Appearance-based Global Similarity Retrieval*

S. Ravela and Chen Luo
Multimedia Indexing and Retrieval Group
Center for Intelligent Information Retrieval
University of Massachusetts, Amherst, MA 01003
Email: ravela@cs.umass.edu

Abstract

Visual appearance is an important cue to judge similarity. We readily classify objects that share a visual appearance as similar, and reject those that do not. The hypothesis is that, image intensity surface features can be used to compute appearance similarity. In the first part of this paper a technique to compute global appearance similarity is described. Images are filtered with Gaussian derivatives to compute two features, namely, local curvatures and orientation. Global image similarity is deduced by comparing distributions of these features. This technique is evaluated on a heterogeneous collection 1600 images. The results support the hypothesis in that images similar in appearance are ranked close together. In the second part of this paper appearance-based retrieval is applied to trademarks. Trademarks are generally binary images containing a single mark against a texture-less background. While moments have been argued as a representation, however, we find that appearance-based retrieval yields better results. Two small databases containing a set of 2,345 parametrically generated shapes, and 10,745 trademarks from the US Patent and Trademark office are used for evaluation. Then a system to retrieve a US Patent and Trademark Office(PTO) trademark database containing 68,000 binary images with textual information is discussed. Text and appearance features are jointly (or independently) queried to retrieve images. Text retrieval is performed using INQUERY and image retrieval using global appearance similarity.

1 Introduction

Libraries of the future will be increasingly digital and will need good tools to navigate, search and retrieve relevant information. While search engines for ASCII text exist, good ways of retrieving images are still not available. In certain instances, textual annotations could be used to index and retrieve images. However, in many cases, textual annotations may not be available and expensive to build. Most importantly, they may

*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.

be inadequate. Annotations alone cannot capture the expressiveness of an image or anticipate the different purposes an image may be used for. Image content itself needs to be used; independently or as a component of a multi-modal system.

Indexing image content is difficult because users seek to find semantically relevant information. For example, a person may be looking for a picture of "Mahatma Gandhi giving a speech". Such queries require solutions to automatic image segmentation, object and context recognition. These are as yet unsolved problems. Therefore, inferring semantic information from image content is as yet difficult to do.

However, in many instances, image attributes like color, texture, shape and "appearance" are 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. Hence color can be used. Fashion designers can use texture and color to search through fabric texture collections. Gandhiji's appearance is uniquely defined. One could use an example of his face to search and browse through the retrieved results.

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. A number of content based retrieval systems have been built using image attributes like color, texture, shape and appearance (see Section 2 for a review). Early systems like QBIC [6] and Virage [5] allow users to combine color, texture and shape to retrieve a database of general images. It is unclear what the purpose of such systems is and whether people would actually search in the fashion described. 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 color similarity. 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 [2] 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 when applied to heterogeneous collections.

1.1 Appearance-based Global Similarity

In the first part of this paper a technique to retrieve images from heterogenous gray level collections is discussed. We develop a method that can be used to find images that share a visual appearance. Visual appearance is an important cue. We readily classify images that share an appearance as similar and reject those that do not. It is distinct from other image attributes. It is not dependent on color because appearance similarity is evident in gray level images. It is not dependent on binary shape, because appearance similarity is evident in images containing no distinct shapes. It is distinct (but somewhat related) from texture because primitive textures need not be present in an image to judge appearance similarity.

It is hard to precisely quantify visual appearance. 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'.

A hypothesis is that the image intensity surface has features that can be used to compute appearance similarity. However, finding these features is non-trivial. Intensity alone cannot be used as a feature, and systems that have used them are either expensive, inapplicable to a heterogeneous domain or intolerant to coordinate deformations and illumination variations. Here, appearance features are developed by considering the differential properties of surfaces and experiments show that images that share an appearance can be retrieved from heterogeneous collections.

A framework for appearance-based retrieval is developed to compute global appearance similarity. That 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, such as trademarks. In order to compute global appearance similarity, features are extracted from pixel neighborhoods and, their distribution over the image are compared. Histograms are used to represent distributions of features and correlation is used to compare histograms.

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 (X,Y, Intensity). A 3-dimensional surface is uniquely determined 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 orientation, and the representation uses histograms of local curvature and orientation.

This technique is demonstrated on two small collections to obtain an evaluation. The first consists of 1600 grey level images of objects such as cars, faces, apes and other miscellaneous objects. The conditions under which these images were taken is unknown.

The second collection is a set of 2345 parametrically generated binary (black and white) shapes of ellipses (including circles), triangles, rectangles (including squares), and pentagons. Images are generated for various values of the parameters that define the shape of the polygon. The purpose of this database is three fold. First, to test whether this technique extends to binary images. There are important applications such as trademark retrieval, which often contain binary images and in many instances contain a geometric shape against a texture-less background. Second, to test whether the proposed method can be used to find similar shapes. That is, does visual appearance extend to shape information. In the context of binary trademarks moment invariants are widely used to compute global similarity [19]. The third purpose of this collection is to provide a test bed to compare the algorithm developed here with moments. It is shown that the proposed method performs significantly better than moments.

1.2 Trademark Retrieval

In the second part of this paper we describe how the techniques described here may be applied to a real user need. The US Patent and Trademark Office has a huge repository that is searched for conflicting (similar) trademarks before one can be awarded to an applicant. Determining conflict is largely visual. A submission in the same category is in conflict if it is visually similar to one that has been awarded. Determining conflict is also labor intensive. Examiners have to leaf through large number of trademarks and the associated textual descriptions of them before making a decision. A system that even partly automates these functions by exploiting text and image information would be immensely valuable. From a research perspective, 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.

A collection of 63,718 trademark images with associated text is 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).

A framework for multi-modal retrieval is developed to retrieve trademarks. All searches begin with a text query such as “apple”. This provides a solution the the problem “where does the example image come from ?”, also referred to as the page zero problem. The INQUERY [1] search engine is used to find trademarks whose associated text match the query. Subsequent searches can use the appearance attribute independently or in conjunction with text. For example, one can pick a picture of an apple and restrict it with the word “computer”.

At the time of this writing relevance judgments for this collection are not available. Hence it is not possible to evaluate the performance. Instead, we use a collection of 10,745 trademarks from the UK Patent and Trademark office, provided by Dr. John Eakins at the University of Northumbria at Newcastle, UK. These images come with a set of 24 queries for which relevance judgments have been obtained from various examiners. This collection is tested with exactly the same parameters used to build the trademark retrieval system. Hence, a modest evaluation can be obtained.

The rest of the paper is organized as follows. Section 2 surveys related work in the literature. In section 3, the notion of appearance is developed further and characterized using Gaussian derivative filters and the derived global representation is discussed. Section 4 shows how the representation may be scaled to retrieve images from a database of about 63,718 trademark images. A discussion and conclusion follows in Section 5.

2 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 [27]. In [25] 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 [24] 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, 29, 26] 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 moment 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.

3 Global appearance similarity

It has been argued that the image intensity surface is robustly characterized using features obtained from responses to multi-scale Gaussian derivative filters. Koenderink [14] and others [7] have shown 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, the response vector locally describes the structure of the intensity surface.

There are several features that can be constructed from derivative vectors. The choice of these features depends on several factors, primary among which is a consideration to the factors affecting appearance. As has been argued earlier, local curvatures and orientation have nice properties with respect to deformations and illumination changes. Principal curvatures and orientation are computed at several scales from responses to multi-scale Gaussian derivative filters.

Then histograms of the curvature ratios [13, 3] and orientation are generated and compared to compute global similarity. 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. 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.

Thus, three steps are involved to compute global similarity. First, local derivatives are computed at several scales. Second, derivative responses are combined to generate local features, namely, the principal curvatures and orientation and, their histograms are generated. Third, the 1D curvature and orientation histograms are 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 mismatches. 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, 29, 26, 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(middle-row), the background in most of these face 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 [3]. A histogram is then computed of the valid index values over an entire image.

The second feature used is orientation. The orientation is simply defined as $P(\mathbf{p},\sigma) = atan2(I_y(\mathbf{p},\sigma), I_x(\mathbf{p},\sigma))$. 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 $V = \langle H_c(\sigma_1)\ldots H_c(\sigma_n), H_p(\sigma_1)\ldots H_p(\sigma_n)\rangle$ where H_p and H_c are the curvature and orientation 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.

3.1 Experiments

The curvature-orientation method is tested using two databases. The first is a a collection of 1561 assorted gray-level images and the second, a collection of 2345 parametrically generates shapes. These two collections are discussed next.

3.2 Assorted Collection of Gray level images

This database has digitized images of cars, steam locomotives, diesel locomotives, apes, faces, people embedded in different background(s) and a small number of other miscellaneous objects such as houses. These images were obtained from the Internet and the Corel photo-cd collection and were taken with several different cameras of unknown parameters, and under varying uncontrolled lighting and viewing geometry.

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 [28]. It is a standard widely used in the information retrieval community and is one that is adopted here.

Figure 1: Image retrieval using Curvature and Orientation

For each query in the number of 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. That is, they are not considered retrieved. These ranks were used to interpolate and extrapolate precision at all recall points.

Six queries submitted for recall-precision are shown in Figure 1. The left most image in each row is the query and is also the first retrieved. The rest from-left to right are seven retrievals depicted in rank order. Note that, flat portions of the background are never considered because the principal curvatures are very close to zero and therefore do not contribute to the final score. Thus, for example, the flat background in Figure 1(second row) is not used. Notice that visually similar images are retrieved even when there is some change in the background (row 1). This is because the dominant object contributes most to the histograms. In using a single scale poorer results are achieved and background affects the results more significantly. The results of these examples are discussed below, with the precision over all recall points depicted in parentheses.

1. Find similar cars(65%). Pictures of cars viewed from similar orientations appear in the top ranks because of the contribution of the orientation histogram. This result also shows that some background variation can be tolerated. The eighth retrieval although a car is a mismatch and is not considered.

2. Find same face(87.4%) and find similar faces: In the face query the objective is to find the same face. In experiments with a University of Bern face database of 300 faces with a 10 relevant faces each, the

Table 1: Precision at standard recall points for six Queries: Gray-scale collection

Recall	0	10	20	30	40	50	60	70	80	90	100
Precision %	100	92.6	90.0	88.3	87.0	86.8	83.8	65.9	21.3	12.0	1.4

average	66.3%

average precision over all recall points for all 300 queries was 78%. It should be noted that the system presented here works well for faces with the same representation and parameters used for all the other databases. There is no specific "tuning" or learning involved to retrieve faces. The query "find similar faces" resulted in a 100% precision at 48 ranks because there are far more faces than 48. Therefore, it was not used in the final precision computation.

3. Find dark textured apes (64.2%). The ape query results in several other light textured apes and country scenes with similar texture. Although these are not mis-matches they are not consistent with the intent of the query which is to find dark textured apes.

4. Find other patas monkeys. (47.1%) Here there are 16 patas monkeys in all and 9 within a small view variation. However, here the whole image is being matched so the number of relevant patas monkeys is 16. The precision is low because the method cannot distinguish between light and dark textures, leading to irrelevant images. Note, that it finds other apes, dark textured ones, but those are deemed irrelevant with respect to the query.

5. Given a wall with a Coca Cola logo find other Coca Cola images (63.8%). This query clearly depicts the limitation of global matching. Although all three database images that had a certain texture of the wall (also had Coca Cola logos) were retrieved (100% precision), two other very dissimilar images with coca-cola logos were not.

6. Scenes with Bill Clinton (72.8%). The retrieval in this case results in several mismatches. However, three of the four are retrieved in succession at the top and the scenes appear visually similar.

The recall/precision results over all queries are summarized 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. The Coca Cola retrieval is a good example of this. Second, mismatches can occur at the algorithmic level. Histograms coarsely represent spatial information and therefore will admit images with non-trivial deformations.

3.3 Parametric Shape Collection

The second collection contains 2345 parametrically generated shape images. The shapes are binary in intensity (0 background ,255 foreground) and contain four types of shapes as shown in Figure 2. There are several reasons for using this collection.

First, to test the performance of the Curvature/Orientation (CO-1) algorithm on binary intensities. Binary images don't span the full range of intensity values. Thus, it would be interesting to examine the effectiveness of curvature histograms in this case. Second, to examine if appearance similarity is effective for shapes. The predominant visual information depicted in the objects (see Figure 2) is shape. Thus, one can evaluate the utility of appearance representation for shape similarity. Third, there are several situations where retrieving binary shape images is necessary. Trademark images are usually binary and several of the them have distinct shapes (See Figure 10). Thus this database provides a feasibility study for applying CO-1 to trademark retrieval. Finally, a parametrically generated collection provides ground truth that can be used to determine shape relevance. Therefore, different algorithms can be systematically generated. In this context it should be noted that moments have been widely used. Thus, this collection provides a test-bed for comparison.

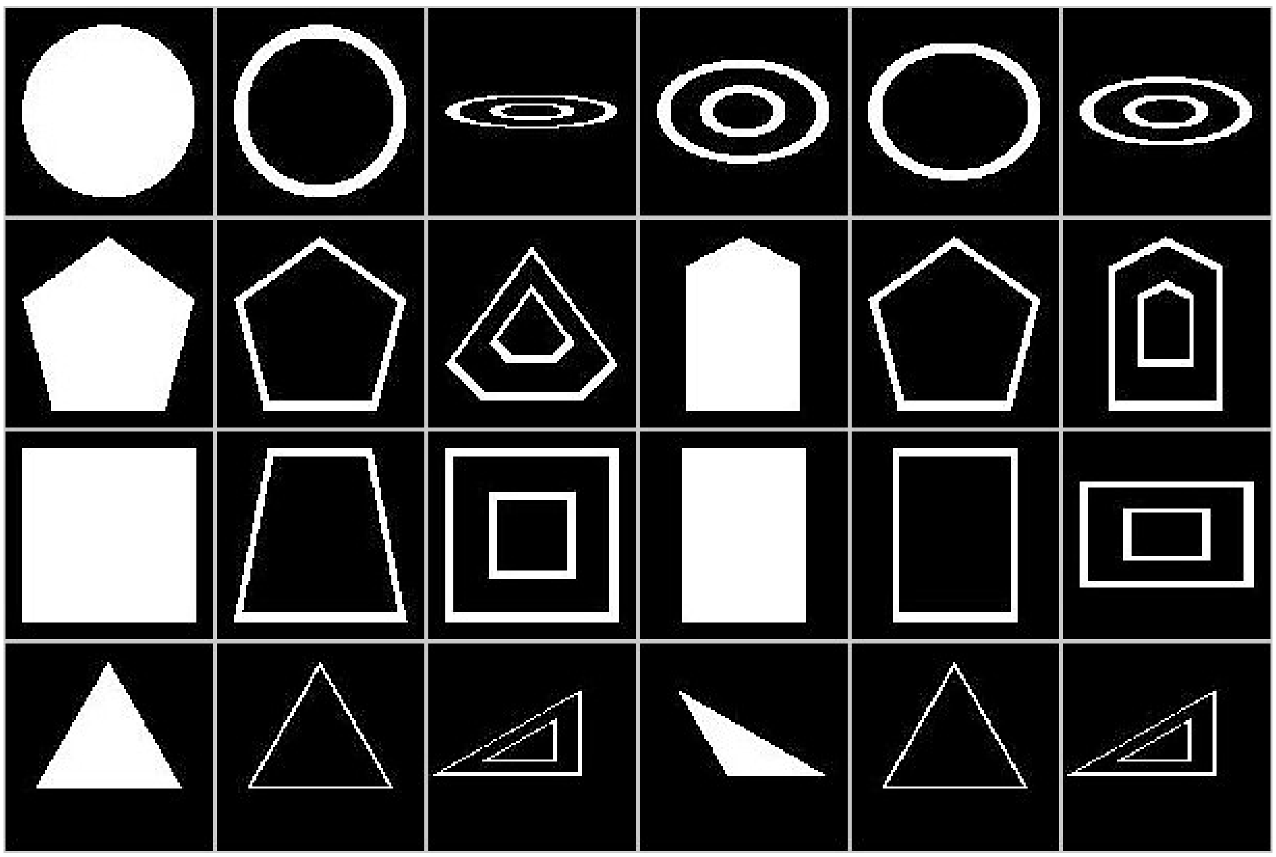

Figure 2: Queries submitted to the Parametric shape collection

Four families of shapes are generated, namely, ellipses, triangles, quadrilaterals and pentagons. The parameters are chosen so a small set of images are generated yet, with a reasonable diversity for evaluation. Each of these shapes are defined by parameters specific to the family and those that are common across families, enumerated in Table 2 and Table 3 respectively. Table 2 enumerates the parameters specific to each shape. All the ellipses in this collection have a constant horizontal diameter, and the eccentricity is changed to generate others. Thus shapes from circles to highly eccentric ellipses are generated. All the

Table 2: Variable parameters individual shape classes

Shape	Parameter 1	Parameter 2	Comment
Ellipse	Eccentricity 0.2 0.4 0.6 0.8 1.0		Horizontal diameter constant
Triangle	Internal angle 1 30, 60, 90, 120	Internal angle 2 30, 60, 90, 120	Base length constant angle 1 + angle 2 ¡= 150
Quadrilateral	Ratio of top edge to bottom edge 0.2, 0.4, 0.6, 0.8, 1.0	Ratio of perpendicular height to bottom edge length 0.6, 0.8, 1.0, 1.2, 1.4	Base length constant
Pentagon	Upper notch angle 72, 90, 108, 126, 144	Angle of bottom notches 90, 105, 120, 135, 150	Base length constant

Table 3: Fixed parameters all shape classes

Parameter	Comment
Edge Thickness	solid, 5, 10, 15 pixels wide.
Inscriptions	One, two or three loops
Noise	0, 2 or 3 points of noise. Points randomly chosen solid shapes have no noise copies

ellipses are oriented horizontally, that is, the horizontal axis has the longer diameter. Triangles are defined by a base length that is constant and two internal angles that are variable. Thus, acute, obtuse, and right angled triangles of various angles (shown in the table) are generated. Quadrilaterals are defined by a fixed base (lower) length and varying both the perpendicular height and length of the side parallel to the base. Thus, various rectangles, squares and trapezoids are generated, but parallelograms and rhombuses are not. Pentagons are defined by a constant base length, and two variable angles. The angle of the edges adjoining the base (they are the same) and the angle of the apex. This constructs regular pentagons and pentagons with obtuse angles.

All shapes are generated with additional parameters constant across families. These parameters are enumerated in Table 3. The first is the thickness of an edge. Shapes can be solid, or with varying edge thickness. The second is the number of inscriptions. Foe example, a pentagon could be inscribed with another pentagon of the same variable parameters. Likewise for other shapes. Finally, each shape has copies with noise added to it. Noise is added by selecting a point randomly (uniform random distribution) and changing its value. Two and three pixel noises are added.

In these set of experiments the CO-1 algorithm is compared with moment invariants. Moments invariant to similarity transformations are generated using the technique shown in [23] and first developed by Hu [9]. Moment invariants have been used by several authors [4, 19, 30] for shape similarity.

Twenty four queries are posed and relevance is shape related. For example, a query number 1, a Circle was posed with find all circles and ellipses as the relevance criteria. Table 4 describes the queries, The first column is the query number whose image is shown in Figure 2 when numbered left to right and then top to bottom. The second column describes the relevance criteria. That is, what is the user looking for. The third

Table 4: Per query average precision over all recall: Parametric shape collection

Query No.	Relevance Criteria	Number Relevant	CO-1	Moments
1	All ellipses and circles	230	90.6	24.1
2	All circles	46	79.1	16.0
3	Ellipses with same eccentricity	46	100.0	24.0
4	Ellipses with same eccentricity	46	88.4	17.0
5	Ellipses with same eccentricity and inscriptions	15	72.7	28.2
6	Ellipses with same eccentricity and inscriptions	15	86.5	28.8
7	All Pentagons	1145	76.8	58.2
8	All pentagons of same regularity	46	92.4	14.3
9	All pentagons of same regularity	45	91.9	21.7
10	All pentagons of same regularity	46	64.0	11.3
11	All pentagons of same regularity and inscription	15	58.6	28.5
12	All pentagons of same regularity and inscription	15	60.4	28.2
13	All quadrilaterals	690	61.9	38.3
14	All quadrilaterals with same shape	28	94.2	25.3
15	All quadrilaterals with same shape	28	56.3	19.9
16	All quadrilaterals with same shape	28	65.0	10.7
17	All quadrilaterals with same shape and inscriptions	9	50.9	36.9
18	All quadrilaterals with same shape and inscriptions	9	58.2	36.8
19	All triangles	280	46.0	19.9
20	All triangles with same angles	28	100.0	21.0
21	All triangles with same angles	28	93.0	21.3
22	All triangles with same angles	28	72.9	10.2
23	All triangles with same angles and inscriptions	9	68.5	37.3
24	All triangles with same angles and inscriptions	9	67.1	37.0

depicts the number of images in the collection (out of 2345) that are actually relevant. The fourth column depicts the percentage average precision of this query over all recall points and the fifth the corresponding number for moments.

Relevance criteria are designed to isolate an increasingly narrow set of images that can be relevant. We start from a very broad criteria. Given a shape find all instances of that family (Query 1, 2, 7, 13, 19). For example, given a pentagon find all pentagons. Then, we make it progressively narrow. Given a shape find find the shape with the same variable parameters (Queries 2,3,4,8,9,10,14,15,16,20,21,22) . For example, given an ellipse find all ellipses of the same eccentricity (other eccentricities would be invalid, but the same with a different edge thickness or number of inscriptions would not). Finally, the narrowest. Given a shape find all shapes of the same variable parameters and inscriptions (Queries 5, 6, 11, 12, 17, 18, 23, 24). For example, given a two loop circle find all two loop circles.

By a large margin the CO-1 technique performs better than moments for all the proposed set of relevance criteria as can be seen from Table 4. The aggregate recall/precision results are summarized in 5. Here we discuss the best and worst performing CO-1 and moment queries.

Figure 3 depicts the top eight ranks for Query 20, the best performing query. Find all triangles with the

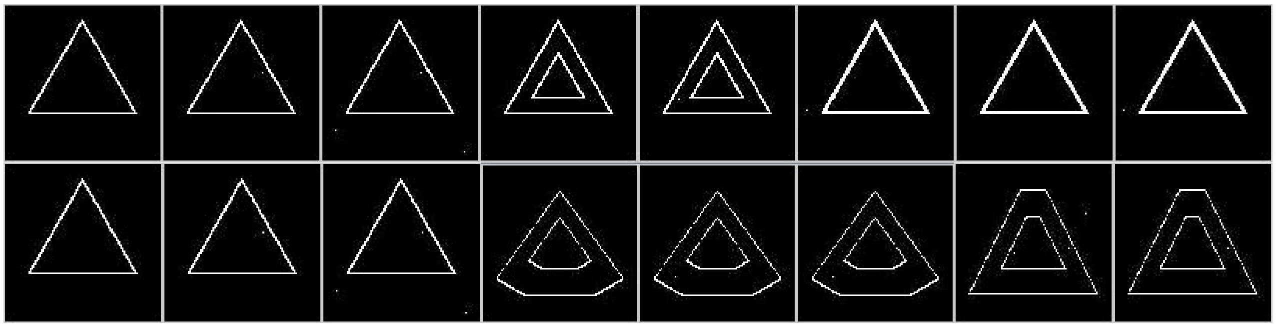

Figure 3: Best performing CO-1 query(20).Top row: top eight CO-1 retrievals. Bottom row: top 8 Moment retrievals

same angles (variable parameters) as the query (shown in Figure 2, second image, bottom row). The CO-1 technique gets all three triangles of the same shape. The next two ranks are of inscribed triangles, whose thicknesses are (5) narrower (5 pixels) compared with the following three (15 pixels). This can be expected because at larger scales the image is smoother, thus making little distinction between a the inscribed triangle and thicker triangle. In contrast, the moments technique retrieves completely different shapes in the top ranks. While it can be argued that these images are somewhat visually similar to the query there are others that are far more similar that have not been retrieved.

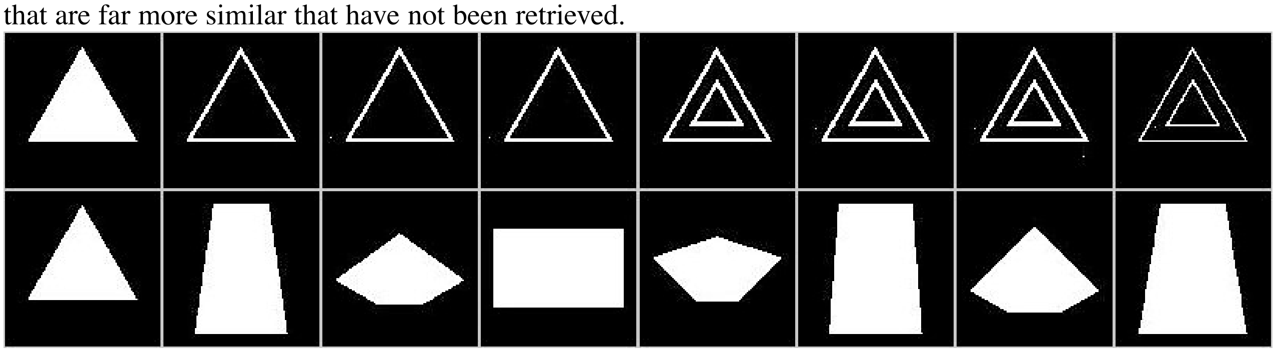

Figure 4: Worst performing CO-1 query(19).Top row: top eight CO-1 retrievals. Bottom row: top 8 Moment retrievals

Figure 4 depicts the worst performing CO-1 query (Query 19) and the corresponding moment ranks. The query triangle is the only solid triangle with the same angle parameters. That is, there are no noise versions of solid objects in the collection (see Table 3). The top row depicting the top 8 CO-1 ranks look very good. However, several triangles in the collection are obtuse (see Figure 6 and were ranked after other shapes such as the three non solid pentagons shown in Figure 5 (ranks 4, 5, 6 top row). In contrast the moments technique performs very differently, and poorly with respect to shape criteria.

This query also depicts a primary difference between the two techniques. Curvature and orientation are differential features and hence flatness in the image intensity surface does not contribute to the representation. Thus the solid "painted" query is treated similar to the "line drawn" triangles. In contrast moments are integrals of the image intensity function and hence work best when the shapes to be compared are all solid. Thus the retrievals prefer solid polygons instead of other triangles and fail shape-based criteria for similarity.

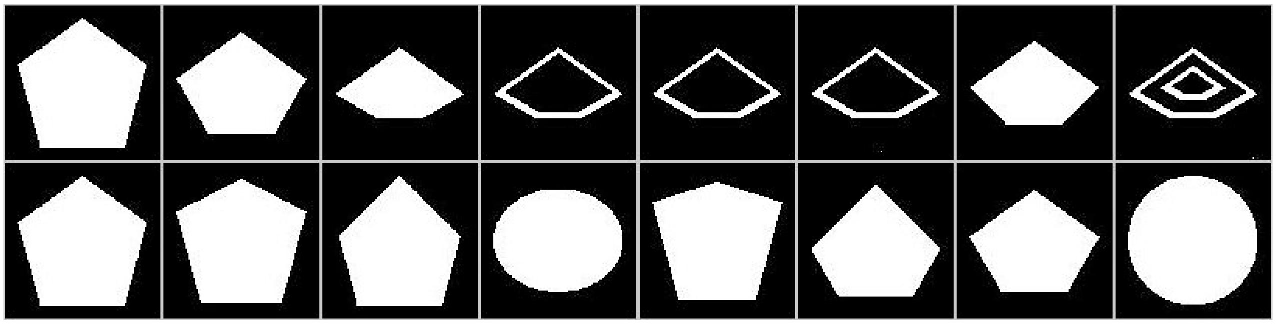

Figure 5: Best performing Moment query(07).Top row: top eight CO-1 retrievals. Bottom row: top 8 Moment retrievals

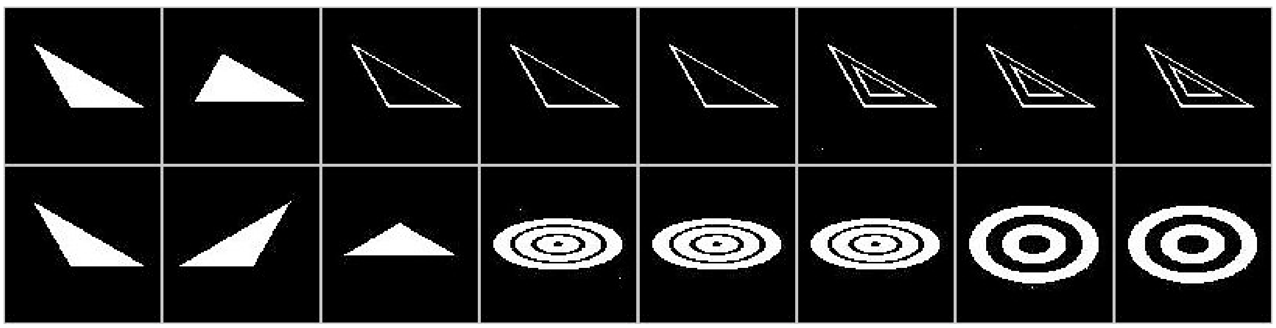

Figure 6: Worst performing Moment query(22).Top row: top eight CO-1 retrievals. Bottom row: top 8 Moment retrievals

Figure 5 and Figure 6 depict the best and worst performing moment queries. The CO-1 technique performs better on both these than moments. These results depict a similar pattern of results as explained above.

The overall results of all twenty four queries are presented in Table 5. This table depicts the precision at standard recall points over the 24 queries. The conclusion from this experiment is that, in so far as shape based criteria are concerned the proposed appearance based technique (CO-1) is better than a commonly used shape based representation namely, moments. In the next section a trademark retrieval system is constructed with application to trademarks.

4 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 7. 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 [1] 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 relevant

Table 5: Precision at standard recall points for 24 queries of the Parametric shape collection

	Precision – 24 queries	
Recall	CO-1	Moments
10	100.0	100.0
20	98.9	70.5
30	95.8	45.1
40	88.1	24.1
50	78.3	7.0
60	74.8	6.4
70	73.8	6.2
80	65.7	5.8
90	58.5	5.7
100	51.8	5.5
110	37.1	5.4
average	74.8	25.6

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 6 lists all the fields associated with a trademark. These were indexed and used for retrieval using the INQUERY search engine [1]. 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 7.

In this section we describe the curvature/orientation 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 orientation histograms are computed for each tile at three scales (1,5,9). 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.

Table 6: 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

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.

4.1 Examples

In Figure 7, 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.

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 8. 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

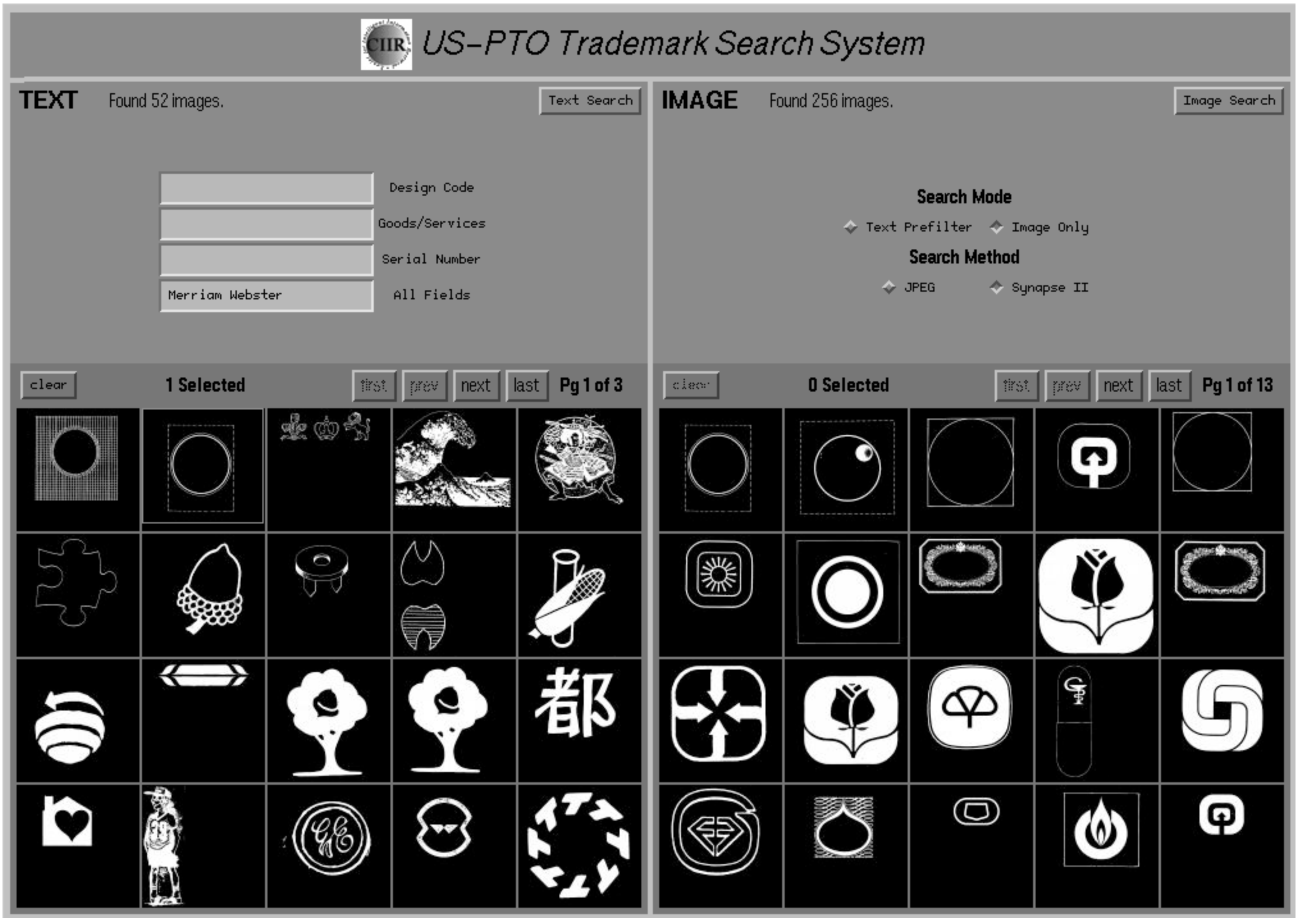

Figure 7: Retrieval in response to a "Merriam Webster" query

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 orientation as is clear by looking at it.

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 9 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.

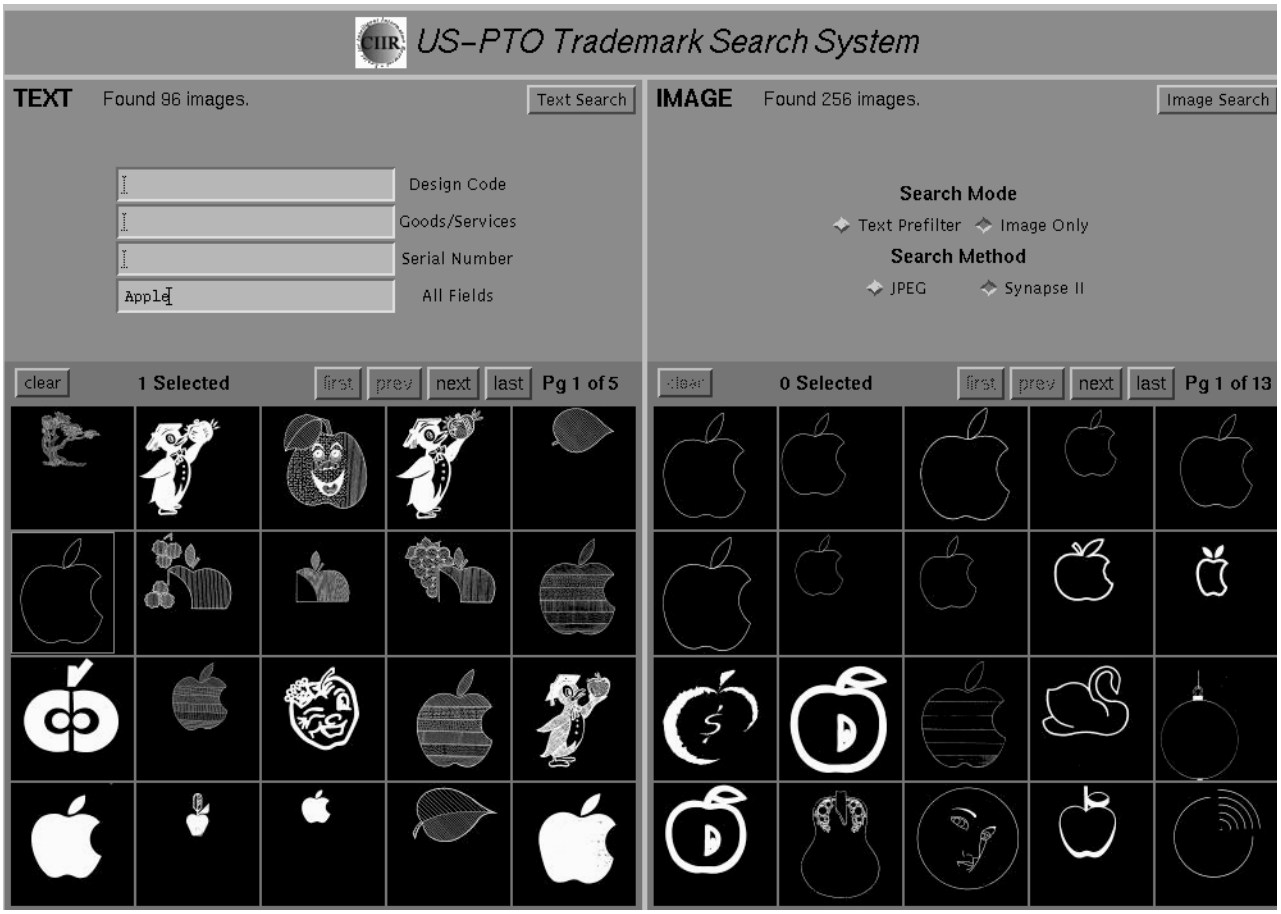

Figure 8: Retrieval in response to the query "Apple"

4.2 Evaluation

At the time of this writing relevance judgments are not available for the PTO collection. As a substitute a collection of 10,745 trademarks are evaluated using the same parameters used to construct the PTO system. These trademarks have been provided by Dr. John Eakins at the University of Northumbria at Newcastle, UK. They come with a set of 24 queries for which relevance judgments have been obtained from various examiners. Hence, a modest, but realistic, evaluation can be obtained. However, the contribution of multi-modality to retrieval effectiveness cannot be gauged.

Twenty four queries are provided for the purpose of evaluation. All these queries are depicted in Figure 10 and Table 7 describes them. Column 2 shows the number of relevant images for the corresponding query, column 3 the precision over all recall for the query using CO-1 and column 4 using moments. Below the results are discussed for the best and worst performing queries which bring up some issues concerning relevance.

Figure 11 shows the top eight ranked retrievals for query 12 using the CO-1 (top row) and moments (bottom row) techniques. The CO-1 technique gives a 100% precision. It should be noted that only four images are provided relevant for this query (first four, top row, Figure 11). However, ranks 5, 6 and 7 are

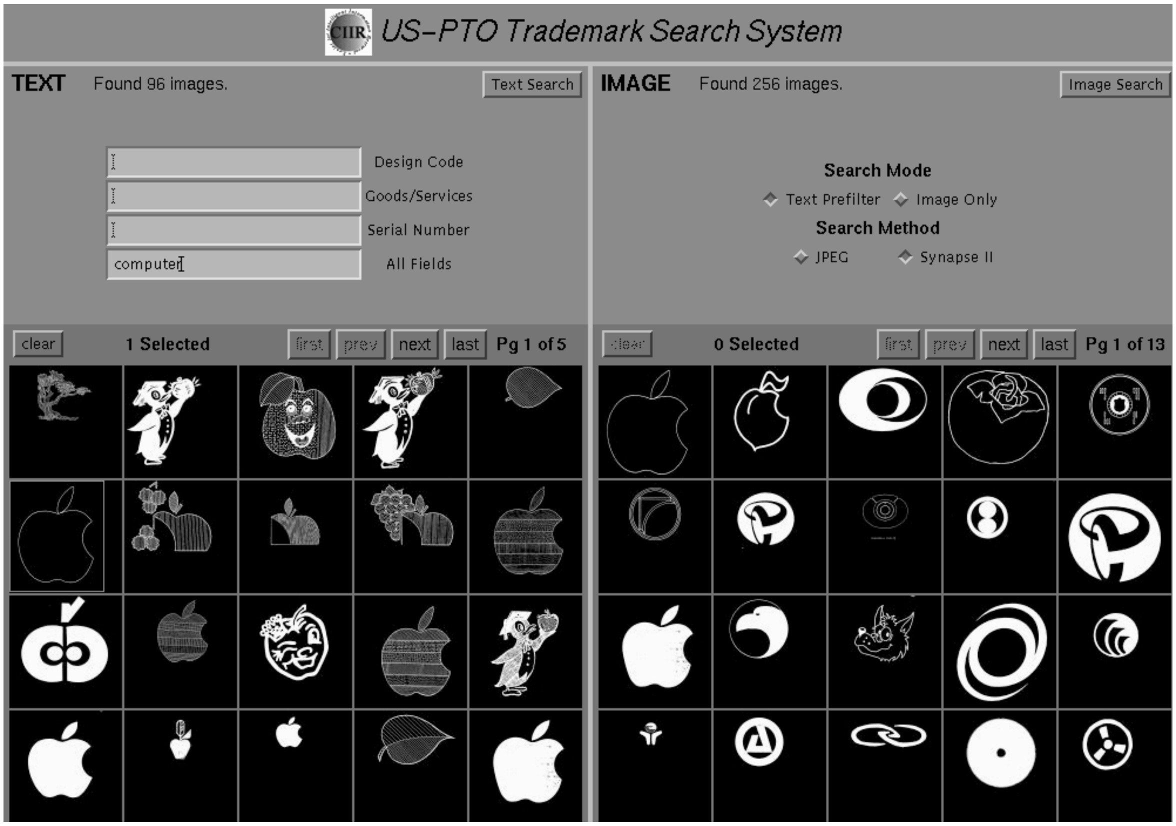

Figure 9: Retrieval in response to the query "Apple" limited to text searches

also very similar visually, but these are not part of the relevant list. In contrast, the moments technique performs poorly. This is also the best performing moment retrieval over all the queries in terms of precision because a near identical copy was retrieved. While the third rank can be considered somewhat visually similar the rest do not bear any resemblance to the query.

Figure 12 and Figure 13 depict the relevant images for query 18 and the retrievals using CO-1 (top row) and moments (bottom row) respectively. This result, in terms of precision, is the worst performing CO-1 query with a meager average precision of 9.5%. However there are some issues that emerge. First the very second rank which is near identical to the query (Image 2047233 in the collection) is not on the relevant list. Second, many images considered relevant actually bear very little "visual resemblance" or "shape resemblance" to the query. We would argue that retrievals 6,7,8 (top row Figure 13) are actually more similar visually than say the first four pictures in row three of the relevant images(Figure 12). Thus, the relevance judgment in this case is based on other criteria. The CO-1 technique cannot be expected to work in this instance. While one can argue that the moment technique will have the same problems, however, even the top 8 ranks of this query (Figure 13, bottom row) contain images that bear very little resemblance to the query visually. It has an average precision of 9.3%.

Figure 14 and Figure 15 show the relevant images and ranked retrievals for the worst performing moment query (Query 19). The moment ranks (Figure 15, bottom row) show that except for rank 6, 8 and perhaps

Figure 10: Queries submitted to the UK Trademark collection

Figure 11: Best performing CO-1 query(12). This is also the best performing moment query. Top row: top eight CO-1 retrievals. Bottom row: top 8 Moment retrievals

2 the others do not bear a significant similarity to the query (Figure 14, bottom row, first picture). In contrast the CO-1 query does perform better. The top eight ranks all have circular shapes, although none with inscribed triangles. All of these image can be considered visually similar, but they are irrelevant. From a matching perspective this is not unexpected. The two flags (triangles) in the query do not provide sufficient discriminating information beyond the circularity of the image to make them prominent. However, an observation of the relevant images shows that the flags are what are deciding in terms of relevance.

The precision at standard recall points for all queries put together is tabulated in Table 8. The CO-1 technique performs better than moments, reasonably well, however not as well as in the parametric shape

Table 7: Per query average precision over all recall: UK trademark collection

Query No.	Number Relevant	CO-1	Moments
1	26	64.3	51.3
2	18	63.8	23.2
3	12	44.7	09.5
4	13	46.8	29.6
5	10	37.0	28.9
6	18	36.5	09.9
7	11	11.1	09.3
8	19	36.9	12.8
9	25	54.7	15.1
10	12	52.0	09.8
11	11	22.2	31.0
12	04	100.0	55.6
13	17	12.4	09.3
14	06	36.4	18.5
15	08	26.5	18.4
16	13	37.6	12.5
17	23	12.2	09.3
18	26	09.5	09.5
19	13	09.6	09.3
20	26	13.9	09.3
21	08	37.0	44.4
22	15	23.6	14.3
23	25	15.8	36.6
24	12	66.4	69.0

collection. The conclusion to be drawn is that in some instances the relevance judgments were incomplete, in some instances the relevant images bore little visual similarity (or shape similarity) to the query and in others, there were other images which bore a gross similarity but didn't match the relevance criteria. Overall, from an algorithmic viewpoint and system performance perspective the results are quite reasonable.

5 Conclusions and Limitations

There are several contributions made in this paper enumerated below

1. The appearance based global similarity technique uses a new type of feature set that is distinct from other appearance based techniques.
2. It is well suited for heterogenous gray level collections.
3. It can be implemented efficiently for large image collections.
4. The technique extends to binary-shape images.
5. Appearance representations can be use to judge shape similarity.

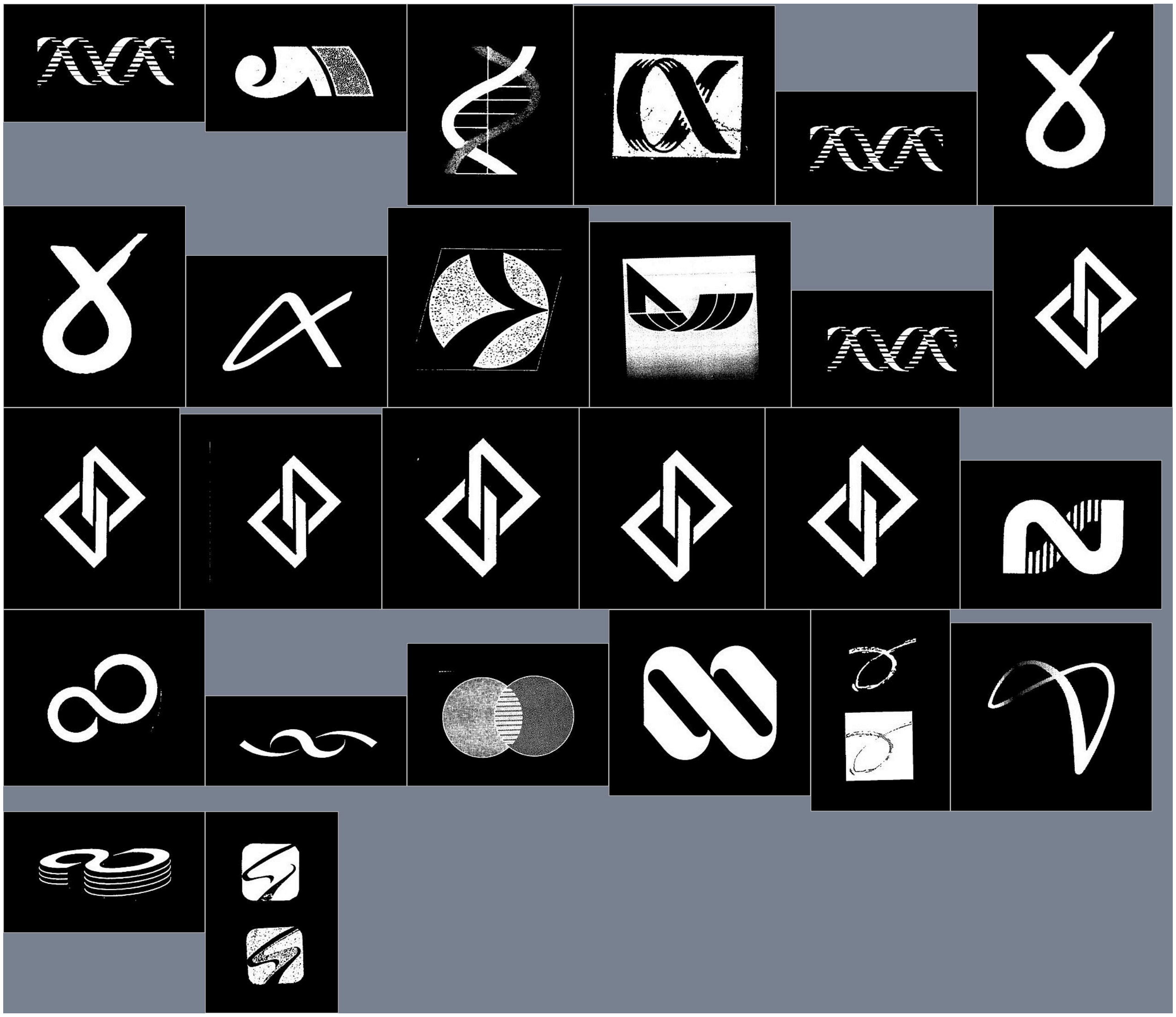

Figure 12: Relevant set for Query 18. Note the second retrieved in Figure 13(Image 2047233) is not in this set.

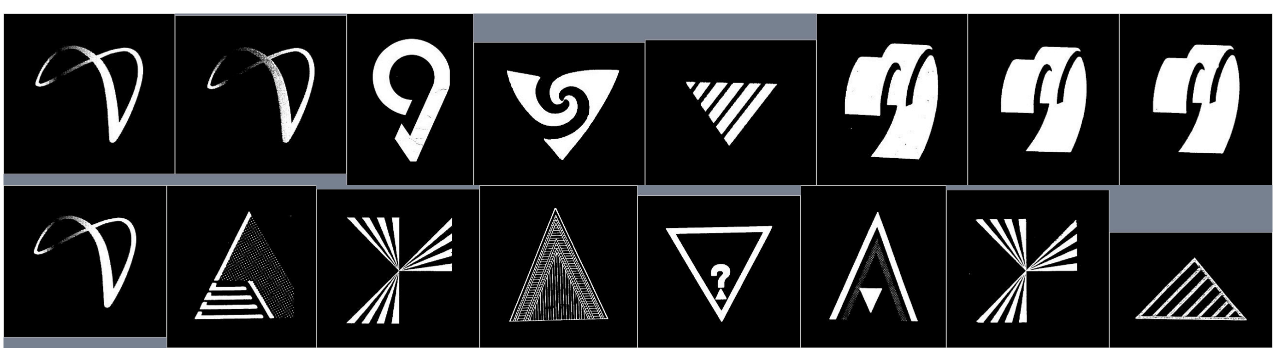

Figure 13: Worst performing CO-1 query(18).Top row: top eight CO-1 retrievals. Bottom row: top 8 Moment retrievals

6. This paper demonstrates multi-modal retrieval and a solution for a practical application.

7. In the context of binary-shape images and trademarks the proposed appearance representation (al-

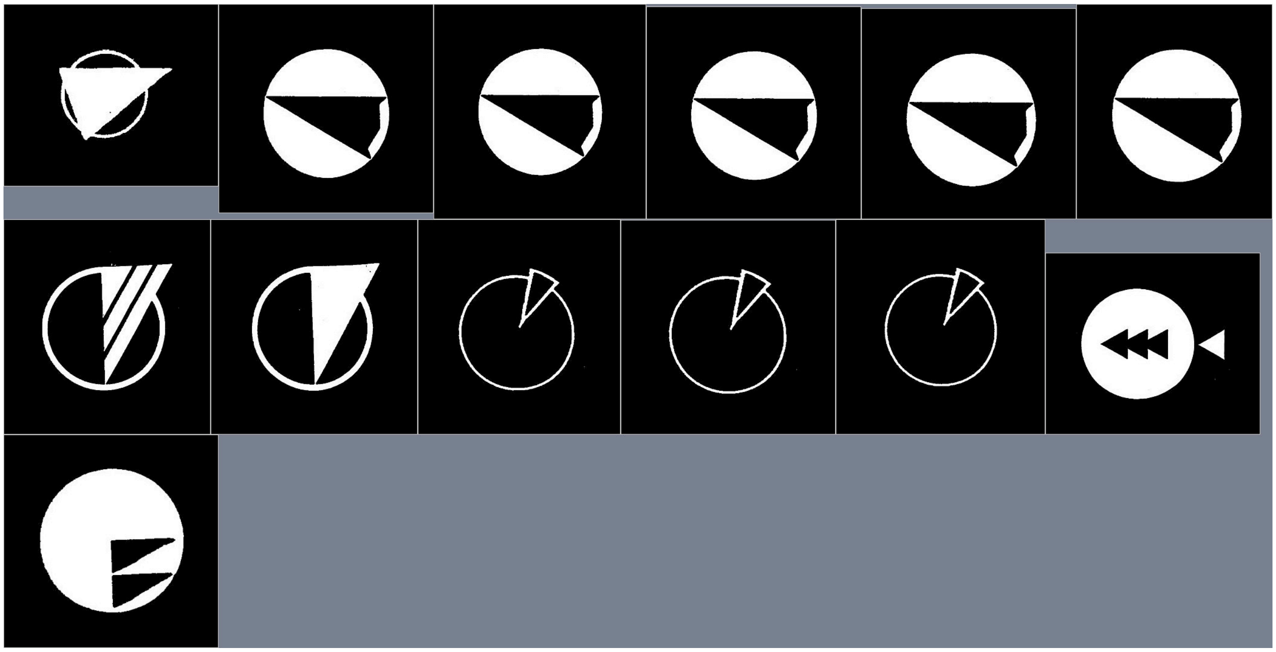

Figure 14: Relevant set for Query 19.

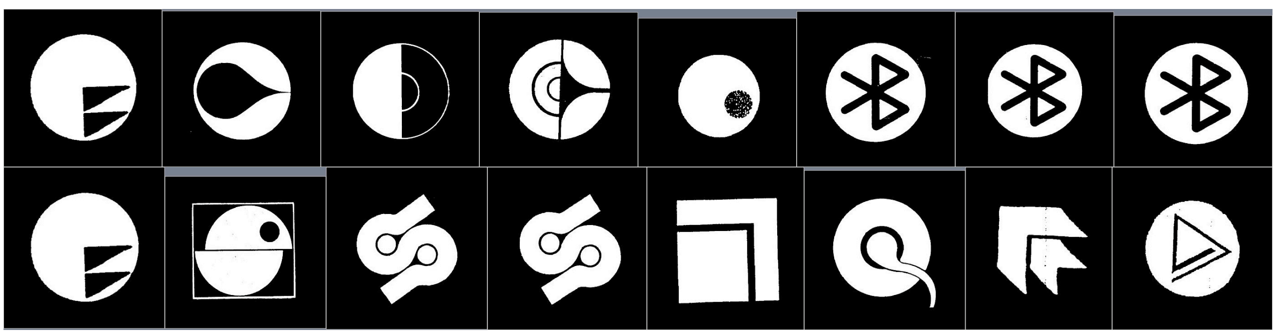

Figure 15: Worst performing Moment query(19).Top row: top eight CO-1 retrievals. Bottom row: top 8 Moment retrievals

though somewhat more expensive to compute and store) provides a far better performance for similarity measurements, be it shape-based or appearance-based, than moments.

There are also several issues that are being examined.

1. In some instances it seems that an image that should be ranked higher is not. We believe this to be because of a bias-variance problem in histogram binning. Thus, kernel estimation techniques can be used to address it.

2. We have been experimenting with relevance feedback, and although it is a part of the system, we do not believe a good model exists yet. This is a hard issue because it is unclear how relevance feedback must be carried out. Should successive runs bring out images that look equally similar to all the images marked relevant ? any of them ? and so on.

3. The parametric shape collection is being expanded. This collection hasn't been exhaustively queried at

Table 8: Precision at standard recall points for 24 queries of the UK Trademark collection

	Precision – 24 queries	
Recall	CO-1	Moments
10	100.0	100.0
20	75.5	51.7
30	62.3	36.1
40	60.7	22.5
50	31.1	15.9
60	26.6	14.2
70	19.0	5.4
80	08.5	3.7
90	05.5	0.4
100	05.4	0.3
110	04.6	0.2
average	36.3	22.8

the time of this writing. It is conceivable to evaluate every image in the collection against several more relevance criteria. It is possible to ask questions such as, all triangles within a 20 degree variation.

4. the multi-modality described here is rudimentary. We are working on techniques to combine text retrieval and image retrieval in a principled manner.

5. Estimating local scale. Scale is a local property. Thus, if one can estimate a natural scale at which a local curvature should be computed, then, a representation that is far more compact and tolerant to scale changes can be computed.

6 Acknowledgements

We would like to thank Tom Michel for designing the graphical user interface. We would also like to thank CIIR for supporting this work. We would like to thank Joe Daverin for his work on this project.

References

[1] 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.

[2] 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.

[3] 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.

[4] J.P. Eakins, K. Shield, and J.M. Boardman. Artisan: A Shape Retrieval System Based on Boundary Family Indexing. In J.K. Sethi and R.C. eds. Jain, editors, *Storage and Retrieval for Image Video and Databases IV*, volume **2670** of *Proc. SPIE*, pages 17–28, 1996.

[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] M.K. Hu. Visual pattern recognition by moment invariants. *IRE Trans. on Information Theory*, IT–8, 1962.

[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] T.H. Reiss. *Recognizing Planar Objects Using Invariant Image Features.*, volume 676 of *Lecture Notes in Computer Science*. Springer-Verlag, 1993.

[24] 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.

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

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

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

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

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

[30] J.K. Wu, B.M. Mehtre, Y.J. Gao, P.C. Lam, and A.D. Narasimhalu. Star – a multimedia database system for trademark registration. In *Lecture Notes in Computer Science: Application of Database*, volume 819, pages 109–122, 1994.