An Efficient Image Similarity Measure based on Approximations of KL-Divergence Between Two Gaussian Mixtures

Jacob Goldberger

Cute Systems
Tel Aviv
Israel

Shiri Gordon   Hayit Greenspan

The Engineering Department
Tel-Aviv University
Israel

Abstract

In this work we present two new methods for approximating the Kullback-Liebler (KL) divergence between two mixtures of Gaussians. The first method is based on matching between the Gaussian elements of the two Gaussian mixture densities. The second method is based on the unscented transform. The proposed methods are utilized for image retrieval tasks. Continuous probabilistic image modeling based on mixtures of Gaussians together with KL measure for image similarity, can be used for image retrieval tasks with remarkable performance. The efficiency and the performance of the KL approximation methods proposed are demonstrated on both simulated data and real image data sets. The experimental results indicate that our proposed approximations outperform previously suggested methods.
Keywords: image similarity; Kullback-Liebler divergence, mixture of Gaussians, unscented transformation.

1 Introduction

Image matching is an important component in many applications that require comparing images based on their content. The most important examples are image data base retrieval systems. Image matching techniques can be classified according to two parameters. The first is the image representation method and the second is a definition of appropriate distance measure to compare between images in the selected representation space. A standard representation method is color histogram. The advantages and disadvantages of this method are well studied and many variations exist. A variety of measures have been proposed for dissimilarity between two histograms (e.g. χ^2 statistics, KL-divergence) [9]. An alternative image representation is a continuous probabilistic framework based on a Mixture of Gaussians model (MoG) [1] [3]. The KL-divergence is a natural dissimilarity measure between two images represented by mixture of Gaussians. However, since there is

no closed form expression for the KL-divergence between two MoGs, computing this distance measure is done using Monte-Carlo simulations. Monte-Carlo simulations may cause a significant increase in computational complexity which can be a major drawback in real content based image retrieval systems. In this work we aim to solve this drawback by presenting two new methods for the approximation of the KL-divergence between two mixtures of Gaussians. The first one is an improved version of the approximation suggested by Vasconcelos [10]. The method is based on matching between the Gaussian elements of the two MoG densities and on the existence of a closed form solution for the KL-divergence between two Gaussians. The second method demands a little more processing time but gives much better results. It is based on the unscented transform introduced by Juiler and Uhlmann [4]. The rest of the paper is organized as follows. Image modeling via a mixture of Gaussians is reviewed in Section 2. In Section 3 we propose an easily computed approximation of the KL-distance between two mixtures of Gaussians. In Section 4 we propose an alternative approximation based on the unscented transform mechanism. In Section 5 we compare both the performance and the computational efficiency of the various KL-divergence approximations. The comparison is performed on both simulated data and MoG densities obtained from modeling real images.

2 Image Modelling via MoG

We model an image as a set of coherent regions. Each homogeneous region in the image plane is represented by a Gaussian distribution, and the set of all the regions in the image is represented by a Gaussian mixture model. The image is, therefore, viewed as an instance of the generative mixture of Gaussians model. We focus here 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 incorpo-

Figure 1: Input image (left). Image modeling via a mixture of Gaussians (center). Image segmentation using the learned model (right).

rate any desired feature space (such as texture, and shape) or combination thereof. 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 by Wyszecki and Stiles [11], thus distances in this space are meaningful. In order to include spatial information, the (x, y) position of the pixel is appended to the feature vector. Including the position enables a localized representation. Each pixel is represented by a five-dimensional feature vector (L,a,b,x,y). Pixels are grouped into homogeneous regions, by grouping the feature vectors in the selected five-dimensional feature space. The Expectation-Maximization (EM) algorithm is used to determine the maximum likelihood parameters of a mixture of k Gaussians. The Minimum Description Length (MDL) principle serves to select among values of k. In our experiments, k ranges from 4 to 8. Figure 1 shows an example of learning a MoG model for a given input image. In this visualization the 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. Using the learned model (center) each pixel of the original image is affiliated with the most probable Gaussian, providing for a probabilistic image segmentation (right).

Given the representation of an image by a density function, we can define a similarity measure between two images as the Kullback-Leibler divergence [8] between the respective density models of the images. In the case of discrete (histogram) representations, the KL-divergence can be easily obtained. However, there is no closed-form expression for the KL-divergence between two mixtures of Gaussians. We can use, instead, Monte-Carlo simulations to approximate the KL-divergence between two MoGs, f and g:

$$KL(f||g) = \int f \log \frac{f}{g} \approx \frac{1}{n} \sum_{t=1}^{n} \log \frac{f(x_t)}{g(x_t)}$$

such that $x_1, ..., x_n$ are sampled from $f(x)$. The problem with this approach is that it can not be used in image retrieval systems due to its large complexity. In the following sections we present two alternative deterministic approximations that can be computed much more efficiently than the Monte-Carlo based method.

3 Matching based Approximation

Let $f(x) = \sum_{i=1}^{n} \alpha_i f_i(x)$ and $g(x) = \sum_{j=1}^{m} \beta_j g_j(x)$ be two mixture densities such that $\alpha = \{\alpha_1, ..., \alpha_n\}$ and $\beta = \{\beta_1, ..., \beta_m\}$ are discrete distributions and f_i, g_j are arbitrary continuous densities. Assume that it is not possible to obtain a closed-form expression for the Kullback-Liebler divergence $KL(f||g)$ but there is an analytical way to compute the KL-divergence between each pair of components f_i, g_j. In this section we present and motivate an approximated expression for $KL(f||g)$ based on the KL-divergence between the mixtures components $KL(f_i||g_j)$.

The convexity of the KL-divergence [2] implies that:

$$KL(\sum_{i=1}^{n} \alpha_i f_i || \sum_{j=1}^{m} \beta_j g_j) \leq \sum_{i,j} \alpha_i \beta_j KL(f_i||g_j).$$

The resultant weighted average approximation is one possible approximation for the KL-divergence. This approximation is too crude, however, especially when each mixture density is composed of distributions which are unimodal and the modes are far apart. A better approximation can be obtained by matching a *single* component of $g(x)$ to each component of $f(x)$. A matching function between the components of $f(x)$ and $g(x)$ is needed.

We propose the following, matching-based approximation:

$$KL(f||g) = \sum_{i=1}^{n} \alpha_i \int f_i \log f - \sum_{i=1}^{n} \alpha_i \int f_i \log g \approx$$

$$\sum_{i=1}^{n} \alpha_i \int f_i \log \alpha_i f_i - \sum_{i=1}^{n} \alpha_i \max_j \int f_i \log \beta_j g_j$$

$$= \sum_{i=1}^{n} \alpha_i \min_j \left(KL(f_i||g_j) + \log \frac{\alpha_i}{\beta_j} \right).$$

This approximation is based on the assumption that the term in the sum $\sum_j \beta_j g_j$ that is most proximal to f_i dominates the integral $\int f_i \log g$.

The proposed approximation yields a matching function between elements of f and elements of g. Define the matching function $\pi : \{1, ..., n\} \to \{1, ..., m\}$ between the components of $f(x)$ and the components of $g(x)$ as follows:

$$\pi(i) = \arg\min_j (KL(f_i||g_j) - \log \beta_j). \qquad (1)$$

Utilizing π we can write the suggested approximation in the following way:

$$KL_{match}(f||g) = \sum_{i=1}^{n} \alpha_i \left(KL(f_i||g_{\pi(i)}) + \log \frac{\alpha_i}{\beta_{\pi(i)}} \right). \qquad (2)$$
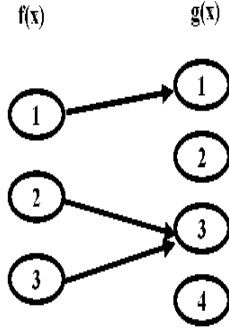
Figure 2: A possible match between a mixture of 3 Gaussians and a mixture of 4 Gaussians

Figure 2 shows a possible matching function. Several components of f may be matched to the same component of g. There can be components of g that no component of f is matched to.

We focus next on the image retrieval application. The following situation is common in image retrieval systems. Given a query mixture density f and a family of mixture densities $\{g_t\}$, we want to affiliate f to the density that minimizes the distance criterion $KL(f||g_t)$ (for a content-based image retrieval system in which mixture densities are used as the image models and the KL measure is used as the distance measure, see [3]). Since

$$\arg \min_t KL(f||g_t) = \arg \max_t \sum_{i=1}^{n} \alpha_i \int f_i \log g_t \quad (3)$$

for each MoG g, we need only to evaluate $\int f \log g$. We can apply the approximation:

$$\int f_i \log(\sum_{j=1}^{m} \beta_j g_j) \geq \max_j \int f_i \log \beta_j g_j$$

to obtain a lower bound approximation:

$$\int f \log g = \sum_{i=1}^{n} \alpha_i \int f_i \log g \approx \sum_{i=1}^{n} \alpha_i \int f_i \log(\beta_{\pi(i)} \cdot g_{\pi(i)})$$

where π is the matching function defined in expression (1).

The suggested approximation will be justified empirically in section 5. As a motivation for the approximation, we show next that the proposed approximation (Equation 2) can be viewed as a KL-divergence between the *complete versions* of the two MoGs.

A mixture model $f(x) = \sum_{i=1}^{n} \alpha_i f_i(x)$ can be viewed as a two step model. In the first step we sample a latent discrete random variable I according to $p(I = i) = \alpha_i$. In the second step we sample the observed continuous random variable x according to $f(x|i) = f_i(x)$. The complete

data is the union of the latent and the observed data. The complete data density function associated with the mixture density $f(x)$ is $f(i, x) = f(i)f(x|i) = \alpha_i f_i(x)$. Note that if the latent variables of f and g share the same alphabet (i.e. $n = m$) then the KL-divergence between the complete data densities associated with f and g is well defined and has the following closed form expression:

$$KL(f(i,x)||g(i,x)) = KL(f(i)||g(i)) + KL(f(x|i)||g(x|i))$$

$$= KL(\alpha||\beta) + \sum_{i=1}^{n} \alpha_i KL(f_i||g_i)$$

such that:

$$KL(\alpha||\beta) = \sum_{i=1}^{n} \alpha_i \log \frac{\alpha_i}{\beta_i}.$$

The chain rule for relative entropy [2] implies that:

$$KL(f(x)||g(x)) \leq KL(f(i,x)||g(i,x)).$$

Thus we obtain an upper bound for $KL(f||g)$. Since the MoG $g(x)$ is invariant to a permutation of the alphabet of the hidden random variable we can obtain a tighter bound:

$$KL(f(x)||g(x)) \leq \min_s \sum_{i=1}^{n} \alpha_i (KL(f_i||g_{s(i)}) + \log \frac{\alpha_i}{\beta_{s(i)}})$$

such that the minimization is performed over all the $n!$ permutations on the set $\{1, ..., n\}$. This approximation, which is suitable only for the special case $n = m$, can be computed by the assignment algorithm [7] whose complexity is high ($O(n^3)$).

We return to the general case where $g = \sum_{j=1}^{m} \beta_j g_j$. Let π be the matching function defined in expression (1). We can build a new mixture density:

$$g_\pi(x) = \frac{1}{C_\pi} \sum_{i=1}^{n} \beta_{\pi(i)} \cdot g_{\pi(i)}(x)$$

such that C_π is the normalization scalar $\sum_{i=1}^{n} \beta_{\pi(i)}$. The MoG g_π is a mixture density composed of the components of g and it has the same number of components as $f(x)$. Standard information theory manipulations reveal that the proposed approximation (Equation 2) can be rewritten in the following way:

$$KL_{match}(f||g) = KL(f(i,x)||g_\pi(i,x)) - \log(C_\pi)$$

such that $f(i, x)$ is the density of the complete data including the hidden discrete variable of the mixture density, i.e. $f(i, x) = \alpha_i f_i(x)$ and $g_\pi(i, x) = \frac{1}{C_\pi} \beta_{\pi(i)} g_{\pi(i)}(x)$.

Therefore, the proposed approximation is based on two principles. The first one is a matching between each component of $f(x)$ to one of the components of $g(x)$. The matching function ensures that the hidden variables of the two

mixture models are defined on the same alphabet such that it is meaningful to consider the KL-divergence between the complete data version of the densities. The second point in the suggested formula is approximating the distance between two mixture densities by the distance of the density functions of the associated complete data densities.

So far we developed an approximation method for a general mixture model. We shall now concentrate on the case which is mixture of Gaussians (MoG). The KL-divergence between the Gaussians $N(\mu_1, \Sigma_1)$ and $N(\mu_2, \Sigma_2)$ is:

$$\frac{1}{2}(\log \frac{|\Sigma_2|}{|\Sigma_1|} + Tr(\Sigma_2^{-1}\Sigma_1) + (\mu_1 - \mu_2)^T\Sigma_2^{-1}(\mu_1 - \mu_2)). \quad (4)$$

Given two mixture of Gaussians

$$f = \sum_{i=1}^{n} \alpha_i N(\mu_{1,i}, \Sigma_{1,i}) \quad \text{and} \quad g = \sum_{j=1}^{m} \beta_j N(\mu_{2,j}, \Sigma_{2,j})$$

we can plug expression (4) into approximation (2) to obtain the approximation of the KL-divergence for the case of MoG.

Another approximation of the KL-divergence between two MoGs was suggested by Vasconcelos [10]. The method is similar to the one presented in this section. The only difference is that the matching function π between the elements of the two MoGs, used in [10], is based on the Mahalanobis distance:

$$\pi(i) = \arg \min_{j} \left((\mu_{1,i} - \mu_{2,j})^T\Sigma_{2,j}^{-1}(\mu_{1,i} - \mu_{2,j}) \right) \quad (5)$$

where as in our approach:

$$\pi(i) = \arg \min_{j} \big(\frac{1}{2}(\log \frac{|\Sigma_{2,j}|}{|\Sigma_{1,i}|} + Tr(\Sigma_{2,j}^{-1}\Sigma_{1,i}) +$$

$$(\mu_{1,i} - \mu_{2,j})^T\Sigma_{2,j}^{-1}(\mu_{1,i} - \mu_{2,j})) - \log \beta_j \big).$$

In section 5 we empirically compare the performance of these two variants.

4 Unscented Transform based Approximation

The matching based method approximates well the KL-divergence if the Gaussian elements are far apart. However, if there is a significant overlap between the Gaussian elements, then the match of a single component of $g(x)$ with each component of $f(x)$ becomes less accurate. Replacing the deterministic matching by a stochastic one doesn't help since we can easily observe that the matching approximation (2) can be written as:

$$KL_{match}(f||g) = \min_{\Psi} \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \Psi_{ij}(\log \frac{\alpha_i}{\beta_j} + KL(f_i||g_j))$$

such that Ψ is a $n \times m$ stochastic matrix, i.e. the minimization over all the stochastic matrices yields a deterministic matching.

To handle overlapping situations we propose another approximation based on the unscented transform. The unscented transformation is a method for calculating the statistics of a random variable which undergoes a non-linear transformation [4]. It is successfully used for nonlinear filtering. The Unscented Kalman filter (UKF) [5] is more accurate, more stable and far easier to implement than the extended Kalman filter (EKF). In cases where the process noise is Gaussian it is also better than the particle filter which is based on Monte-Carlo simulations. Unlike the EKF which uses the first order term of the Taylor expansion of the non-linear function, the UKF uses the true nonlinear function and approximates the distribution of the function output. In this section we show how we can utilize the unscented transform mechanism to obtain an approximation for the KL-divergence between two MoGs. Let x be a d-dimensional normal random variable $x \sim f(x) = N(\mu, \Sigma)$ and let $h(x) : R^d \to R$ be an arbitrary non-linear function. We want to approximate the expectation of $h(x)$ which is $\int f(x)h(x)dx$. The unscented transform approach is the following. A set of $2d$ "sigma" points are chosen as follows:

$$x_k = \mu + (\sqrt{d\Sigma})_k \qquad k = 1, ..., d$$
$$x_{d+k} = \mu - (\sqrt{d\Sigma})_k \qquad k = 1, ..., d$$

such that $(\sqrt{\Sigma})_k$ is the k-th column of the matrix square root of Σ. Let UDU^T be the singular value decomposition of Σ, such that $U = \{U_1, ..., U_d\}$ and $D = diag\{\lambda_1, ..., \lambda_d\}$ then $(\sqrt{\Sigma})_k = \sqrt{\lambda_k}U_k$. These sample points completely capture the true mean and variance of $f(x)$ (see Figure 3). The uniform distribution over the points $\{x_k\}_{k=1}^{2d}$ has mean μ and covariance matrix Σ. Given the "sigma" points we define the following approximation:

$$\int f(x)h(x)dx \approx \frac{1}{2d} \sum_{k=1}^{2d} h(x_k). \quad (6)$$

Although this approximation algorithm resembles a Monte-Carlo method, no random sampling is used thus only a small number of points are required. The basic unscented method can be generalized. The mean of the Gaussian distribution μ can be also included in the set of sigma points. Scaling parameters can provide an extra degree of freedom to control the scaling of the sigma points further or towards μ [6]. In the implementation presented in this paper in which the dimensionality of the distributions is five (see Section 2), including μ in the set of sigma point did not cause any improvement in performance.

It can be verified that if $h(x)$ is a quadratic function then the approximation is precise. For example in the case of
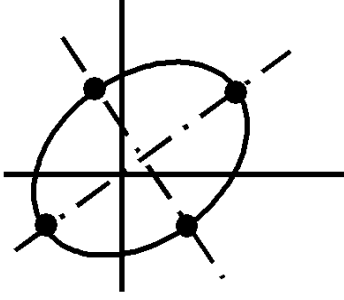
Figure 3: The sigma points of the unscented transform

ϵ	MC-100	Mahalanobis match	KL match	unscented
0.0001	99	99	99	100
0.0004	98	98	98	99
0.0016	97	96	97	99
0.0064	96	91	95	99
0.0256	94	83	91	96
0.1024	92	68	84	93
0.4096	86	58	74	90
0.8192	83	53	73	86
Relative time	1.0	0.04	0.07	0.46

Table 1: Comparison between retrieval simulation results using different approximations to the KL-divergence between mixture of Gaussians

$h(x) = \log N(\mu_2, \Sigma_2)$, $h(x)$ is a quadratic function of x. Hence expression (6) is a method, alternative to expression (4), to compute the exact KL-divergence between two Gaussian distributions. In the case that $h(x)$ is the log density function of MoG, expression (6) is an approximation. Given two mixtures of Gaussians:

$$f = \sum_{i=1}^{n} \alpha_i N(\mu_{1,i}, \Sigma_{1,i}) \quad \text{and} \quad g = \sum_{j=1}^{m} \beta_j N(\mu_{2,j}, \Sigma_{2,j})$$

the approximation of $\int f \log g$ based on the unscented transform is:

$$\frac{1}{2d} \sum_{i=1}^{n} \alpha_i \sum_{k=1}^{2d} \log g(x_{i,k})$$

such that:

$$
\begin{aligned}
x_{i,k} &= \mu_{1,i} + (\sqrt{d\Sigma_{1,i}})_k && k = 1, ..., d, \quad (7) \\
x_{i,d+k} &= \mu_{1,i} - (\sqrt{d\Sigma_{1,i}})_k && k = 1, ..., d.
\end{aligned}
$$

5 Experimental Evaluation

In order to compare the accuracy of the proposed approximations as well as their processing efficiency we conducted the following simulation of a retrieval task based on image similarity. In each retrieval session we sample a random MoG f as a query object and four other random MoGs $\{g_t\}$ as a data-set. The task is to find for a given f, the member of the data-set that is most similar to it, i.e., the retrieval task is to find:

$$\arg\min_t KL(f||g_t) = \arg\max_t \int f \log g_t.$$

Gaussian mixture models were randomly sampled according to the following rules. The number of Gaussians within the MoG was chosen uniformly in the range 4-8. The dimension of all the Gaussians was 5. For each Gaussian $N(\mu, \Sigma)$, μ was sampled from $N(0, I)$ and Σ was sampled from the Wishart distribution as follows. The entries of a matrix $A_{5\times 5}$ were independently sampled from $N(0, 1)$ and we set $\Sigma = \epsilon AA^T$. The parameter ϵ controls the size of the covariance matrices. As we decrease ϵ, the Gaussians that compose the MoG are further apart. Hence approximating the KL-divergence using matching between the Gaussians becomes more relevant. The approximations we compare are: the Mahalanobis based matching method, denoted as Mahalanobis-match (Expression (5)), our matching based approximation, denoted as KL-match, (Expressions (2) and (4)), the unscented transform approximation described in section 4, and a Monte-Carlo simulation based on 100 samples. We considered the retrieval results based on a Monte-Carlo simulation using 10,000 samples as the ground truth. For each of the four approximation methods we count the percentage of retrieval results that are consistent with the ground truth. For each ϵ that appears in Table 1, the simulated retrieval task was repeated 10,000 times. All the approximations were done to the expression $\int f \log g_t$.

Table 1 summarizes the comparison results. The best results were obtained via the unscented approximation, followed by the results obtained via the 100-sample Monte-Carlo simulation. As ϵ is increased the Gaussians become closer to each other and the overlapping between them increases. Approximating the KL-divergence by matching a single Gaussian of g to each Gaussian component of f becomes less accurate in that case, as can be seen from the results of the Mahalanobis-match and the KL-match. In all the tests that were conducted, the KL-match variant of approximation via Gaussian matching outperforms the Mahalanobis-match. The bottom row of Table 1 indicates the relative processing time needed to compute $\int f \log g$ for each approximation method. The most efficient results are

5

ϵ	MC-100	Mahalanobis match	KL match	unscented
0.025	0.042	1.17	0.18	0.025
0.100	0.062	1.48	0.42	0.043
0.400	0.089	3.27	0.67	0.067

Table 2: Proximity to the true KL-divergence

those obtained by the Mahalanobis-match approximation [1]. A trade-off exists between efficiency and retrieval performance. This trade-off strongly depends on the nature of the Gaussian mixtures.

A more straight-forward criterion for the quality of an approximation is its proximity to the true value of the KL-divergence. The proximity is measured as the average of the following metric:

$$\frac{|KL_{approximate}(f||g) - KL_{true}(f||g)|}{KL_{true}(f||g)}.$$

Table 2 presents the accuracy result for several ϵ values. As in the former experiment it can be seen that as the value of ϵ increases the accuracy of each of the KL-divergence approximations decreases. The most accurate approximation is the unscented based approximation. The worst results are those obtained by the Mahalanobis-match approximation.

In the final set of experiments we evaluate the retrieval results created by the various approximations using precision versus recall (PR) curves. Recall measures the ability of retrieving all relevant or perceptually similar items in the database. It is defined as the ratio between the number of perceptually similar items retrieved and the total relevant items in the database. Precision measures the retrieval accuracy and is defined as the ratio between the number of relevant or perceptually similar items retrieved and the total number of items retrieved. The database used throughout the experiments consists of 1460 images selectively hand-picked from the COREL database to create 16 categories. The images within each category have similar colors and color spatial layout, and can be labelled with a high-level semantic description. In the following experiment we averaged retrieval results for 320 images, 20 images drawn randomly from each of the 16 labelled categories we have in the database. PR curves were calculated for 10,20,30,40,50, and 60 retrieved images. Figure 4 shows the PR curves obtained by each of the approximations. It can be seen that the retrieval results obtain via the unscented based approximation are similar to the retrieval results obtained via 500-sample Monte-Carlo simulations.

---

[1]It should be noted that when comparing computational complexity between the proposed approximations, we assume that any computational step needed for the approximation that can be done on a single image (e.g. inverting the Gaussian matrix, choosing the "sigma" point) is considered as a pre-processing step which is done as part of the MoG model learning.

The proposed matching based approximation (KL-match) significantly outperforms the Mahalanobis-matching based method (Mahalanobis-match).
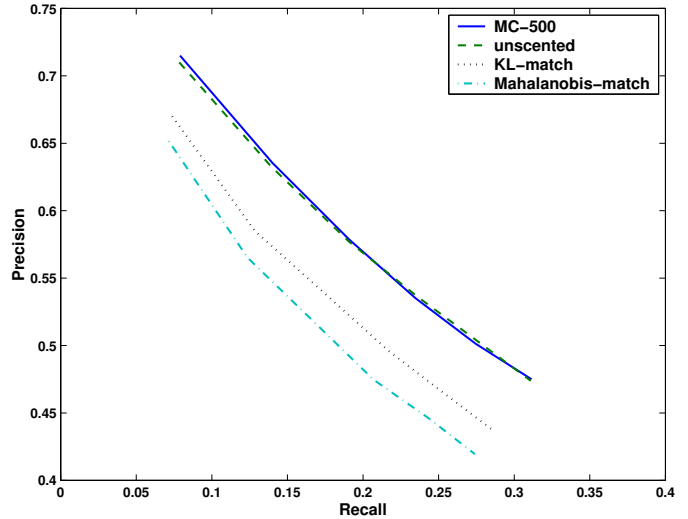

Figure 4: Precision vs recall for evaluating retrieval results for different KL-divergence approximations.

Figure 5 displays the first 20 images retrieved by each of the KL approximations, for an example query taken from the $Tigers$ category. The best retrieval results are obtained as before via the Monte-Carlo simulation and the unscented based approximation. In these two cases more images within the first 20 retrieved images are from the same category as the query.

6   Conclusions

In this paper we described two new methods for approximating the KL-divergence between mixture densities. The first (match-based) method can be applied to any mixture density while the second (unscented) is tailored for mixtures of Gaussian densities. The efficiency and the performance of these methods were demonstrated on image retrieval tasks on a large database. In all the experiments conducted, the unscented approximation achieves the best results, results that are very close to large sample Monte-Carlo based ground truth. The Kl-match based approximation is faster but less accurate than the unscented based method. A future research direction can be to combine the two approximation methods into a single scheme: In order to approximate $KL(f||g)$, we can check separately for each Gaussian component f_i whether the matching based approximation is accurate enough. This is the case if the component of g matched to f_i is significantly closer to f_i as compared to the other components of g. Otherwise we can utilize the
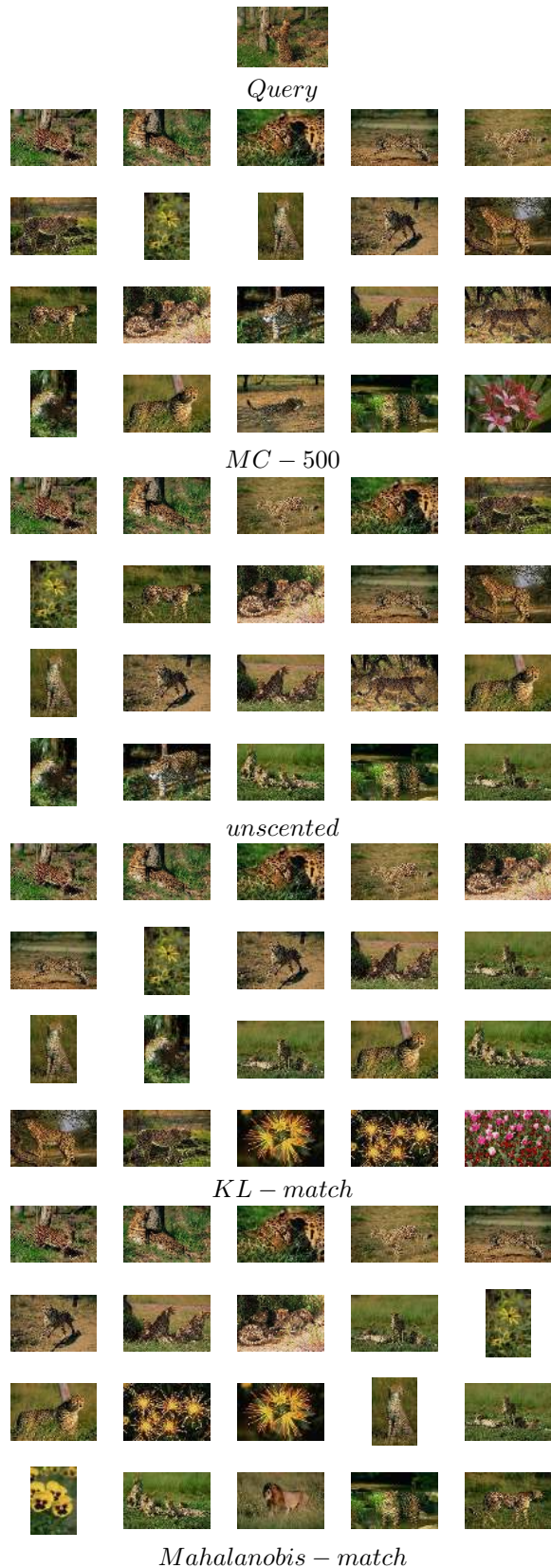
unscented based approximation to compute $KL(f_i||g)$. Efficient approximation of the KL-divergence between MoGs is a major step towards continuous and probabilistic image retrieval systems.

References

[1] C. Carson, S. Belongie, H. Greenspan and J. Malik, "Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying", *IEEE Transactions on PAMI*, 24(8): 1026-1038, 2002.

[2] T. Cover and J.Thomas, "Elements of information theory", *Wiley Series in Telecommunications*, Jhon Wiley and Sons, New-York, USA, 1991.

[3] H. Greenspan, J. Goldberger and L. Riddel, "A continuous probabilistic framework for image matching", *Journal of Computer Vision and Image Understanding* 84, pp. 384-406, 2001.

[4] S. Julier and J. K. Uhlmann, "A general method for approximating nonlinear transfromations of probability distributions", *Technical report*, RRG, Dept. of Engineering Science, University of Oxford, 1996.

[5] S. Julier and J. K. Uhlmann, "A new extension of the Kalman filter to non-linear systems", *Proc of AeroSense: The 11th International Symposium on Aerospace/Defence Sensing, Simulation and Control*, Florida, 1997.

[6] S. Julier, "The scaled unscented transformation", to appear in Automatica, 2000.

[7] H. W. Kuhn, "The Hungarian method for the assignment problem", *Naval Research Logistics Quarterly*, Vol., pp 83-97,1955.

[8] S. Kullback, "Information theory and statistics", Dover Publications, New York, 1968.

[9] J. Puzicha, Y. Rubner, C. Tomasi and J. Buhmann, "Empirical evaluation of dissimilairty measure for color and texture", *Proc. of the Int. Conference on Computer Vision*, 1999.

[10] N. Vasconcelos, "On the complexity of probabilistic Image Retrieval", *Proc. of the Int. Conference on Computer Vision*, 2001.

[11] G. Wyszecki and W. Stiles, "Color Science: Concepts and Methods, Quantitative Data and Formulae", Jhon Wiley and Sons, New-York, USA, 1982.

Figure 5: Retrieval example for a query image taken from the $Tigers$ category.