ELECTRONIC WORKSHOPS IN COMPUTING

Series edited by Professor C.J. van Rijsbergen

Jonathan Furner, School of Information and Media Studies, and David Harper, School of Computer and Mathematical Studies, The Robert Gordon University, Aberdeen, Scotland. (Eds)

Information Retrieval Research

Proceedings of the 19th Annual BCS-IRSG Colloquium on IR Research,
Aberdeen, Scotland, 8-9 April 1997

Paper:

Evaluation of a Trademark Image Retrieval System

J.P. Eakins, M.E. Graham and J.M. Boardman

Published in collaboration with the
British Computer Society

Evaluation of a trademark image retrieval system

J P Eakins

Department of Computing, University of Northumbria at Newcastle, Newcastle NE1 8ST, UK

M E Graham

Department of Information and Library Management, University of Northumbria at Newcastle

J M Boardman

Department of Computing, University of Northumbria at Newcastle

Abstract

This paper describes the evaluation of ARTISAN, a system designed to provide automatic retrieval of abstract trademark images by shape feature. The system depends for its operation on analyzing each image to characterize key shape components, grouping image regions into families which potentially mirror human image perception, and then deriving characteristic indexing features from these families. A variety of run-time search options is provided, allowing the user to select alternative sets of shape features and similarity matching paradigms.

The system's retrieval effectiveness has been evaluated by measuring its retrieval effectiveness using a set of 12 real queries with known results put to a collection of over 10 000 abstract geometric shapes from the UK Trade Marks Registry. Normalized recall and precision scores averaged 0.93 and 0.65 respectively. The results suggest strongly that the basic ARTISAN approach is valid, though the present version has significant limitations, particularly when handling badly-scanned images or images with implied shape features. Possible ways of overcoming these limitations are discussed.

1 Introduction

Interest in exploiting the potential of electronically-stored images has increased enormously over the last five years, as more and more image collections have become available in this form [1]. Users in many professional fields are exploiting electronic images in all kinds of new and creative ways, opening up a whole new range of opportunities for providers of such resources. But it has brought problems, too - particularly the task of successfully locating a desired image in a large and varied collection. Retrieval of images by characteristics such as shape, colour or texture - often known as *content-based image retrieval* (CBIR) - is now a flourishing research field [2].

Many experimental CBIR systems have been described in the literature, the best-known probably being QBIC [3], Virage [4] and Photobook [5]. However, few systematic studies of their retrieval effectiveness have been conducted (exceptions being Eakins [6] and Faloutsos et al [7]). The main reasons for this are the limited availability of substantial collections of suitable images, and the lack of image queries with associated relevance judgements from real end-users. Virtually all evaluation studies conducted to date have relied on simulated queries put to artificial image collections - severely limiting the validity of their conclusions.

Working image collections which provide the essential ingredients for worthwhile retrieval experiments do exist, however. The patent offices of most advanced countries hold substantial collections of registered trademark images. Such trademark registries routinely receive requests for registration or infringement queries, which require searching of their existing image collections by shape similarity. The registries thus have large image collections, and examples of past queries together with judgements on the relevance of retrieved images. They therefore provide an ideal test environment for experimental systems.

2 Trade mark image retrieval and the ARTISAN¹ project

The UK Patent Office is responsible for registering all UK trademarks, and now holds over 300 000 current trademarks in their Trade Marks Registry. Around 40% of these contain some form of image data. Before a new trademark is registered, the Registry has to ensure that it is sufficiently distinctive to avoid confusion with existing marks. Any new candidate trademark that is considered to be confusingly similar to an existing mark is 'cited' and

¹ Automatic Retrieval of Trademark Images by Shape Analysis

referred back to its originator for modification. Every time a new candidate mark is submitted, trademark examiners thus have to search through the Registry to identify any existing marks with which it might be confused.

To reduce this task to a manageable level, the Trade Marks Registry currently classifies trademark images by shape feature and type of object depicted, using an elaborate system of manually-assigned codes. These codes work well with trademark images depicting animate or inanimate objects, but less well with abstract geometric designs of the type shown in Fig. 1. Classifying such images on the basis of geometric shape constituents such as circles, triangles, or squares provides only a partial solution to the problem, since there can be several hundred images in each category. Registry staff attempting to establish the novelty of a trademark based on an abstract design are therefore faced with a difficult, time-consuming and potentially error-prone task.

Fig. 1 Examples of typical abstract trademark images. Crown copyright reserved.

The aim of the ARTISAN project is to overcome this problem by developing and evaluating an automatic shape retrieval system for abstract trademark images. Our objective has been to develop a system which meets the Patent Office's specific image retrieval needs, but which is capable of extension to a wider range of image and query types in the future. The current system has been designed to handle any monochrome image capable of being represented as a number of regions with relatively well-defined boundaries. This includes two-dimensional engineering and architectural drawings, the silhouettes often used as test images by image researchers, and many clip-art images, as well as virtually all abstract trademark images - but *excludes* colour images, images depicting natural objects or scenes, and images consisting primarily of texture, without well-defined boundaries.

The underlying philosophy of ARTISAN has been described elsewhere [8]. Essentially it aims to reproduce the judgements of an experienced trademark examiner. Discussions with examiners, and observation of a number of trademark image searches, suggested that examiners assess similarity by identifying key features in the query image (such as a single large shape or a group of objects making up a recognizable pattern), and then looking for stored images containing similar features. ARTISAN thus uses principles derived from Gestalt psychology (summarized in [9]) to segment trademark images into such components before extracting shape features for use in retrieval.

3 Architecture of the ARTISAN system

ARTISAN is a modular system with capabilities for accepting bitmap images in a standard format, processing these images to extract salient components, creating descriptions of these image components, extracting and storing retrieval features from these descriptions, allowing formulation of visual queries, matching query and stored images, and displaying query results on the screen. At present, it consists of the following modules:

- (a) ***Extraction of region boundaries from bitmap images and approximation by straight-line and circular-arc segments.*** This module identifies regions of interest within each image, and characterizes each region by an approximation of its boundary, using a representation based on line segments which can be interpreted *either* as straight lines *or* circular arcs, using a technique based on that of Rosin and West [10].
- (b) ***Reprocessing of boundary representations to remove anomalies caused by noise in the original image.*** This module aims to remove spurious segments generated by noise in the original image, by invoking a set of boundary redrawing rules to remove or reclassify them. Our approach is based on the Gestalt principles outlined above, and uses a rule set based on the redrawing rules developed for our earlier SAFARI system [11].
- (c) ***Grouping of region boundaries into families.*** This module groups boundaries into families which potentially mirror human image perception, as illustrated in Fig. 2. It uses clustering techniques to group boundaries into two separate classes of family. *Proximal families* are identified by clustering image

boundaries on the basis of proximity, parallelism and concentricity scores, and *shape families* by clustering boundaries on the basis of shape similarity.

Fig 2. Processing of a typical trademark image (a) by ARTISAN. The first step is to extract region boundaries, as shown in (b). Boundaries are then grouped into families (c) on the basis of perceptual similarity - in this case grouping the outer two arcs into one family, and the inner twelve lozenge shapes into another.

- (d) **Construction of envelopes for proximal boundary families.** This module constructs an envelope for every proximal family identified in the previous step, preserving concavities in the family's constituent boundaries.
- (e) **Extraction and storage of global shape features.** This module derives a set of shape vectors (arrays of shape features) from the image at three different levels: the entire image, each proximal family, and each individual boundary. At present, we compute a *boundary shape vector* (aspect ratio, circularity, transparency, and relative area) directly from each boundary envelope, and a *family characteristics vector* (complexity, right-angleness, sharpness, straightness and directedness) from individual boundaries within each family, though we are still experimenting with alternatives.
- (f) **Database query.** This module allows the user to select a query image and run-time search parameters, extracts appropriate shape features from the query image, computes appropriate similarity scores between query and stored images by shape feature matching, and displays the most similar retrieved images on the screen. The similarity matching algorithm used is an adaptation of the matching algorithm used for SAFARI [11], but extended to allow alternative methods of comparing stored and query shape elements. Its key elements are:
 - (i) **Multi-level matching.** Query and stored images can be matched either by comparing families alone, or by comparing individual boundaries within each family;
 - (ii) **Alternative feature sets.** Similarity can be computed from any combination of boundary shape vector, family characteristics vector, and relative boundary positions.
 - (iii) **Alternative matching paradigms.** Similarity scores between query and stored images are based on comparing the shape vectors of their components. Since query and stored images normally have different numbers of components, several alternative ways of combining component similarity measures into an overall similarity measure are possible. The ARTISAN prototype offers the following:
 - **nearest**, which takes the overall similarity score between query and stored images as the similarity between their two most closely-matching components;
 - **symmetric strict**, which averages similarity scores for the $\min(q,s)$ closest component matches, where q and s are the numbers of components in query and stored images respectively;
 - **symmetric spread**, which averages similarity scores for the $\max(q,s)$ closest component matches;
 - **asymmetric simple**, which averages similarity scores for the q closest component matches, with no restrictions on which stored image components are matched;
 - **asymmetric spread**, which averages similarity scores for the q closest component matches, with the restriction that no stored image component can be matched more than once until all have been matched.

A detailed exposition of our approach can be found in reference [12].

4 Implementation

The current ARTISAN prototype was implemented in Visual C++ to run under Windows v.3.1 on 66MHz 486DX2 PC compatibles with 16 Mb of RAM and 540 Mb hard discs. This almost certainly represents the minimum system configuration needed to run the system. Fig. 3 shows a typical query screen from ARTISAN, with a query image already selected. The user selects run-time matching options by clicking on the appropriate boxes.

Fig. 3 A typical ARTISAN query screen

Fig. 4 shows some of the results from the above query, with images displayed in order of similarity with the query image. Further retrieved images can be viewed by using the scroll bar on the right of the screen. Note that the query image itself is always included in the search results as a check on the accuracy of the search process.

Fig 4. Results of the query illustrated in Fig 3.

5 Evaluation

The crucial test of the effectiveness of the ARTISAN approach is whether it provides effective retrieval. Thanks to the co-operation of the UK Patent Office, we were able to take the opportunity to evaluate ARTISAN's

retrieval performance on a set of live queries put to a subset of the UK Trade Marks Registry. The evaluation was done in two phases.

Firstly, we obtained informal feedback on the effectiveness of our initial prototype by putting together a small database which contained a selection of genuine trademark images plus four series of similar images created for the purpose by Trade Marks Registry staff. Queries were run against this database using a wide range of different search parameter combinations. Feedback from this phase enabled us to select a small set of promising search parameter combinations for the formal evaluation studies.

Secondly, a formal evaluation experiment was conducted, in which a set of query images which had already been put to the Patent Office's existing system TRIMS - and for which samples of desired output were thus already available - were run against our improved ARTISAN prototype. Both ARTISAN searching and analysis of results were performed at Northumbria University. Where ARTISAN searching retrieved further potentially relevant images, these were referred back to the Patent Office for an experienced trademark examiner to judge whether they were indeed sufficiently similar to the query.

As in previous studies of this kind, retrieval performance was measured using normalized recall R_n and normalized precision P_n as defined by Salton [13]. A further measure used in this evaluation was *last-place* ranking L_n , defined as:

$$L_n = 1 - \frac{R_n - n}{N - n}$$

This measure gives an indication of the number of retrieved items through which a user has to search in order to have a reasonable expectation of finding *all* relevant items, thus providing an indication of the level of confidence that the Patent Office could place in the system.

5.1 Pilot evaluation tests

A pilot database of 268 images was built using the techniques described above. This database contained 231 randomly selected trademark images, plus four series of test images provided by the Trade Mark Registry. Each series contained one genuine trademark image, which served as the image query, plus a number of modified images close enough to the original to be considered 'cites'. It was necessary to resort to this subterfuge because genuine examples of such images are extremely rare, and the Patent Office wanted to retain these for use in the formal system trials. The four query images are illustrated in Fig. 5.

Fig 5. The four query images used in preliminary trials

Each of the four queries was run against the sample database using the five search paradigms described above:

- **nearest** (score based on best match between *any* pair of image components)
- **symmetric strict** (score based on one match for each component in image with *fewer* boundaries)
- **symmetric spread** (one match for each component in image with *more* boundaries)
- **asymmetric simple** (one match for each component in query image; no restrictions on which stored boundary is matched)
- **asymmetric spread** (one match for each component in query image; each stored boundary matched at least once if possible)

with various combinations of the following search parameters:

- **boundary shapes** (aspect ratio, circularity, transparency & relative area)
- **boundary positions** (relative positions of boundary centroids with respect to overall image centroid)
- **family characteristics** (right-angledness, sharpness, complexity, directedness & straightness)

boundary shape matching being applied at one or both of the following levels of image description:

- **family envelopes** (shape measures compared for the bounding envelope of each family)

- *explicit boundaries* (shape measures compared for each individual boundary in a family).

5.2 Results from pilot evaluation

It rapidly became obvious that some combinations of search were incapable of delivering acceptable results whatever search paradigm was used, so these were dropped from consideration after the first few experiments. Average P_n scores for the best four search combinations for the four query images are presented in Table 1 below; R_n and L_n scores tell a similar story.

Search parameters:				
Boundary shapes	*	*	*	
Boundary positions	*			
Family characteristics	*	*	*	*
Boundary matching level:				
Family envelopes	*		*	
Individual boundaries	*	*	*	
Matching paradigm:				
Nearest	0.68	0.69	0.68	0.65
Symmetric strict	0.60	0.64	0.64	0.50
Symmetric spread	0.72	0.70	0.71	0.54
Asymmetric simple	0.70	0.77	0.76	0.72
Asymmetric spread	0.63	0.67	0.66	0.54

Table 1. Effect of different search combinations on retrieval effectiveness, shown as average P_n scores for the four test queries

On the basis of these preliminary findings, we drew the following provisional conclusions:

1. Even in its prototype form, the system appears to be capable of delivering respectable retrieval results.
2. The most promising combinations of matching paradigm and search option were:
 - (a) asymmetric simple matching using a combination of boundary shape features at the individual boundary level, plus family characteristic features;
 - (b) asymmetric simple matching using a combination of boundary shape features both at the family and individual boundary level, plus family characteristic features;
 - (c) asymmetric simple matching using family characteristic features alone;
 - (d) symmetric spread matching using all available types of feature.

However, it should be noted that given the small sample size, none of these differences were statistically significant.

5.3 Evaluation of ARTISAN on a full-sized database

The main experiments on retrieval performance were performed on a collection of 10745 abstract trademark images provided by the Patent Office, which were loaded into ARTISAN to form the test database. They also provided twelve queries (including the four used for preliminary trials) for which TRIMS output was already available. For each of these queries, they listed the retrieved images which Patent Office examiners considered citeable. Each of these twelve queries was put to ARTISAN, using the most promising search strategies identified in the preliminary trials. The rank at which ARTISAN retrieved each cited image was used to calculate R_n , P_n and L_n scores using the same method as before.

Table 2 summarizes results for all 12 queries used in the full evaluation of ARTISAN, using the most successful search combination identified in our preliminary experiments - asymmetric simple matching on boundary shape features at the individual boundary level, plus family characteristics. Given that the current version of ARTISAN is an early prototype, with a number of design features not yet fully implemented, the results are highly encouraging.

The generally high values of R_n (emphasizing retrieval performance at higher ranks over those lower in the scale) suggest that ARTISAN is generally very successful in retrieving closely similar images at high ranks. In virtually all cases, at least one cited image was retrieved within the first few places on the list. The lower values of P_n and L_n indicate that ARTISAN in its present form is not so successful at retrieving less similar images.

Query no	R_n	P_n	L_n
392632	0.88	0.34	0.56
665322	0.98	0.84	0.86
1037814	0.96	0.85	0.60
1138103	0.89	0.66	0.46
1138293	1.00	0.88	0.97
1259886	0.84	0.59	0.24
1279931	0.90	0.64	0.34
1376861	0.90	0.49	0.55
1525429	0.99	0.63	0.96
1575286	0.99	0.84	0.96
2010916	0.90	0.61	0.54
2018809	0.90	0.44	0.65
Mean	0.93	0.65	0.64
S.D.	0.05	0.18	0.25

Table 2. Summary of evaluation results for queries to full database

A typical example of ARTISAN’s retrieval results is shown in Table 3.

Device no		Retrieval rank	Device no		Retrieval rank
Query image		-	1572567		14
1575272		1	2014724		47
1575269		2	2040904		91
1575270		3	1576428		183
1575271		4	1581395		205
2014722		5	1282138		461
1572494		6			

Table 3. Retrieval results for query image 1575286, using asymmetric simple matching on boundary shape features at the individual boundary level plus family characteristics.

5.4 Analysis of retrieval failures

Future modifications to ARTISAN need to be guided by an understanding of its current weaknesses. Hence a detailed analysis of cases of retrieval failure - defined as inability to retrieve a citeable image within the top 10% of ranked output - was conducted, revealing the following causes:

A. *Failure to recognize implied shape features (13 images)*

Not all shape features which the eye recognizes in an image are explicitly present as image regions. Although ARTISAN successfully copes with many such cases, its ability to recognize implied features is clearly not yet sufficiently developed, as the example in Fig. 6 shows.

Fig 6. Failure to match on implied shape features. Query 2018809 (a) matched well with image (b), but almost completely failed to match image (c), despite that each was perceived as containing a triangular shape. ARTISAN correctly recognizes the explicit triangle in (b), but not the striated triangle perceived by most observers in (c).

B. *Failure to cope with border around query or cited image (10 images)*

Query images with a marked border often fail to retrieve images with similar components, but lacking any border. This is an unfortunate side-effect of the otherwise successful *asymmetric simple* matching paradigm adopted for the evaluation experiments. Fig 7 illustrates the problem.

Fig 7. Failure to recognize basically similar shapes because they lack the same bounding envelope as the query. Query 1138103 (a) failed to match either of the images shown in (b) and (c).

C. *Inappropriate boundary detection in query or cited images (9 images)*

This problem was caused by failure of the initial image segmentation algorithms to partition closely similar images in the same way. An example of its adverse effect on retrieval can be seen in Table 4, which compares ARTISAN's retrieval performance with the badly-scanned query image 392632 (illustrated in fig 5(b)) and the more accurately-scanned but otherwise identical query image 392633. Image quality clearly has a marked effect on retrieval performance.

Query image	Retrieval effectiveness		
	<i>Rn</i>	<i>Pn</i>	<i>Ln</i>
392632	0.88	0.34	0.56
392633	0.99	0.77	0.98

Table 4. Effect of query image quality on retrieval effectiveness

D. *Miscellaneous problems (6 images)*

No common thread could be identified in the remaining cases of retrieval failure. Problems included software bugs in ARTISAN's boundary approximation and redrawing modules, failure to cope with textured regions in an image, and one case where a series of multiple images was processed as a single image, resulting in a rather bizarre overall shape.

6 Conclusions

The evaluation results presented here suggest strongly that the basic approach used in the development of ARTISAN has been justified. In particular, retrieval performance with closely-similar shapes is generally very good. Analysis of retrieval failures has revealed that the majority fall into a relatively small number of classes, making it easy to target future development effort. The principle of grouping of boundaries into families from which shape measures are derived - perhaps the most novel aspect of ARTISAN's design - appears to be of significant value, though the concept is still in need of refinement. Use of the *family characteristics* matching option appeared to be particularly successful.

However, the prototype system does not yet offer reliable enough performance to form the basis of an image retrieval system which could be put into routine use. In particular, its ability to retrieve marginally similar images needs considerable improvement. In the short term, we aim to improve ARTISAN's retrieval performance by correcting the anomalies in boundary and family representation discussed above, improving image cleanup routines, and extending texture-handling facilities. Further research is also needed to investigate the effectiveness of alternative types of shape feature in retrieval, including moment invariants [14], axes of symmetry [15], shape family pattern features [8], local shape features such as the angle-segment-angle triplets [11] or longer sequences of contiguous segments [16], and the symbolic features shown by Dyson et al to form a basis for human shape discrimination [17].

In the longer term, the most serious problem we need to address is ARTISAN's limited ability to recognize similarities involving implied shape features in an image. As discussed above, this is the commonest cause of retrieval failure. It is also probably the hardest to tackle. We feel the most promising approach to this problem is to attempt to create a multi-view representation of each image, using an extension of the Gestalt principles discussed earlier, drawing on ideas first proposed by Marr [18]. Many trademark are capable of a number of different interpretations. For example, the image in fig 5(a) could be interpreted as a set of irregular white regions bounded by thick black lines, a series of overlapping straight lines bounded by a circle, or as the letters A, O and W. Similarly, the overall impression conveyed by fig 1(d) is of a circle within an ellipse, even though it consists only of long thin bars. ARTISAN can correctly identify the ellipse (which gives it the advantage over most image retrieval systems), but not the circle.

We therefore aim to incorporate what we propose to call *perceptual shape representations* into future image retrieval systems - allowing them to judge shape similarity on the basis of how the eye actually perceives an image, not just from explicit features extracted from the image. This perceptual representation will comprise multiple views of image components, allowing alternative interpretations of their content. We expect this technique to yield significant improvements in retrieval effectiveness.

7. Acknowledgements

We are grateful for the financial support from the British Library and the UK Patent Office for this work, and the advice and assistance given to us by Patent Office staff. The views expressed here are purely those of the authors.

8. References

1. Cawkell A E "Indexing collections of electronic images: a review" British Library Research Review 15. British Library, London (1993)
2. Gudivada V N & Raghavan V V, eds "Content-based image retrieval systems" *IEEE Computer*, **28**(9), 18-22 (1995)
3. Flickner M et al "Query by image and video content: the QBIC system" *IEEE Computer*, **28**(9), 23-32 (1995)

4. Gupta A et al "The Virage image search engine: an open framework for image management" in *Storage and Retrieval for Image and Video Databases IV*, (Sethi I K & Jain R C, eds), Proc SPIE 2670, pp 76-87 (1996)
5. Pentland A et al "Photobook: tools for content-based manipulation of image databases" in *Storage and Retrieval for Image and Video Databases II* (Niblack W R & Jain R C, eds), Proc SPIE 2185, pp 34-47 (1994)
6. Eakins J P "Design and evaluation of a shape retrieval system" PhD thesis, University of Newcastle upon Tyne (1990)
7. Faloutsos C et al "Efficient and effective querying by image content" *Journal of Intelligent Information Systems* **3**, 231-262 (1994)
8. Eakins J P, Shields K, & Boardman J M "ARTISAN - a shape retrieval system based on boundary family indexing" in *Storage and Retrieval for Image and Video Databases IV*, (Sethi, I K & Jain, R C, eds), Proc SPIE 2670, pp 17-28 (1996)
9. Lowe D G "Perceptual organization and visual recognition" Kluwer, Boston (1985)
10. Rosin P L & West G A W "Segmentation of edges into lines and arcs" *Image and Vision Computing* **7**(2), 109-114 (1989)
11. Eakins J P "Design criteria for a shape retrieval system" *Computers in Industry* **21**, 167-184 (1993)
12. Eakins J P et al "Retrieval of trade mark images by shape feature" British Library Research & Innovation Report 26. British Library, London (1996)
13. Salton G "The SMART retrieval system - experiments in automatic document processing" Prentice-Hall, Englewood Cliffs, New Jersey (1971)
14. Hu M K "Visual pattern recognition by moment invariants" *IRE Transactions on Information Theory* **IT-8**, 179-187 (1962)
15. Rosenfeld A "Axial representation of shape" *Computer Vision, Graphics and Image Processing* **33**, 156-173 (1986)
16. Mehrotra R & Gary J E "Similar-shape retrieval in shape data management" *IEEE Computer* **28**(9), 57-62 (1995)
17. Dyson M et al "The perception of symbols on screen and methods of retrieval from a database" British Library R & D Report 6163. British Library, London (1994)
18. Marr D "Vision" Freeman, San Francisco (1982)