

PROFI



Project number:	FP6-511572
Project acronym:	PROFI
Title:	Perceptually-relevant Retrieval Of Figurative Images

Deliverable No: D4.1:	Algorithms to match sets of regions
-----------------------	-------------------------------------

Short description:

Randomized algorithms are developed for matching two shapes A and B under translations, rigid motions and similarities. Shapes are modeled by regions, more precisely, sets of simple polygons. The measure of dissimilarity is the area of the symmetric difference of the two sets of regions.

The major idea is to take random samples of points from both shapes and give a “vote” for that transformation matching one sample with the other. If that experiment is repeated frequently we obtain by the votes a certain distribution of points in the space of transformations. Clusters of this point set indicate transformations with a high area of overlap or, equivalently, a low area of the symmetric difference of the shapes.

We also developed a matching algorithm based on geometric primitives. The main idea here is to identify primitive geometric objects, e.g., triangles, rectangles, ellipses, contained within the shapes and their relative positions. The matching is then evaluated based on the similarity of the primitives and the similarity of the relations between the primitives.

Due month:	27
Delivery month:	27
Lead partner:	FUB
Partners contributed:	FUB
Classification:	Public



Project funded by the European Community under the “Information Society Technologies” Programme

1 Introduction

In this work package we develop algorithms for *matching* two planar shapes. We assume that shapes are modeled by sets of plane simple polygons. Other than in previous work on objectives 3 and 4 the interiors of the polygons are considered as part of the shape, as well, not only their boundaries. As possible classes of transformations we will consider translations, rigid motions (i.e., translations and rotations), and similarities (i.e., translations, rotations and scalings).

The general situation is that we are given two objects A and B and a set T of allowable transformations and we want to transform B optimally so that the transformed image of B is as close to A as possible. Usually the quality of match is measured by some distance function or similarity measure $\delta(A, B)$ which assigns a distance value to any pair of objects A and B . An established distance measure for planar regions that is insensitive to noise is the area of the symmetric difference of these regions, which is defined as the area of the union of A and B minus the area of the intersection of A and B . Note, that for matching under translations and rigid motions minimizing the area of the symmetric difference is equivalent to maximizing the area of overlap.

Most previous work on the problem has been restricted to matching under translations. Mount et al. [22] analyze the combinatorial complexity of the function mapping a translation vector to the area of overlap of a translated simple n -vertex polygon with another simple m -vertex polygon. They show that this function is continuous, piecewise polynomial of degree at most two, has $O(n^2m^2)$ pieces and can be computed within the same time bound. De Berg et al. [7] consider the case of convex polygons and give an $O(n \log n)$ time algorithm, where n is the total number of vertices of two polygons. Alt et al. [3] give a linear time constant factor approximation algorithm for minimizing the area of the symmetric difference of two convex polygons under translations and homotheties (i.e., translations and scalings).

Surprisingly little is known about the problem if arbitrary rigid motions are allowed. Cheong et al. [6] gave a general probabilistic framework for maximizing the overlap of two shapes. This framework computes an approximation with prespecified absolute error in near quadratic time for translations and near cubic time for rigid motions. Finally, Ahn et al. [2] describe a $(1 - \epsilon)$ -approximation algorithm for maximizing the area of overlap of two convex polygons under translations and rigid motions.

We developed a probabilistic algorithm that maximizes the area of overlap of two given shapes under translations and rigid motions, which is equivalent to minimizing the area of the symmetric difference for these two classes of transformations. The method we introduce is related to the generalized Hough transform, and similar techniques called pose clustering or evidence gathering, which are widely used in pattern recognition. But instead of discrete sets of features in image and model, as in a typical setting in a pattern recognition task, we are dealing with compact plane regions. The idea is to take random samples of points of suitable size from each region set and record a vote for the transformation matching one sample to the other. The sample size depends on the allowable class of transformations and is such, that there exists a unique transformation that maps one sample to the other. Repeating this experiment many times we obtain a certain distribution of votes in transformation space. We show that transformations with the highest number of votes are exactly those, maximizing the area of overlap of the given sets of regions for the cases of translations and rigid motions. For similarity transformations the description of the resulting distribution is a bit more complex, we will give a detailed analysis.

On the other hand, evaluating algorithms for matching sets of curves, in collaboration with Actor Knowledge Technology we found out that the intuitive and widely used concept of matching as finding a transformation that maps one shape to the other, does not cover all aspects of perceived similarity (See Figure 1 for an example.).



Figure 1: Two images perceived as similar, for which no affine transformation exists that maps B onto A .

Therefore, we developed an algorithm for matching based on image primitives. It disregards the concrete position of the objects contained in the images, but concentrates on their shape and on their relative positions. An image is decomposed into a set of salient primitive shapes, such as ellipses, rectangles, triangles as well as convex polygons. Parts of the image that do not belong to one of these simple classes are grouped together and classified as complex shapes. The identified objects get weights according to their size and “prägnanz”, as understood in gestalt theory. Moreover, we observed that some trademark images contain frames and that these frames are hardly relevant for perceived similarity. When an object is recognized as being a frame, its weight is decreased. Two images are compared based on the similarity between the objects and on their relations. The experiments we made with a preliminary implementation show convincing results encouraging further research in this direction.

2 Results

2.1 Probabilistic matching

2.1.1 The similarity measure

We want to compare shapes that are regions in the plane. An established similarity measure for regions is the area of overlap. Two shapes are understood as similar if they overlap much. Psychological studies show that the area of overlap is an essential factor for perceived similarity [23].

The first row of Figure 2 shows two discs, one of them with noise added. Figures extracted from digital images often contain noise that should not affect the similarity of the figures. Since noise is expected to have a very small area, this is not the case if similarity is measured by the area of overlap. So one advantage of the area of overlap as similarity measure is its insensitivity to noise.

Nevertheless, there are examples where the area of overlap does not quite capture perceived similarity. An instance is shown in the second row of Figure 2, where the two shapes are perceived as quite different but the area of overlap is quite large if the blot is moved on top of the star.

The last row of Figure 2 indicates that comparing regions is a different matter than comparing curves. All common similarity measures for curves would assign a very small distance to the disc and the ring but the distance measured by area of overlap is very large.

If no scaling is allowed maximizing the area of overlap is the same as minimizing the symmetric difference which is a metric on the set of all shapes.

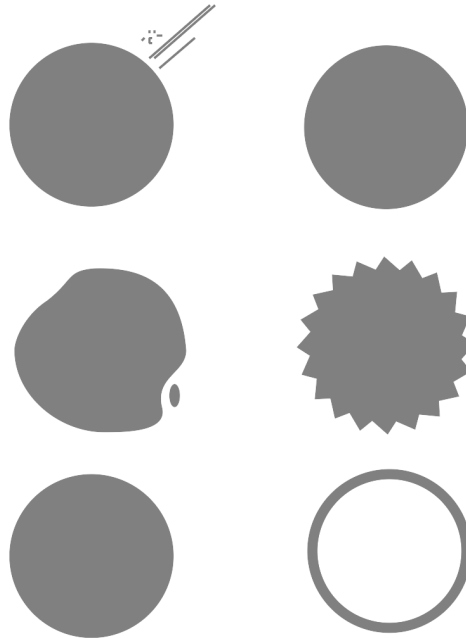


Figure 2: Some examples for the area of overlap as similarity measure.

2.1.2 The algorithm

Suppose that two regions A and B in \mathbb{R}^2 and a set of transformations F are given. We present a randomized algorithm finding a transformation $f \in F$ that approximately maximizes the area of intersection of the transformed shape $f(A)$ and B . We will investigate the cases where the set of transformations F consists of translations, rigid motions, and similarities. A rigid motion consists of a rotation angle and a translation vector, a similarity is composed of rotation, translation and scaling.

We define shapes quite generally as Lebesgue measurable sets in \mathbb{R}^2 . Suppose, there are given a fixed, small $\delta > 0$, two shapes A and B , and a set of transformations F which is identified with a subspace of some \mathbb{R}^k . We examine the following algorithmic scheme:

1. For $i = 1, \dots, n$ choose uniformly distributed a random point sample of suitable size in A and a random point sample in B .
2. Give one vote to the unique transformation in F that maps the sample from A onto the one from B .
3. Determine and return the transformation whose δ -neighborhood in F obtained the most votes.

Of how many random points one sample consists depends on the selected set of transformations. Size and form of a sample are chosen in a way that there exists exactly one transformation that maps one sample onto another sample.

The votes approximate the probability distribution in F resulting from the experiment. The output of the algorithm is a transformation whose δ -neighborhood in F has approximately highest probability. If δ is small enough this transformation is one at which the density function of the probability distribution is maximal.

In the following it is shown that for translations and rigid motions the density function of the probability distribution in $f \in F$ equals the area of overlap of $f(A)$ and B . Thus, the algorithm gives a transformation that approximately maximizes the area of overlap.

For similarities we do not have an equally simple characterization. Nevertheless, we will derive a formula of the density function and give an intuitive explanation. Furthermore, it is shown that the induced similarity measure is not symmetric in case of similarities.

All proofs are listed in Appendix A.

2.1.3 Translations

In case of translations a random point sample consists of one point in each shape because if we choose a point $a \in A$ and a point $b \in B$, there is exactly one translation that maps a onto b , namely $t = b - a$. The transformation space T equals \mathbb{R}^2 where a point $(t_1, t_2) \in T$ means the translation $(x, y) \mapsto (x + t_1, y + t_2)$. Note that $(-A) \oplus B$ contains exactly those translations that map at least one point of A onto a point of B , where \oplus denotes the Minkowski sum.

For a region $G \subset \mathbb{R}^2$ we denote its area by $|G|$. For a translation $t \in T$ we denote the δ -neighborhood of t with respect to the maximum norm by $B_\delta(t)$.

We are interested in the area of the intersection of A translated by t with the shape B denoted by

$$\mu(t) = |(A + t) \cap B|.$$

Let S be a set of n translations resulting from the experiment. The number of votes for the δ -neighborhood of t is given by $|S \cap B_\delta(t)|$. Our estimate of the probability that a translation resulting from one random experiment lies in the δ -neighbourhood of t is the ratio of the number of votes for the δ -neighborhood of t to the number of all votes, namely $\frac{|S \cap B_\delta(t)|}{|S|}$. We need to normalize this value with the sizes of the shapes and the size of the neighborhood of t to get an estimate $e_S(t)$ of $\mu(t)$:

$$e_S(t) = \frac{|S \cap B_\delta(t)| \cdot |A| \cdot |B|}{|S| \cdot |B_\delta(t)|}.$$

It can be shown that the error of the estimate becomes small if n is large enough. More precisely, it can be shown that for each $\varepsilon > 0$, $p < 1$ and each $\delta > 0$ there is a $n \in \mathbb{N}$ such that the error $|\mu(t) - e_S(t)|$ is less than ε with probability at least p . We achieve this by proving the following lemma:

Lemma 2.1. *For all shapes A and B , $\delta > 0$ and $t \in T$ with probability at least $1 - 2e^{-\frac{\varepsilon^2 \cdot n}{2}}$ it holds*

$$|\mu(t) - e_S(t)| \leq \frac{|A| \cdot |B| \cdot \varepsilon}{4\delta^2} + \sqrt{2}\Delta\delta$$

where Δ denotes the maximum of the lengths of the boundaries of A and B .

From this lemma it follows that in case of translations the algorithm returns a translation that approximately maximizes the area of overlap of a translated copy of A with the shape B for all shapes A and B . More formally: Let t^* be the map that maximizes the estimate $e_S(t)$

and let t_{opt} be an optimal solution. If the approximation error $|\mu(t) - e_S(t)|$ is less than ε then the difference between approximation and optimum is at most 2ε :

$$\mu(t^*) \geq e_S(t^*) - \varepsilon \geq e_S(t_{\text{opt}}) - \varepsilon \geq \mu(t_{\text{opt}}) - 2\varepsilon.$$

A bound for the required number of random samples in case of translations. For translations we can deduce a concrete bound for the required number of random samples from Lemma 2.1.

Corollary 2.2. *Let $\tilde{\varepsilon} > 0$ and $p \in [0, 1]$. Let t^* be the map that maximizes the estimate $e_S(t)$ and let t_{opt} be an optimal solution. If*

$$n \geq \frac{|A|^2|B|^2(-\log(\frac{1-p}{2}))}{8\delta^4(|B|\tilde{\varepsilon} - \sqrt{2}\Delta\delta)^2}$$

the approximation error $|\mu(t_{\text{opt}}) - e_S(t^)|$ is less than $2\tilde{\varepsilon}$ with probability at least $1 - p$. The bound is minimal for $\delta = \frac{\sqrt{2}|B|\tilde{\varepsilon}}{3\Delta}$.*

The theoretical bound for n seems to be quite pessimistic. Our test implementation shows that the real convergence is much faster than the bound indicates. In fact, the results that our preliminary implementation shows for translations are very promising.

2.1.4 Rigid Motions

The space of rigid motions is given by $R = [0, 2\pi) \times T \subseteq \mathbb{R}^3$ the first coordinate defining the anti-clockwise rotation angle. A point $(\alpha, (t_1, t_2)) \in R$ denotes the map defined by $(x, y) \mapsto M_\alpha \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$.

Observe that there exists exactly one translation that maps a point $p \in A$ onto a point $q \in B$ while there exist infinitely many rigid motions that map p onto q . To be more precise, for each $\alpha \in [0, 2\pi)$ there is exactly one rigid motion whose anti-clockwise rotation angle is given by α and that maps p onto q . This rigid motion is given by $r = (\alpha, q - M_\alpha p)$ where M_α is the rotation matrix $\begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix}$.

If we choose two random points p and p' in A and two random points q and q' in B there is exactly one similarity that maps p onto q and p' onto q' . (To determine one affine transformation we need 3 points from each shape.) If we want to determine exactly one rigid motion there is no size of the random sample that works. Therefore we pick uniformly distributed a random angle α in $[0, 2\pi)$, one point a in A and one point b in B . We give one vote to the unique rigid motion with anti-clockwise rotation angle α that maps a onto b .

If we do this experiment very often we get an approximation of a probability distribution on R . The algorithm returns a rigid motion whose δ -neighborhood in R got the most votes and thus has approximately highest probability. If δ is small enough the rigid motion whose δ -neighborhood in R has highest probability is the rigid motion at which the density function of the probability distribution is maximal.

Lemma 2.3. *The density function on R is given by $f(r) = \frac{|r(A) \cap B|}{2\pi \cdot |A| \cdot |B|}$.*

Because of this lemma as in case of translations it holds that also in case of rigid motions the output of the algorithm is a rigid motion that approximately maximizes the area of overlap. For matching under rigid motions we can derive similar approximation error bounds as in the case of translations.

2.1.5 Similarities

For matching under similarities, our experiment consists of choosing two random points a and a' in A and two random points b and b' in B . Then there is exactly one similarity that maps a onto b and a' onto b' . We parameterize the space of similarities $S \subseteq \mathbb{R}^4$ in the following way: $(s_1, s_2, s_3, s_4) \in S$ denotes the similarity $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} s_1 & s_2 \\ -s_2 & s_1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} s_3 \\ s_4 \end{pmatrix}$, so s_1 is the product of scaling factor and cosine of the anti-clockwise rotation angle, s_2 is the product of scaling factor and sine of the rotation angle and $\begin{pmatrix} s_3 \\ s_4 \end{pmatrix}$ is the translation vector. Having chosen a, a', b, b' the scaling factor λ is given by $\frac{\|b'-b\|}{\|a'-a\|}$, the cosine of the rotation angle α is $\frac{\langle a'-a, b'-b \rangle}{\|a'-a\| \cdot \|b'-b\|}$, the sine is $\frac{\det \begin{pmatrix} a'-a \\ b'-b \end{pmatrix}}{\|a'-a\| \cdot \|b'-b\|}$ (where $(a'-a)$ and $(b'-b)$ are written as row vectors) and the translation vector is given by $b - \lambda M_\alpha a$. The following formula for the density function on S can be shown:

Lemma 2.4. *The density function f of the probability distribution resulting from the experiment on S is given by*

$$f(s) = \frac{1}{|A|^2 \cdot |B|^2} \int_{(A \cap s^{-1}(B))^2} \|a' - a\|^2 d(a', a).$$

Intuitively, the density function is large and thus the probability of a small neighborhood is large as well if essential parts of A have a large distance from each other. The density function is large if A has a elongated shape in contrast to a circular shape.

Since for all $a, a' \in C$ the squared distance $\|a' - a\|^2$ is less than the squared diameter $\text{diam}(C)^2$ the following holds

Proposition 2.5. *For shapes $A, B \subseteq \mathbb{R}^2$ and a similarity $s \in S$ it holds*

$$f(s) \leq \frac{\text{diam}(A \cap s^{-1}(B))^2 \cdot |A \cap s^{-1}(B)|^2}{|A|^2 \cdot |B|^2}.$$

For similarities the algorithm is not symmetric as the example in Figure 3 shows. If A is a dumbbell and B is a disc the algorithm returns a similarity that shrinks A into B because the density is maximal for this similarity. If A is a disc and B is a dumbbell the two similarities with highest density are the translations that put A on top of one of the discs of the dumbbell.

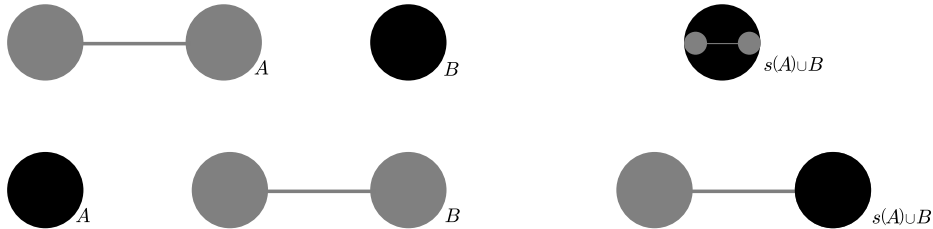


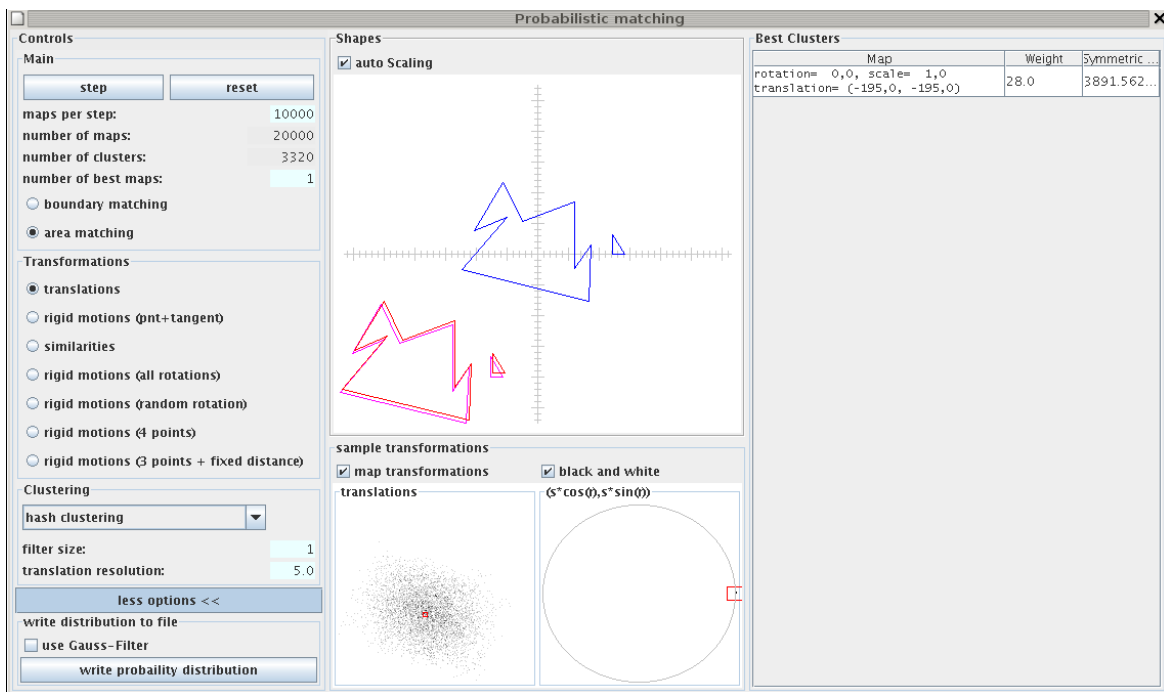
Figure 3: Matching A with B under similarities gives a different result than matching B with A .

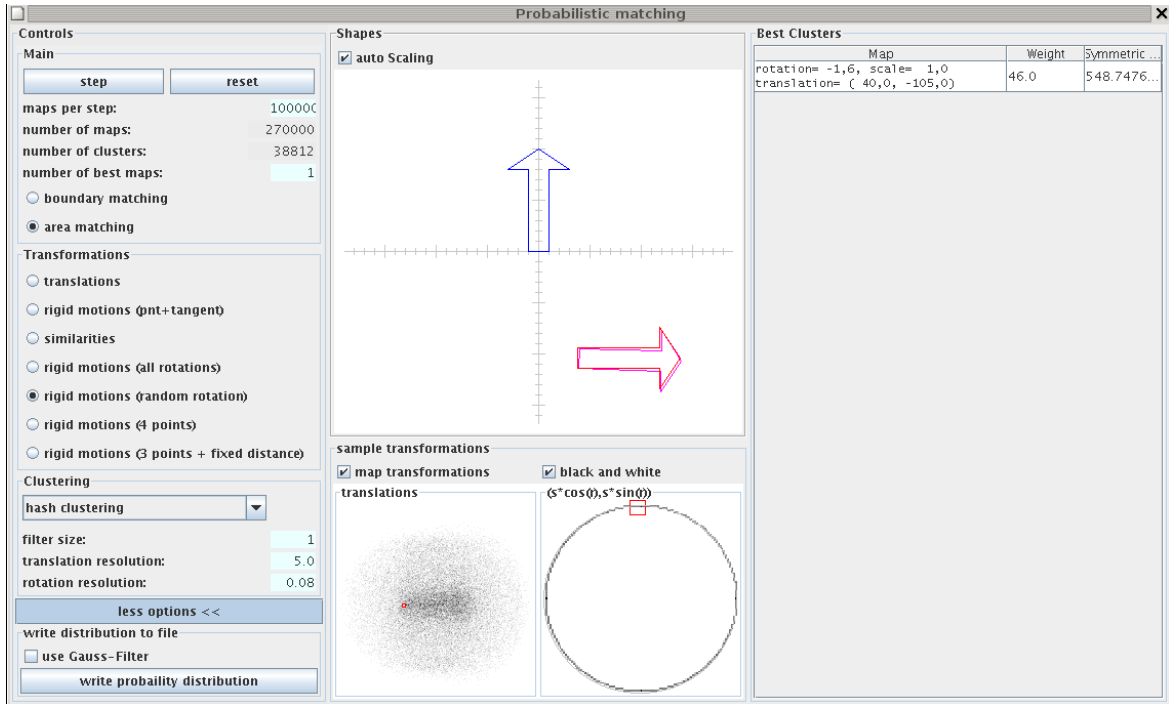
2.1.6 Experiments

We did a preliminary implementation of our algorithm for translations, rigid motions and similarities. In this test implementation the clustering of the transformations is done by bucketing: the transformation space is divided into cubes of suitable dimension and the cube with the highest number of votes is returned.

Even with the simpler clustering method the experiments confirm our theoretical results for the cases of translations and rigid motions. Our experiments also show that the theoretical bound for the number of necessary samples in case of translations is quite pessimistic, we observed a much faster convergence of the approximation error in practice.

Two screenshots below show examples of matching under translations and rigid motions. The center of the window shows the shapes A and B and the image of A under the transformation computed by the algorithm. The pictures below show the distribution of the votes. On the left hand side a point is drawn for each translation that obtained a vote and on the right hand side the votes for angles are drawn on a circle, in the case of rigid motions.





2.2 Matching based on Image Primitives

2.2.1 Motivation

For the comparison of figurative images represented by sets of (polygonal) curves we developed algorithms (objective 3) that were based on finding similarity transformations that match the two images. This works well for images whose parts lie close together and whose parts do not differ much in perceived significance.

Although one of the laws invented by gestalt theory states that configurations cannot be analyzed into parts and relations [14], for multi-component images the comparison based on the individual image components is more effective than a comparison based on the whole image [11].

With regard to the ground truth provided by professional trademark examiners (see section 2.2.3), 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 appendix B figs. 5 and 6). 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 [9] (cited in [12]).
 - "there is an unconscious effort to simplify what is perceived into what the viewer can understand". [13] (cited in [5])
- If the image consists of spatially independent parts, the size of the gaps in between plays a minor role (see appendix B fig. 7).

- If an essential part of the image is framed, the shape of the frame and even the existence of the frame play a minor role (see appendix B fig. 8). In [16] 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 whether 16 circles form a 4×4 grid or whether 25 circles form a 5×5 grid.

Our approach for 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 the region as given by the bounding polygonal curve. Many patent offices use such a

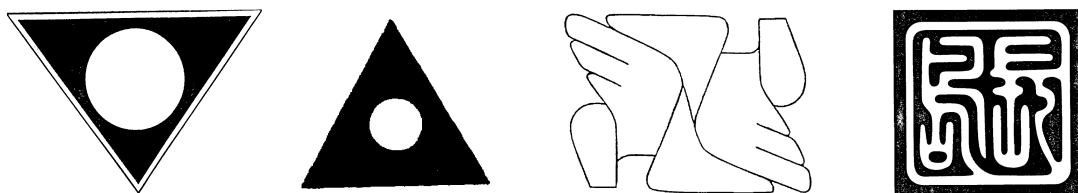


Figure 4: actual trademark images — some easy to describe by geometric primitives and some not.

characterization based on the so called Vienna classification [1]. The codes for the examples given in fig. 4 would possibly be '*26.3.10 Triangles 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 arbitrarily composed shapes.

In [17] 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 discard 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.

2.2.2 Technique

It is assumed that figurative images are given as a set S of polygonal regions $s_1 \dots s_m$. Based on these polygonal regions a set P of figures $p_1 \dots p_n$ is extracted and “relations” $R = r_{1,2} \dots r_{n,n-1}$ among them 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. 2.2.3 however, a simple proof-of-concept implementation was used.

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 parts of the image that cannot be represented by these primitives are categorized as

- convex polygons
- arbitrary sets of polygons and polylines

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

Relations For a pair $(p_i, p_j) \in P \times P, i \neq j$ of figures the relation $r_{i,j}$ consists of numerical values reflecting

- the size of p_j relative to the size of p_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 p_j to p_i (The distance of the centers relative to the size of p_i .)
- the qualitative relation, i.e., whether p_i and p_j are very similar, just mirrored, or just rotated etc.

Comparison of two images For the comparison of two images I^1 and I^2 the *relevance* w_P of the figures and the *relevance* w_R of the relations is preset such that $w_P + w_R = 1$ — 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 $w(p_i)$ and $w(r_{i,j})$ such that for each image all weights sum up to 1, namely: $\sum_{p \in P} w(p) = w_P$ and $\sum_{r \in R} w(r) = w_R$.

For every pair $(p_i^1, p_k^2) \in P^1 \times P^2$ of figures and every pair $(r_{i,j}^1, r_{k,l}^2) \in R^1 \times R^2$ of relations a value of similarity $s \in [0, 1]$ is computed, using simple measures of similarity as described later.

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 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:

$$s(I^1, I^2) = \max_{M \in \mathcal{M}} \left\{ \sum_{(p^1, p^2) \in M} s(p^1, p^2) \cdot \frac{w(p^1) + w(p^2)}{2} + \sum_{\substack{(p_i^1, p_k^2) \in M \\ (p_j^1, p_l^2) \in M}} s(r_{i,j}^1, r_{k,l}^2) \cdot \frac{w(r_{i,j}^1) + w(r_{k,l}^2)}{2} \right\}$$

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. [21]) and therefore is NP-complete¹. 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 2.2.3). Thus, the value of similarity $s(I^1, I^2)$ may even be computed using methods having a high asymptotic runtime which would not be acceptable in a different context:

1. using a weighted association graph $G_{1,2} = (V, E)$.

The vertex set $V = P^1 \times P^2 \cup R^1 \times R^2$ consists of all possible pairs of figures of I^1 and I^2 and of all pairs of relations of I^1 and I^2 . The weight $w(v^1, v^2)$ of a vertex $v = (v^1, v^2)$ is the product of the similarity and the arithmetic mean of the weights of the two figures or relations respectively: $w(v^1, v^2) = s(v^1, v^2) \cdot 0.5 \cdot [w(v^1) + w(v^2)]$

Between any two vertices v_1, v_2 an edge is inserted into the graph $G_{1,2}$, if and only if choosing the two underlying pairs would lead to a contradiction (e.g. if a figure p_a^1 of I^1 is matched to a figure p_b^2 of I^2 it cannot be matched to a figure p_c^2 at the same time).

The value $s(I^1, I^2)$ equals the weight of the maximum (weight) independent set of vertices in $G_{1,2}$ which can be approximated using the greedy algorithm described in [18].

2. using a branch and bound algorithm for enumeration of the promising matches.

Weights Every figure p_i gets an absolute weight $w_a(p_i)$ 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 weights of the figures p_i and p_j . Before the comparison the weights are normalized such that $\sum w(p) = w_P$ and $\sum w(r) = w_R$. If we compare two images I^1 and I^2 with different numbers n^1, n^2 of figures, then, of course, we can match at most as many figures as the smaller of two images has. Therefore, only the relations for $n_{min} = \min(n^1, n^2)$ figures may be selected in the matching. In this case, the weights of the relations of the image consisting of more figures are adjusted, such that the maximum over all n_{min} -subsets of the figures of the sum of the relation weights between the figures in a subset equals w_R .

Frames For every figure the likeliness of being a frame is rated based on the following propositions:

- frames are convex and symmetric

¹The NP-completeness can easily be seen by reduction of the problem whether a graph contains a Hamiltonian cycle.

- 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. 6) and some trademark images even contain miscellaneous variants of the actual logo (see the appendix fig. 5). 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 [4] 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 p^1 into the proximity of corresponding parts of the other figure p^2 and the similarity is rated based on proximity and parallelism of $t(p^1)$ and p^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 two 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. whether p_i and p_j are very similar, just mirrored, or just rotated etc.

Dealing with different representations Whenever such 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). To diminish that effect, the images are also compared using the simple similarity measure that is used for the figures. The maximum of the two values is taken as the similarity of the images.

2.2.3 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 [10]. 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. 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 eliminated². The remaining closed contours for which every

²This noise reduction is important but it is not in the main focus of our work. Therefore, a very simple

vertex corresponds to a pixel, were then simplified using the Douglas-Peucker algorithm [8] (cited in [15]).

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 decomposition 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 [10]:

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 for the 24 queries was 0.96 (0.90 early artisan, 0.94 late artisan).

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 for the 24 queries was 0.79 (0.63 early artisan, 0.70 late artisan).

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 for the 24 queries was 0.79 (0.56 early artisan, 0.72 late artisan).

Number of Retrieved Images $n_{0.01}$ The number of relevant images, as defined by ground truth, that are ranked by the system within the top 1 percent of the entire collection. The sum for the 24 queries was 229 (168 early artisan).

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

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.

3 Deviations from plan

There have been no deviations from plan.

4 Project publications

Sven Scholz. Similarity Evaluation based on Image-Primitives. Submitted to International Conference on Image and Video Retrieval, CIVR 2007.

Helmut Alt, Ludmila Scharf. Probabilistic matching of polygonal curves. To be published in Dagstuhl Seminar Proceedings “Computational Geometry”, 2007, in preparation.

References

- [1] International classification of the figurative elements of marks (vienna classification) fifth edition. WORLD INTELLECTUAL PROPERTY ORGANIZATION, 2002. ISBN 92-805-1054-7.
- [2] Hee-Kap Ahn, Otfried Cheong, Chong-Dae Park, Chan-Su Shin, and Antoine Vigneron. Maximizing the overlap of two planar convex sets under rigid motions. *Computational Geometry*, 37(1):3–15, 2007.
- [3] Helmut Alt, Ulrich Fuchs, Günter Rote, and Gerald Weber. Matching convex shapes with respect to the symmetric difference. *Algorithmica*, 21:89–103, 1998.
- [4] Helmut Alt, Ludmila Scharf, and Sven Scholz. Profi deliverable 4.3: Algorithms to match sets of curves, 2006.
- [5] 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.
- [6] Otfried Cheong, Alon Efrat, and Sarel Har-Peled. On finding a guard that sees most and a shop that sells most. In *Proc. 15th ACM-SIAM Sympos. Discrete Algorithms (SODA)*, pages 1091–1100, 2004.
- [7] Mark de Berg, Olivier Devillers, Marc J. van Kreveld, Otfried Schwarzkopf, and Monique Teillaud. Computing the maximum overlap of two convex polygons under translations. In *ISAAC '96: Proceedings of the 7th International Symposium on Algorithms and Computation*, pages 126–135, London, UK, 1996. Springer-Verlag.
- [8] 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.
- [9] Mary C. Dyson, Hilary Box, and Michael Twyman. The perception of symbols on screen and methods of retrieval from database, 1994.

- [10] John P. Eakins, Jago M. Boardman, and Margaret E. Graham. Similarity retrieval of trademark images. *IEEE MultiMedia*, 5(2):53–63, 1998.
- [11] John P. Eakins, K. Jonathan Riley, and Jonathan D. Edwards. Shape feature matching for trademark image retrieval. In *CIVR*, pages 28–38, 2003.
- [12] 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.
- [13] Mercedes M. Fisher and Karen Smith-Gratto. Gestalt theory: A foundation for instructional screen design. *Journal of Educational Technology Systems*, 27(4), 1998-1999.
- [14] H. Helson. The fundamental propositions of gestalt psychology. *Psychological Review*, 40(1):13–32, 1933.
- [15] John Hershberger and Jack Snoeyink. Speeding up the douglas-peucker line-simplification algorithm. Technical report, Vancouver, BC, Canada, Canada, 1992.
- [16] Victoria J. Hodge, Garry Hollier, John P. Eakins, and Jim Austin. Eliciting perceptual ground truth for image segmentation. In *CIVR*, pages 320–329, 2006.
- [17] Hui Jiang, Chong-Wah Ngo, and Hung-Khoon Tan. Gestalt-based feature similarity measure in trademark database. *Pattern Recognition*, 39(5):988–1001, 2006.
- [18] Akihisa Kako, Takao Ono, Tomio Hirata, and Magnús M. Halldórsson. Approximation algorithms for the weighted independent set problem. In *WG*, pages 341–350, 2005.
- [19] Ulrich Krengel. *Einführung in die Wahrscheinlichkeitstheorie und Stochastik*. vieweg, Braunschweig/Wiesbaden, fifth edition, 2000.
- [20] Viktor Kurotschka. Elementare wahrscheinlichkeitstheorie. Lecture notes, 2002. FU Berlin, Germany.
- [21] Eugene L. Lawler. The quadratic assignment problem. *anagement Science*, 9:586–599, 1963.
- [22] David M. Mount, Ruth Silverman, and Angela Y. Wu. On the area of overlap of translated polygons. *Comput. Vis. Image Underst.*, 64(1):53–61, 1996.
- [23] Eleanor Rosch, Carolyn B. Mervis, Wayne D. Gray, David M. Johnson, and Penny Boyes-Braem. Basic objects in natural categories. *Cognitive Psychology*, 8(3):382–439, 1976.

A Proofs

Proof. (of Lemma 2.1) Let $t, t' \in T$ be translations and let Δ denote the maximum of the lengths of the boundaries of the shapes A and B . Then the following holds

$$\|t - t'\|_{\max} < \delta \implies |t(A) - t'(A)| < \sqrt{2}\Delta\delta.$$

This implies

$$\left| |B_\delta(t)| \cdot \mu(t) - \int_{B_\delta(t)} |A + g \cap B| dg \right| \leq |B_\delta(t)| \cdot \sqrt{2}\Delta\delta.$$

Since the random sample points in A and B are uniformly distributed the density functions are $f_A(a) = \frac{1}{|A|} \cdot \chi_A(a)$ and $f_B(b) = \frac{1}{|B|} \cdot \chi_B(b)$ where χ_A denotes the indicator function of A . Because of the independence of the distributions the density function $f_{(-A) \oplus B}$ of the distribution of $b - a$ is the convolution of $f_{(-A)}$ and f_B , namely

$$f_{(-A) \oplus B}(y) = \int_{\mathbb{R}^2} f_{(-A)}(y - s) f_B(s) ds.$$

See for example [19] for a proof. Now the probability of a δ -neighborhood can be determined:

$$\begin{aligned} \Pr(B_\delta(t)) &= \int_{B_\delta(t)} \int_{\mathbb{R}^2} f_{(-A)}(y - s) f_B(s) ds dy \\ &= \int_{B_\delta(t)} \int_{y-s \in (-A), s \in B} \frac{1}{|A| \cdot |B|} ds dy \\ &= \frac{1}{|A| \cdot |B|} \int_{B_\delta(t)} \int_{s \in A+y, s \in B} 1 ds dy \\ &= \frac{1}{|A| \cdot |B|} \int_{B_\delta(t)} |A + y \cap B| dy \end{aligned}$$

For a given set of random translations $S = \{s_1, \dots, s_n\}$ consider a random variable $X(t)$ which is defined as follows

$$\begin{aligned} X_i(t) &= \begin{cases} 1 & \text{if } s_i \in B_\delta(t) \\ 0 & \text{otherwise} \end{cases} \text{ for each } i \in \{1, \dots, n\} \\ X(t) &= \sum_{i=1}^n X_i(t). \end{aligned}$$

Thus $X(t)$ counts the number of votes for the δ -neighborhood of t . Let $E(X(t))$ denote the expected value of $X(t)$. Then by linearity of the expected value the following equations hold

$$E(X(t)) = n \cdot \Pr(B_\delta(t))$$

For $\hat{\varepsilon} > 0$ the Chernoff bound as described in [6] gives

$$\Pr\left(\left|\frac{E(X(t))}{n} - \frac{X(t)}{n}\right| > \hat{\varepsilon}\right) < 2e^{-\frac{\hat{\varepsilon}^2 n}{2}}$$

and by this and the above the following holds with high probability

$$\left| \int_{B_\delta(t)} |A + g \cap B| dg - |B_\delta(t)| \cdot e_S(t) \right| = |A| \cdot |B| \cdot \left| \frac{E(X(t))}{n} - \frac{X(t)}{n} \right| \leq |A| \cdot |B| \cdot \hat{\varepsilon}.$$

Now we have

$$\begin{aligned}
|B_\delta(t)| \cdot |\mu(t) - e_S(t)| &\leq \left| |B_\delta(t)| \cdot \mu(t) - \int_{B_\delta(t)} |A + g \cap B| dg \right| + \\
&\left| \int_{B_\delta(t)} |A + g \cap B| dg - |B_\delta(t)| \cdot e_S(t) \right| \\
&\leq |A| \cdot |B| \cdot \hat{\varepsilon} + 4\sqrt{2}\Delta\delta^3 \text{ with high probability.}
\end{aligned}$$

□

Proof. (of Lemma 2.3)

We need to show that the density function f on R equals $f(r) = \frac{|r(A) \cap B|}{2\pi \cdot |A| \cdot |B|}$. We choose the random samples uniformly distributed from $[0, 2\pi) \times A \times B$ but we are interested in the density function on $R = [0, 2\pi) \times \mathbb{R}^2$ where the \mathbb{R}^2 is understood as translation space. To determine the density on R we define probability spaces $\Omega_1 = [0, 2\pi)$ and $\Omega_2 = \mathbb{R}^2$. The distribution on Ω_1 is uniform, so the density function f_1 is given by $f_1(\alpha) = \frac{1}{2\pi}$. The probability distribution on Ω_2 depends on the chosen $\alpha \in \Omega_1$ and is given by the experiment: if the chosen random point sample is (α, a, b) we get the rigid motion $(\alpha, b - M_\alpha a) \in \Omega_1 \times \Omega_2$. We understand the probability on Ω_2 as conditional probability. To prove the lemma we determine the conditional density functions $f_2(\alpha, \cdot)$. The density function f is then given by $f((\alpha, t)) = f_1(\alpha) \cdot f_2(\alpha, t)$ as is explained in [20].

To determine $f_2(\alpha, \cdot)$ we use the following theorem which is a special case of the general transformation rule for density functions (Theorem A.2) and is proven in [20].

Theorem A.1. *Given a probability space Ω , a random variable $X : \Omega \rightarrow \mathbb{R}^n$ with density function f and an affine map $h : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $h : x \mapsto Mx + t$ such that $\det M \neq 0$. Then the random variable $h \circ X$ has the density function g defined by $g(y) = \frac{f(M^{-1}(y-t))}{|\det M^{-1}|}$.*

$$\text{We choose } X = id_{A \times B} \text{ and } h : \mathbb{R}^4 \rightarrow \mathbb{R}^4, h(x) = Mx \text{ where } M = \begin{pmatrix} \cos \alpha & -\sin \alpha & 0 & 0 \\ \sin \alpha & \cos \alpha & 0 & 0 \\ \cos \alpha & \sin \alpha & 1 & 0 \\ -\sin \alpha & \cos \alpha & 0 & 1 \end{pmatrix}.$$

Obviously the determinant of M is 1 and $M^{-1} = \begin{pmatrix} \cos \alpha & \sin \alpha & 0 & 0 \\ -\sin \alpha & \cos \alpha & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$. $h \circ X$ is the

random variable that maps (a, b) onto $(M_\alpha a, b - M_\alpha a)$ and its density function g equals

$$\begin{aligned}
g(a, y) &= \frac{1}{|A| \cdot |B|} \cdot \chi_A(M_{-\alpha}(a)) \cdot \chi_B(a + y) \\
&= \frac{1}{|A| \cdot |B|} \cdot \chi_{M_\alpha A}(a) \cdot \chi_{B-y}(a).
\end{aligned}$$

We are interested in the probability distribution of $b - M_\alpha a$ on \mathbb{R}^2 whose density function

$f_2(\alpha, \cdot)$ is given by

$$\begin{aligned}
f_2(\alpha, t) &= \int_{\mathbb{R}^2} g(a, t) da \\
&= \int_{\mathbb{R}^2} \frac{1}{|A| \cdot |B|} \cdot \chi_{M_\alpha A}(a) \cdot \chi_{B-y}(a) da \\
&= \frac{1}{|A| \cdot |B|} \int_{\mathbb{R}^2} \chi_{M_\alpha A + y \cap B}(a) da \\
&= \frac{|M_\alpha A + y \cap B|}{|A| \cdot |B|}.
\end{aligned}$$

□

Proof. (of Lemma 2.4)

Let $\bar{A} = A^2 \setminus \{(a, a) : a \in A\}$ and $\bar{B} = B^2 \setminus \{(b, b) : b \in B\}$. We define a map φ that is differentiable with continuous derivative and bijective by

$$\begin{aligned}
\varphi : \bar{A} \times \bar{B} &\rightarrow \bar{A} \times S \subseteq \mathbb{R}^8 \\
\varphi : (a, a', b, b') &\mapsto (a, a', \underbrace{\langle a' - a, b' - b \rangle}_{=\lambda \cos \alpha}, \underbrace{\frac{\det \begin{pmatrix} a' - a \\ b' - b \end{pmatrix}}{\|a' - a\|^2}}_{=\lambda \sin \alpha}, b - \lambda M_\alpha a)
\end{aligned}$$

φ models the experiment but is defined in a way that it is bijective. The inverse is given by $\varphi^{-1}(a, a', s) = (a, a', s(a), s(a'))$ where $(a, a') \in \bar{A}, s \in S$. We use the following transformation rule for density functions written down in [19]

Theorem A.2. *Let $X : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a random variable with density function f and $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ a bijective, differentiable map with continuous derivative. Let $\Delta(x) = \det\left(\frac{\partial \varphi_j}{\partial x_i}\right)_{i,j=1,\dots,n}$. The density function g of $\varphi \circ X$ is given by*

$$g(y) = \begin{cases} \frac{f(\varphi^{-1}(y))}{|\Delta(\varphi^{-1}(y))|} & \text{if } y \in \varphi(G) \\ 0 & \text{otherwise} \end{cases}$$

where $G \subseteq \mathbb{R}^n$ such that $\Pr(X \in G) = 1$.

Let X be $id_{\mathbb{R}^8}$ with uniform distribution. The density function f is given by $f(a, a', b, b') = \frac{\chi_{A^2}(a, a') \cdot \chi_{B^2}(b, b')}{|A|^2 \cdot |B|^2}$. The determinant of the Jacobi matrix of φ is $\Delta(a, a', b, b') = \frac{1}{\|a' - a\|^2}$. The density function g of $\varphi \circ X$ is

$$\begin{aligned}
g(a, a', s) &= \frac{\|a' - a\|^2}{|A|^2 \cdot |B|^2} \cdot \chi_{A^2}(a, a') \cdot \chi_{B^2}(s(a), s(a')) \\
&= \frac{\|a' - a\|^2}{|A|^2 \cdot |B|^2} \cdot \chi_{A^2}(a, a') \cdot \chi_{(s^{-1}(B))^2}(a, a') \\
&= \frac{\|a' - a\|^2}{|A|^2 \cdot |B|^2} \cdot \chi_{(A \cap s^{-1}(B))^2}(a, a')
\end{aligned}$$

Integration gives the density k on S

$$k(s) = \frac{1}{|A|^2 \cdot |B|^2} \int_{(A \cap s^{-1}(B))^2} \|a' - a\|^2 d(a, a')$$

□

B Examples of trademark images

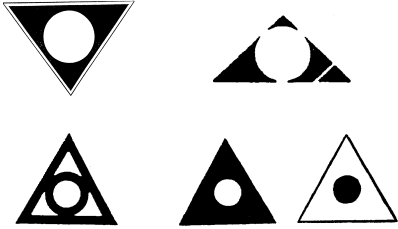


Figure 5: Query image (top left) and images to retrieve having different proportions.



Figure 6: Query image (left) and images to retrieve having different arrangements.

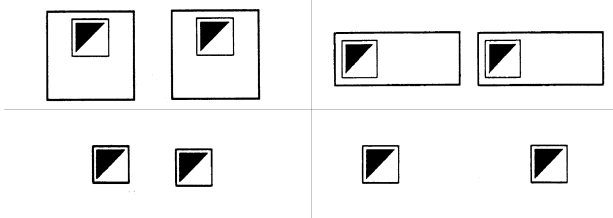


Figure 7: Query image (top left) and images to retrieve having different gaps.



Figure 8: Query image (left) and images to retrieve having different frames.

C Experimental results

























	query	size	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 recall, precision, last place ranking, and number of relevant images ranked within the top 1 percent of the entire collection