Multi-modal retrieval of trademark images using global similarity*

S. Ravela R. Manmatha
Multimedia Indexing and Retrieval Group
Center for Intelligent Information Retrieval
University of Massachusetts, Amherst, MA 01003
Email: {ravela,manmatha}@cs.umass.edu

Abstract

In this paper a system for multi-modal retrieval of trademark images is presented. Images are characterized and retrieved using associated text and visual appearance. A user initiates retrieval for similar trademarks by typing a text query. Subsequent searches can be performed by visual appearance or using both appearance and text information. Textual information associated with trademarks is searched using the INQUERY search engine. Images are searched visually using a method for global image similarity by appearance developed in this paper. Images are filtered with Gaussian derivatives and geometric features are computed from the filtered images. The geometric features used here are curvature and phase. Two images may be said to be similar if they have similar distributions of such features. Global similarity may, therefore, be deduced by comparing histograms of these features. This allows for rapid retrieval. The system's performance on a database of 2000 trademark images is shown. A trademark database obtained from the US Patent and Trademark Office containing 63000 design only trademark images and text is used to demonstrate scalability of the image search method and multi-modal retrieval.

1 Introduction

Retrieval of similar trademarks is an interesting application for multimedia information retrieval. Consider the following example. The US Patent and Trademark Office has a repository that has to be searched

*This material is based on work supported in part by the National Science Foundation, Library of Congress and Department of Commerce under cooperative agreement number EEC-9209623, in part by the United States Patent and Trademarks Office and the Defense Advanced Research Projects Agency/ITO under ARPA order number D468, issued by ESC/AXS contract number F19628-95-C-0235, in part by the National Science Foundation under grant IRI-9619117 and in part by NSF Multimedia CDA-9502639. Any opinions, findings and conclusions or recommendations expressed in this material are the author(s) and do not necessarily reflect those of the sponsors.

for conflicting (similar) trademarks before one can be awarded to a company or individual. There are several issues that make this task attractive for multi-modal information retrieval techniques. First, current searches are labour intensive. The number of trademarks stored is enormous and examiners have to leaf through large number of trademarks before making a decision. Second, there is a distinct notion of visual similarity used to compare trademarks. This is usually a decisive factor in an award decision. Third, there is readily available text information describing and categorizing a trademark. A system that automates these functions and helps the examiner decide faster would be immensely valuable. Clearly trademarks need to be searched both by text and image content. Text retrieval is a better understood problem, and there are several search engines that are applicable.

However, the indexing and retrieval of images using their content is a difficult problem. A person using an image retrieval system usually seeks to find semantically relevant information. For example, a person may be looking for a picture of a leopard from a certain viewpoint. Or alternatively, the user may require a picture of Abraham Lincoln from a particular viewpoint. Since the automatic segmentation of an image into objects is a difficult and unsolved problem in computer vision, inferring semantic information from image content is difficult to do. However, many image attributes like color, texture, shape and "appearance" are often directly correlated with the semantics of the problem. For example, logos or product packages (e.g., a box of Tide) have the same color wherever they are found. The coat of a leopard has a unique texture while Abraham Lincoln's appearance is uniquely defined. These image attributes can often be used to index and retrieve images.

In this paper, a system for multi-modal retrieval combining textual information and visual appearance is presented. The system combines text search using INQUERY [2] and image search. The image search was originally developed for general (heterogeneous) grey-level image collections [18]. Here, it is applied to trademark images. Trademark images are large binary images rather than grey-level images. Trademark images may consist of geometric designs, more realistic pictures (for example, animals and people) as well as abstract images making them a challenging domain. Trademark images are also an example of a domain where there is an actual user need to find "similar" trademarks to avoid conflicts. Trademarks for this paper were obtained from the US Patent and Trademark office. The 63000 design trademarks used here contain images of trademarks and associated text describing the trademark.

Multi-modal retrieval begins with a user requesting trademarks that match a text query. The INQUERY search engine is used to find trademarks whose associated text match the query. The images associated with these trademarks are then displayed. Once an initial query is processed subsequent searches can be carried out by selecting the returned images and submitting them for retrieval by visual appearance or a combination of visual appearance and associated text.

INQUERY is a well known search engine for retrieving text which is based on a probabilistic retrieval

model called an inference net The reader is referred to [2] for details about the INQUERY engine. The current paper focuses on visual appearance representation, its quantitative evaluation with respect to trademarks, scalability to a large collection, and feasibility to multi-modal retrieval.

The visual appearance of an image is characterized here using the shape of the intensity surface. The images are filtered with Gaussian derivatives and geometric features are computed from the filtered images. The geometric features used here are the image shape index (which is a ratio of curvatures of the three dimensional intensity surface) and the local orientation of the gradient. Two images are said to be similar if they have similar distributions of such features. The images are, therefore, ranked by comparing histograms of these features. Recall/Precision results with this method is tabulated with a database of about 2000 trademark images. Then multi-modal retrieval is demonstrated on a collection of 63000 trademark images.

The rest of the paper is organized as follows. Section 2 provides some background on the image retrieval area as well as on the appearance matching framework used in this paper. Section 3 surveys related work in the literature. In section 4, the notion of appearance is developed further and characterized using Gaussian derivative filters and the derived global representation is discussed. Section 5 shows how the representation may be scaled for multi-modal retrieval from a database of about 63,000 trademark images. A discussion and conclusion follows in Section 6.

2 Motivation and Background

The different image attributes like color, texture, shape and appearance have all been used in a variety of systems for retrieving images similar to a query image (see 3 for a review). Systems like QBIC [6] and Virage [5] allow users to combine color, texture and shape to retrieve a database of general images. One weakness of such a system is that attributes like color do not have direct semantic correlates when applied to a database of general images. For example, say a picture of a red and green parrot is used to retrieve images based on their similarity in color with it. The retrievals may include other parrots and birds as well as red flowers with green stems and other images. While this is a reasonable result when viewed as a matching problem, clearly it is not a reasonable result for a retrieval system. The problem arises because color does not have a good correlation with semantics when used with general images. However, if the domain or set of images is restricted to say flowers, then color has a direct semantic correlate and is useful for retrieval (see [3] for an example).

Some attempts have been made to retrieve objects using their shape [6, 22]. For example, the QBIC system [6], developed by IBM, matches binary shapes. It requires that the database be segmented into objects. Since automatic segmentation is an unsolved problem, this requires the user to manually outline the objects in the database. Clearly this is not desirable or practical.

Except for certain special domains, all methods based on shape are likely to have the same problem. An object's appearance depends not only on its three dimensional shape, but also on the object's albedo, the viewpoint from which it is imaged and a number of other factors. It is non-trivial to separate the different factors constituting an object's appearance and it is usually not possible to separate an object's three dimensional shape from the other factors. For example, the face of a person has a unique appearance that cannot just be characterized by the geometric shape of the 'component parts'. In this paper a characterization of the shape of the intensity surface of imaged objects is used for retrieval. The experiments conducted show that retrieved objects have similar visual appearance, and henceforth an association is made between 'appearance' and the shape of the intensity surface.

Similarity can be computed using either local or global methods. In local similarity, a part of the query is used to match a part of a database image or images. One approach to computing local similarity [18] is to have the user outline the salient portions of the query (eg. the wheels of a car or the face of a person) and match the outlined portion of the query with parts of images in the database. Although, the technique works well in extracting relevant portions of objects embedded against backgrounds it is slow. The slow speed stems from the fact that the system must not only answer the question "is this image similar" but also the question "which part of the image is relevant".

This paper focuses on a representation for computing global similarity. That is, the task is to find images that, as a whole, appear visually similar. The utility of global similarity retrieval is evident, for example, in finding similar scenes or similar faces in a face database. Global similarity also works well when the object in question constitutes a significant portion of the image.

2.1 Appearance based retrieval

The image intensity surface is robustly characterized using features obtained from responses to multi-scale Gaussian derivative filters. Koenderink [14] and others [7] have argued that the local structure of an image can be represented by the outputs of a set of Gaussian derivative filters applied to an image. That is, images are filtered with Gaussian derivatives at several scales and the resulting response vector locally describes the structure of the intensity surface. By computing features derived from the local response vector and accumulating them over the image, robust representations appropriate to querying images as a whole (global similarity) can be generated. One such representation uses histograms of features derived from the multi-scale Gaussian derivatives. Histograms form a global representation because they capture the distribution of local features (A histogram is one of the simplest ways of estimating a non parametric distribution). This global representation can be efficiently used for global similarity retrieval by appearance and retrieval is very fast.

The choice of features often determines how well the image retrieval system performs. Here, the task is to robustly characterize the 3-dimensional intensity surface. A 3-dimensional surface is uniquely de-

termined if the local curvatures everywhere are known. Thus, it is appropriate that one of the features be local curvature. The principal curvatures of the intensity surface are invariant to image plane rotations, monotonic intensity variations and further, their ratios are in principle insensitive to scale variations of the entire image. However, spatial orientation information is lost when constructing histograms of curvature (or ratios thereof) alone. Therefore we augment the local curvature with local phase, and the representation uses histograms of local curvature and phase.

Local principal curvatures and phase are computed at several scales from responses to multi-scale Gaussian derivative filters. Then histograms of the curvature ratios [13, 4] and phase are generated. Thus, the image is represented by a single vector (multi-scale histograms). During run-time the user presents an example image as a query and the query histograms are compared with the ones stored, and the images are then ranked and displayed in order to the user.

2.2 The choice of domain

There are two issues in building a content based image retrieval system. The first issue is technological, that is, the development of new techniques for searching images based on their content. The second issue is user or task related, in the sense of whether the system satisfies a user need. While a number of content based retrieval systems have been built ([6, 5]), it is unclear what the purpose of such systems is and whether people would actually search in the fashion described.

In this paper we describe how the techniques described here may be scaled to retrieve images from a database of about 63000 trademark images provided by the US Patent and Trademark Office. This database consists of all (at the time the database was provided) the registered trademarks in the United States which consist only of designs (i.e. there are no words in them). Trademark images are a good domain with which to test image retrieval. First, there is an existing user need: trademark examiners do have to check for trademark conflicts based on visual appearance. That is, at some stage they are required to look at the images and check whether the trademark is similar to an existing one. Second, trademark images may consist of simple geometric designs, pictures of animals or even complicated designs. Thus, they provide a test-bed for image retrieval algorithms. Third, there is text associated with every trademark and the associated text maybe used in a number of ways. One of the problems with many image retrieval systems is that it is unclear where the example or query image will come from. In this paper, the associated text is used to provide an example or query image. In addition associated text can also be combined with image searches. Using trademark images does have some limitations. First, we are restricted to binary images (albeit large ones). As shown later in the paper, this does not create any problems for the algorithms described here. Second, in some cases the use of abstract images makes the task more difficult. Others have attempted to get around it by restricting the trademark images to geometric designs [9].

3 Related Work

Several authors have tried to characterize the appearance of an object via a description of the intensity surface. In the context of object recognition [21] represent the appearance of an object using a parametric eigen space description. This space is constructed by treating the image as a fixed length vector, and then computing the principal components across the entire database. The images therefore have to be size and intensity normalized, segmented and trained. Similarly, using principal component representations described in [11] face recognition is performed in [26]. In [24] the traditional eigen representation is augmented by using most discriminant features and is applied to image retrieval. The authors apply eigen representation to retrieval of several classes of objects. The issue, however , is that these classes are manually determined and training must be performed on each. The approach presented in this paper is different from all the above because eigen decompositions are not used at all to characterize appearance. Further, the method presented uses no learning and, does not require constant sized images. It should be noted that although learning significantly helps in such applications as face recognition, however, it may not be feasible in many instances where sufficient examples are not available. This system is designed to be applied to a wide class of images and there is no restriction per se.

In earlier work we showed that local features computed using Gaussian derivative filters can be used for local similarity, i.e. to retrieve parts of images [18]. Here we argue that global similarity can be determined by computing local features and comparing distributions of these features. This technique gives good results, and is reasonably tolerant to view variations. Schiele and Crowley [23] used such a technique for recognizing objects using grey-level images. Their technique used the outputs of Gaussian derivatives as local features. A multi-dimensional histogram of these local features is then computed. Two images are considered to be of the same object if they had similar histograms. The difference between this approach and the one presented by Schiele and Crowley is that here we use 1D histograms (as opposed to multi-dimensional) and further use the principal curvatures as the primary feature.

The use of Gaussian derivative filters to represent appearance is motivated by their use in describing the spatial structure [14] and its uniqueness in representing the scale space of a function [15, 12, 28, 25] The invariance properties of the principal curvatures are well documented in [7]. Nastar [20], has independently used the image shape index to compute similarity between images. However, in his work curvatures were computed only at a single scale. This is insufficient.

In the context of global similarity retrieval it should be noted that representations using moment invariants have been well studied [19]. In these methods global representation of appearance may involve computing a few numbers over the entire image. Two images are then considered similar if these numbers are close to each other (say using an L2 norm). We argue that such representations are not able to really capture the "appearance" of an image, particularly in the context of trademark retrieval where mo-

ment invariants are widely used. In other work [18] we compared moment invariants with the technique presented here and found that moment invariants work best for a single binary shape without holes in it, and, in general, fare worse than the method presented here. Jain and Vailaya [10] used edge angles and invariant moments to prune trademark collections and then use template matching to find similarity within the pruned set. Their database was limited to 1100 images.

Texture based image retrieval is also related to the appearance based work presented in this paper. Using Wold modeling, in [16] the authors try to classify the entire Brodatz texture and in [8] attempt to classify scenes, such as city and country. Of particular interest is work by [17] who use Gabor filters to retrieve texture similar images.

The earliest general image retrieval systems were designed by [6, 22]. In [6] the shape queries require prior manual segmentation of the database which is undesirable and not practical for most applications.

4 Global representation of appearance

Three steps are involved in order to computing global similarity. First, local derivatives are computed at several scales. Second, derivative responses are combined to generate local features, namely, the principal curvatures and phase and, their histograms are generated. Third, the 1D curvature and phase histograms generated at several scales are matched. These steps are described next.

A. Computing local derivatives: Computing derivatives using finite differences does not guarantee stability of derivatives. In order to compute derivatives stably, the image must be regularized, or smoothed or band-limited. A Gaussian filtered image $I_\sigma = I * G$ obtained by convolving the image I with a normalized Gaussian $G(\mathbf{r}, \sigma)$ is a band-limited function. Its high frequency components are eliminated and derivatives will be stable. In fact, it has been argued by Koenderink and van Doorn [14] and others [7] that the local structure of an image I at a given scale can be represented by filtering it with Gaussian derivative filters (in the sense of a Taylor expansion), and they term it the N-jet.

However, the shape of the smoothed intensity surface depends on the scale at which it is observed. For example, at a small scale the texture of an ape's coat will be visible. At a large enough scale, the ape's coat will appear homogeneous. A description at just one scale is likely to give rise to many accidental mis-matches. Thus it is desirable to provide a description of the image over a number of scales, that is, a scale space description of the image. It has been shown by several authors [15, 12, 28, 25, 7], that under certain general constraints, the Gaussian filter forms a unique choice for generating scale-space. Thus local spatial derivatives are computed at several scales.

B. Feature Histograms: The normal and tangential curvatures of a 3-D surface (X,Y,Intensity) are defined as [7]:

$$N(\mathbf{p},\sigma) = \left[\frac{I_x^2 I_{yy} + I_y^2 I_{xx} - 2 I_x I_y I_{xy}}{\left(I_x^2 + I_y^2\right)^{\frac{3}{2}}}\right](\mathbf{p},\sigma)$$

$$T(\mathbf{p},\sigma) = \left[\frac{\left(I_x^2 - I_y^2\right) I_{xy} + \left(I_{xx} - I_{yy}\right) I_x I_y}{\left(I_x^2 + I_y^2\right)^{\frac{3}{2}}}\right](\mathbf{p},\sigma)$$

Where $I_x(\mathbf{p},\sigma)$ and $I_y(\mathbf{p},\sigma)$ are the local derivatives of Image I around point $\mathbf{p}$ using Gaussian derivative at scale σ. Similarly $I_{xx}(\cdot,\cdot)$, $I_{xy}(\cdot,\cdot)$, and $I_{yy}(\cdot,\cdot)$ are the corresponding second derivatives. The normal curvature N and tangential curvature T are then combined [13] to generate a shape index as follows:

$$C(\mathbf{p},\sigma) = atan\left[\frac{N+T}{N-T}\right](\mathbf{p},\sigma)$$

The index value C is $\frac{\pi}{2}$ when $N = T$ and is undefined when either N and T are both zero, and is, therefore, not computed. This is interesting because very flat portions of an image (or ones with constant ramp) are eliminated. For example in Figure 1, the background in most of these images does not contribute to the curvature histogram. The curvature index or shape index is rescaled and shifted to the range $[0,1]$ as is done in [4]. A histogram is then computed of the valid index values over an entire image.

The second feature used is phase. The phase is simply defined as $P(\mathbf{p},\sigma) = atan2\left(I_y(\mathbf{p},\sigma), I_x(\mathbf{p},\sigma)\right)$. Note that P is defined only at those locations where C is and ignored elsewhere. As with the curvature index P is rescaled and shifted to lie between the interval $[0,1]$.

At different scales different local structures are observed and, therefore, multi-scale histograms are a more robust representation. Consequently, a feature vector is defined for an image I as the vector

$$V_i = \langle H_c(\sigma_1) \ldots H_c(\sigma_n), H_p(\sigma_1) \ldots H_p(\sigma_n) \rangle$$

where H_p and H_c are the curvature and phase histograms respectively. We found that using 5 scales gives good results and the scales are $1 \cdots 4$ in steps of half an octave.

C. Matching feature histograms: Two feature vectors are compared using normalized cross-covariance defined as

$$d_{ij} = \frac{V_i^{(m)} \cdot V_j^{(m)}}{\left\|V_i^{(m)}\right\| \left\|V_j^{(m)}\right\|}$$

where $V_i^{(m)} = V_i - mean(V_i)$.

Retrieval is carried out as follows. A query image is selected and the query histogram vector V_q is correlated with the database histogram vectors V_i using the above formula. Then the images are ranked by their correlation score and displayed to the user. In this implementation, and for evaluation purposes, the ranks are computed in advance, since every query image is also a database image.

4.1 Experiments

The curvature-phase method is evaluated on a small database of 2048 images obtained from the US Patent and Trademark Office (PTO). The images obtained from the PTO are large, binary and are converted to gray-level and reduced for the experiments. This smaller set is used because relevance judgments can be obtained relatively easily.

In the following experiments an image is selected and submitted as a query. The objective of this query is stated and the relevant images are decided in advance. Then the retrieval instances are gauged against the stated objective. In general, objectives of the form 'extract images similar in appearance to the query' will be posed to the retrieval algorithm. A measure of the performance of the retrieval engine can be obtained by examining the recall/precision table for several queries. Briefly, recall is the proportion of the relevant material actually retrieved and precision is the proportion of retrieved material that is relevant [27]. It is a standard widely used in the information retrieval community and is one that is adopted here.

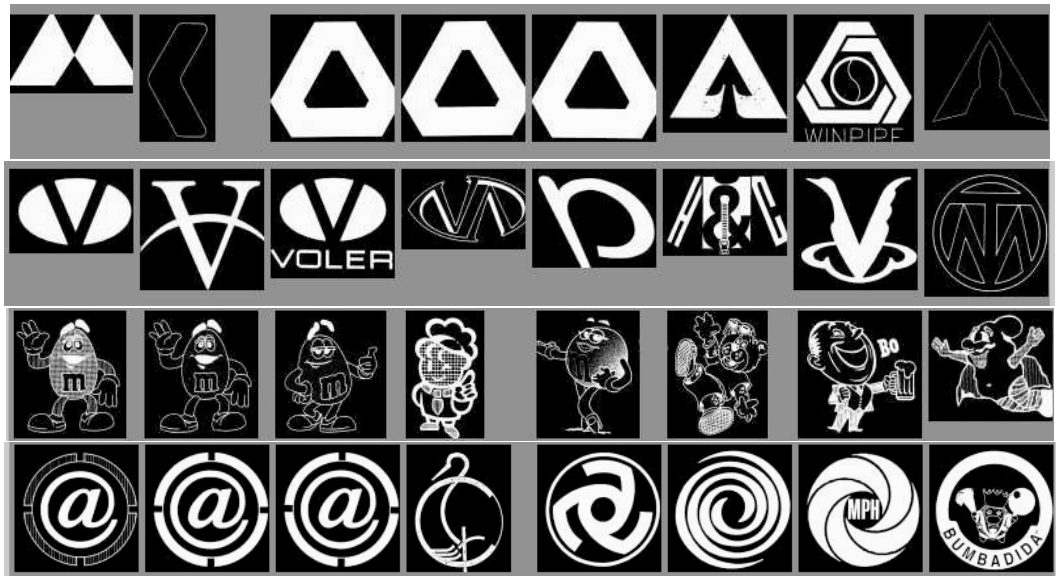

Figure 1: Trademark retrieval using Curvature and Phase

Queries were submitted for the purpose of computing recall/precision. The judgment of relevance is qualitative. For each query in both databases the relevant images were decided in advance. These were restricted to 48. The top 48 ranks were then examined to check the proportion of retrieved images that were relevant. All images not retrieved within 48 were assigned a rank equal to the size of the database.

Table 1: Precision at standard recall points for six Queries

Recall	0	10	20	30	40	50	60	70	80	90	100
Precision(trademark) %	100	93.2	93.2	85.2	76.3	74.5	59.5	45.5	27.2	9.0	9.0
Precision(assorted) %	100	92.6	90.0	88.3	87.0	86.8	83.8	65.9	21.3	12.0	1.4

average(trademark)	61.1%
average(assorted)	66.3%

That is, they are not considered retrieved. These ranks were used to interpolate and extrapolate precision at all recall points. In the case of assorted images relevance is easier to determine and more similar for different people. However in the trademark case it can be quite difficult and therefore the recall-precision can be subject to some error. The recall/precision results are summarized in Table 1 and both databases are individually discussed below.

Figure 1 shows the performance of the algorithm on the trademark images. Each strip depicts the top 8 retrievals, given the leftmost as the query. Most of the shapes have roughly the same structure as the query. Note that, outline and solid figures are treated similarly (see rows one and two in Figure 1). Six queries were submitted for the purpose of computing recall-precision in Table 1. Tests were also carried out with an assorted collection of 1561 grey-level images. These results are discussed elsewhere [1], and the recall/precision table is shown in Table 1.

While the queries presented here are not "optimal" with respect to the design constraints of global similarity retrieval, they are however, realistic queries that can be posed to the system. Mismatches can and do occur. The first is the case where the global appearance is very different. Second, mismatches can occur at the algorithmic level. Histograms coarsely represent spatial information and therefore will admit images with non-trivial deformations. The recall/precision presented here compares well with text retrieval. The time per retrieval is of the order of milli-seconds. In the next section we discuss the application of the presented technique to a database of 63000 images.

5 Trademark Retrieval

The system indexes 63,718 trademarks from the US Patent and Trademark office in the design only category. These trademarks are binary images. In addition, associated text consists of a design code that designates the type of trademark, the goods and services associated with the trademark, a serial number and a short descriptive text.

The system for browsing and retrieving trademarks is illustrated in Figure 2. The netscape/Java user interface has two search-able parts. On the left a panel is included to initiate search using text. Any or all of the fields can be used to enter a query. In this example, the text "Merriam Webster' is entered. All images associated with it are retrieved using the INQUERY [2] text search engine. The user can then use any of the example pictures to search for images that are similar visually or restrict it to images with

Table 2: Fields supporting the text query

Field	Description
Goods & services	The business this trademark is used in
Mark drawing code	All are of type DESIGN ONLY
Design code	An assigned code category
Serial number	Serial number assigned to trademark
File date	Date trademark application was filed
Registration number	Number assigned to trademark
Registration date	Date trademark was registered
Owner	Owner of the trademark
Description	A textual description of the trademark. Example, "The mark consists of the silhouette of an apple with a bite removed."
Section 44	No description available
Type of mark	All are of type TRADEMARK.
Register	Who registered the trademark
Affidavit text	The file numbers for affidavits filed
Live/ dead	Whether the trademark is active or not

relevant text, thereby combining the image and text searches. In the specific example shown, The second image is selected and retrieved results are displayed on the right panel. The user can then continue to search using any of the displayed pictures as the query.

Text was provided for each image in the collection of design only trademark category from the Patent and Trademark Office. This information contained specific fields such as the design code, the goods and services provided, the serial number, the manufacturer, among others. Table 2 lists all the fields associated with a trademark. These were indexed and used for retrieval using the INQUERY search engine [2]. The queries that can be submitted are conjunctive (and) of all the words that are entered in all the fields allowed within the interface with equal weighting. These fields are shown at the top of the left panel in Figure 2.

In this section we describe the curvature/phase histograms to retrieve visually similar trademarks and demonstrate searches using text and visual information. The following steps are performed to retrieve images.

Preprocessing: Each binary image in the database is first size normalized, by clipping. Then they are converted to gray-scale and reduced in size.

Computation of Histograms: Each processed image is divided into four equal rectangular regions. This is different than constructing a histogram based on pixels of the entire image. This is because in scaling the images to a large collection, we found that the added degree of spatial resolution significantly improves the retrieval performance. The curvature and phase histograms are computed for each tile at three scales

(1,4,8). A histogram descriptor of the image is obtained by concatenating all the individual histograms across scales and regions.

These two steps are conducted off-line.

Execution: The image search server begins by loading all the histograms into memory. Then it waits on a port for a query. A CGI client transmits the query to the server. Its histograms are matched with the ones in the database. The match scores are ranked and the top N requested retrievals are returned.

5.1 Examples

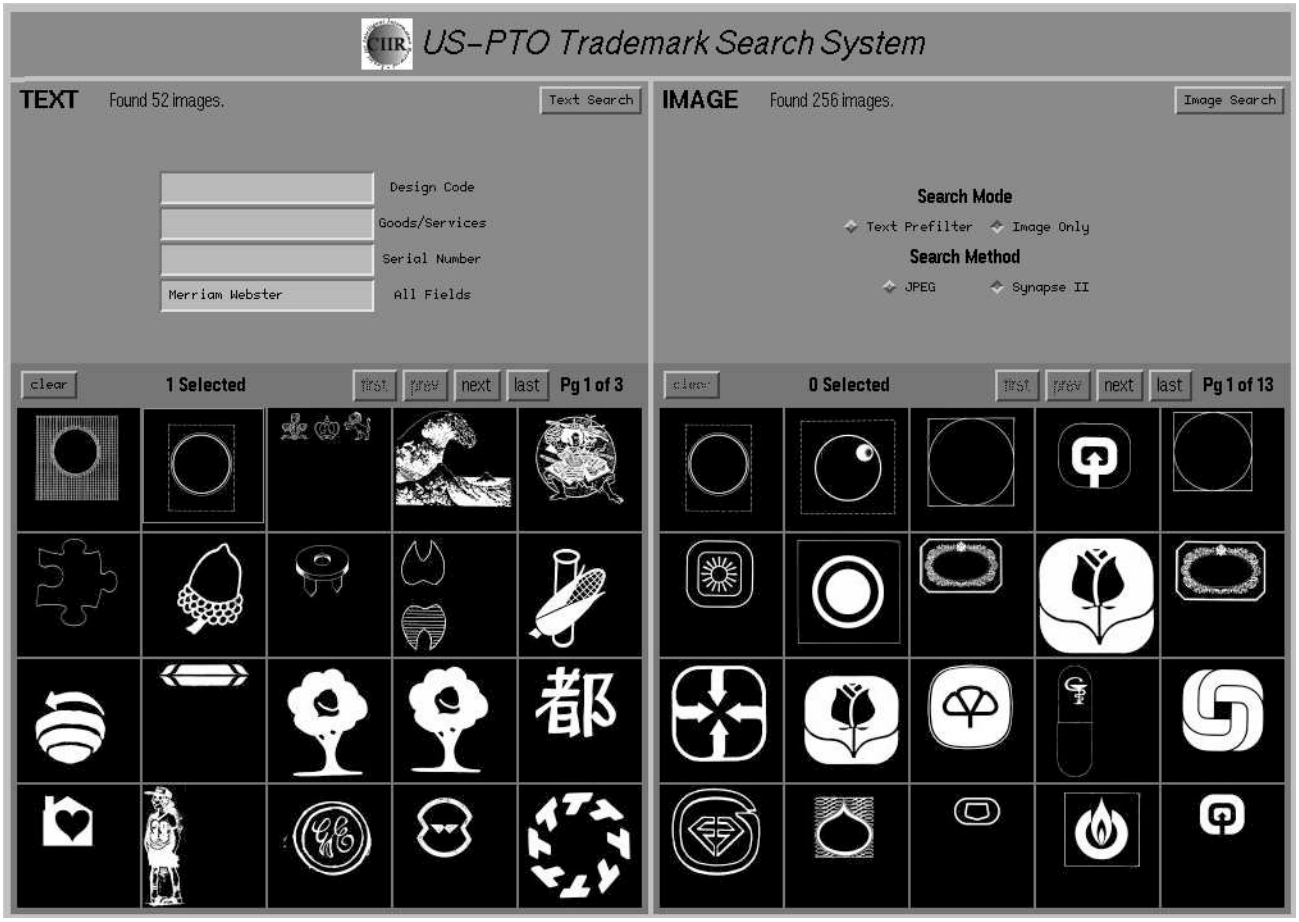

Figure 2: Retrieval in response to a "Merriam Webster" query

In Figure 2, the user typed in Merriam Webster in the text window. The system searches for trademarks which have either Merriam or Webster in th associated text and displays them. Here, the first two trademarks (first two images in the left window) belong to Merriam Webster. In this example, the user has chosen to 'click' the second image and search for images of similar trademarks. This search is based entirely on the image and the results are displayed in the right window in rank order. Retrieval takes a few seconds and is done by comparing histograms of all 63,718 trademarks on the fly.

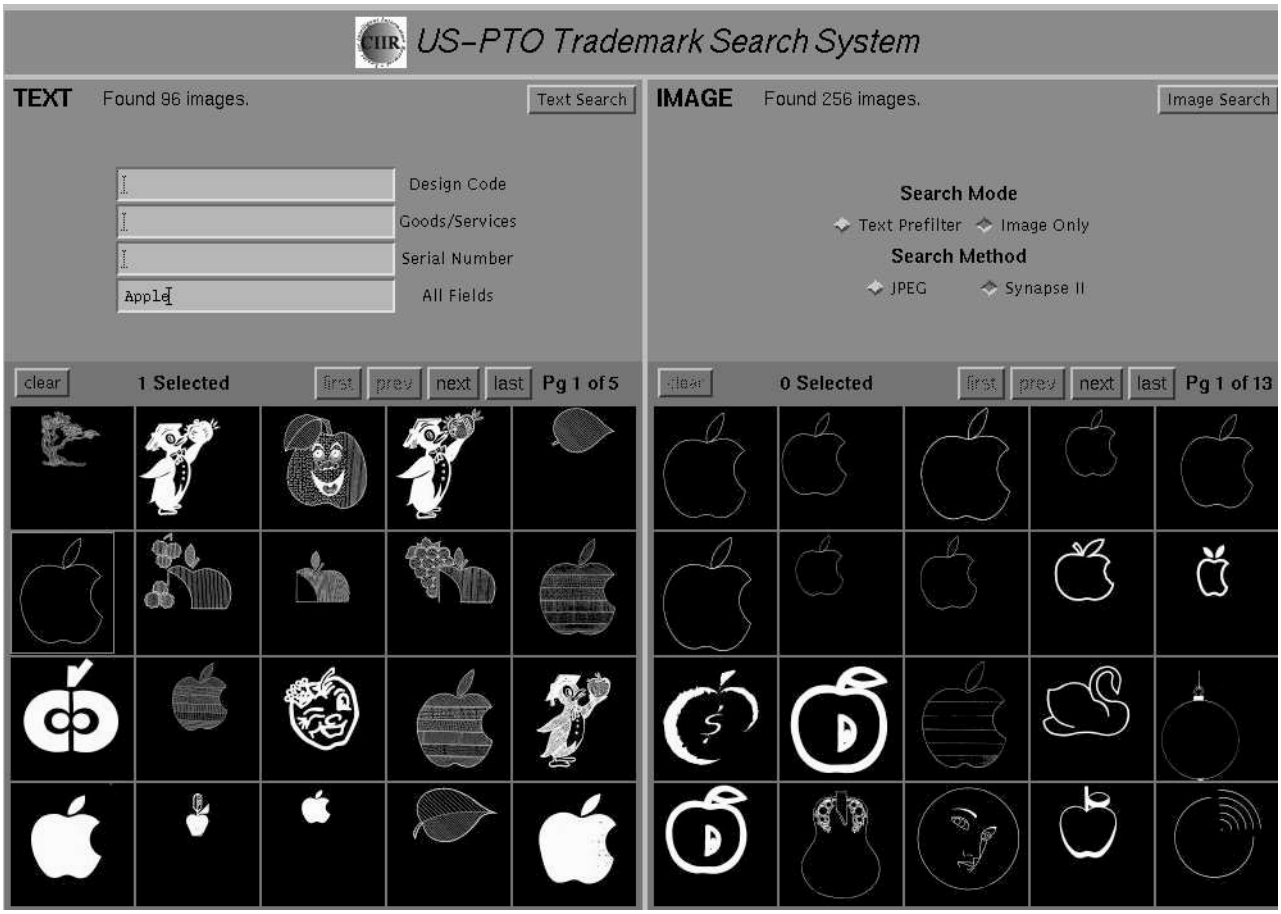

Figure 3: Retrieval in response to the query “Apple”

The original image is returned as the first result (as it should be). The images in positions 2,3 and 5 in the second window all contain circles inside squares and this configuration is similar to that of the query. Most of the other images are of objects contained inside a roughly square box and this is reasonable considering that similarity is defined on the basis of the entire image rather than a part of the image.

The second example is shown in Figure 3. Here the user has typed in the word Apple. The system returns trademarks associated with the word Apple. The user queries using Apple computer’s logo (the image in the second row, first column of the first window). Images retrieved in response to this query are shown in the right window. The first eight retrievals are all copies of Apple Computer’s trademark (Apple used the same trademark for a number of other goods and so there are multiple copies of the trademark in the database). Trademarks number 9 and 10 look remarkably similar to Apple’s trademark. They are considered valid trademarks because they are used for goods and services in areas other than computers. Trademark 13 is another version of Apple Computer’s logo but with lines in the middle. Although somewhat visually different it is still retrieved in the high ranks. Image 14 is an interesting example of a mistake made by the system. Although the image is not of an apple, the image has similar distributions of curvature and phase as is clear by looking at it.

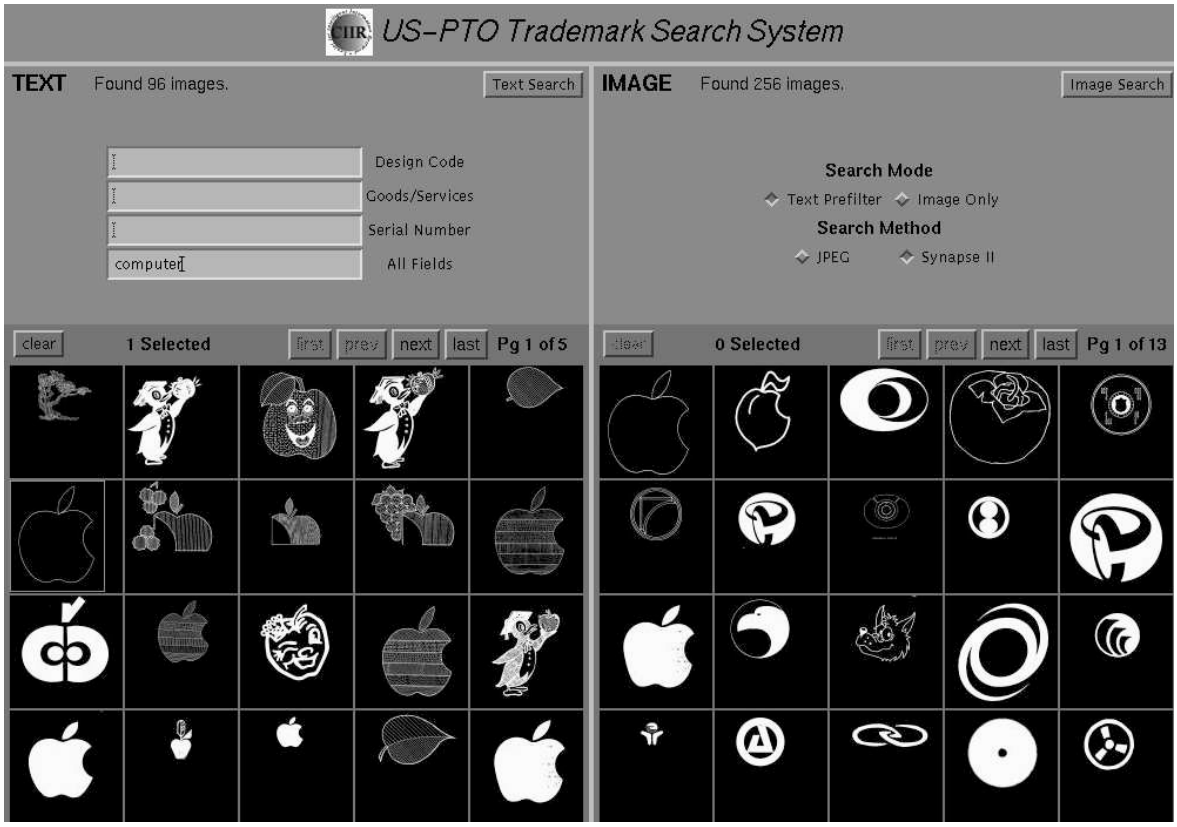

Figure 4: Retrieval in response to the query "Apple" limited to text searches

The third example demonstrates combining text and visual appearance for searching. We use the same apple image obtained in the previous image as the image query. However, in the text box we now type "computer" and turn the text combination mode on. We now search for trademarks which are visually similar to the apple query image but also have the words computer associated with them. The results are shown in Figure 4 on the right-hand side. Notice that the first image is the same as the query image. The second image is an actual conflict. The image is a logo which belongs to the Atlanta Macintosh User's Group. The text describes the image as a peach but visually one can see how the two images may be confused with each other (which is the basis on which trademark conflicts are adjudicated). This example shows that it does not suffice to go by the text descriptions alone and image search is useful for trademarks. Notice that the fourth image which some people describe as an apple and others as a tomato is also described in the text as an apple.

The system has been tried on a variety of different examples of both two dimensional and three dimensional pictures of trademarks and had worked quite well. Clearly, there are issues of how quantitative results can be obtained for such large image databases (it is not feasible for a person to look at every image in the database to determine whether it is similar). In future work, we hope to evolve a mechanism for quantitative testing on such large databases. It will also be important to use more of the textual information

to determine trademark conflicts.

6 Conclusions and Limitations

This paper demonstrates multi-modal retrieval of trademarks. Both text and images separately and together are used to retrieve trademarks. Text search is done using INQUERY while image search is done on the basis of similarity in visual appearance. Visual appearance is characterized using filter responses to Gaussian derivatives over scale space. In addition, we claim that global representations are better constructed by representing the distribution of robustly computed local features. The paper shows that it is not sufficient to use text or image search alone to retrieve trademarks.

Currently we are investigating three issues. First is to scale the database up to about 600000 images. The second is to incorporate user feedback or preferences of retrieved images. The third is to combine text retrieval and image retrieval in a principled manner.

References

[1] *On Computing Global Similarity in Images*, Oct 1998.

[2] J. P. Callan, W. B. Croft, and S. M. Harding. The inquery retrieval system. In *Proceedings of the 3rd International Conference on Database and Expert System Applications*, pages 78–83, 1992.

[3] M. Das, R. Manmatha, and E. M. Riseman. Indexing flowers by color names using domain knowledge-driven segmentation. In *In the Proc. of the 4th IEEE Workshop on Applications of Computer Vision (WACV'98), Princeton, NJ.*, pages 94–99, Oct 1998.

[4] Chitra Dorai and Anil Jain. Cosmos - a representation scheme for free form surfaces. In *Proc. 5th Intl. Conf. on Computer Vision*, pages 1024–1029, 1995.

[5] J.R. Bach et al. The virage image search engine: An open framework for image management. In *SPIE conf. on Storage and Retrieval for Still Image and Video Databases IV*, pages 133–156, 1996.

[6] Myron Flickner et al. Query by image and video content: The qbic system. *IEEE Computer Magazine*, pages 23–30, Sept. 1995.

[7] L. Florack. *The Syntactical Structure of Scalar Images*. PhD thesis, University of Utrecht, Utrecht, Holland, 1993.

[8] M. M. Gorkani and R. W. Picard. Texture orientation for sorting photos 'at a glance'. In *Proc. 12th Int. Conf. on Pattern Recognition*, pages A459–A464, October 1994.

[9] K. Shields J. P. Eakins and J. M. Boardman. Artisan - a shape retrieval system based on boundary family indexing. In *In Proc. SPIE conf. on Storage and Retrieval for Image and Video Databases IV, vol. 2670, San Jose*, pages 17–28, Feb 1996.

[10] A. K. Jain and A. Vailaya. Shape-based retrieval: A case study with trademark image databases. *Pattern Recognition*, 31(9):1369–1390, 1998.

[11] M Kirby and L Sirovich. Application of the kruhnen-loeve procedure for the characterization of human faces. *IEEE Trans. Patt. Anal. and Mach. Intel.*, 12(1):103–108, January 1990.

[12] J. J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–396, 1984.

[13] J. J. Koenderink and A. J. Van Doorn. Surface shape and curvature scales. *Image and Vision Computing*, 10(8), 1992.

[14] J. J. Koenderink and A. J. van Doorn. Representation of local geometry in the visual system. *Biological Cybernetics*, 55:367–375, 1987.

[15] Tony Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, 1994.

[16] Fang Liu and Rosalind W Picard. Periodicity, directionality, and randomness: Wold features for image modeling and retrieval. *IEEE Trans. PAMI*, 18(7):722–733, July 1996.

[17] W. Y. Ma and B. S. Manjunath. Texture-based pattern retrieval from image databases. *Multimedia Tools and Applications*, 2(1):35–51, January 1996.

[18] R. Manmatha, S. Ravela, and Y. Chitti. On computing local and global similarity in images. In *Proc. SPIE conf. on Human and Electronic Imaging III,*, 1998.

[19] B.M. Methre, M.S. Kankanhalli, and W.F. Lee. Shape Measures for Content Based Image Retrieval: A Comparison. *Information Processing and Management*, 33, 1997.

[20] C. Nastar. The image shape spectrum for image retrieval. Technical Report Technical Report 3206, INRIA, June 1997.

[21] S. K. Nayar, H. Murase, and S. A. Nene. Parametric appearance representation. In *Early Visual Learning*. Oxford University Press, February 1996.

[22] A. Pentland, R. W. Picard, and S. Sclaroff. Photobook: Tools for content-based manipulation of databases. In *Proc. Storage and Retrieval for Image and Video Databases II,SPIE*, volume 185, pages 34–47, 1994.

[23] Bernt Schiele and James L. Crowley. Object recognition using multidimensional receptive field histograms. In *Proc. 4th European Conf. Computer Vision*, Cambridge, U.K., April 1996.

[24] D. L. Swets and J. Weng. Using discriminant eigen features for retrieval. *IEEE Trans. Patt. Anal. and Mach. Intel.*, 18:831–836, August 1996.

[25] Bart M. ter Har Romeny. *Geometry Driven Diffusion in Computer Vision*. Kluwer Academic Publishers, 1994.

[26] M. Turk and A. Pentland. Eigenfaces for recognition. *J. of Cognitive NeuroScience*, 3:71–86, 1991.

[27] C. J. van Rijsbergen. *Information Retrieval*. Butterworths, 1979.

[28] A. P. Witkin. Scale-space filtering. In *Proc. Intl. Joint Conf. Art. Intell.*, pages 1019–1023, 1983.