Pattern Recognition Letters 29 (2008) 2174–2181

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Combining similarity measures in content-based image retrieval

Miguel Arevalillo-Herráez [a,*], Juan Domingo [b], Francesc J. Ferri [a]

[a] Department of Computer Science, University of Valencia, Avda. Vicente Andrés Estellés, 1. 46100-Burjasot, Spain
[b] Institute of Robotics, University of Valencia, Spain

ARTICLE INFO

Article history:
Received 26 August 2007
Received in revised form 5 August 2008
Available online 15 August 2008

Communicated by M.-J. Li

Keywords:
CBIR
Combining descriptors
Similarity function
Score normalization
Probabilistic

ABSTRACT

The purpose of content based image retrieval (CBIR) systems is to allow users to retrieve pictures from large image repositories. In a CBIR system, an image is usually represented as a set of low level descriptors from which a series of underlying similarity or distance functions are used to conveniently drive the different types of queries. Recent work deals with combination of distances or scores from different and usually independent representations in an attempt to induce high level semantics from the low level descriptors of the images. Choosing the best method to combine these results requires a careful analysis and, in most cases, the use of ad-hoc strategies. Combination based on or derived from product and sum rules are common approaches. In this paper we propose a method to combine a given set of dissimilarity functions. For each similarity function, a probability distribution is built. Assuming statistical independence, these are used to design a new similarity measure which combines the results obtained with each independent function.

1. Introduction

CBIR systems (Datta et al., 2008; Lew et al., 2006) aim to recover pictures from large image repositories, according to the user's interest. Usually, a CBIR system represents the images in the repository as a multi-dimensional feature vector extracted from a series of low level descriptors, such as color, texture or shape. The subjective similarity between two pictures is usually quantified in terms of a particular measure of distance defined on the corresponding multi-dimensional feature space.

Most retrieval systems including CBIR ones explicitly rely on distance, similarity or score functions aiming at relating descriptors to perceptual or subjective resemblance to some extent (Neumann and Gegenfurtner, 2006; Li and Chang, 2003). The Minkowski-form distance, the Manhattan distance, the Euclidean distance, the Hausdorff distance, the Quadratic Form (QF) distance, the Mahalanobis' distance, the Kullback-Leibler divergence (Do and Vetterli, 2002) and the Jeffrey divergence are some of the most commonly used functions to estimate the similarity between pictures. These are applied on low level descriptors such as the color histogram (Swain and Ballard, 1991; Pass et al., 1996), the co-occurrence matrix (Haralick et al., 1973), morphological features (Ayala et al., 2001), wavelet-based descriptors (Chuang and Kuo, 1996) or Zernike moments (Khotanzad and Hong, 1990). Further details on these and other dissimilarity measures in the context of image retrieval are provided in (Rubner et al., 2001; Kamarainen et al., 2003). For more information on feature extraction, the reader is referred to e.g. (Long et al., 2003).

System performance heavily depends on the descriptors and underlying similarity function used. It is a common practice to design a number of similarity measures as described above, each using a different function acting on different or even disjoint subsets of the available features. These similarities are later combined to produce a definitive result that may consist of a composite similarity or score value. Different strategies to appropriately perform this combination have been proposed in general information retrieval (Fernandez et al., 2006) and biometric recognition (Jain et al., 2005; Prabhakar and Jain, 2002). In the particular case of CBIR, a multi-objective optimization technique based on a Pareto Archive Evolution Strategy (PAES) (Knowles and Corne, 2000) has been used in (Zhang et al., 2006) to define a global measure as an optimal linear combination of partial similarity functions; in (Iqbal and Aggarwal, 2002), color, texture and structure distances were pre-processed using a technique based on Gaussian normalization, and a global measure was defined as a weighted linear combination of the normalized distances; in (Giacinto and Roli, 2004) a simple linear normalization so that features are mapped to the range $[0, 1]$ was applied; and in (Torres et al., 2005) a genetic algorithm was used to derive a set of weights for each descriptor.

Evidence combination has also been used in complete CBIR systems usually to fuse possibly multimodal information taken from user feedback (Urban and Jose, 2004; Bruno et al., 2007).

In this paper, a novel probabilistic strategy to combine similarity measures is proposed. The method takes ideas from different

* Corresponding author. Tel.: +34 96 354 39 62; fax: +34 96 354 47 68.
E-mail addresses: Miguel.Arevalillo@uv.es (M. Arevalillo-Herráez), Juan.Domingo@uv.es (J. Domingo), Francesc.Ferri@uv.es (F.J. Ferri).

0167-8655/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2008.08.003

contexts and builds upon preliminary ideas presented in (Arevalillo-Herráez et al., 2008). This approach can be regarded as optimal from the point of view that it best represents empirically assessed user preferences under certain independence assumptions.

The remainder of this paper is organized as follows. First, a method to convert a similarity function into another one which has a self contained probabilistic meaning is described. Next, a probabilistic reasoning is followed to yield a simple method to combine a series of such functions. Then, some implementation issues related to the computation of the probabilities and the method used to gather the empirical data required by the method are discussed. After, a series of experiments that evidence the potential of the technique are presented. Finally, some conclusions are drawn.

2. From similarities to probabilities

Let us assume our objects are represented in a particular feature space where the whole set of available (vector) features is designated as $\mathbb{F}$ and $\{\mathbb{F}^{(i)}\}_{i=1}^{n}$ is a family of subsets of $\mathbb{F}$.

We define a similarity measure as one which makes use of a subset of features to produce a value which is as related as possible to the subjective perception of similarity, in such a way that the higher the value the more similar (or less similar for dissimilarity measures). A typical example of a (dis)similarity measure is the application of the Euclidean distance to color histograms.[1]

Mathematically, a similarity measure using a subset of the features $\mathbb{F}^{(i)}$ can be expressed as

$$s_i : \mathbb{F}^{(i)} \times \mathbb{F}^{(i)} \rightarrow \mathbb{R}^{\geqslant 0}$$

In the context of this work, a probabilistic similarity measure is a function that produces values that can be directly translated into probabilities. This can refer to probabilities of belonging to the same category (Jain et al., 2005), of being relevant to a query (Nottelmann and Fuhr, 2003) or, as in our case, being subjectively similar.

Converting a measure which only has a meaning when two measurements are compared (such as Euclidean distance) into another one which has a self-contained probabilistic meaning provides a significant advantage specially when these values are to be further processed.

Similar transformations are usually referred to in the information retrieval literature as normalization. Normalization methods play usually an homogenization role prior either to combining or post-processing the results of one or several queries. Some normalization methods have also been related to a probabilistic interpretation (Fernandez et al., 2006).

In this particular work, the key fact is to model the similarity between images according to the preferences of a generic user which are assumed to be roughly the same for a large number of users. Obviously, it is implicitly assumed that this subjective similarity can be modelled as a probability and that there is at least a weak relation or dependence between this and the family of similarity measures considered.

Let us assume that the fact that a user considers that any two images in a given repository are similar can be conveniently modelled in terms of the particular similarity value obtained using the function s_i. Similarly to other works (Manmatha et al., 2001; Nottelmann and Fuhr, 2003; Fernandez et al., 2006) we will denote this (posterior) probability simply by $p(similar|x_i)$ where the subindex refers to using the ith basic similarity measure (using ith feature subset) and x_i represents a given similarity value.

[1] From this point on, and unless otherwise stated, we will assume without loss of generality that we work with similarity measures.

If we use the Bayes rule, this probability can be written as

$$p(similar|x_i) = \frac{p(x_i|similar) \cdot P(similar)}{p(x_i)} \quad (1)$$

where $p(x_i|similar)$ is the conditional probability density function associated with similarity values produced by s_i, $P(similar)$ is the prior probability of images being similar and $p(x_i)$ is the unconditional probability that s_i produces a value x_i. This last probability could also be written as $P(similar)p(x_i|similar) + (1 - P(similar))\, p(x_i|dissimilar)$.

In the equation above, and for a given repository and application, $P(similar)$ can be estimated from data or even fixed for convenience while the probability distribution functions $p(x_i|similar)$, $p(x_i|dissimilar)$ and $p(x_i)$ could in principle be independently or jointly estimated using either parametric or non-parametric methods (Fernandez et al., 2006; Jain et al., 2005; Manmatha et al., 2001).

3. A composite probabilistic similarity measure

In our particular context, the goal consists of introducing a composite similarity measure that uses all available features by combining and aggregating all the information provided by all the particular basic similarity measures considered.

Let us assume that we have a set of particular similarity measures $S = \{s_1, s_2, \ldots, s_n\}$ each defined on a different subset of features, $\mathbb{F}^{(i)}$.

We wish to design a new similarity function on the entire set of features

$$s : \mathbb{F} \times \mathbb{F} \rightarrow \mathbb{R}^{\geqslant 0}$$

It would be fair to associate the value of s for any two images with the probability that a user judges them as being similar. To compute this probability, we can reuse the information produced by each of the basic similarity measures in the set S. Let us denote by $\boldsymbol{x}$ the vector $(x_1, \ldots, x_n)$, containing the values produced by each of the similarity functions s_i.

Then, we can consider the probability that two images in the given repository are considered similar conditioned to the fact that the basic similarity measures $s_1, \ldots, s_n$ have produced values $x_1, \ldots, x_n$, respectively. As in the previous section, we write this composite probability as $p(similar|\boldsymbol{x})$.

Applying the Bayes rule this probability turns into

$$p(similar|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|similar) \cdot P(similar)}{p(\boldsymbol{x})}$$

where the new $p(\boldsymbol{x}|similar)$ and $p(\boldsymbol{x})$ represent as in the previous section, the conditional and unconditional probability density functions corresponding to obtaining a particular similarity vector, $\boldsymbol{x}$ when using the family of similarities S.

In order to make use of the above expression we will assume from now on the mutual independence of the values obtained through the similarity measures in S. In this work, as in most closely related prior work, the similarity measures defined on the different subsets of features can be considered as both conditionally and unconditionally independent. This is merely an approximation for particular pairs of closely related similarities. The effect of this assumption on the behavior of the proposed method for different sets of similarities will be empirically assessed in Section 5.

By using the independence assumption, the above expression for the composite probability may be rewritten as

$$p(similar|\boldsymbol{x}) = P(similar) \cdot \prod_{i=1}^{n} \frac{p(x_i|similar)}{p(x_i)} \quad (2)$$

and simplified further using Eq. (1):

$$p(similar|\boldsymbol{x}) = P(similar)^{(1-n)} \cdot \prod_{i=1}^{n} p(similar|x_i) \quad (3)$$

This expression is in fact the well-known product rule for the combination of classifiers when they act on conditionally independent representations (Kittler et al., 1998). Other combination rules apart from the product one would be possible by adding further assumptions but we will restrict ourselves only to the product rule and the independence assumption. On the other hand, by reasons that will become clear, we will prefer the expression in Eq. (2) instead of the more familiar one that explicitly uses the basic posterior probabilities in Eq. (3).

Notice that the final similarity function is proportional to the product of the individual distances between each set of components. This is consistent with the fact that independence between them has been assumed and also that a logical 'and' operator has indeed been used.

4. Effectively combining similarities into a composite measure

Once the estimated composite probabilistic similarity has been written in terms of the basic similarity functions, one of the possible ways of using this consists of estimating the conditional probabilities from empirical data. This would allow the evaluation of the expression in Eq. (2) for new data.

Even though posterior probabilities could be directly estimated (as e.g. in k-NN methods Duda et al., 2000) we consider here the separate estimation of $p(x_i|similar)$ and $p(x_i)$. The motivation for this comes partially from the fact that relatively much more data is available to estimate the unconditional probability.

Basically for the same reasons explained in (Prabhakar and Jain, 2002) all estimates in this work will be produced using the Kernel Density or Parzen Windows method (Duda et al., 2000; Silverman, 1986). In particular, the kernel density estimate $g_m(x)$ obtained from m samples $\{X_j\}_{j=1}^m$ drawn from an unknown density g is given by:

$$g_m(x) = \frac{1}{m \cdot h} \sum_{j=1}^{m} K\left(\frac{x - X_j}{h}\right)$$

where K is the kernel or window function and h is the window size or smoothing parameter. Gaussian windows are commonly used because of their flexibility and ability to give smooth estimates. The smoothing parameter is critical for different applications and must be carefully selected based on the data at hand. This has been addressed by using a plug-in method (Sheather and Jones, 1991) which uses an iterative procedure to solve a non-linear equation, each iteration involving the use of the density estimator with a different smoothing parameter.

A numerical problem arises with this estimation. Since the estimate of $p(x_i)$ is later used in the denominator of a fraction (see Eq. (2)) we must be sure that it does not vanish or becomes so small that it causes numerical accuracy problems. This happens when the probability distribution of the values produced by the function s_i concentrates in a small interval of its range. The case of a peaked probability distribution also produces a sampling problem. Since the whole interval of possible similarity values is sampled at the same rate, such a distribution leads to a poor definition of the most frequent values.

To better deal with numerical problems and further simplify the corresponding expressions, we propose to replace the division in Eq. (2) by an equalization of the values produced by each of the similarity functions s_i, in such a way that $p(x_i)$ becomes uniform in its range. This equalization operation takes away the denominator in the equation and does not change the meaning of the previous expressions, that only suffer a change of variable. Equalizing probability densities (or histograms in the context of image analysis) (Gonzales and Wintz, 1987) is a well known technique that has also been used previously in the context of score normalization in information retrieval (Fernandez et al., 2006).

Each of the equalizing functions E_i is given by the cumulative density function corresponding to $p(x_i)$ as a function of x_i. With this transformation, $y_i = E_i(x_i)$ is a new real-valued random variable which is unconditionally uniformly distributed in $[0, 1]$ and whose conditional distribution is given by $p(y_i|similar)$. With this change of variable, the Eq. (2) becomes

$$p(similar|\boldsymbol{x}) = P(similar) \cdot \prod_{i=1}^{n} p(E_i(x_i)|similar) \tag{4}$$

When this composite probability is to be used to account for the similarity between a given pair of images, the constant term $P(similar)$ can be safely ignored; note that we are not interested in the particular probability value but only in the order it induces in a given set of the images, i.e., it is used to rank images by their similarity to a given query.

Fig. 1 shows an illustrative example of the described procedure to transform the original similarity value x_i into a more meaningful measure $p(similar|y_i)$. Fig. 1a shows the histogram of the distance values produced by a typical dissimilarity function. In Fig. 1b the cumulative histogram which is used to equalize the distances is illustrated. In Fig. 1c, the histogram once the distances have been equalized is depicted. A kernel approximation of the distribution of $p(y_i|similar)$ is also shown, scaled so that its area matches that of the histogram.

5. Experimental results

To demonstrate the validity of the technique, a number of experiments have been carried out, using three different data sets:

- A database containing 1508 pictures, some of which were extracted from the web and others were taken by the members of the research group. These have been manually classified as belonging to 28 different semantic concepts such as flowers, horses, paintings, skies, textures, ceramic tiles, buildings, clouds, trees, etc. The number of images in each of these categories varies from 24 to 300. This database and corresponding labels have also been used in (de Ves et al., 2006 and León et al., 2007), where further details can be found. The features which have been computed for these pictures are a flattened 10×3 HS (Hue-Saturation) histogram, and the horizontal and vertical granulometries (Soille, 2003), calculated according to (de Ves et al., 2006). We have used the histogram intersection (Swain and Ballard, 1991) to estimate color similarities. For the rest of the features and in general for the rest of the paper unless otherwise specified, the Euclidean distance has been used.
- A far larger database composed of a total of 102 894 royalty free photographs extracted from a commercial collection called "Art Explosion", distributed by the company Nova Development (http://www.novadevelopment.com). The images in this repository are organized in 201 thematic folders. Six texture features have been computed for this repository, namely Gabor Convolution Energies (Smith and Burns, 1997), Gray Level Co-occurrence Matrix (Conners et al., 1984), Gaussian Random Markov Fields (Chellappa and Chatterjee, 1985), the coefficients of fitting the granulometry distribution with a B-spline basis (Chen and Dougherty, 1994) and two versions of the Spatial Size distribution (Ayala et al., 2001), one using a horizontal segment and another with a vertical segment.
- A subset of the previous database, composed of 5476 images classified into 62 categories. The themes in the previous database have been replaced by categories where images have been carefully selected so that the ones in the same category represent a similar semantic concept.

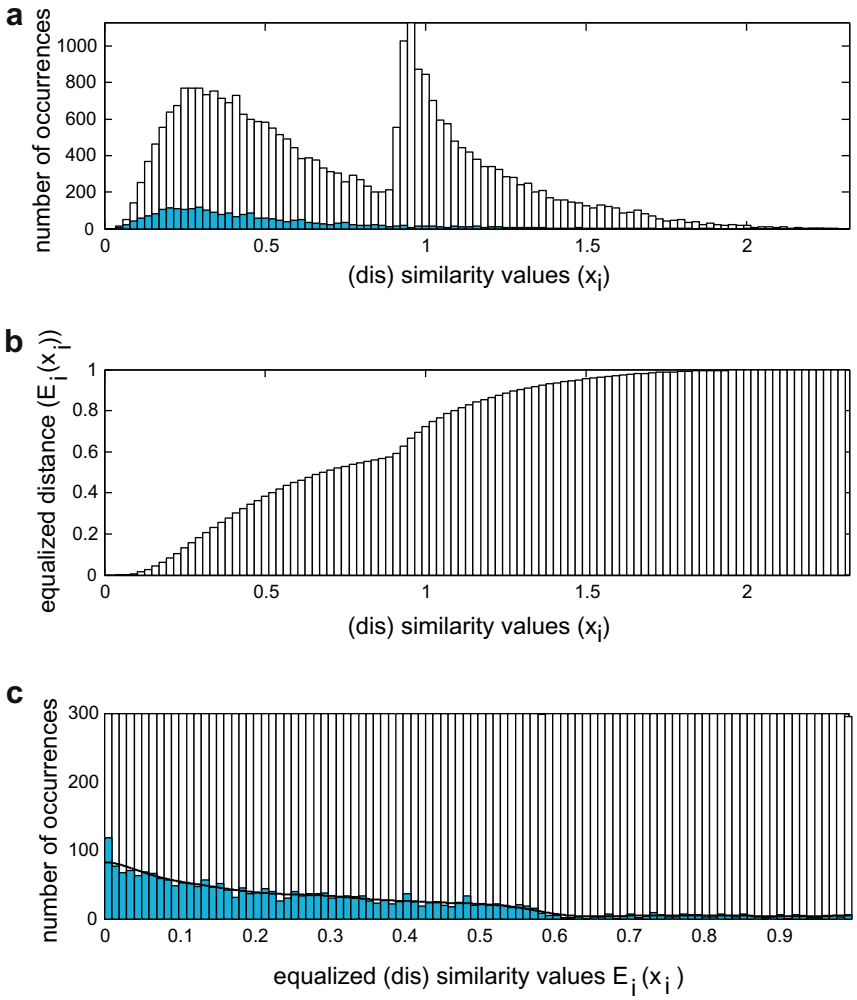

Fig. 1. Histograms representing the frequency of distance values x_i between similar (dark) and dissimilar (white) pairs shown for evaluation; (a) before equalization, (b) equalizing function $E(x_i)$ and (c) after equalization using $E(x_i)$, along with a scaled version of the corresponding kernel estimate.

For experimental purposes, and with the exception of a last experiment, the available categories have been used as concepts. User judgments about similarity have been simulated considering that all pictures under the same category are similar, and all images under different categories represent different concepts and are then dissimilar. This allows us to easily study the behavior of the technique as the number of training samples increases and obtain quantitative performance measures.

Two experiments have been carried out using the first data set. In a first experiment, we study the behavior of the method for different sizes of the training set. Then, a second experiment compares the algorithm to classical combination approaches. For the first experiment we have used a fixed and increasing number of random samples to produce the corresponding composite similarity measure given by Eq. (4). Then this new function has been used on a different and independent set of 50 000 image pairs extracted from the same repository, and its associated kernel density estimate (after equalization) has been calculated using the categories of these pairs as a ground truth. This is equivalent to estimating the function $p(E_i(x_i)|similar)$ for the composite measure, as if it was a basic function. The shape of this function can provide a good indication of the expected discriminatory power of the similarity function. Highly skewed curves suggest a better performance, as they imply higher relative probabilities that closer images are judged similar by a generic user.

This entire procedure has been repeated for training sets of increasing sizes to study the influence of the size on the performance of the resulting composite function. In order to quantify these results, two measures on the corresponding distribution have been considered: the (negative) skewness and the probability that a similar pair is within the 10% most similar pairs according to the ranking established by the composite function. For the sake of robustness, the experiments have been run 10 times for each different size of the training set and the results have been averaged. Fig. 2a shows these data. As expected, the performance of the measure grows with the size of the training set. In Fig. 2b the actual probability distributions are shown to allow for a relative visual comparison for different values of the size of the training set. For clarity reasons, only seven functions have been plotted.

Although it is not shown in the plots, the experiments performed have also revealed that with a relatively small training set of only 256 judgments, the composite measure performs better than any of the three basic functions.

A more exhaustive and objective experiment consists of evaluating precision and recall graphs on a large amount of independent data. Precision vs recall curves are a common method to present results in the context of information retrieval, and they provide a good indication of performance (Müller et al., 2001). Precision and recall can be measured for different numbers of retrieved images, a value which is usually referred to as the cut-off. In particular

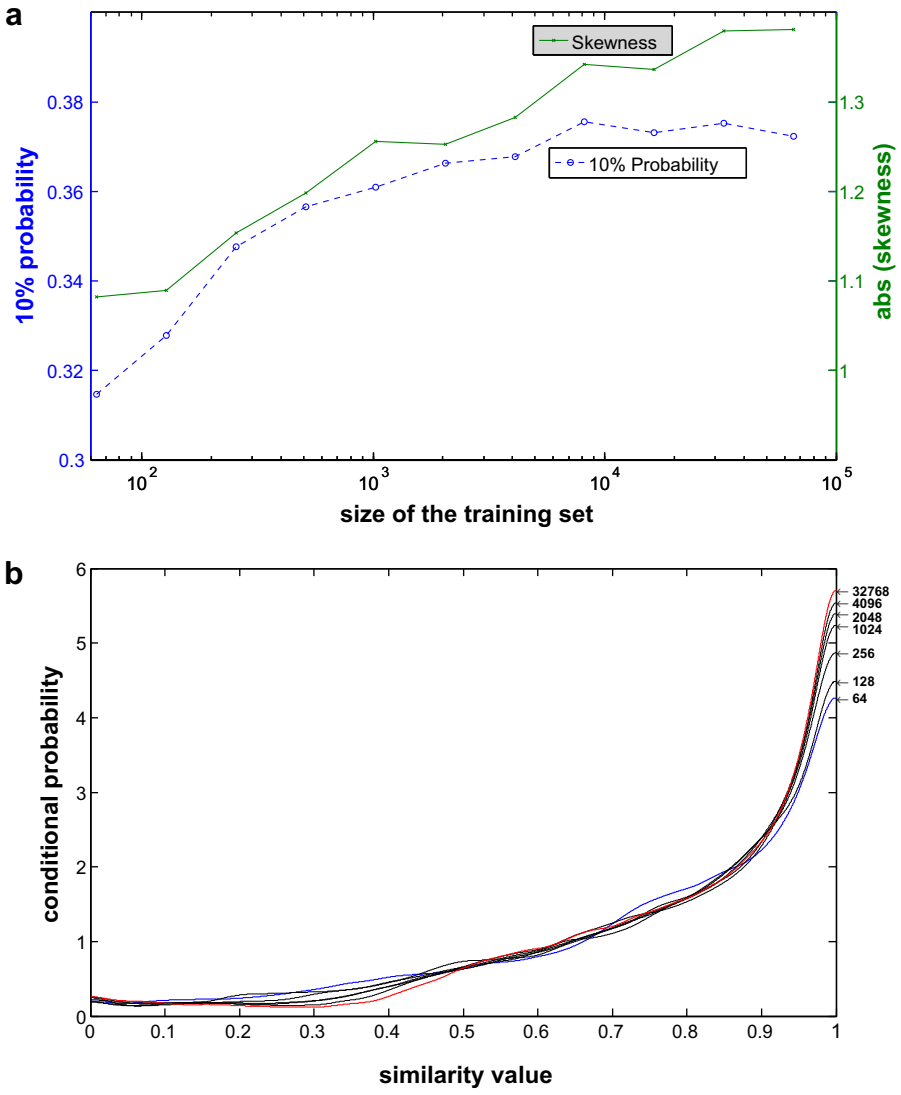

Fig. 2. Behavior of the composite similarity measure as the training data increases (a) probability that a pair judged as similar is within the 10% most similar pairs according to the composite similarity measure, and skewness of the resulting probability distribution. (b) Estimated density plot for different number of training pairs.

$$precision = \frac{\text{number of relevant images retrieved}}{\text{total number of images retrieved}} \quad (5)$$

$$recall = \frac{\text{number of relevant images retrieved}}{\text{total number of relevant images in the collection}} \quad (6)$$

In this experiment, randomly selected sets of 30000 pairs of pictures (less than 3% of the pairs that can be built from the whole set) have been considered for training while the rest of the over one million possible pairs of pictures have been kept for testing. The training data is used to estimate both the equalizing functions and the conditional probabilities to construct the composite similarity. Then, the proposed composite similarity is used to evaluate the probability of being similar on the test data while the user judgments on all possible pairs from the database is used as a ground truth. These judgments allow us to rank the test data by probability and compute precision and recall, considering that pairs under the same category are relevant results.

In this case we have run each experiment 20 times and averaged the results obtained for each cut-off value. The resulting plots conveniently summarize the behavior of the different similarity measures considered over the whole range of situation. The outcome is shown in Fig. 3 where it can be observed that the proposed composite measure significantly outperforms any of the original measures and their sum using either a linear (Giacinto and Roli, 2004) or a Gaussian normalization (Iqbal and Aggarwal, 2002), which have also been implemented. Notice that at recall values below 0.015 the Gaussian normalization obtains a slightly better performance. However, the differences observed are not significant and the normalization proposed in this paper quickly provides better results.

To test the effectiveness of the approach when applied to larger databases and with a larger number of distance functions, a similar experiment has been performed on the other two repositories presented above. On the second database, it has been necessary to use an alternative approach because of the unfeasible amount of memory which would otherwise be required to store the over 10^9 possible pairs. In particular, sets of 150000 image pairs have been randomly chosen to train the function, and the evaluation has been carried out on sets composed of other different 200000 image pairs, also chosen at random. These results have been averaged over 20 runs. The low precision values obtained for all approaches evidences that the initial folder-based classification provided in "Art explosion" does not correspond well to differentiated semantic concepts. Still, the use of the composite measure proposed

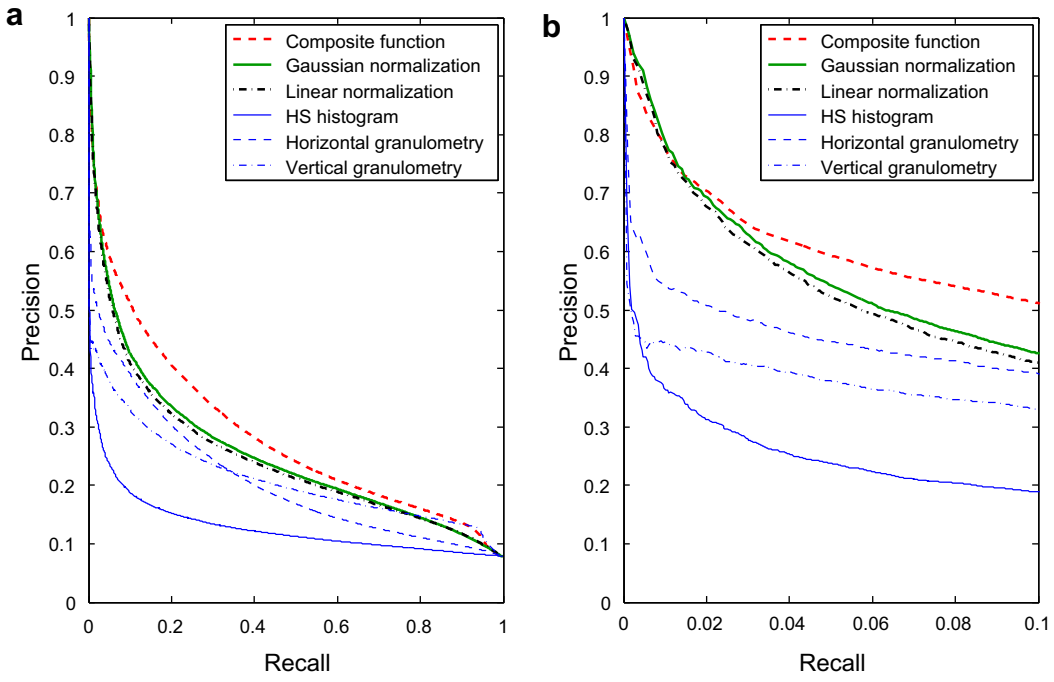

Fig. 3. Precision-recall graphs on independent test data for the original measures and their sum using either a linear or a Gaussian normalization (a) entire graph; (b) zoomed version to appreciate details which cannot be easily observed in (a).

yields better results. An experiment that followed the same set-up as for the first repository has also been performed on the third database. This time, the composite function was built using nearly 300 000 pairs (1% of the total number of pairs) and the results were averaged over 20 runs. Both results are shown in Fig. 4 in the form of precision-recall graphs, appropriately zoomed to appreciate the relevant details. Again, it can be observed that the composite measure proposed significantly outperforms the other two combination methods.

A last experiment has been carried out to evaluate the new similarity function in a more practical and realistic setting, using the third picture database. An application has been built that chooses

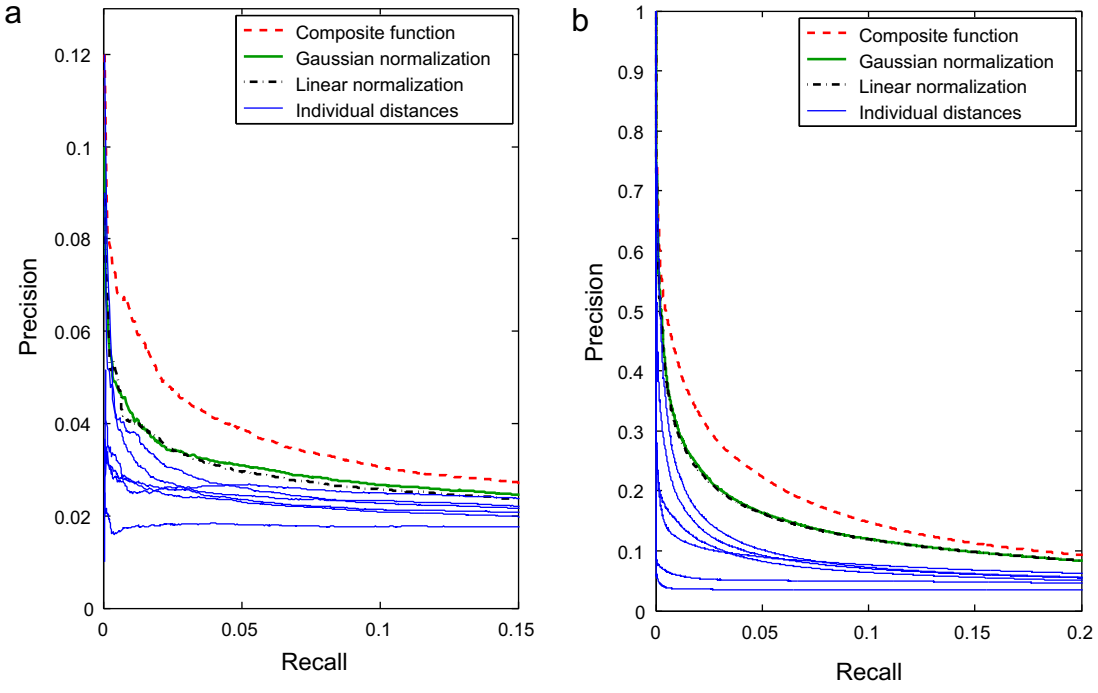

Fig. 4. Precision-recall graphs obtained using (a) the second repository and (b) the third repository.

Table 1
Precision obtained for the four similarity functions

Distance function	Precision (%)
Composite measure	23.50
Gaussian normalization	17.75
Linear normalization	16.94

a target image at random and then presents three groups of pictures to the user, each containing the closest 16 images according to the composite function, and their sum after applying Gaussian and linear normalization. The user is requested to report how many images on each set are relevant to the query. This data is recorded and averaged, to produce the average precision at a cut-off value of 16.

A total of 10 users participated in the process, each evaluating the performance on 10 different queries. Table 1 shows the results.

As it can be observed, the proposed similarity measure correlates extremely well with the previous results in which subjective similarity was merely simulated by performing a manual classification.

6. Concluding remarks

This paper has presented a technique which allows one to combine a set of distance functions into a composite measure that presents a higher performance than each of the individual functions. The technique uses empirical data and kernel density estimates to convert the basic distance functions into the probability that the user considers that two images are subjectively similar for each possible value the function may produce. The values obtained using these distributions are then combined to yield a new global similarity measure, according to a consistent probabilistic reasoning.

Exhaustive experimentation has been carried out to evaluate the approach. The performance of the new global similarity function has been compared to that of the individual distance measures, and to other normalization approaches, according to precision and recall values in several ways. In these experiments it has been shown that the new composite measure reaches acceptable levels of performance even when using small training sets, and that the algorithm outperforms other widely used combination approaches.

A number of open questions and extensions arise. The present work has been developed and tested using a Single Instance representation (SI) of the pictures (a set of global features extracted from the entire image). Recent works have combined global and regional features for image annotation (Wang et al., 2008; Tang et al., in press). The latter are produced by segmenting the picture and then processing each region independently, resulting in a Multiple Instance representation (MI) of the image. In (Tang et al., in press), the MI representation is converted into a SI representation as part of the process. This same approach could in principle be adopted to allow combining both global and regional features.

Besides, the equalization performed as part of our method could also be integrated with other combination rules which are more commonly used in information retrieval systems. This could lead both to fully understand the relative benefits of equalization and combination and also to improve the proposed measure further. The most obvious but in principle unfeasible extension could consist on stating our proposal in a per query basis, as a first step to incorporate the approach into a user feedback mechanism. To this end possibly parametric techniques as in (Manmatha et al., 2001) and/or some sort of query aggregation and evidence combination would need to be integrated into the approach.

Acknowledgements

This work has been partially funded by UVEG, FEDER and Spanish MEC through projects UV-AE-20070220, TIN2006-10134, TIN2006-12890, DPI2006-15542-C04-04 and Consolider Ingenio 2010 CSD2007-00018.

References

Arevalillo-Herráez, M., Domingo, J., Zacarés, M., 2008. Probabilistic normalization: an approach to normalizing similarity measures in content based image retrieval. In: Proceeding of the Fifth IASTED International Conference on Signal Processing, Pattern Recognition and Applications, Innsbruck, Austria, pp. 30–35.

Ayala, G., Domingo, J., 2001. Spatial size distributions. Applications to shape and texture analysis. IEEE Trans. Pattern Anal. Machine Intell. 23 (12), 1430–1442.

Bruno, E., Kludas, J., Marchand-Maillet, S., 2007. Combining multimodal preferences for multimedia information retrieval. In: MIR'07: Proceedings of the International Workshop on Workshop on Multimedia Information Retrieval, ACM, New York, NY, USA, pp. 71–78.

Chellappa, R., Chatterjee, S., 1985. Classification of textures using gaussian markov random fields. IEEE Trans. Acoust. Speech Signal Process. 33, 959–963.

Chen, Y., Dougherty, E., 1994. Gray-scale morphological granulometric texture classification. Opt. Eng. 33 (8), 2713–2722.

Chuang, G., Kuo, C., 1996. Wavelet descriptor of planar curves: Theory and applications. IEEE Trans. Image Process. 5 (1), 56–70.

Conners, R.W., Trivedi, M.M., Harlow, C.A., 1984. Segmentation of a high-resolution urban scene using texture operators. Comput. Vision Graphics Image Process. 25 (3), 273–310.

Datta, R., Joshi, D., Li, J., Wang, J.Z., 2008. Image retrieval: Ideas, influences, and trends of the new age. ACM Comput. Surveys 40 (2), 1–60.

de Ves, E., Domingo, J., Ayala, G., Zuccarello, P., 2006. A novel bayesian framework for relevance feedback in image content-based retrieval systems. Pattern Recognition 39, 1622–1632.

de Ves, E., Benavent, X., Ayala, G., Domingo, J., 2006. Selecting the structuring element for morphological texture classification. Pattern Anal. Applicat. 9, 48–57.

Do, M.N., Vetterli, M., 2002. Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. IEEE Trans. Image Process. 11 (2), 146–158.

Duda, R.O., Hart, P.E., Stork, D.G., 2000. Pattern Classification, second ed. Wiley-Interscience.

Fernandez, M., Vallet, D., Castells, P., 2006. Probabilistic score normalization for rank aggregation. LNCS/Advances in Information Retrieval, vol. 3936. Springer, Berlin, pp. 553–556.

Giacinto, G., Roli, F., 2004. Nearest-prototype relevance feedback for content based image retrieval. In: ICPR'04: Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04), vol. 2. IEEE Computer Society Washington, DC, USA, pp. 989–992.

Gonzales, R.C., Wintz, P., 1987. Digital Image Processing, second ed. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

Haralick, R., Shanmugam, K., Distein, I., 1973. Textual features for image classification. IEEE Trans. Systems Man Cybernet. 3, 610–621.

Iqbal, Q., Aggarwal, J., 2002. Combining structure, color and texture for image retrieval: A performance evaluation. In: 16th International Conference on Pattern Recognition (ICPR), Quebec City, QC, Canada, pp. 438–443.

Jain, A.K., Nandakumar, K., Ross, A., 2005. Score normalization in multimodal biometric systems. Pattern Recognition 38, 2270–2285.

Kamarainen, J.-K., Kyrki, V., Ilonen, J., KSlviSinen, H., 2003. Improving similarity measures of histograms using smoothing projections. Pattern Recognition Lett. 24 (12), 2009–2019.

Khotanzad, A., Hong, Y.H., 1990. Invariant image recognition by zernike moments. IEEE Trans. Pattern Anal. Machine Intell. 12 (5), 489–497.

Kittler, J., Hatef, M., Duin, R.P.W., Matas, J., 1998. On combining classifiers. IEEE Trans. Pattern Anal. Machine Intell. 20 (3), 226–239.

Knowles, J., Corne, D., 2000. Approximating the nondominated front using the pareto archived evolution strategy. Evolutionary Computat. 8 (2), 149–172.

León, T., Zuccarello, P., Ayala, G., de Ves, E., Domingo, J., 2007. Applying logistic regression to relevance feedback in image retrieval systems. Pattern Recognition 40 (10), 2621–2632.

Lew, M.S., Sebe, N., Djeraba, C., Jain, R., 2006. Content-based multimedia information retrieval: State of the art and challenges. ACM Trans. Multimedia Comput. Commun. Applicat. 2 (1), 1–19.

Li, B., Chang, E.Y., 2003. Discovery of a perceptual distance function for measuring image similarity. ACM Multimedia J. Special Issue Content-Based Image Retrieval 8 (6), 512–522.

Long, F., Zhang, H., Feng, D., 2003. Multimedia Information Retrieval and Management. Technological Fundamentals and Applications. Springer-Verlag, Berlin, Heidelberg, New York. Ch. Fundamentals of content-based image retrieval, pp. 1–26.

Manmatha, R., Rath, T., Feng, F., 2001. Modeling score distributions for combining the outputs of search engines. In: SIGIR'01: Proceedings of the 24th Annual

International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, New York, NY, USA, pp. 267–275.
Müller, H., Müller, W., Squire, D.M., Marchand-Maillet, S., Pun, T., 2001. Performance evaluation in content-based image retrieval: Overview and proposals. Pattern Recognition Lett. 22 (5), 593–601.
Neumann, D., Gegenfurtner, K.R., 2006. Image retrieval and perceptual similarity. ACM Trans. Appl. Perception 3 (1), 31–47.
Nottelmann, H., Fuhr, N., 2003. From retrieval status values to probabilities of relevance for advanced ir applications. Information Retrieval 6 (3–4), 363–388.
Pass, G., Zabih, R., 1996. Histogram refinement for content-based image retrieval, In: IEEE Workshop on Applications of Computer Vision, pp. 96–102.
Prabhakar, S., Jain, A.K., 2002. Decision-level fusion in fingerprint verification. Pattern Recognition 35 (861–874), 861–874.
Rubner, Y., Puzicha, J., Tomasi, C., Buhmann, J.M., 2001. Empirical evaluation of dissimilarity measures for color and texture. Computer Vision Image Understanding 84 (1), 25–43.
Sheather, S., Jones, M., 1991. A reliable data-based bandwidth selection method for kernel density estimation. J. Royal Statistical Soc. B 53, 683–690.
Silverman, B., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.
Smith, G., Burns, I., 1997. Measuring texture classification algorithms. Pattern Recognition Lett. 18 (14), 1495–1501.
Soille, P., 2003. Morphological Image Analysis: Principles and Applications. Springer-Verlag, Berlin.
Swain, M.J., Ballard, D.H., 1991. Color indexing. Int. J. Comput. Vision 7 (1), 11–32.
Tang, J., Li, H., Qi, G.-J., Chua, T.-S., in press. Integrated Graph-based Semi-supervised Multiple/Single Instance Learning Framework for Image Annotation, ACM International Conference on Multimedia. Available from: <http://jhtang.googlepages.com/>.
da S. Torres, R., Falcpo, A.X., Goncalves, M.A., Zhang, B., Fan, W., Fox, E.A., Calado P., 2005. A new framework to combine descriptors for content-based image retrieval. In: Fourteenth Conference on Information and Knowledge Management, Bremen, Germany, pp. 335–336.
Urban, J., Jose, J.M., 2004. Evidence combination for multi-point query learning in content-based image retrieval. In: ISMSE'04: Proceedings of the IEEE Sixth International Symposium on Multimedia Software Engineering, IEEE Computer Society, Washington, DC, USA, pp. 583–586.
Wang, Y., Mei, T., Gong, S., Hua, X.-S., Global, Combining, 2008. Regional and contextual features for automatic image annotation. Pattern Recognition. doi:10.1016/j.patcog.2008.05.010.
Zhang, Q., Izquierdo, E., 2006. Optimizing metrics combining low-level visual descriptors for image annotation and retrieval. In: Proceeding of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 06), Toulouse, France, pp. 405–408.