International Journal of Natural and Engineering Sciences 7 (1): 23-25, 2013
ISSN: 1307-1149, E-ISSN: 2146-0086, www.nobel.gen.tr

Effective Frame work for Hierarchical Indexing Scheme using Expectation Maximization based on Full Automatic Algorithm

Shahin SHAFEI[1*] Tohid SEDGHI[2]
[1] Department of Electrical Engineering, Mahabad Branch, Islamic Azad University, Mahabad, Iran
[2] Department of Electrical Engineering, Urmia Branch, Islamic Azad University, Urmia, Iran

*Corresponding author:
E-mail: Shahin_shafei1987@yahoo.com

Received: July 03, 2012
Accepted: August 12, 2012

Abstract
The intersection method had a higher performance as shown by the ROC curves in our paper. We extended the EM-variant algorithm to model each object as a Gaussian mixture, and the EM-variant extension outperforms the original EM-variant on the image data set having generalized labels. Intersecting abstract regions was the winner in our experiments on combining two different types of abstract regions. However, one issue is the tiny regions generated after intersection. The problem gets more serious if more types of abstract regions are applied. Another issue is the correctness of doing so. In some situations, it may be not appropriate to intersect abstract regions. For example, a line structure region corresponding to a building will be broken into pieces if intersected with a color region. In future works, we attack these issues with two phase approach the classification problem.

Keywords: correctness, regions, Local descriptors, Gaussian mixture.

INTRODUCTION

In the web page of the Viper project, a framework to evaluate the performance of CBIR systems, about 70 academic systems and 11 commercial systems are listed. Prominent systems include [1], [2],[3]. In the CBIR, only a small number of researchers have worked on retrieval via object recognition and many of these efforts have been limited to a single class of object, such as people or horses. The SIMPLIcity system extracts features by a wavelet-based approach and compares images using a region-matching scheme. It classifies images into categories, such as textured or nontextured, graphic or non-graphic. Barnard and Forsyth [1] utilize a generative hierarchical model to automatically annotate images. Duygulu et al. [3] classifies image regions as blobs and finds the relationship between blobs and annotations as a machine translation problem. Jeon et al. [4] from University of Massachusetts uses cross-media relevance models to learn the translation between blobs and words. In ALIP [4] concepts are modeled by a two-dimensional multi-resolution hidden Markov model. Color features and texture features based on small rigid blocks are extracted. A new and very promising approach to object classes [5] models objects classes as flexible configurations of parts, where the parts are merely square regions selected by entropy- based feature detector [5]; a Bayesian classifier is used for the final recognition task. Image annotation has received a lot of recent attention. Maron and Ratan [5] formalized the image annotation problem as a multiple-instance learning model [2]. Duygulu et al. [3] described their model as machine translation. One problem with both of these approaches is the assumption of a one-to-one mapping between image regions and objects, which is not always true. Instead, some objects span multiple regions, and some regions contain multiple objects. For the same reason, these approaches cannot use context information to assist in recognition. Yet context is an important cue that is often very helpful. The fundamental difference between these approaches and ours is that they map a point in feature space to the target object, while we map a set of points in feature space to the target. In the SIMPLIcity system, the authors recognized the problem with one-to-one mappings and solved it with an approach called “integrated region matching,” which measures the similarity between two images by integrating properties of all regions in the images. This approach takes all the regions within an image into account, which can bring in regions that are not related to the target object. Our approach first discovers which regions are related to the target object and makes its decision based on those regions. Clearly there is no single feature suitable for all object recognition tasks. A robust system should be able to combine the power of many different features to recognize many different objects. Carson et al. [1] and Berman and Shapiro [2] provide sets of different features and allow users to adjust their weights, which passes the burden of feature selection to the user. In Wang et al. [1],The

objective is to develop a technique which captures local texture descriptors in a coarse segmentation framework of grids. The new method has a shape descriptor in terms of invariant moments computed on the edge image. The image is partitioned into different sizes of non-overlapping tiles. A new framework is used for texture analysis. The features computed on these tiles serve as local descriptors of texture. Invariant moments are used to serve as shape features. The combination of these features forms a robust feature set in retrieving applications. Then, an integrated matching procedure based on the adjacency matrix of a bipartite graph between the image tiles is provided, similar to the one discussed in [1], which yields image similarity. Our method is similar to IRM, but so simple and less time consumer since all of the RBIR systems are complicated due to using different kinds of complicated algorithms which makes them be time consumption. We note that in any CBIR system, fast retrieval is the main objective. The experimental results are compared with the methods of [1], [2], [3], and [4]. The results indicate that the new method performs better. We developed a new semi-supervised EM-like algorithm that is given the set of objects present in each training image, but does not know which regions correspond to which objects. We have tested the algorithm on a dataset of 860 hand-labeled color images using only color and texture features, and the results show that our EM variant is able to break the symmetry in the initial solution. We compared two different methods of combining different types of abstract regions, one that keeps them independent and one that intersects them.

RESULTS AND DISCUSSION

Let T be the set of training images and O be a set of m object classes. Suppose that we have a particular type a of abstract region and that this type of region has a set of n^a attributes which have numeric values. Then any instance of region type a can be represented by a feature vector of values $r^a = (v_1, v_2, ..., v_{n^a})$. Each image I is represented by a set F_I^a of type a region feature vectors. Furthermore, associated with each training image I $\in$ T is a set of object labels O_I, which gives the name of each object present in I. Finally, associated with each object o is the set $R_o^a = \bigcup_{I:o \in O_I} F_I^a$, I, the set of all type a regions in training images that contain object class o. Our approach assumes that each image is a set of regions, each of which can be modeled as a mixture of multi-variate Gaussian distributions. We assume that the feature distribution of each object o within a region is a Gaussian $N_o(\mu_o, \sum_o)$, o $\in O$ and that the region feature distribution is a mixture of these Gaussians. We have developed a variant of the EM algorithm to estimate the parameters of the Gaussians. Our variant is interesting for several reasons. First, we keep fixed the component responsibilities to the object priors computed over all images. Secondly, when estimating the parameters of the Gaussian mixture for a region, we utilize only the list of objects that are present in an image. We have no information on the correspondence between image regions and object classes. The vector of parameters to be learned is:

$$\lambda = (\mu_{o1}^a, ..., \mu_{om}^a, \mu_{bg}^a, \sum_{o1}^{a}, ..., \sum_{om}^{a}, \sum_{bg}^{a}) \quad (1)$$

where $\{\mu_{oi}^a, \sum_{oi}^{a}\}$ are the parameters of the Gaussian for the ith object class and $\{\mu_{bg}^a, \sum_{bg}^{a}\}$ are the parameters of an additional Gaussian for the background.

Users of commercial CBIR systems prefer to pose their queries in terms of key words. To help automate the indexing process, we represent images as sets of feature vectors of multiple types of abstract regions, which come from various segmentation processes. With this representation, we have developed an algorithm to recognize classes of objects and concepts in outdoor scenes. We have developed a new method for object recognition that uses whole images of abstract regions, rather than single regions for classification. The comparison of experimental results of the proposed method with the other retrieval systems reported in the literature [1], [2], and [3] is presented in Table 1.

Suppose that the outer mixture has (m+ 1) components and that the outer EM algorithm converges after i iterations. The inner mixtures require re-estimation for each of the I iterations. If the number of components of the inner Gaussian mixtures is m′, then there are i × m m′-component inner Gaussian mixtures plus one (m + 1)-component complex outer mixture to calculate, which is much heavier work than that of the original EM-variant. The aligned Gaussian parameters are chosen for the EM-variant extension to relieve the system burden. The other objective of using aligned Gaussian parameters is to reduce the number of parameters to learn. Suppose the feature vectors are d-dimensional. For each Gaussian component, there are d^2 parameters for the covariance matrix, d for the mean, and 1 for its probability. Thus with general Gaussian parameters, the original EM-variant has (m+1) ×(d^2 +d + 1) parameters to learn. Using general Gaussian parameters with the EM-variant extension, there are (m + 1) × [m′ × (d^2 + d + 1) + 1] parameters to learn, and the number is roughly m′ times of that of the original EM-variant. Having more parameters means a higher likelihood of over fitting unless a large number of training samples are provided. Therefore, we chose aligned Gaussian parameters for the EM-variant extension, and the number of parameters reduces to (m + 1) × [m′ × (2 × d + 1) + 1]. We performed a series of experiments to explore the effect of the parameter m′, the number of components of the inner Gaussian mixtures, on the performance. The ROC scores of experiments with different value of m′ are shown in Figure 1. In the figure, the ROC score of the original EM-variant is also plotted for comparison.

It shows that when m′ is less than 4, the performance of the EM-variant extension is worse than the EM-variant and this suggests that for this particular task, using a mixture of a few Gaussians with the aligned Gaussian parameters to model a object is not as good as just using a single Gaussian with the general Gaussian parameters. When m′ increases, the performance of the EM-variant extension outperforms the original EM-variant. The ROC scores settle at a level between

85% and 86% when m′ is greater than 10, which is about 2.4% higher than that of the original EM-variant.

It is worth mentioning that having a fixed m′ is not the best solution. Although the major trend shows that the higher the value of m′, the better the performance, a bigger m′ does not always lead to a better performance, since the quality of the clustering also plays an important role here. It is better to have a smart clustering algorithm to adaptively calculate m′ for different objects and to discover the optimal clusters. This task is challenging and deserves more research by itself. The ROC scores for individual objects for the original EM-variant and the EM-variant extension with m′ set to 12 are listed in Table 1. The average score on the ten labels for the original EM-variant with single Gaussian models was 82.6%; while the average score for the EM-variant extension was 86.0%.

Furthermore, if only the labels of combined classes are considered, the EM-variant extension approach achieved a score of 83.1%, about 5% higher than that of the EM-variant approach, which achieved a score of 78.2%.

Table 1. ROC Scores for EM-variant with single Gaussian models and EM-variant extension with 12-component Gaussian mixture for each object.

	EM varient (%)	EM Varient extension (%)
African animal	71.8	86.1
Arctic	80	82.9
Beach	88	93.2
Grass	76.9	67.7
Mountaions	94.0	96.3
Primate	74.4	86.7
Sky	91.9	84.8
Stadium	95.2	98.4
Tree	70.7	76.6
Water	82.9	87.1
Mean	82.6	86
Mean of Combined Class	78.2	83.1

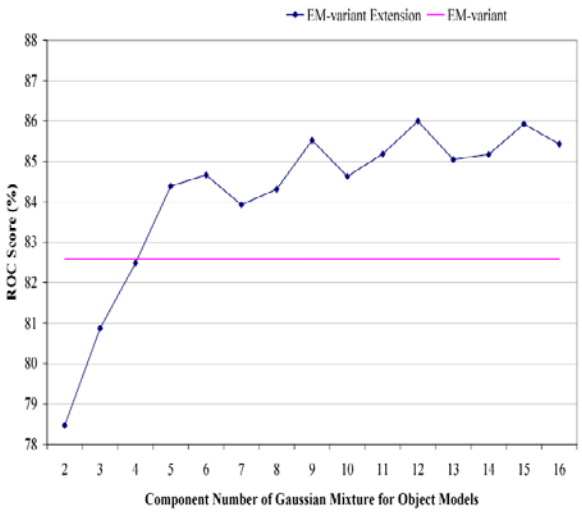

Figure 1. The ROC scores of experiments with different value of the parameter, m′, the component number of Gaussian mixture for each object model.

REFERENCES

[1] T. Gevers and A.W.M. Smeuiders., "Combining color and shape invariant features for image retrieval", Image and Vision computing, vol.17(7),pp. 475-488 , 2009. [2] M.Banerjee, M,K,Kundu and P.K.Das, "Image Retrieval with Visually Prominent Features using Fuzzy set theoretic Evaluation", ICVGIP, India, Dec 2011.

[3] D.Hoiem, R. Sukhtankar, H. Schneiderman, and L.Huston, "Object-Based Image retrieval Using Statistical structure of images", Proc CVPR, 2011.

[4] P. Howarth and S. Ruger, "Robust texture features for still-image retrieval", IEE. Proceedings of Visual Image Signal Processing, Vol. 152, No. 6, December 2009.

[5] Dengsheng Zhang, Guojun Lu, "Review of shape representation and description techniques", Pattern Recognition Vol. 37,pp 1-19, 2010.