

Probabilistic models for generating, modelling and matching image categories

Hayit Greenspan
Faculty of Engineering,
Tel-Aviv University, Israel
hayit@eng.tau.ac.il

Shiri Gordon
Faculty of Engineering,
Tel-Aviv University, Israel
gordonha@post.tau.ac.il

Jacob Golberger
CUTE Systems Ltd.,
Tel-Aviv, Israel
jacob@cute.co.il

Abstract

In this paper we present a probabilistic and continuous framework for supervised image category modelling and matching as well as unsupervised clustering of image space into image categories. A generalized GMM-KL framework is described in which each image or image-set (category) is represented as a Gaussian mixture distribution and images (categories) are compared and matched via a probabilistic measure of similarity between distributions. Image-to-category matching is investigated and unsupervised clustering of a random image set into visually coherent image categories is demonstrated.

Keywords: Image Categories; Image and Category Matching; Unsupervised Clustering; Image Grouping; Gaussian mixture modelling; Kullback-Leibler distance.

1 Introduction

Image categorization is a means for high-level description of image content. In classification algorithms, the goal is to affiliate an input image to one of a predefined set of categories. In content-based search, the goal is to retrieve the “most-similar” images to a query image introduced to the system. The images belonging to the query-image “category” are the images we wish to retrieve first. Often, categories are defined a-priori by the user (supervised) and rules are learned to characterize each category’s unifying characteristics. A second approach to categorization of image information is to view it as a clustering task (unsupervised or supervised) in which the goal is to find relationships among the images and the best way to characterize the content within a given image archive.

In this work we approach the categorization task on both fronts. We start by learning models for categories (supervised) and show image-category matching along with category-category matching. In the second part of the paper we treat the categorization as an unsupervised clustering, or grouping task. We group image information bottom-up, to generate clusters, or “image sets”. The generated image

sets provide a concise summarization and visualization of the image content.

2 The GMM-KL framework

In recent work [2] we presented the GMM-KL framework, as a continuous and probabilistic framework for image representation and matching. Each image is first represented as a Gaussian mixture distribution (GMM) and images are compared and matched via a probabilistic measure of similarity between distributions. In the representation phase, each homogeneous region in the image plane is represented by a Gaussian distribution, and the set of regions in the image is represented by a Gaussian mixture model (GMM). Pixels are grouped into homogeneous regions in the image plane, by grouping feature vectors in the five-dimensional feature space of color and space (L, a, b, x, y) . The underlying assumption is that the image colors and their spatial distribution in the image plane are generated by a mixture of Gaussians.

The distribution of a random variable $X \in R^d$ is a mixture of k Gaussians if its density function is:

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

such 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 matrix.

Given a set of feature vectors x_1, \dots, x_n , the maximum likelihood estimation of θ is :

$$\theta_{ML} = \arg \max_{\theta} f(x_1, \dots, x_n|\theta). \quad (2)$$

Since a closed form solution for this maximization problem is not possible, we utilize the Expectation-Maximization (EM) algorithm as an iterative method to obtain θ_{ML} [2].

The iterative EM algorithm is initialized via the K-means algorithm [1], and is repeated until the log-likelihood measure is increased by less than a predefined threshold (1%)

from one iteration to the next. Once we associate a Gaussian mixture model to an image, the image can be viewed as a set of independently identically distributed (IID) samples from the Gaussian mixture distribution.

In the GMM-KL framework, the distance measure between two images is taken as a distance measure between the two Gaussian mixture distributions obtained from the images. The matching between images is treated as a distribution matching task, using the information-theoretic motivated Kullback-Leibler (KL) distance [3]. Denote the Gaussian mixture models computed from the two images by f_1 and f_2 . Given two distributions f_1 and f_2 the (non-symmetric version) of the KL distance is :

$$D(f_1||f_2) = E_{f_1} \log \frac{f_1(x)}{f_2(x)} \quad (3)$$

where E is the expected value function. The KL distance between two Gaussian mixture distributions can not be analytically computed. An approximation is possible through Monte-Carlo procedures.

3 Modelling image categories

In this paper we present a generalization of the GMM-KL framework to image-sets, or categories. In this section we focus on category modelling and matching. We assume that the image sets have been selected a-priori (supervised). Let I_1, \dots, I_N denote an image set for class C . Alternatively, the image set may be represented by a set of the corresponding GMMs of each input image: GMM_1, \dots, GMM_N .

Given the set of image GMMs we define next the underlying model for all the images in the set. The image-set model can be learned as follows: Extract random samples from each image of the image set. Samples can be extracted from the pixel space or from the GMM model representing the image. The extracted feature set can be viewed as a set of independently identically distributed (IID) samples from the Gaussian mixture distribution representing the class. A GMM model is next generated for the combined collected sample set, using the iterative EM algorithm as in the single image case.

Figure 2 shows examples of image sets (left) and the corresponding image category GMMs (right). Each Gaussian in the model is displayed as a localized colored ellipsoid. Some of the Gaussians overlap spatially and thus are not explicitly shown in the image. The class model allows for a certain amount of variability in the colors per spatial location, as well as a certain amount of variability in the spatial location of the colored blobs.

We next define distance measures between an image and a class, as well as between two separate image classes (image categories). Denote the Gaussian mixture model com-

puted for the *image class*, C , by f_C . Given an input image distributions f_I and the class distribution f_C , the (non-symmetric version) of the KL distance is :

$$D(f_I||f_C) = E_{f_I} \log \frac{f_I(x)}{f_C(x)} \quad (4)$$

In the image-to-class case the distances may be directional, from the image to the class. A symmetric version may be the more appropriate measure for between category distances. For example, if categories C_1 and C_2 are to be compared:

$$\begin{aligned} d(C_1, C_2) &= \frac{1}{2} (D(f_{C_1}||f_{C_2}) + D(f_{C_2}||f_{C_1})) \quad (5) \\ &\cong \frac{1}{n_1} \sum_{t=1}^{n_1} \log \frac{f_{C_1}(x_{1t})}{f_{C_2}(x_{1t})} + \frac{1}{n_2} \sum_{t=1}^{n_2} \log \frac{f_{C_2}(x_{2t})}{f_{C_1}(x_{2t})} \end{aligned}$$

where f_{C_1} and f_{C_2} are the centroids of categories C_1 and C_2 , $x_{i1} \dots x_{in_i}$ is the feature set extracted from category C_i , ($i = 1, 2$), and n_i is the size of this set.

3.1 Experiments in category modelling and matching

It is of interest to investigate the following questions: are the category models representative of the underlying image set? Can image-to-category matching enable image classification? An initial investigation was conducted on a pre-selected set of 6 categories: monkey, snow, sunset, flowers, pyramid, and coral, taken from the COREL database (10 images per class). The database used can be seen in Figures 1 and 2. Category-to-category distances are computed following equation (5) and are listed in Table 1. Similarity between categories may be learned automatically using the category model distances. For example, the distances between the three categories of sunsets, flowers and coral, are quite small, corresponding to the visual similarity of the image sets. The category-to-category distances can be used for classification error prediction. Table 2 lists distances between *image* models (rows) and *category* models (columns). Each table entry is calculated as an average of 10 ‘‘leave-one-out’’ experiments, in each such experiment a single image is used as query, the other 9 are used to learn a GMM. Image-to-category distance is computed as in equation (4). Note that Table 1 is symmetric while Table 2 is not.

A tight correlation exists between the category-to-category models and the image-to-category distances, suggesting that image models are very similar in behavior to category models. Category models do in fact represent their image building blocks. Extracting information from pre-learned category models may provide a reasonable prediction of image classification performance.

Class	(1)	(2)	(3)	(4)	(5)	(6)
(1)monkey	0	20.0	21.0	8.5	16.1	11.9
(2)snow	20.2	0	21.6	16.8	21.5	13.3
(3)sunset	21.4	20.9	0	13.4	18.3	6.3
(4)flowers	8.7	16.7	12.9	0	18.8	5.3
(5)pyramid	16.3	20.5	18.7	17.8	0	18.5
(6)coral	13.2	12.8	6.5	5.1	17.3	0

Table 1. Category Model to category Model analysis

Image	(1)	(2)	(3)	(4)	(5)	(6)
(1)monkey	6.5	32.5	34.8	16.4	26.2	20.7
(2)snow	29.6	10.4	42.1	30.4	29.9	21.2
(3)sunset	30.2	36.3	14.2	27.7	28.6	14.6
(4)flowers	14.4	28.7	29.1	8.5	27.6	12.1
(5)pyramid	23.1	29.8	34.8	28.0	11.7	23.8
(6)coral	20.8	24.7	20.1	15.1	26.3	7.7

Table 2. Image Model to category Model analysis

4 Unsupervised Clustering of Image GMMs into Image Categories

In this section we focus on *grouping* image content in an automatic *unsupervised* scheme. The grouping framework is hierarchical: The first phase of the grouping process is the basic image modelling phase, as discussed in section 2. Image pixels are grouped into coherent regions (“blobs”) in feature space; these are modelled via Gaussian mixture models (GMMs). Here we focus on the higher-level grouping process, from image GMMs to *image categories*. Image matching, using the KL distance, is a necessary component in each stage of the grouping process.

We use an iterative clustering algorithm, following the well-known K-means algorithm [1]. Each iteration includes two major steps: (1) estimating cluster centroids, and (2) assigning each sample to the closest centroid. In the classical K-means algorithm, a set of real-valued vectors is given and the Euclidean distance is used. Centroids are computed as the affiliated sample mean. The Euclidean distance (or variations thereof) is used to measure distances between the samples and the cluster centers. Here, we shift from a set of vectors to a set of images; we thus need to define the *centroid* of an *image set* and the distances between images and between an image and the centroid.

We use the GMM-KL framework: Given the set of image GMMs, the cluster C centroid is defined as the underlying model, GMM_C , for all the images in the set. This is similar to learning a model representative of the class (see section

3). Correspondingly, the distance measures between an image sample and the centroid is equivalent to the distance between an image model and a class model. Such distance measures have been defined within the GMM-KL framework, as in equation 4.

The iterative clustering algorithm we propose combines the two main steps discussed above: *Initialization*: Start with $K=1$ (single cluster case); compute centroid of all images in the image archive. Perturb centroid to generate two cluster means ($K=2$). The perturbation of the cluster model involves shifting the mean of each blob of the centroid-GMM by a small quantity in two opposite directions.

For $K = 2, \dots$ do :

- Iterate until there is no change in image-cluster affiliations:

- Given cluster centers, $f_{C_i}, i = 1..K$, assign each image, indexed by I , to the cluster j , for which image to center distance is the smallest using KL-dist:

$$j = \arg \min_i D(f_I || f_{C_i}) \quad (6)$$

- Update cluster centroid using the image set that is associated with each cluster. Cluster centroid is the GMM model for all images in the set, GMM_C .

- Split one of the clusters into two. The splitting criterion involves selecting the cluster C_L in which the intra-class distances are of the largest spread. Varying other criteria may be used, such as the size of the clusters and more. The centroid of the selected cluster C_L is perturbed into two cluster means (see initialization step).

The number of clusters K is increased until a stopping criterion is met. In the current implementation K is increased until there is no significant change in intra-class distances.

4.1 Experiments in Image Clustering

We have experimented with several natural scenery image databases from the COREL database. In the following we show experimental results of unsupervised clustering a given image set. The database used consists of 10 images from 6 of the COREL database categories (same database used in section 3.1). We start the unsupervised clustering process with the entire image set ($K = 1$). The K-means algorithm is run from $K = 2$, increasing K until a stopping criterion is met. Each iteration of the algorithm (differing K) can be viewed as a level of a descriptive hierarchy of the image content within the archive.

Figure 1 and Figure 2 show the clusters that are extracted unsupervised, for $K = 2$ and $K = 6$, respectively

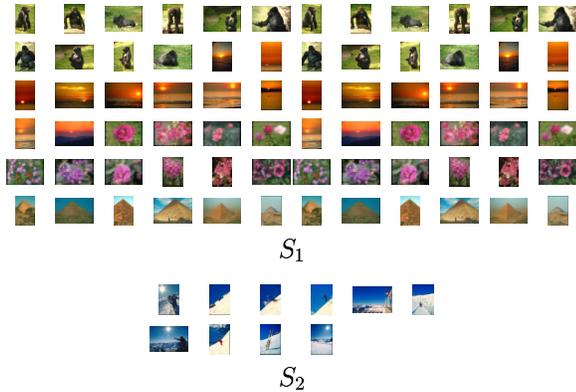


Figure 1. Results: Two cluster case (K=2).

(clusters can be viewed for any K). In Figure 1 the input set is grouped, unsupervised, into two image sets. The more bluish colors seem to be separated from the rest of the images. After several additional iterations, the algorithm reaches the six cluster case ($K = 6$), the results of which are shown in Figure 2. According to the stopping criterion chosen, this is an optimal grouping. Note that the number of categories (6) is found unsupervised. This number matches with the number of COREL categories included in the original database. The clustering of the images to the different image sets is visually pleasing and also matches the ground-truth of the pre-labeled set (except for the flower image that is affiliated with the “monkey” image set). It is important to remember that the clustering is based on the limited representation of colors and their spatial layouts (no higher-level information is present). The GMM cluster representation is shown on the right of each image set. A clear distinction between the groups is evident in the Gaussian mixture characteristics, in blob color features and their spatial layouts.

5 Discussion

In this work we present results of using the proposed GMM-KL framework in category modelling, generation (unsupervised) and matching. The GMM-KL framework provides a continuous and probabilistic framework for image and category representation and matching. Matching between categories and between images and categories is defined. A high degree of correlation between the two scenarios suggests that image models are very similar in behavior to category models. It may be concluded therefore, that learning category models and matching between the models can provide a prediction of the performance of any future image classification task within the given image archive. Category generation is presented as an hierarchical grouping process in which we use unsupervised clustering of image content to shift from image GMMs to image categories.

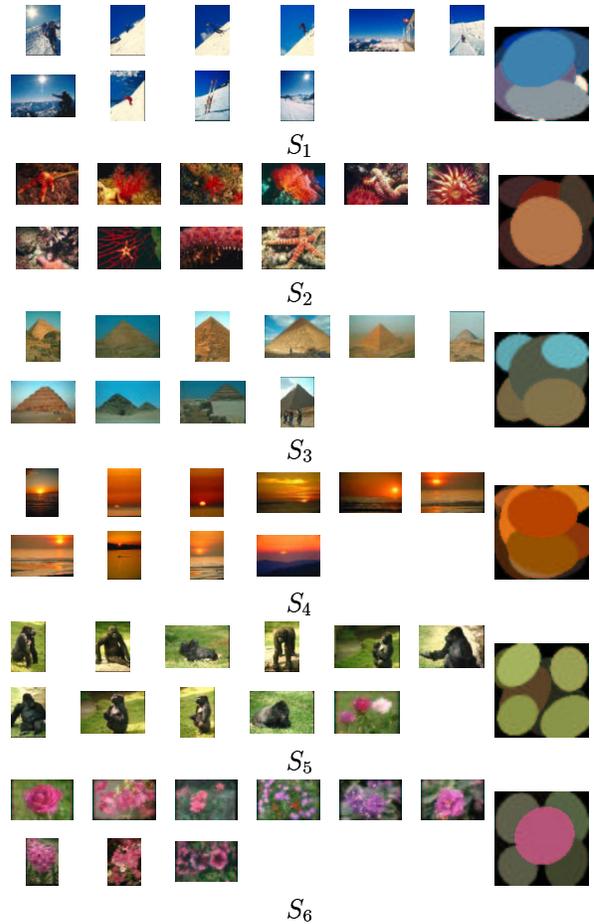


Figure 2. Results: Six cluster case (K=6)

The generated image sets provide a concise summarization and visualization of the image content within a given image archive.

References

- [1] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis*. John Wiley and Sons Inc., 1973.
- [2] H. Greenspan, J. Goldberger, and L. Ridel. A continuous probabilistic framework for image representation and matching. *Journal of Computer Vision and Image Understanding*. In press.
- [3] S. Kullback. *Information theory and Statistics*. Dover, 1968.