PERGAMON

Pattern Recognition 34 (2001) 1639–1655

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Content-based similarity retrieval of trademarks using relevance feedback

G. Ciocca, R. Schettini*

Istituto Tecnologie Informatiche Multimediali, Consiglio Nazionale delle Ricerche, Via Ampere 56, 20131 Milano, Italy

Received 11 March 1999; received in revised form 14 March 2000; accepted 14 March 2000

Abstract

This paper addresses the problem of how to efficiently and effectively retrieve images similar to a query from a trademark database purely on the basis of low-level feature analysis. It investigates the hypothesis that the low-level image features used to index the trademark images can be correlated with image contents by applying a relevance feedback mechanism that evaluates the feature distributions of the images the user has judged relevant, or not relevant and dynamically updates both the similarity measure and query in order to better represent the user's particular information needs. Experimental results on a database of 1100 trademarks are reported and commented.

Keywords: Content-based retrieval; Relevance feedback; Image similarity; Trademarks

1. Introduction

The content-based retrieval of trademarks is "*extremely challenging and instructive to study*", due to the high complexity and diversity of the data involved, also often composed of several distinct components [1]. Our study has addressed the problem of how to efficiently and effectively retrieve images similar to a query from a trademark database purely on the basis of low-level feature analysis. As already pointed out by several authors [1,2], perceptually similar images are not necessarily similar in terms of low-level features. We have investigated the hypothesis that the low-level features used to index the images can be correlated with their semantic contents by applying a relevance feedback mechanism.

A few applications designed specifically for the registration of trademarks are available. Wu et al. have developed a prototype system, STAR, using their content-based retrieval engine for multimedia information systems [2–4]. Eakins et al. have developed a prototype system (ARTISAN) for the UK Patent Office Trade Marks Registry, to retrieve trademarks when these consist of abstract geometric designs [5]. Another system, called TRADEMARK, operates on the trademark database of the Patent Office of Japan [6]. A detailed analysis of the problems involved in trademark registration can be found in a recent paper by Jain and Vailaya [1]. These authors propose a computational strategy in which multiple feature description schemes of the same visual cue (shape) are used to improve retrieval accuracy without significantly increasing computational costs. At the first stage of processing (pruning) Jain and Vailaya represent the trademark images in terms of invariant moments and the histogram of the edge directions, integrating the dissimilarity of these features by a weighted mean. A small set of plausible candidates is then presented to a detail matcher based on deformable templates to eliminate false matches. This second phase makes it possible to eliminate the false matches, but cannot cope with trademarks that have not been retrieved in the first stage, although actually perceptually similar to the query. We have attempted to improve the effectiveness of the first stage of retrieval by relevance feedback, i.e. by allowing the user to progressively refine the system's response to

* Corresponding author. Tel.: + 39-2-706-43288; fax: + 39-2-706-43292.

E-mail addresses: ciocca@itim.mi.cnr.it (G. Ciocca), centaura@itim.mi.cnr.it (R. Schettini).

0031-3203/01/$20.00
PII: S0031-3203(00)00055-8

his query. The key concept of the relevance feedback we propose is the statistical analysis of the feature distributions of the retrieved images the user has judged relevant, or not relevant, in order to understand what features he has taken into account (and to what extent) in formulating this judgment, so that we can then accentuate their influence in the overall evaluation of image similarity, as well as in the formulation of a new query iteration.

Section 2 of the paper briefly describes the feature sets used to index the images. Section 3 presents the relevance feedback mechanism implemented. Experimental results are reported in Section 4, followed by our conclusions.

2. Image indexing

The features used for indexing have been selected for three basic properties [7]:

(i) perceptual similarity (the feature distance between two images is large only if the images are not "similar"),
(ii) efficiency (the features can be rapidly computed), and
(iii) economy (small dimensions that do not affect retrieval efficiency).

Loncavic [8] has recently published a review of shape analysis, while Mehthre et al. [9] and Scasselati et al. [10] have reported experimental comparisons of methods of image retrieval based on shape similarity. Jain and Vailaya have experimented different feature sets for trademark indexing, finding that invariant moments and the histogram of the edge directions are the most effective [1]. These features are applied here, together with the mean and variance of the absolute values of the coefficients of the sub-images of the first three levels of the multiresolution wavelet transform of the image. We have used this somewhat redundant image description as none of the features can univocally identify an image: completely different images may yield similar feature values. The very different natures of the indices chosen here should limit the possibility of different trademarks corresponding to very close points in the feature space.

2.1. Moments

Moments, in general, describe numerical quantities at some distance from a reference point or axis [11,12]. Their use in image analysis is straightforward if we consider the image a two-dimensional density distribution function. However, characterizing all the information contained in an image would require an infinite number of moment values. The challenge, therefore, is to select a meaningful subset of moment values that contains sufficient information to describe image appearance. For an image $f(x,y)$ the central moments are given by

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q f(x,y), \tag{1}$$

where $\bar{x}$ and $\bar{y}$ are the coordinates of the center of mass. From the following combinations of second and third moments:

$$\begin{aligned}
M_1 &= (\mu_{20}+\mu_{02}),\\
M_2 &= (\mu_{20}-\mu_{02})^2 + 4\mu_{11}^2,\\
M_3 &= (\mu_{30}-3\mu_{12})^2 + (3\mu_{21}-\mu_{03})^2,\\
M_4 &= (\mu_{30}+\mu_{12})^2 + (\mu_{21}+\mu_{03})^2,\\
M_5 &= (\mu_{30}+\mu_{12})(\mu_{30}-3\mu_{12})[(\mu_{30}+\mu_{12})^2\\
&\quad -3(\mu_{21}+\mu_{03})^2] + (3\mu_{21}-\mu_{03})(\mu_{21}+3\mu_{03})\\
&\quad \times[3(\mu_{03}+\mu_{21})^2 - (\mu_{21}-\mu_{03})^2],\\
M_6 &= (\mu_{20}-\mu_{02})[(\mu_{30}+\mu_{12})^2 - (\mu_{21}+\mu_{03})^2]\\
&\quad + 4\mu_{11}(\mu_{30}+\mu_{12})(\mu_{21}+\mu_{03}),\\
M_7 &= (3\mu_{21}-\mu_{03})(\mu_{30}+\mu_{12})[(\mu_{30}+\mu_{12})^2\\
&\quad -3(\mu_{21}+\mu_{03})^2] + (\mu_{30}-3\mu_{12})(\mu_{21}+\mu_{03})\\
&\quad \times[3(\mu_{03}+\mu_{21})^2 - (\mu_{21}-\mu_{03})^2],
\end{aligned} \tag{2}$$

a set of invariant moments which have the useful properties of being invariant to the object's scale, rotation, and position has been derived [11]:

$$\begin{aligned}
&M'_1 = M_1/\mu_{00}, \quad M'_2 = M_2/r^4, \quad M'_3 = M_3/r^6,\\
&M'_4 = M_4/r^6, \quad M'_5 = M_5/r^{12}, \quad M'_6 = M_6/r^8,\\
&M'_7 = M_7/r^{12},
\end{aligned} \tag{3}$$

where $r = (\mu_{20}+\mu_{02})^{1/2}$ is the radius of gyration of the object.

Moments, however, do not suffice to completely describe the perceptual appearance of trademarks. The two trademarks shown in Fig. 1, for example, are actually similar from a perceptual point of view, although they have quite different moment values.

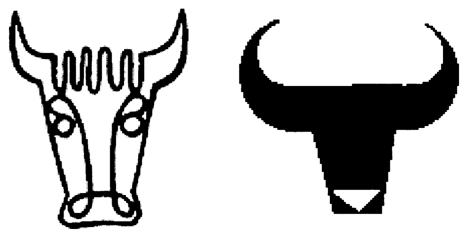


Fig. 1. Two similar trademarks having very different descriptions in terms of invariant moments.

2.2. Edge directions

A histogram of the edge directions is also used to describe the trademarks' shape [13]. Edges are extracted by Canny's edge detectors [14], and the corresponding edge directions are quantized in 72 bins at 2.5° intervals. To compensate for different image sizes, histograms are normalized with respect to the total number of edge pixels detected in the image. The histograms of the edge directions of the two trademarks shown in Fig. 2 are quite similar.

2.3. Wavelets

Multiresolution wavelet analysis provides representations of image data in which both spatial and frequency information are present. It has recently been used in content-based retrieval for similarity retrieval and target search (e.g. [15]).

In multiresolution wavelet analysis we have four bands for each level of resolution (see Fig. 3): a low-pass filtered version of the processed image, and three bands of details. Each band corresponds to a coefficient matrix one-fourth the size of the processed image. In our procedure the features are extracted using a three-step Daubechies multiresolution wavelet expansion producing 10 subbands [16]. Two energy features, the mean and variance of the coefficient's absolute values, are then computed for each subband. These features provide a concise description of the trademark's texture and shape.

Fig. 2. Two similar trademarks having very similar descriptions in terms of the histogram of the edge directions.

3. Our relevance feedback mechanism

Relevance feedback has been widely studied in textual information retrieval. In image retrieval, it has been exploited for target search by Minka and Picard [17] and for similarity retrieval by Cox et al. [18], and by Rui et al. [19,20] and Sclaroff et al. [21]. In Ciocca and Schettini [22] we designed a new algorithm that through the statistical analysis of the feature distributions of the retrieved images the user has judged relevant, or not relevant, identifies what features (and to what extent) the user has taken into account in formulating his judgement, and then updates the weights of the different features in the overall evaluation of image similarity, as well as in the reformulation of a new query, accordingly.

Applying this algorithm to a database containing n images indexed as described in Section 2, the corresponding vectors of features are indicated by $\mathbf{X}_h^i$, where i is the image index, and h the index of the feature. $\mathbf{X}_h^i(j)$ indicates the jth value of the sub-vector $\mathbf{X}_h^i$, and D_h is the distance associated with the feature hth.

The global metric used to evaluate the dissimilarity between two images i and j of the database is, usually, a linear combination of the distances between the individual features

$$\text{Dissimilarity}\,(\mathbf{X}^i, \mathbf{X}^j) = \sum_{h=1}^{p} w_h D_h(\mathbf{X}_h^i, \mathbf{X}_h^j) \tag{4}$$

in which the w_h are weights. There are two problems in this formulation of image similarity. First, since the single distances may be defined on intervals of widely varying values, they must be normalized to a common interval so that equal emphasis is placed on every feature score. Second, the weights must often be set heuristically by the user, and this may be rather difficult, as there may be no clear relationship between the features used to index the

(a) (b) (c)

Fig. 3. (a) original image; (b) two-step multiresolution wavelet transform; and (c) three-step multiresolution wavelet transform.

image database and those evaluated by the user in a subjective image similarity evaluation. Moreover, image similarity is user- and task-dependent [17,23], and this dependence is not in general understood well enough to permit careful, a priori selection of the optimal measure.

To cope with the problem of distances defined on different intervals of values, we proceed as described here below.

3.1. Distance normalization

The average distance between features of database items, assuming that the database contains n images, is computed as follows:

$$\mu_h = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} D_h(\mathbf{X}_h^i, \mathbf{X}_h^j). \tag{5}$$

In our system all the features are compared with the city block distance measure (L_1) by default, as it is statistically more robust than the Euclidean distance measure (L_2) [24]. Other distance measures could be employed without requiring any change in the algorithm.

The vector of the normalized distance between two images having indices i and j, respectively, is

$$\tilde{\mathbf{D}}(\mathbf{X}^i, \mathbf{X}^j) = \left[\frac{D_1(\mathbf{X}_1^i, \mathbf{X}_1^j)}{\mu_1}, \ldots, \frac{D_h(\mathbf{X}_h^i, \mathbf{X}_h^j)}{\mu_h}, \ldots, \frac{D_p(\mathbf{X}_p^i, \mathbf{X}_p^j)}{\mu_p}\right]^{\mathrm{T}}. \tag{6}$$

The advantage of this type of normalization is that, if the database is large enough, the averages are computed only once, when the database is indexed and, it is not necessary to recompute their values when new items are added. Moreover, the computational cost is low [21]. This normalization, however, does not guarantee that all feature distances will be defined on a common interval, but simply that half of the values will lie within the range of [0,1], and the other half within the range of $[1, x]$, where x is a function of the maximum value of the set. Another possibility would be to normalize the distances on the basis of the smallest ($\min_h$) and biggest ($\max_h$) values among the $n(n-1)/2$ possible image pairs, as follows:

$$\tilde{\mathbf{D}}(\mathbf{X}^i, \mathbf{X}^j) = \left[\frac{D_1(\mathbf{X}_1^i, \mathbf{X}_1^j) - \min_1}{\max_1 - \min_1}, \ldots, \frac{D_h(\mathbf{X}_h^i, \mathbf{X}_h^j) - \min_h}{\max_s - \min_h}, \ldots, \frac{D_p(\mathbf{X}_p^i, \mathbf{X}_p^j) - \min_p}{\max_p - \min_p}\right]^{\mathrm{T}}. \tag{7}$$

This approach, however, may compress the feature distance values into a very small range if even a single abnormally large distance is present [21].

The approach proposed by Ortega [20] to overcome this drawback applies Gaussian normalization as follows:

$$\tilde{\mathbf{D}}(\mathbf{X}^i, \mathbf{X}^j) = \left[\frac{D_1(\mathbf{X}_1^i, \mathbf{X}_1^j) - \mu_1}{K\sigma_1}, \ldots, \frac{D_h(\mathbf{X}_h^i, \mathbf{X}_h^j) - \mu_h}{K\sigma_h}, \ldots, \frac{D_p(\mathbf{X}_p^i, \mathbf{X}_p^j) - \mu_p}{K\sigma_p}\right]^{\mathrm{T}}, \tag{8}$$

where the mean values are computed as above (Eq. (5)) and the standard deviation is computed as follows

$$\sigma_h = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} [D_h(\mathbf{X}_h^i, \mathbf{X}_h^j) - \mu_h]^2. \tag{9}$$

Assuming that the features' distance distributions have a Gaussian distribution, it can be shown that there is a 68% probability that the feature values will lie within the range of $[-1, \ 1]$ if $K = 1$, and a 99% probability if $K = 3$ [25]. Unfortunately, in content-based retrieval applications we cannot assume a priori that the distances will have Gaussian distributions. The following general relationship holds without requiring any assumption about feature distributions:

$$P\left[-1 \leqslant \frac{\tilde{D}_h - \mu_h}{K\sigma_h} \leqslant 1\right] \geqslant 1 - \frac{1}{K^2}. \tag{10}$$

According to this relationship, there is an 89% probability that the distance will fall within the range of $[-1, 1]$ if we set K at 3, and a 94% probability when k is set at 4 [25]. Therefore, we apply Eq. (8) with k set at 4 for distance normalization; a simple additional shift moves the normalized distances into the [0, 1] range:

$$d_h = \frac{\tilde{D}_h + 1}{2}. \tag{11}$$

Out-of-range values are mapped to the extreme values, without bias to further processing.

At this point our similarity function has the following form:

$$\text{Dissimilarity}(\mathbf{X}^i, \mathbf{X}^j) = \sum_{h=1}^{p} w_h \, d_h(\mathbf{X}_h^i, \mathbf{X}_h^j). \tag{12}$$

3.2. Distance weights

The algorithm must now determine the weights for the individual distances by a statistical analysis of the feature distances of the images the user has judged "relevant" or "not-relevant" in order to model his information needs [22].

We have observed that users do not find it difficult to provide examples of similar and dissimilar images interactively. However, since the image database queried is heterogeneous and the retrieval task particularly complex, users may not find enough examples of images that are truly similar to the query in the first screens, and, to avoid the time-consuming visual browsing of the database, may mark as relevant images that are only partially similar. Users' information needs may also be rather vague, such as: find all the images containing animals. In both cases the images judged relevant may differ widely. Consequently, treating all these images in the same way — for example, averaging the features of the relevant images to update the similarity measure — may produce very poor results, while processing all the relevant images as single queries, and then combining the retrieval outputs may create an unacceptable computational burden when the database is large. To cope with these problems we proceed as follows.

Let $\mathbf{R}^+$ be the set of relevant images selected by the user ($\mathbf{R}^+$ is usually only an approximation of the set of images relevant to the query in the whole database); $\mathbf{d}_h^+$, the set of normalized distances (computed on feature h) among the elements of $\mathbf{R}^+$; and μ_h^+, the mean of the values of $\mathbf{d}_h^+$ computed according to Eq. (5). Similarly, we define $\mathbf{R}^-$ as the set of non relevant images selected by the user as negative examples, and $\mathbf{d}_h^-$ the corresponding set of distances. We then use $\mathbf{R}^+$ and $\mathbf{R}^-$ to determine whether the influence of a feature must be reduced in the computation of the dissimilarity by reducing the corresponding weight: let $\mathbf{R}^\pm$ be the union of $\mathbf{R}^+$ with $\mathbf{R}^-$, and $\mathbf{d}_h^\pm$ the corresponding set of distances among its elements. Since we can not make any assumptions about the statistical distribution of the features of non-relevant images by analyzing $\mathbf{R}^-$ (the non-relevant images selected may be not representative of all the non-relevant images in the database), we exclude set $\mathbf{d}_h^-$ from $\mathbf{d}_h^\pm$, obtaining a new set of distances: $\mathbf{d}_h^* = \mathbf{d}_h^\pm \backslash \mathbf{d}_h^-$. Letting μ_h^* be the mean of the elements in $\mathbf{d}_h^*$, we can now determine the weights to use in Eq. (15) as follows:

$$w_h^+ = \begin{cases} \dfrac{1}{\varepsilon} & \text{if } |\mathbf{R}^+| < 3, \\ \dfrac{1}{\varepsilon + \mu_h^+} & \text{otherwise,} \end{cases} \tag{13}$$

$$w_h^* = \begin{cases} 0 & \text{if } |\mathbf{R}^+| + |\mathbf{R}^-| < 3 \quad \text{or} \quad |\mathbf{R}^-| = 0 \text{ or } |\mathbf{R}^+| = 0, \\ \dfrac{1}{\varepsilon + \mu_h^*} & \text{otherwise,} \end{cases} \tag{14}$$

$$w_h = \begin{cases} 0 & \text{if } w_h^+ < w_h^*, \\ w_h^+ - w_h^* & \text{otherwise,} \end{cases} \tag{15}$$

where ε is a positive constant, set at 0.01 in our experiments.

Looking at these formulas, we observe that

- At least three examples (relevant or non-relevant images) are required for updating weights, otherwise the weights' values are all set at $1/\varepsilon$.
- If the user selects only relevant images, the weights are computed according to Eq. (13). For any given feature, w_h^+ is high when there is some form of agreement among the feature values of the selected images. We have already seen that treating all the relevant images selected in the same way may produce very poor results when these images resemble the query image only in some pictorial features, but are actually quite different from each other.
- For any given feature w_h^*, of Eq. (14), is high when there is some form of agreement among the feature values of positive and negative examples. This should mean that the feature is not discriminant for the query; consequently its corresponding weight is decreased (Eq. (15)).

3.3. Query reformulation

When more than one image is selected by the user as relevant, the query processing algorithm can compute a new query vector that better represents the user's information needs. One way of doing this is to take a weighted average of the query feature vector and of the relevant images [26]. But in this case the algorithm cannot provide for the fact that relevant images may differ from the original query with respect to some features. Our approach is similar to that adopted for the estimation of the weights: we let $\mathbf{R}^+$ be the set of relevant images the user has selected (including the original query), indicating with $\bar{\mathbf{Q}}$ the average query and with $\bar{\sigma}$ the vector of the standard deviations:

$$\bar{\mathbf{Q}} = \frac{1}{|\mathbf{R}^+|} \sum_{\mathbf{X}^i \in \mathbf{R}^+} \mathbf{X}^i \; \bar{\boldsymbol{\sigma}} = \sqrt{\frac{1}{|\mathbf{R}^+|} \sum_{\mathbf{X}^i \in \mathbf{R}^+} (\mathbf{X}^i - \bar{\mathbf{Q}})^2}. \tag{16}$$

We then compute a new query $\tilde{\mathbf{Q}}$ as follows. We let

$$Y_h(j) = \{X_h^i(j) \mid |X_h^i(j) - \bar{Q}_h(j)| \leqslant 3\bar{\sigma}_h(j)\} \quad \forall h,\, j \text{ and } \mathbf{X}^i \in \mathbf{R}^+ \tag{17}$$

obtaining

$$\tilde{Q}_h(j) = \frac{1}{|Y_h(j)|} \sum_{X_h^i(j) \in Y_h(j)} X_h^i(j) \tag{18}$$

that is, the query processing formulates a new query $\tilde{\mathbf{Q}}$ that incorporates the new information supplied by user feedback without allowing single instances of a difference in feature values to bias query computation.

4. Experimental results

The retrieval process is graphically portrayed in Fig. 4. When a query is submitted, the system reranks the database images by decreasing similarity with respect to the query, and then returns to the user, displaying

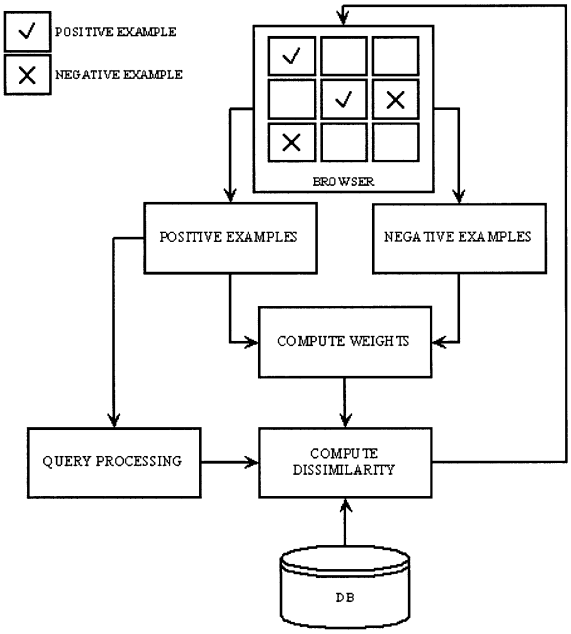


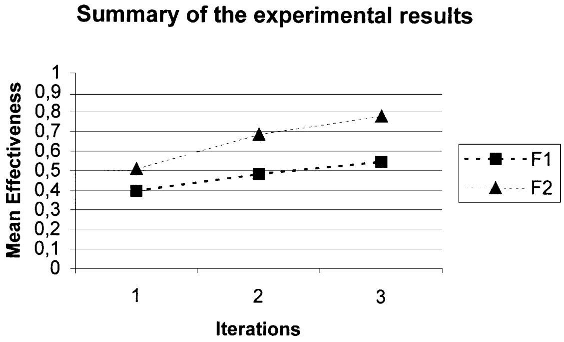
Fig. 4. Diagram of the retrieval process.

the 24 most similar images. In subsequent iterations the user may again mark some of these retrieved images as relevant, or not relevant. A new query vector is then computed, taking into account the features of the newly indicated relevant images, and the overall evaluation of the dissimilarity function is updated, on the basis of the features of both relevant and non-relevant images. A new query is then submitted, starting a new iteration of retrieval. There is no limit to the number of images that can be selected as relevant or non-relevant, nor to the number of relevance feedback iterations. Each retrieval iteration takes about 5 s on a Pentium Pro 200 MHz.

The evaluation of performance for text retrieval has been extensively studied [27], and some of the methods proposed can be adapted to image content-based retrieval [28]. In order to quantify the improvement in performance obtained by applying the relevance feedback mechanism a measure called effectiveness (efficiency of retrieval, or fill ratio), was applied here. This measure, proposed by Methre et al. [29], has also been applied recently to compare the performance of shape similarity measures [9,30] and color similarity measures [31] in content-based image retrieval.

Fig. 5. Typical trademarks in the database.

Table 1
Summary of experimental results



	Feature set F1			Feature set F2		
	First iteration	Second iteration	Third iteration	First iteration	Second iteration	Third iteration
Min eff.	0.2	0.25	0.267	0.25	0.466	0.533
Mean eff.	0.398	0.487	0.554	0.511	0.689	0.785
Max eff.	0.708	0.75	0.83	0.75	1	1

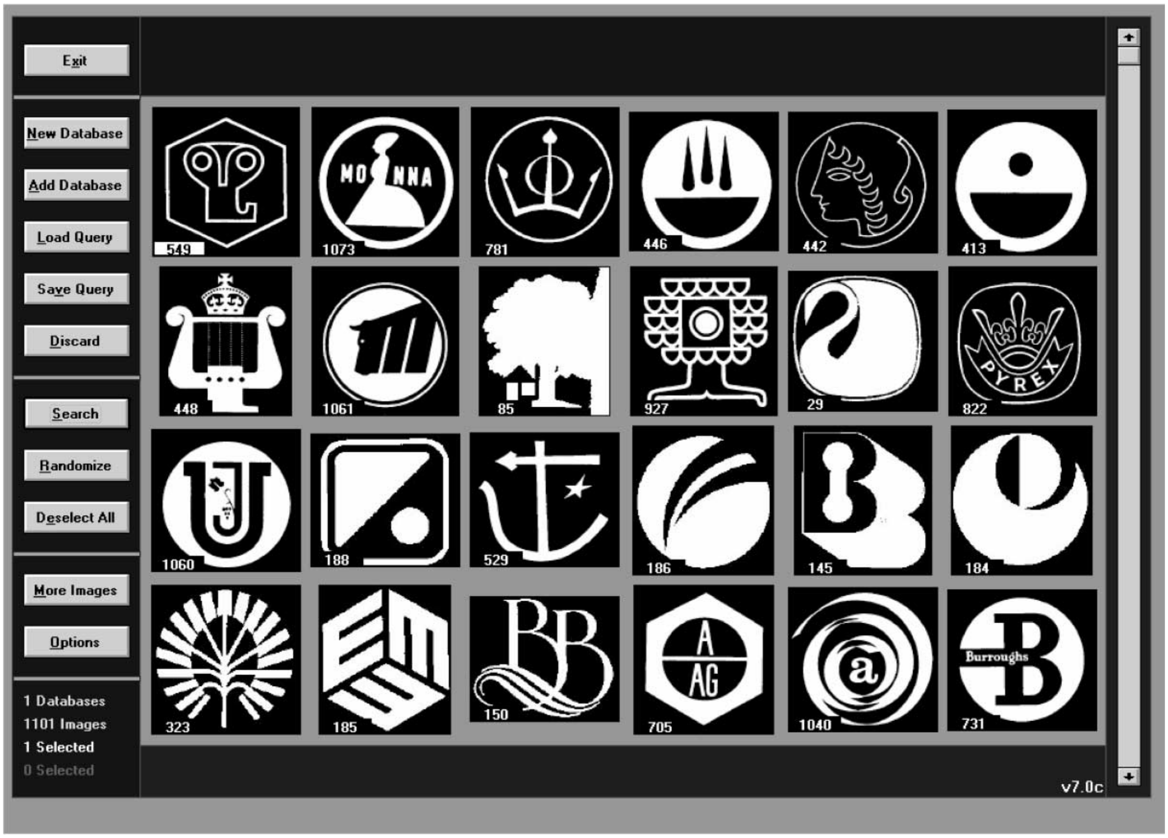


Fig. 6. Retrieval results using moments and histograms of directions as the indexes (F1).

Fig. 7. (a) Initial retrieval results using moments, histograms of directions, and wavelets as the indexes (F2); (b) retrieval results after the first iteration of relevance feedback (F2); and (c) retrieval results after the second iteration of relevance feedback (F2).


Exit
New Database
Add Database
Load Query
Save Query
Discard
Search
Randomize
Deselect All
More Images
Options
1 Databases
1101 Images
1 Selected
0 Selected
549
781
822
827
778
892
442
545
692
1073
150
927
1063
1008
1061
547
698
640
1059
448
693
1019
1040
646
v7.0c

(a)


Exit
New Database
Add Database
Load Query
Save Query
Discard
Search
Randomize
Deselect All
More Images
Options
1 Databases
1101 Images
1 Selected
2 Selected
549
781
827
545
692
892
778
547
1063
822
640
1008
1019
150
698
927
442
690
961
814
646
825
759
434
v7.0c

(b)

(c)

Fig. 7. (*Continued*).

We considered a relevant image — that is an image truly similar to the query — correctly retrieved if it appeared within the first set of displayed images (short list). We let S be the number of images retrieved in the short list when posing a query; $\mathbf{R}_q^I$, the set of relevant images in the database; and $\mathbf{R}_q^E$, the set of images retrieved in the short list (considered "relevant" by the system). The effectiveness measure is defined as

$$\eta_S = \begin{cases} \dfrac{|\mathbf{R}_q^I \cap \mathbf{R}_q^E|}{|\mathbf{R}_q^I|} & \text{if } |\mathbf{R}_q^I| \leqslant S, \\ \dfrac{|\mathbf{R}_q^I \cap \mathbf{R}_q^E|}{|\mathbf{R}_q^E|} & \text{if } |\mathbf{R}_q^I| > S. \end{cases} \tag{19}$$

If $|\mathbf{R}_q^I| \leqslant S$, the effectiveness was reduced to the traditional recall measure, while if $|\mathbf{R}_q^I| > S$, the effectiveness corresponded to precision (In our implementation S was set at 24).

The database we used is the Jain and Vailaya trademark database [1] of 1100 binary images of trademarks acquired from several collections [32–34]. The images are 200×200 pixels large. The database contains a wide range of objects depicting animals, humans, the Sun, the Earth, letters of the alphabet, abstract symbols, etc. Several trademarks are composed of many distinct elements combined to form more complex shapes. Many of the symbols have a similar perceptual meaning, but hardly match in appearance, rendering the retrieval problem even more challenging. The query set used in evaluating the system's performance is shown in Fig. 5. It includes first the three queries considered as given by Jain and Vailaya, together with seven queries selected by the authors as representative of the range of the database contents, and 10 more randomly chosen. In Table 1 we have summarized the experimental results — the ground truth similarity was assessed jointly by the paper's authors — for all the queries, using two different sets of features.

The first three columns correspond to the first three retrieval iterations, using only moments and histograms of directions (F1) as the indexes; the last three columns regard moments, histograms of directions, and wavelets (F2). The first retrieval iteration always corresponds to a similarity measure in which all the features have the same importance (the weights in Eq. (9) are all set at $1/\varepsilon$); the following iterations correspond to queries in which the user has marked at least three of the retrieved trademarks as relevant, or not relevant. In Table 1, for sake of completeness, the minimum, average and maximum effectiveness value at each of the first three retrieval iterations are also reported.

Relevance feedback improves the effectiveness of the retrieval for both feature sets; in general, the second iteration (that is the first relevance feedback iteration) corresponds to the largest single improvement. We observed, to the contrary, little benefit in repeating the procedure more than five, or six times. It can be reasonably argued that this is not due to any deficiency in the mechanism itself but to the low-level features used which can not exhaustively describe the image content. Some examples may better explain how the system works. A first example (Figs. 6 and 7) concerns the retrieval of trademarks composed of thin lines surrounding a symbol or text. In this case only the overall appearance of the image is considered; the image semantics was not taken into account. In a second example the objective was to retrieve all the trademarks depicting a bear. Here the relevance feedback was used to model the image semantics and the retrieval task is much more difficult. The shape of the bears differs greatly, although they are all depicted as a white object against a black background. The last example, retrieving all the trademarks depicting a swan, is the most complex. There are eight swans in the database (see Fig. 11), although we have no difficulty in relating all these trademarks to the concept of swan, images are very different in terms of shape.

Fig. 6 shows the retrieval results with a query image composed by line drawings (image 549), and using the F1 feature set as the trademarks' index. This did not produce good results, due to the limited capacity of the moments and histogram of directions alone to describe a trademark composed of very thin lines. With the F2 feature set the results were much better as many images with line drawings are retrieved (Fig. 7a). These results were further improved by relevance feedback: selecting two trademarks composed of non-thin structures (1073 and 1061) as non-relevant images gave the results shown in Fig. 7b.

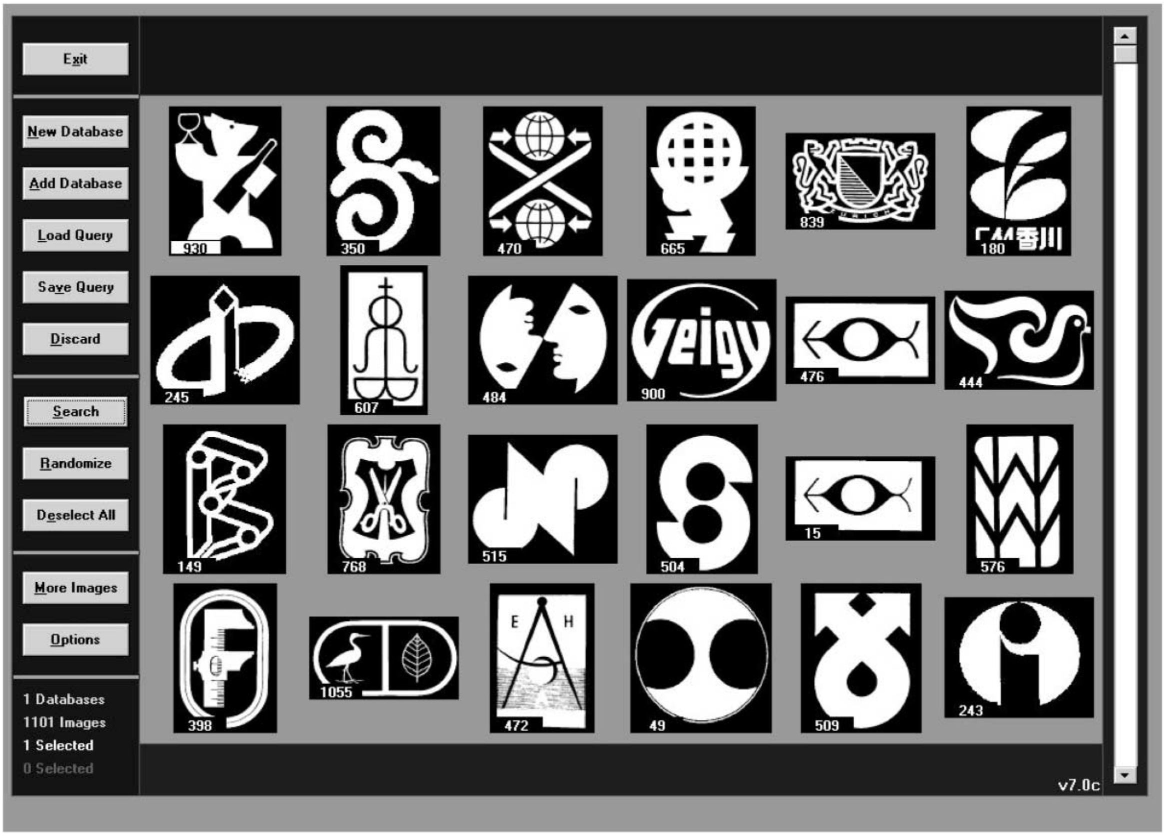
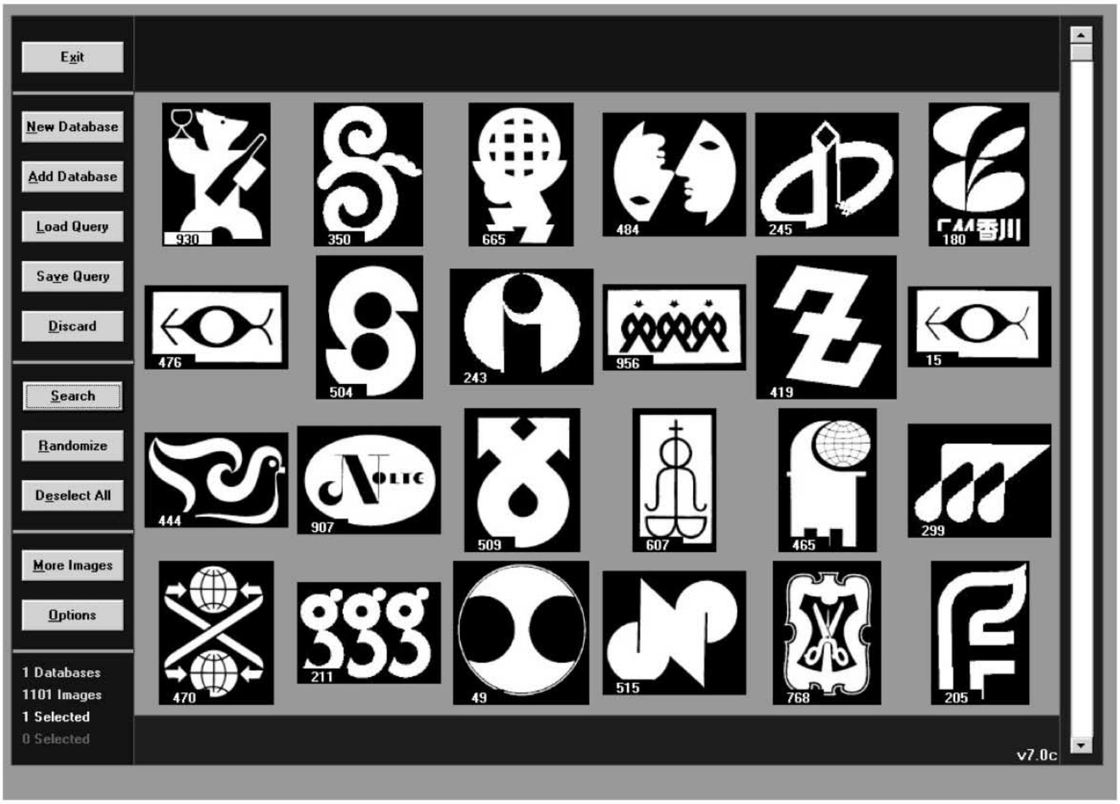
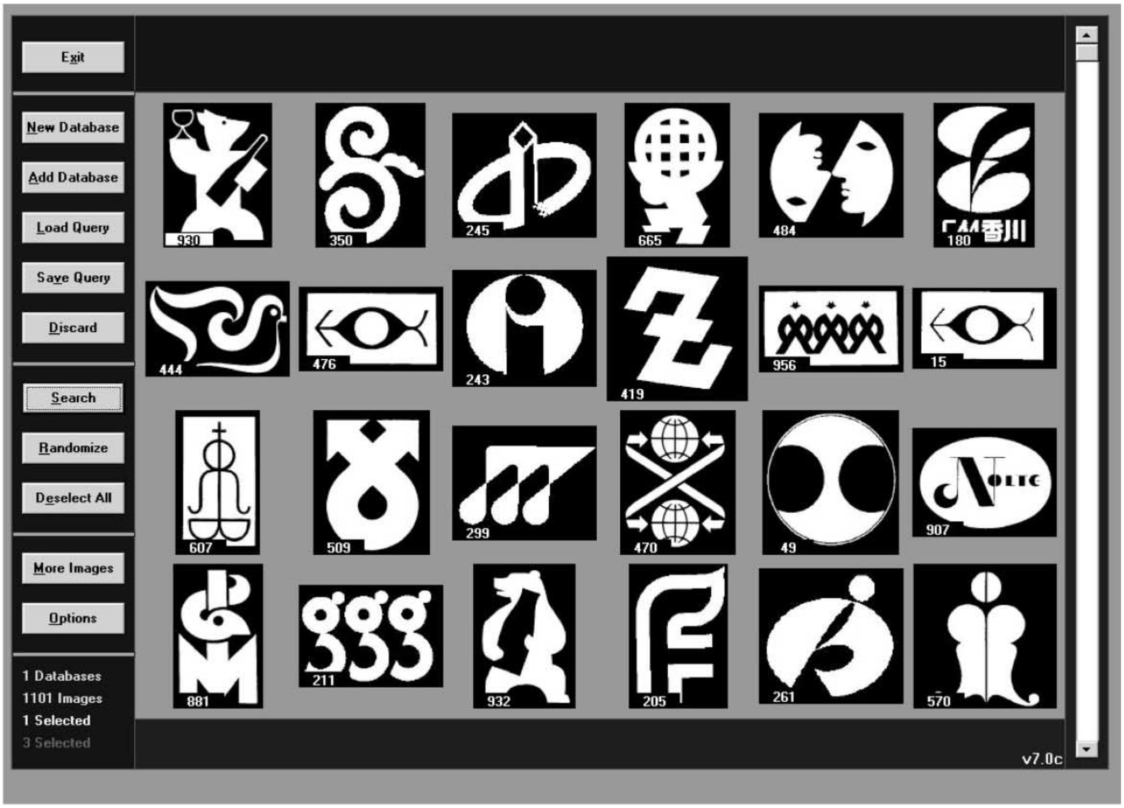


Fig. 8. Retrieval results using moments and histograms of directions as the indexes (F1).

Fig. 9. (a) Initial retrieval results using moments, histograms of directions, and wavelets as the indexes (F2); (b) retrieval results after the first iteration of relevance feedback; and (c) retrieval results after the second iteration of relevance feedback.


Exit
New Database
Add Database
Load Query
Save Query
Discard
Search
Randomize
Deselect All
More Images
Options
1 Databases
1101 Images
1 Selected
0 Selected
930 350 665 484 245 180
476 504 243 956 419 15
444 907 509 607 465 299
470 211 49 515 768 205
v7.0c

(a)


Exit
New Database
Add Database
Load Query
Save Query
Discard
Search
Randomize
Deselect All
More Images
Options
1 Databases
1101 Images
1 Selected
3 Selected
930 350 245 665 484 180
444 476 243 419 956 15
607 509 299 470 49 907
881 211 932 205 261 570
v7.0c

(b)

(c)

Fig. 9. (*Continued*).

And adding 547 and 1019 to the set of relevant images, together with 961 as a not relevant image, we obtained the results shown in Fig. 7c.

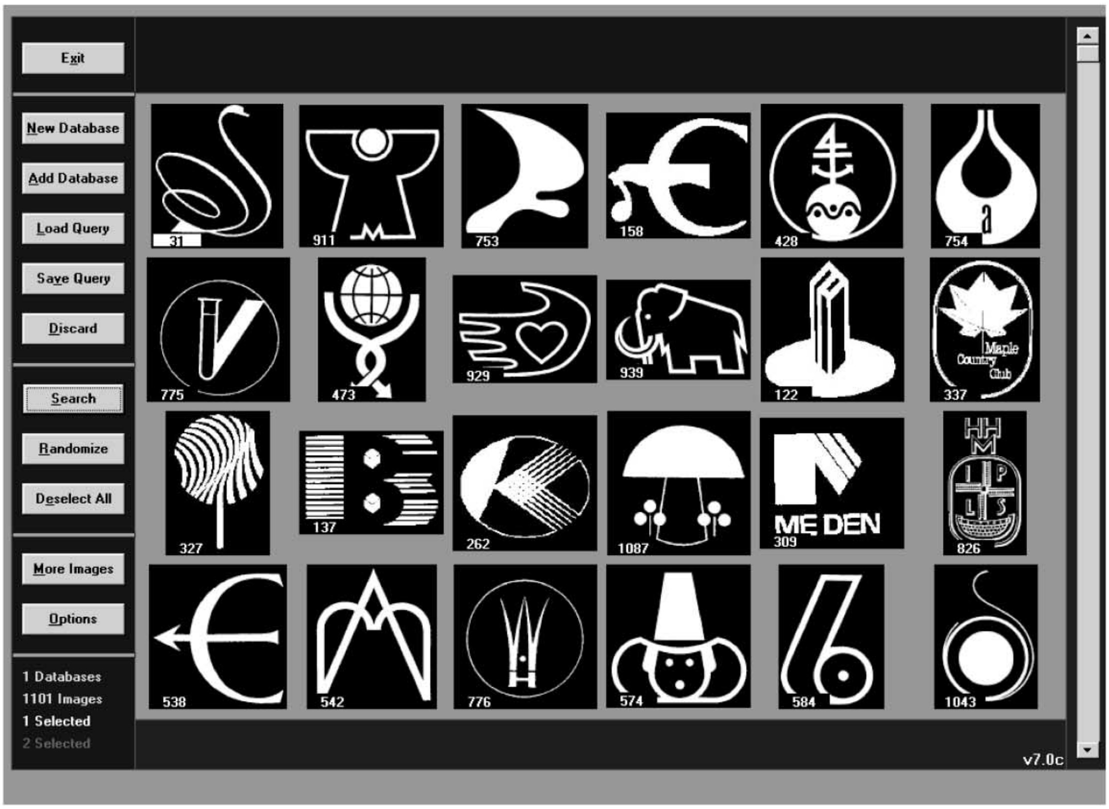
In Fig. 8 another, more complex, example is given. Here we wanted to retrieve all three bears in the database. Fig. 8 shows the results using the F1 feature set. Using F2 does not actually improve on the results (Fig. 9a); however selecting images 515, 504, and 465 as non-relevant, and then querying the system again we obtain the result shown in Fig. 9b. Selecting image 942 as relevant, and images 211, 956, and 607as not relevant, we obtained the optimal result, retrieval all the bears in the database. These images are only perceptually similar, but present in very different shapes: the relevance feedback, that is, user's interaction with the system, has made it possible to cope with this problem.

Fig. 10a depicts the initial retrieval results (using the F2 feature set) of a query using the swan image number 31: no relevant images were retrieved from the database. When images 240 and 704 are selected as non-relevant, and the system queried again, image 1043, depicting an object that resembles a stylized swan is retrieved (Fig. 10b). Adding this image to the set of relevant images allowed the retrieval of image 30 as well (Fig. 10c). By adding image 30 to the set of relevant images we retrieved other three swans (Fig. 10d, images 409, 897 and 32). Further iterations did not make it possible to retrieve all the relevant images (compare the results shown in Fig. 11). Similar results (no all relevant image retrieved) were obtained when we used the other swans depicted in Fig. 11 as initial query images. In all cases, however, relevance feedback improved retrieval results. More sophisticated image indexing strategies would probably have increased the system's performance. Jain and Vailaya [1] reported that their method is more robust for line drawings when the features are computed on filled in images. We plan to enlarge the feature set used for indexing, not because we believe that an enlarged set of visual features could capture the whole meaning of images, but simply that it could be better correlated with

Fig. 10. (a) Initial retrieval results using moments, histograms of directions, and wavelets as the indexes (F2); (b) retrieval results after the first iteration of relevance feedback; (c) retrieval results after the second iteration of relevance feedback; and (d) retrieval results after the third iteration of relevance feedback.


Exit
New Database
Add Database
Load Query
Save Query
Discard
Search
Randomize
Deselect All
More Images
Options
1 Databases
1101 Images
1 Selected
0 Selected
31
911
158
753
428
929
775
754
473
939
538
826
148
122
337
704
137
262
1087
574
776
925
240
327
v7.0c

(a)


Exit
New Database
Add Database
Load Query
Save Query
Discard
Search
Randomize
Deselect All
More Images
Options
1 Databases
1101 Images
1 Selected
2 Selected
31
911
753
158
428
754
775
473
929
939
122
337
327
137
262
1087
309
826
538
542
776
574
584
1043
v7.0c

(b)

(c)

(d)

Fig. 10. (*Continued*).

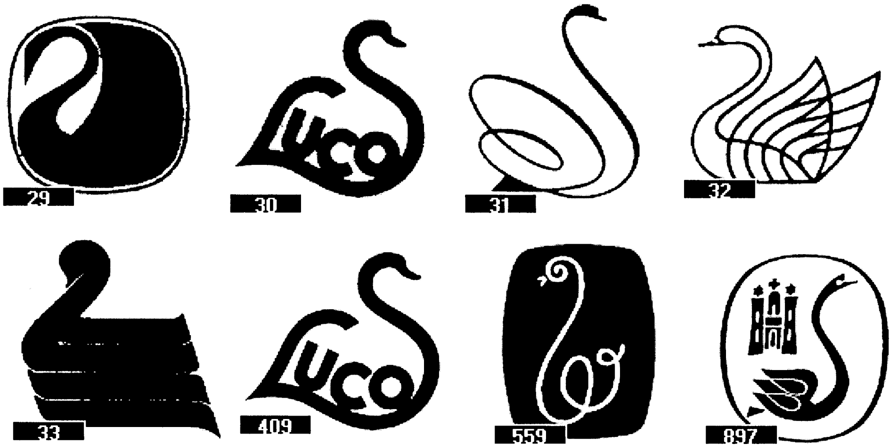


Fig. 11. Database trademarks depicting a swan.

the images' semantic contents by the relevance feedback mechanism.

5. Conclusions

We have addressed the problem of retrieving perceptually "similar" images in a database of trademarks containing a great variety of objects. The performance of an image retrieval system is closely related to the nature and quality of the features used to represent image content, but it is also strongly influenced by the measure adopted to quantify image similarity. We have shown that the use of relevance feedback greatly improves retrieval results, making it possible, in many cases, and with no significant effort by the user, to tune the similarity measure used by the system to the user's notion of image similarity.

Designing the whole system we have assumed that the user will identify at least one relevant example in the database by random browsing and, that some relevant images will be retrieved at the first iteration (in which all the features have the same importance) within the first set of displayed images. In our experiments we found that these conditions were almost always met for the database used, but we realized that this might not be the case if we scale up the size of the database.

Finding an initial query image will be further addressed in our system by allowing the user to perform the query with a sketch, or using an image from outside the database, and by offering a database preview not only by random access, but also by image clustering. The lack of relevant images retrieved at the first iteration, could be addressed pragmatically by allowing the user to select relevant and not-relevant images not only within the first 24 retrieved images (we found that in many cases relevant images were ranked within the first two of three screens). For faster tuning of the similarity function, which would also deal in part with this problem, we could also exploit previous query sessions performed by the user on the same database. The user would be allowed to register satisfactory queries, together with the corresponding weights in the similarity measure. When the user has already formulated a query "similar" to the new one, the algorithm could set the initial weights of the similarity function at the value of the earlier query, reducing the time and effort needed to adapt the similarity measure by applying the relevance feedback algorithm.

Trademarks are often also colored [13,30]. We are now organizing a project to determine what color features could be applied to index the images. The integration of multiple and heterogeneous features will be straightforward for us, as the structure of the relevance feedback mechanism is actually description-independent, that is, the index can be modified, or extended to include other features without requiring any change in the algorithm. More demanding will be the integration of text-based image annotation to further increase retrieval effectiveness.

6. Summary

We addressed here the problem of how to efficiently and effectively retrieve images similar to a query from a trademark database purely on the basis of low-level feature analysis. As already pointed out by several authors, perceptually similar images are not necessarily similar in terms of low-level features. We have investigated the hypothesis that the low-level image features, used to index the trademark images, can be well correlated with image contents by applying a relevance feedback mechanism that evaluates the feature distributions

of the images judged relevant, or not relevant, by the user, and dynamically updates both the similarity measure and query in order to better represent the user's particular information needs. The features used to index an image are the invariant moments, the histogram of the edge directions, and the mean and variance of the absolute values of the coefficients of the sub-images of the first three levels of the multi-resolution wavelet transform of the image. We have used this somewhat redundant image description as none of the features can univocally identify an image. The key concept of the relevance feedback we propose is the statistical analysis of the feature distributions of the images the user has judged relevant, or not relevant, in order to understand what features the user has considered (and to what extent) in formulating this judgment, so that we can then accentuate the influence of these features in the overall evaluation of image similarity, as well as in the formulation of a new query. Experimental results on a database of 1100 trademarks confirm the feasibility of this approach.

Acknowledgements

The image database used here for testing has been kindly provided by Dr. Vailaya.

References

[1] A.K. Jain, A. Vailaya, Shape-based retrieval: a case study with trademarks databases, Pattern Recognition 31 (1998) 1369–1390.

[2] J.K. Wu, C.P. Lam, B.M. Mehtre, Y.J. Gao, A. Desai Narasimhalu, Content-based retrieval for trademark registration, Multimedia Tools Appl. 3 (1996) 245–2567.

[3] J.K. Wu, A. Desai Narasimhalu, B.M. Mehtre, C.P. Lam, Y.J. Gao, CORE: a content-based retrieval engine for multimedia information systems, Multimedia Systems 3 (1995) 25–41.

[4] http://cmchome.cmcltd.com/cmc/ip/trims/trims.htm.

[5] J.P. Eakins, J.M. Boardman, K. Shields, Retrieval of trademark images by shape features - the Artisan Project, Proceedings of IEE Colloquium on Intelligent Image Database http://vision.hanyang.ac.kr/ ~ yskim/rpaper.html, 1996.

[6] T. Kato, T. Kurita, H. Shimogaki, M. Hoga, Z.M. Salim, T. Sota, Trademark system, http://www.aist.go.jp/ETL/etl/taiwa/HumanMedia/HM-ETL/trademark/, 1999.

[7] Del Bimbo, P. Pala, Visual image retrieval by elastic matching of user sketches, IEEE Trans, Pattern Anal. Mach. Intell. 19 (1997) 121–132.

[8] S. Loncaric, A survey on shape analysis techniques, Pattern Recognition 3 (1) (1998) 983–1001.

[9] B.M. Mehtre, M.S. Kankanhalli, W.F. Lee, Shape measures for content based image retrieval: a comparison, Inform. Process. Manage. 33 (1997) 319–337.

[10] B. Scasselati, S. Alexopoulos, M. Flickner, Retrieving images 2D shape: a comparison of computational methods with human perceptual judgments. SPIE Conference on Storage and Retrieval for Images and Videos, 1994.

[11] M. Hu, Visual pattern recognition by moment invariants, IRE Trans. Inform. Theory 8 (1962) 179–187.

[12] M. Belkasim, M. Shridhar, M. Ahmadi, Pattern recognition with moment invariants: a comparative study and new results, Pattern Recognition 24 (1991) 1117–1138.

[13] A.K. Jain, A. Vailaya, Image retrieval using color and shape, Pattern Recognition 29 (8) (1996) 1233–1244.

[14] J. Canny, A computational approach to edge detection, IEEE Trans, Pattern Anal. Mach. Intell. IEEE-8 (1986) 679–698.

[15] F. Idris, S. Panchanathan, Storage and retrieval of compressed images using wavelet vector quantization, J. Visual Languages Comput. 8 (1997) 289–301.

[16] P. Scheunders, S. Livens, G. Van de Wouwer, P. Vautrot, D. Van Dyck, Wavelet-based texture analysis, Int. J. Comput. Sci. Inform. Manage. 1 (2) (1998) 22–34.

[17] T. Minka, R.W. Picard, Interactive learning with a Society of Models, Pattern Recognition 30 (4) (1997) 565–581.

[18] I.J. Cox, M.L. Miller, S.O. Omohundro, P.N. Yianilos, PicHunter: Bayesian relevance feedback for image retrieval, Proceedings of ICPR'96 1996, pp. 361–369.

[19] Y. Rui, T.S. Huang, M. Ortega, S. Mehrotra, A relevance feedback architecture in content-based multimedia information retrieval systems, Paper presented at the IEEE Workshop on Content-based Access of Image and Video Libraries, 1997.

[20] Y. Rui, T.S. Huang, M. Ortega, S. Mehrotra, Relevance feedback: a power tool in interactive content-based retrieval, IEEE Trans. Circuits Systems Video Technol. 8 (5) (1998) 644–655. (Special Issue on Interactive Multimedia Systems for the Internet.)

[21] S. Sclaroff, L. Taycher, M. La Cascia, ImageRover: a content-based image browser for the world wide, Paper presented at the IEEE Workshop on Content-based Access Image and Video Libraries, 1997.

[22] G. Ciocca, R. Schettini, A relevance feedback mechanism for content-based image retrieval, Inform. Process. Manage. 35 (1999) 605–632.

[23] E. Binaghi, I. Gagliardi, R. Schettini, Image retrieval using fuzzy evaluation of color similarity, Int. J. Pattern Recognition Artif. Intell. 8 (1994) 945–968.

[24] P.J. Rousseeuw, A.M. Leroy, Robust Regression and Outlier Detection, Wiley, New York, 1987.

[25] A.M. Mood, F.A. Graybill, D.C. Boes, Introduzione alla statistica, McGraw-Hill, New York, 1998.

[26] M. Mitra, J. Huang, S.R. Kumar, Combining supervised learning with color correlograms for content-based image retrieval, Proceedings of the Fifth ACM Multimedia 97 Conference, 1997.

[27] M.J. McGill, G. Salton, Introduction to Modern Information Retrieval, McGraw-Hill, New York, 1983.

[28] A. Narasimhalu Desai, M.S. Kankanhalli, Benchmarking multimedia databases, Multimedia Tools Appli. 4 (1997) 333–356.

[29] B.M. Mehtre, M.S. Kankanhalli, A. Desai, C.G. Narasimhalu, Man, Color matching for image retrieval, Pattern Recognition Lett. 16 (1995) 325–331.

[30] B.M. Mehtre, M.S. Kankanhalli, W.F. Lee, Content-based image retrieval using a composite color-shape approach, Inform. Process. Manage. 34 (1) (1998) 109–120.
[31] I. Gagliardi, R. Schettini, A method for the automatic indexing of color images for effective image retrieval, New Rev. Hypermedia Multimedia 3 (1997) 201–224.
[32] T. Igarashi (Ed.), Trademarks and Logotypes, Graphic-sha, Tokyo, 1983.
[33] T. Igarashi (Ed.), Trademarks and Logotypes II: A collection of International Symbols and their Applications, Graphic-sha, Tokyo, 1983.
[34] A Collection of Trademarks and Logotypes in Japan. Graphic-sha, Tokyo, 1973.

About the Author—GIANLUIGI CIOCCA took his degree (*Laurea*) in Computer Science at the University of Milan in 1998, and he has since then been a fellow at the Institute of Multimedia Information Technologies of the Italian National Research Council, where his research has focused on multimedia system design, and the content-based retrieval of images and videos.

About the Author—RAIMONDO SCHETTINI took his degree (*Laurea*) in Physics at the University of Milan in 1986. He has been associated with the Italian National Research Council (CNR) since 1987. In 1989 he joined the Image Processing Group of the Institute of Cosmic Physics and Related Technologies. In 1994 he moved to the CNR Institute of Multimedia Information Technologies, where he is currently in charge of the Image and Color Analysis Lab. He has published more than 90 refereed papers on image processing, analysis and reproduction, and on the content-based indexing and retrieval of images. Since 1997, he also teaches a course on multimedia design at the School of Industrial Design of the Polytechnic of Milan.