Trademark Image Retrieval Using Multiple Features

Sujeewa Alwis and Jim Austin

Department of Computer Science

University of York

York, YO10 5DD, UK

email: sujeewa@cs.york.ac.uk and austin@cs.york.ac.uk

Abstract

This paper describes an ongoing research project aimed at implementing a trademark retrieval system using an associative memory neural network. The novel aspect presented in this paper is the proposed integrated framework for image retrieval using multiple representations of images based on gestalt principles. In this paper we summarise the methods we followed in extracting local perceptual features as well as features of the closed figures of images. In designing the search engine of the system we have adopted a novel similarity assessment criteria based on local features as well as features of the closed figures, which is being implemented using an associative memory neural network to achieve high performance in retrieval. Then we describe the strategy we followed in combining multiple similarity measures and present the results obtained from the first phase of evaluation of the system.

1 Introduction

There has been considerable progress in the area of content based image retrieval during the last two decades. But capturing perceptual similarity of images is a relatively under-explored area of research [1]. Trademark image retrieval provides a good avenue of investigation in this regard since an effective trademark retrieval system should necessarily be able to retrieve images which humans perceive as similar.

Trademarks play an important role in providing unique identity for products and services in the marketing environment and trademark classification systems should be able to ensure that the existing trademarks are distinct to avoid confusions. Traditionally, classification of trademarks is based on limited vocabulary descriptions. Most of the patent offices use manually assigned codes to represent these descriptions such as human beings, animals or geometrical figures. But it has been shown that these methods suffer a number of problems. The assignment of classes to trademarks is subjective, the classes become either too specific or too broad depending on how users use the classes, there is no mechanism to handle the generation of new classes, and there is a large fraction of images with little or no representational meaning making such a classification extremely difficult. This motivates the need to investigate the potential of content based image retrieval techniques in solving this problem.

In this study, we investigate a new trademark image retrieval system based on features extracted using Gestalt feature extraction methods. During retrieval, we utilize alternative feature interpretations in four different modules. To obtain the final similarity scores, we combine evidence from these modules. Though this framework may be able to capture perceptual similarity of trademark images, the high computational requirements creates the need for an efficient and low cost computational platform.

There have been numerous attempts to solve a range of problems using neural networks. However, many neural network architectures suffer from a long training time or an inefficient hardware implementation. Associative memory architectures perform better than many other methods in this respect. Pattern matching capabilities offered by correlation matrix memory networks(CMM) [2] under the framework of AURA [3] provide a number of features that have been exploited to obtain an efficient search engine for the proposed system. Apart from a fast, low-cost hardware implementation of the network, it offers the ability to parallelise the search mechanism by presenting input patterns

and obtaining output pattern simultaneously. With this integration we hope to use findings from visual cognitive psychology under the neural network framework, in an attempt to integrate advantages from both fronts.

The content of this paper differs from our earlier reported work in [9] in several ways: we present the strategy we followed in extracting an alternative feature collection for our image collection and the strategy we followed in combining multiple feature measurements. Also the results which we present in this paper are based on a larger collection of 1000 trademark images.

The rest of the paper is organised as follows: in section 2, we summarise previous work done on trademark image retrieval and in section 3, we summarise the method followed in extraction features and how they are represented in the model. We describe the proposed methods for feature extraction in section 4 and in section 5, we describe how multiple similarity measures can be combined. We finally present results of the first phase of evaluation of the system and conclusions.

2 Related Work

Several research groups have been working on finding a solution to this problem using automatic shape retrieval systems. Kato *et al* [4] described a system known as TRADE MARKS. Their approach is based on mapping normalised trademarks to an 8X8 pixel grid and calculating a various pixel distributions for each image. The query phase consists of comparing pixel distributions between the query image and the stored images. The STAR system [5] offers a collection of modules for trademarks image retrieval based on shape, the meaning of the trademark and the words in the trade mark. They use Fourier descriptors, gray level projection and moment invariants to capture the shape features. The ARTISAN system [6] presents an approach which incorporates principles derived from Gestalt psychology. It uses a boundary family extraction method based on gestalt principles in an attempt to capture human perceptual similarity judgements about images. In this approach, gestalt principles of proximity and similarity are used to group boundaries into perceptually significant families. Feature vectors obtained from these families and boundaries alone are used for database indexing. Retrieval is based on calculating the Euclidean distance between feature vectors using several different strategies.

We investigate a framework for image retrieval which use multiple feature measurements. The search engine of our system is implemented using CMMs to obtain high performance during retrieval. A typical query process using our image retrieval system is shown in figure 1.

Figure 1: A typical query process using our image retrieval system.

3 Feature Extraction and Representation

During feature extraction, we aim at extracting features from different analytical levels of images. One of our recent experiments conducted using human subjects justifies this approach as we have observed different human interpretations for the same image. In deciding the features to be extracted we are motivated by visual cognitive psychology. Biederman [7] argues that edge extraction is the initial phase of human object recognition and then it goes through the detection of a number of non-accidental properties identified by Witkin and Tenenbaum as follows: co-linearity of points or lines, co-curvilinearity of points or arcs, parallelism of lines and arcs, symmetry under reflection or rotation, convergence of lines or arcs at vertices. The Gestalt psychologists observed and emphasised the importance of organization in vision. They demonstrated that shapes have some illusive, immeasurable collective properties that do not appear when analysed by their constituent parts. Gestalt perceptual organization phenomena are based on proximity and similarity of features. It can be seen that the gestalt feature extraction methods could be used to extract the the above mentioned non-accidental properties and closure feature as well as to group the image into perceptually significant segments.

According to Gestalt theorists, humans tend to use collective properties of closed figures or contours in making similarity judgements. But, retrieval of images using only features of closed boundaries would not be always successful for a number of reasons. For example, objects under occlusion would not give the expected feature vector of the closed figure, the objects that do not have a closed contour would not have such a feature vector either. On the other hand, local features-based matching alone would also not always be robust enough due to the fact that it would be too sensitive to small changes in the object. These facts show the necessity of performing analysis based on features of closed figures as well as local features, in image retrieval.

In our system, the above mentioned non-accidental relationships are used in local feature representation while a widely used set of features is utilised in closed-figures based feature representation.

We summarize the feature extraction process in figure 2.

Figure 2: The overview of the feature extraction phase.

During our study, we first extracted features using the bin voting method for gestalt feature extraction suggested by

Sarkar and Boyer [8] as explained in our earlier work [9]. Our experience with this approach showed that it requires a certain amount of user interaction in adjusting thresholds for bin space voting. During the process, we tried to use an empirically driven set of thresholds but we used user interaction in re-adjusting them when it gave a better set of features with some images. The main disadvantage of this method is its inability to change the thresholds according to the properties of pairs of segments. For example, the same distance threshold is used in assessing the end-point proximity relationship between pairs of shorter lines and longer lines.

It can be seen that the pairwise feature extraction process suggested by Lowe [10] does not suffer from this limitation. In deriving this framework he assumes that the line segments are uniformly distributed in the image with respect to orientation, position and scale. According to his criteria, end point proximity relationship can be extracted by considering the significance of the relationship between two lines of length of l and l_1 ($l < l_1$) separated by a distance of r . He suggests that the number of expected lines for joining can be measured by $\rho_1 r^2 / l^2$ where ρ_1 is a unitless constant. Since the measure of significance is relative, the absolute value of ρ is not important. Accordingly he suggests that the proximity factor for parallelism between two lines within the same angular range having lengths of l and l_1 ($l < l_1$) can be measured by $\rho \theta s l_1 / l^2$ where s is the perpendicular distance from the longer line to the mid-point of the shorter line. The proximity factor for co-linearity is $\rho \theta s (g + l_1) / l^2$ where g is the separation of the two lines. We extend the endpoint proximity relationship extraction method for curves by replacing the length of lines with the perimeter of curves.

We conducted an experiment to compare the capabilities of these two methods in extracting perceptual features of trademark images. During the experiment, we used 20 images which consisted of different perceptual relationships. We used the bin voting method with bin size equal to average length of lines (as suggested by Sarkar and Boyer). Also we used bin voting method under bin sizes based on our earlier experience in extracting these features with some human intervention. In extracting these features using the pair wise method we used a set of three different proximity thresholds for each feature relationship and in this paper, we present only the best results obtained using them.

With each image we counted unexpected groupings (false positives) and missed groupings (false negatives) as a fraction of expected groupings. This expectations were however subjective decisions as they were made by us based on our earlier experience in extracting perceptual features with human intervention.

The table 1, table 2 and table 3 show average performance of these methods in extracting closure, parallel lines and co-linear lines, respectively.

Bin voting (7 pixel units)		Bin voting (average length)		Pair wise method (proximity factor < 0.01)	
false positives	false negatives	false positives	false negatives	false positives	false negatives
3 %	19 %	646 %	24 %	28 %	13 %

Table 1: Performance of different strategies in extracting closure.

Bin voting (50 pixel units)		Bin voting (average length)		Pair wise method (proximity factor < 0.01)	
false positives	false negatives	false positives	false negatives	false positives	false negatives
9 %	22 %	28 %	15 %	14 %	25 %

Table 2: Performance of different strategies in extracting parallel lines.

Bin voting (60 pixel units)		Bin voting (average length)		Pair wise method (proximity factor < 0.04)	
false positives	false negatives	false positives	false negatives	false positives	false negatives
7 %	25 %	5 %	21 %	21 %	32 %

Table 3: Performance of different strategies in extracting co-linear lines.

It can be seen that pair wise method gives the smallest number of false negatives in extracting closed figures. The use of average length, as suggested by Sarkar and Boyer performed reasonably well against other methods in extracting co-linear and parallel lines. But it produced poor results in extracting closure. This observation justifies our earlier approach in using separate bin spaces for end-point proximity and co-linearity and co-curvilinearity as described in [9]. However, we could not observe any improvement of performance in extracting parallelism and co-linearity using the pair wise approach.

Then we extracted a second set of features for the same image collection using Lowe's the end-point proximity extraction method and bin voting method for parallelism, co-linearism and co-curvilinearism. During this, the feature selection process was completely free from user intervention.

We group images based on co-linearism and co-curvilinearism and obtain a new image (gestalt image) as shown in figure 4, which is again subjected to the earlier process of extraction of end-point proximity and parallelism. We obtain closed figures by grouping the features of the image based on end-point proximity and continuity. This method extracts alternative interpretations of closed figures which may not be obtainable using standard pixel based linking methods as shown in figure 5. In the next step, we extract features of closed figures, namely circularity, directionality, straightness, complexity, right-angledness, aspect-ratio, sharpness and stuffedness.

The local feature arrangement is then represented as a graph in which nodes represent the local attributes (length, orientation, curvature) of segments and arcs represent different perceptual relationships between the segments, as shown in figure 6.a. Each image is represented in two such graphs representing the perceptual structure before and after grouping based on co-linearism and co-curvilinearism.

Closed figure arrangement can also be visualised as a graph structure in which nodes represent the feature vectors and every node is connected to each other denoting the fact that they are constituents of the same image, as shown in figure 6.b.

Figure 3: Figure 3.b shows the co-linear and co-curvilinear segments while figure 3.c parallel segments extracted using the image in figure 3.a.

Figure 4: Figures 4.b and 4.d show the grouped images obtained using images in figures 4.a and 4.c respectively.

4 Similarity assessment

In this phase, we use local features as well as the features of closed figures of both original and grouped images, in separate modules.

Figure 5: Some of the closed figures extracted using the image in figure 5.a.

Figure 6: (a) Local perceptual feature representation. (b) Closed figures based representation.

4.1 Using local features

The main drawback of many of the local feature based retrieval algorithms is that they are expensive in terms of computational time. The common approach taken by many researchers to avoid this problem is to use a subset of images obtained from a hierarchical refining process. But this limits the independence of local feature based algorithms. Relaxation by elimination algorithm (RBE) [11] provides an alternative framework within which a local feature based retrieval mechanism can be implemented using CMMs which comes with a low cost and efficient hardware implementation. The RBE process obtains initial matching possibilities between nodes of the query graph and the model graphs using local features of the segments. Then, elimination of unfeasible candidates at the nodes is performed using upper bound probability estimations from neighbours.

Biederman [7] postulates that edge extraction is the initial phase of human object recognition and then it goes through the detection of a number of non-accidental properties. Motivated by this argument, we bring the RBE process into a perceptual features based framework by obtaining evidence from perceptual neighbours rather than all the neighbours in the structure as performed in the standard RBE process.

Similarity assessment using local features is based on graph representations of the image in which nodes represent the segments of the image and arcs represent different perceptual relationships between the segments. During this process the query graph is compared against all the model graphs in the database. Initial matching possibilities between nodes of the query graph and the model graphs are obtained using local features (length, orientation, curvature) of the segments. Then, elimination of unfeasible candidates at the nodes is performed using upper bound probability

estimates from the perceptual neighbours. This is an iterative process which is stopped when it obtains stability (ie. no more eliminations). The remaining candidates at the nodes are then used to calculate a final similarity score.

4.2 Using features of closed figures

In assessing similarity based on closed figures we have proposed a new method which use feature vectors of each figure as well as the knowledge on other figures which constitute the image [9]. This method allows partial matching of feature vectors and also is less affected under inclusion of additional closed figures in the query image, a drawback we observed in *asymmetric simple matching* method [6]. Moreover this method has been implemented using CMMs to achieve better efficiency [12].

Similarity assessment between a query image and the model image is performed by considering the feature vectors of each closed figure to obtain candidate matching possibilities for each query figure. This can be performed in two different ways; either in a symbolic fashion by discretising the feature components or calculating distance measures between feature vectors of the corresponding query figure and the model figure. In the next step upper bound support for each matching possibility is obtained from the contextual neighbourhood. Evidence obtained from the context and the feature vectors are thus used to calculate final similarity scores.

5 Combination of multiple similarity measures

Our aim is to investigate a unified framework for image retrieval using multiple image representations since results obtained from different representations shows that no single representation is always better than the others. Most researchers in the image retrieval community have either followed hierarchical approach [13] or used composite feature vectors [5]. The main drawback of a hierarchical combination is that it limits the decision making ability of the lower modules to the image space agreed by the higher modules. On the other hand composite feature vectors may comprise features which do not exhibit linear exchange. In unifying the multiple representations we follow an approach based on ‘results level’ evidence. Observations from document retrieval give evidence that this strategy give better results than using single representations [14]. Researchers in favour of using ranks argue that this method is better since combining scores involves combination of incommensurates which differ in range, mean and variance since they come from different representations and mechanisms. They suggest techniques based on simple data fusion logics for this task. Most common methods are based on obtaining the mean, minimum and maximum ranks to re-order the list. We observed that different retrieval modules retrieve different sets of similar images based on the nature of features used in the modules. To give more weight to the top few retrievals in the list we propose another method for combination using reciprocals of ranks [9] which gives inversely proportional weighting for ranks. In this we calculate a combined score as follows to obtain a score in re-ordering the retrieved images.

$$\text{combined score} = \sum_i \frac{1}{rank_i}$$

where $rank_i$ is the rank of the retrieved images obtained using the retrieval module i .

A recent study on retrieval mechanisms for photographic collections [15] proposes a framework for combining similarity scores using Dempster-Shafer method. In this, they provide a simplified expression for situations where there is possible evidence for only singleton hypotheses. In this framework, combination of similarity scores from two modules can be performed using the following equation.

$$CS_k = \mu_1(s_k^1/S^1)\mu_2(s_k^2/S^2) + (1 - \mu_2)\mu_1(s_k^1/S^1) + (1 - \mu_1)\mu_2(s_k^2/S^2)$$

where CS_k combined similarity score for image k

μ_i belief on module i

s_k^i similarity score for image k obtained using module i

$$S^i = \sum_k s_k^i$$

Since Dempster-Shafer method is commutative this can be repeatedly applied to any number of modules.

6 Experiments

During the experiments discussed in this paper we used two feature collections extracted using the same trademark image collection of 1000 images. Feature collection 1 was obtained using Sarkar and Boyer's voting methods and the feature collection 2 was obtained using a combination of Sarkar and Boyer's methods and Lowe's methods as explained in section 3. Both methods were aimed at extracting local perceptual features (end-point proximity, parallelism and co-linearism) as well closed figures. For the evaluation results presented in this paper, we used 10 query images for which we had similarity judgement data from trademark officers obtained during the evaluation experiments of the ARTISAN system. Figures 7 and 8 show the query images we used for this task. In calculating the retrieval effectiveness we use widely cited recall - precision graphs, and averaging of graphs is based on the macro-evaluation method suggested in [16].

Figure 7: (a) 1 (b) 2 (c) 3 (d) 4 (e) 5

Figure 8: (a) 6 (b) 7 (c) 8 (d) 9 (e) 10

For a given query image x , we can calculate the recall and precision as

recall (x) = number of objects found and relevant to x / the total number of objects relevant to x

precision (x) = number of objects found and relevant to x / the total number of objects found

According to this criteria, we can obtain pairs of recall-precision values which indicate the fraction of relevant items retrieved and the fraction of retrieved items that are relevant respectively, as we traverse from the top to the bottom of the list.

Figures 9.a and 9.b show recall-precision distributions obtained for each of the different retrieval modules used in our system. Figure 9.a shows the distribution obtained using feature collection 1, and figure 9.b shows the distribution with feature collection 2. It can be seen that the pairwise feature extraction methods have improved effectiveness of retrieval based on local features but at the expense of relatively poor performance from the retrieval module based on closed figures. This is mainly due to relatively higher number of false closed figures extracted using this method. Feature collection 1 was obtained under some user intervention and as a result some effort was made to reduce the number of false positives and negatives in extracting closure. In general retrieval modules which use features of raw images perform better than retrieval modules which use features from gestalt images. But we have observed that in retrieving some of the images, retrieval modules which use gestalt images perform better than the other modules. Experimental results shown in figures 12.a and 12.b show that combination of retrieval modules which use both image representations

Figure 9: Recall-precision distribution of different retrieval modules.

gives the best results despite better performance of retrieval modules which use features from raw images over retrieval modules which use features from gestalt images.

Figure 10: Performance of different combination strategies.

Figures 10.a and 10.b show the results obtained in combining multiple feature measurements using different strategies. Again, figure 10.a shows performance on collection 1 and figure 10.b shows performance on collection 2. It can be seen that the best performance is obtained using the combination method based on Dempster-Shafer theory. We assigned belief values of 0.6 and 0.4 for raw and gestalt image modules respectively, under feature collection 1 and belief values of 0.6 and 0.3 for raw and gestalt image modules respectively, under feature collection 2. It was observed that these belief values gave the best performance under the experimental conditions. It can be also seen that our rank based combination method using reciprocal ranks performs better than other standard rank based combination methods. The worst performance was obtained by combining using the lowest rank. We can observe that these conclusions are valid

with both feature collections.

Figure 11: Comparison of performance between the best single module and the best method for combination (a) using feature collection 1 (b) using feature collection 2.

Figures 11.a and 11.b show the comparison of performance between the best individual modules and best method of combination. These results show that as one would anticipate, we can obtain better performance than any of the single modules by combining multiple similarity measures.

Figure 12: Combination of different subsets of modules.

We conducted another experiment to compare performance in combining different subsets of modules. We separately combined retrieval modules which use feature representations (local feature representations and features of closed figures) of raw images and gestalt images. And also we separately combined retrieval modules which use local feature representations and features of closed figures. We used the same belief values for the modules as assigned in the earlier experiment. In this paper, we present results obtained using only the best combination method (Dempster-Shafer), as

observed earlier. Figures 12.a and 12.b show that best average performance is obtained in combining all the modules. These results may be summarised as follows. For the given performance metrics (recall and precision) and the given test set, the combination of outputs from separate modules using Dempster-Shafer theory gives the best average performance. Under the same conditions the best average performance is obtained by combining all the modules. Local perceptual feature extraction using Lowe's pair wise method has improved retrieval effectiveness of local feature based retrieval modules over bin voting methods, at the expense of degradation of performance of closed-figure modules.

7 Conclusions and Future Work

We have presented a novel integrated framework for image retrieval using multiple feature interpretations obtained using gestalt principles. In this, we have shown that we can obtain better performance than any of the single modules by combining multiple similarity measures. We have presented the results using two feature collections, one obtained under certain amount of user interaction and the other obtained without any user interaction. We are currently evaluating the performance of our system using similarity judgement data obtained with a set of 20 human subjects. We are also in the process of implementing the whole system using correlation matrix memory neural networks.

Acknowledgements

We would like to thank Dr. John Eakins at the University of Northumbria at Newcastle and Dr. Philip Quinlan at the University of York for the substantial benefits obtained from the discussions with them. Financial support for this project comes from the association of common wealth universities.

References

- [1] S. Santini and R. Jain. Similarity matching. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, , 1996.
- [2] M. Turner and J. Austin. Matching Performance of Binary Correlation Matrix Memories. *Neural Networks*, Elsevier Science, 1997.
- [3] J. Austin , J. Kennedy, K. Lees "The Advanced Uncertain Reasoning Architecture. In *Proceedings of the Artificial Neural Networks and Expert Systems Conference*, June 1995
- [4] T. Kato. Database Architecture for content based image retrieval. *Image storage and Retrieval Systems-SPIE*, vol 2185, pages 112-123, 1992.
- [5] J. K. Wu, C. P. Lam, B. E. Mehtre et al. Content-based retrieval for trade mark registration. *Multimedia Tools and Applications*, vol 3, pages 245-267, 1996.
- [6] J. P. Eakins, M. E. Graham, J. M. Boardman et al. Retrieval of trademark images by shape feature. *British Library Research and Innovation Report* 26, 1996.
- [7] I. Biederman. Recognition by components - A theory of human image understanding. *Psychological Review*, vol 94, no 2, pages 115-147, 1987.
- [8] S. Sarkar and K. Boyer. *Computing perceptual organization in computer vision*. World Scientific Publishers, 1994.
- [9] S. Alwis and J. Austin. A novel architecture for trademark image retrieval systems. In *Proceedings of the challenge of image retrieval*, 1998.
- [10] D. G. Lowe. Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, vol 31, pages 355-395, 1987.
- [11] M. Turner and J. Austin. A neural relaxation technique for chemical graph matching. In *Proceedings of the Fifth International Conference on Artificial Neural Networks*, Cambridge, UK, Editor: M Niranjana, IEE Publishers, July 1997.
- [12] S. Alwis. Content based retrieval of trademark images. In *Thesis Proposal*, University of York, UK, 1998.
- [13] A. K. Jain and A. Vailaya. Shape-based retrieval: A case study with trademark image databases. vol. 31, no 9, pages 1369-1390, 1998.
- [14] N. J. Belkin, P. Kantor, E. A. Fox, J. B. Shaw. Combining evidence of multiple query representations for information retrieval. *Information processing and management*, no 31, pages 431-448, 1995.
- [15] J. M. Jose, D. J. Harper. A retrieval mechanism for semi-structured photographic collections. *Lecture notes in computer science* 1308, pages 276-292, 1997.
- [16] C. J. Van Rijsbergen. *Information retrieval*. Butterworths, London, 1979.