

SIMILARITY EVALUATION BASED ON IMAGE PRIMITIVES

Sven Scholz
Institute of Computer Science
Freie Universität Berlin
Berlin, Germany
email: scholz@inf.fu-berlin.de

ABSTRACT

A new framework for the perceptually relevant comparison of figurative images, especially trademark logos is presented in this paper. Images are divided into salient geometric figures such as rectangles, ellipses, and triangles. Parts not fitting into any of those simple classes are represented by their boundaries. The figures are classified, related, and weighted according to their perceptual relevance. For the comparison of two images the figures and the relations are compared independently from each other. For the comparison of single figures a simple measure of similarity based on registration techniques is applied, which is noise tolerant and shows good results for figurative images that have no spatially independent parts. The similarity of the images is then determined by the similarities of the figures and the relations for the best match. The algorithms were tested with a collection of 10 745 trademark images from the UK PTO, with the same set of 24 reference queries that were used to test the ARTISAN System. Each query consists of a query image and a list of relevant images, compiled by experienced trademark examiners. The experiments show that the presented approach allows for a considerable improvement of content based image retrieval in trademark images.

KEY WORDS

shape, content based image retrieval, trademark images

1. Introduction

For the comparison of figurative images that can be represented by a single closed (polygonal) curve, a variety of methods were invented that show respectable performance. Most trademarks on the other hand are way more complex and therefore the comparison has to consider many more aspects. Although one of the laws invented by Gestalt Theory states that configurations cannot be analyzed into parts and relations [1], for such multi-component images the comparison based on the individual image components is more effective than a comparison based on the whole image [2].

With regard to the ground truth provided by professional trademark examiners (see section 3), some observations can be made which are formulated as follows:

- People look for figures in the image that can easily be memorized. These figures may be abstract figures such as squares, circles, and triangles or figures of everyday life such as letters, digits, and stylized eyes or paperclips. If such figures exist within the image, their concrete proportions and positions play a minor role (see the appendix figs. 2 and 3).

This is supported by the facts that:

- a small number of common shape elements can form a basis for humans to discriminate between a wide variety of images [3] (cited in [4]).
- "there is an unconscious effort to simplify what is perceived into what the viewer can understand". [5] (cited in [6])
- If the image consists of spatially independent parts, the size of the gaps inbetween plays a minor role (see the appendix fig. 4).
- If an essential part of the image is surrounded by a frame, the shape of the frame and even the existence of the frame play a minor role (see the appendix fig. 4). In [7] experiments on the way humans decompose figurative images were made. 5 of the images had a frame, for 3 of them all subjects completely ignored the frame and for 1 image only the second least significant decomposition (out of 9) contained the frame.
- Looking at a figurative image, the number of essential parts that are perceived is typically very small. For example in a regular pattern of little circles, one does normally not discriminate between the different circles, but group them together to a 'pattern of circles'. Moreover when comparing such patterns it plays only a minor role if 16 circles form a 4×4 grid or if 25 circles form a 5×5 grid.

Our Framework for improving the comparison of figurative images is based on a very simple idea: try to characterize a figurative image the same way humans would do. If there is a circle in a triangle, characterize it as 'a circle in a triangle', if there is something never seen before, characterize it as 'something never seen before' and describe it by what is known about it — in our case its boundaries. Many patent offices use such a characterization based on the so called Vienna classification [8]. The codes for the examples given in fig. 1 would possibly be '26.3.10 Triangles



Figure 1. actual trademark images — some easy to describe by geometric primitives and some not.

containing one or more circles, ellipses or polygons’ and ‘26.13.25 Other geometrical figures, indefinable designs’ respectively.

Following this idea in our approach, an image is divided into a set of (not necessarily spatially independent) parts — preferably simple and salient geometric figures. These parts are classified, weighted, and related. The relationships are weighted as well. Comparing two images is accomplished by searching for subsets of the parts and their relations that match well.

The comparison of the parts is done independently, leaving aside their relative sizes and positions. It can be done using a similarity measure that works well for shapes whose parts lie close together whereas the resulting measure can handle arbitrary composed shapes.

In [9] a similar approach of dividing the images into geometric primitives and finding a match between these primitives is proposed. Its main drawbacks are 1.) that the comparison of the primitives does not prescind from their concrete positions and 2.) that the similarity between primitives belonging to different categories is defined as being zero, which is contrary to human perception e.g. when comparing a circle and a regular 12-gon.

We do not assume that all parts of all images can be replaced by high level primitives in a meaningful way. Analysis of annotations of trademark images shows that a considerable number of images needs different treatment (see 2.1). In addition, whenever a measure of similarity depends on the way the images are decomposed, there is the risk of underestimating the similarity just because two images get decomposed in different ways (e.g. two triangles forming a square vs. a square plus its diagonal).

For these reasons the comparison based on image primitives is not used as a stand-alone measure of similarity, but it is used in a framework to improve the results of the underlying, simple measure of similarity. Images are first compared using the underlying similarity measure and only if the decomposition leads to a higher value of similarity it is used. In this way the advantages of using high level features is combined with the robustness of the simple, low level comparison.

2. Comparison based on Image Primitives

For the comparison based on image primitives an underlying measure of similarity (e.g., the measure mentioned in

section 2.4) is used, that assigns every pair of images or image parts their value of similarity $s \in [0, 1]$.

It is assumed that figurative images are given as a set P of polygonal boundary curves $p_1 \dots p_m$. Based on these polygonal curves a set F of figures $f_1 \dots f_n$ is extracted and their relations $R = r_{1,2} \dots r_{n,n-1}$ are computed.

The process of figure detection is not described in detail here, but the decomposition is assumed to be part of the input. For the experiments in sec. 3 however, a simple proof-of-concept implementation was used.

2.1 Figures

The figures can either be simple geometric objects (*image primitives*) or more complex objects. The primitives considered in our implementation are:

- ellipses (as a generalization of circles)
- rectangles (as a generalization of squares)
- triangles

The choice of these three types of primitives is based on their frequency of occurrence: In a collection of 1 762 395 trademark images for which we had access to the frequencies of the vienna codes, more than 23% of the images contain rectangles (as a special case of quadrilaterals) and 15% contain circles. These two topmost frequencies are followed by ‘lines, bands’ (which leaves open how to deal with geometrically), and by triangles.

Although these primitives occur very often, more than one half of the images is not annotated with one of them at all. Even with an increased set of primitive types, there will be unclassifiable parts remaining for which even humans have no proper category. The parts of the image that cannot be represented by the three types of primitives are categorized as

- convex polygons
- arbitrary sets of polylines

Analogously to concentric circles, ‘concentric’ ellipses, rectangles, triangles, and convex polygons resp. are conflated to a single figure with multiple layers.

2.2 Relations

For a pair $(f_i, f_j) \in F \times F, i \neq j$ of figures the relation $r_{i,j}$ consists of numerical values reflecting

- the size of f_j relative to the size of f_i (The size of a figure is defined to be the perimeter of the bounding box that maximizes the aspect ratio.)
- the relative distance of f_j to f_i (The distance of the bounding boxes’ centers relative to the size of f_i .)
- the qualitative relation, i.e., the similarity of f_i and f_j under translations, rigid motions and under reflections.

2.3 Comparison of two Images

For the comparison of two images I^1 and I^2 the relevance w_F of the figures and the relevance w_R of the relations is preset such that $w_F + w_R = 1$ — for images consisting only of one type of figures, e.g., only squares, the relations between these figures are of greater importance than for images consisting of totally different figures. The figures and relations get weights $w(f_i)$ and $w(r_{i,j})$ according to their salience, such that for each image all weights sum up to 1, namely: $\sum_{f \in F} w(f) = w_F$ and $\sum_{r \in R} w(r) = w_R$.

For every pair $(f_i^1, f_k^2) \in F^1 \times F^2$ of figures a value of similarity $s(f_i^1, f_k^2) \in [0, 1]$ is computed, using the underlying measure of similarity. For every pair $(r_{i,j}^1, r_{k,l}^2) \in R^1 \times R^2$ of relations a value of similarity $\tilde{s}(r_{i,j}^1, r_{k,l}^2) \in [0, 1]$ is computed, using a simple measure of similarity.

Let \mathcal{M} be the set of all one-to-one matchings between figures of image I^1 and image I^2 . The value of similarity S of the two images is then defined as the weighted sum of the similarities of the matched figures, plus the weighted sum of the similarities of the (implicitly) matched relations:

$$S(I^1, I^2) = \max_{M \in \mathcal{M}} \left\{ \sum_{(f_i^1, f_j^2) \in M} s(f_i^1, f_j^2) \cdot \frac{w(f_i^1) + w(f_j^2)}{2} + \sum_{\substack{(f_i^1, f_k^2) \in M \\ (f_j^1, f_l^2) \in M}} \tilde{s}(r_{i,j}^1, r_{k,l}^2) \cdot \frac{w(r_{i,j}^1) + w(r_{k,l}^2)}{2} \right\}$$

The problem of determining whether $S(I^1, I^2) \geq \theta$ for a given threshold $0 < \theta \leq 1$ is an extension of the *quadratic assignment problem* (see e.g. [10]) and therefore is NP-complete. Since the number of essential parts that are perceived is typically very small, the admissible number of figures that represent an image can be bounded by a small constant (see section 3). Thus, the value of similarity $S(I^1, I^2)$ may be computed using a branch and bound algorithm for enumeration of the promising matches.

2.4 Proof of Concept Implementation

Several estimates in the implementation are arbitrarily fixings. Since comprehensive psychological studies on e.g. the relationship between the size and the perceived relevance of figures or on the effect of repeated figures were not available (or at least unknown to the author), the formulas used stem from qualitative considerations but do not necessarily comply with reality in their quantitative behavior.

Weights Every figure f_i gets an absolute weight $w_a(f_i)$ which equals the square root of the figure's size (perimeter of the figure's bounding box that maximizes the aspect ratio). Every relation $r_{i,j}$ gets an absolute weight $w_a(r_{i,j})$

based on the absolute weights of the figures f_i and f_j . The weights w used in the comparison are derived from these absolute weights by normalizing them such that $\sum w(f) = w_F$ and $\sum w(r) = w_R$. If two images I^1 and I^2 with different numbers n^1, n^2 of figures are compared, only the relations for $n_{min} = \min(n^1, n^2)$ figures may be selected. In this case the weights of the relations of the image consisting of more figures are adjusted such that the maximum sum of the weights of relations between a n_{min} -subset of the figures equals w_R .

Frames A frame is a — mostly rectangular — part of an image that only surrounds the essential parts, but has only very limited or no influence on the perception of the image. For every figure the likeliness of being a frame is rated based on the following propositions:

- frames are convex and symmetric
- frames contain at least one complex figure or two primitive figures
- frames are not too small compared with surrounding frames
- frames are not surrounded by something that is not a frame

Based on this likeliness the weight of a frame figure is decreased by a factor $\in [1.0, 2.0]$.

Repetitions If a logo contains groups of identical figures, the concrete number of these identical figures plays only a minor role in comparison (see the appendix fig. 3) and some trademark images even contains miscellaneous variants of the actual logo (see the appendix fig. 2). Therefore the weights of such copies are reduced.

Underlying Measures of Similarity For the underlying measures of similarity between figures or relations respectively, values between 0 and 1 are required so that the resulting value will range from 0 to 1. In [11] such a normalized measure of similarity is described which works respectably well for figurative images whose parts lie close together. The basic idea behind this approach is to find a (similarity) transformation $t : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ that maps parts of the one figure f^1 into the proximity of corresponding parts of the other figure f^2 and the similarity is rated based on proximity and parallelism of $t(f^1)$ and f^2 . For the comparison of image primitives (ellipses, rectangles, triangles) the values of similarity may be predefined, for the comparison of primitives with complex figures the values may be precomputed so that only the values for the comparison of complex figures have to be computed online.

The similarity of 2 relations $r_{i,j}^1$ and $r_{k,l}^2$ is computed by a formula based on the difference in relative distances, the difference in relative sizes, and the qualitative relations i.e. the similarity $s(f_f, f_j)$ under translations, rigid motions, or reflections.

3. Experimental Results

The retrieval performance was tested with the same set of 10 745 trademark images and the same 24 reference queries that were used to test the ARTISAN System [12]. Each query consists of a query image and a list of relevant images from the test set (including the query image). The lists of relevant images had been compiled by experienced trademark examiners (examples of query images with some relevant images can be found in Appendix A). Most of the images depict abstract geometrical figures — black shapes on white background — but some of the figures are hatched or have texture: the number of closed contours (distinguishable black and white areas) exceeds 1 000 for about 800 images (7 %) and the maximum observed is even 92 436.

From every image the set of polygonal boundary curves was extracted and polygons belonging to noise and texture were eliminated¹. The remaining closed contours for which every vertex corresponds to a pixel, were then simplified using the Douglas-Peucker algorithm [13] (cited in [14]).

The segmented images were automatically decomposed by detecting image primitives and grouping the remaining parts based on their proximity. For images with more than one possible decompositions a value of *simplicity* was computed for every decomposition (based on regularity of the figures, symmetries, and number of figures). More than 90 % of the images were decomposed into at most 6 figures, the maximum number of perceptually relevant figures in an image that were identified by the segmentation was 14.

For each of the 24 queries, all images were compared to the query image and they were ranked according to the resemblance values. Let N be the number of images, n the number of relevant images for a query, r_i the rank of the i -th relevant image, and r_l the maximum rank of a relevant image for a query. The retrieval performance was rated based on the following values as defined in [12]:

Normalized Recall R_n Value in the range from 0 (worst case) to 1 (perfect retrieval).

$$R_n = 1 - \frac{\sum_{i=1}^n r_i - \sum_{i=1}^n i}{n(N - n)}$$

The recall gives a higher weight to success in retrieving the first few items.

The average value for the 24 queries achieved by the combined approach was 0.96 (0.90 early artisan, 0.94 late artisan). The average value achieved by the underlying measure of similarity alone was 0.93, so the framework yields an improvement of 0.03.

¹This noise reduction is important but it is not in the main focus of our work. Therefore, a very simple implementation was used, that was not able to process the entire collection of images. In 116 cases out of 10 745, the texture in the image had to be removed by hand and the segmentation was redone.

Normalized Precision P_n Value in the range from 0 (worst case) to 1 (perfect retrieval).

$$P_n = 1 - \frac{\sum_{i=1}^n \log(r_i) - \sum_{i=1}^n \log(i)}{\log\left(\frac{N!}{(N-n)! \cdot n!}\right)}$$

The precision gives equal weight to all retrievals.

The average value for the 24 queries achieved by the combined approach was 0.79 (0.63 early artisan, 0.70 late artisan). The average value achieved by the underlying measure of similarity alone was 0.71, so the framework yields an improvement of 0.08.

Normalized Last-Place-Ranking L_n Value in the range from 0 (worst case) to 1 (perfect retrieval).

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

The last-place-ranking indicates the number of retrieved items a user has to search in order to have reasonable expectation of finding all relevant items.

The average value for the 24 queries achieved by the combined approach was 0.79 (0.56 early artisan, 0.72 late artisan). The average value achieved by the underlying measure of similarity alone was 0.68, so the framework yields an improvement of 0.11.

Number of Retrieved Images $n_{0.01}$ The number of relevant images ranked within the top 1 percent of the entire collection.

The sum for the 24 queries achieved by the combined approach was 229 (168 early artisan). The sum achieved by the underlying measure of similarity alone was 191, so the framework yields an improvement of 20 %.

For the detailed values of all 24 queries see the appendix table 1.

4. Conclusion

A new framework for content based image retrieval (esp. for trademark images) is presented which does not so much bank on sophisticated computation, but on taking account of some observations concerning perception: Familiar figures in the images are mostly perceived separately and their relevance may differ considerably. According to these observations the computation of image similarity is proceeded as follows: Images are divided up into sets of simple figures and the figures are weighted according to their relevance. The comparison of images is based on comparing the figures as well as their relations separately and on summing up the weighted similarities for the best matching of figures. The results of the experiments encourage further efforts in this direction, e.g., for improving the partitioning of the images, extending the set of image primitives, and refining the underlying measures of similarity for figures and relations.

Acknowledgements

This work was supported by the European Union under contract No. IST-511572-2, Project Perceptually-Relevant Retrieval of Figurative Images (PROFI).

References

[1] H. Helson. The fundamental propositions of gestalt psychology. *Psychological Review*, 40(1):13–32, 1933.

[2] John P. Eakins, K. Jonathan Riley, and Jonathan D. Edwards. Shape feature matching for trademark image retrieval. In *CIVR*, pages 28–38, 2003.

[3] Mary C. Dyson, Hilary Box, and Michael Twyman. The perception of symbols on screen and methods of retrieval from a database. British Library Research and Development Department Report 6163. British Library, London, 1994.

[4] John P. Eakins, Kevin Shields, and Jago Boardman. ARTISAN – a shape retrieval system based on boundary family indexing. In *Storage and Retrieval for Still Image and Video Databases IV. Proceedings SPIE 2670*, pages 17–28, 1996.

[5] Mercedes M. Fisher and Karen Smith-Gratto. Gestalt theory: A foundation for instructional screen design. *Journal of Educational Technology Systems*, 27(4), 1998-1999.

[6] Dempsey Chang, Laurence Dooley, and Juhani E. Tuovinen. Gestalt theory in visual screen design: a new look at an old subject. In *CRPIT '02: Proceedings of the seventh world conference on computers in education: Australian topics*, pages 5–12, Darlinghurst, Australia, 2002. Australian Computer Society, Inc.

[7] Victoria J. Hodge, Garry Hollier, John P. Eakins, and Jim Austin. Eliciting perceptual ground truth for image segmentation. In *CIVR*, pages 320–329, 2006.

[8] International classification of the figurative elements of marks (vienna classification) fifth edition. WORLD INTELLECTUAL PROPERTY ORGANIZATION, 2002. ISBN 92-805-1054-7.

[9] Hui Jiang, Chong-Wah Ngo, and Hung-Khoon Tan. Gestalt-based feature similarity measure in trademark database. *Pattern Recognition*, 39(5):988–1001, 2006.

[10] Eugene L. Lawler. The quadratic assignment problem. *anagement Science*, 9:586–599, 1963.

[11] Helmut Alt, Ludmila Scharf, and Sven Scholz. Probabilistic matching and resemblance evaluation of

shapes in trademark images. In *CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval*, pages 533–540, New York, NY, USA, 2007. ACM Press.

[12] John P. Eakins, Jago M. Boardman, and Margaret E. Graham. Similarity retrieval of trademark images. *IEEE MultiMedia*, 5(2):53–63, 1998.

[13] D. Douglas and T. Peucker. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. In *The Canadian Cartographer*, volume 10, pages 112–122, 1973.

[14] John Hershberger and Jack Snoeyink. Speeding up the douglas-peucker line-simplification algorithm. In *Proceedings of the 5th International Symposium on Spatial Data Handling*, volume 1, pages 134–143, Charleston, South Carolina, 1992.

Appendix A Examples of Trademark Images

Some examples of query images together with relevant images that can not be handled properly with a simple registration based approach.



Figure 2. Query image (left) and images to retrieve having different proportions.



Figure 3. Query image (left) and images to retrieve having different arrangements.

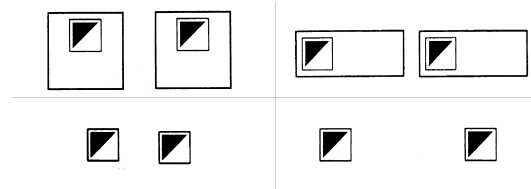


Figure 4. Query image (top left) and images to retrieve having different gaps and different frames.

Appendix B Experimental Results

























query	relevant images	R_n	P_n	L_n	$n_{0.01}$	
1.		26	0.99	0.87	0.93	19
2.		16	0.99	0.87	0.89	13
3.		12	0.96	0.89	0.60	10
4.		10	0.92	0.81	0.34	7
5.		10	0.99	0.72	0.97	4
6.		18	0.94	0.80	0.36	12
7.		11	0.97	0.71	0.89	6
8.		20	0.98	0.86	0.73	16
9.		25	1.00	1.00	1.00	25
10.		11	0.92	0.54	0.76	5
11.		10	1.00	0.91	0.98	8
12.		4	1.00	0.99	1.00	4
13.		16	0.97	0.62	0.89	6
14.		6	0.94	0.70	0.74	4
15.		13	0.99	0.85	0.93	10
16.		13	1.00	0.97	0.99	13
17.		17	0.94	0.66	0.72	9
18.		12	0.97	0.60	0.87	6
19.		21	0.67	0.28	0.11	1
20.		8	0.97	0.79	0.85	6
21.		8	1.00	0.91	0.98	7
22.		10	0.99	0.74	0.91	8
23.		23	0.99	0.87	0.93	20
24.		13	0.97	0.86	0.67	10
mean / sum		333	0.96	0.79	0.79	229
standard deviation			0.07	0.16	0.23	

Table 1. Results achieved: 24 query images plus values for normalized recall, normalized precision, normalized last place ranking, and number of relevant images ranked within the top 1 percent of the entire collection