Probabilistic shape-based image indexing and retrieval

Konstantinos Valasoulis and Aristidis Likas
University of Ioannina
Department of Computer Science
GR 45110, Ioannina, Greece
{merit,arly}@cs.uoi.gr

Abstract

In this paper we present a probabilistic framework for shape-based indexing and retrieval of images. In our framework shape-based features are extracted from each image and then a statistical model of the image is constructed using an effective determininstic method for Gaussian mixture modeling. In this way, each image is finally represented as a mixture of Gaussians and shape-based similarity between images is computed by measuring the distance between the corresponding mixture distributions. Several distance measures are presented and experimentally compared. Experimental results on the retrieval of logo images indicate that the method is very effective and exhibits robustness to the presence of various types of edge-related noise in the query image.

1. Introduction

In content-based image search, the goal is to retrieve the "most similar" images to a query image introduced to the system. The most commonly used properties of images for content-based retrieval are color, texture and shape. In this work we focus on shape-based similarity although the approach can be easily adapted to the case of other image characteristics. Several methods have been proposed for indexing using shape-based information: for example, Huet & Hancock [2] use shape-based attributes to perform object recognition with the use of robust error kernels and Suganthan [3] extracts shape-based attributes from images and builds the shape index using a combination of two Self-Organizing Maps (SOMs). In this work we elaborate on the problem of shape-based image indexing and retrieval using a probabilistic modeling approach based on Gaussian Mixture Models (GMMs).

Acknowledgement: This research was partially funded by the European Commission and the Hellenic Ministry of Education through EPEAEK II.

To apply our method, first image segmentation and linking is applied to an image and then a dataset of shape-related features is computed for each image. The statistics of this dataset are subsequently modelled by constructing a Gaussian Mixture Model (GMM) using an efficient deterministic incremental training method [4]. The GMM constructed for each image constitutes the corresponding index and threfore, the task of shape-based image retrieval can be performed with the use of appropriately defined similarity measures (eg. Kullback-Liebler distance) between GMMs. We have studied the performance of three probabilistic similarity measures under three types of edge corruption error.

2. Probabilistic Shape-Based Image Modeling

We have chosen to represent the shapes in images by line segments. Therefore, via a preprocessing step we map each image in the library to a line segment list representing its shape. Each line segment list is then used to create a set of five dimensional attribute vectors for each image. In the image representation phase, we associate the set of relational attribute vectors of each image to a GMM. The mixture models are used instead of the traditional histogram-based approach for three reasons. First, they can effectively model high dimensional data as happens in our case where each image is represented as a set of five dimesnional vectors. Second they provide a compact representation of each image that uses a small number of model parameters and third they lead to the definition of image similarity measures based on the distance between probability distributions.

Once every image in the library has been modeled by a GMM the indexing task is completed. In order to be able to perform a shape-based search in the image library (given a query image) first a similarity measure between GMMs is defined and then the following procedure is applied. We preprocess the query image and obtain its line segment list.

0-7695-2128-2/04 $20.00 (C) 2004 IEEE

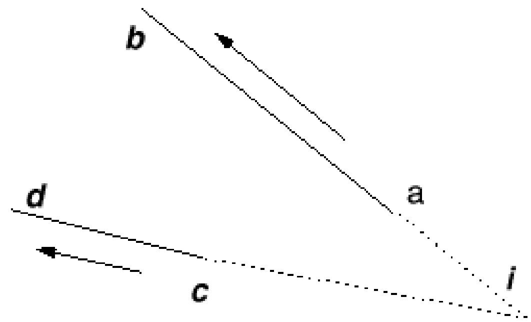

Figure 1. A pair of line segments and the labels of the end points.

Then, we extract the set of five dimensional attribute vectors that coresponds to the line list and use this data set to associate a new GMM to the query image. Finally, by computing the values of the distance between the query GMM and all the GMMs in the library we can can identify the images that are most similar to the query based on shape characteristics.

2.1. Image Preprocessing

In order to obtain the line segments for each image in the library, first an edge detection routine is applied to each image producing a corresponding binary image. Then an edge linking algorithm is applied to the binary image that also elliminates all short edges. With the use of a simple line extraction algorithm a list of all the line segments that exist in the binary image is obtained. At the end of the preprocessing phase the shape of each image in the library is represented by a list of line segments.

2.2. Relational Attribute Vectors

In this phase the line segments extracted for each image are used to produce a set of relational attribute vectors [3]. For an arbitrary pair of line segments in the list (Fig. 1) we first compute the intersection point "i" between the two line segments. Then the end point of the first line segment (which is the line segment that is closer to point "i") which is closer to the intersection point is labeled as "a", and the corresponding end point of the other line segment as "b". The other end points of the two line segments are labeled as "c" and "d" respectively. After labeling the four end points the following five translation, rotation and scale invariant relational attributes can be produced: (a) $\theta_{ab,cd} = arccos\{\vec{\mathbf{ab}} \cdot \vec{\mathbf{cd}}/|\vec{\mathbf{ab}}| \cdot |\vec{\mathbf{cd}}|\}$, (b) Relative position ratio: $1/((1/2) + (l_{ib}/l_{ab}))$, (c) Line length ratio: $\min\{l_{ab}, l_{cd}\}/\max\{l_{ab}, l_{cd}\}$, (d) End point ratio: $\min\{l_{ac}, l_{bd}\}/\max\{l_{ac}, l_{bd}\}$, (e) Cross end point ratio: $\min\{l_{ad}, l_{bc}\}/\max\{l_{ad}, l_{bc}\}$. The angle $\theta_{ab,cd}$ is between zero and π. However, if we identify the rotation from $\vec{ab}$ to $\vec{cd}$ as clockwise or counter-clockwise by evaluating the vector product between $\vec{ab}$ and $\vec{cd}$, then we can compute the angle between $-\pi$ to π in order to improve the discrimination quality of this attribute.

Using the above procedure for every pair of line segments existing in the list, we may produce a number of relational attribute vectors. In this way, the shape of each image in the library can be represented by a set of relational attribute vectors. However, the shape on some images may be considerably complex, which means that after the preprocessing phase the produced list of line segments of an image may contain thousands of entities. In this case the number of line pairs will be enormous leading to the construction of huge datasets for each image and making the subsequent statistical modelling of each dataset a time consuming task. To overcome this diificulty, we conisder only the relational attribute vectors corresponding to the line pairs that arise from the six nearest line segments of each line in the list as suggested in [1]. The procedure mentioned above is applied to every line segment list L_i produced during the preprocessing phase of each image I_i in the library. Thus, each image is represented by a set S_i of five dimensional relational attribute vectors with $|S_i| = 6 \cdot |L_i|$.

2.3. Image Modelling

In the image modelling phase, we model every set S_i corresponding to an image I_i using a Gaussian Mixture Model. It is well known that if the distribution of a random variable $x \in R^d$ is a mixture of k Gaussians then its density function is:

$$f(x|\theta) = \sum_{j=1}^{k} \alpha_j \frac{1}{\sqrt{(2\pi)^d |\Sigma_j|}} exp\{-0.5(x-\mu_j)^T \Sigma_j^{-1} (x-\mu_j)\}$$

where that the parameter set $\theta = \{\alpha_j, \mu_j, \Sigma_j\}_{j=1}^{k}$ consists of: $\alpha_j > 0$, $\sum_{j=1}^{k} \alpha_j = 1$, $\mu_j \in R^d$ and Σ_j is a $d \times d$ positive definite covariance matrix. In our case $d = 5$ and we fixed $k = 6$.

Given a set of relational attribute vectors S_i, the maximum likelihood estimation of θ is:

$$\theta^i_{ML} = \arg\max_{\theta} L(S_i|\theta)$$

where $L(S_i|\theta) = \sum_{x_j \in S_i} logf(x_j|\theta)$. A first approach would be to use the well-known Expectation-Maximization (EM) algorithm as an iterative method to obtain θ^i_{ML} for each image I_i. However, a serious problem of EM is its dependence on the initial values of parameters θ. In our system we require a training method that always provides the same GMM solution for a specific image. Consequently, the

EM should always start from the same initial parameter values for all images and in general such initial values might lead to inferior training results for many of them. This serious problem also exists in other apporaches that train parametric models for shape-based image retrieval [3].

In our system, the mixture models are built using the *Greedy EM* algorithm for Gaussian Mixture Learning [4]. This incremental method starts with one Gaussian component and sequentially adds a new component unitl a maximum number of components has been added. It does not depend on initial parameter values and provides significantly better solutions than the typical EM algorithm.

Using the Greedy EM algorithm, we associate a GMM with six components to every set S_i and therefore to every image I_i. In other words, we represent each image I_i that exists in the library by GMM_i.

2.4. Probabilistic Similarity Measures

Once a GMM has been associated to every image, measuring the distance between images can be viewed as measuring the distance between GMMs. In this work we studied the performance of three different probabilistic similarity measures presented below. Let I_1 and I_2 denote two arbitrary images with sets of relational attribute vectors $S_1 = \{x_{11}, \ldots, x_{1n_1}\}$ and $S_2 = \{x_{21}, \ldots, x_{2n_2}\}$ respectively. Let also f_1 and f_2 two probability density functions corresponding to S_1 and S_2. We can define:

(a) The non-symmetric average log-likelihood measure ($\mathbf{d_{ns}}$):

$$d_{ns}(S_1||f_2) = \frac{1}{n_1}\sum_{t=1}^{n_1}\log f_2(x_{1t})$$

(b) The symmetric average log-likelihood measure ($\mathbf{D_S}$):

$$D_S(f_1||f_2) = \frac{1}{n_1}\sum_{t=1}^{n_1}\log f_2(x_{1t}) + \frac{1}{n_2}\sum_{t=1}^{n_2}\log f_1(x_{2t})$$

(c) The symmetric version of the Kullback-Liebler (KL) distance ($\mathbf{D_{KL}}$):

$$D_{KL}(f_1||f_2) \cong \frac{1}{n_1}\sum_{t=1}^{n_1}\log\frac{f_1(x_{1t})}{f_2(x_{1t})} + \frac{1}{n_2}\sum_{t=1}^{n_2}\log\frac{f_2(x_{2t})}{f_1(x_{2t}}$$

Using any of the above three measures d_{ns}, D_S and D_{KL} we can easily perform the image retrieval task as follows: Given a query image I_q, we compute the corresponding L_q and S_q. In case the d_{ns} measure is used, we simply compute the values $d_{ns}(S_q||f_i)$ for $i = 1, \ldots, M$, where M is the number of images in the library. The most similar image to the query is

$$i^\star_{ns} = \arg\max_i d_{ns}(S_q||f_i)$$

In case we use D_S or D_{KL} as a similarity measure, we associate a GMM ($f_q = GMM_q$) to the query image I_q and we compute the values $D_S(f_q||f_i)$ or $D_{KL}(f_q||f_i)$. The most similar image in the library to the query image is:

$$i^\star_S = \arg\max_i D_S(f_q||f_i)$$

or

$$i^\star_{KL} = \arg\min_i D_{KL}(f_q||f_i)$$

respectively.

3. Experimental Results

We have experimented with an image dataset of 182 *logos and trademarks* found on the Web. Note that some images in the database are very similar on the front of shape. We modelled each image in the dataset using a mixture of six Gaussians with diagonal covariance matrices. Then we performed shape-based retrieval experiments using each time as a query one of the images in the dataset and tested our system with the three similarity measures defined in the previous section. We have found that regardless of the similarity measure that was used, *the system always provided the correct image*.

Figure 2. Query Image with segment error.

We also studied the sensitivity of the proposed approach under various types of pattern corruption error on the query. We have simulated the segmentation errors that can occur when line-segments are extracted from realistic query image data, investigating the four different cases listed below [1] :

- Extra lines: Additional line segments are placed at random image locations. The lengths and angles of the added line segments heve been generated randomly.
- Missing lines: Existing line segments are randomly selected and removed from the image.
- Segment errors: For a predefined fraction of randomly selected existing line segments, we have introduced random displacements in the end points positions. The end point errors are random.

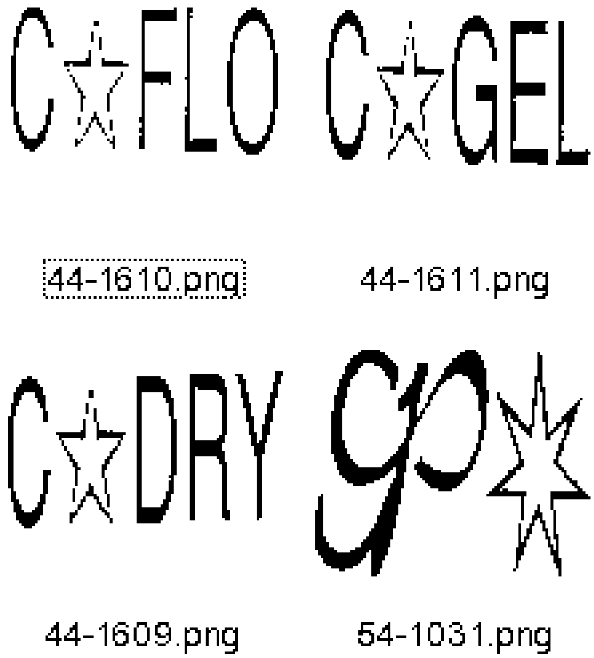

Figure 3. The four most similar images to the query of Fig. 2 that were retrieved by the system.

- Mixed errors: We have introduced the above three types of errors to the query image in equal proportion.

We applied the above types of pattern corruption error, on each image in the dataset. A typical query example is presented in Fig. 2 and the retrieved images for this query are presented in Fig. 1. Table 1 shows the recognition errors that occured in 182 different query searches for 5%-20% segmentation error. Columns labeled as A show the percentage of experiments where the target image was not the *most probable* returned by the system. Columns labeled as B show the percentage of experiments where the target image was not present in the *eight (ie. 5%) most probable* images returned by the system. It is clear that the performance of the system is very good especially for error levels until 5%. It is also quite clear that the system is more sensitive to *missing lines*. Moreover, the use of distance D_S leads to better performance than D_{KL} and d_{NS}, for all types of segmentation error.

4. Conclusions

In this work we have presented a framework for shape-based indexing and retrieval on datasets of images. The proposed approach is based on the statistical modeling of the shape characteristics of each image using gaussian mixture models. For an image database with 182 logo images we have shown that the retrieval task is accurate when no segmentation error is present in the query image. Moreover, the experimental results also indicate that the method exhibits robust retrieval performance in the presence of several types of noise in the query image. In what concerns our future work, first, we aim to use the above framework to perform shape-based image clustering and classification. Another aim is to expand the approach in order to construct the statistical image models using not only shape-based features but also a combination of additional features such as color, texture or wavelet-based features. Finally, we also plan to enchance the functionality our shape-based retrieval system in order to take into account relevance feedback provided by the user.

		$\mathbf{d_{ns}}$		$\mathbf{D_{KL}}$		$\mathbf{D_S}$	
Type		A	B	A	B	A	B
Missing lines	5%	6	**1**	4	**1**	4	**0**
	10%	50	**20**	34	**15**	27	**5**
	15%	81	**48**	71	**42**	58	**24**
	20%	113	**74**	97	**63**	83	**43**
Segment errors	5%	1	**1**	1	**0**	0	**0**
	10%	17	**3**	7	**1**	4	**0**
	15%	40	**12**	24	**5**	12	**2**
	20%	81	**30**	52	**16**	29	**7**
Extra lines	5%	6	**1**	4	**1**	2	**1**
	10%	52	**20**	38	**12**	24	**2**
	15%	111	**50**	97	**32**	62	**6**
	20%	150	**73**	139	**68**	92	**23**
Mixed	15%	64	**30**	43	**21**	28	**8**

Table 1. Retrieval errors for various types of segmentation noise.

References

[1] B. Huet and E. Hancock. Line pattern retrieval using relational histograms. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21:1363–1370, 1999.

[2] B. Huet and E. Hancock. Relational object recognition from large structural libraries. *Pattern Recognition*, 35:1895–1915, 2002.

[3] P. N. Suganthan. Shape indexing using self-organizing maps. *IEEE Trans. on Neural Networks*, 13:835–840, July 2002.

[4] N. Vlassis and A. Likas. A greedy em algorithm for gaussian mixture learning. *Neural Processing Letters*, 15:77–87, 2002.