density models of the features that are used to describe images, e.g. [2], [9], [13]. In this framework,

each image is described as a bag of feature vectors which are computed locally (e.g. a feature vector for each pixel or region of the image). This bag of feature vectors is subsequently used to train, in a maximum likelihood manner, a GM that models the probability density of the image features in the feature space. A GM model for the image feature vectors $x \in R^d$ is defined as

$$p(x) = \sum_{j=1}^{K} \pi_j \phi(x \mid \theta_j) \tag{1}$$

$$\theta_j = (\mu_j, \Sigma_j) \tag{2}$$

$$\phi(x \mid \theta_j) = N(x \mid \theta_j) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_j|}} e^{-\frac{1}{2}(x-\mu_j)^T \Sigma_j^{-1}(x-\mu_j)} \tag{3}$$

where K is the number of Gaussian components in the model, $0 \le \pi_j \le 1$ the mixing probabilities with $\sum_{j=1}^{K} \pi_j = 1$, and $\phi(x \mid \theta_j)$ a Gaussian pdf with mean μ_j and covariance Σ_j.

In order to retrieve images from an image database, a distance measure between the image models is needed. The KL divergence cannot be computed analytically for two GMs. Thus, for efficient retrieval using GM models, one has to resort to approximations such as those discussed next.

3 Approximations of the KL Divergence

A distance measure between images represented as GMs, which will be used for CBIR, must have good separation properties and must allow fast computation. This imposes the requirement that the distance can be defined in closed form for the case of GMs, which is not easy to achieve. In this spirit, several distance measures have been proposed, with the aim to address these requirements.

The distance measure introduced in [18] is adapted to the case of mixture models. It is known that the KL divergence between two Gaussian pdfs can be computed in closed form. In particular,

$$KL(\phi(x \mid \theta_1) \| \phi(x \mid \theta_2)) = \frac{1}{2}\left(tr(\Sigma_2^{-1}\Sigma_1) - \log\frac{|\Sigma_1|}{|\Sigma_2|} - d \right) + \frac{1}{2}(\mu_1 - \mu_2)^T \Sigma_2^{-1}(\mu_1 - \mu_2) \tag{4}$$

Based on this fact and assuming that we have two GMs $p_1(x)=\sum_{j=1}^{K_1}\pi_{1j}\phi(x\mid\theta_{1j})$ and $p_2(x)=\sum_{m=1}^{K_2}\pi_{2m}\phi(x\mid\theta_{2m})$, an approximation of the KL divergence between them can be defined using the KL divergence between the Gaussian components of the mixtures:

$$KL_{gkl}(p_1\parallel p_2)=\sum_{j=1}^{K_1}\pi_{1j}\min_{m=1\ldots K_2} KL(\phi(x\mid\theta_{1j})\parallel\phi(x\mid\theta_{2m})) \quad (5)$$

Using the above definition, we can introduce the symmetric version of this distance measure as

$$SKL_{gkl}(p_1,p_2)=\frac{1}{2}KL_{gkl}(p_1\parallel p_2)+\frac{1}{2}KL_{gkl}(p_2\parallel p_1) \quad (6)$$

In [12], the Asymptotic Likelihood Approximation (ALA) was proposed as a similarity measure for GMs. In particular, for the same GMs $p_1(x)$ and $p_2(x)$ as above, the ALA measure can be computed as

$$ALA(p_1\parallel p_2)=\sum_{j=1}^{K_1}\pi_{1j}\left\{\log\pi_{2\beta(j)}+\left[\log\phi(\mu_{1j}\mid\theta_{2\beta(j)})-\frac{1}{2}tr(\Sigma_{2\beta(j)}^{-1}\Sigma_{1j})\right]\right\} \quad (7)$$

$$\|x-\mu\|_{\Sigma}^2=(x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu) \quad (8)$$

$$\beta(j)=k\Leftrightarrow\|\mu_{1j}-\mu_{2k}\|_{\Sigma_{2k}}^2-\log\pi_{2\kappa}<\|\mu_{1j}-\mu_{2l}\|_{\Sigma_{2l}}^2-\log\pi_{2l},\ \ \forall l\neq k \quad (9)$$

In [12], it is proven that under certain assumptions $ALA(p_1\parallel p_1)-ALA(p_1\parallel p_2)$ approximates $KL(p_1\parallel p_2)$. Thus one can define

$$KL_{ala}(p_1\parallel p_2)=ALA(p_1\parallel p_1)-ALA(p_1\parallel p_2) \quad (10)$$

and

$$SKL_{ala}(p_1,p_2)=\frac{1}{2}KL_{ala}(p_1\parallel p_2)+\frac{1}{2}KL_{ala}(p_2\parallel p_1) \quad (11)$$

as approximations of the KL and the symmetric KL divergence, respectively.

A careful inspection of the measures defined in [18] and [12] (Eq. (5) and Eq. (7)-(9), respectively) shows that they have several similarities. For example, both are based on the computation of a correspondence between the Gaussian components of the two mixtures. Moreover, the final value of these measures for GMs is given by the convex combination (using the mixing weights π_{1j}) of some pairwise measures between Gaussian components.

4 Support Vector Machines

Consider the binary classification problem $\{(x_i, y_i)\}_{i=1}^{N}$ with $y_i \in \{-1,+1\}$ and x_i the labeled patterns based on which we want to train the SVM classifier. The patterns are mapped to a new space, called kernel space, which can be non-linear and of much higher dimension than the initial one, using a transformation $x \mapsto \phi(x)$. Then a linear decision boundary is computed in the kernel space. The SVM methodology addresses the problem of classification by maximizing the margin, which is defined as the smallest distance in the kernel space between the decision boundary and any of the samples. This can be achieved by solving a quadratic programming problem:

$$\max_{a=(a_1,\ldots,a_N)^{\mathrm{T}}} \sum_{i=1}^{N} a_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} a_i a_j y_i y_j k(x_i, x_j) \tag{12}$$

$$s.t.\ 0 \le a_i \le C \text{ and } \sum_{i=1}^{N} a_i y_i = 0 \tag{13}$$

where

$$k(x_i, x_j) = \phi^{\mathrm{T}}(x_i)\phi(x_j) \tag{14}$$

is the kernel function and C is a parameter controlling the trade-off between training error and model complexity. Then, the decision function for a new pattern x is defined by

$$y(x) = \sum_{i=1}^{N} a_i y_i k(x, x_i) + b \tag{15}$$

where b is a bias parameter the value of which can be easily determined after the solution of the optimization problem (see [17]). After training, 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").

It is obvious that the patterns under classification need not be in vectorial form, but they can be any data objects for which an appropriate kernel function expressing their pair-wise similarity can be defined.

5 Combining SVMs with GM Models

In the framework of CBIR with RF, and assuming that we model each image using a GM, in each round of RF we have a number of images, represented as GMs, which correspond to the feedback examples provided by the user until now. Each

of these images is labeled by -1 or +1 in case the user considers it as irrelevant or relevant to the initial query, respectively. The initial query is considered to be one of the relevant images and is labeled by +1, of course. The aim of the task is to train an SVM classifier to distinguish between the classes of relevant and irrelevant images.

As mentioned above, the kernel function is defined as the inner product of the patterns in the kernel space (Eq. (14)), namely, it is a similarity measure. The most popular non-linear kernel functions used for SVMs belong to the class of Radial Basis Functions (RBFs). From all RBF functions, now, the most commonly used are the Gaussian, $\exp\left(-\gamma\|x-y\|^2\right)$, and the Laplacian, $\exp\left(-\gamma\|x-y\|\right)$. A straightforward generalization of this concept in the GM framework, is to use as kernel function between GMs a function of the form $\exp\left(-\gamma d\left(p,q\right)\right)$, where $p,\ q$ are two GMs and $d\left(p,q\right)$ is a distance measure between them.

The distances presented in Section 3 fulfill the requirement for effective separation and closed-form computation. Thus, based on the above considerations, we can define the functions

$$k_{gkl}\left(p,q\right)=\exp\left(-\gamma SKL_{gkl}\left(p,q\right)\right) \tag{16}$$

$$k_{ala}\left(p,q\right)=\exp\left(-\gamma SKL_{ala}\left(p,q\right)\right) \tag{17}$$

as kernel functions between GMs *p* and *q*.

After the SVM classifier has been trained, each image in the database is presented to the classifier and the value of the decision function (Eq. (15)) is 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.

6 Experiments

In order to test the validity of the proposed method, an image set containing 3740 images from the image database in [19] is used. These images are classified in 17 semantic categories. This categorization corresponds to the ground truth.

To model each image, several features are extracted including position, color and texture information. As position features we use the pixel coordinates, as color features we use the 3 color coordinates (L*,a*,b*) in the CIE-Lab color space and as texture features we use the contrast (c), the product of anisotropy with contrast (ac) and the product of polarity with contrast (pc) as described in [9].

Consequently, for each image a set of feature vectors is extracted, which is subsequently used as input to the Greedy EM algorithm [11] to produce a GM model of the image features distribution. For all GM models, 10 Gaussian components are adopted, and each Gaussian component is assumed to have full covariance matrix.

For reasons of comparison, we also applied the SVM-RF approach using the same image feature sets but the standard Gaussian RBF function, which is the most commonly used kernel function for SVMs. This kernel function requires a global vectorial representation of the images. Thus, in this case, we represent each image by the joint position-color and position-texture histogram. The position-color histogram consists of 3x3x4x8x8 (x-y-L*-a*-b*) bins, whereas the position-texture histogram consists of 3x3x4x4x4 (x-y-ac-pc-c) bins.

In order to quantify the performance of the compared methods, we implemented an RF simulation scheme. As a measure of performance we use Precision which is the ratio of relevant images in top N retrieved images. An image is assessed to be relevant or irrelevant according to the ground truth categorization of the image database. In this simulation scheme, 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 100 images of the ranking. These images are used in combination with the examples provided in the previous RF rounds to train a new SVM classifier. Based on this new classifier, the ranking of the database images is updated.

For the experiments presented below, average Precision in scope N = 10, 20, 30 is shown. The values of the SVM parameter C and of the kernel parameter γ are empirically chosen for each method so as to obtain the best performance. As SVM implementation we used the one provided in [20].

In Figures 1-3 we can see that the SVM-RF method based on GMs constantly outperforms the common SVM-RF method which uses histograms and the Gaussian RBF kernel function. Moreover, it can be observed that the method which is based on the distance measure defined in [18] results in slightly superior performance when compared to that obtained by the method which uses the ALA based distance measure.

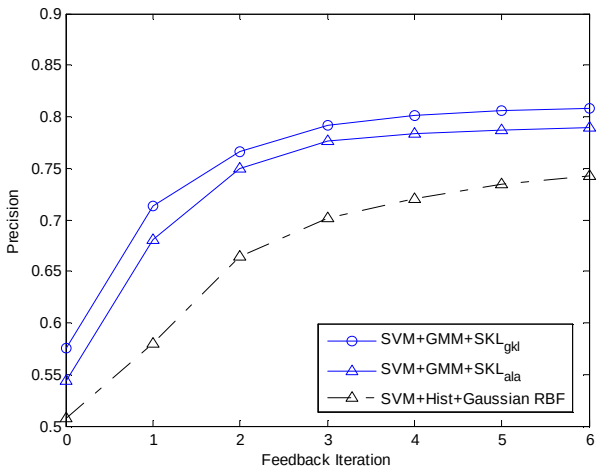

Fig. 1. Average Precision in scope N = 10 during different rounds of RF

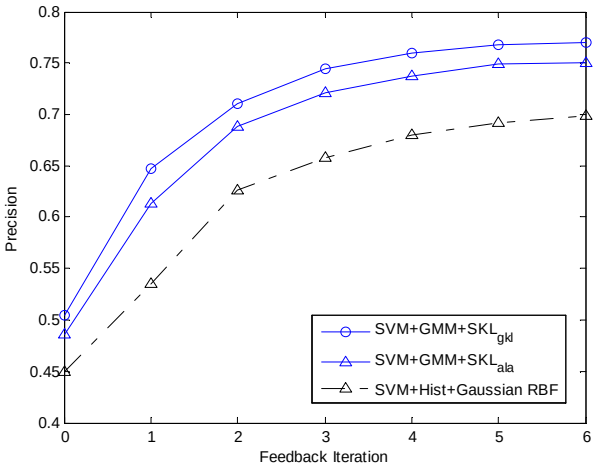

Fig. 2. Average Precision in scope N = 20 during different rounds of RF

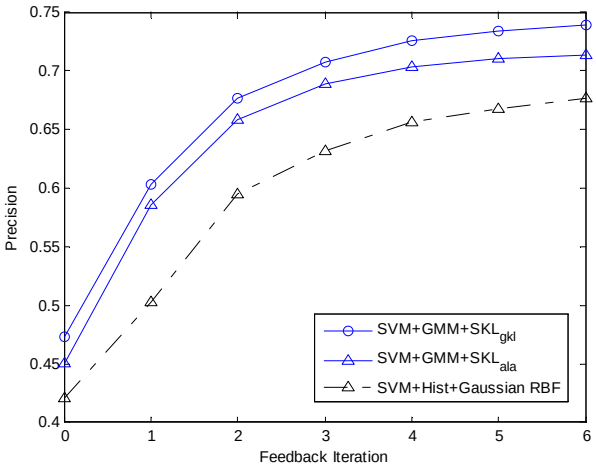

Fig. 3. Average Precision in scope N = 30 during different rounds of RF

7 Conclusions – Future Work

A new relevance feedback approach for CBIR is presented in this paper. This approach uses GMs to model the image content and SVM classifiers to distinguish between the classes of relevant and irrelevant images. To combine these method-

ologies, a new SVM kernel function is introduced based on distance measures between GMs which can be computed efficiently, i.e. in closed form. The main advantages of the proposed methodology are accuracy as indicated by our experimental results, speed, due to the distance measures used, and flexibility. As indicated by our experiments, very promising results can be obtained using GMs as SVM patterns, even if we are forced to use an approximation and not the exact KL divergence. In particular, for the two KL approximations tested, the performance does not differ significantly. However, the distance measure introduced in [18] gives slightly better results.

In the future, we would like to adapt and test our method using other efficiently computable distance measures for GMs. Moreover, we aim to use more sophisticated image features to represent the image content. In addition, we plan to generalize our RF scheme to support region-based image descriptions. Furthermore, we aim to apply techniques for automatic determination of the appropriate number of components for each GM. 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. Y. Ishikawa, R. Subramanya, and C. Faloutsos, "MindReader: Querying databases through multiple examples", *Proceedings International Conf*erence *on Very large Data Bases (VLDB)*, 1998.
2. N. Vasconcelos, "Minimum Probability of Error Image Retrieval", *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2322-2336, Aug. 2004.
3. Ritendra Datta, Jia Li and James Ze Wang, "Content-based image retrieval: approaches and trends of the new age", *Multimedia Information Retrieval*, pp. 253-262, 2005.
4. G. D. Guo, A. K. Jain, W. Y. Ma, and H. J. Zhang, "Learning similarity measure for natural image retrieval with relevance feedback", *IEEE Transactions on Neural Networks*., vol. 13, no. 4, pp. 811-820, Jul. 2002.
5. C.T. Hsu, and C. Y. Li, "Relevance Feedback Using Generalized Bayesian Framework With Region-Based Optimization Learning", *IEEE Transactions on Image Processing*., Vol. 14, No. 10, pp. 1617-1631, October 2005.
6. F. Qian, M. Li, L. Zhang, H. J. Zhang, and B. Zhang, "Gaussian mixture model for relevance feedback in image retrieval", *Proceedings IEEE ICME*, Aug. 2002.
7. F. Jing, M. Li, H-J. Zhang, and B. Zhang, "Relevance Feedback in Region-Based Image Retrieval", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 5, pp. 672-681, May 2004.
8. Wei Jiang, Guihua Er, Qionghai Dai and Jinwei Gu, "Similarity-Based Online Feature Selection in Content-Based Image Retrieval", *IEEE Transactions on Image Processing*, vol. 15, no. 3, pp. 702-712, March 2006.

9. C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image segmentation using expectation-maximization and its application to image querying", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1026-1038, Aug. 2002.
10. A. Marakakis, N. Galatsanos, A. Likas and A. Stafylopatis, "A Relevance Feedback Approach for Content Based Image Retrieval Using Gaussian Mixture Models", *Proceedings International Conference Artificial Neural Networks* (*ICANN*), Athens, Greece, September 2006.
11. N. Vlassis and A. Likas, "A greedy EM algorithm for Gaussian mixture learning", *Neural Processing Letters*, vol. 15, pp. 77-87, 2002.
12. N. Vasconcelos, "On the Efficient Evaluation of Probabilistic Similarity Functions for Image Retrieval", *IEEE Transactions on Information Theory*, vol. 50, no. 7, pp. 1482-1496, July 2004.
13. N. Vasconcelos and A. Lippman, "Learning from user feedback in image retrieval systems", *Advances in Neural Information Processing Systems*, 1999.
14. S. Tong and E. Chang, "Support vector machine active learning for image retrieval", *ACM Multimedia*, 2001.
15 Dacheng Tao, Xiaoou Tang and Xuelong Li, "Which Components Are Important for Interactive Image Searching?", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 1, pp. 3-11, Jan. 2008.
16. C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford Univ. Press Inc., New York, 1995.
17. C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
18. Jacob Goldberger and Sam Roweis, "Hierarchical Clustering of a Mixture Model", *Neural Information Processing Systems 17 (NIPS'04)*, pp 505-512, 2004.
19. Microsoft Research Cambridge Object Recognition Image Database, version 1.0. http://research.microsoft.com/research/downloads/Details/b94de342-60dc-45d0-830b-9f6eff91b301/Details.aspx
20. LIBSVM – A Library for Support Vector Machines. http://www.csie.ntu.edu.tw/~cjlin/libsvm