A continuous probabilistic framework for image matching

Hayit Greenspan
Tel-Aviv University,
Tel-Aviv 69978, Israel

Jacob Goldberger *
The Weizmann institute of Science,
Rehovot, 76100, Israel

Lenny Ridel
Tel-Aviv University,
Tel-Aviv 69978, Israel

September 9, 2001

Corresponding author:
Dr. Hayit Greenspan
Department of Biomedical Engineering
Faculty of Engineering
Tel Aviv University
Tel Aviv 69978, Israel
Phone: +972-3-6407398
Fax: +972-3-6407939
email : hayit@eng.tau.ac.il

*Currently with CUTe Ltd. Tel-Aviv, Israel

Abstract

In this paper we describe a probabilistic image matching scheme in which the image representation is continuous, and the similarity measure and distance computation are also defined in the continuous domain. Each image is first represented as a Gaussian mixture distribution and images are compared and matched via a probabilistic measure of similarity between distributions. A common probabilistic and continuous framework is applied to the representation as well as the matching process, ensuring an overall system that is theoretically appealing. Matching results are investigated and the application to an image retrieval system is demonstrated.

Keywords: Image matching; Image representation; Gaussian mixture modeling; Kullback-Leibler distance; probabilistic matching.

1 Introduction

Image matching is an important component in many applications that require comparing images based on their content. Examples include image database retrieval systems [5, 1, 14, 17, 24, 10] and a variety of video related applications, such as scene break detection and video parsing (e.g. [30, 29]). There are two main phases in image matching. The first phase involves choosing an image representation space, and the second phase is the definition of an appropriate metric to compare between images in the selected representation space.

Varying resolutions of image representation may be used in the image matching task. One may use the very low-level, pixel representation. In this case, matching between images is based on a distance measure between corresponding pixels (e.g. template matching using the Euclidean distance). The computational effort is minimal in the representation stage, with substantial effort (computational cost) in the matching process. A second option is to shift to a very high-level image representation, in which each image is labeled as belonging to a category (categories such as "sunset", "animals", "indoors" vs "outdoors"). A substantial computational effort is needed in the representation stage, such as using supervised learning techniques to classify the images, and this enables a very simplistic matching stage

of grouping similar content images by the category labels. A mid-level representation exists, that balances the above two, in which a transition is made from pixels to features. Feature vectors are used to compactly represent image content and the image matching phase translates to matching of features. Similarity measures or distance metrics are required to match images in the feature spaces chosen, and across feature spaces.

Most of the works in image retrieval applications belong to the mid-level representation, including the most frequently used histogram methods (see section 2). The framework we present herein also uses the mid-level representation scheme. A shift is made from pixels to a selected feature space. Moreover, pixels are grouped into homogeneous regions in the feature space. The extracted representation is a localized representation both in image as well as feature space.

Additional characteristics of the proposed framework include:

- The image representation is continuous, and the similarity measure and distance computations are also defined in the continuous domain. This provides for a novel continuous and probabilistic framework for image matching.

- A given image is viewed as a particular instantiation of a distribution model representing the image class, and the image matching problem is treated as a distribution matching problem.

- A direct correspondence is shown between the image representation space and the image plane, enabling probabilistic image segmentation.

- A novel statistical evaluation methodology is introduced to enable benchmarking in robustness experiments.

- The image matching framework is extended to include image category modeling and image-to-category matching.

The paper is organized as follows. In section 2 we describe some of the related work in the literature. In section 3 we focus on the representation phase of the proposed framework in which we transition from pixels to coherent regions in feature space, via Gaussian mixture

modeling. In section 4 we present the probabilistic image similarity measure, the Kullback-Leibler (KL) measure, as applied in the extracted image representation space. Analysis of the combined Gaussian mixture and Kullback-Leibler distance framework (GMM-KL) is presented in section 5. A discussion concludes the paper in section 6.

2 Related Work

Histograms are the classical means of representing image content, widely used as the chosen image representation in systems such as IBM's QBIC [5] and Virage's VIR Engine [1]. A histogram is a discrete representation of the continuous feature space. It is generated by a fixed partitioning of the feature space. The partitioning of the feature space is determined by the feature space chosen (e.g. the color space representation), the quantization scheme chosen (such as uniform or vector quantization), as well as computational and storage considerations. Color histograms advantages and disadvantages are well studied [27] and many variations exist [16, 26, 9].

Several measures have been proposed for the dissimilarity between two histograms. In general, they may be divided into two categories [23, 18]: "bin-by-bin" measures, that compare contents of corresponding histogram bins, and "cross-bin" measures that enable comparisons across non-corresponding bins as well. In the first category are included the Minkowski-form distance, as well as the histogram intersection (H.I.) measure [27, 23]. Additional "bin-by-bin" measures include the χ^2 statistics, as well as the Kullback-Leibler (KL) divergence and Jeffrey divergence [11, 4, 19].

"Cross-bin" measures include additional information about the distance between individual features (e.g. between colors represented by the histogram bins). Such measures include the Quadratic-form distance [7], in which a similarity matrix is included to represent similarity between bins. The Earth mover's distance measure [20] extracts dominant modes from histogram as a signature, and defines a measure of similarity between signatures. Additional dissimilarity measures for image retrieval are evaluated and compared in [20, 23, 18].

A common characteristic of the approaches discussed above is the discretization of the feature space with the histogram representation. The binning of the space involves a loss of information. A binning that is too coarse will not have sufficient discriminative power,

while a binning that is too fine will place similar features in different bins which will never be matched. A second major characteristic of the approaches above is that histograms capture only global color distributions of the images, and lack information about spatial relationships of the image colors. A shift to a more localized representation, which reflects spatial information from the image plane, may be desired.

The histogram representation has been extended recently to include additional features as well as spatial information. In [16] each entry of a "joint" histogram contains the number of pixels in the image that are described by a particular combination of feature values. In [26] local information is included by dividing an image into five fixed overlapping blocks and extracting the first three color moments of each block to form a feature vector for the image. In [9] correlograms are proposed to take into account the local color spatial correlation as well as the global distribution of the spatial correlation.

A separate set of works in image representation include "region-based" approaches. Image regions are the basic building blocks in forming the visual content of an image, and thus have great potential in representing the image content and enabling image matching. In [25] Smith and Chang store the location of each color that is present in a sufficient amount in regions computed using histogram backprojection. Ma and Manjunath [13] perform retrieval based on segmented image regions. The segmentation is not fully automatic, as it requires some parametric tuning and hand pruning of regions. Unsupervised segmentation of an image into homogeneous regions in feature space, such as the color and texture space, can be found in the "blobworld" image representation [3, 2]. In [3] a naive Bayes algorithm is used to learn image categories from the blob representation in a supervised learning scheme. The framework suggested entails learning blob-rules per category. Thus, one may argue that there is a shift to a high-level image description (image labeling). Each query image is next compared with the extracted category models, and associated with the closest matching category. The comparison with global color histograms is non-conclusive. In [2] the user composes a query by viewing the blobworld representation, selecting the blobs to match along with possible weighting of the blob features. A query may include a combination (conjunction) of two blobs. In essence, the image matching problem is shifted to a (one or two) blob matching problem. Each blob is compared with all blobs in each database image. Spatial information is thus included, yet in a very concise manner. It should be noted

that each blob is represented by a color histogram, thus the representation is a discrete representation (in the image plane as well as in feature space).

In our approach we combine the following. The image representation is a localized region representation, in which the image is first segmented into homogeneous regions in feature space. Each homogeneous region is represented by a Gaussian distribution in feature space. The set of regions in an image is represented by a Gaussian mixture model (GMM). GMM provides for a continuous representation. Images are compared and matched via a probabilistic measure of similarity between the Gaussian mixture distributions. In the following sections we will elaborate on the image representation and the proposed similarity measure.

3 Image Representation

The overall framework of the image representation and matching phases is represented in the block-diagram of Figure 1. In this section we focus on the representation phase of the system. We transition from the pixel representation to a mid-level representation of an image, in which the image is represented as a set of coherent regions in feature space. In this work we focus on the color feature. In particular we model each image as a mixture of Gaussians in the color feature space. It should be noted that the representation model is a general one, and can incorporate any desired feature space (such as texture, shape, etc) or combination thereof.

3.1 Feature extraction

An initial transition is made from pixels to the selected feature space. Color features are extracted by representing each pixel with a three-dimensional color descriptor in a selected color space. In this work we choose to work in the $L*a*b$ color space which was shown to be approximately perceptually uniform; thus distances in this space are meaningful [28]. In order to include spatial information, the (x, y) position of the pixel is appended to the feature vector. Including the position generally decreases oversegmentation and leads to smoother regions.

Following the feature extraction stage, each pixel is represented with a five-dimensional feature vector, and the image as a whole is represented by a collection of feature vectors in the

five-dimensional space. Note that the dimensionality of the feature vectors, and the feature space, is dependent on the features chosen and may be augmented if additional features are added.

3.2 Grouping pixels into regions

In this stage, pixels are grouped into homogeneous regions, by grouping the feature vectors in the selected five-dimensional feature space. The feature space is searched for dominant clusters and the image samples in feature space are then represented via the modeled clusters. The underlying assumption is that the image colors and their spatial distribution in the image plane are generated by a mixture of Gaussians. Note that although image pixels are placed on a regular (uniform) grid, this fact is not relevant to the probabilistic clustering model in which the posterior of a cluster given a pixel value is of interest. In general, a pixel is more likely to belong to a certain cluster if it is located near the cluster centroid. This observation implies a unimodal distribution of pixel positions within a cluster. A natural choice for a unimodal distribution within a GMM framework is a Gaussian distribution. The posterior is not influenced by the parametric form of the mixture distribution for the space coordinates as long as it is the same for all components and it is unimodal. Each homogeneous region in the image plane is thus represented by a Gaussian distribution, and the set of regions in the image is represented by a Gaussian mixture model. Learning a Gaussian mixture model is in essence an unsupervised clustering task.

The Expectation-Maximization (EM) algorithm is used (similar to [2]) to determine the maximum likelihood parameters of a mixture of k Gaussians in the feature space. The image is then modeled as a Gaussian mixture distribution in feature space. We briefly describe next the basic steps of the EM algorithm for the case of Gaussian mixture model. The distribution of a random variable $X \in R^d$ is a mixture of k Gaussians if its density function is :

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

such that the parameter set $\theta = \{\alpha_j, \mu_j, \Sigma_j\}_{j=1}^{k}$ consists of :

- $\alpha_j > 0 \quad , \quad \sum_{j=1}^{k} \alpha_j = 1$

- $\mu_j \in R^d$ and Σ_j is a $d \times d$ positive definite matrix.

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

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

The EM algorithm is an iterative method to obtain θ_{ML}. Given the current estimation of the parameter set θ, each iteration of the EM algorithm re-estimates the parameter set according to the following two steps :

- Expectation step :

$$w_{tj} = \frac{\alpha_j f(x_t|\mu_j, \Sigma_j)}{\sum_{i=1}^{k} \alpha_i f(x_t|\mu_i, \Sigma_i)} \qquad j = 1, ..., k \quad , \quad t = 1, ..., n \tag{3}$$

- Maximization step :

$$\begin{aligned} \hat{\alpha}_j &\leftarrow \frac{1}{n} \sum_{t=1}^{n} w_{tj} \\ \hat{\mu}_j &\leftarrow \frac{\sum_{t=1}^{n} w_{tj} x_t}{\sum_{t=1}^{n} w_{tj}} \\ \hat{\Sigma}_j &\leftarrow \frac{\sum_{t=1}^{n} w_{tj} (x_t - \hat{\mu}_j)(x_t - \hat{\mu}_j)^T}{\sum_{t=1}^{n} w_{tj}} \end{aligned} \tag{4}$$

The first step in applying the EM algorithm to the problem at hand is to initialize the mixture model parameters. The K-means algorithm is utilized to extract the data-driven initialization. The update scheme defined above allows for full covariance matrices; variants include restricting the covariance to be diagonal or scalar matrix. The updating process is repeated until the log-likelihood is increased by less than a predefined threshold from one iteration to the next. In this work we choose to converge based on the log-likelihood measure and we use a 1% threshold. Other possible convergence options include using a fixed number of iterations of the EM algorithm, or defining target measures, as well as using more strict convergence thresholds. We have found experimentally that the above convergence methodology works well for our purposes. Using EM, the parameters representing the Gaussian mixture are found. K-Means and EM are calculated for $k \geq 1$, with k corresponding to the model size.

3.3 Image model selection

It is common knowledge that the number of mixture components (or number of means), k, although often ignored, is of great importance in accurate representation of a given image. Ideally, k is to represent the value that best suits the natural number of groups present in the image. Note that each of these feature groups may include several spatially disjoint regions in the image. It is often accepted that the Minimum Description Length (MDL) principle may serve to select among values of k. This can be operationalized as follows. Choose k to maximize :

$$\log L(\theta|X) - \frac{l_k}{2} \log n \tag{5}$$

where l_k is the number of free parameters needed for a model with k mixture components. In the case of a Gaussian mixture with full covariance matrices, we have :

$$l_k = (k-1) + kd + k(\frac{d(d+1)}{2}) \tag{6}$$

As a consequence of this principle, when models using two values of k fit the data equally well, the simpler model will be chosen. In our experiments, k ranges from 3 to 6.

An example is shown in Figure 2 in which we see an input image (top) and a set of localized Gaussians representing the image for differing mixtures (different k values), bottom. In this visualization each localized Gaussian mixture is shown as a set of ellipsoids. Each ellipsoid represents the support, mean color and spatial layout, of a particular Gaussian in the image plane. This example shows that a larger number of localized Gaussians introduces important information, such as the sun, yet may over-fragment a conceptually homogeneous region, such as the sky.

3.4 Probabilistic image segmentation

An immediate transition is possible between the image representation using a Gaussian mixture model, and probabilistic image segmentation. A direct correspondence can be made between the mixture representation and the image plane. Each pixel of the original image is now affiliated with the most probable Gaussian cluster. The labeling of each pixel is done in the following manner. Suppose that the parameter set that was trained for the image is

$\theta = \{\alpha_j, \mu_j, \Sigma_j\}_{j=1}^{k}$ Denote :

$$f_j(x|\alpha_j, \mu_j, \Sigma_j) = \alpha_j \frac{1}{\sqrt{(2\pi)^d |\Sigma_j|}} \exp\{-\frac{1}{2}(x-\mu_j)^T \Sigma_j^{-1}(x-\mu_j)\} \tag{7}$$

Then the labeling of the pixel related to the feature vector x is chosen as follows:

$$\text{Label}(x) = \arg\max_j f_j(x|\alpha_j, \mu_j, \Sigma_j) \tag{8}$$

In addition to the labeling, a confidence measure can be computed. The confidence measure is a probabilistic label that indicates the uncertainty that exists in the labeling of the pixel. The probability that a pixel x is labeled j is :

$$p(\text{Label}(x) = j) = \frac{f_j(x|\alpha_j, \mu_j, \Sigma_j)}{f(x|\theta)} \tag{9}$$

Equations (7-9) provide for a probabilistic image segmentation, as shown in Figure 3. Each pixel from the original image is displayed with the color of the most-probable corresponding Gaussian. The segmentation results provide a visualization tool for better understanding the image model. Uniformly colored regions represent homogeneous regions in feature space. The associated pixels are all linked (unsupervised) to the corresponding Gaussian characteristics.

The EM algorithm, along with the model selection described above, ensures a Gaussian mixture in color and space. In essence, we have found the most dominant colors in the image, as present in homogeneous localized regions, making up the image composition. There is a significant dependency between adjacent pixels in the image plane. This dependency can be well modeled with Markov random fields [6] (which are, however, difficult to manipulate). The GMM approach models the image as an IID process. However incorporating the spatial information into the feature vector does not only supply local information. It is also imposing a correlation between adjacent pixels in such a manner such that pixels that are not far apart tend to be associated (labeled) with the same Gaussian component. Figure 3 clearly demonstrates this fact, as can be seen in the smooth nature of the segmentation that results in labeling the image according to the GMM.

Our method treats image segmentation and image matching in a unified manner. Note that in matching systems that are based on histogram representations, labeling the image

according to the histogram bins results with segmentation with no semantic meaning. On the other hand state-of-art image segmentation methods (e.g. normalized cuts [21]) can not be applied to image matching problems in a straightforward manner.

4 Image similarity and matching

The localized Gaussian mixture representation provides for a compact representation in feature space, with which we transition to the next stage of image comparison. It is our interest to define a distance measure between distributions. In this section we focus on one such measure, the Kullback-Leibler (KL) distance, and demonstrate its effectiveness.

The Kullback-Leibler distance (or relative entropy) is a measure of the distance between two distributions based on information theoretic motivation [11]. It is consistent with the probabilistic modeling technique and can be efficiently evaluated through Monte-Carlo procedures. The discrete version of the KL distance has been mentioned as a possible distance measure between histograms [15, 18]. In these studies, comparisons are conducted across a variety of distance metrics, in a world in which the image representation is global (no spatial information included), and discretized in histogram form. In such scenarios, the KL distance compares favorably with other standard measures, with no special advantages evident. The approach we are presenting in this work is a shift from the above mentioned works, in that we are preserving a continuous image representation and utilizing the KL measure as an information theoretic measure of distance between *continuous distributions*. We believe that the framework presented is a novel one in the computer-vision arena. We were able to find an interesting equivalence within the audio domain, as will be discussed in section 6.

4.1 KL Distance between images

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. Hence, a reasonable distance measure between two images is a distance measure between the two Gaussian mixture distributions obtained from the images. 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 :

$$D(f_1||f_2) = E_{f_1} \log \frac{f_1(x)}{f_2(x)} \tag{10}$$

where E is the expected value function. Since the KL distance between two Gaussian mixture distributions can not be analytically computed, we can instead apply the image data to approximate it. Denote the feature set extracted from the first image by $x_1 \ldots x_n$. The KL distance can be approximated as follows:

$$D(f_1||f_2) \cong \frac{1}{n} \sum_{t=1}^{n} \log \frac{f_1(x_t)}{f_2(x_t)} \tag{11}$$

Another possible approximation is to use synthetic samples produced from the Gaussian mixture distribution, f_1, instead of the image data. This enables us to compute the KL distance without referring to the images from which the models were built. Image retrieval experiments show no significant difference between these two proposed approximations of the KL distance.

The distance between the two models may be directional, such as in the case in which a target image is provided, and query images are matched to the target image. Distances may also be non-directional, providing for a symmetric distance measure. The symmetric version of the KL distance is :

$$\begin{aligned} d(f_1, f_2) &= \frac{1}{2}\left(D(f_1||f_2) + D(f_2||f_1)\right) \\ &\cong \frac{1}{n_1} \sum_{t=1}^{n_1} \log \frac{f_1(x_{1t})}{f_2(x_{1t})} + \frac{1}{n_2} \sum_{t=1}^{n_2} \log \frac{f_2(x_{2t})}{f_1(x_{2t})} \end{aligned} \tag{12}$$

such that $x_{i1} \ldots x_{in_i}$ is the feature set extracted from image i, $(i = 1, 2)$, and n_i is the size of this set.

Variations on the KL measure exist in the literature and may be considered. The KL divergence, $D(f_1||f_2)$, is undefined if f_1 is not absolutely continuous with respect to f_2 (i.e. the support of f_1 is not a subset of the support of f_2). A variant distance measure that overcomes this problem is the Jensen-Shannon (JS) distance, defined as [12]:

$$D^{JS}(f_1\|f_2) = 1/2[D^{KL}(f_1\|(f_1 + f_2)/2) + D^{KL}(f_2\|(f_1 + f_2)/2)] \tag{13}$$

The JS distance is a statistical test that two images are generated by the same underlying source. In contrast to the KL distance, using the average distribution ensures that the

JS divergence is always defined. In this work we continue with the KL distance measure as defined in 10. An experimental comparison between the two KL-based measures is conducted.

A major question is if the representation and the distance measure are strong enough to enable perception-like similarity. As shown in Figure 2, we may encounter situations in which similar content images are represented by differing number of regions and varying layouts. We term this problem the *fragmentation* problem. Our goal in the matching framework is to have images compared and matched regardless of this variability (dependent on k), and show robustness to it.

4.2 KL Distance between image categories

So far we have treated each image as a unique (separate) entity, an instantiation of a distribution model representing the image as a class. An interesting question arises: can we model an *image category?* Moreover, can we *match* between images and categories, and between categories?

The term category requires further definition. Based on the representation in the current implementation, a category is defined as a set of images with similar high-level content, such as "waterfall", "desert", "sunsets" etc. that exhibit visual similarity in the spatial relationships between colored regions (note that global color content does not suffice). The term "similar" is kept fuzzy to accommodate for flexibility in absolute positions, relative orientations, and differing sizes of the corresponding regions in the image set. Let $I_1, \ldots, I_n$ denote an image set for *class* C_l. Following a transition of each image to the feature space, the extracted feature set can be viewed as a set of independently identically distributed (IID) samples from the Gaussian mixture distribution representing class C_l.

Examples of category modeling can be seen in Figure 4. Three categories are presented: "desert", "monkey" and "waterfall". Ten images for each category were handpicked from the COREL database. A sample of three of the images in each class are shown left. Gaussian mixture modeling of the 10 images per class is shown right. The mixture model is learned from a combined sample set extracted from the 10 images per class. 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 category 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. In the desert example of Figure 4, we note that the lower part of the image plane may have colorings that vary from more pinkish to more yellowish (accommodating for the two leftmost sample images). A variety of bluish colors model the top part of the image, accommodating for varying shades of the sky region. In the waterfall example, we can see the fuzziness in the spatial extent of the Gaussians, as they enable variability in the location of the waterfall in the image plane. The wide support of the bluish and white Gaussians in Figure 4 is quite different from the more finely tuned modeling of a single image model, as shown in Figure 3.

We conclude that the concept of a continuous and probabilistic representation of an image can be extended to represent an image class. We can now define distance measures between an image and a class, as well as between two separate image classes (image categories). Denote the Gaussian mixture model computed for the *image class* by f_{C_l}. Given an input image distributions f_I and the class distribution f_{C_l}, the (non-symmetric version) of the KL distance is :

$$D(f_I||f_{C_l}) = E_{f_I} \log \frac{f_I(x)}{f_{C_l}(x)} \tag{14}$$

In the image to class case the distances may be directional, from the image to the class (such as a query image matched to a target class). A symmetric version may be the more appropriate measure for between category distances. For example, if categories C_l and C_h are to be compared:

$$\begin{aligned} d(C_l, C_h) &= \frac{1}{2}\left(D(C_l||C_h) + D(C_h||C_l)\right) \\ &\cong \frac{1}{n_1}\sum_{t=1}^{n_1} \log \frac{C_l(x_{1t})}{C_h(x_{1t})} + \frac{1}{n_2}\sum_{t=1}^{n_2} \log \frac{C_h(x_{2t})}{C_l(x_{2t})} \end{aligned} \tag{15}$$

such that $x_{i1} \ldots x_{in_i}$ is the feature set extracted from category i, $(i = 1, 2)$, and n_i is the size of this set. Experimental results of matching across categories will be shown in section 5.3.

5 Experimental Results

In this section we present an investigative analysis of the proposed scheme, in which we combine the GMM representation with the KL distance measure. We term the combined contin-

uous and probabilistic framework, the *GMM-KL* framework. We investigate the framework's robustness in the image matching task, we demonstrate characteristics of the framework, and show initial applications to the image retrieval task. The database used throughout is extracted from the COREL database. A set of 245 images (512*512) were chosen randomly. In addition to the random set, 70 images were hand-picked as comprising 7 different classes or categories (10 images per class). Labeled categories include: "car", "desert", "field", "monkey" , "pyramid", "snow" and "waterfall". Figure 5 shows a selection of images from the 7 labeled categories as well as the random set.

The definition of ground truth in real world imagery is a difficult and challenging task. The standard collection of images of the COREL database is categorized into high-level semantic categories. Such categories, however, are far from satisfactory in terms of indicating image-plane similarity between images. The database categorization problem is experienced by works using global representation schemes, and even more severely in works using localized image representations, once an attempt is made to provide benchmarking on large image data sets. Once a data set is chosen (e.g., the selected set in this work), an immediate question arises as to its generality across any other image set. If images are selectively hand-picked so as to abide by visual similarity constraints, an immediate legitimate issue is how general the experiment is - as compared with choosing a random set of images. In consideration of the above issues we have selected a random set of images, as well as images in a variety of labeled categories. For benchmarking experiments we suggest a new ground-truth methodology that is not sensitive to the actual category used or the number of images in the set. In each presented experiment, we address a specific scenario and investigation task, and we limit our conclusions to the scope of the respective data-sets used. [1]

[1]Note that each image in the data-sets is represented by k Gaussians in a 5-dimensional space. Each Gaussian is represented by 20 parameters (15 parameters of the covariance matrix and 5 centroids). Overall we get $20 \times k$ parameters per image that we need to learn. In most of the experiments conducted we extract approximately 2000 samples (pixels) per training image. Each such sample is 5-dimensional. We get a total of $2000 \times 5 = 10,000$ samples per image. With k in the range of $5 - 10$, we conclude that we have a sufficient number of samples per model parameter.

5.1 KL measure analysis

We start with a computational proof-of-concept for the KL distance measure. We introduce a novel *intra-inter class* statistical evaluation methodology as a benchmarking procedure to numerically evaluate the KL distance measure. The *intra-class set* of images corresponds with similar content image samples, and the *inter-class set* corresponds to pairing of images with different content.

5.1.1 Robustness to fragmentation in the image representation

Semantically similar content images may be represented by differing number of regions via the Gaussian mixture model (see Figure 2). The goal is to have images compared and matched regardless of this variability (dependent on k parameter), and show robustness to it. Note that definition (10) does not require same number of Gaussians for the two distribution f_1 and f_2. Theoretically, the continuous version of the KL distance quantifies the distance between two continuous distributions regardless of their parametric representation. Hence the combination of Gaussian mixture modeling of an image and the KL distance can overcome the problem caused by different segmentations of similar images. Our goal is to demonstrate this characteristic in practice.

In this experiment we use the random set of 245 images extracted from the COREL database. The ground-truth is generated by choosing four mixture representations (4 values of k, $k = 3, 4, 5, 6$) per input image. The "intra-class" distance set is computed as the distances between all combinations of representation models *per image*. Note that the similarity of the models within the "intra-class" set is an objective one, not dependent on subjective labeling. As each image is the source of a set of models (similarity is per image rather than per labeled class of images), the size of the dataset (number of images per class) is less relevant. We have overall a set of 12 non-zero distances per image. This process is repeated for each of the 245 images in the database for an overall $12 * 245$ distances. A second set of distances is computed *across images*, with each image represented by the MDL chosen mixture representation (the optimal k value). We term this set of distances (with $245 * 244$ distances) the "inter-class" distance set.

A histogram of the "intra-class" and "inter-class" distances is plotted in the graph pre-

sented in Figure 6 (a). The x-axis is the KL distance and the y-axis is the frequency of occurrence of the respective distance in each of the two distance sets. Two distinct modes are present, demonstrating the clear separation between the sets. The "intra-class" distances are vary narrowly spread at the lower end of the axis (close to zero), as compared to the wide-spread and larger distance values of the "inter-class" set.

The result presented in Figure 6 (a) indicates the strong similarity between same class (same image) models, as measured by the KL measure, regardless of the variability in the representation. The KL distance metric is in fact robust to fragmentation in the representation space.

5.1.2 Sensitivity of the KL distance measure to sample noise

A second benchmarking experiment is conducted to evaluate the sensitivity of the distance measure to sample noise. Here, the ground-truth database is comprised of a random set of 100 images and then randomly sampling pixels from the chosen images. For each image we create 8 disjoint sample sets, where each sample consists of 2000 pixels.

In this experiment each sample is represented by an MDL chosen Gaussian mixture (the optimal k value). Gaussian noise of 0 mean and 0.2 variance (20% deviation) was added to the ground truth samples just before the computation of the Gaussian mixture. We have overall 8 different models per image, 100 different images. "Intra-class" and "inter-class" distances are computed as before. In the "intra-class", $8 * 7$ distances are computed for each of the 100 images. The "inter-class" set distances are computed between different images.

A histogram of the "intra-class" and "inter-class" distances, with and without noise, is plotted in the graph presented in Figure 6 (b). Two distinct modes are present as before, even though the addition of noise decreased somewhat the separation between the classes. We again see that the "intra-class" distances are vary narrowly spread at the lower end of the axis (close to zero), as compared to the wide-spread and larger distance values of the "inter-class" set. Figure 6 (b) indicates the robustness of the GMM and KL distance to noise in sample space.

5.2 The combined GMM-KL framework in image retrieval

We next demonstrate the applicability of the presented framework to the image retrieval task. Each image in the database is processed to extract the localized Gaussian mixture representation. The KL distance (non-symmetric) is next computed between each of the images and an input query image. The images are sorted based on the distance and the closest ones are presented as the retrieval results.

An example is shown in Figure 7. On the top are the database images, with their respective mixture representations shown on the bottom. The input query image is presented top left, with the retrieved images sorted top down, left to right, with increasing distance values. Viewing the resultant images (query results) we note that the first two rows are very similar in their color and layout composition. Different color components (more redish, yellowish) are mostly found in the bottom right corner. Note the respective mixture representations, in which we see the corresponding region color content. The distance metric proposed is able to compare successfully between differences in the number of mixture components and their exact layout.

5.2.1 Statistical performance evaluation

Retrieval results are evaluated by precision versus recall (PR) curves. Recall measures the ability of retrieving all relevant or similar information items in the database. It is defined as the ratio between the number of relevant or perceptually similar items retrieved and the total relevant items in the database (in our case 10 relevant images per each of the labeled classes). Precision measures the retrieval accuracy and is defined as the ratio between the number of relevant or perceptually similar items and the total number of items retrieved.

Precision vs. recall (PR) curves are extracted for each of the 7 categories. A comparison with global histogram representation and several histogram distance measures is conducted. The histogram measures include the bin-2-bin Euclidean distance (Euc.), the histogram intersection measure (H. I.) and the discrete KL measure (Disc. KL) [27, 23, 18]. A binning of $8 * 8 * 8$ is used in the histogram representation. This resolution (512 quantization levels) is commonly found in the literature. This resolution is also in the same order of magnitude (and favorably so) with the GMM representation. Six of the curves are presented in Figures

8 and 9. In Figure 8 the set of images used is comprised of the 70 "labeled" images (i.e. 10 images in 7 categories). Each plot is an average of the results of the 10 query images in the class. In black is the PR curve of the GMM-KL framework. The purple, red and green curves correspond to histogram representation and Euc., H. I., and Disc. KL distance measures, respectively. In all cases we note the increase in performance with the GMM-KL. Some categories seem to be more difficult than others, such as the "snow" category, that seems to be much more difficult than the "field" category.

In Figure 9 the histogram-based distance measures are compared with the GMM and KL-based distance measures, on a dataset of 315 images (the dataset of 70 images is combined with the "random" set). In this experiment we include a plot that shows the performance of combining GMM with the JS distance of equation 13 (GMM-JS). The advantage of the GMM-KL and GMM-JS distance measures, over the histogram-based methods is evident. There is no significant difference between the performances of the two KL-based distance measures.

An additional performance evaluation, using the *rank* measure, is presented in Tables 1 and 2. For a given query image, we sort all the images in the database (according to any chosen similarity function), and define the *rank* of an image as its location in the sorted list. A valuable performance criteria is the rank of the first image in the sorted list that belongs to the query class (or the first "correct" answer). A second measure of interest is the rank of the last image in the set, i.e. the minimal number of images that need to be retrieved so that all (10) images in the set are present. We term the rank of the first true image, *rank 1*, and the rank in which the entire set is retrieved, *rank 10*. In Tables 1 and 2, the average *rank 1* location and average *rank 10* location are listed, with the average taken over 10 query cases per class. Included in the Tables are 6 different categories and varying distance measures. Note that the first position (location "1") is the query image. Position 2 is therefore the first retrieval response possible. Table 1 summarizes the 70 image set case, while Table 2 summarizes the 315 image set case. The GMM-KL achieves much better results overall (in some cases similar results). Most of the rank 1 results are in position 2 (i.e. first image retrieved is in the "true" class). Rank 10 results are at substantially smaller-number locations in the sorted list (i.e., all "true" responses are found at much smaller image sets). The advantage becomes more evident as the database is increased. Note in particular the

rank 10 results in Table 2 that show 50% to more than 100% increase in the position of the 10th category image, between the GMM-KL and histogram measures. Among the histogram distance measures, the Euclidean is very clearly the worst (as expected), with the Disc. KL measure quite consistently the best.

5.3 The combined GMM-KL framework in category modeling and category matching

The final set of experiments deals with the concept of category modeling and matching. An example of category modeling was shown in Figure 4. 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 with the results listed in Tables 3 and 4 and plotted in Figure 10. Table 3 lists distances between category models, following equation (15). In the field category, for example, the closest category model is seen to be the "waterfall" category (see Figure 5). Similar relationships may be learned automatically by the system, and used for later analysis, error prediction etc. Table 4 lists distances between *image* models (columns) and *category* models (rows). 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 (14). Note that Table 3 is symmetric while Table 4 is not.

In Figure 10 (a) corresponding rows from Tables 3 and 4 are plotted, in 6 of the categories. Category-to-category distances are shown in blue; Image-to-category distances are plotted in red. A high degree of correlation between the two scenarios is clearly visible, as shown in Figure 10 (b). Such a tight correlation suggests that image models are very similar in behavior to category models (though recall earlier examples in which specific details are clearly different between the two models). Category models do in fact represent their image building blocks. Extracting information from prelearned category models may provide a reasonable prediction of image classification performance.

6 Discussion

In this work we focus on a probabilistic image matching scheme in which the image representation is continuous, and the distance measure is also defined in the continuous domain. Each image is first represented as a Gaussian mixture distribution and images are compared and matched via the probabilistic KL distance between distributions.

We are presenting here a different approach from the well-researched approach of discrete histogram representations. We are also enabling the transition to a representation that includes spatial information with localized clustering in the spatial domain as well as in the feature (color) domain. The combination of the continuous representation of the image, along with a continuous distance measure between continuous distributions is novel. The proposed framework is theoretically appealing. The results of experiments pursued are encouraging.

The proposed methodology in this paper is a general one. We were not able to find equivalent approaches in the computer vision literature. We did in fact find an interesting equivalence with the audio literature. The audio analogy to the problem of finding the distance between two images, is the problem of computing the distance between two acoustic segments. The first step towards the solution of such an audio task is the representation of the acoustic data with a Gaussian mixture model. Distinct Gaussians are used to model the acoustic variability within a single speaker, that are caused mainly by various phoneme types. Given the Gaussian mixture model representation, the KL distance described in section 4 may be used to define a distance measure between the audio segments. A typical application is automatic audio broadcast news segmentation based on detecting changes in the acoustics [22]. Another application of this method is for speaker clustering problems [8]. The success of this general framework in solving speech problems motivates its attractiveness in approaching computer vision problems.

We view this work as a first step in an extensive research effort ahead, in which we evaluate the proposed framework on larger data-sets, as well as compare it with other more localized, image representations and distance metrics proposed in the literature. Several research questions are open with respect to the KL distance, such as the differences between the symmetric and non-symmetric variations of the metric. The KL divergence may be sensitive to cases in which the distribution f_2 vanishes where f_1 is finite, causing infinite KL-

divergence. Empirically we have found this not to be the case. Using Gaussian distribution models, with supports that cover the entire space, we do not expect to reach an event in which one of the Gaussian supports vanishes at a particular sample. Variations on the KL distance exist in the literature. One such distance measure that theoretically answers the above concern is the Jensen Shannon distance. Initial investigation indicates consistent behavior with the KL distance.

Image variations that include illumination irregularities, texture and other artifacts are not accounted for in the proposed model. Our basic assumption is that "real-world" images are smoothly-varying in feature space and in the spatial domain. Texture characteristics of regions, as well as shape and other features, can be extracted as additional features augmenting the feature space dimensionality. Different models need to be learned in this augmented representation.

A Gaussian model is a suitable representation for homogeneous regions with an ellipsoid-like shape. However, non convex regions (e.g. the yellow sky around the sun in Figure 2) are poorly represented by a Gaussian distribution. One of our future research goals is to extend the model family in a manner such that non convex regions can be better represented.

One of the main difficulties is the benchmarking process. We have overcome this problem by introducing an intra-inter class statistical evaluation methodology, in which the intra-class composition ensures sets of similar content image samples. The statistical evaluation results as presented in this work reflect the ability of the GMM-KL framework to cope with critical robustness issues.

Acknowledgment

Hayit Greenspan was supported by the Eshkol Grant of the Ministry of Science. Part of the work was supported by the Israeli Ministry of Science, Grant number 05530462.

References

[1] J. R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Jain, and C.F. Shu. Virage image search engine: an open framework for image management. In *Jain R. (ed) Symposium on Electronic Imaging:Science and Technology - Storage and Retrieval for Image and Video databases IV*, volume IV, pages 76–87, 1996.

[2] S. Belongie, C. Carson, H. Greenspan, and J. Malik. Color and texture-based image segmentation using em and its application to content based image retrieval. In *Proc. of the Int. Conference on Computer Vision*, pages 675–82, 1998.

[3] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Region-based image querying. In *Proc. of the IEEE Workshop on Content-based Access of Image and Video libraries (CVPR'97)*, pages 42–49, 1997.

[4] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley and Sons, 1991.

[5] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, and B. Dom et al. Query by image and video content: the qbic system. *IEEE Computer*, 28(9):23–32, 1995.

[6] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Inteligence*, 6(6):721–741, 1984.

[7] J. Hafner, H. Sawhney, W. Equitz, M. Flickner, and W. Niblacket. Efficient color histogram indexing for quadratic from distance functions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(7):729–739, 1995.

[8] L. Heck and A. Sankar. Acoustic clustering and adaptation for improved speaker recognition. In *Proc. of Speech Recognition Workshop, ARPA*, 1997.

[9] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih. Image indexing using color correlograms. In *Proc. of the IEEE Comp. Vis. And Patt. Rec.*, pages 762–768, 1997.

[10] P. Kelly, M. Cannon, and D. Hush. Query by image example: the candid approach. In *Proceedings of SPIE Storage and Retrieval for Image and Video Databases*, pages 238–248, 1995.

[11] S. Kullback. *Learning Textures*. Dover, 1968.

[12] J. Lin. Divergence measures based on the shannon entropy. *IEEE Trans. on Information Theory*, 37(1):145–151, 1991.

[13] W. Ma and B. Manjunath. Netra: A toolbox for navigating large image databases. In *Proceedings of IEEE Int. Conf. On Image Proc.*, pages 568–571, 1997.

[14] V. Ogle and M. Stonebraker. Chabot: Retrieval from a relational database of images. *IEEE Computer*, 28(9):40–48, 1995.

[15] T. Ojala, M. Pietikainen, and D. Harwood. A comparative study of texture measures with classification based feature distribution. *Pattern Recognition*, 29(1):51–59, 1996.

[16] G. Pass and R. Zabih. Comparing images using joint histograms. *Multimedia Systems*, 7:234–240, 1999.

[17] A. Pentland, R. Picard, and S. Sclaroff. Photobook: tools for content based manipulation of image databases. In *Proceedings of SPIE Conference on Storage and Retrieval of Image and Video Databases II*, volume 2185, pages 34–47, San-Jose, CA, Feb 1994.

[18] J. Puzicha, J.M. Buhmann, Y. Rubner, and C. Tomasi. Empirical evaluation of dissimilarity measures for color and texture. In *Proceedings of the Int. Conference on Computer Vision*, pages 1165–72, 1999.

[19] J. Puzicha, T. Hofmann, and J. M. Buhmann. Non-parametric similarity measures for unsupervised texture segmentation and image retrieval. In *Proceeding of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 267–272, 1997.

[20] Y. Rubner. *The Earth Mover's Distance as a Metric for Image retrieval.* PhD thesis, Stanford University, 1999.

[21] J. Shi and J. Malik. Normalized cuts and image segmentation. In *Proc IEEE Conf. on Computer Vision and Pattern Recognition*, pages 731–737, 1997.

[22] M. Siegler, U. Jain, B. Ray, and R. Stern. Automatic segmentation, classification and clustering of broadcast news audio. In *Proceeding of the ARPA Speech Recognition Workshop*, pages 97–99, 1997.

[23] J. R. Smith. *Integrated Spatial and Feature Image Systems: Retrieval, Analysis and Compression.* PhD thesis, Columbia University, 1997.

[24] J. R. Smith and S-F Chang. Tools and techniques for color image retrieval. In *Proceedings of SPIE Storage and Retrieval for Image and Video Databases*, volume 2670, pages 426–437, 1996.

[25] J. R. Smith and S-F Chang. Integrated spatial and feature image query. *Multimedia Systems*, 7:129–140, 1999.

[26] M. Stricker and A. Dimai. Spectral covariance and fuzzy regions for image indexing. *Machine Vision and Applications*, 10(2):66–73, 1997.

[27] M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

[28] G. Wyszecki and W. Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulae.* Wiley, 1982.

[29] W. Xiong, J. C-M Lee, and R-H Ma. Automatic video data structuring through shot partitioning and key-frame computing. *Machine Vision and Applications*, 10:51–65, 1997.

[30] H. J. Zhang, A. Kankanhalli, and S. W. Smoliar. Automatic partitioning of full-motion video. *Multimedia Systems*, 1:10–28, 1993.

FIGURE CAPTIONS

- Figure 1: A block diagram of the image matching system.
- Figure 2: Example of an input image (top) with the corresponding set of representative Gaussian mixtures (bottom). The mixtures are composed of: K=3,4,5 and 6 components.
- Figure 3: Image modeling via Gaussian mixture (center) provides for a probabilistic image segmentation (right).
- Figure 4: Category modeling using Gaussian mixture distributions. A sample of 3 images per category is shown left (a) and the generated category model (based on 10 images) is shown right (b). Each Gaussian is displayed as a localized colored ellipsoid. Note that some ellipsoids overlap in the spatial domain (thus not all are explicitly present in the image).
- Figure 5: Sample images from data-set. Each row is a category: (a) car; (b) desert; (c) field; (d) monkey; (e) pyramid; (f) snow; (g) waterfall; (h) noise.
- Figure 6: Statistical analysis of intra-class distances vs. inter-class distances. (a) Robustness to model fragmentation. (b) Sensitivity to noise. Red and blue curves are intra-inter distances without noise; green and black curves are intra-inter distances with noise.
- Figure 7: Retrieval example: Images and image representations (top and bottom, respectively). Top left image is the input query image. Ordered from top to bottom, left to right, are the retrieved images.
- Figure 8: Precision vs. Recall for 6 categories. The set of images used is comprised of the 70 "labeled" images (10 images in 7 categories). Each plot is an average of the results of the 10 query images in the class. In black is the PR curve of the GMM-KL framework. The purple, red and green curves correspond to histogram representation and Euc., H. I., and Disc. KL distance measures, respectively.
- Figure 9: Precision vs. Recall. 315 images in database. Each plot is an average of the results of the 10 query images in the class. In black is the PR curve of the GMM-KL framework. In cyan is the PR curve of GMM-JS framework. The purple, red, green and blue curves correspond to histogram representation and Euc., H. I., and Disc. KL distance measures, respectively.
- Figure 10: Statistical analysis: (a) Category model to category model matching, and image model to category model matching. (b) Correlation between image models and category models.

Class	Rank 1				Rank 10			
	GMM-KL	Euc.	H. I.	Disc. K.L.	GMM-KL	Euc.	H. I.	Disc. K. L.
field	2	3	4	3	41	46	55	52
snow	2	3	2	2	28	55	35	26
car	2	3	2.5	2.5	27.6	36.5	38	35
desert	2	5	3.5	3	60	68	66	64
monkey	2	2	2	2	15	31	23	21
waterfall	2	4	3.5	3	20	36	25	22

Table 1: Statistical Results - 70 images. Average *rank 1* location and average *rank 10* location are listed, with the average taken over 10 query cases per class. Included are 6 different categories and varying distance measures. Note that position 2 is the first retrieval response possible (position 1 is left for the query image itself).

Class	Rank 1				Rank 10			
	GMM-KL	Euc.	H. I.	Disc. K.L.	GMM-KL	Euc.	H. I.	Disc. K. L.
field	2.5	6	6	4	54.5	103	112	105
snow	2	2	2	2	11	22	18	12
car	5.8	9.5	7	7	130	195	171	161
desert	2	6	4	3	74	105	84	77
monkey	2	3	2	2	30	137	87	78
waterfall	2	13	8	6	51	158	99	85

Table 2: Statistical Results - 315 images. Average *rank 1* location and average *rank 10* location are listed, with the average taken over 10 query cases per class. Included are 6 different categories and varying distance measures. Note that position 2 is the first retrieval response possible (position 1 is left for the query image itself).

Class Model	car	desert	field	monkey	pyramid	snow	waterfall
car	0	17.77	20.02	8.01	17.21	21.10	15.26
desert	17.72	0	13.97	22.29	15.30	15.80	19.77
field	20.16	14.57	0	18.20	17.33	22.67	12.89
monkey	7.94	22.31	17.20	0	20.62	28.01	13.27
pyramid	17	15.33	17.28	21.41	0	32.87	20.21
snow	20.87	15.31	21.91	27.84	32.04	0	20.77
waterfall	15.17	19.45	12.81	13.28	20.50	21.49	0

Table 3: Category Model to category Model analysis. Listed are distances between *category* models (columns) and *category* models (rows).

Image model	car	desert	field	monkey	pyramid	snow	waterfall
car	14.42	27.03	32.32	21.56	25.07	31.06	25.41
desert	38.79	24.89	39.22	50.16	34.82	34.81	38.88
field	41.24	35.67	22.60	40.68	37.28	41.02	32.06
monkey	20.32	33.80	30.83	11.30	29.93	41.62	22.11
pyramid	31.97	28.50	34.43	38.30	20.27	42.59	30.35
snow	35.71	30.67	38.80	48.59	44.12	18.22	33.23
waterfall	29.98	33.33	31.68	28.13	27.77	34.11	13.09

Table 4: Image model to Category Model analysis. Listed are distances between *image* models (columns) and *category* models (rows). 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.

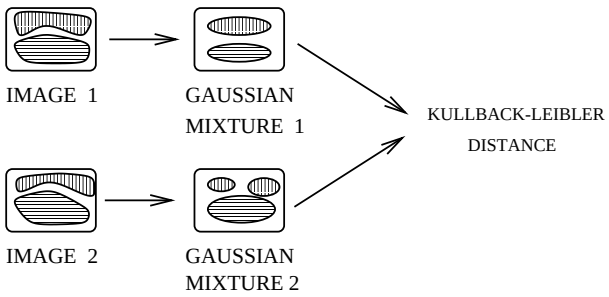


Figure 1: A block diagram of the image matching system.

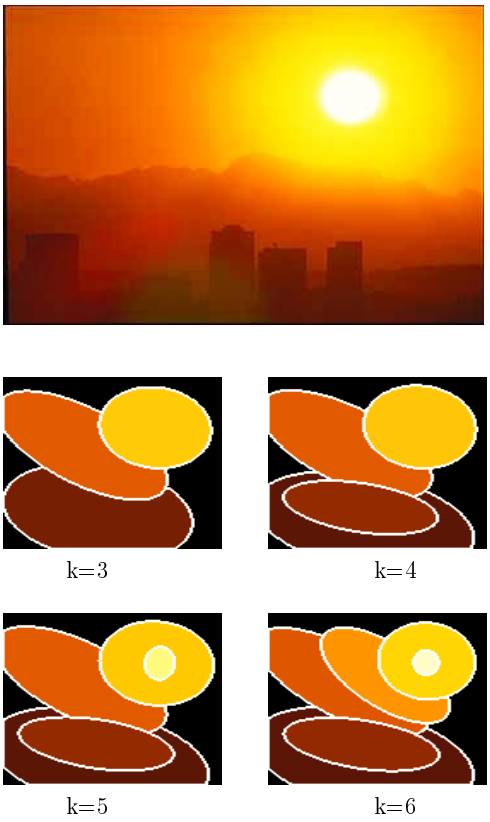


Figure 2: Example of an input image (top) with the corresponding set of representative Gaussian mixtures (bottom). The mixtures are composed of: K=3,4,5 and 6 components.

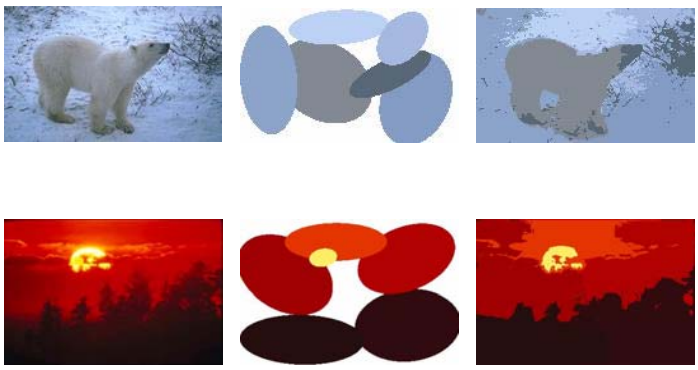


Figure 3: Image modeling via Gaussian mixture (center) provides for a probabilistic image segmentation (right).

(a) (b)

Figure 4: Category modeling using Gaussian mixture distributions. A sample of 3 images per category is shown left (a) and the generated category model (based on 10 images) is shown right (b). Each Gaussian is displayed as a localized colored ellipsoid. Note that some ellipsoids overlap in the spatial domain (thus not all are explicitly present in the image).

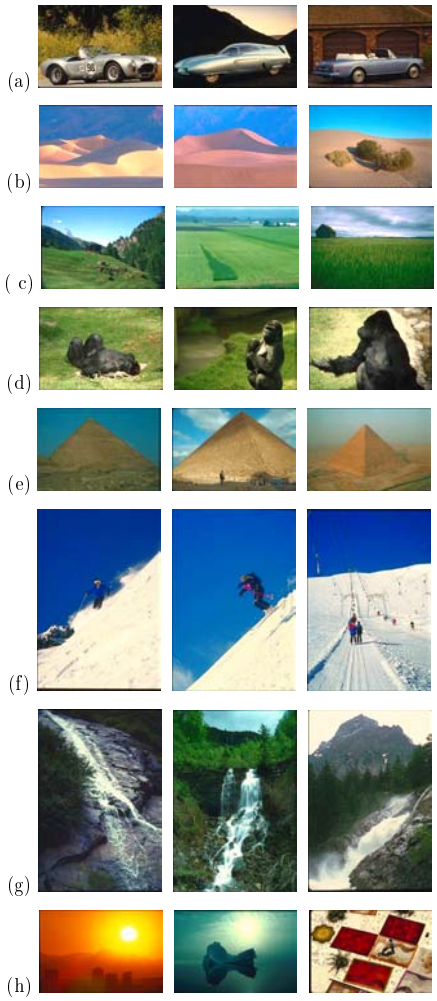
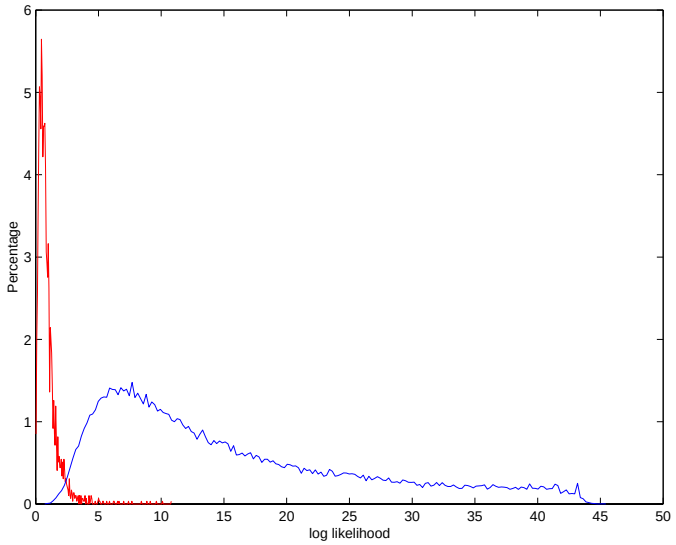
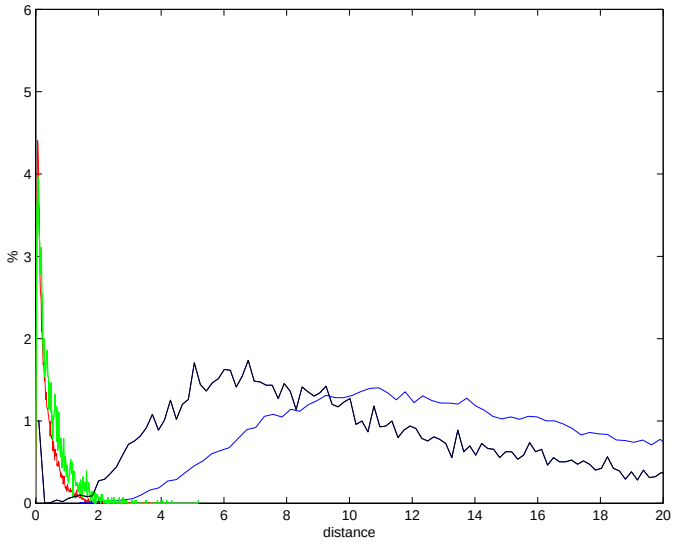


Figure 5: Sample images from data-set. Each row is a category: (a) car; (b) desert; (c) field; (d) monkey; (e) pyramid; (f) snow; (g) waterfall; (h) noise.



(a)



(b)

Figure 6: Statistical analysis of intra-class distances vs. inter-class distances. (a) Robustness to model fragmentation. (b) Sensitivity to noise. Red and blue curves are intra-inter distances without noise; green and black curves are intra-inter distances with noise.

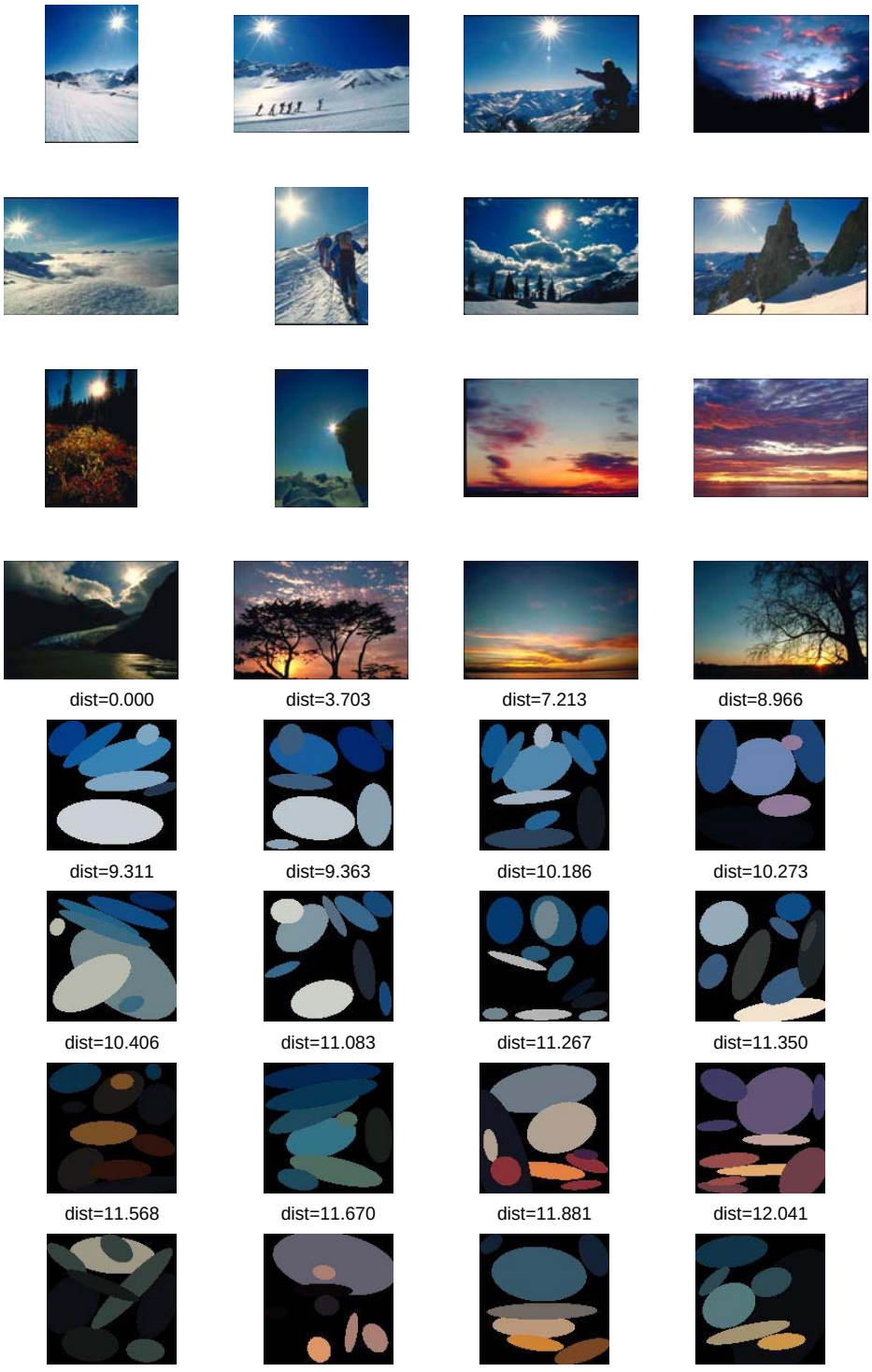


Figure 7: Retrieval example: Images and image representations (top and bottom, respectively). Top left image is the input query image. Ordered from top to bottom, left to right, are the retrieved images.

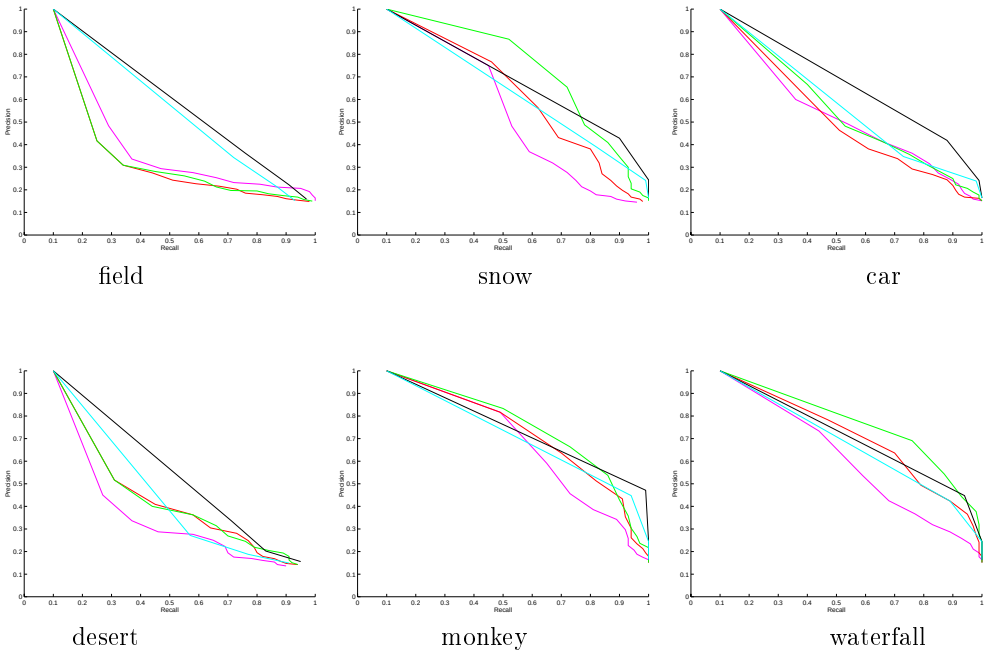


Figure 8: Precision vs. Recall for 6 categories. The set of images used is comprised of the 70 "labeled" images (10 images in 7 categories). Each plot is an average of the results of the 10 query images in the class. In black is the PR curve of the GMM-KL framework. The purple, red and green curves correspond to histogram representation and Euc., H. I., and Disc. KL distance measures, respectively.

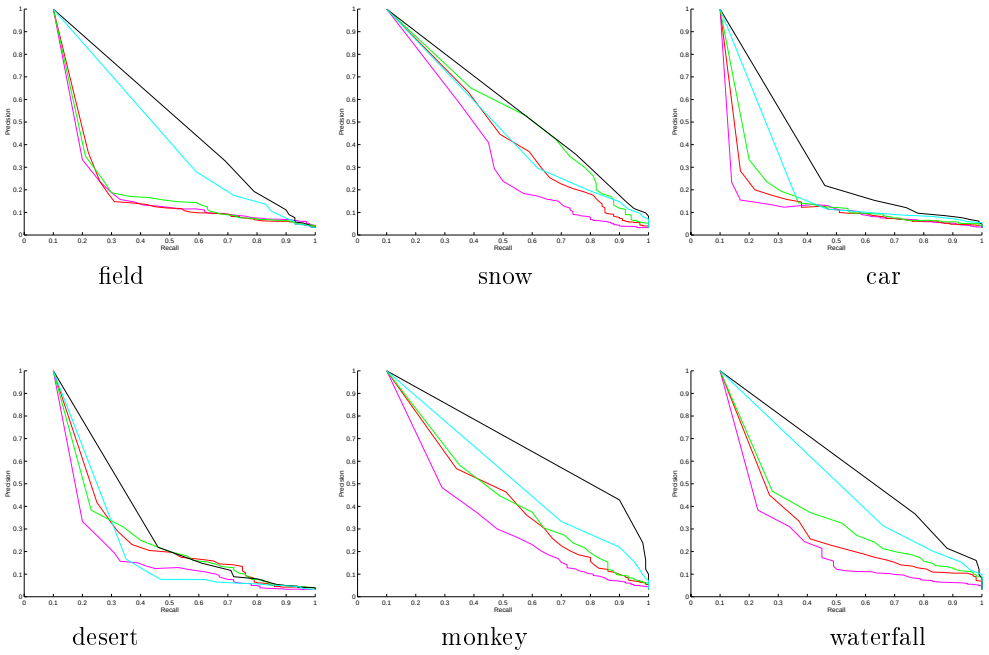


Figure 9: Precision vs. Recall. 315 images in database. Each plot is an average of the results of the 10 query images in the class. In black is the PR curve of the GMM-KL framework. In cyan is the PR curve of GMM-JS framework. The purple, red, green and blue curves correspond to histogram representation and Euc., H.I., and Disc. KL distance measures, respectively.

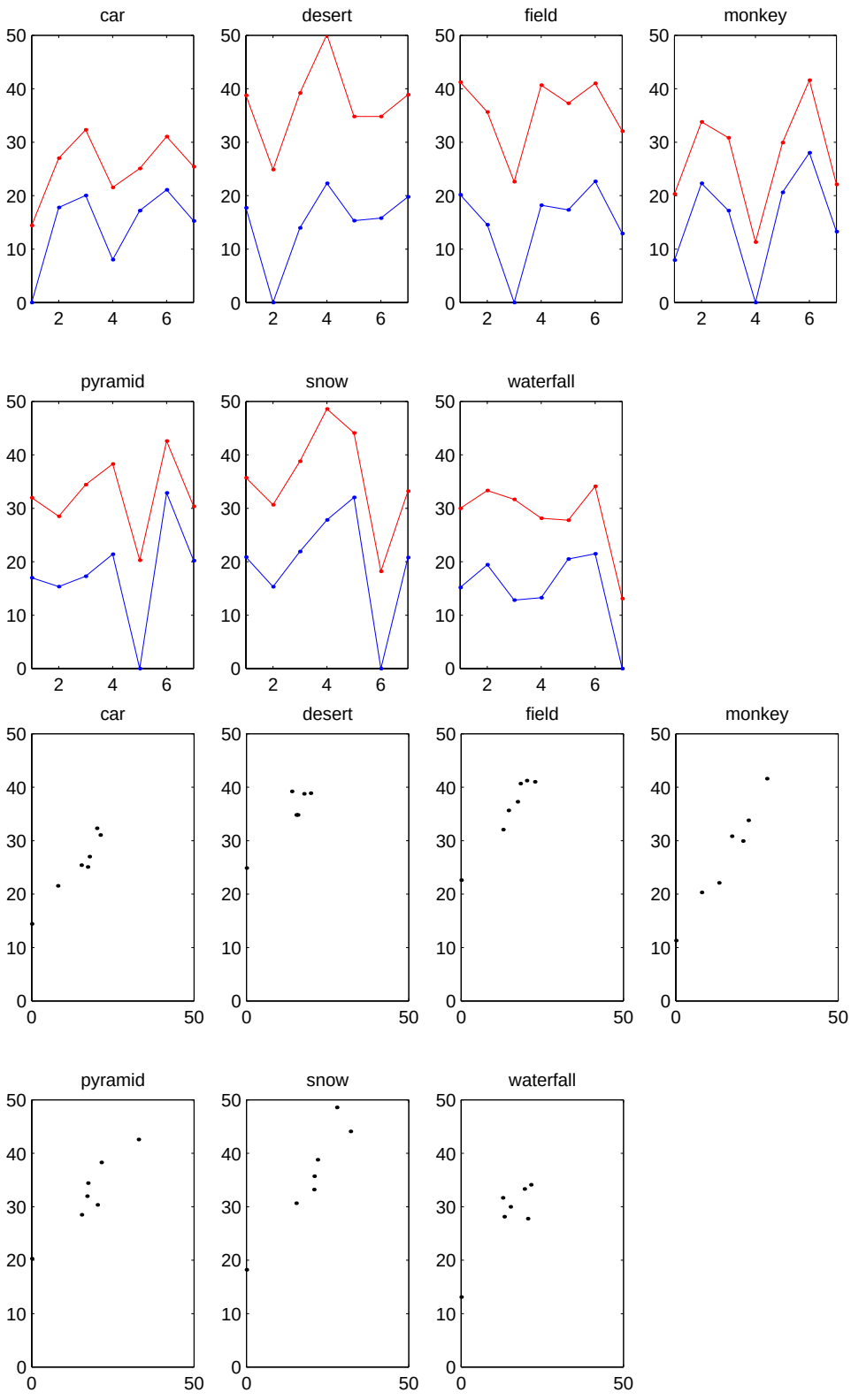


Figure 10: Statistical analysis: (a) Category model to category model matching, and image model to category model matching. (b) Correlation between image models and category models.