Technical Report

Query by Sketch for Trademark Image Retrieval

Liang Shuang[1]

Department of Computer Science and Technology, Nanjing University, P R China

1. Introduction

Despite over a decade of research into content-based image retrieval (CBIR), the task of finding a desired image in a large collection remains problematic. Even in application areas where there is a clear need for effective image retrieval, such as medical diagnosis and trademark registration, current technology fails to meet user needs. Much existing research has concentrated on retrieval techniques for natural images (typically photographs of natural scenes or objects), using various combinations of extracted colour, texture and layout feature [VT01]. Techniques for the retrieval of figurative images, a class of artificially-produced images which includes icons, trademarks, coats of arms, and clip-art images, have received less attention, even though there is evidence, e.g. [WLMGDN96], that these images require different techniques for effective retrieval.

Usually, trademark images do not readily lend themselves to retrieval on the basis of name. And current techniques for the retrieval of such images are demonstrably inadequate [E01]. In order to improve the retrieval performance and find out the most perceptually-relevant trademark images, we aim to invent and develop new techniques for incorporating free-hand sketches into retrieval of trademark images. This is because, hand-drawn sketch is the most natural and useful graphic media for representation human-perceptual information. Users may draw sketches to interact with the retrieval system, thus signify their perception or understanding of trademark images.

Our work will be based on the sketch interface and perceptual information extraction. Following aspects will be covered:

- Design of interaction with sketches.
- Sketch processing and human-perceptual information exaction.
- Incorporating sketch information and image processing results.
- Refining the similarity matching to obtain an optimal overall similarity.

Finally, the new manner of trademark retrieval with sketches will be experimentally verified in a prototype system.

2 Related works

Incorporating sketches into image retrieval system falls into the category of content-based visual information retrieval, and there has been some previous work about it.

IBM's Query-By-Image-Content (QBIC) system [ABFH95] was the first CBIR system, and it offers a sketch input interface and allows queries using multiple features based on image contents. Users draw a sketch on an input panel, and global features such as area, circularity,

[1] This work was done while at Utrecht University.

eccentricity, etc., are then used in shape matching. [MDB98] proposed another approach to sketch-based image retrieval by exploring Curvature Scale Space (CSS) to match the shape contours. The [SM99] system uses Fourier descriptors for shape matching, and they use relevance feedback to refine the retrieval performance. However, all these approaches have an assumption that the query consists of only one single shape.

Other approaches like Leung [L03] proposed to create multiform representations of a sketch under user variations, and he used relevance feedback to improve the retrieval performance. But his approach pays too much attention on stroke information and is very sensitive to the shape deformation. Fonseca et al. [FJ02] used stroke hierarchy and shape geometry to retrieve complex drawings. But only stroke hierarchies may not always be the most efficient method for comparison. Meanwhile, they do not deal well with curve segments.

Within a trademark retrieval system, Leung & Chen [LC02] aim to bridge the gap between shape's boundary and the shape's interior (fill area) by unifying skeletons and edge detection approaches. Their system segments the image into regions. For each region, the system then either performs edge detection or performs thinning to produce a skeleton. They believe that it is advantageous to use different methods under different situations. However, Carlin [C01] noted that skeleton features perform well on application specific criteria but are not robust to shape deformation.

All of the above systems use a complete sketch as the query image. The gap between the sketch and image processing are neglected. Therefore, we intend to develop a multi-channel query interaction and combine the sketch and image information together to obtain a better query representation.

3 Trademark Retrieval Framework

This section describes the overall objectives and working framework of the retrieval with sketch interface and relevance feedback for trademarks, which undertakes several steps.

3.1 Objectives

We aim to focus on several aspects, including sketch interaction design, sketch and image processing, similarity measure refinement and performance evaluation. So the prime objectives are:

1. Sketch interaction - Sketch representing the users' perception could be drawn on both the query and the result sides. There are various drawing manners, such as labeling, editing, skeleton, etc, among which we should design the specific interaction manner.
2. Sketch processing
    - 2.1. Sketching together with image as query - Combining sketches with original image query or the automatic segmentation results to develop a more specific query.
    - 2.2. Sketching with return set to carry out feedback - Labeling or editing the candidate results set, and explore the common information from the drawn results to refine the query to get better results.
3. Refining the similarity measure - use relevance feedback to capture the positive examples and the negative ones, and then adjust the similarities according to the effectiveness of these labeled samples.
4. Performance evaluation - carry out experiments to verify the effectiveness of sketch interface.

3.2 Retrieval Framework

Considering the Gestalt[2] principles [W38, K63], our framework begins by segmenting trademark images into distinct shapes (or primitives) using a closed shape identification algorithm. A simple edge detector would not be sufficient as many images in our test set are noisy and this noise causes small gaps in the shape boundaries so these gaps need to be closed. We chose to refine and adapt Saund's closed shape identification algorithm [S03] within the PROFI project [HEA07] to find the perceptual shapes present in an image.

Due to noises in the original trademark images, some segmentation results of the images may not be perfect. We then incorporate a query refining interface of sketching based on the initial automatic segmenting to overcome this problem. Users may edit the automatic segmentation results with freehand strokes and gestures to create a more perceptually suitable segmentation for the query image, therefore refining the query to form a better shape representation of the image and get improved results. These perceptual closed shapes combined together by automatic segmentation and sketching will be considered as the representation of the trademark images.

A probabilistic shape matching algorithm [ASS07] will be taken for calculating shape similarities. While because the geometric shape matching algorithms are slow, they cannot be used directly during a query on a large database. Hence, standard database indexing techniques cannot be used. Rather than using heuristic feature vector methods, which are not very discriminating, we will use indexing of the pattern metric space [LDHA07, DLV07], which might result in a few false positives, but guarantees avoiding false negatives. Similar trademark images will then be returned based on their distances in the metric space.

Here we can again use a sketch interface to carry out a relevance feedback section. Users will label each shape (or primitive) of the image to indicate the system the relevant and irrelevant parts of a result image to the query image. Then for each image in the return set, a set of relevant shapes and a set of irrelevant shapes will be obtained. These labelling information acquired on the whole candidate set will be analyzed and employed to refine the similarity measure and generate a refined result set. This feedback procedure could be carried out repeatedly to improve the performance step by step.

In the following, we will discuss the sketch interaction manner first, and then carry out the corresponding retrieval process in detail.

4 Sketch Interaction Design

First of all, we should design the sketch interaction manner, since the system must get an understanding of what the user want to represent. There are different ways of drawing a sketch when interacting with the trademark retrieval system, and meanwhile different drawing manners often indicate different meanings.

Usually, the electronic drawing of sketches takes place in the basic format of strokes and gestures. Strokes are the freestyle lines or curves people depict on the drawing panel to represent their ideas about image outlines, while gestures are some predefined symbols people draw on the existing image to carry out certain operating commands, e.g. copy, move. These gestures will improve the ability and flexibility of sketch representation.

[2] The Gestalt principles refer to the shape-forming capability of human vision. In particular, they refer to the visual recognition of figures and whole shapes rather than just "seeing" a simple collection of lines and curves.

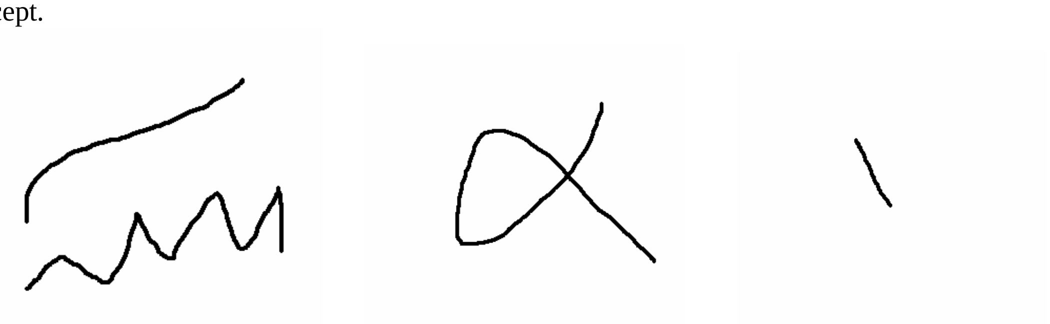
In our trademark retrieval system, we will use the left/right button of mouse drawing to indicate the stroke/gesture mode respectively. And the sketching interface could both be employed on the query end to refine the segmentation of the query and on the result end to carry out feedbacks.

4.1 Sketching on the Query

When a user draws a sketch on the query, he/she may extremely construct a complete new sketch and use it as query. In this manner, user could fully express the query intention in his/her minds and retrieve whatever he/she likes. But considering that the drawing skill of different people is different, the same query drawn by sketches usually differs a lot from time to time and from person to person. This inaccuracy and ambiguous of sketch properties of course allow more flexibility and variety kinds of query representation, however they may conflict with the ground truth. This is because the ground truth is pre-built based on the common sense, while sketches usually represent individual user's query interests.

In order to stand with the existing ground truth and make use of the automatic segmentation of image processing, we develop a multi-channel interaction manner for submitting a query. In our trademark retrieval system, both the Gestalt-based automatic image segmentation results and sketches are incorporated to construct a better representation of the image shapes.

Here we design two kinds of gestures named delete and break. Figure 1 shows the sketch input manners. Generally, these sketch interaction manners play different roles in representing the concept.



a) Stroke b) Delete Gesture c) Break Gesture

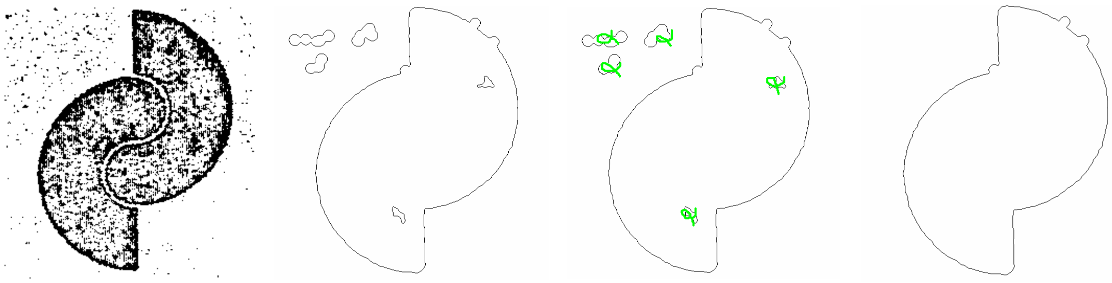
Figure 1. Demonstration of user sketches

a) is freehand strokes made of lines and curves; b) is a "delete" gesture to delete a whole or part of the shape; and c) is a "break" gesture to indicate a breaking point on the shape

The detail operating manners are as follows:

- Users will first examine the segmentation results to check if the image processing has found out the main perceptual closed shapes correctly. If the system has successfully identified the main shapes of the query, then these shapes could be used as the representation of the query directly and submitted to the server.
- Otherwise:
  - For unwanted shapes – shape deletion

    Because of the noises and poor quality of the original trademark images, some unwanted parts of the trademark maybe wrongly detected by the segmentation process. In these cases, we will use the "delete" gesture to delete these wrong shapes. Users could just draw the "delete" gesture on top of the redundant shapes and they will be removed by the system automatically. This deleting operation is very convenient to carry out. Figure 2 is a demonstration of the shape deletion operation.



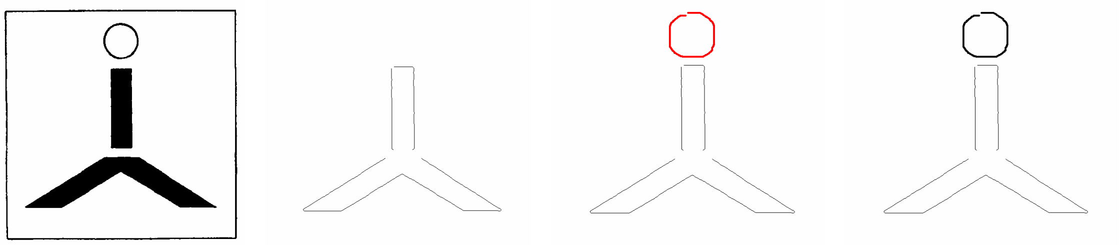
a) Original image b) Segmentation c) Shape deletion d) Refined query

Figure 2. Shape deletion.

a) is an original trademark image with poor quality; b) is the automatic segmentation result obtained by image processing; c) is the shape deletion carried out with sketching; and d) is the final refined shape of the query.

- For missing shapes – shape addition

  If some main parts of the query image are missing in the automatic segmentation detection, we could outline the missing constitutes of the query using sketch strokes. Users could draw strokes along the shapes of the image to signify the main contours or skeleton of a given image query. We will set the original image as the background of the drawing panel, so these user outlines will be relatively consistent with the existing query image. Users could also add detail strokes to the current query to form a more specific query. Figure 3 shows what happens when the shape deletion operation is carried out.



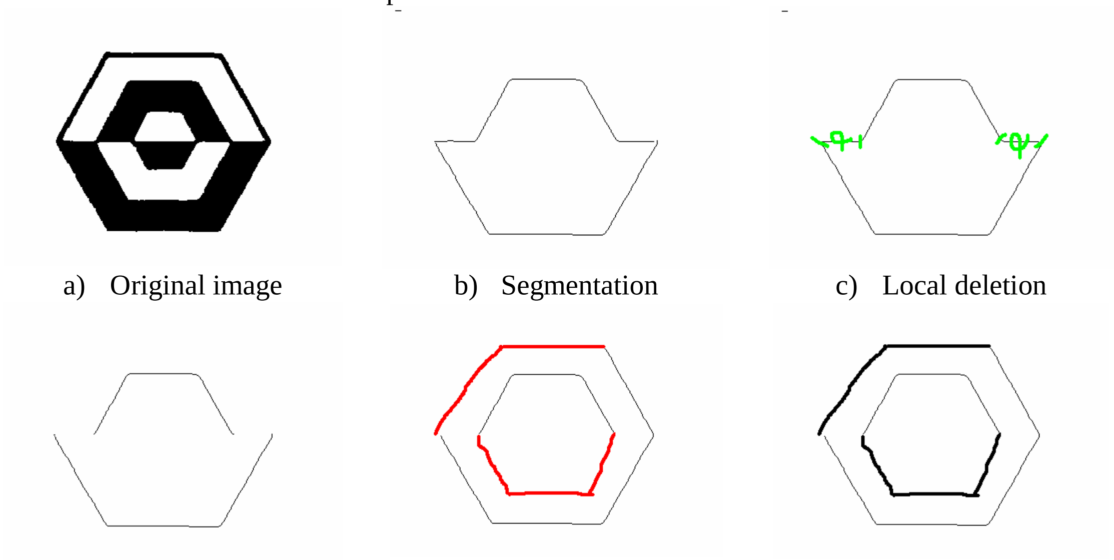
a) Original image b) Segmentation c) Shape addition d) Refined query

Figure 3. Shape addition.

a) is an original trademark image; b) is the automatic segmentation result with a circle missing; c) is the shape addition by adding shape strokes; and d) is the final refined shape of the query.

- For inaccurate shapes – shape editing

  We could do a shape editing in case of the shape is detected but wrongly depicted. This means that the some parts of the trademark image are identified properly, but others are not. In order to deal with this kind of situation, we could employ a shape editing operation which is made up by two steps of local deletion, which first deletes the error parts, and local addition, which then use sketch to fill the missing parts in a correct way. As we could see in figure 4, when doing local deletion on the automatic segmentation results, users should first signify two “break” gestures to indicate the section of the shape which he/she wants to delete, and then draw a “delete” gesture on this section to perform a delete operation, as in figure 4c. Here only the right parts remains for the following editing (figure 4d). Local addition is then carried out by filling in the missing parts by freehand strokes, as shown in figure 4e. The newly added strokes will be merged to the remaining shapes after local deletion to form new closed shapes which form the shape representation of the trademark image query, and could be used for the following similarity measure. To some extent, the shape editing operation is a combination of shape deletion and addition, only on local sections of the current closed shape.



a) Original image b) Segmentation c) Local deletion

d) Shape after local deletion e) Local addition f) Refined query

Figure 4. Shape editing.

a) is an original trademark image; b) is the automatic segmentation result with parts of it being correct and others not; c) is local deletion to remove the inaccurate parts; d) is the intermediate results after local deletion; e) is local addition to correct the shapes; and f) is the final refined query shape.

The image segmentation offers an initial shape representation standing consistent with the database ground truth and based on which the sketch interface will improve the representation in case of segmentation failures. After the automatic segmentation has identified out the main perceptual closed shapes of the query image, the operations of shape deletion, addition and editing by sketching could be employed together to refine the image processing results. Finally, the refined segmentation result will be considered as the representation of the query trademark and submitted to the server to retrieve similar images. In this way, the retrieval system could make use of both the image processing results and the user sketching refinement to achieve more accurate query intentions.

4.2 Sketching on the Results

When the retrieval results are obtained, users could also sketch on the result set to perform feedbacks. We could label on the different shapes of each image to indicate relevant and irrelevant primitives. Some pre-defined symbols maybe drawn to mark these parts in the image, such as cross, circles, etc. Here marking out the main region by these kinds of strokes or gestures is enough.
(underway…)

5 Query Refinement

Given a query image and its segmentation result, we could sketch on the interface to edit the main parts of the query. This way, the users' perception information will be obtained online, which will be incorporated with the automatic image processing results to form a new specific query. Two sub-steps are taken here, in which the trademark images are first automatically segmented into closed shapes, and then combined with sketch editing to refine the segmentation to achieve a more accurate query representation.

We chose the segmentation algorithm in PROFI project [HEA07] to find the perceptual shapes present in an image.When we refine the query by sketch interaction, sketch segments will be incorporated with the automatic image segmentation results.

If the sketch format is stroke, new user information included in sketches will added to the current query to form a new feature representation. For every closure in sketches, it represents some closure in the image. Therefore, we will first extract the sketch closures which are made up of strokes by means of determine their end-point proximities. Two strokes are considered as the same component if there is a pair of end points between these strokes that are close. Or in other words, if the end points of two strokes are proximate, they are merged into one new stroke. The gap between the adjacent end points will be filled by a line segment connecting the two end points. Moreover, if one of the stroke’s end points is adjacent to the other end point of itself, then it will be classified as a sketch closure.

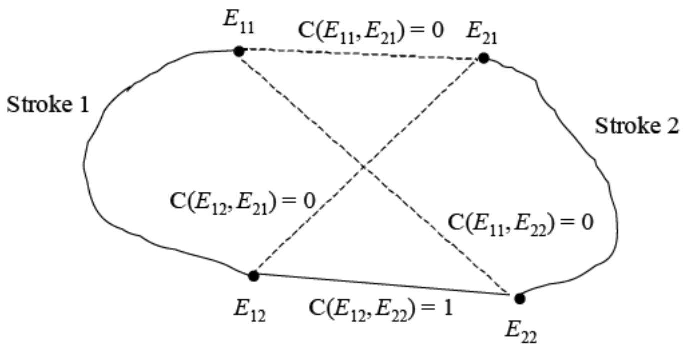
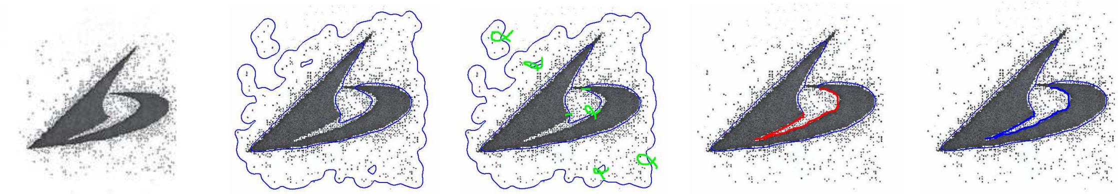


Figure 5. Adjacency between a pair of strokes

Figure 5 shows an example of a pair of adjacent strokes, where $C(E_a,E_b)$ denotes the connectivity between the end points E_a and E_b, where $C(E_a,E_b)=1$ means that E_a and E_b are adjacent and $C(E_a,E_b)=0$ means that E_a and E_b are not. There is a tradeoff in choosing the threshold for the end point proximity. When the threshold is too small, then the number of strokes in the same component will be smaller, thus reducing the complexity of the search space. However, this also means that it is more sensitive to user variation since some people may leave a larger space between neighbouring strokes. On the other hand, when the threshold is too large, it is less sensitive to user variation but at the same time the number of strokes in the same component may be large such that it may defeat the purpose of dividing the sketch into components. In our implement, if the distance between the stroke end points is smaller than 1/10 of the length of the shorter stroke of the two, then they are considered to be adjacent. This approach is invariant to translation, scale, rotation, and arbitrary drawing orders.

If the sketch format is gesture, then Rubine features [R91] are employed to recognize the type of the gesture. If the gesture type is “delete”, then the intersect shape or parts of the shape will be found out by the retrieval system and removed from the segmentation. If the gesture type is “break”, then the intersecting point of the break gesture and the shape outline will be calculated and recorded. Another break point should also be marked out to indicate a section of the shape closure. If there is a following “delete” operation happens on a section of the shape between to break points, then this section will be deleted from the closure. The new strokes filling up the deleted part will be merged to the remaining section of the shape to form a more accurate closure for the following shape matching process. Figure 6 is a demonstration of the sketch interface.



a) Query image b) segmentation c) Gestures d) Strokes e) Results

Figure 6. Sketching interface.

6 Similarity Measure Refinement

After the retrieval system returns the results, users may also use sketches to label the result set to offer more feedback information. Since images are made up of multi-shapes, and the similarity between different images are calculated based on the weighted sum of the individual shape similarities, we are going to label out the relevant and irrelevant parts of the images. These information could first be extracted within single object or image, and then be combined to get an overall feedback information. And this feedback information could be employed to refine the weighting scheme of similarity measurement to get the optimal combination.

Currently, every shape gets an absolute weight value, which equals to the square root of the shape's size (perimeter of the shape's bounding box that maximizes the aspect ratio). For the new segments which are added from sketch closures, their weights should be calculated in the same way with the original image segments. Meanwhile, all new weights should be normalized again. In this manner, the weighting scheme represents the user understanding to a larger extent and gets better retrieval results.
(underway…)

7 Performance Evaluation

Finally, we aim to test both the efficiency and effectiveness of our method through an experimentation platform. Both the database and ground truth are important. We use a subset of the trademark images from the UK PTO dataset used in the ARTISAN project [ESB96] as our database.

A demo of the sketch interface could be found on http://profi.cs.uu.nl/shuang/index.html. Evaluation metrics such as recall and precision should be calculated based on average retrieval performance.

Acknowledgement

I would like to thank Aktor Knowledge Technology for their help with establishing the ground truth. This work is supported by the following grants from:

- The FP6 IST project 511572-2 PROFI (Perceptually-Relevant Retrieval of Figurative Images), funded by the European Community under the "Information Society Technologies" Programme.
- The National Natural Science Foundation of China [Project No. 69903006 and 60373065].
- The Program for New Century Excellent Talents in University of China.

References

[ABFH95] J. Ashley, R. Barber, M.D. Flickner, J.L. Hafner, D. Lee, W. Niblack, and D. Petkovic. Automatic and Semiautomatic Methods for Image Annotation and Retrieval in QBIC, *Proc. Storage and Retrieval for Image and Video Databases Conf.*, Feb. 1995.

[ASS07] Helmut Alt, Ludmila Scharf , Sven Scholz. Probabilistic Matching and Resemblance Evaluation of Shapes in Trademark Images, ACM International Conference on Image and Video Retrieval, (CIVR07). July 9-11 2007, 2007.

[C01] M Carlin. Measuring the performance of shape similarity retrieval methods. Computer Vision and Image Understanding, 84(1): 44-61, October 2001, special issue on empirical evaluation of computer vision algorithms, ISSN: 1077-3142.

[DLV07] Fatih Demirci, Reinier van Leuken, Remco Veltkamp. Shape Indexing through Laplacian Spectra, VMDL'07, 13-14 September 2007.

[E01] Eakins J P. Trademark Image Retrieval. Principles of Visual Information Retrieval, M. Lew (ed.), Springer, Berlin. ISBN 1-85233-381-2, 319-350.

[ESB96] J. P. Eakins, K. Shields, and J. M. Boardman. Artisan: A shape retrieval system based on boundary family indexing. Storage and Retrieval for Image and Video Databases (SPIE), pages 17–28, 1996.

[FJ02] M.J. Fonseca and J.A. Jorge. Towards Content-Based Retrieval of Technical Drawings through High Dimensional Indexing Proc. of the 1st Ibero-American Symposium in Computer Graphics, Portugal, July 2002, pp263-270.

[HEA07] V. Hodge, J. Eakins, and J. Austin. Inducing a perceptual relevance shape classifier. ACM International Conference on Image and Video Retrieval, (CIVR07). July 9-11 2007, 2007.

[K63] K. Koffka. Principles of Gestalt Psychology. Harcourt Brace. New York, 1963.

[L03] Howard Leung.Representations, Feature Extraction, Matching and Relevance Feedback for Sketch Retrieval, Ph.D dissertation, CMU, June 2003.

[LC02] W.H. Leung, and T. Chen. "Trademark retrieval using contour-skeleton stroke classification", *IEEE Intl. Conf. on Multimedia and Expo.* (ICME 2002), Lausanne, Switzerland, August 2002.

[LDHA07] Reinier van Leuken, Fatih Demirci, Vicky Hodge, Jim Austin. Lay-out indexing of trademark images. ACM International Conference on Image and Video Retrieval, (CIVR07). July 9-11 2007, 2007.

[MDB98] Matusiak S., Daoudi M. and Blu T. Sketch-Based Images Database Retrieval. Proc. of the Fourth Intl. Workshop on Multimedia Information Systems (MIS'98), pp. 185-191, Istanbul, Turkey, September 1998.

[R91] Rubine, D. Specifying Gestures by Example. Computer Graphics, 25,4, (1991) pp. 329-337.

[S03] E. Saund. Finding Perceptually Closed Paths in Sketches and drawings. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 25(4): 475-491 (2003).

[SM99] Sciascio E. D. and Mongiello M. Query by Sketch and Relevance Feedback for Content-Based Image Retrieval over the Web. *Journal of Visual Languages and Computing, 10(6), pp. 565-584, 1999.*

[VT01] Veltkamp, R C and Tanase, M. Features in Content-Based Image Retrieval Systems: A Survey. *State-of-the-Art in Content-Based Image and Video Retrieval. Kluwer, NY*

[W38] M. Wertheimer. Laws of organization in perceptual forms (1923)., 1938.

[WB91] D.M. Wuescher, and K.L. Boyer. Robust contour decomposition using a constant curvature criterion. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*: 13(1): 41-51 (1991).

[WLMGDN96] J. K. Wu, C. P. Lam, B. M. Mehtre, Y. J. Gao and A. Desai Narasimhalu Content-based retrieval for trademark registration. Multimedia Tools and Applications 3, 245-267.