Shape-Based Retrieval: A Case Study with Trademark Image Databases

Anil K. Jain and Aditya Vailaya
Department of Computer Science
Michigan State University
East Lansing, MI 48824
jain@cps.msu.edu, vailayaa@cps.msu.edu

Abstract

Retrieval efficiency and accuracy are two important issues in designing a content-based database retrieval system. We propose a method for trademark image database retrieval based on object shape information that would supplement traditional text-based retrieval systems. This system achieves both the desired efficiency and accuracy using a two-stage hierarchy: in the first stage, simple and easily computable shape features are used to quickly browse through the database to generate a moderate number of plausible retrievals when a query is presented; in the second stage, the candidates from the first stage are screened using a deformable template matching process to discard spurious matches. We have tested the algorithm using hand drawn queries on a trademark database containing $1,100$ *images. Each retrieval takes a reasonable amount of computation time (*$\sim$ *4-5 seconds on a Sun Sparc* 20 *workstation). The top most image retrieved by the system agrees with that obtained by human subjects, but there are significant differences between the ranking of the top-*10 *images retrieved by our system and the ranking of those selected by the human subjects. This demonstrates the need for developing shape features that are better able to capture human perceptual similarity of shapes. An improved heuristic has been suggested for more accurate retrievals. The proposed scheme matches filled-in query images against filled-in images from the database, thus using only the gross details in the image. Experiments with database images used as query images have shown that matching on the filled-in database extracts more images within the top-*20 *retrievals that have similar content. We believe that developing an automatic retrieval algorithm which matches human performance is an extremely difficult and challenging task. However, considering the substantial amount of time and effort needed for a manual retrieval from a large image database, an automatic shape-based retrieval technique can significantly simplify the retrieval task.*

— Key words: Image database, trademarks, logos, deformable template, moment invariants, shape similarity.

1 Introduction

Digital images are a convenient media for describing and storing spatial, temporal, spectral, and physical components of information contained in a variety of domains (e.g., aerial/satellite images in remote sensing, medical images in telemedicine, fingerprints in forensics, museum collections in art history, and registration of trademarks and logos) [1]. A typical database consists of hundreds of thousands of images, taking up gigabytes of memory space. While advances in image compression algorithms have alleviated the storage requirement to some extent, large volumes of these images make it difficult for a user to quickly browse through the entire database. Therefore, an efficient and automatic procedure is required for indexing and retrieving images from databases.

Traditionally, textual features such as filenames, captions, and keywords have been used to annotate and retrieve images. But, there are several problems with these methods. First of all, human intervention is required to describe and tag the contents of the images in terms of a selected set of captions and keywords. In most of the images there are several objects that could be referenced, each having its own set of attributes. Further, we need to express the spatial relationships among the various objects in an image to understand its content. As the size of the image databases grow, the use of keywords becomes not only cumbersome but also inadequate to represent the image content. The keywords are inherently subjective and not unique. Often, the preselected keywords in a given application are context dependent and do not allow for any unanticipated search. If the image database is to be shared globally then linguistic barriers will render the use of keywords ineffective. Another problem with this approach is the inadequacy of uniform textual descriptions of such attributes as color, shape, texture, layout, and sketch.

Although content-based image retrieval is extremely desirable in many applications, it is a very difficult problem. The ease with which humans capture the image content and how they do it have not been understood at all to automate the procedure. Problems arise in segmenting the images into regions corresponding to individual objects, extracting features from the images that capture the perceptual and semantic meanings, and matching the images in a database with a query image based on the extracted features. Due to these difficulties, an isolated image content-based retrieval method can neither achieve very good results, nor will it replace the traditional text-based retrievals in the near future.

In this paper we address the problem of efficiently and accurately retrieving images from a database of trademark images purely based on shape analysis. Since we desire a system that has both high speed and high accuracy of retrievals, we propose a two-tiered hierarchical image retrieval system. Figure 1 shows a block diagram of our proposed image retrieval scheme. We assume that a prior text-based retrieval stage exists that reduces the search space from hundreds of thousands of trademarks to a few thousand. The content-based scheme is applied to this pruned database. The first stage computes simple image features to prune the database to a reduced set of plausible matches. As simple shape features are used in screening the image database, this first stage can be performed very fast. The small set of plausible candidates generated by the fast screening stage is then presented to a detailed matcher in the second stage. This stage uses a deformable template model to eliminate false matches. The proposed hierarchical content-based image retrieval algorithm has been tested on a trademark image database containing $1,100$ images. Our retrieval system is insensitive to variations in scale, rotation, and translation. In other words, even if a query image differs from its stored representation in the database in its orientation, position, or size, the image retrieval system is able to correctly retrieve it.

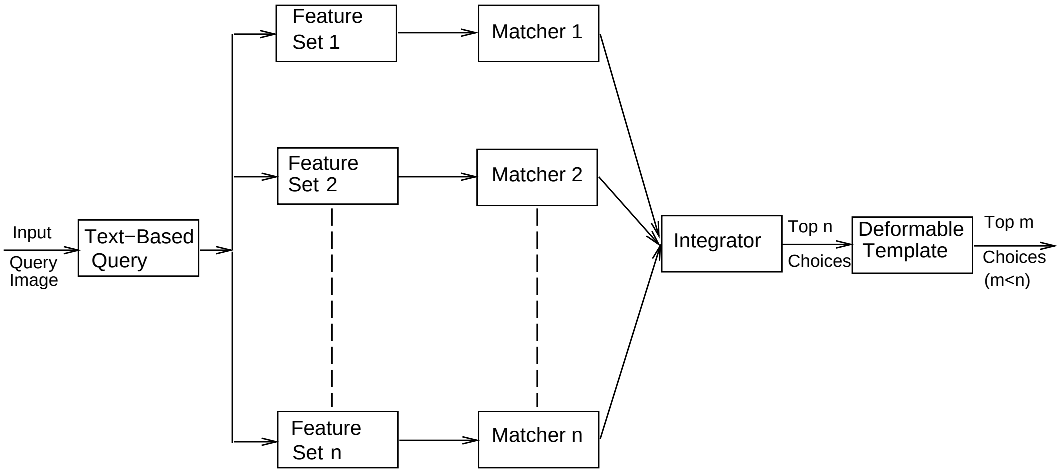


Figure 1: A hierarchical image retrieval system.

The outline of the paper is as follows. Section 2 briefly reviews relevant literature on content-based image database retrieval. In section 3 we present our image database and describe the challenges in matching trademark images based on their shape. We describe the proposed hierarchical retrieval system in section 4. Experimental results on a digital trademark database are presented in section 5. Section 6 presents the conclusions and some ideas for future research.

2 Literature Review

Much of the past research in content-based image retrieval has concentrated on the feature extraction stage. For each database image, a feature vector which describes various visual cues, such as shape, texture, and color is computed. Given a query image, its feature vector is calculated and those images which are most similar to this query based on an appropriate distance measure in the feature space are retrieved. Traditional image matching schemes based on image distance or correlations of pixel values are in general too expensive and not meaningful for such an application.

Various schemes have been proposed in the literature for shape-based representation and retrieval. These include shape representation using polygonal approximation of the shape [2] and matching using the polygonal vertices; shape matching using relaxation techniques [3] to find acceptable combinations of matches between pairs of angles on two shapes; image representation on the basis of strings [4, 5] and employing string matching techniques for retrieval; comparing images using the Hausdorff distance [6] that measures the extent to which each point in a stored database image lies near some point of the query and vice versa; point matching techniques [7] that extract a set of distinctive local features from the query and the model, and then match the resulting point patterns; image registration by matching relational structures [8] that are used to represent images; shape matching based on chord distributions [9] that uses chord length distribution for image matching; image representation using Codons [10] that uses continuous curve segments in an image that are separated by concave cusps to represent the object shape; matching objects using Fourier descriptors [11]; and object matching using invariant moments [12].

The above techniques rely on a single model and its associated features to describe the object shape. A major limitation of using a single shape model in image database retrieval is that it might not be possible to extract the corresponding features in a given application domain. Moreover, many of the shape features are not invariant to large variations in image size, position, and orientation. When we consider large image databases, retrieval speed is an important consideration. We, therefore, need to identify shape features which can be efficiently computed and which are invariant to 2D rigid transformations.

Retrieval based on a single image attribute often lacks sufficient discriminatory information. As we consider large databases, we need to extract multiple features for querying the database. Recently, attempts have been made to develop general purpose image retrieval systems based on multiple features that describe the image content. These systems attempt to combine shape, color, and texture cues for a more accurate retrieval. We briefly review a few of these systems reported in the literature.

- QBIC: The QBIC (Query By Image Content) system allows users to search through large online image databases using queries based on sketches, layout or structural descriptions, texture, color, and sample images. Therefore, QBIC techniques serve as a database filter and reduce the search complexity for the user. These techniques limit the content-based features to those parameters that can be easily extracted, such as color distribution, texture, shape of a region or an object, and layout. The system offers a user a virtually unlimited set of unanticipated queries, thus allowing for general purpose applications rather than catering to a particular application. Color- and texture-based queries are allowed for both images and objects, whereas shape-based queries are allowed only for individual objects and layout-based queries are allowed only for an entire image.

- Photobook: *Photobook* is a set of interactive tools for browsing and searching an image database. The features used for querying can be based on both text annotations and image content. The key idea behind the system is *semantics-preserving image compression*, which reduces images to a small set of perceptually significant coefficients. These features describe the shape and texture of the images in the database. Photobook uses multiple image features for querying general purpose image databases. The user is given the choice to select features based on the appearance, shape, and texture to browse through large databases. These features can be used in any combination and with textual features to improve the efficiency and accuracy of the retrievals.

- STAR: STAR (System for Trademark Archival and Retrieval) [13] uses a combination of color and shape features for retrieval purposes. The color of an image is represented in terms of the R, G, and B color components, whereas the shape is represented in terms of combination of outline-based features (sketch of the images) and region-based features (objects in an image). The features used to describe both the color and shape in an image are non-information preserving or ambiguous in nature and hence they cannot be used to reconstruct the image. However, they are useful as approximate indicators of shape and color.

- Learning-based Systems: A general purpose image database system should be able to automatically decide on the image features that are useful for retrieval purposes [14]. Minka and Picard [15] describe an interactive learning system using a society of models. Instead of requiring universal similarity measures or manual selection of relevant features, this approach provides a learning algorithm for selecting and combining groupings of the data, where these groupings are generated by highly specialized and context-dependent features. The selection process is guided by a rich user interaction where the user generates both positive and negative retrieval examples. A greedy strategy is used to select a combination of existing groupings from the set of all possible groupings. These modified groupings are generated based on the user interactions, which over a period of time, replace the initial groupings that have very low weights. Thus, the system performance improves with time through user interaction and feedback.

While multiple cue-based schemes prove more effective than single cue-based retrieval systems, content-based retrieval still faces many problems and challenges. There is no one way to decide what cues (color, shape, and texture) to use or what feature models (color histograms, reference colors, etc.) to use for a particular application. Another challenge is the problem of integration of retrievals from multiple independent cues. In the case of trademark images, color does not play a useful role in distinguishing between various marks. The design marks are registered as binary images with the United States Patent and Trademarks Office (USPTO). When searching for a conflict, the USPTO bases its decision on the shape information present in the binary images. Thus, cues like color and texture are not applicable for query purposes. We feel that multiple feature models for a particular cue (shape) can improve the retrieval accuracy just as the use of multiple cues does. We, therefore, need to investigate multiple shape models for the retrieval of trademark images. In the next section, we describe the various challenges in matching trademark images based on their shape.

3 Trademark Database

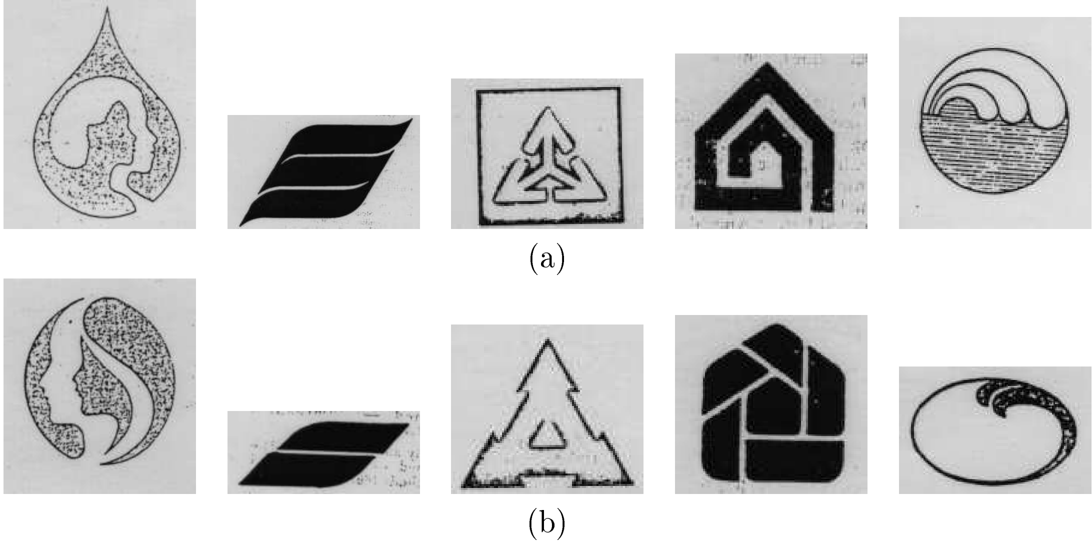
Trademarks represent a gamut of pictorial data. There are over one million registered trademarks in the U.S. alone. *A trademark is either a word, phrase, symbol or design, or combination of words, phrases, symbols or designs, which identifies and distinguishes the source of goods or services of one party from those of others. A service mark is the same as a trademark except that it identifies and distinguishes the source of a service rather than a product* [16]. Most of the trademarks are abstract representations of concepts in the world, like abstract drawings of animals, or physical objects (Sun, Moon, etc.). It is extremely challenging and instructive to study and address the issue of image database retrieval on this huge source of complex pictorial data.

3.1 Search for Conflicting Marks

Before a mark is registered with the USPTO, an examining attorney conducts a search for conflicting marks. Usually, it is not necessary for an applicant to conduct a search for conflicting marks prior to filing an application. The application fee covers processing and search costs, and is not refunded in the event a conflict is found and the mark cannot be registered.

To determine whether there is a conflict between two marks, the USPTO determines whether there would be a likelihood of confusion, i.e., *whether relevant consumers would be likely to associate the goods or services of one party with those of the other party as a result of the use of the marks at issue by both parties* [17]. The principal factors to be considered in reaching this decision are the similarity of the marks and the commercial relationship between the goods and services identified by the marks [17]. In order for a conflict to occur, it is not necessary that the marks be very similar when subjected to side-by-side comparison, but the issue is whether the marks are sufficiently similar to produce a likelihood of confusion with regard to the source of the goods or services. Thus, marks representing similar meanings and used to sell similar goods or services may cause a confusion among the general public. While evaluating the similarities between the marks, emphasis must be placed on the recollection of the average purchaser who may normally retain a general impression rather than any specific detail of the trademark.

In case of *design marks* (trademarks consisting of symbols and designs), the issue of similarity is primarily based on *visual similarity*. Here consideration must be placed on the fact that the purchaser's recollection of the mark is of a general and hazy nature. Figure 2 shows a few pictures of trademarks that were considered for *opposition* based on their visual similarity. For further details on these and other cases, please refer to the trademark manual for examining procedures [17].


(a)

(b)

Figure 2: Trademarks considered for opposition by the USPTO; (a) five registered marks, (b) registration for the first three new trademarks was refused by the USPTO because of their similarity to images in (a), whereas the last two were accepted.

The TRADEMARKSCAN Design Code Manual [18] has been developed by Thomson and Thomson to assist an applicant in finding trademarks that have, or are, specific designs. A search mechanism based on design codes associated with the trademarks is incorporated in the CD-ROMS stored at depository libraries in each of the 50 States. Trademarks are organized in a three-layered hierarchical structure. The highest level consists of 30 main *categories* (including celestial bodies, human beings, animals, geometric figures and solids, foodstuffs, etc). Each *category* is further divided into *divisions* (the second level of the hierarchy), which are further divided into *sections*. Every trademark is assigned as many six-digit design codes (two digits corresponding to each of the levels) as possible. For example, the *category* celestial bodies is divided into *divisions* such as stars and comets (01.01), constellations and starry sky (01.03), sun (01.05), etc. Similarly, the *division* sun is divided into sections such as sun rising or setting (01.05.01), sun with rays or halo (01.05.02), sun representing a human face (01.05.03), etc. It can be seen that a trademark can be assigned more than one design code. For example, figure 3 shows a trademark with a sun in a filled square background. It will thus have two design codes corresponding to the filled square and sun with rays. The CD-ROM based search procedure uses these design codes to search the database of current and pending trademarks to retrieve other marks with a similar design code as that of the query.

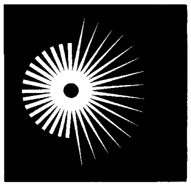


Figure 3: Example of a trademark with multiple design codes.

The current search strategy is based on manually assigning as many design codes as possible to the query (by referring to the design code manual) and then conducting the search based on the design codes at the Patent Depository Libraries. This search quite often retrieves tens of thousands of marks for a single query due to the presence of multiple design codes for a trademark. An additional, content-based scheme is therefore required to rank order these retrieved marks according to their visual similarity to the query trademark. We address this issue of an efficient content-based (shape-based) retrieval. A drawback of such a system is that it cannot handle similarities between a design mark and a word mark. For example, a design mark of a lion can be confused with a word mark, "lion". Identifying semantics from binary images is an extremely difficult task and we do not attempt to solve the problem. We assume that the design code based search would identify logically similar marks. Our goal is to extract visually similar marks from the large set of design marks that are retrieved after a search based on the design codes.

3.2 Image Database

The image database used in this study was created by scanning a large number of trademarks from several books [19, 20, 21]. About 400 images were scanned at MSU while 700 more were scanned at Siemens Corporate Research, Princeton. The database consists of these $1,100$ images. A design mark is registered in a binary form at the USPTO. Therefore, we have converted our scanned gray level images to binary images using a global threshold. The threshold was determined empirically. The 400 trademarks gathered at MSU were scanned using a Hewlett Packard Scanjet IIcx flatbed scanner at a resolution of roughly 75 dpi. The images are approximately 200×200 pixels in size. The images scanned at Siemens Corporate Research are 500×500 pixels in size. The trademarks in our database were selected so that many of them have similar perceptual meaning to make the shape-based retrieval problem more challenging. These trademarks encompass a wide variety of objects; some are based on the alphabets of the English language, while others represent the Sun, Earth, humans, eyes, animals, etc.

3.3 Difficulties in Shape-based Matching

In practice, an attorney decides whether two trademarks are sufficiently similar to cause a (legal) infringement. The goal of the proposed retrieval system is to present to the user a subset of database images that are visually similar to the query. The aim is that all the registered trademarks that may cause a conflict to the query image be included in the retrieved images. We assume that a design code-based search has already been implemented, so some pruning of the database has been done. As mentioned earlier, a major drawback of the search based on the design codes is that it produces too many marks that have the same design code, regardless of the visual appearance of these marks. Thus, there is a need for an efficient content-based (shape-based) retrieval to present only meaningful database images to the user.

A major challenge of the content-based retrieval is that trademarks that appear to be perceptually similar need not be exactly similar in their shape. In fact, it is extremely difficult to define and capture perceptual similarity. Experiments in classification of visual information [22] have illustrated how different people classify the same pictorial data into different classes. These experiments further illustrate how different samples of visual representations and different classification goals can produce different taxonomies. We define perceptual similarity in terms of the concepts the trademarks represent. Thus, two images of a bear are termed as similar and an ideal system should be able to retrieve them when an image of a bear is presented to the system. In fact, two trademarks that might produce a conflict need not have a high pixel-based correlation value. As an example, figure 4(a) shows two images of a bullhead. Although, these two images have the same perceptual meaning, they have a low correlation value. While one image is a filled head, the other is just an outline with some interior details. Figure 4(b) shows another pair of perceptually similar marks that have a low image correlation. While both the marks in figure 4(b) show a bear holding similar objects (a bottle in the first case, and a glass of wine and a bottle in the second), they are markedly different because of the direction in which the head is pointing, the difference in the shape of bottles they are holding, and the manner in which the bottles are being held. While humans can easily identify these images as containing a bear, it is extremely difficult for a computer vision system to identify such objects automatically. As another example, figure 4(c) shows an image of a Panda bear that is made up of a number of parts rather than a single component. Though, we use Gestalt principles and top-down processing to identify the presence of a bear in this figure, such identification capabilities are not present in the state-of-the-art computer vision systems. As a final example, figure 5 presents several trademark images containing corn and wheat, that are perceptually similar but are markedly different in their appearance. While some of these images are made up of a single component, others are made up of multiple components. Since, a content-based retrieval system cannot by itself retrieve perceptually similar images, we believe that such a system should be augmented by traditional text-based approaches to facilitate the search and retrieval tasks.

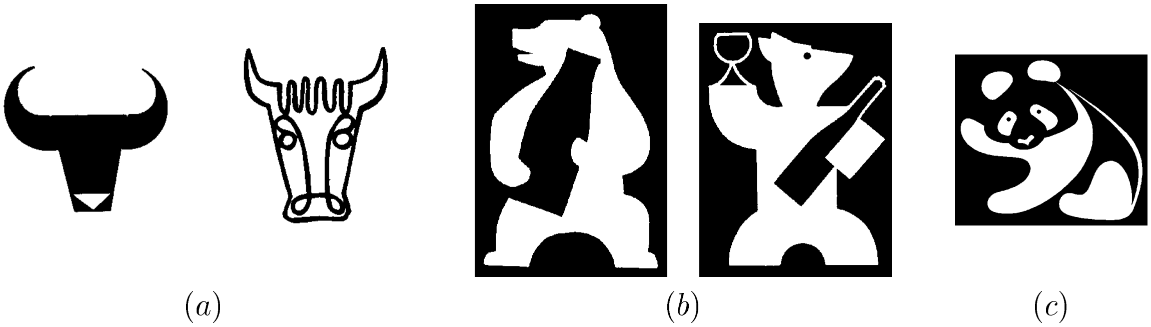


Figure 4: Perceptually similar images; (a) two images of bull head, (b) two images of a bear holding a bottle, (c) a Panda bear.

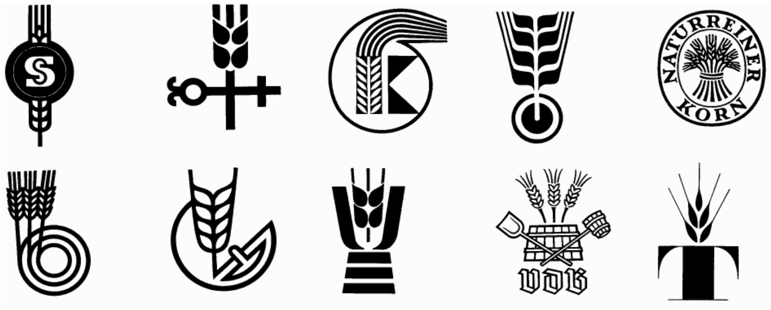


Figure 5: Images of wheat and corn that vary in their appearance.

4 Hierarchical System for Efficient Retrieval

We describe a system to extract *visually similar* trademarks from a database of design marks. The goal of the system is to present to the user all possible similar design marks that resemble in shape to the query trademark. Logically equivalent marks (such as a design of a lion and the word mark, lion) are extremely difficult to identify and our system does not tackle the issue.

Retrieval *speed* and *accuracy* are two main issues in designing image databases. System accuracy can be defined in terms of *precision* and *recall* rates. A *precision rate* can be defined as the percent of retrieved images similar to the query among the total number of retrieved images. A *recall rate* is defined as the percent of retrieved images which are similar to the query among the total number of images similar to the query in the database. It can be easily seen that both precision and recall rates are a function of the total number of retrieved images. In order to have a high accuracy, the system needs to have both high precision and high recall rates. Although, simple image features can be easily and quickly extracted, they lack sufficient expressiveness and discriminatory information to determine if two images have a similar content. Thus, there usually exists a trade-off between speed and accuracy.

In order to build a system with both high speed and accuracy, we use a hierarchical two-level feature extraction and matching structure for image retrieval (Fig. 1). Our system uses multiple shape features for the initial pruning stage. Retrievals based on these features are integrated [23] for better accuracy and higher system recall rate. The second stage uses deformable template matching to eliminate false retrievals present among the output of the first stage, thereby improving the precision rate of the system.

4.1 Image Attributes

In order to retrieve images, we must be able to efficiently compare two images to determine if they have a similar content. An efficient matching scheme further depends upon the discriminatory information contained in the extracted features.

Let $\{\mathbf{F}(x, y); x, y = 1, 2, ..., N\}$ be a two-dimensional image pixel array. For color images, $\mathbf{F}(x, y)$ denotes the color value at pixel (x, y). Assuming that the color information is represented in terms of the three primary colors (Red, Green, and Blue), the image function can be written as $\mathbf{F}(x, y) = \{F_R(x, y), F_G(x, y), F_B(x, y)\}$. For black and white images, $\mathbf{F}(x, y)$ denotes the gray scale intensity value at pixel (x, y). Let f represent a mapping from the image space onto the n-dimensional feature space, $\mathbf{x} = \{x_1, x_2, ..., x_n\}$, i.e.,

$$f : \mathbf{F} \rightarrow \mathbf{x},$$

where n is the number of features used to represent the image. The difference between two images, $\mathbf{F}_1$ and $\mathbf{F}_2$, can be expressed as the distance, D, between the respective feature vectors, $\mathbf{x}_1$ and $\mathbf{x}_2$. The choice of this distance measure, D is critical and domain-dependent. The problem of retrieval can then be posed as follows: Given a query image $\mathbf{P}$, retrieve a subset of the images, $\mathcal{M}$ from the image database, $\mathcal{S}$ ($\mathcal{M} \subset \mathcal{S}$), such that

$$D(f(\mathbf{P}), f(\mathbf{M})) \ \leq t, \qquad \mathbf{M} \in \mathcal{M},$$

where t is a user-specified threshold. Alternatively, instead of specifying the threshold, a user can ask the system to output, say, the top-twenty images which are most similar to the query image.

4.2 Fast Pruning Stage

It is desirable to have an image retrieval system which is insensitive to large variations in image scale, rotation, and translation. Hence, the pruning stage has to be not only fast but should also extract invariant features for matching. We have tried a few schemes reported in the literature, but found them inadequate for the given application. For example, shape features based on the turning edge angles [24] on the object boundary require the object to have only one closed boundary. Design marks are usually stylistic with many fine details and turning angle features are not sufficient to capture the fine details in the image due to coarse sampling. Increasing the sampling rate improves the system accuracy to an extent, but drastically effects the speed of comparisons. Moreover, at higher sample rates, the turning angles cannot be computed efficiently. We also tried to use the Principal Component Analysis (PCA) method to extract the shape features, which has been used to organize and index large image databases [25]. However, this approach also does not adequately capture the shape characteristics of our binary images. The PCA method merely computes the image distance in a *reduced* space (compared to the N^2-dimensional feature space for an $N \times N$ image), Perceptually similar images as shown in Figures 4 and 5 cannot be classified as those belonging to the same class with a simple image correlation method. Based on a detailed empirical study, we have decided to use the following shape features.

- Edge Angles: A histogram of the edge directions [23, 26] is used to describe global shape information.

- Invariant Moments: The global image shape is also described in terms of seven invariant moments [12, 26].

4.2.1 Edge Directions

A histogram of the edge directions is used to represent the shape attribute. The edge information contained in the database images is extracted off-line using the Canny edge operator [27] (with $\sigma = 1$ and Gaussian masks of size = 9). The corresponding edge directions are quantized into 72 bins of 5° each. The Euclidean distance metric is used to compute the dissimilarity value between two edge direction histograms. A histogram of edge directions is invariant to translations in an image. The positions of objects in an image have no effect on the edge directions. In order to achieve invariance to scale, we normalize the histograms with respect to the number of edge points in the image. A shift of the histogram bins during matching partially takes into account a rotation of the image. But, due to the quantization of the edge directions into bins, the effect of rotation is more than a simple shift in the bins. To reduce this effect of rotation, we smooth the histograms as follows:

$$I_s[i] = \frac{\sum_{j=i-k}^{i+k} I[j]}{2k+1}, \tag{1}$$

where I_s is the smoothed histogram, I is the original histogram, and the parameter k determines the degree of smoothing. In our experiments we used $k = 1$.

Fig. 6 shows three database images ((*a*) and (*b*) are perceptually similar to each other but (*c*) is different from (*a*) and (*b*)), their respective edge images ((*d*)-(*f*)), and the edge angle histograms ((*g*)-(*i*), 36 bin histograms are shown here). Let D_e denote the edge direction-based dissimilarity between two images. The pairwise dissimilarity value are as follows: $D_e(g, h) =$

0.065, $D_e(g, i) = 0.69$, $D_e(h, i) = 0.70$. Note that $D_e(g, h) < D_e(g, i)$ and $D_e(g, h) < D_e(h, i)$. These pairwise dissimilarity values capture the perceptual similarity for these three images.

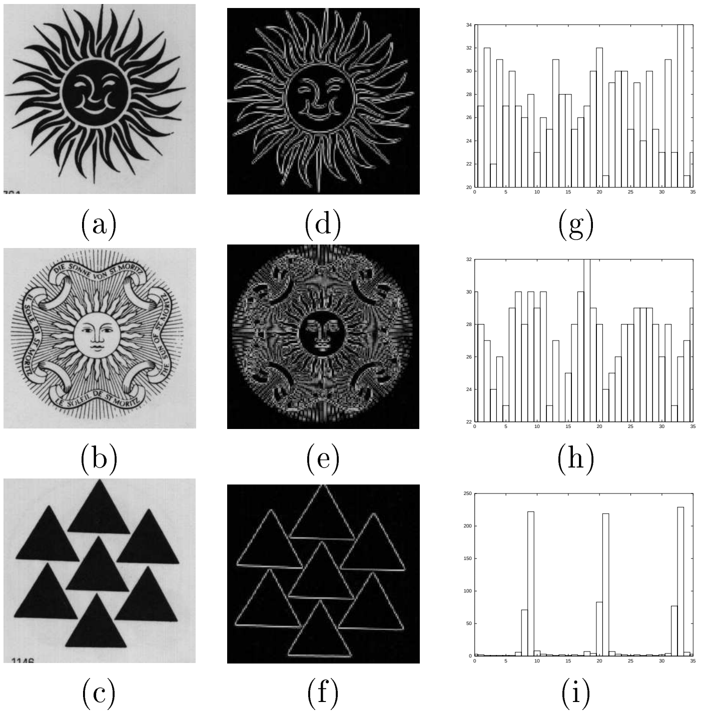


Figure 6: An example of shape representation using edge directions; (a)-(c) show 3 database images, (d)-(f) their corresponding edge images, and (g)-(i) their corresponding edge direction histograms.

4.2.2 Invariant Moments

We also represent the shape of an image in terms of 7 invariant moments. These features are invariant under rotation, scale, translation, and reflection of images and have been widely used in a number of applications due to their invariance properties [12]. For a 2-D image, $f(x, y)$, the central moment of order $(p + q)$ is given by

$$\mu_{pq} = \sum_x \sum_y (x - \overline{x})^p (y - \overline{y})^q f(x, y). \tag{2}$$

Seven moment invariants (M_1-M_7) based on the 2nd- and 3rd-order moments are given as follows:

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

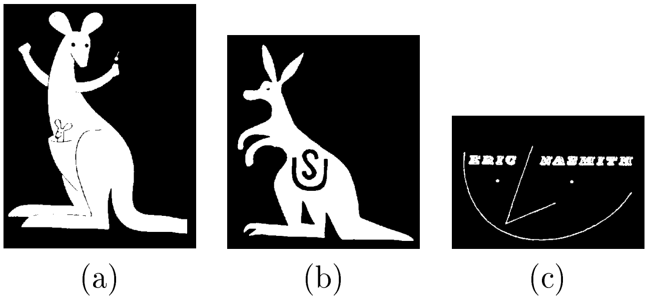
M_1 through M_6 are invariant under rotation and reflection. M_7 is invariant only in its absolute magnitude under a reflection. Scale invariance is achieved through the following transformations.

$$\begin{aligned}
&M_1' = M_1/n, \quad M_2' = M_2/r^4, \quad M_3' = M_3/r^6, \quad M_4' = M_4/r^6, \\
&M_5' = M_5/r^{12}, \quad M_6' = M_6/r^8, \quad M_7' = M_7/r^{12},
\end{aligned}$$

where n is the number of object points and r is the radius of gyration of the object:

$$r = (\mu_{20} + \mu_{02})^{1/2}.$$

Fig. 7 shows 3 database images ((a) and (b) are perceptually similar to each other but (c) is different from (a) and (b)). Let D_m denote the moments-based dissimilarity between a pair of images. The pairwise dissimilarity values are as follows: $D_m(a,b) = 0.033$, $D_m(a,c) = 0.85$, $D_m(b,c) = 0.85$. Note that $D_m(a,b) < D_m(a,c)$ and $D_m(a,b) < D_m(b,c)$ which correctly follows the perceptual similarity for these three images.



(a) (b) (c)

Figure 7: An example of shape representation using invariant moments; (a)-(c) show 3 database images.

4.2.3 Integration of Image Attributes

Use of a single image attribute for retrieval may lack sufficient discriminatory information and might not be able to support large variations in image orientation and scale. In order to increase the accuracy of the retrievals, it is often necessary to integrate the results obtained from the query based on individual shape features.

The edge direction-based matching takes into consideration the boundary of the objects whereas invariant moments are defined over an entire object region. By integrating the two different object shape attributes, we can, therefore, retrieve those images that resemble the query in either the boundary or its entirety.

We have integrated the results of the two different shape-based retrievals by combining the associated dissimilarity values. Let Q be a query image and I be a database image. Let D_e be the dissimilarity index between Q and I on the basis of edge directions and D_m be the dissimilarity index between Q and I on the basis of invariant moments. We define an integrated dissimilarity index D_t between Q and I as,

$$D_t = \frac{w_e * D_e + w_m * D_m}{w_e + w_m}, \tag{3}$$

where w_e and w_m are the weights assigned to the edge direction-based dissimilarity and the invariant moment-based dissimilarity, respectively. Given a query, the set of top-twenty retrieved images on the basis of the total dissimilarity index D_t is presented to the user. In the current implementation, we have used equal weights ($w_c = w_s = 1$). Another method of determining the weights is based on the accuracies of the individual feature-based retrievals. We observed that the retrievals based on edge direction histograms are generally more accurate than moment-based features and this fact can be used to assign a higher weight to the edge direction-based dissimilarity values.

One of the difficulties involved in integrating different distance measures is the difference in the range of associated dissimilarity values. In order to have an efficient and robust integration scheme, we normalize the two dissimilarity values to be within the same range of $[0, 1]$. The normalization is done as follows:

$$D'_s(i,j) = \frac{(D_s(i,j) - distmin)}{(distmax - distmin)}, \tag{4}$$

where D_s is the dissimilarity value between the ith query and the jth database image, and *distmin* and *distmax* are the minimum and the maximum dissimilarity values of the query image to the database images, respectively.

In order to measure the accuracy of the retrievals of the pruning stage, we conducted the following experiments with rotated, scaled, and noisy versions of each database image [23].

- *Rotated (R)*: Every image in the database was rotated arbitrarily and then presented as the query image.

- *Scaled (S)*: Every image in the database was scaled and presented as the query image.
- *Noisy (N)*: A uniform i.i.d. additive noise model was used to change either 5% of the pixel values (regions for moments) or 5% of the edge orientations (addition of noise to edges in the case of edge directions).

Tables 1 and 2 present the results of the retrieval using shape features based on the edge direction histograms and the invariant moments. We notice that both the individual shape features are not very effective in retrieving rotated images. The performance of these shape-based features is better for scaled and noisy images. The results of the integrated query are presented in Table 3. For retrieving scaled and noisy images, the topmost retrieved image is the correct result for each of the presented query. In the case of rotated images, the query matches the correct image within the top-20 positions in all but 17 of the $1,100$ cases. Figure 8 shows the 17 database images that were not retrieved in our rotation experiments. Most of these 17 images are line drawings and present very few image points for robust calculation of the invariant moments. Moreover, the edge directions also cannot be computed accurately for thin line drawings and small isolated blobs. We feel that both the edge directions and invariant moments are not very robust shape measures for line drawings. Integrating the two dissimilarity measures reduces the number of false retrievals as it is highly unlikely that a pair of perceptually different images is assigned a high similarity value in both the schemes.

4.3 Object Matching based on Deformable Templates

Both the edge direction histogram and the seven invariant moments used in section 4.2 are necessary but not sufficient features for shape matching. In other words, two dramatically different shapes can have very similar edge direction histograms and invariant moment features. We have observed that, using the above features, similar shapes are likely to be among the top-20 retrievals; however, the top retrievals also contain some trademarks that seem to be perceptually very different from the query. To further refine the retrievals to ensure that only visually similar shapes are reported to the user, we use a more elaborate matching technique based on deformable templates [28, 29]. During the refined matching stage, the edge map of the query trademark is deformed to match the edge maps of the top-N retrieved trademark images (referred to as test trademarks) as much as possible. The edge map of the query trademark is used as a prototype template; this template is deformed towards the edge map of the test trademark. The goodness of the matching is determined by an energy function which depends on the amount of deformation of the template and how well the deformed template fits the test edge map.

Query Nature	$n = 1$ (%)	$n \leq 2$ (%)	$n \leq 5$ (%)	$n \leq 20$ (%)	Not Retrieved (%)
R	73	79	86	94	6
S	100	100	100	100	0
N	92	95	96	100	0

Table 1: Edge directions-based retrieval results for the $1,100$ database images.

Query Nature	$n = 1$ (%)	$n \leq 2$ (%)	$n \leq 5$ (%)	$n \leq 20$ (%)	Not Retrieved (%)
R	44	55	67	88	12
S	100	100	100	100	0
N	88	95	97	100	0

Table 2: Invariant moments-based retrieval results for the $1,100$ database images.

Query Nature	$n = 1$ (%)	$n \leq 2$ (%)	$n \leq 5$ (%)	$n \leq 20$ (%)	Not Retrieved (%)
R	85	91	95	98.5	1.5
S	100	100	100	100	0
N	96	99	100	100	0

Table 3: Integrated shape-based retrieval results for the $1,100$ database images; n refers to the position of the correct retrieval, R: rotated query, S: scaled query, N: query with a noisy image; The last column indicates percentage of time the query image was not retrieved in the top-20 matches.

In the matching scheme defined in Jain et al. [28], the deformation model consists of *(i)* a prototype template which describes the representative shape of a class of objects, and *(ii)* a set of parametric transformations which deforms the template. The query image is used as the prototype, which is deformed to match the test trademarks. An energy function is then calculated to evaluate the quality of the match.

4.3.1 Representation of the Prototype Template

A prototype template consists of a set of points on the object contour, which is not necessarily closed, and can consist of several connected components. Only the edge map of the query trademark image is used (Fig. 9 (b)) for its representation. Since the images in the database are binary, the edge map is obtained as the set of foreground pixels which have at least one neighboring background pixel.

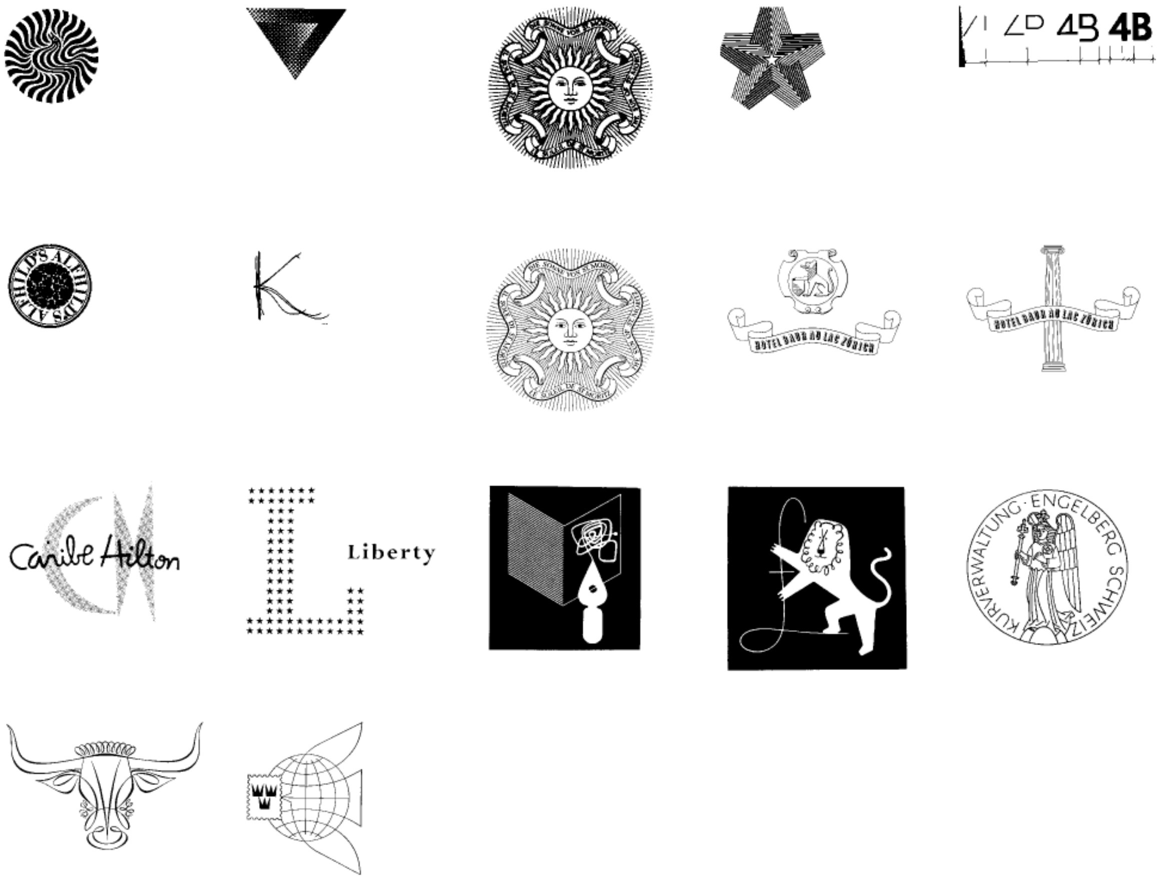


Figure 8: Database images not retrieved correctly in the presence of rotation.

4.3.2 Deformation Transformation

The prototype template describes only one of the possible (though most likely) instances of the shape of interest. Therefore, it has to be deformed to match similar trademarks. A deformation of the template is performed by introducing a displacement field in the 2D template image. Without any loss of generality, it is assumed that the template is drawn on a unit square $\mathcal{S} = [0,1]^2$. The points in the square are mapped by the function $(x,y) \mapsto (x,y) + (\mathcal{D}^x(x,y), \mathcal{D}^y(x,y))$, where the displacement functions $\mathcal{D}^x(x,y)$ and $\mathcal{D}^y(x,y)$ are continuous and satisfy the following boundary conditions: $\mathcal{D}^x(0,y) \equiv \mathcal{D}^x(1,y) \equiv \mathcal{D}^y(x,0) \equiv \mathcal{D}^y(x,1) \equiv 0$. The space of such displacement functions is spanned by the following orthogonal bases [30]:

$$\begin{aligned} \mathbf{e}^x_{mn}(x,y) &= (2\sin(\pi n x)\cos(\pi m y), 0) \\ \mathbf{e}^y_{mn}(x,y) &= (0, 2\cos(\pi m x)\sin(\pi n y)), \end{aligned} \tag{5}$$

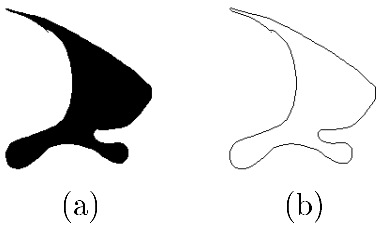


Figure 9: Constructing the query template for deformable template matching. **(a)** A hand drawn query trademark, and **(b)** the query template obtained by calculating the edge map of the image in **(a)**.

where $m, n = 1, 2, \ldots$. Specifically, the displacement function is chosen as follows:

$$\mathcal{D}(x, y) = (\mathcal{D}^x(x, y), \mathcal{D}^y(x, y)) = \sum_{m=1}^{M} \sum_{n=1}^{N} \frac{\xi^x_{mn} \cdot \mathbf{e}^x_{mn} + \xi^y_{mn} \cdot \mathbf{e}^y_{mn}}{\lambda_{mn}}, \tag{6}$$

where $\lambda_{mn} = \alpha\pi^2(n^2 + m^2)$, $m, n = 1, 2, \ldots$ are the normalizing constants. The parameters $\underline{\boldsymbol{\xi}} = \{(\xi^x_{mn}, \xi^y_{mn}), m, n = 1, 2, \ldots\}$, which are the projections of the displacement function on the orthogonal basis, uniquely define the displacement field, and hence the deformation. The deformations of the template shown in Figure 9 are shown in Figure 10.

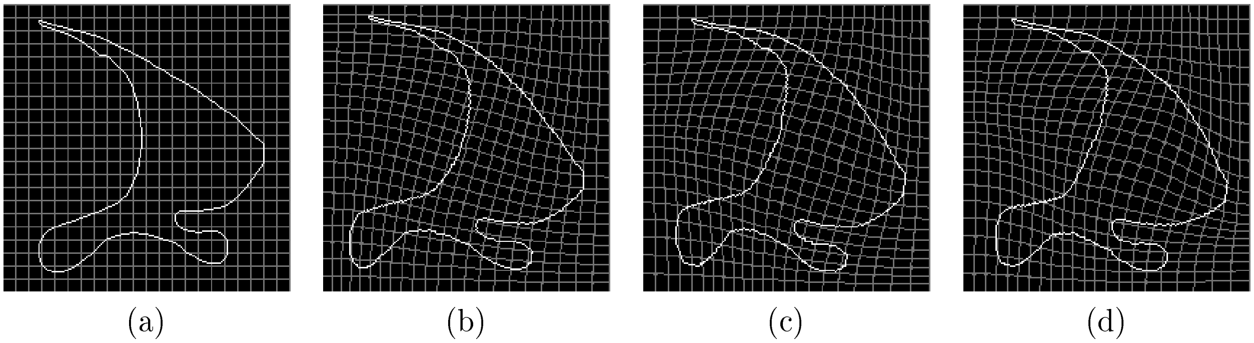


Figure 10: Deformation transformation for the boomerang template; (a)-(d) show the template under increasing deformation.

4.3.3 Dissimilarity Measure

The dissimilarity between a trademark edge map and the query template is described by two terms: (*i*) the amount of deformation of the template; the larger the deformation, the more the deformed template deviates from the prototype, and (*ii*) the discrepancy between the deformed template and the edge map of the test trademark.

Formally, the dissimilarity measure of a trademark I and query template Q using deformable template is defined as:

$$D_{dt}(Q, I) = \min_{s,\Theta,\underline{\boldsymbol{\xi}},\underline{d}} \{\gamma \sum_{m=1}^{M} \sum_{n=1}^{N} ({\xi_{mn}^{x}}^2 + {\xi_{mn}^{y}}^2) + \mathcal{E}(Q_{s,\Theta,\underline{\boldsymbol{\xi}},\underline{d}}, I)\}, \tag{7}$$

where $Q_{s,\Theta,\underline{\boldsymbol{\xi}},\underline{d}}$ denotes the deformed template Q with scale s, orientation Θ, deformation $\underline{\boldsymbol{\xi}}$, and position $\underline{d}$. The first term on the right hand side of Eq. (7), which is the sum of squares of the deformation parameters, measures the deviation of the deformed template from the prototype template; the second term is the energy function that relates the deformed template $Q_{s,\Theta,\underline{\boldsymbol{\xi}},\underline{d}}$ to the edges in the test trademark image I:

$$\mathcal{E}(Q_{s,\Theta,\underline{\boldsymbol{\xi}},\underline{d}}, I) = \frac{1}{n_Q} \sum (1 + \Phi(x, y) | \cos(\beta(x, y))|), \tag{8}$$

where the potential field $\Phi(x, y) = -\exp\{-\rho(\delta_x^2 + \delta_y^2)^{1/2}\}$ is defined in terms of the displacements (δ_x, δ_y) of a template pixel (x, y) to its nearest edge pixel, $\beta(x, y)$ is the angle between the tangent of the nearest edge and the tangent direction of the template at (x, y), ρ is a smoothing factor which controls the degree of smoothness of the potential field, the summation is over all the pixels on the deformed template, n_Q is the number of pixels on the template, and the constant 1 is added so that the potentials are positive and take values between 0 and 1. Intuitively, the first term on the right hand side in Eq. (7) favors small deformations, and the second term requires that the deformed template be in the proximity of and aligned with the edge directions of the test image. The parameter γ provides a relative weighting of the two penalty measures; a larger value of γ implies a lower variance of the deformation parameters, and as a result, a more rigid template. This dissimilarity is always nonnegative and it is zero if and only if the query template matches the edge map of the test trademark image exactly.

The dissimilarity measure D_{dt} defined in Eq. (7) is minimized *w.r.t.* the pose and deformation parameters ($s, \Theta, \underline{\boldsymbol{\xi}}$, and $\underline{d}$) of the template. The optimization is carried out by finding an initial guess for the pose parameters (s, Θ and $\underline{d}$) using the generalized Hough transform, and then performing a gradient descent search in the parameter space. It is this iterative process which makes the deformable template matching a computationally expensive operation. The proposed deformable template-based matching scheme has been successfully applied to retrieve a template from complex scenes. For details, readers are requested to refer to [28].

We have presented our framework for shape-based image retrieval. The retrieval consists of a two-stage process, the fast pruning stage for retrieving a small subset of database images, and a comprehensive matching strategy based on deformable template analysis to discard false retrievals.

5 Experimental Results

We have applied the hierarchical shape-based retrieval algorithm to a trademark image database. Our goal is to present the user with a subset of images that are most similar to the query. We have conducted experiments on the original trademark images as well as some hand drawn sketches of trademark images. The system integrates the retrievals based on the edge directions and invariant moments and presents to the deformable template matching stage a small subset of images ranked based on their similarity to the query image. The deformable template matching further prunes the subset of database images presented to the user.

The various stages involved in the image retrieval are mentioned below.

- *Preprocessing*: The edge direction and invariant moment features are pre-calculated and stored for all the images in the database. As a result, two feature vectors are associated with each database image.

- *Query image*: A query consists of a hand drawn image of a shape. It can be disconnected, or may contain holes. Fig. 11 shows some of the hand drawn query trademarks used in the experiments.

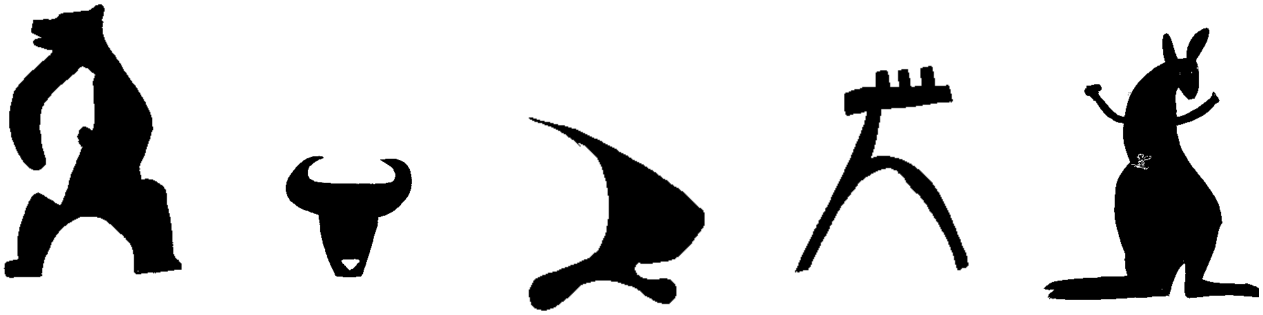
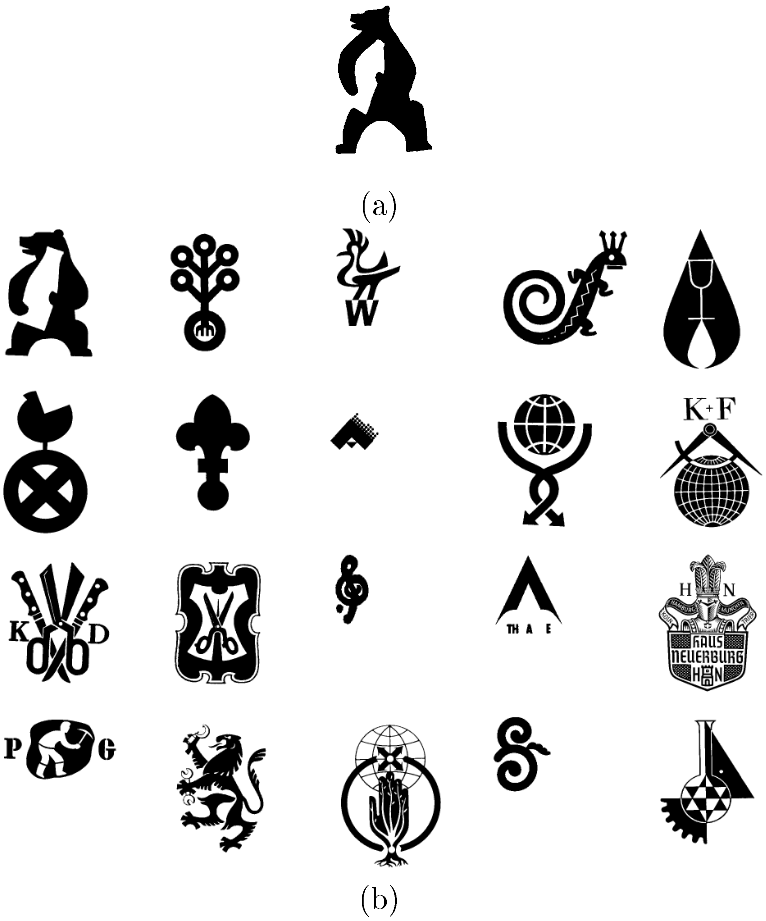


Figure 11: Examples of hand drawn query trademarks.

- *Fast pruning*: The query image is compared to the database images based on the edge direction histogram and invariant moment shape features using the integrated dissimilarity index D_t (Eq. (3)). Figures 12 to 14 show the top-20 retrieved images in the order of increasing dissimilarity for the bear, bull, and kangaroo query images, respectively. The correct database image that is retrieved has the smallest dissimilarity value for queries involving bear, bull, boomerang, and deer, and the second smallest dissimilarity value in the case of kangaroo query. Although the query image is linearly compared to all the images in the database,the pruning process is still reasonably fast. For each query, this stage takes about 4-5 seconds on a Sun Sparc 20 workstation (for a database containing $1,100$ images).



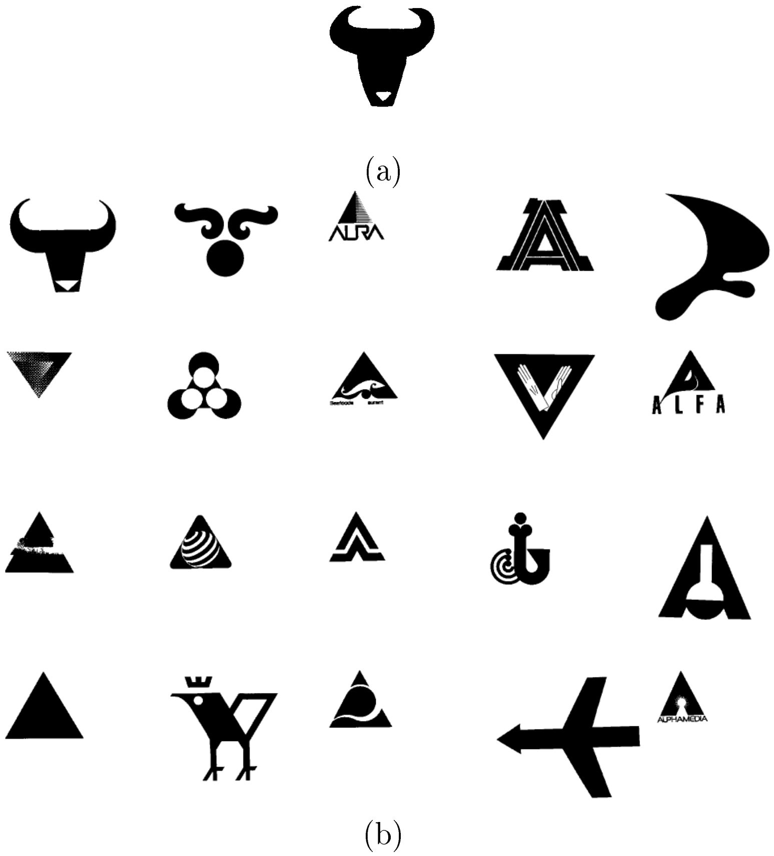
(a)

(b)

Figure 12: Database pruning results for the hand drawn bear ((a)), and ((b)) the top-20 retrievals given in the increasing order of dissimilarity.

- *Matching based on Deformable Template*: Under the assumption that all plausible candidates for a query image are contained in the top-10 retrievals in the fast pruning stage, we apply the deformable matching scheme on these candidates only to further refine the results. The initial pose parameters of the deformable template (position, scale, and orientation) are estimated using the generalized Hough transform. Figures 15(a) and (b) illustrate the initial and final configurations of the deformable template match for the bull trademark. It typically takes $5-8$ seconds to calculate the initial configuration using the generalized Hough transform. The iterative deformable matching process takes about 6 seconds on a Sun Sparc 20 workstation.

Table 4 presents the dissimilarity measures of the five hand drawn logos (Fig. 11) to the top-10 retrieved images by the pruning stage. In four out of the five queries, the simple integrated shape dissimilarity index ranks the correct logo in the first place, and in one case, the correct logo is ranked in the second place. The dissimilarity score using the deformable matching ranks the desired images (underlined) in the first place for all the five queries. An incorrect match



(a)

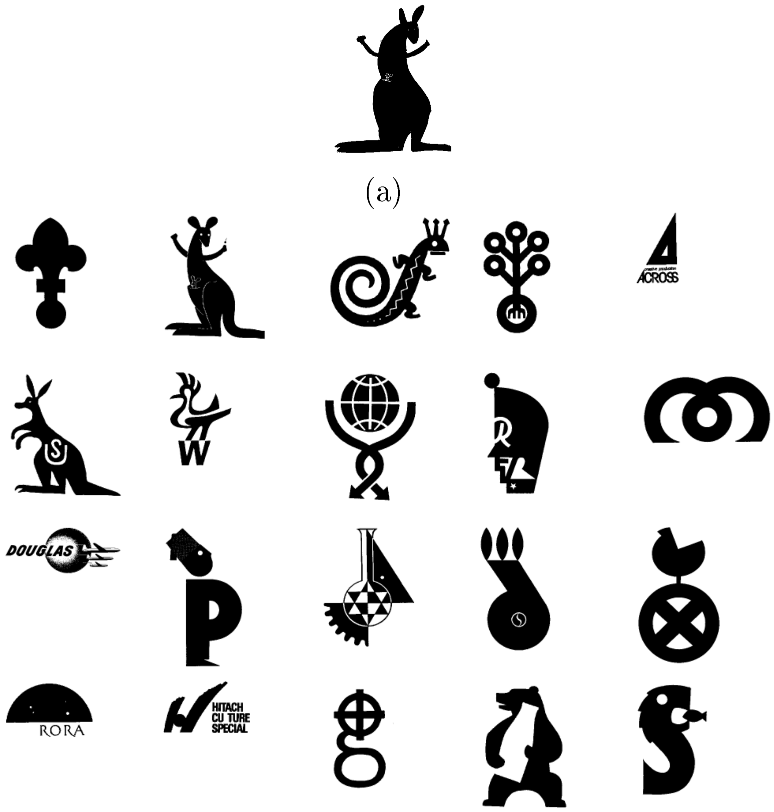
(b)

Figure 13: Database pruning results for the hand drawn bull ((a)), and ((b)) the top-20 retrievals given in the increasing order of dissimilarity.

generally results in a large dissimilarity value so that the corresponding matching hypothesis can be rejected.

5.1 Retrieval by Human Subjects

It is instructive to compare the automated retrieval results with those obtained by human subjects. For this purpose, we asked five subjects to retrieve images from the same database using the same five queries shown in Fig. 11. Fig. 16 shows the top-nine retrievals for the bull query by the human subjects. It took each subject between 1 to 2 hours to find the top-ten retrievals for the five query images. By comparing Figs. 13 and 16, one can make the following observations. The top most retrieval for the bull query and the human respondents is the same. In fact, the top most retrievals for all the five queries are consistent for both the human respondents and the proposed algorithm. This is expected since the hand drawn queries closely resemble one of the database images. As there is a substantial amount of diversity in the image database, only two to three of the top-ten retrievals are in common between the outputs of the algorithm and human



(a)

(b)

Figure 14: Database pruning results for the hand drawn kangaroo (a), and ((b)) the top-20 retrievals given in the increasing order of dissimilarity.

respondents for each query. We note that for all the queries, the retrievals obtained by the five respondents are somewhat consistent for the following reasons: *(i)* human subjects can easily decide the foreground greyscale of the object, no matter whether it is 0 or 255, and *(ii)* human subjects tend to abstract the query image for some conceptual information. As an example, for the bull query, most human respondents retrieved all the trademark images in the database which contain a bull head, even though the shape of the bull in the retrieved images is quite different from the query shape. Our system, on the other hand, cannot understand the concept of a bull head and retrieves most of the images that resemble a triangle. As another difference, while we (humans) can extract other images with similar semantic content (like bull head in this example), we lack the ability to easily match objects at different orientations. The proposed system retrieves images invariant to orientation, and hence, retrieves images that resemble a triangle (a convex hull approximation of the bull head) at arbitrary orientations. A drawback with our system is its inability to adjust to changing contexts. We (humans) have the ability to retrieve different images for the same query based on the context we are looking for. We thus feel that it is not possible with the current state-of-the-art in computer vision to build a system that can

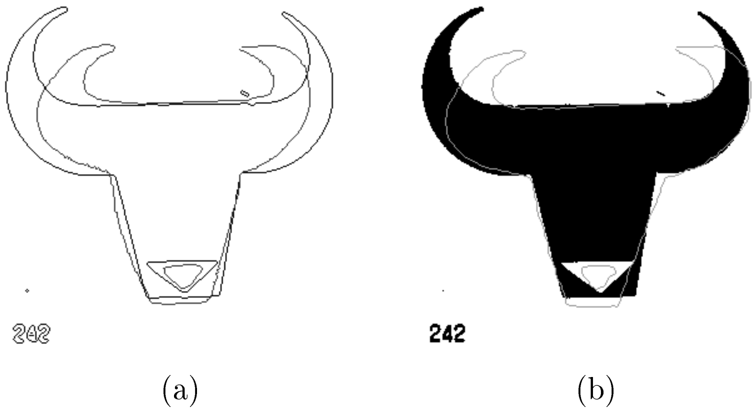


Figure 15: Deformable template matching; (a) the initial position of the bull template overlaid on the edge map of a bull logo using the generalized Hough transform, (b) the final match.

template	*Top* 10 *retrievals from the fast pruning stage*									
	1	2	3	4	5	6	7	8	9	10
bull	.149	.670	.959	.856	.847	.862	.803	.784	.820	.913
boomerang	.137	.596	.731	.820	.628	.785	.794	.857	.771	.804
bear	.425	.639	.504	.509	.705	.688	.640	.669	.574	.609
kangaroo	.751	.422	.521	.630	.877	.725	.639	.628	.645	.559
deer	.392	.457	.662	.857	.677	.665	.488	.787	.686	.425

Table 4: Dissimilarity values for the five query images when the deformable template matching is applied to the top-10 retrieved images from the fast pruning stage.

completely mimic human behavior, but we feel that a consistent system that presents efficient and accurate results (upto a certain tolerance range) is an important step towards automation of image database retrieval. These observations explain the difference between the retrievals by human subjects and the proposed algorithm.

5.2 Retrieval based on Image Filling

Retrieval results in Figs. 13 and 16 appear to indicate that trademark similarity is based more on the general global shape of the marks rather than on fine details. When the query is a filled bull image, the system tends to retrieve database images that are filled, rather than the other bull images made up of line drawings. Can the object outline itself be used for retrieving perceptually similar images? We extracted the outline of objects within the trademark images and computed edge direction histogram and moment-based features on the outlines. This approach did not yield good retrieval results since both the methods are not very robust for line drawings (Section 4.2.3). Thus, instead of using just the outline, we filled in the objects and computed the features on the filled-in images. Figure 17 presents five bull images in the database, and their

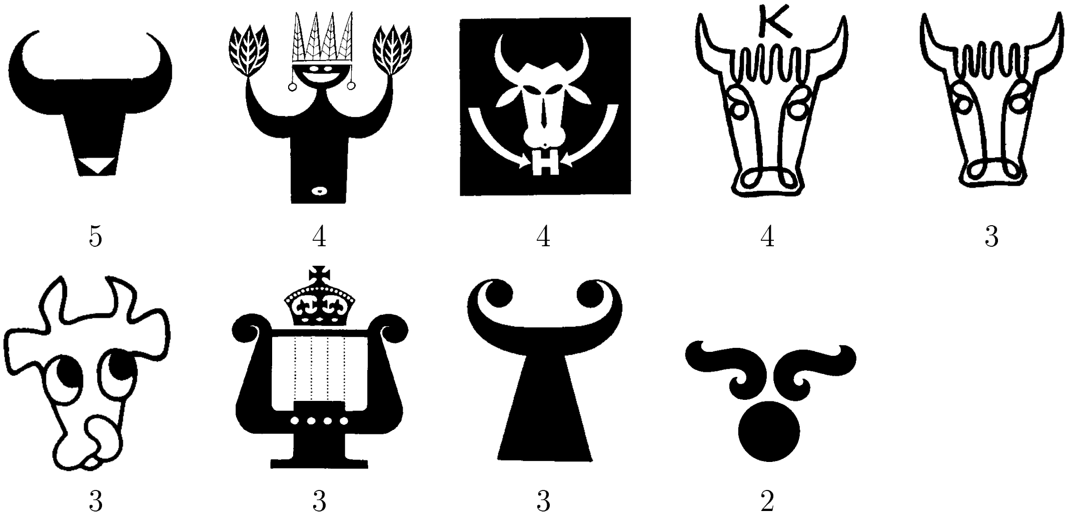
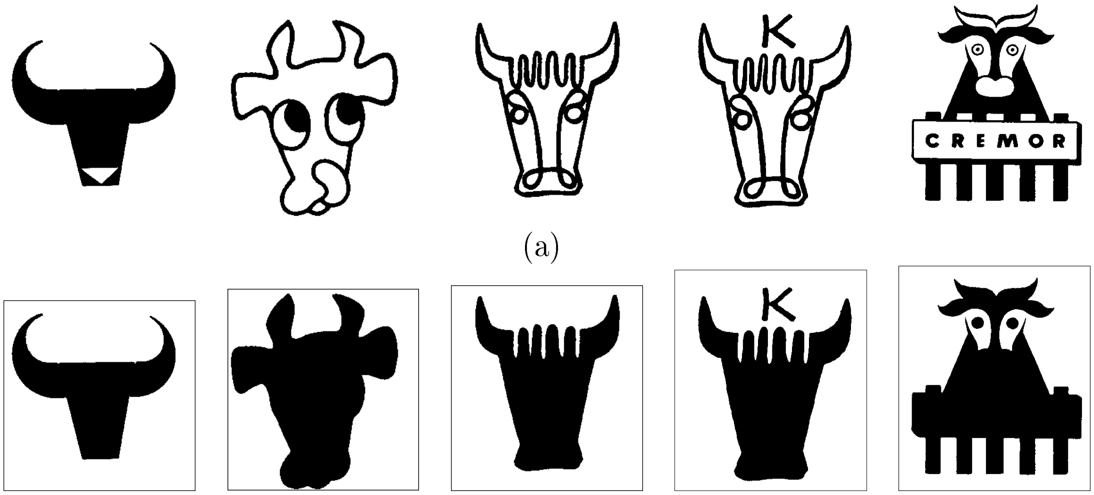


Figure 16: Nine top most retrieved trademark images by the five human subjects for the bull template. The number below each retrieved image is the number of human respondents (out of 5) who placed that image in the top-10 retrievals.

corresponding filled-in images. The image filling is done by extracting the connected components in the background pixels in an image and setting each one of them except the actual background (the rest are the holes) to the foreground value.

We have conducted experiments to match the filled-in query images against the filled-in images from the database. The aim of these experiments was to study the effect of using only gross details in the images (filling-in leads to a coarser representation of the images in the database) on the efficiency and accuracy of the fast pruning stage. Applying this technique on the hand drawn sketches, we found that image filling improved the rankings of other similar images, but these similar images were still not present in the top-20 retrievals. We also conducted experiments with database images as the query (as opposed to hand drawn images) and compared the top-20 retrievals based on the filled and unfilled databases. Figures 18(b), 19(b), and 20(b) show retrieval results for three query images (a bull image, a bear image, and a kangaroo image shown in Figures 18(a), 19(a), and 20(a)). It can be seen that in each case, the retrieval based on the filled-in database extracts more images that have similar semantic content. In the case of the query on a bull image (figure 18(a)), the system extracts four other images of a bull head other than the query one. These are bull head images retrieved in the second, third, sixth, and fourteenth places. Similarly, we see that with the bear and kangaroo queries (Figures 19(b) and 20(b)), we retrieve two more images of a bear and a kangaroo, respectively. As a basis of comparison, Figures 18(c), 19(c), and 20(c) show the retrieval results when the bull, bear and kangaroo images from the database were retrieved without filling the holes in the images. It is extremely difficult to quantify the results of the system in terms of the actual recall or



(a)

(b)

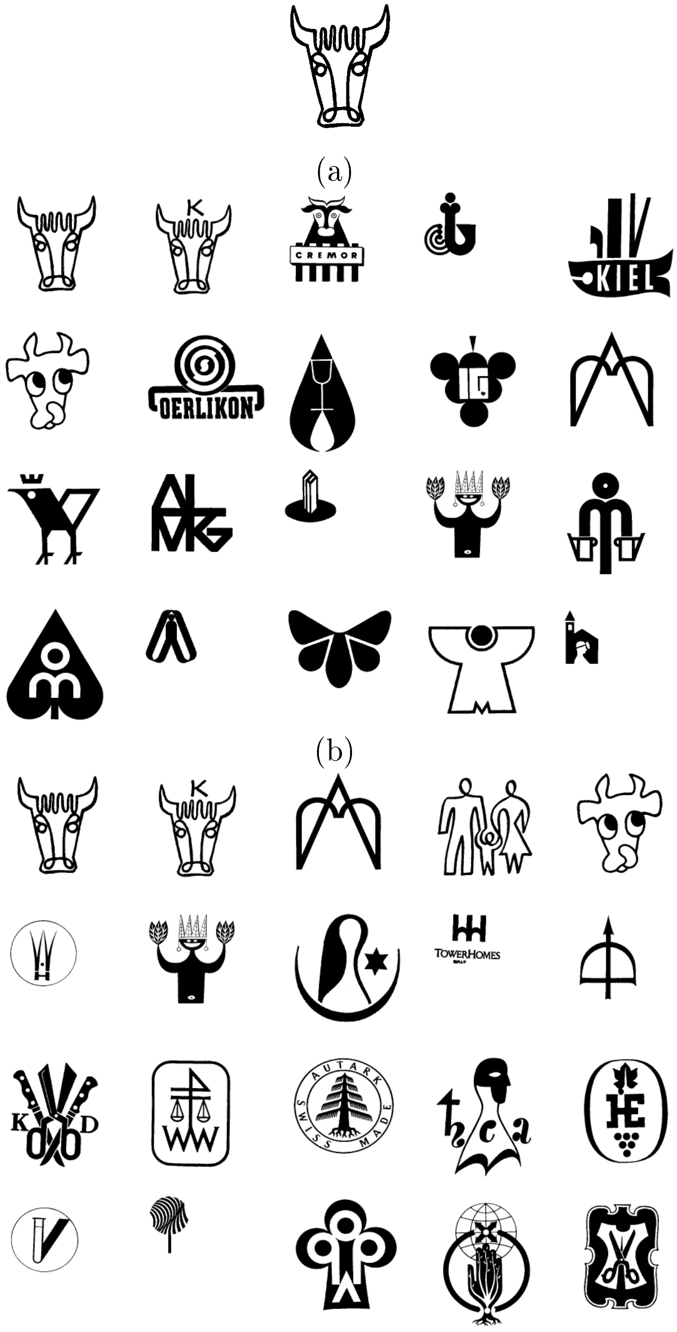
Figure 17: Image Filling; (a) Five images of a bullhead, (b) Filled-in images for the five bull images.

precision rates. A major drawback in quantifying the precision and recall rates is the need to identify images that are *similar* to the query image. We have found that it is time consuming and tedious for human subjects to identify all images in the database that are *similar* to a query image. Therefore, we compute the precision and recall rates using our (one of the authors) definition of similarity to be the perceptual similarity. Table 5 presents the recall and precision rates for three queries as shown in Figures 18, 19, and 20. We have been able to identify ten images of a bullhead, four images of a bear, and three images of a kangaroo in the database. The precision and recall rates are based on these values. A more detailed study is needed to have a better understanding of the system performance in terms of the precision and recall rates.

Query	n_1 (F)	n_1 (U)	n_2	Recall Rate (F)	Recall Rate (U)	Precision Rate (F)	Precision Rate (U)
Bullhead	4	3	10	40%	30%	20%	15%
Bear	3	1	4	75%	25%	20%	5%
Kangaroo	3	2	3	100%	66%	15%	10%

Table 5: Recall and precision rates for three queries (shown in Figures 18, 19, and 20); n_1 refers to the number of images retrieved in the top-20 positions that are *similar* to the query, n_2 refers to the number of images in the database that are *similar* to the query, (F) refers to results on filled images, and (U) refers to the results on the unfilled images; Recall rate is (n_1/n_2), precision rate is $(n_1/20)$.

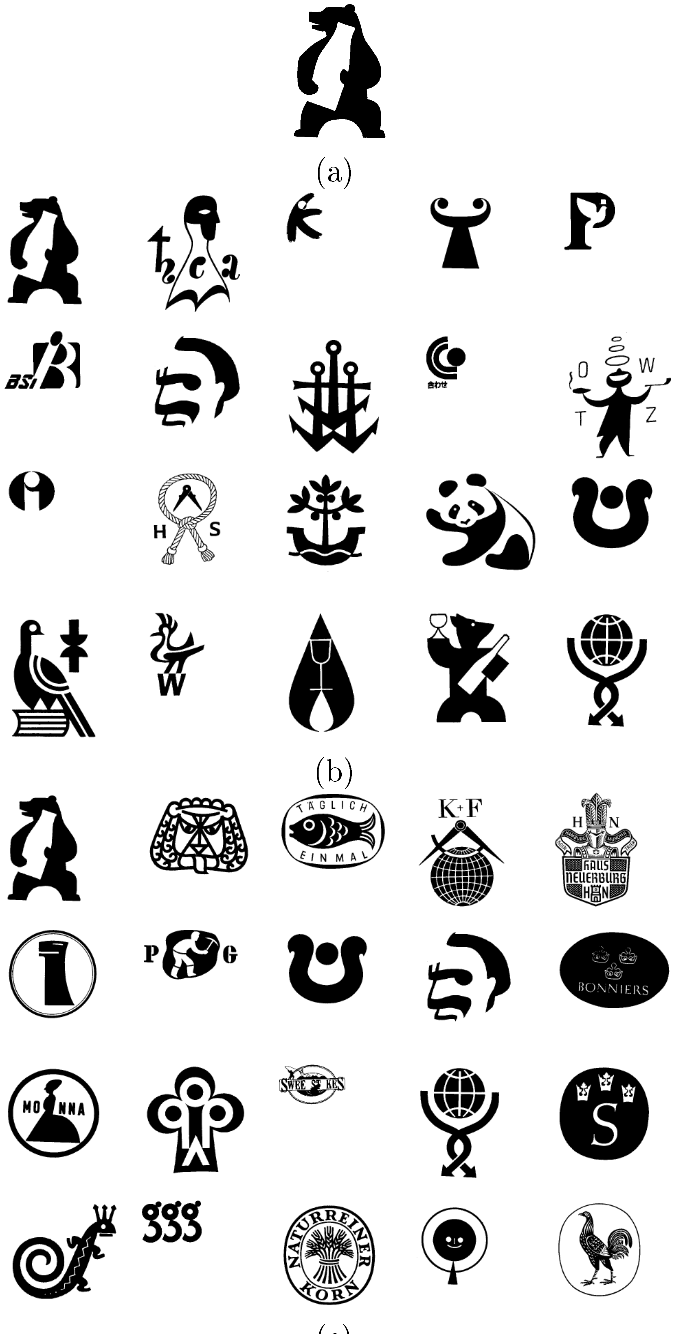
Filling-in the holes in the image improves the efficiency of a match and makes the system more robust, since the finer details are neglected (in the case of edge directions, holes contribute to the edge direction histograms and in the case of invariant moments, they are evaluated over larger


(a)

(b)

(c)

Figure 18: Database pruning results for a bull image ((a)), (b) top-20 results based on filled images, and (c) top-20 results based on unfilled images.

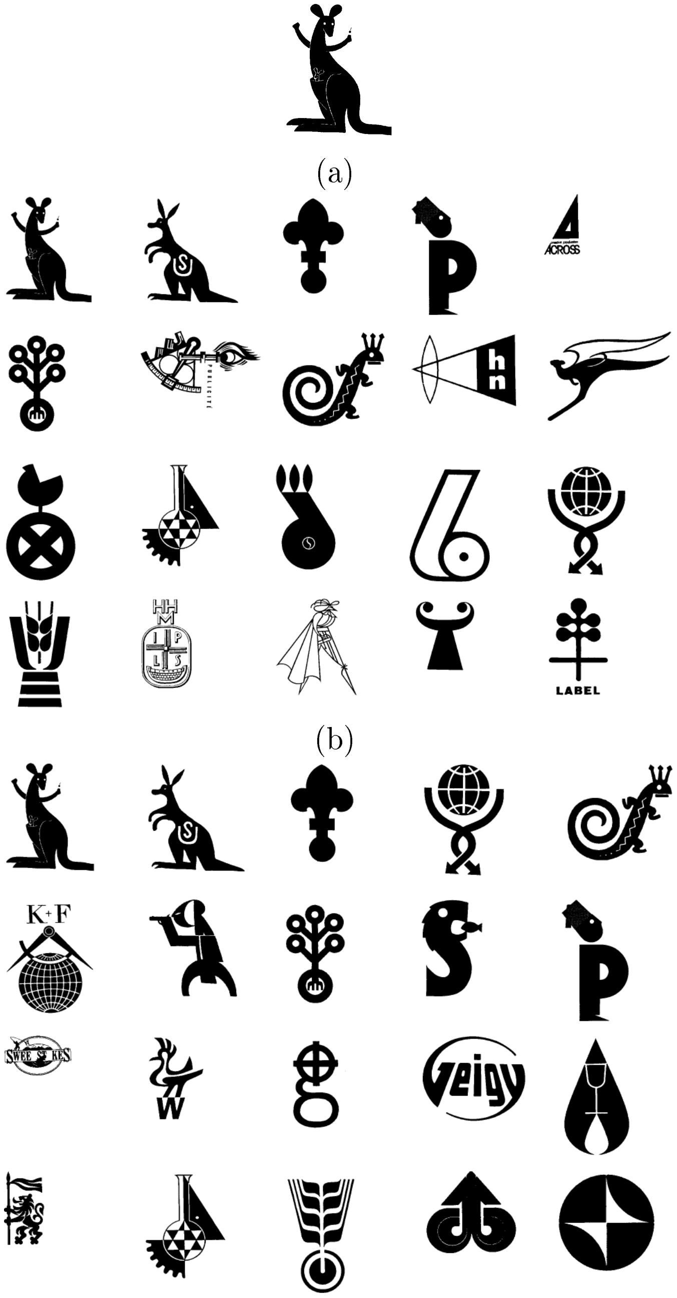


(a)

(b)

(c)

Figure 19: Database pruning results for a bear image ((a)), (b) top-20 results based on filled images, and (c) top-20 results based on unfilled images.



(a)

(b)

(c)

Figure 20: Database pruning results for a kangaroo image ((a)), (b) top-20 results based on filled images, and (c) top-20 results based on unfilled images.

number of points improving their robustness). With the proposed filling-in, we can now match images which are either just outlines or regions. A major drawback of the scheme though, is non-utilization of information contained in the holes. Many a times, certain trademarks contain useful information in their holes. Figure 21 shows two database images where the holes contain significant information. Note that, for the sake of clarity, the holes are marked in white where as the foreground for these trademarks is black. When filling-in the database images, these holes are not utilized, leading to false matches (here matching is based on squares). A human can very easily identify the foreground and background in an image and extract various objects from the image. Our proposed system is not able to do this automatically. Another difference is the ease of humans to identify objects in an image. Figure 22 shows a trademark which uses an image of a bee as the design mark. The object (bee) is made up of a number of components. It is easy for humans to identify the object, but it is extremely difficult for an automatic system to group the components into a single object. Automatic object segmentation is a promising area for future research. We believe that if the system is presented with segmented objects extracted from images, then it can perform matching based on these objects. At present, the system extracts images similar to the entire query image, rather than on the basis of the individual objects present. Figure 23 presents the retrieval results based on a query containing an image which has significant information in the holes; all the images in the database with a square border are retrieved.

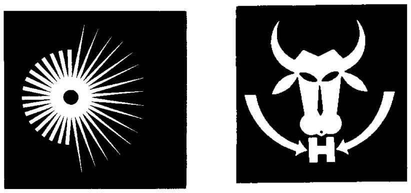


Figure 21: Two images containing a square border; information contained in the holes (white) is neglected when filling in the database.

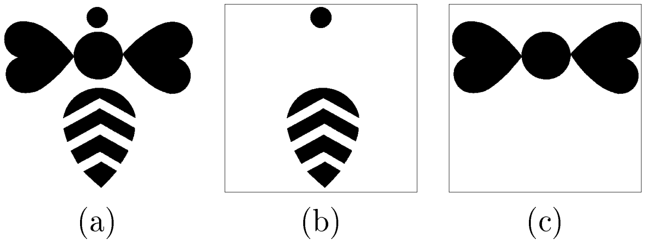


Figure 22: Object consisting of multiple components; (a) trademark image of a bee; (b) & (c) segmented objects.

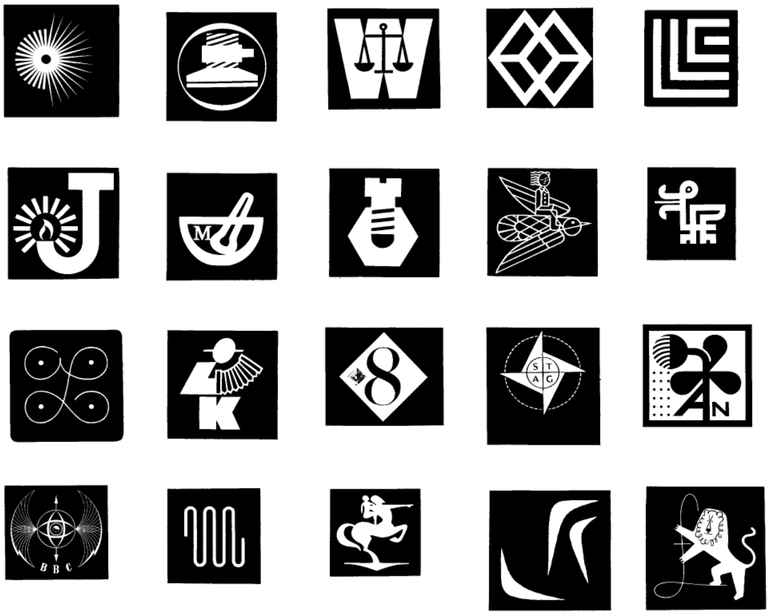


Figure 23: Database pruning results for an image with a square border (query image corresponds to the top most retrieval). The top-20 retrievals are given in the increasing order of dissimilarity.

6 Conclusions and Future Work

An efficient shape-based retrieval algorithm has been developed to retrieve trademark images. Efficiency and accuracy of retrievals are achieved by designing a two-stage hierarchical retrieval system: *(i)* a simple statistical feature-based process quickly browses through a database for a moderate number of plausible retrievals; and *(ii)* a deformable template matching process screens the candidate set for the best matches. Preliminary results on a trademark image database show that this is a promising technique for content-based image database retrieval. The technique is robust under rotated, scaled and noisy versions of the database images [26].

We note that image retrieval based on user-provided information such as hand drawn sketches is a challenging problem in multimedia applications. A human's perception of shape can be rather subjective and as a result, the representation of the desired object has a large variance. Our goal is to extract images which have a semantic content that is similar to the query image. Figure 16 demonstrates that semantically similar images may actually be visually very different from each other. In order to retrieve these images in the fast pruning stage, we first need to somehow extract salient shape features from the images. We present a heuristic that extracts the outline of the objects in an image and fills-in the interior in order to generalize the shape of the objects in the image. This scheme improves the retrieval results for the trademark images. Another method to extract the semantic content is through a semi-automatic (manual intervention during preprocessing) scheme using textual description of the trademark images. Text-based search can then be incorporated prior to the pruning stage.

There are a number of research issues which need to be addressed to improve the system performance. Our system can be extended to match on the basis of image components rather than the entire image. This would require a robust and automatic segmentation algorithm. A further extension could be to allow partial matchings of objects based on local features. Currently, we only use global features in the fast pruning stage. Local information can be very useful in eliminating many false matches. Our system currently has no learning capabilities; it cannot improve its performance with time. A future extension may be in the direction of allowing the system to learn through positive and negative examples of each query. Another area of research is in automatic grouping or clustering of the trademark images to improve the retrieval speed.

References

[1] A. K. Jain and C. Dorai, "Practicing vision: Integration, evaluation and applications," *Pattern Recognition*, vol. 30, no. 2, pp. 183–196, 1997.

[2] R. Schettini, "Multicolored object recognition and location," *Pattern Recognition Letters*, vol. 15, pp. 1089–1097, November 1994.

[3] L. S. Davis, "Shape matching using relaxation techniques," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, pp. 60–72, January 1979.

[4] G. Cortelazzo, G. A. Mian, G. Vezzi, and P. Zamperoni, "Trademark shapes description by string-matching techniques," *Pattern Recognition*, vol. 27, no. 8, pp. 1005–1018, 1994.

[5] P. W. Huang and Y. R. Jean, "Using 2D C^+-strings as spatial knowledge representation for image database systems," *Pattern Recognition*, vol. 27, no. 9, pp. 1249–1257, 1994.

[6] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 850–863, September 1993.

[7] D. J. Kahl, A. Rosenfeld, and A. Danker, "Some experiments in point pattern matching," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 10, pp. 105–116, February 1980.

[8] J. K. Cheng and T. S. Huang, "Image registration by matching relational structures," *Pattern Recognition*, vol. 17, no. 1, pp. 149–159, 1984.

[9] S. P. Smith and A. K. Jain, "Chord distribution for shape matching," *Computer Graphics and Image Processing*, vol. 20, pp. 259–271, 1982.

[10] W. Richards and D. D. Hoffman, "Codon representation on closed 2D shapes," *Computer Vision, Graphics, and Image Processing*, no. 31, pp. 265–281, 1985.

[11] C. T. Zahn and R. Z. Roskies, "Fourier descriptors for plane closed curves," *IEEE Transactions on Computer*, vol. C-21, no. 1, pp. 269–281, 1972.

[12] S. A. Dudani, K. J. Breeding, and R. B. McGhee, "Aircraft identification by moment invariants," *IEEE Transactions on Computers*, vol. C-26, pp. 39–45, January 1977.

[13] C. P. Lam, J. K. Wu, and B. Mehtre, "STAR - a system for trademark archival and retrieval," in *Proceedings of the 2nd Asian Conference on Computer Vision*, vol. III, pp. 214–217, December 1995, Singapore.

[14] J. Weng, "SHOSLIF: A framework for sensor-based learning for high-dimensional complex systems," in *IEEE Workshop on Architectures for Semantic Modeling and situation analysis in Large Complex Systems*, August 27-29 1995. Invited Paper.

[15] T. P. Minka and R. W. Picard, "Interactive learning using a 'society of models'," in *CVPR96*, pp. 447–452, 1996.

[16] USPTO, *Basic Facts About Registering a Trademark.* US Patent and Trademark Office, http://www.uspto.gov/web/trad_reg_info/basic_facts.html, 1995.

[17] USPTO, *Trademark Manual of Examining Procedures.* US Patent and Trademark Office, http://www.uspto.gov/web/info/ftp.html, 1995.

[18] Thomson & Thomson, 500 Victory Road, North Quincy, MA, *The TRADEMARKSCAN Design Code Manual*, August 1995.

[19] T. Igarashi, Ed., *World Trademarks and Logotypes.* Tokyo: Graphic-sha, 1983.

[20] T. Igarashi, Ed., *World Trademarks and Logotypes II : A Collection of International Symbols and their Applications.* Tokyo: Graphic-sha, 1987.

[21] *Collection of Trademarks and Logotypes in Japan.* Tokyo: Graphic-sha, 1973.

[22] G. L. Lohse, K. Biolsi, N. Walker, and H. H. Rueter, "A classification of visual representations," *Communications of the ACM*, vol. 37, pp. 36–49, December 1994.

[23] A. K. Jain and A. Vailaya, "Image retrieval using color and shape," *Pattern Recognition*, vol. 29, pp. 1233–1244, August 1996.

[24] W. Niblack and J. Yin, "A pseudo-Distance measure for 2D shapes based on turning angle," *Proc 2nd IEEE Int. Conf. on Image Processing (ICIP)*, vol. III, pp. 352–355, 1995.

[25] Y. Cui, D. Swets, and J. Weng, "Learning-based hand sign recognition using SHOSLIF-M," *Proc. 5th Int. Conf. on Computer Vision (ICCV)*, pp. 631–636, 1995.

[26] A. Vailaya, "Shape-based image retrieval," Master's thesis, Department of Computer Science, Michigan State University, 1996.

[27] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, pp. 679–698, November 1986.

[28] A. Jain, Y. Zhong, and S. Lakshmanan, "Object matching using deformable templates," *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 18, pp. 267–279, March 1996.

[29] A. Vailaya, Y. Zhong, and A. K. Jain, "A Hierarchical System for Efficient Image Retrieval," in 13*th International Conference on Pattern Recognition*, (Vienna), pp. C356–360, August 1996.

[30] Y. Amit, U. Grenander, and M. Piccioni, "Structural image restoration through deformable template," *J. American Statistical Association*, vol. 86, pp. 376–387, June 1991.