

Published in IET Image Processing
 Received on 21st January 2008
 Revised on 16th September 2008
 doi: 10.1049/iet-ipr:20080012



Probabilistic relevance feedback approach for content-based image retrieval based on gaussian mixture models

A. Marakakis¹ N. Galatsanos² A. Likas² A. Stafylopatis¹

¹School of Electrical and Computer Engineering, National Technical University of Athens, 15773 Athens, Greece

²Department of Computer Science, University of Ioannina, 45110 Ioannina, Greece

E-mail: amara@central.ntua.gr

Abstract: A new relevance feedback (RF) approach for content-based image retrieval is presented. This approach uses Gaussian mixture (GM) models of the image features and a query that is updated in a probabilistic manner. This update reflects the preferences of the user and is based on the models of both the positive and negative feedback images. The retrieval is based on a recently proposed distance measure between probability density functions, which can be computed in closed form for GM models. The proposed approach takes advantage of the form of this distance measure and updates it very efficiently based on the models of the user-specified relevant and irrelevant images. It is also shown that this RF framework is fairly general and can be applied in case other image models or distance measures are used instead of those proposed in this work. Finally, comparative numerical experiments are provided, which demonstrate the merits of the proposed RF methodology and the use of the distance measure, and also the advantages of using GMs for image modelling.

1 Introduction

The target of content-based image retrieval (CBIR) [1] is to retrieve relevant images from an image database based on the similarity of their visual content with one or more query images. These query images are submitted by the user as examples of his/her preferences. Then, the CBIR system ranks the database images and displays the retrieved results ordered with respect to their similarity with the query images. Most CBIR systems [2–9] model each image using a combination of low-level features and, then, define a distance measure that is used to quantify the similarity between the image models. A lot of effort has been devoted to developing features and strategies that capture the human perception of image similarity in order to enable efficient indexing and retrieval [6, 10–12]. Nevertheless, low-level image features cannot always capture the human perception of image similarity. In other words, it is difficult using only low-level image features to describe the semantic content of an image.

This is known in the CBIR community as the ‘semantic gap’ problem [13].

Relevance feedback (RF) has been proposed as a methodology to alleviate this problem [2–4, 7–9]. RF attempts to insert the subjective human perception of image similarity into a CBIR system. Thus, RF is an interactive process that refines the distance between the query and the database images through interaction with the user and taking into account his/her preferences. To accomplish this, during a round of RF, the user is required to rate the relevance of the retrieved images according to his/her preferences. Then, the retrieval system updates the matching criterion based on the user’s feedback [2–4, 7–9, 12, 14].

Gaussian mixtures (GM) are a well-established methodology to model probability density functions (pdf). The advantages of this methodology, such as adaptability to the data, modelling flexibility and robustness, have made

GM models attractive for a wide range of applications e.g. [15, 16]. Because of the above merits, GM models have already been employed for the CBIR problem [5, 15, 17]. The main challenge when using a GM model in CBIR is to define a distance measure between GMs, which separates the different models well and, in addition, can be computed efficiently. The traditionally used distance measure between pdfs is the Kullback–Leibler (KL) divergence that cannot be computed in closed form for GM models. Thus, we have to resort to random sampling Monte Carlo methods to compute this measure for GMs. This makes its use impractical for CBIR, where retrieval time is an important issue. In [17], the earth movers distance (EMD) was proposed as an alternative distance metric for GM models. Although the EMD metric has good separation properties and is much faster to compute than the KL divergence (in the GM case), it still requires the solution of a linear program. Consequently, it is not computable in closed form and it is not fast enough for an RF-based CBIR system, which requires online interaction with the user. Moreover, in [18], the asymptotic likelihood approximation (ALA) was proposed as a measure that, under certain assumptions, approximates the KL divergence and can be computed in closed form for GMs.

With regard to RF approaches proposed in the literature, much work has been done during the last years. This work can be classified in two main categories. The first category concerns classifier-based methods, that is, it includes the methods which are based on some learning model (usually SVMs) in order to train a classifier to distinguish between the positive and negative feedback examples. Several SVM variations have been proposed for RF and a number of different SVM kernels have been adopted, from typical ones, like radial basis functions based on some L_p norm, to more sophisticated ones that are based on the EMD [7, 19, 20]. The main drawback of the classifier-based approach is that, for every feedback round, a new classifier must be trained taking into account both the previously presented examples and the new ones presented in the last feedback round.

The second category of RF methods (data distribution-based methods) includes those that attempt to model the statistical distribution of feedback examples in the feature space. These methods can be further divided in two subcategories.

The first subcategory includes methods that make the assumption that the feedback examples form one cluster in the feature space. The cornerstone of such methods is MindReader [2]. Other methods that work under this assumption are presented in [12, 21–23]. Usually, in these methods, the new query is formed as a linear combination of the positive feedback examples and the distance between the query and the database images is expressed in the form of a Mahalanobis distance. Unfortunately, in most cases, the single cluster assumption is very restrictive even for the

set of positive examples. Moreover, the negative feedback examples cannot be taken into account, because they naturally spread across different semantic categories and, thus, it cannot be claimed that they form one cluster. Nevertheless, in [22], a solution to this problem is proposed based on a ‘two-step’ retrieval approach. In the first step, only the positive examples are used to determine a reduced set of database images very similar to them. In the second step, these images are re-ranked based on both the positive and negative examples, for the final ranking to be produced.

The second subcategory of the data distribution-based RF techniques includes methods, which assume that the feedback examples (either positive or negative) form more than one cluster [15, 24–27]. The methods, which are based on GM models, belong to this category. GM models are used in the aforementioned methods in two different ways. First, given an image description in a multi-dimensional space, GM models are used in order to approximate the distribution of the user examples, which are encountered as points in the feature space. The ranking of the database images can be performed based on the likelihood of the formed GM model for each of the images [15, 24, 25]. The drawbacks associated with these methods are, on the one hand, that we must re-train a new GM model for every RF round based on both the new and the old examples and, on the other hand, that usually a global image description based either on a single or on a few vectors per image is adopted. Alternatively, one GM can be used to model each image based on a set of feature vectors extracted from the image pixels or regions. This is the approach adopted in this work. In [26, 27], the posterior probability of the model of each image, given the feedback, is used in order to rank the database images. However, the main problem regarding these methods is the definition of an easy to compute distance measure between GM models, which can be expressed in a closed form, in order to accelerate the retrieval and feedback process.

In this paper, as a solution to the above problem, we propose the use of an alternative distance measure between pdfs, called $C2$ divergence, which was recently proposed in [28]. This measure can be computed in a closed form for GM models. Moreover, we analyse thoroughly the properties of this measure and provide a bound on the difference between the $C2$ divergence and the symmetric KL divergence. Additionally, we propose an efficient probabilistic RF technique, which relies on a suitable, straightforward and intuitive update of the GM model of the query after each RF round, using the relevance of the retrieved images. Furthermore, we propose an effective strategy that requires very few computations to update the distance measure after each RF round. In particular, the distance of the new query from the database images can be easily computed using some pre-computed quantities related to the $C2$ divergence, and is incrementally updated during the RF rounds. This query update methodology and

efficient distance computation strategy can be applied not only in the case of GM models, but also with other pdf models of the images. For example, and for reasons of comparison with GMs, we have also applied this framework when the images are modelled using histograms. Moreover, the proposed query update methodology, because of its elegant form, can be applied very efficiently using other distance measures as well. To obtain a comparative evaluation of the $C2$ divergence in this context, we also develop and test an incremental scheme to update the ALA measure.

A preliminary version of this work has been presented in [29]. In the present paper, we further elaborate on the approach and provide an in-depth experimental study for performance assessment of the proposed method. The experimental evaluation is based on comparative results on large-scale data sets using an enriched set of image features.

The rest of this paper is organised as follows. In Section 2, we describe GMs in the context of image modelling for CBIR, and we give more details about the most popular distance measures between GM models. We also elaborate further on the difficulties that arise from the use of these distance measures for RF based on GMs. In Section 3, we present the $C2$ divergence as a promising alternative distance measure for GMs, we analyse its properties and we present the proposed $C2$ -based RF scheme. In Section 4, we first present the other methods with which our method is compared and, then, we provide the details and the results of the experiments. Finally, in Section 5, we present conclusions and directions for future research.

2 CBIR using GM models

2.1 Definition of GM models

GM models have been used extensively in many data-modelling applications. Using them for the CBIR problems allows us to bring to bear several powerful features of the GM-modelling methodology, such as modelling flexibility and easy training, that make it attractive for a wide range of applications [15, 16, 30]. GM models have been used previously for CBIR [5, 17] as probability density models of the features that are used to describe the images. In this framework, each image is described as a bag of feature vectors that are computed locally (e.g. a feature vector for each pixel or region of the image). This bag of feature vectors is, subsequently, used to train (in a maximum likelihood manner) a GM that models the probability density of the image features in the feature space. A GM model for the image feature vectors $\mathbf{x} \in R^d$ is defined as

$$p(\mathbf{x}) = \sum_{j=1}^K \pi_j \phi(\mathbf{x}|\theta_j) \quad (1.1)$$

$$\theta_j = (\mu_j, \Sigma_j) \quad (1.2)$$

$$\begin{aligned} \phi(\mathbf{x}|\theta_j) &= N(\mathbf{x}|\theta_j) \\ &= \frac{1}{\sqrt{(2\pi)^d |\Sigma_j|}} e^{-\frac{1}{2}(\mathbf{x} - \mu_j)^T \Sigma_j^{-1} (\mathbf{x} - \mu_j)} \end{aligned} \quad (1.3)$$

where K is the number of Gaussian components in the model, $0 \leq \pi_j \leq 1$ the mixing probabilities with $\sum_{j=1}^K \pi_j = 1$, and $\phi(\mathbf{x}|\theta_j)$ a Gaussian pdf with a mean μ_j and covariance Σ_j .

2.2 Distance measures between GM models

To describe the similarity between images in this context, a distance measure must be defined. The KL divergence [11] is the most commonly used distance measure between pdfs. For two pdfs, $p_1(\mathbf{x})$ and $p_2(\mathbf{x})$, the KL divergence is defined as

$$KL(p_1||p_2) = \int p_1(\mathbf{x}) \log \frac{p_1(\mathbf{x})}{p_2(\mathbf{x})} d\mathbf{x} \quad (2)$$

However, this distance measure cannot be computed in a closed form for GMs. Thus, one has to resort to time consuming random sampling Monte Carlo methods, which make the use of this distance measure impractical for CBIR. As far as RF is concerned, the situation is even worse, because RF is based on online interaction with the user, which demands rapid distance updates at every RF round. To overcome these difficulties, some alternatives have been proposed.

In [17], the EMD metric between GMs was proposed. This metric is based on considering the probability mass of the one GM as piles of earth and of the other GM as holes in the ground and, then, finding the least work necessary to fill the holes with the earth in the piles. More specifically, for two GMs, $p_1(\mathbf{x}) = \sum_{i=1}^{K_1} \pi_{1i} \phi(\mathbf{x}|\theta_{1i})$ and $p_2(\mathbf{x}) = \sum_{j=1}^{K_2} \pi_{2j} \phi(\mathbf{x}|\theta_{2j})$

$$EMD(p_1, p_2) = \frac{\sum_{i=1}^{K_1} \sum_{j=1}^{K_2} f_{ij} d_{\text{ground}}(\phi(\mathbf{x}|\theta_{1i}), \phi(\mathbf{x}|\theta_{2j}))}{\sum_{i=1}^{K_1} \sum_{j=1}^{K_2} f_{ij}} \quad (3.1)$$

where $f_{ij} \geq 0$ are selected so as to minimise the numerator in the EMD definition subject to the following constraints

$$\begin{aligned} \sum_{j=1}^{K_2} f_{ij} &\leq \pi_{1i}, \quad \sum_{i=1}^{K_1} f_{ij} \leq \pi_{2j}, \\ \sum_{i=1}^{K_1} \sum_{j=1}^{K_2} f_{ij} &= \min \left(\sum_{i=1}^{K_1} \pi_{1i}, \sum_{j=1}^{K_2} \pi_{2j} \right) = 1 \end{aligned} \quad (3.2)$$

$d_{\text{ground}(s)}$ denotes a ground distance metric quantifying the distance between two Gaussians. The EMD is an effective

metric for CBIR, however, it cannot be computed in a closed form and requires the solution of a linear program each time that must be computed. Thus, it is slow and cumbersome to use for RF, where the query changes after each RF epoch.

In [18], the ALA was proposed as a similarity measure for GMs. More specifically, for two GMs, $p_1(\mathbf{x}) = \sum_{i=1}^{K_1} \pi_{1i} \phi(\mathbf{x}|\theta_{1i})$ and $p_2(\mathbf{x}) = \sum_{j=1}^{K_2} \pi_{2j} \phi(\mathbf{x}|\theta_{2j})$

$$ALA(p_1||p_2) = \sum_{i=1}^{K_1} \pi_{1i} \left\{ \log \pi_{2\beta_{12}(i)} + \left[\log \phi(\mu_{1i}|\theta_{2\beta_{12}(i)}) - \frac{1}{2} \text{trace}(\Sigma_{2\beta_{12}(i)}^{-1} \Sigma_{1i}) \right] \right\} \quad (4.1)$$

where

$$\|\mathbf{x} - \mu\|_{\Sigma}^2 = (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) \quad (4.2)$$

$$\beta_{12}(i) = k \Leftrightarrow \|\mu_{1i} - \mu_{2k}\|_{\Sigma_{2k}}^2 - \log \pi_{2k} < \|\mu_{1i} - \mu_{2l}\|_{\Sigma_{2l}}^2 - \log \pi_{2l}, \quad \forall l \neq k \quad (4.3)$$

This measure has the favourable property that, under certain assumptions, it approximates the KL divergence. The difficulty with the ALA measure is that it requires the computation of a correspondence function, $\beta_{12}(i)$, between the components of the query and those of the database image model. As the query model changes after every feedback round, there is a need for re-computation of this correspondence function, which can burden the distance computation taking into account that the new query model may be much more complex than the initial one, in order to incorporate the models corresponding to the feedback examples. Nevertheless, in Section 4.1, we present an incremental scheme based on the proposed methodology of query model update, which enables the fast update of this measure after each round of RF.

3 The C2 divergence and its application to RF

3.1 Definition of the C2 divergence

To alleviate the aforementioned difficulties, which are encountered by most of the popular distance measures in the case of GMs, a new distance measure was proposed in [28]. This measure between two pdfs, $p_1(\mathbf{x})$ and $p_2(\mathbf{x})$, is defined as

$$C2(p_1, p_2) = -\log \frac{2S_{p_1 p_2}}{S_{p_1 p_1} + S_{p_2 p_2}} \quad (5.1)$$

with

$$S_{p_m p_l} = S_{ml} = \int p_m(\mathbf{x}) p_l(\mathbf{x}) d\mathbf{x} \quad (5.2)$$

and it can be computed in a closed form when $p_1(\mathbf{x})$ and $p_2(\mathbf{x})$ are GMs. In this case, we have (see proof in Appendix 8.1)

$$S_{ml} = \frac{1}{\sqrt{(2\pi)^d}} \sum_{i=1}^{K_m} \sum_{j=1}^{K_l} \frac{\pi_{mi} \pi_{lj}}{\sqrt{|C_{ml}(i, j)| e^{k_{ml}(i, j)}}} \quad (6.1)$$

where

$$C_{ml}(i, j) = \Sigma_{mi} + \Sigma_{lj} \quad (6.2)$$

$$k_{ml}(i, j) = (\mu_{mi} - \mu_{lj})^T C_{ml}^{-1}(i, j) (\mu_{mi} - \mu_{lj}) \quad (6.3)$$

π_{mi} , μ_{mi} and Σ_{mi} are the mixing weight, the mean and the covariance matrix, respectively, of the i th Gaussian kernel of p_m ; d is the dimension of the feature vector \mathbf{x} , and K_m is the number of Gaussian components in p_m .

3.2 Properties of the C2 divergence

For the C2 divergence, the following properties hold:

1. $C2(p_1, p_2) \geq 0$
2. $C2(p_1, p_2) = 0 \Leftrightarrow p_1(\mathbf{x}) = p_2(\mathbf{x})$
3. $C2(p_1, p_2) = C2(p_2, p_1)$
4. The triangular inequality $C2(p_1, p_3) \leq C2(p_1, p_2) + C2(p_2, p_3)$ does not hold; therefore the C2 is not a metric as is also the case with the KL divergence.

With regard to the relation between the C2 divergence and the symmetric KL divergence defined by

$$SKL(p_1, p_2) = \frac{1}{2} [\text{KL}(p_1||p_2) + \text{KL}(p_2||p_1)] \quad (7)$$

the following inequality can be proved for arbitrary pdfs, p_1 and p_2 (the proof is given in Appendix 8.2).

Remark 1: The difference between the SKL and the C2 is bounded by

$$SKL(p_1, p_2) - C2(p_1, p_2) \leq \frac{1}{2} [\Delta_1(p_1) + \Delta_2(p_2)] + \log 2 \quad (8.1)$$

where

$$\Delta_i(p_i) = E[\log p_i]_{p_i} - \log E[p_i]_{p_i} \geq 0 \quad (8.2)$$

because of the Jensen's inequality. It must be noted that the Jensen's inequality becomes an equality when the log function is replaced by a linear function. Since the log function can be locally approximated by a linear function, it can be inferred that the bound becomes tighter (i.e. $\Delta_i(p_i)$ becomes small) when p_i exhibits small variance.

Furthermore, with regard to the relation between the $C2$ divergence and the L_2 norm, it is straightforward to show that

$$C2(f, g) = -\log \frac{2 \langle f, g \rangle}{L_2^2(f) + L_2^2(g)} \quad (9.1)$$

$$= -\log \left(1 - \frac{L_2^2(f - g)}{L_2^2(f) + L_2^2(g)} \right)$$

where

$$L_2(f) = \sqrt{\int f^2(x) dx} \quad (9.2)$$

and

$$\langle f, g \rangle = \int f(x)g(x) dx \quad (9.3)$$

are the definitions of the L_2 norm and the inner product, respectively, for real continuous functions. Given that $\log x \simeq x - 1$ for $x \simeq 1$

$$C2(f, g) \simeq \frac{L_2^2(f - g)}{L_2^2(f) + L_2^2(g)} \quad \text{for} \quad (9.4)$$

$$L_2^2(f - g) \ll L_2^2(f) + L_2^2(g)$$

3.3 RF based on the $C2$ divergence

For a distance measure to be useful in RF, it is crucial that it be easily updated based on the feedback images provided by the user. Thus, assume we have a query, q , modelled by $q(\mathbf{x})$ (e.g. a GM model) and that the i th database image is modelled by $i(\mathbf{x})$, for $i = 1, \dots, N$. The search based on this query requires the calculation of an $N \times 1$ table with the values $C2(q, i)$. Also, assume that, from the retrieved images, the user decides that the images with models $r_m(\mathbf{x})$, $m = 1, 2, \dots, M$, are the most relevant and desires to update his query based on them. One simple and intuitive way to go about it, is to generate a new query model given by

$$q'(\mathbf{x}) = (1 - \Lambda)q(\mathbf{x}) + \sum_{m=1}^M \lambda_m r_m(\mathbf{x}) \quad (10)$$

where $0 \leq \lambda_m \leq 1$, $\sum_{m=1}^M \lambda_m = \Lambda$, $0 \leq \Lambda \leq 1$, λ_m is the relevance assigned by the user to the image r_m and $1 - \Lambda$ is the weight of contribution of the previous query to the formation of the new query. The attractive feature of this definition of the new query model is that it is consistent with the probabilistic framework adopted for image modelling, with each λ_m having a clear probabilistic meaning. Moreover, λ_m has a physical meaning as well; it is proportional to the relevance degree assigned by the user to the image r_m . From the above considerations, it becomes

apparent that (10) defines a 'composite model' that incorporates the preferences of the user.

Furthermore, it is desirable to efficiently compute the distances between the database image models, $i(\mathbf{x})$, $i = 1, 2, \dots, N$, and the new query model, $q'(\mathbf{x})$. Taking into account (5.x) and (10), the updated distance measure for the new query, q' , is given by (the proof is given in Appendix 8.3)

$$C2(q', i) = -\log \frac{2S_{q'i}}{S_{q'q'} + S_{ii}} \quad (11.1)$$

$$S_{q'i} = (1 - \Lambda)S_{qi} + \sum_{m=1}^M \lambda_m S_{r_m i} \quad (11.2)$$

$$S_{q'q'} = (1 - \Lambda)^2 S_{qq} + 2(1 - \Lambda) \sum_{m=1}^M \lambda_m S_{qr_m}$$

$$+ \sum_{m=1}^M \sum_{m'=1}^M \lambda_m \lambda_{m'} S_{r_m r_{m'}} \quad (11.3)$$

Since the S_{ml} values can be computed (and stored) a priori for all the images of the database, and, additionally, S_{qq} and all S_{qi} have already been computed for the previous query, it is obvious that the computation of the distance between $q'(\mathbf{x})$ and the database image models is very fast. This constitutes a notable advantage of our method. Indeed, computing the distance for the new query involves only re-scaling operations based on the relevance probabilities λ_m . Actually, the new query model in (10) does not need to be constructed. To update the distance between the query and the database images, only the computations of (11.x) are needed, and those computations only implicitly involve the new query model. It should be mentioned that, even if we do not pre-compute the S_{ml} values for all the database images (because of the lack of space or time when huge image databases are considered, for example), these quantities can be computed online very fast for GMs, since they are defined in a closed form in (6.x). Thus, in this case too, by incrementally updating S_{qq} and S_{qi} using (11.2) and (11.3), the $C2$ divergence between the new query and the database images can be computed fast and without the need of constructing the new query model.

Another favourable property of the proposed RF scheme is that it can be easily used with many types of pdf models and not just with the GMs, proposed in this work. In other words, the models of the database images and the query need not be GMs. Any other image model for which the computation of the S -measure, defined by (5.2), is not prohibitively time consuming can be adopted without any modification to the RF scheme.

The images retrieved by the system at each retrieval epoch, which are not selected by the user as relevant, can be considered as irrelevant. Except for this implicit way of

selection of irrelevant images, another one, more rigorous and accurate, can be considered in which the user will be asked to define them explicitly. The user-selected irrelevant images can also be incorporated, as negative feedback, into the RF process. In this context, we define a query model for irrelevant images, which we call 'negative' query model, and update it by a method similar to (10). Specifically, the new negative query model is given by

$$n'(x) = (1 - \Lambda_n)n(x) + \sum_{m=1}^{M_n} \lambda_m^n r_m^n(x) \quad (12)$$

where n and n' correspond to the previous and new negative queries, respectively; Λ_n and λ_m^n are analogous to the previously mentioned Λ and λ_m ; and r_m^n , $m = 1, 2, \dots, M_n$, are the negative examples. The negative query is initially 'empty', contrary to the positive one that includes the initial query selected by the user.

The best images to retrieve can be found by combining both the positive and negative RF. This can be done by minimising the following distance measure as

$$c(i) = a_{\text{pos}}d(q, i) - (1 - a_{\text{pos}})d(n, i) \quad (13.1)$$

$$d(q, i) = C2(q, i) \quad (13.2)$$

$$d(n, i) = C2(n, i) \quad (13.3)$$

with $0 \leq a_{\text{pos}} \leq 1$ being the relative weight given to the positive feedback. After computing the measure $c(i)$ for every database image, we can retrieve the images with the lowest values for this measure. Such images will be similar to the user 'ideal' query, which is determined by the initial query and the positive examples, and dissimilar to the user-provided negative examples.

4 Experiments

4.1 Methods of comparison

In this work, we propose the use of GMs in order to model the database images, along with the RF scheme based on the C2 divergence, described in Section 3.3. To demonstrate the merits of this method, we conducted a number of experiments.

As we mentioned above, our RF scheme can also be directly applied when the database images have been modelled using some other pdf model instead of GMs. To demonstrate the strength of GMs for image modelling, we compared our method with a histogram-based variation. In other words, we implemented a variation of our method, which uses histograms instead of GMs as image models.

In the case of modelling the pdf of image features using histograms, the range of values of each feature is divided into a number of disjoint intervals. The simplest choice for

those intervals is to assume that they have equal width. Thus, for image features f_1, f_2, \dots, f_L ($a_i \leq f_i \leq b_i$) and with the range of each feature f_i divided into z_i intervals, we obtain an L -dimensional space quantised into $Z = z_1 z_2 \dots z_L$ bins. Then, assuming that we describe each image with one feature vector per image pixel, we can approximate the pdf of the image features by simply counting the pixels the features of which fall into each bin and normalising by dividing with the total number of pixels. Thus, we can represent the histogram of an image in vectorial form as $\mathbf{p} = (b_1, b_2, \dots, b_Z)^T$, with $\sum_{i=1}^Z b_i = 1$.

Taking into account (5.x), in order to compute the C2 divergence between histogram models \mathbf{p}_1 and \mathbf{p}_2 , it is sufficient to compute $S_{ml} = \int \mathbf{p}_m(x)\mathbf{p}_l(x)dx$, $m, l = 1, 2$. It is trivial to show that, for histograms

$$S_{ml} = \sum_{i=1}^Z b_{mi}b_{li} \quad (14)$$

where $\mathbf{p}_m = (b_{m1}, b_{m2}, \dots, b_{mZ})^T$ and $\mathbf{p}_l = (b_{l1}, b_{l2}, \dots, b_{lZ})^T$. Using (14) to compute S_{ml} and under the assumption that we use (10) and (12) to update the positive and negative queries (based on the feedback given by the user), RF, in the case of histograms, is performed in exactly the same manner as for GMs. Thus, again, the histograms for the new query models do not need to be constructed.

To demonstrate the benefits of using the C2 divergence for an RF based on GMs, we compared our method with a variation that uses the ALA measure to rank the images. Using (10) and (12) to update the models of the queries, it is easy to prove (the proof is given in Appendix 8.4) that the ALA measure for the new query models can be incrementally updated by

$$\begin{aligned} \text{ALA}(q' || i) &= (1 - \Lambda)\text{ALA}(q || i) \\ &+ \sum_{m=1}^M \lambda_m \text{ALA}(r_m || i) \end{aligned} \quad (15)$$

A similar equation holds in the case of the negative query. To obtain this equation, the only thing to be done is to replace $q, q', \lambda_m, \Lambda, M$ and r_m by $n, n', \lambda_m^n, \Lambda_n, M_n$ and r_m^n , respectively, in (15). Taking into account that the ALA is a similarity measure, the definitions of the distances $d(q, i)$ and $d(n, i)$ of (13.x) were modified, taking the following form

$$d(q, i) = -\text{ALA}(q || i) \quad (16.1)$$

$$d(n, i) = -\text{ALA}(n || i) \quad (16.2)$$

Additionally, for a comparative evaluation of the proposed RF methodology, we have implemented the method described in [22]. This method incorporates both positive and negative examples using a 'two step' retrieval scheme.

In the first step, only the positive examples are used in order to rank the database images. Then, only a relatively small number of the database images, which are very similar to the positive examples, are retained and re-ranked in the second step which takes into account both the positive and negative examples, to produce the final ranking from which the top images, placed near to the positive and far from the negative examples, are presented to the user. Both the retrieval steps are based on the Lagrange optimisation in order to estimate the distance parameters that minimise the within class distance and maximise the between class distance of the positive and negative examples.

4.2 Image databases and features

To test the validity of the proposed approach, we used two image databases. The first image set (DB I) contains 3740 colour images of size 640×480 from the image database in [31]. These images have been classified into 17 semantic categories according to their content (e.g. airplanes, cars, birds, windows etc.). Generally, we adopted the categorisation specified by the database provider, except for a few cases where categories, which are semantically very close to each other, were merged. The second database (DB II) is a subset of the Corel image database. It contains 9923 images of size 256×384 , which are classified into 42 semantic categories. Although the Corel database is professionally annotated and categorised, many images containing the same semantic content are distributed across different Corel categories. Hence, we decided to merge some Corel categories, thus producing our own semantic categorisation, which is considered as the ground truth.

For each image pixel in the aforementioned databases, a set of features has been extracted. These features are of three types and are described below:

1. *Position*: the (x, y) -coordinates of each pixel normalised by the image width (x_{\max}) for the x -dimension and by the image height (y_{\max}) for the y -dimension.
2. *Colour*: we chose the CIE-Lab [17] colour space as being approximately perceptually uniform, a property very useful for retrieval and, thus, three colour features (L^* , a^* , b^*) per pixel were used. To take the final values $[(L^*)^{\text{smoothed}}, (a^*)^{\text{smoothed}}, (b^*)^{\text{smoothed}}]$ for these features, local Gaussian smoothing is performed according to the texture scale which is mentioned below. In this way, we decorrelate the colour and the texture of the image.
3. *Texture*: As a scheme to extract texture features from the images, we selected the one presented analytically in [5]. The gradient of the L^* colour component of the image is estimated. The method is based on the eigenstructure of the second moment matrix derived using the aforementioned gradient. One measure of the local texture scale along with the anisotropy (A), the polarity (P) and the contrast (C), corresponding to this scale, are estimated

for each pixel. Taking into account that the polarity and anisotropy values are meaningless in regions of low contrast, the texture features that were selected are: $AC = A * C$, $PC = P * C$ and C .

Thus, finally, we have an eight-dimensional vector, x , for each image pixel, with

$$x = (f_1, f_2, \dots, f_8)^T = \left(x^{\text{norm}}, y^{\text{norm}}, (L^*)^{\text{smoothed}}, (a^*)^{\text{smoothed}}, (b^*)^{\text{smoothed}}, AC, PC, C \right)^T \quad (17.1)$$

$$x^{\text{norm}} = \frac{x}{x_{\max}} \quad (17.2)$$

$$y^{\text{norm}} = \frac{y}{y_{\max}} \quad (17.3)$$

Prior to feature extraction, pixel sub-sampling was performed using a spatially uniform grid. In DB I, only 15% of the image pixels were used for feature extraction and GM training. On the other hand, in DB II, 50% of the image pixels were retained after sub-sampling, because the images are of lower resolution. Prior to sub-sampling, the images were smoothed, using a Gaussian kernel, in order to avoid aliasing. Sub-sampling was necessary in order to reduce the computational load required by the feature extraction and model parameter estimation steps.

4.3 Implementation parameters

Regarding our method, the parameters of the GM model of each image were estimated using a variation of the well known EM algorithm, called greedy EM [32]. This algorithm avoids the problem of parameter initialisation, which is critical for the normal EM. The greedy EM algorithm starts with a single component and adds components sequentially until a maximum number is reached. In all the experiments, we chose to use full covariance parametrisation for the GM components. We also performed several experiments with GM models of different number of components and we concluded empirically that a good choice for this parameter is to have ten components per GM model. Before applying the EM algorithm, we normalise each feature to have zero mean and unit standard deviation taking into account all the feature vectors of the image. After the application of the EM, we convert the produced model to the initial feature space (before feature normalisation). This is necessary so that the subsequent comparison between the produced GM models are fair.

The parameters λ_m and λ_m^n for the positive and negative examples are given equal values regardless of m , because, for the simulations described analytically in the next section, the ground truth of the aforementioned pre-categorised databases is used in order to select the positive and negative examples. Thus, using the strict ground truth

categorisation, it is meaningless to assume different relevance degrees for each example. The exact values of these parameters for each RF round are determined as follows, based on the number of positive and negative examples already provided by the user until the current RF round, and aiming at assigning equal weights to all the examples.

Assuming that we are at the r th RF round and that the positive and the negative examples provided by the user at this RF round are k_r and k_r^n , respectively, the total number of positive and negative examples given during all the RF rounds until now are t_r and t_r^n , respectively, where

$$t_r = 1 + \sum_{j=1}^r k_j \quad (18.1)$$

$$t_r^n = \sum_{j=1}^r k_j^n \quad (18.2)$$

In the case of positive examples, in order to include the initial query in the computation, we increment by one the total number of positive examples given by the user during the RF rounds. Based on the above considerations, the parameters λ_m and λ_m^n for the current feedback round, r , are

$$(\lambda_m)^r = \frac{1}{t_r}, \quad m = 1 \dots k_r \quad (18.3)$$

$$(\lambda_m^n)^r = \frac{1}{t_r^n}, \quad m = 1 \dots k_r^n \quad (18.4)$$

Regarding the histogram-based variation of our method, it is well known that the main difficulty with histograms is the so-called ‘curse of dimensionality’. In other words, as we increase the number of intervals for each feature, the number of histogram bins increases exponentially, resulting in a shortage of data for robust density estimation, given the limited number of available feature vectors per image. For the eight-dimensional feature space used in these experiments, even if we use a small number of intervals for each feature, the resulting total number of bins of an eight-dimensional histogram is prohibitively large.

For this purpose, instead of one eight-dimensional histogram, we decided to use two five-dimensional ones, one with the position and colour features and another with the position and texture features. The decision of using one position-colour histogram and one position-texture histogram was motivated by the fact that, on the one hand, it is meaningless to consider the position features separately from the other types of features and, on the other hand, both the colour and texture feature distributions depend on the position in the image. Thus, the only reasonable choice is to combine the position features with both the colour and texture features. Finally, it should be mentioned that the straightforward inclusion of the position features in the histograms is consistent with the previously described GM

modelling, which makes no discrimination between position, colour and texture features.

In this context, in order to compute the overall distance between the query and the database images, a convex combination of the computed distances for the two histogram models was used. More specifically, the definition of the distances $d(q, i)$ and $d(n, i)$ of (13.x) was modified taking the following form

$$d(q, i) = l_{\text{pos-colour}} C2^{\text{pos-colour}}(q, i) + l_{\text{pos-text}} C2^{\text{pos-text}}(q, i) \quad (19.1)$$

$$d(n, i) = l_{\text{pos-colour}} C2^{\text{pos-colour}}(n, i) + l_{\text{pos-text}} C2^{\text{pos-text}}(n, i) \quad (19.2)$$

$$l_{\text{pos-colour}} + l_{\text{pos-text}} = 1 \quad (19.3)$$

with $C2^{\text{pos-colour}}(\cdot, \cdot)$ and $C2^{\text{pos-text}}(\cdot, \cdot)$ denoting the $C2$ distances for the position-colour and position-texture histograms, respectively, and $l_{\text{pos-colour}}$ and $l_{\text{pos-text}}$ being two positive constants summing up to one.

Regarding the quantisation of the feature space, several experiments were performed. We present here the results obtained for two distinct quantisation settings (denoted Hist1, Hist2), as being the most representative. Specifically, Hist1 uses a $3 \times 3 \times 4 \times 8 \times 8$ histogram for the position-colour feature combination, $x^{\text{norm}} - y^{\text{norm}} - (L*)^{\text{smoothed}} - (a*)^{\text{smoothed}} - (b*)^{\text{smoothed}}$, and a $3 \times 3 \times 4 \times 4 \times 4$ histogram for the position-texture feature combination, $x^{\text{norm}} - y^{\text{norm}} - AC - PC - C$. Respectively, Hist2 uses a $5 \times 5 \times 8 \times 16 \times 16$ position-colour histogram and a $5 \times 5 \times 8 \times 8 \times 8$ position-texture histogram.

Regarding the ALA-based variation of our method, the image models and the RF scheme are exactly the same as those used in our method, except for the ranking measure incremental update, which is performed using (15), and the definition of the distances $d(q, i)$ and $d(n, i)$ of (13.1), which is given by (16.x).

To extract features appropriate for the method proposed in [22], we partitioned each database image in nine sub-images, dividing each of the x and y image axes in three equal width intervals, and, for each sub-image, we estimated a $4 \times 8 \times 8 \left((L*)^{\text{smoothed}} - (a*)^{\text{smoothed}} - (b*)^{\text{smoothed}} \right)$ colour histogram and a $4 \times 4 \times 4(AC - PC - C)$ texture histogram. Thus, we obtain an image description of 18 feature vectors, nine 256-dimensional colour histogram vectors and nine 64-dimensional texture histogram vectors. As it can easily be seen, this feature extraction process is in clear correspondence with the previously described Hist1 quantisation setting (the features corresponding to the Hist2 quantisation cannot be used for the method proposed in [22], because of the rapidly increasing with the feature space dimensionality time complexity of this

method). Moreover, the relevance degrees π_n^k [22] are given equal values, independent of n , for the same reason for which we give equal values to the parameters λ_m and λ_m^n of our method. Finally, for reasons of comparison with our method, we selected to set $\tilde{\pi}^1 = a_{\text{pos}}$ and $\tilde{\pi}^2 = 1 - a_{\text{pos}}$, because $\tilde{\pi}^1$ and $\tilde{\pi}^2$, in [22], are in essence the relative weights given to the positive and negative examples, respectively. For our method, this is exactly the role of the parameters a_{pos} and $1 - a_{\text{pos}}$.

4.4 Simulations and results

To quantify the performance of the proposed RF system, we have resorted to RF simulation. Specifically, two simulation schemes were implemented. In both schemes, a query set is formed including for each database category a percentage of the images belonging to it. Each image in this set is used as an initial query. In this context, the relevant (irrelevant) images are identified, using the ground truth categorisation of the image databases, as those which (do not) belong to the database category of the initial query. In the first simulation scheme, for each query, similar images are retrieved until a specific recall level [33] is reached. Then, for this recall level, the corresponding precision [33] is computed. The retrieved relevant and irrelevant images are also identified, with the percentage of them used for RF in the simulation denoted by rprc and nrprc, respectively. The relevant and irrelevant images used in RF are selected at random from the sets of retrieved relevant and irrelevant images, respectively, when the aforementioned percentages are < 1 . In Figs. 1 and 2, we show the progression in precision averaged over the set of initial queries during different rounds of RF, for a recall level equal to 0.3. In these figures, 'GMM + C2' indicates the method proposed in this paper and 'Hist1 + C2' ('Hist2 + C2') indicates the histogram-based variation when the Hist1 (Hist2) feature

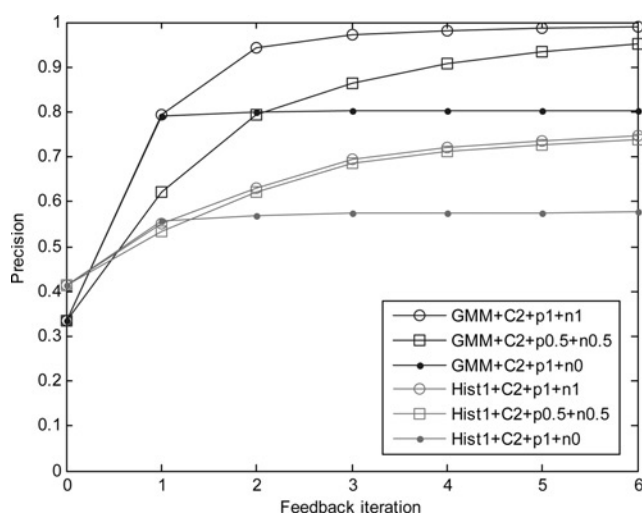


Figure 1 Image models evaluation (first experiment): average precision for recall level = 0.3, during different rounds of RF (DB I)

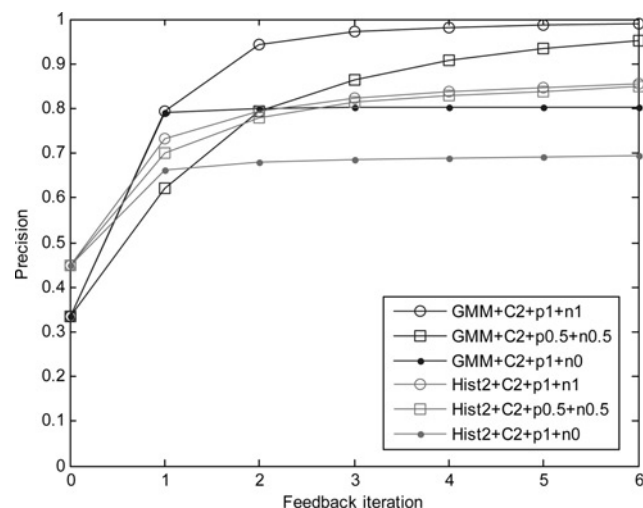


Figure 2 Image models evaluation (second experiment): average precision for recall level = 0.3, during different rounds of RF (DB I)

space quantisation is used. Also, ' pN ' indicates that rprc equals N and ' nN ' indicates that nrprc equals N . The results in Figs. 1 and 2 were obtained using DB I, where the set of initial queries included every image in the database. When negative feedback was given, a_{pos} equals 0.55. With regard to the histogram-based variation of our method, we set $l_{\text{pos-colour}} = l_{\text{pos-text}} = 0.5$.

The second simulation scheme is similar to that proposed in [12]. The accuracy [21] is measured as the ratio of relevant images among the top T retrieved images. At each feedback step, at most K_p relevant images are selected from the top R retrieved images, as positive feedback. If the number of relevant images in the top R retrieved images is greater than K_p , then, we select randomly K_p of them as feedback, else we select all the relevant retrieved images. In case we wish to provide the system with negative feedback, we follow a similar procedure providing as feedback at most K_n of the irrelevant images retrieved in the first R retrieved images. We provide experiments with this methodology in both databases. In DB I, each image was used once as the initial query, whereas in DB II, the query set included 40% of the database images (a percentage sufficient from a statistical point of view). The average accuracy was computed for all the images in the set of initial queries and for each RF round. In both databases, for reasons of comparison, we considered several choices regarding the image models, the RF methodology and the distance measures between the images. The results are given in Figs. 3–6. The choices for the simulation parameters were: $T = 20$, $R = 150$, $K_p = 10$, $K_n = 10$. When negative feedback was provided, $a_{\text{pos}} = 0.65$ (except for the case of the ALA-based variation, which gives the best results when $a_{\text{pos}} = 0.8$). For the histogram-based variation, we set again $l_{\text{pos-colour}} = l_{\text{pos-text}} = 0.5$. For the method proposed in [22], when negative feedback was provided, the R (=150)

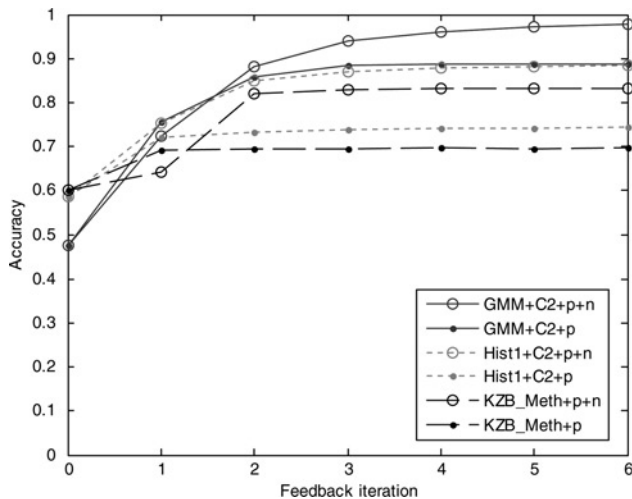


Figure 3 RF schemes evaluation: average accuracy in scope $T = 20$ during different rounds of RF (DB I)

top-ranked images in the first step of retrieval, were retained in the second step. In Figs. 3–6, ‘GMM + C2’ denotes our method, ‘Hist1 + C2’ the histogram-based variation with the Hist1 feature space quantisation, ‘GMM + ALA’ the ALA-based variation and ‘KZB_Meth’ the method proposed in [22]. Also, ‘p’ indicates the use of positive feedback and ‘n’ indicates the use of negative feedback.

From Figs. 1–3, it can be observed that, when GMs are used, the precision (first simulation) or the accuracy (second simulation) is always higher than in the histogram case, except for the initial round (before RF) of retrieval and rarely the initial one or two RF rounds. This is true when either the first (Hist1) or the second (Hist2) quantisation for the histograms is considered. After several RF rounds, the precision (accuracy) level obtained using GMs is significantly higher compared with that of the histogram-based variation.

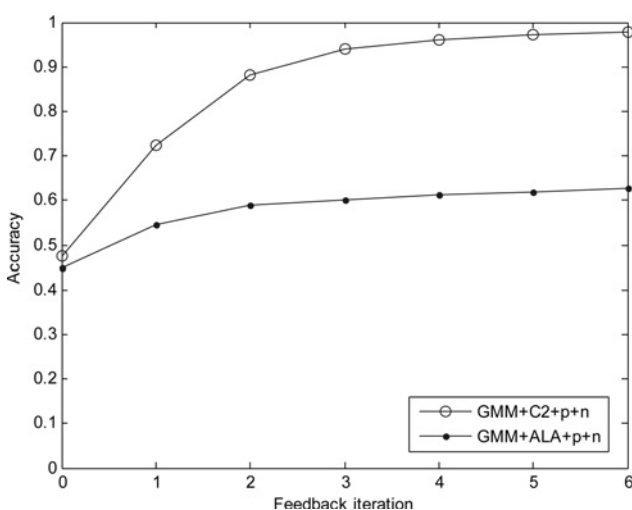


Figure 4 Distance measures evaluation: average accuracy in scope $T = 20$ during different rounds of RF (DB I)

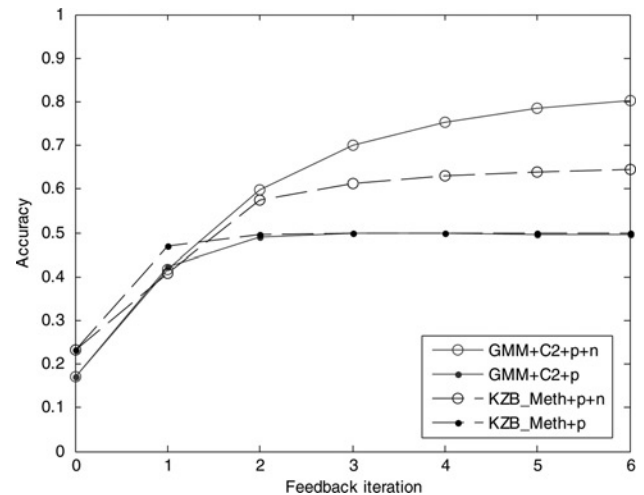


Figure 5 RF schemes evaluation: average accuracy in scope $T = 20$ during different rounds of RF (DB II)

Furthermore, from Figs. 4 and 6, it can be observed that the use of the C2 divergence as a distance measure between GMs leads to a very significant improvement compared with the accuracy obtained using the ALA measure in exactly the same context.

Additionally, as illustrated in Figs. 3 and 5, our method almost always significantly outperforms the method proposed in [22] after the initial (one or two) RF rounds. Moreover, when compared with the method proposed in [22], the histogram-based variation of our method (using the Hist1 quantisation) also results in higher levels of accuracy. Taking into account that the image features used for both the histogram-based variation and the method proposed in [22] are exactly the same in this experiment, this is an indication of the superiority of our RF scheme.

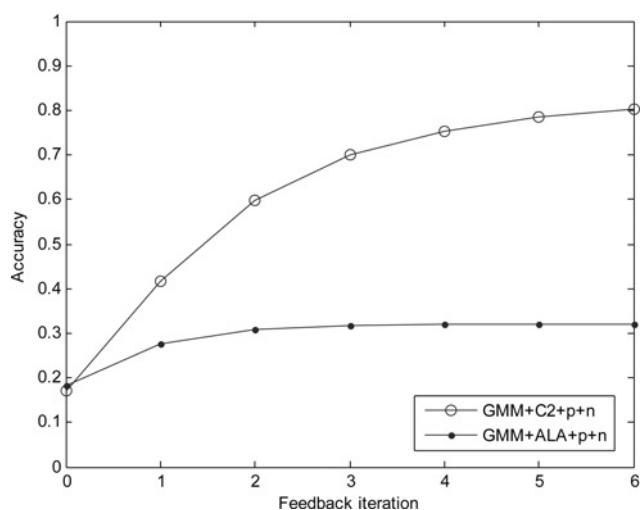


Figure 6 Distance measures evaluation: average accuracy in scope $T = 20$ during different rounds of RF (DB II)

Finally, it should be mentioned that, when we use only positive feedback (and, more generally, when we assign a high weight (a_{pos}) to the positive feedback), we observe a rapid increase in precision (accuracy) at the initial RF rounds and, then, the precision (accuracy) stabilises to the maximum attainable level. On the other hand, when both the positive and negative examples are used (and as the weight of the positive feedback decreases, always remaining >0.5), the increase in precision (accuracy) becomes smoother, a greater number of rounds are required for convergence, however, in many cases, we can achieve a higher level of precision (accuracy) than before. Additionally, in the first simulation scheme, the use of a subset of the available images for feedback (when r_{prc} or nr_{prc} is <1) leads, in many cases, to slower convergence and to lower levels of precision.

To provide an estimate of the retrieval time of the proposed method, it can be noted that, for DB I and for the second simulation scheme (with both positive and negative feedback and with the simulation parameters having values as specified above), it takes about 166 s of computation time (on a 3 GHz PC using Matlab) to execute, for 3740 images presented as initial queries, the initial retrieval plus six epochs of RF. This means that the average retrieval time per query image with six rounds of RF is 0.044 s. Regarding the histogram-based variation, it takes about double the time for the same simulation scenario, because each image is described by two histogram models instead of the one GM model used by the proposed method. Moreover, the ALA-based variation, using the incremental update formula of (15), results in an execution as fast as that of our method. On the other hand, using the same simulation scenario, the method proposed in [22] takes about 19 000 s for 3740 images used as initial queries, which means an average of about 5 s per query. This significant difference in the time complexity demonstrates the noticeable ability of our RF scheme to rapidly update the distances between the database images and the query and to provide an instantaneous response to the user.

5 Conclusions – future work

A probabilistic framework for RF based on GM models was proposed in this paper. The main advantages of the proposed methodology are accuracy as indicated by our simulation study results, speed of implementation and flexibility. Incorporation of both positive and negative feedback examples is performed in a very intuitive manner which, in combination with the simple and easy to update form of the proposed $C2$ divergence, allows for real-time evaluation of the image-ranking criterion. In this way, fast retrieval is achieved after the user feedback has been provided.

Furthermore, we compared GM and histogram-based image models, in order to access the value of GM modelling within the context of RF. To use the same

number of features in both the histogram and GM cases, we were forced to partition the feature space in two subspaces and to use two histograms for each image because of the ‘curse of dimensionality’. In contrast, when using GMs, there is no need for such partitioning. As far as the experimental results are concerned, it is clear that significant performance improvement is obtained when we use GM models, compared with the performance achieved when we use histograms to model the images. Furthermore, the combination of our RF scheme either with GMs or with histograms leads to superior results when compared to the method in [22]. Finally, we compared our method with a variation using the ALA similarity measure instead of the proposed $C2$ divergence, in order to demonstrate the advantages of the latter. The comparative experimental results demonstrate that a superior performance can be achieved with the $C2$ divergence.

In the future, we intend to provide the user with the possibility to determine explicitly the degree of relevance of the feedback examples, by implementing a sophisticated interactive system based on GMs and using our RF scheme. In addition, we aim to generalise our RF scheme to support region-based similarity and retrieval. Furthermore, we aim to attempt to apply techniques to determine automatically the appropriate number of kernels for each mixture. Moreover, we plan to test the performance of our method using more sophisticated image features. Finally, we would like to test the scalability of the proposed method using even larger image databases.

6 Acknowledgments

This work was partly supported by Public Funds under the PENED 2003 Project. The Project is co-funded by the European Social Fund (80%) and National Resources (20%) from the Hellenic Ministry of Development - General Secretariat for Research and Technology.

7 References

- [1] DATTA R., LI J., ZE WANG J.: ‘Content-based image retrieval: approaches and trends of the new age’. *Multimedia Information Retrieval*, 2005, pp. 253–262
- [2] ISHIKAWA Y., SUBRAMANYA R., FALOUTSOS C.: ‘MindReader: querying databases through multiple examples’. *Proc. VLDB Conf*, 1998
- [3] COX I.J., MILLER M.L., MINKA T.P., PAPATHOMAS T.V., YIANILOS P.N.: ‘The Bayesian image retrieval system, PicHunter: theory, implementation, and psychophysical experiments’, *IEEE Trans. Image Process.*, 2000, **9**, (1), pp. 20–37

- [4] RUI Y., HUANG T.S., ORTEGA M., MEHROTRA S.: 'Relevance feedback: a power tool for interactive content-based image retrieval', *IEEE Trans. Circuits Syst. Video Technol.*, 1998, **8**, (5), pp. 644–655
- [5] CARSON C., BELONGIE S., GREENSPAN H., MALIK J.: 'Blobworld: image segmentation using expectation-maximization and its application to image querying', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, (8), pp. 1026–1038
- [6] CHEN Y., WANG J.Z.: 'A region-based fuzzy feature matching approach to content-based image retrieval', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, (9), pp. 1252–1267
- [7] GUO G.D., JAIN A.K., MA W.Y., ZHANG H.J.: 'Learning similarity measure for natural image retrieval with relevance feedback', *IEEE Trans. Neural Netw.*, 2002, **13**, (4), pp. 811–820
- [8] AGGARWAL G., ASHWIN T.V., GHOSAL S.: 'An image retrieval system with automatic query modification', *IEEE Trans. Multimedia*, 2002, **4**, pp. 201–214
- [9] JING F., LI M., ZHANG H.J., ZHANG B.: 'An efficient and effective region-based image retrieval framework', *IEEE Trans. Image Process.*, 2004, **13**, (5), pp. 699–709
- [10] MANJUNATH B.S., OHM J.R., VASUDEVAN V.V., YAMADA A.: 'Color and texture descriptors', *IEEE Trans. Circuits Syst. Video Technol.*, 2001, **11**, (6), pp. 703–715
- [11] DO M.N., VETTERLI M.: 'Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance', *IEEE Trans. Image Process.*, 2002, **11**, (2), pp. 146–158
- [12] HSU C.T., LI C.Y.: 'Relevance feedback using generalized bayesian framework with region-based optimization learning', *IEEE Trans. Image Process.*, 2005, **14**, (10), pp. 1617–1631
- [13] RUI Y., HUANG T.S., CHANG S.F.: 'Image retrieval: current techniques, promising directions, and open issues', *J. Vis. Commun. Image Represent.*, 1999, **10**, pp. 39–62
- [14] EL-NAQA I., YANG Y., GALATSANOS N., WERNICK M.: 'A Similarity learning approach to content based image retrieval: application to digital mammography', *IEEE Trans. Med. Imaging*, 2004, **23**, (10), pp. 1233–1244
- [15] QIAN F., LI M., ZHANG L., ZHANG H.J., ZHANG B.: 'Gaussian mixture model for relevance feedback in image retrieval'. Proc. IEEE ICME, August 2002
- [16] BISHOP C.M.: 'Neural networks for pattern recognition' (Oxford University Press Inc, New York, 1995)
- [17] GREENSPAN H., DVIR G., RUBNER Y.: 'Context-dependent segmentation and matching in image databases', *Comput. Vis. Image Underst.*, 2004, **93**, pp. 86–109
- [18] VASCONCELOS N.: 'On the efficient evaluation of probabilistic similarity functions for image retrieval', *IEEE Trans. Inf. Theory*, 2004, **50**, (7), pp. 1482–1496
- [19] JING F., LI M., ZHANG H.-J., ZHANG B.: 'Relevance feedback in region-based image retrieval', *IEEE Trans. Circuits Syst. Video Technol.*, 2004, **14**, (5), pp. 672–681
- [20] TONG S., CHANG E.: 'Support vector machine active learning for image retrieval'. ACM Multimedia, 2001
- [21] SU Z., ZHANG H., LI S., MA S.: 'Relevance feedback in content-based image retrieval: Bayesian framework, feature subspaces, and progressive learning', *IEEE Trans. Image Process.*, 2003, **12**, (8), pp. 924–937
- [22] KHERFI M.L., ZIOU D., BERNARDI A.: 'Combining positive and negative examples in relevance feedback for content-based image retrieval', *J. Vis. Commun. Image Represent.*, 2003, **14**, pp. 428–457
- [23] RUI Y., HUANG T.S.: 'Optimizing learning in image retrieval'. IEEE Int. Conf. Computer Vision and Pattern Recognition, Hilton Head, SC, USA, 2000
- [24] KHERFI M.L., ZIOU D.: 'Relevance feedback for CBIR: a new approach based on probabilistic feature weighting with positive and negative examples', *IEEE Trans. Image Process.*, 2006, **15**, (4), pp. 1017–1030
- [25] MUNEESAWANG P., GUAN L.: 'Automatic machine interactions for content-based image retrieval using a self-organizing tree map architecture', *IEEE Trans. Neural Netw.*, 2002, **13**, (4), pp. 821–834
- [26] VASCONCELOS N., LIPPMAN A.: 'Learning over multiple temporal scales in image databases'. European Conf. Computer Vision, 2000
- [27] VASCONCELOS N., LIPPMAN A.: 'Learning from user feedback in image retrieval systems'. Advances in Neural Information Processing Systems, 1999
- [28] SFIKAS G., CONSTANTINOPOULOS C., LIKAS A., GALATSANOS N.P.: 'An analytic distance metric for Gaussian mixture models with application in image retrieval'. Proc. Int. Conf. Artificial Neural Networks (ICANN 2005), Warsaw, Poland, 2005
- [29] MARAKAKIS A., GALATSANOS N., LIKAS A., STAFYLOPAPIS A.: 'A relevance feedback approach for content based image retrieval using Gaussian mixture models'. Proc. Int. Conf. Artificial Neural Networks (ICANN 2006), Athens, Greece, September 2006

[30] McLACHLAN G.M., PEEL D.: 'Finite mixture models' (John Wiley and Sons Inc, New York, 2001)

[31] Microsoft Research Cambridge Object Recognition Image Database: 'Version 1.0'. <http://research.microsoft.com/research/downloads/Details/b94de342-60dc-45d0-830b-9f6eff91b301/Details.aspx>

[32] VLASSIS N., LIKAS A.: 'A greedy EM algorithm for Gaussian mixture learning', *Neural Process. Lett.*, 2002, **15**, pp. 77–87

[33] BIMBO A.D.: 'Visual information retrieval' (Morgan Kaufmann, San Mateo, CA, 1999)

8 Appendix

8.1 Proof of (6.x)

We have

$$p_m(x) = \sum_{i=1}^{K_m} \pi_{mi} \phi(x|\theta_{mi}) \quad \text{and} \quad p_l(x) = \sum_{j=1}^{K_l} \pi_{lj} \phi(x|\theta_{lj})$$

with

$$\begin{aligned} \theta_{nk} &= (\mu_{nk}, \Sigma_{nk}) \quad \text{and} \quad \phi(x|\theta_{nk}) \\ &= N(x|\theta_{nk}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_{nk}|}} e^{-1/2(x-\mu_{nk})^T \Sigma_{nk}^{-1} (x-\mu_{nk})} \end{aligned}$$

Thus

$$\begin{aligned} \int p_m(x)p_l(x)dx &= \int \left[\left(\sum_{i=1}^{K_m} \pi_{mi} \phi(x|\theta_{mi}) \right) \right. \\ &\quad \left. \times \left(\sum_{j=1}^{K_l} \pi_{lj} \phi(x|\theta_{lj}) \right) \right] dx \\ \Rightarrow \int p_m(x)p_l(x)dx &= \sum_{i=1}^{K_m} \sum_{j=1}^{K_l} \pi_{mi} \pi_{lj} \int \phi(x|\theta_{mi}) \phi(x|\theta_{lj}) dx \end{aligned} \quad (20)$$

However

$$\begin{aligned} \int \phi(x|\theta_{mi}) \phi(x|\theta_{lj}) dx \\ = \int \frac{1}{\sqrt{(2\pi)^d |\Sigma_{mi}|}} e^{-1/2(x-\mu_{mi})^T \Sigma_{mi}^{-1} (x-\mu_{mi})} \end{aligned}$$

$$\begin{aligned} &\times \frac{1}{\sqrt{(2\pi)^d |\Sigma_{lj}|}} e^{-1/2(x-\mu_{lj})^T \Sigma_{lj}^{-1} (x-\mu_{lj})} dx \\ &= \frac{1}{(2\pi)^d \sqrt{|\Sigma_{mi}| |\Sigma_{lj}|}} \\ &\times \int e^{-\frac{1}{2}[(x-\mu_{mi})^T \Sigma_{mi}^{-1} (x-\mu_{mi}) + (x-\mu_{lj})^T \Sigma_{lj}^{-1} (x-\mu_{lj})]} dx \\ &= \frac{e^{-1/2(\mu_{mi}^T \Sigma_{mi}^{-1} \mu_{mi} + \mu_{lj}^T \Sigma_{lj}^{-1} \mu_{lj})}}{(2\pi)^d \sqrt{|\Sigma_{mi}| |\Sigma_{lj}|}} \\ &\times \int e^{-1/2(x^T \Sigma^{-1} x - x^T \Sigma^{-1} M - M^T \Sigma^{-1} x)} dx \end{aligned} \quad (21)$$

with

$$\Sigma = (\Sigma_{mi}^{-1} + \Sigma_{lj}^{-1})^{-1} \quad (22)$$

and

$$M = \Sigma (\Sigma_{mi}^{-1} \mu_{mi} + \Sigma_{lj}^{-1} \mu_{lj}) \quad (23)$$

Moreover, given that $\Theta = [M, \Sigma]$, we know that

$$\begin{aligned} \int \phi(x|\Theta) dx &= \int \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} e^{-1/2(x-M)^T \Sigma^{-1} (x-M)} dx = 1 \\ \Rightarrow \int e^{-\frac{1}{2}(x^T \Sigma^{-1} x - x^T \Sigma^{-1} M - M^T \Sigma^{-1} x)} dx &= \frac{\sqrt{(2\pi)^d |\Sigma|}}{e^{-1/2 M^T \Sigma^{-1} M}} \end{aligned} \quad (24)$$

From (21) and (24), we have

$$\begin{aligned} \int \phi(x|\theta_{mi}) \phi(x|\theta_{lj}) dx &= \frac{e^{-1/2(\mu_{mi}^T \Sigma_{mi}^{-1} \mu_{mi} + \mu_{lj}^T \Sigma_{lj}^{-1} \mu_{lj})}}{(2\pi)^d \sqrt{|\Sigma_{mi}| |\Sigma_{lj}|}} \\ &\times \frac{\sqrt{(2\pi)^d |\Sigma|}}{e^{-1/2 M^T \Sigma^{-1} M}} \\ &= \frac{1}{\sqrt{(2\pi)^d}} \sqrt{\frac{|\Sigma|}{|\Sigma_{mi}| |\Sigma_{lj}|}} e^R \end{aligned} \quad (25)$$

with R defined by

$$\begin{aligned} R &= \mu_{mi}^T \Sigma_{mi}^{-1} \mu_{mi} + \mu_{lj}^T \Sigma_{lj}^{-1} \mu_{lj} - M^T \Sigma^{-1} M \\ &= \mu_{mi}^T \Sigma_{mi}^{-1} \mu_{mi} + \mu_{lj}^T \Sigma_{lj}^{-1} \mu_{lj} - (\mu_{mi}^T \Sigma_{mi}^{-1} + \mu_{lj}^T \Sigma_{lj}^{-1}) \\ &\quad \times (\Sigma_{mi}^{-1} + \Sigma_{lj}^{-1})^{-1} (\Sigma_{mi}^{-1} \mu_{mi} + \Sigma_{lj}^{-1} \mu_{lj}) \end{aligned} \quad (26)$$

From the matrix inversion lemma, we have

$$\left(\Sigma_{mi}^{-1} + \Sigma_{lj}^{-1}\right)^{-1} = \Sigma_{mi} - \Sigma_{mi} \left(\Sigma_{mi} + \Sigma_{lj}\right)^{-1} \Sigma_{mi} \quad (27)$$

and

$$\left(\Sigma_{mi}^{-1} + \Sigma_{lj}^{-1}\right)^{-1} = \Sigma_{lj} - \Sigma_{lj} \left(\Sigma_{mi} + \Sigma_{lj}\right)^{-1} \Sigma_{lj} \quad (28)$$

Furthermore

$$\begin{aligned} \Sigma_{mi}^{-1} \left(\Sigma_{mi}^{-1} + \Sigma_{lj}^{-1}\right)^{-1} \Sigma_{lj}^{-1} &= \left[\Sigma_{lj} \left(\Sigma_{mi}^{-1} + \Sigma_{lj}^{-1}\right) \Sigma_{mi}\right]^{-1} \\ &= \left(\Sigma_{mi} + \Sigma_{lj}\right)^{-1} \\ \Rightarrow \left(\Sigma_{mi}^{-1} + \Sigma_{lj}^{-1}\right)^{-1} &= \Sigma_{mi} \left(\Sigma_{mi} + \Sigma_{lj}\right)^{-1} \Sigma_{lj} \end{aligned} \quad (29)$$

Similarly, we obtain

$$\left(\Sigma_{mi}^{-1} + \Sigma_{lj}^{-1}\right)^{-1} = \Sigma_{lj} \left(\Sigma_{mi} + \Sigma_{lj}\right)^{-1} \Sigma_{mi} \quad (30)$$

Combining (26) – (30), we obtain

$$\begin{aligned} R &= \mu_{mi}^T \Sigma_{mi}^{-1} \mu_{mi} + \mu_{lj}^T \Sigma_{lj}^{-1} \mu_{lj} \\ &- \mu_{mi}^T \Sigma_{mi}^{-1} \left(\Sigma_{mi}^{-1} + \Sigma_{lj}^{-1}\right)^{-1} \Sigma_{mi}^{-1} \mu_{mi} - \mu_{lj}^T \Sigma_{lj}^{-1} \\ &\times \left(\Sigma_{mi}^{-1} + \Sigma_{lj}^{-1}\right)^{-1} \Sigma_{lj}^{-1} \mu_{lj} \\ &- \mu_{mi}^T \Sigma_{mi}^{-1} \left(\Sigma_{mi}^{-1} + \Sigma_{lj}^{-1}\right)^{-1} \Sigma_{lj}^{-1} \mu_{lj} - \mu_{lj}^T \Sigma_{lj}^{-1} \\ &\times \left(\Sigma_{mi}^{-1} + \Sigma_{lj}^{-1}\right)^{-1} \Sigma_{mi}^{-1} \mu_{mi} \\ \Rightarrow R &= \left(\mu_{mi} - \mu_{lj}\right)^T \left(\Sigma_{mi} + \Sigma_{lj}\right)^{-1} \left(\mu_{mi} - \mu_{lj}\right) \end{aligned} \quad (31)$$

Finally, from (20), (22), (25) and (31), we obtain

$$\begin{aligned} &\int p_m(x) p_l(x) dx \\ &= \sum_{i=1}^{K_m} \sum_{j=1}^{K_l} \pi_{mi} \pi_{lj} \frac{1}{\sqrt{(2\pi)^d}} \sqrt{\frac{|\Sigma|}{|\Sigma_{mi}| |\Sigma_{lj}|}} e^R \end{aligned}$$

$$\begin{aligned} &= \frac{1}{\sqrt{(2\pi)^d}} \sum_{i=1}^{K_m} \sum_{j=1}^{K_l} \pi_{mi} \pi_{lj} \\ &\times \sqrt{\frac{\left|\left(\Sigma_{mi}^{-1} + \Sigma_{lj}^{-1}\right)^{-1}\right|}{|\Sigma_{mi}| |\Sigma_{lj}|}} e^{(\mu_{mi} - \mu_{lj})^T \left(\Sigma_{mi} + \Sigma_{lj}\right)^{-1} (\mu_{mi} - \mu_{lj})} \\ &= \frac{1}{\sqrt{(2\pi)^d}} \sum_{i=1}^{K_m} \sum_{j=1}^{K_l} \frac{\pi_{mi} \pi_{lj}}{\sqrt{|\Sigma_{mi} + \Sigma_{lj}|}} e^{(\mu_{mi} - \mu_{lj})^T \left(\Sigma_{mi} + \Sigma_{lj}\right)^{-1} (\mu_{mi} - \mu_{lj})} \end{aligned}$$

Thus, from (5.2), we have

$$\begin{aligned} S_{ml} &= \int p_m(x) p_l(x) dx \\ \Rightarrow S_{ml} &= \frac{1}{\sqrt{(2\pi)^d}} \sum_{i=1}^{K_m} \sum_{j=1}^{K_l} \frac{\pi_{mi} \pi_{lj}}{\sqrt{|C_{ml}(i, j)|}} e^{k_{ml}(i, j)} \end{aligned}$$

with

$$\begin{aligned} C_{ml}(i, j) &= \Sigma_{mi} + \Sigma_{lj} \\ k_{ml}(i, j) &= \left(\mu_{mi} - \mu_{lj}\right)^T C_{ml}^{-1}(i, j) \left(\mu_{mi} - \mu_{lj}\right) \end{aligned}$$

8.2 Proof of Remark 1

Because of Jensen's inequality it holds that

$$\begin{aligned} \log \int p_1(x) p_2(x) dx &\leq \int p_1(x) \log p_2(x) dx \quad \text{and} \\ \log \int p_1(x) p_2(x) dx &\leq \int p_2(x) \log p_1(x) dx \end{aligned}$$

From (2) and (5.x), we have

$$\begin{aligned} \text{KL}(p_1 \| p_2) &= \int p_1(x) \log \frac{p_1(x)}{p_2(x)} dx \quad \text{and} \quad C2(p_1, p_2) \\ &= -\log \frac{2 \int p_1(x) p_2(x) dx}{\int p_1^2(x) dx + \int p_2^2(x) dx} \end{aligned}$$

Therefore it holds that

$$\begin{aligned} \text{KL}(p_1 \| p_2) - C2(p_1, p_2) &\leq \log 2 + \int p_1(x) \log p_1(x) dx \\ &- \log \int [p_1^2(x) + p_2^2(x)] dx \end{aligned}$$

which leads to

$$\text{KL}(p_1||p_2) - C2(p_1, p_2) \leq \log 2 + \int p_1(x) \log p_1(x) dx - \log \int p_1^2(x) dx$$

By defining $\Delta_i(p_i) = \int p_i(x) \log p_i(x) dx - \log \int p_i^2(x) dx = E[\log p_i]_{p_i} - \log E[p_i]_{p_i} \geq 0$ (because of Jensen's inequality) we, finally, obtain

$$\text{KL}(p_1||p_2) - C2(p_1, p_2) \leq \log 2 + \Delta_1(p_1)$$

Similarly

$$\text{KL}(p_2||p_1) - C2(p_1, p_2) \leq \log 2 + \Delta_2(p_2)$$

Adding the last two inequalities, we obtain

$$\text{SKL}(p_1, p_2) - C2(p_1, p_2) \leq \frac{1}{2} [\Delta_1(p_1) + \Delta_2(p_2)] + \log 2$$

8.3 Proof of (11.x)

From (10), we have

$$q'(x) = (1 - \Lambda)q(x) + \sum_{m=1}^M \lambda_m r_m(x)$$

From (5.1), the C2 divergence between the new query and the i th database image is given by

$$C2(q', i) = -\log \frac{2S_{q'i}}{S_{q'q'} + S_{ii}}$$

Moreover, from (5.2), we obtain

$$\begin{aligned} S_{q'i} &= \int q'(x) i(x) dx \\ &= \int \left[(1 - \Lambda)q(x) + \sum_{m=1}^M \lambda_m r_m(x) \right] i(x) dx \\ \Rightarrow S_{q'i} &= (1 - \Lambda) \int q(x) i(x) dx + \sum_{m=1}^M \lambda_m \int r_m(x) i(x) dx \\ \Rightarrow S_{q'i} &= (1 - \Lambda)S_{qi} + \sum_{m=1}^M \lambda_m S_{r_m i} \end{aligned}$$

Furthermore

$$\begin{aligned} S_{q'q'} &= \int q'(x) q'(x) dx \\ &= \int \left[(1 - \Lambda)q(x) + \sum_{m=1}^M \lambda_m r_m(x) \right] \times \left[(1 - \Lambda)q(x) + \sum_{m'=1}^M \lambda_{m'} r_{m'}(x) \right] dx \\ &= (1 - \Lambda)^2 \int q^2(x) dx + 2(1 - \Lambda) \times \sum_{m=1}^M \lambda_m \int q(x) r_m(x) dx \\ &\quad + \sum_{m=1}^M \sum_{m'=1}^M \lambda_m \lambda_{m'} \int r_m(x) r_{m'}(x) dx \\ \Rightarrow S_{q'q'} &= (1 - \Lambda)^2 S_{qq} + 2(1 - \Lambda) \sum_{m=1}^M \lambda_m S_{qr_m} \\ &\quad + \sum_{m=1}^M \sum_{m'=1}^M \lambda_m \lambda_{m'} S_{r_m r_{m'}} \end{aligned}$$

Of course, $S_{ii} = \int i^2(x) dx$ and it does not change from one RF round to another.

8.4 Proof of (15)

From (10), we have

$$q'(x) = (1 - \Lambda)q(x) + \sum_{m=1}^M \lambda_m r_m(x)$$

Given that

$$\begin{aligned} q(x) &= \sum_{j=1}^{N_q} \pi_{qj} \phi(x|\theta_{qj}) \quad \text{with } \theta_{qj} = (\mu_{qj}, \Sigma_{qj}) \\ r_m(x) &= \sum_{j=1}^{N_m} \pi_{mj} \phi(x|\theta_{mj}) \quad \text{with } \theta_{mj} = (\mu_{mj}, \Sigma_{mj}) \end{aligned}$$

and

$$i(x) = \sum_{j=1}^{N_i} \pi_{ij} \phi(x|\theta_{ij}) \quad \text{with } \theta_{ij} = (\mu_{ij}, \Sigma_{ij})$$

the new query model is given by

$$\begin{aligned} q'(x) &= \sum_{j=1}^{N_{q'}} \pi_{q'j} \phi(x|\theta_{q'j}) \\ &= \sum_{j=1}^{N_q} (1 - \Lambda) \pi_{qj} \phi(x|\theta_{qj}) + \sum_{j=1}^{N_1} \lambda_1 \pi_{1j} \phi(x|\theta_{1j}) \\ &\quad + \dots + \sum_{j=1}^{N_M} \lambda_M \pi_{Mj} \phi(x|\theta_{Mj}) \end{aligned} \tag{32}$$

with $N_{q'} = N_q + N_1 + \dots + N_M$.

Moreover, from (4.x), we have

$$\text{ALA}(q' \| i) = \sum_{j=1}^{N_{q'}} \pi_{q'j} \left\{ \log \pi_{i\beta_{q'i}(j)} + \left[\log \phi(\mu_{q'j} | \theta_{i\beta_{q'i}(j)}) - \frac{1}{2} \text{trace}(\Sigma_{i\beta_{q'i}(j)}^{-1} \Sigma_{q'j}) \right] \right\}$$

with

$$\begin{aligned} \beta_{q'i}(j) = k &\Leftrightarrow \left\| \mu_{q'j} - \mu_{ik} \right\|_{\Sigma_{ik}}^2 - \log \pi_{ik} \\ &< \left\| \mu_{q'j} - \mu_{il} \right\|_{\Sigma_{il}}^2 - \log \pi_{il}, \quad \forall l \neq k \end{aligned}$$

Taking into account (32), the sets of $\pi_{q'j}$, $\mu_{q'j}$ and $\Sigma_{q'j}$ are

$$C_{\pi} = \left\{ (1 - \Lambda) \pi_{q1}, \dots, (1 - \Lambda) \pi_{qN_q}, \lambda_1 \pi_{11}, \dots, \lambda_1 \pi_{1N_1}, \dots, \lambda_M \pi_{M1}, \dots, \lambda_M \pi_{MN_M} \right\}$$

$$C_{\mu} = \left\{ \mu_{q1}, \dots, \mu_{qN_q}, \mu_{11}, \dots, \mu_{1N_1}, \dots, \mu_{M1}, \dots, \mu_{MN_M} \right\}$$

$$C_{\Sigma} = \left\{ \Sigma_{q1}, \dots, \Sigma_{qN_q}, \Sigma_{11}, \dots, \Sigma_{1N_1}, \dots, \Sigma_{M1}, \dots, \Sigma_{MN_M} \right\}$$

respectively. Moreover, each correspondence function value, $\beta_{q'i}(j)$, depends only on $\mu_{q'j} \in C_{\mu}$ and on

$\pi_{ik}, \mu_{ik}, \Sigma_{ik}, k = 1, \dots, N_i$. Thus, the set of $\beta_{q'i}(j)$ is

$$C_{\beta} = \left\{ \beta_{q'i}(1), \dots, \beta_{q'i}(N_q), \beta_{1i}(1), \dots, \beta_{1i}(N_1), \dots, \beta_{Mi}(1), \dots, \beta_{Mi}(N_M) \right\}$$

and, for the $\text{ALA}(q' \| i)$, it holds that

$$\begin{aligned} \text{ALA}(q' \| i) &= \sum_{j=1}^{N_q} (1 - \Lambda) \pi_{qj} \left\{ \log \pi_{i\beta_{q'i}(j)} \right. \\ &\quad \left. + \left[\log \phi(\mu_{qj} | \theta_{i\beta_{q'i}(j)}) - \frac{1}{2} \text{trace}(\Sigma_{i\beta_{q'i}(j)}^{-1} \Sigma_{qj}) \right] \right\} \\ &\quad + \sum_{j=1}^{N_1} \lambda_1 \pi_{1j} \left\{ \log \pi_{i\beta_{1i}(j)} + \left[\log \phi(\mu_{1j} | \theta_{i\beta_{1i}(j)}) \right. \right. \\ &\quad \left. \left. - \frac{1}{2} \text{trace}(\Sigma_{i\beta_{1i}(j)}^{-1} \Sigma_{1j}) \right] \right\} \\ &\quad + \dots + \sum_{j=1}^{N_M} \lambda_M \pi_{Mj} \left\{ \log \pi_{i\beta_{Mi}(j)} \right. \\ &\quad \left. + \left[\log \phi(\mu_{Mj} | \theta_{i\beta_{Mi}(j)}) \right. \right. \\ &\quad \left. \left. - \frac{1}{2} \text{trace}(\Sigma_{i\beta_{Mi}(j)}^{-1} \Sigma_{Mj}) \right] \right\} \\ &\Rightarrow \text{ALA}(q' \| i) = (1 - \Lambda) \text{ALA}(q \| i) \\ &\quad + \sum_{m=1}^M \lambda_m \text{ALA}(r_m \| i) \end{aligned}$$