International Journal of Natural and Engineering Sciences 5 (1): 19-25, 2011
ISSN: 1307-1149, E-ISSN: 2146-0086, www.nobel.gen.tr

Region and Object Based Image Retrieval Technique Using Textural and Color Expectation Maximization Method

Majid FAKHERI[2*] Mehdi Chehel AMIRANI[1] Tohid SEDGHI[2]
[1]Faculty of Engineering, Urmia University, Urmia, IRAN
[2]Young Researchers Club, Islamic Azad University, Urmia Branch, Urmia, IRAN

***Corresponding Author**
e-mail: st_m.fakheri@urmia.ac.ir

Received : July 04, 2010
Accepted : October 17, 2010

Abstract

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. A key part of our approach is that we do not need to know where in each image the objects lie. We only utilize the fact that objects exist in an image, not where they are located. We have designed an EM-like procedure that learns multivariate Gaussian models for object classes based on the attributes of abstract regions from multiple segmentations of color images. The objective of this algorithm is to produce a distribution for each of the object classes being learned. It uses the label information from training images to supervise EM-like iterations.

Key words: Content based, EM Algorithm, Gaussian models, Gaussian mixture, segmentation, image retrieval,

INTRODUCTION

In recent years, the computer vision community has started to tackle more general, more difficult recognition algorithms using a number of techniques that have been developed over the years. Techniques that use the appearance of an object in its images, instead of its 3D structure, are called appearance-based object recognition techniques [1]. The current limitations of these techniques are that they expect the image to consist of, or be limited to, the object in question and that this object must be presented from the same viewpoint as the images used to train the system. Appearance-based techniques have been able to yield high recognition accuracy in limited domains. Appearance-based techniques do not attempt to segment the image; this is both strength and a weakness of the approach. Region-based techniques [2] require pre segmentation of the image into regions of interest. In most applications, the reliability of image segmentation techniques has been a problem for object recognition, but image segmentation algorithms [3] that use both color and texture can now partition an image into regions that, in many cases, can be identified as having the right colors and textural pattern to be a tiger or a zebra or some other object with a well-known color-texture signature. Related to this approach are algorithms that look for regions in color-texture space that correspond to particular materials, such as human flesh [4]. A different set of color criteria and spatial region relationships can be used to find horses [5]. People's faces have also been successfully detected using only gray-tone features and relying on heavily-trained neural net classifiers [6]. In fact, neural nets and support-vector machines have become an important tool in recognizing several different specific classes of imagery. CBIR has become increasingly popular in the past 10 years. In a publication [7] by Smeulders et al. more than 200 references are reviewed. 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 VISUALSEEK [8], WebSEEK, [9], BLOBWORLD [2], and SIMPLIcity [10]. 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 [11] utilize a generative hierarchical model to automatically annotate images. Duygulu et al. [12] classifies image regions as blobs and finds the relationship between blobs and annotations as a machine translation problem. Jeon et al. [13] from University of Massachusetts uses cross-media relevance models to learn the translation between blobs and words. In ALIP [14] 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 [15] models objects classes as flexible configurations of parts, where the parts are merely square regions selected by entropy- based feature detector [16]; a Bayesian classifier is used for the final recognition task. Image annotation has received a lot of recent attention. Maron and Ratan [17] formalized the image annotation problem as a multiple-instance learning model [18]. Duygulu et al. [12] 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. [19] and Berman and Shapiro [20] 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. [10], the feature set is determined empirically by the developer. Our system learns the best weights for combining different features to recognize different objects. For the most part, generic object recognition efforts have been standalone. There is not yet a unified methodology for generic object class recognition or for concept class recognition [21]. The development of such a methodology is the subject of our research. In Section II we formalize this approach, in Section III we describe our experiments and results, and an extension of this approach aiming at recognizing object classes with different appearances.

FORMALIZING THE APPROACH

Initialization phase of the EM-variant approach, each object is modeled as a Gaussian component, and the weight of each component is set to the frequency of the corresponding object class in the training set. Each object model is initialized using the feature vectors of all the regions in all the training images that contain the particular object, even though there may be regions in those images that do not contribute to that object. From these initial estimates, which are full of errors, the procedure iteratively re estimates the parameters to be learned. The iteration procedure is also supervised by the label information, so that a feature vector only contributes to those Gaussian components representing objects present in its training image. The resultant components represent the learned object classes and one background class that accumulate the information from feature vectors of other objects or noise. With the Gaussian components, the probability that an object class appears in a test image can be computed. This part describes the EM-variant approach and illustrates its use with color and texture regions. We are given a set of training images, each containing one or more object classes, such as grass, trees, sky, houses, zebras, and so on. Each training image comes with a list of the object classes that can be seen in that image. There is no indication of where the objects appear in the images. We would like to develop classifiers that can train on the features of the abstract regions extracted from these images and learn to determine if a given class of object is present in an image.

Single-Feature Case

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 eachobject 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. The purpose of the extra model is to absorb the features of regions that do not fit well into any of the object models, instead of allowing them to contribute to, and thus bias, the true object models. The label bg is added to the set O_I of object labels of each training image I and is thus treated just like the other labels. The initialization step, rather than assigning random values to the parameters, uses the label sets of the training images. For object class $o \in O$ and feature type a, the initial values are:

$$\mu_o^a = \frac{\sum_{r^a \in R_o^a} r^a}{|R_o^a|} \tag{2}$$

$$\sum_o^a = \frac{\sum_{r^a \in R_o^a} [r^a - \mu_o^a \mathbf{I}\ r^a - \mu_0^a]^T}{|R_o^a|} \tag{3}$$

Note that the initial means and covariance matrices most certainly have errors. For example, the Gaussian mean for an object in a region is composed of the average feature vector over all regions in all images that contain that object. This property will allow subsequent iterations by EM to move the parameters closer to where they should be. Moreover, by having each mean close to its true object, each such subsequent iteration should reduce the strength of the errors assigned to each parameter. In the E-step of the EM algorithm, we calculate:

$$p(r^a \mid o, \mu_o^a(t) \sum_o^a(t)) = \begin{cases} 0 \\ \dfrac{1}{\sqrt{(2\pi)^{n^a} \left|\sum^a(t)\right|}} e^{-\frac{1}{2}(r^a - \mu_o^a(t)^T (\sum_o^a(t))^{-1}(r^a - \mu_o^a(t)} & otherwise \end{cases} \tag{4}$$

$$p(o \mid r^a, \lambda(t)) = \frac{p(r^a \mid 0, \mu_o^a(t), \sum_o^a(t))\, p(o)}{\sum_{j \in O_I} p(r^a \mid j, \mu_j^a(t), \sum_j^a(t))\, p(j)} \tag{5}$$

where

$$p(o) = \frac{|\{I \mid o \in O_I\}|}{|T|} \tag{6}$$

Note that when calculating $p(r^a \mid 0, \mu_o^a(t) \sum^a(t))$ in (4) for region vector r_a of image I and object class o and when normalizing in (5), we use only the set of object classes of O_I, which are known to be present in I. The M-step follows the usual EM process of updating μ_o^a and $\sum_0^a$:

$$\mu_o^a(t+1) = \frac{\sum_{r^a} p(o \mid r^a, \lambda(t)\ r^a}{\sum_{r^a} p(o \mid r^a, \lambda(t)} \tag{7}$$

$$\sum_0^a(t+1) = \frac{\sum_{r^a} p(o \mid r^a, \lambda(t)\ [r^a - \mu_0^a(t+1)][r^a - \mu_o^a(t+1]^T}{\sum_{r^a} p(o \mid r^a, \lambda(t)} \tag{8}$$

After multiple iterations of the EM-like algorithm, we have the final values μ_o^a and $\sum_o^a$ for each object class o and the final probability $p(o|r^a)$ for each object class o and feature vector r^a. Now, given a test image I we can calculate the probability of object class o being in image I given all the region vectors r^a in I:

$$p(o \mid F_I^a) = f\{p(o \mid r^a) \mid r_a \in F_I^a\} \tag{9}$$

where f is an aggregate function that combines the evidence from each of the type-a regions in the image. We use max and mean as aggregate functions in our experiments.

Multiple-Feature Case

Since our abstract regions can come from several different processes, we must specify how the different attributes of the different processes will be combined. For the EM-variant, we have tried two different forms of combination:

1. Treat the different types of regions independently and combine only at the time of classification:

$$p(0 \mid \{F_I^a\} = \prod_a p(0 \mid F_I^a) \tag{10}$$

2. Form intersections of the different types of regions and use them, instead of the original regions, for classification.

In the first case, only the specific attributes of a particular type of region are used for the respective mixture models. If a set of regions came from color segmentation, only their color attributes are used, whereas if they came from texture segmentation, only their texture coefficients are used. In the second case, the intersections are smaller regions with properties from all the different processes. Thus an intersection region would have both color attributes and texture attributes.

RESULTS

We tested the EM-variant approach on color segmentations and texture segmentations. The test database of 860 images was obtained from two image databases: creatas.com and our groundtruth database. The images are described by 18 keywords. The keywords and their appearance counts are listed in Table1. We ran a set of cross-validation experiments in each of which 80% of the images were used as the training set and the other 20% as the test set. In the experiments, the recognition threshold was varied to obtain a set of ROC curves to display the percentage of true positives vs. false positives for each object class. The measure of performance for each class was the area under its ROC curve, which we will henceforth call a ROC score. Figure 1 illustrates the ROC curves for each object, treating color and texture independently. Figure 2 illustrates the results for the same objects, using intersections of color and texture regions. Table 2 lists the ROC scores for the 18 object classes for these two different feature combination methods. In general, the intersection method achieves better results than the independent treatment method, a 6.4% performance increase in terms of ROC scores. This makes sense because, for example, a single region exhibiting grass color and grass texture is more likely to be grass than one region with grass color and another with grass texture. Using intersections, most of the curves show a true positive rate above 80% for false positive rate 30%. The poorest results are on object classes “tree,” “grass,” and “water,” each of which has a high variance, for which a single Gaussian model is not sufficient.

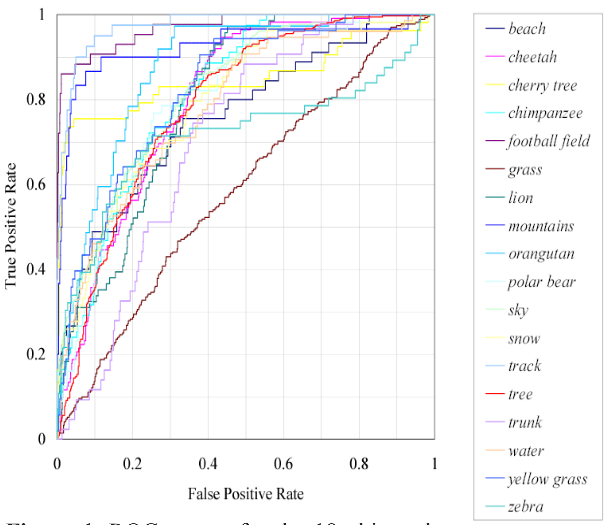

Figure 1. ROC curves for the 18 object classes with independent treatment of color and texture.

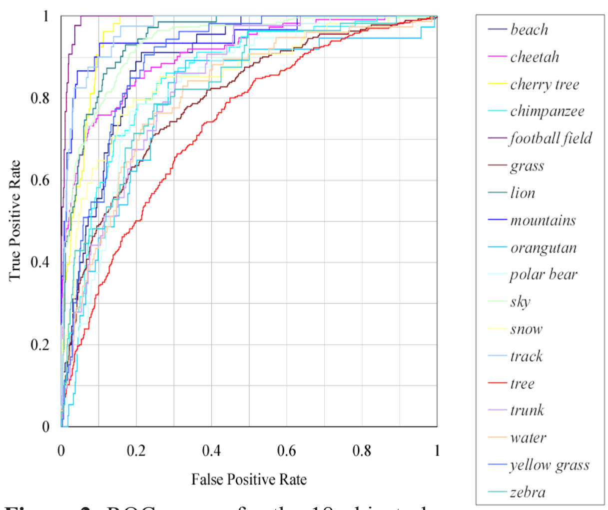

Figure 2. ROC curves for the 18 object classes using intersections of color and texture regions.

Table 1. EM-variant Experiment Data Set Keywords and Their Appearance Counts

Key word	count
Mountains	30
Orangutan	37
Track	40
Tree trunk	43
Football field	43
Beach	45
Prairie grass	53
Cherry tree	53
Snow	54
Zebra	56
Polar bear	56
Lion	71
Water	76
Chimpanzee	79
Cheetah	112
Sky	259
Grass	275
tree	361

Our EM-variant approach, described in Section II, assumes that the feature distribution of each object within a region is a Gaussian. So it has difficulty modeling objects having a high variance or multiple appearances, for which a single Gaussian model is not sufficient. Therefore a justifiable extension of the EM-variant approach is to model the feature distribution of each object as a mixture of Gaussian, instead of a single Gaussian. To compare this extension to the EM-variant approach described in Section II for recognizing objects having multiple appearances, we used the same set of 860 images, but relabeled them with 10 general object classes to replace the 18 more specific classes used in that work. For example, the former classes “tree trunk”, “cherry tree”, and just plain “tree” were merged to form a single “tree” class. The set of 10 classes used were mountains, stadiums, beaches, arctic scenes, water, primates, African animals, sky, grass, and trees. The mapping relationships from the old labels to the new labels are listed in Table 3. We applied both the EM-variant and EM-variant extension to this new labeled image set using color and texture features. The features were combined via region intersections. The EM-variant extension uses a Gaussian mixture to approximate the distribution of each object. While general Gaussian parameters are used for the original EM-variant, aligned Gaussian parameters, in which the covariance matrixes are diagonal matrices,

Table 2. ROC scores for the two different feature combination methods: 1) independent treatment of color and texture and, 2) intersections of color and texture regions

	Independent treatment(%)	Intersection Method(%)
Tree	78.8	73.3
Orangutan	87.4	79.3
Grass	58.5	79.5
Water	78.2	81
Zebra	71.7	82.9
Polar bear	79.9	82.9
Tree trunk	70.6	83.4
Snow	79.6	85.2
Chimpanzee	81.5	85.3
Beach	76.1	89
Prairie grass	82.5	89.4
Cheetah	80.1	90.5
Sky	82	93.3
Lion	79.7	94.4
Mountains	92.6	94.7
Cherry tree	84.8	95.7
Track	97.5	96.7
Football field	97	99.1
Mean	**81**	**87.5**

are adopted for the EM-variant extension. There are two reasons for this decision. The first one is the system efficiency. If there are m objects to learn, the original EM-variant performs the iterations for the convergence of a (m + 1)-component Gaussian mixture in which m Gaussians components are for the objects and one is for the "background". For the EM-variant extension, a region is modeled as a mixture of object models denoted by the outer mixture, which in turn are modeled as Gaussian mixtures denoted by the inner mixtures.

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 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 +d + 1) parameters to learn. Using general Gaussian parameters with the EM-variant extension, there are (m + 1) × [m′ × (d + 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 3. 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 4. 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 3. Mapping of the more specific old labels to the more general new labels. The first column is the new labels and the second column lists their corresponding old labels. The number of images containing each object class is shown in parentheses.

New label	Old label
Mountains(30)	Mountains(30)
Stadium(44)	Track(40, football field(43)
Beach(45)	Beach(45)
Arctic(56)	Snow(54), polar bear(56)
Water(76)	Water(76)
Primate(116)	Orangutan(37), chimpanzee(79)
African animal	ebra(56),lion(71), cheetah(112)
Sky(259)	Sky(259)
Grass(321)	Prairie grass(53), grass(272)
Tree(378)	Tree trunk (43), cherry tree(53), tree(361)

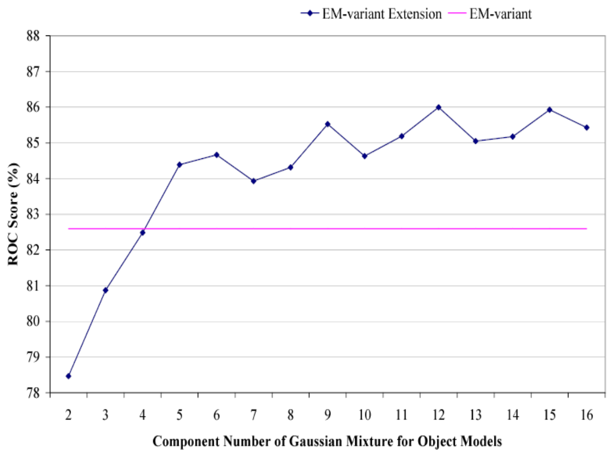

Figure 3. The ROC scores of experiments with different value of the parameter, m′, the component number of Gaussian mixture for each object model

Table 4. 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

CONCLUSION

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. 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.

REFERENCE

[1] H. Murase and S.K. Nayar. Visual learning and recognition of 3-d objects from appearence. International Journal of Computer Vision, 14:5–24, 1995.

[2] Chad Carson, Megan Thomas, Serge Belongie, Joseph M. Hellerstein, and Jitendra Malik. Blobworld: A system for region-based image indexing and retrieval. In Third International Conference on Visual Information Systems. Springer, 1999.

[3] Wei-Ying Ma and B. S. Manjunath. Netra: A toolbox for navigating large image databases. Multimedia Systems, 7(3):184–198, 1999.

[4] Margaret M. Fleck, David A. Forsyth, and Chris Bregler. Finding naked people. In Proceedings of the European Conference on Computer Vision (ECCV) (2), pages 593–602, 1996.

[5] D. Forsyth and M. Fleck. Body plans. In Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, pages 678–683, 1997.

[6] Henry A. Rowley, Shumeet Baluja, and Takeo Kanade. Neural network-based face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(1):23–38, 1998.

[7] A. W. M. 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, 2000.

[8] John R. Smith and Shih-Fu Chang. Visualseek: A fully automated content-based image query system. In ACM Multimedia, pages 87–98, 1996.

[9] John R. Smith and Shih-Fu Chang. Image and video search engine for the world wide web. In Storage and Retrieval for Image and Video Databases (SPIE), pages 84–95, 1997.

[10] James Ze Wang, Jia Li, and Gio Wiederhold. SIMPLIcity: Semantics-sensitive integrated matching for picture LIbraries. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(9):947–963, 2001.

[11] Kobus Barnard and David Forsyth. Learning the semantics of words and pictures. In International Conference on Computer Vision, volume 2, pages 408–415, 2001.
[12] P. Duygulu, K. Barnard, J.F.G. de Freitas, and D.A. Forsyth. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In ECCV02, volume 4, pages 97–112, 2002.
[13] J. Jeon, V. Lavrenko, and R. Manmatha. Automatic image annotation and retrieval using cross-media relevance models. In Proceedings of SIGIR '03 Conference, pages 119–126, 2003.
[14] J. Li and J.Z. Wang. Automatic linguistic indexing of pictures by a statistical modeling approach. PAMI, 25(9):1075–1088, September 2003.
[15] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In CVPR 2003, pages 264–271, 2003.
[16] Markus Weber, Max Welling, and Pietro Perona. Unsupervised learning of models for recognition. In Proceedings of the European Conference on Computer Vision (ECCV)(1), pages 18–32, 2000.
[17] Oded Maron and Aparna Lakshmi Ratan. Multiple-instance learning for natural scene classification. In ICML, 1998.
[18] Thomas G. Dietterich, Richard H. Lathrop, and Toms Lozano-Prez. Solving the multiple instance problem with axis-parallel rectangles. Artificial Intelligence, 89:31 – 71, 1997.
[19] Chad Carson, Serge Belongie, Hayit Greenspan, and Jitendra Malik. Blobworld: Image segmentation using expectation-maximization and its application to image querying. PAMI, 24(8):1026–1038, 2002.
[20] A. P. Berman and L. G. Shapiro. A flexible image database system for content-based retrieval. Computer Vision and Image Understanding, 75(1-2):175–195, 1999.
[21] Amirani, M., Sadeghi Gol, Z., Beheshti, A., "Efficient Feature Extraction for Shape-based Image Retrieval," Journal of Applied Science, pp.2378-2386, 2008.