Shape Recognition Via an a Contrario Model for Size Functions

Andrea Cerri[1], Daniela Giorgi[1], Pablo Musé[2], Frédéric Sur[3], and Federico Tomassini[1]

[1] Dipartimento di Matematica
Università di Bologna
Piazza di Porta S. Donato, 5
I-40126 Bologna, Italy
[2] Centre de Mathématiques et de Leurs Applications
École Normale Supérieure de Cachan
61, avenue du président Wilson
94235 Cachan Cedex, France
[3] Loria & INPL
Loria, Campus scientifique BP 239
54506 Vandoeuvre-lès-Nancy Cedex, France

Abstract. Shape recognition methods are often based on feature comparison. When features are of different natures, combining the value of distances or (dis-)similarity measures is not easy since each feature has its own amount of variability. Statistical models are therefore needed. This article proposes a statistical method, namely an a contrario method, to merge features derived from several families of size functions. This merging is usually achieved through a touchy normalizing of the distances. The proposed model consists in building a probability measure. It leads to a global shape recognition method dedicated to perceptual similarities.

1 Global Shape Recognition

1.1 A Brief State of the Art

Shape recognition is a central problem in computer vision. Geometrical shapes can be defined as outlines of objects and are thus formally connected bounded domains of the plane. Many shape recognition methods are global in the sense that the extracted features are computed over the whole solid shape. Since they mix global and local information, they are sensitive to occlusions (a part of the shape is hidden) or insertion (a part is added to the shape). There are however a large number of applications in which the shapes that are to be identified are not occluded, for which global methods are undoubtly useful. The present method actually enables global shape recognition.

The global features are in general scalar numbers computed over the whole shape. Two classical global features are based on Fourier descriptors (after Zahn and Roskies [1]) or invariant moments [2,3] (following a founding work by Hu [4]). Affine invariant scalars for global shape representation can also be derived from

A. Campilho and M. Kamel (Eds.): ICIAR 2006, LNCS 4142, pp. 410–421, 2006.

wavelet coefficients (see for instance [5]). Scale-space representations can also be used to derive invariant representations. In this class of methods the *Curvature Scale Space* by Mokhtarian and Mackworth [6] is certainly the most popular approach. It consists in smoothing the shape boundary by curvature motion, while tracking the position of its inflexion points across the scales. This method yields similarity invariant representations and it has a certain amount of robustness to noise. Another invariant shape representation based on scale spaces can be found in Alvarez *et al.* [7] where shape invariants are built from the evolution of area and perimeter of the shapes undergoing the affine scale space.

All these methods deal with similar shape recognition where *similar* is understood with respect to geometrical invariance (similar shapes are sought up to an affine transformation for example). On the contrary, a well-known, moment related, global method which deals with perceptual similarity is *modal matching*, by Sclaroff and Pentland [8]. In this method, solid shapes are represented by eigenmodes associated to a physical elastic model. This method permits relatively realistic shape deformations where the thin parts of the shape can alter more than the bulk (for example two human silhouettes are retrieved as similar, whatever the pose could be).

1.2 Contribution: Mixing Size Functions Through a Statistical Model

The descriptors that are considered in the present shape recognition method are based on Size Functions (SFs), which are geometrical-topological descriptors, conceived for formalizing qualitative aspects of shapes. Although SFs carry both local and global information and cannot be classified as a purely global method, in this paper we shall use them as global features. We are indeed interested in the comparison of shapes when occlusions are not involved. When applied to shape recognition, SFs theory leads to (quasi) invariant descriptions. These descriptions have proven to be particularly useful for perceptual matching, mainly when no standard geometric templates are available [9]. See for example [10] where SFs are applied to trademark retrieval. As a single SF cannot provide a *complete* shape description (two dissimilar shapes could share similar SFs) several SFs have to be extracted from each shape. Each one of them enables to capture geometrical-topological information of different natures. The question is: how to merge these informations? The aim of this article is precisely to propose an appropriate statistical framework. The shape recognition algorithm derived from this model is dedicated to retrieve global shapes that have not undergone occlusions.

1.3 Plan of the Article

We first give guidelines on Size Theory and explain how to extract geometrical information from shapes via several measuring functions (Sec. 2). In Sec. 3 we address the problem of retrieving shapes from a database that are similar to a query shape by merging the extracted information through a statistical framework, namely an a contrario methodology. This methodology is governed by a single parameter on a *Number of False Alarms* (NFA). In Sec. 4 experiments are discussed.

2 Size Theory

2.1 Size Functions

Size Functions are shape descriptors suitable for comparison, that are able to capture both topological and geometric properties of shapes. Let us recall the formal definition of a SF (more details can be found in the papers cited above). Consider a continuous real-valued function $\varphi : \mathcal{M} \to \mathbb{R}$, called *measuring function*, defined on a subset $\mathcal{M}$ of a Euclidean space; it is often implicitly defined as the restriction of a function defined on the whole Euclidean space.

The (reduced) *Size Function* of the pair $(\mathcal{M}, \varphi)$ is a function $\ell_{(\mathcal{M},\varphi)}$ from $\{(x, y) \in \mathbb{R}^2 \mid x < y\}$ to $\mathbb{N}$ defined by setting $\ell_{(\mathcal{M},\varphi)}(x, y)$ equal to the number of connected components of the set $\mathcal{M}_y = \{P \in \mathcal{M} \mid \varphi(P) \leq y\}$ that contain at least one point of the set $\mathcal{M}_x = \{P \in \mathcal{M} \mid \varphi(P) \leq x\}$. A discrete version of the theory exists, which substitutes the subset of the plane with a graph $G = (V, E)$ (vertices V, edges E), the function $\varphi : \mathcal{M} \to \mathbb{R}$ with a function $\varphi\prime : V \to \mathbb{R}$, and the concept of connectedness with the usual connectedness notion for graphs.

Figure 1 shows a simple example of SF. In this case the topological space $\mathcal{M}$ is a curve while the measuring function φ is the distance from point C.

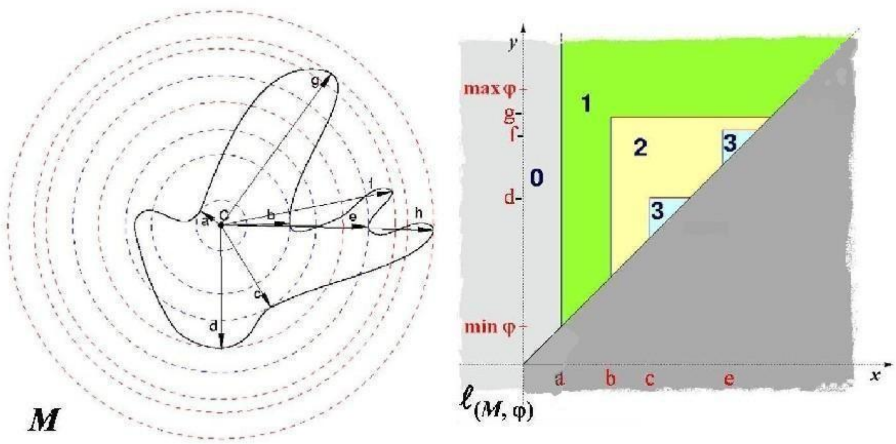

Fig. 1. Left: A pair $(\mathcal{M}, \varphi)$ where $\mathcal{M}$ is the curve depicted by a solid line and φ is the distance from point C. Right: the corresponding reduced size function.

As can be seen in Fig. 1, SFs have a typical structure: They are linear combination (with natural numbers as coefficients) of characteristic functions of triangular regions. That implies that each SF can be described by a formal linear combination of right-angle vertices (called *cornerpoints* and *cornerlines*). Due to this kind of representation, the original complex issue of comparing shapes can be turned into a simpler problem: Each distance between formal series naturally produces a distance between SFs. A detailed treatment of this subject can be found in [11]. Of the many available distances between formal series, the one which is used in this paper is the Hausdorff distance.

Let us moreover notice that SFs inherit the invariance properties, if any, of the underlying measuring function [12].

2.2 Measuring Functions

According to the proposed framework, the representation of shapes by means of SFs is based on the definition of some measuring functions. Such functions are adequate if they satisfy the required invariance properties and produce representations, i.e. SFs, able to distinguish between different shapes. Since for a given measuring function two different graphs can produce the same SF, a representation scheme induced by a single measuring function may not be sufficient to distinguish between shapes: This implies the need for a representation based on different families of measuring functions, in order to increase the discriminatory power. Three different sets of measuring functions were implemented for the shape recognition application described in this paper, that we shall now describe.

First of all, for each image we compute the principal axes. Then each image is normalized, so that the mean distance from the barycenter is equal to one and its center of mass is taken to be the origin O of the reference frame. As a consequence the corresponding SFs turn out to be invariant by rotation, translation and scale. The main drawback of such representation is of course the lack of robustness toward occlusion, which is not the point of this paper. Vectors e_1 and e_2 being the directing unit vectors of the principal axes, (O, e_1, e_2) fix a Cartesian reference frame. From now on, points are identified with their coordinate pairs.

- The first set consists of eight measuring functions defined as the distances from eight points taken on a spiral in the plane of the shape.
  Let $p = (x_p, y_p) \in \mathbb{R}^2$. We define the measuring function $\varphi_p : \mathbb{R}^2 \to \mathbb{R}$ as $\varphi_p(x, y) = \mathrm{d}(p, (x, y))$ with d the Euclidean distance. The formal definition of this first set of measuring functions is:
  $$\phi_1 = \left\{ \varphi_p \text{ s.t. } p = \frac{i}{8}\left(\cos\left(\frac{2\pi \cdot i}{8}\right), \sin\left(\frac{2\pi \cdot i}{8}\right)\right),\ i = 1, \ldots, 8 \right\}.$$
- The second set contains again eight measuring functions, corresponding to the Euclidean distances from eight lines passing through the origin of the reference frame. Denoting $l : y = mx + q$ a line in the plane and defining $\varphi_l : \mathbb{R}^2 \to \mathbb{R}$ as $\varphi_l(x, y) = \mathrm{d}(l, (x, y))$, this second set is defined as:
  $$\phi_2 = \left\{ \varphi_l \text{ s.t. } l : y = \tan\left(\frac{\pi \cdot i}{8}\right) x,\ i = 1, \ldots, 8 \right\}.$$
- The last set ϕ_3 consists of five measuring functions, each one of them having a segment as domain. Five lines are considered, passing through the origin

and defined by the polar angles 0, $\pi/8$, $\pi/4$, $3\pi/8$, $\pi/2$. For each line, the same procedure is applied: the whole image is fibered into a set of n stripes orthogonal to the line; to each stripe k ($k = 1, \ldots, n$), the number of pixels in its intersection with the shape is associated; the final measuring function is obtained by convolving these values with a narrow Gaussian kernel.

3 Merging Measuring Functions Through an a Contrario Model

In a series of articles (see [13] for a comprehensive account), Desolneux, Moisan and Morel recently proposed a new detection methodology for vision based on the so-called *a contrario models*. These authors state that a detection has to be considered as perceptually meaningful as soon as it is not likely to occur "just by chance". The main advantage of this approach is that no a priori model for the structures that are to be detected is needed, and detection can be based just on a model for the statistical *background*. More precisely, a background model (for what is expected to occur "by chance") is first built, then the expectation of a given observation in this model is estimated and called *Number of False Alarms* (NFA). Observations with a corresponding low NFA are then meaningful events, in the sense that they are not likely to have been generated by the background model.

A major difficulty here is to accurately estimate such small probabilities or expectations. Indeed, the number of occurrences of rare events is by definition very low. One way to overcome this difficulty is to estimate these probabilities as a combination of probabilities of independent, more frequently observed events. In computer vision, D. Lowe first proposed such a framework based on the computation of *accidental occurrences* [14]. The a contrario methodology has been recently developed in various situations such as stereo images comparison [15] or contrasted edges detection [16]. It has also been very recently developed towards shape recognition: In [17], a Number of False Alarms is computed for *shape elements* (pieces of level lines), leading to an affine invariant shape recognition algorithm which is robust to noise and occlusions. We shall here derive a Number of False Alarms which is no more adapted to shape descriptors such as pieces of level lines, but to descriptors based on measuring functions.

We give in what follows a simplified goal-driven version of the theory. A rigorous framework for a contrario models dedicated to shape matching is given in [17].

Suppose that the problem is to search a query shape $\mathcal{S}$ among shapes in a database $\mathcal{B}$ (whose cardinality is N). Given $\mathcal{S}$ and a database shape $\mathcal{S}'$ from $\mathcal{B}$, the three distances d_i ($i \in \{1, 2, 3\}$) between groups of size functions are computed:

$$d_i(\mathcal{S}, \mathcal{S}') = \sum_{\varphi \in \phi_i} d_H \left(\ell_{(\mathcal{S},\varphi)}, \ell_{(\mathcal{S}',\varphi)} \right),$$

where d_H is the Hausdorff distance and φ is one of the measuring functions in ϕ_i. Each d_i is then obtained by summing the Hausdorff distances between the size functions corresponding to the group i ($i \in \{1, 2, 3\}$). The problem we are coming up against is to mix the information provided by d_1, d_2, and d_3. Since these distances do not share the same distribution at all, the distance between two shapes cannot simply be defined as the average, or as the maximum of the three distances d_1, d_2, and d_3. A statistical model is needed which we derive from the a contrario paradigm.

For each $i \in \{1, 2, 3\}$ and $\delta > 0$, let us note:

$$H_i(\delta) = 1/N \cdot \# \{\mathcal{S}' \in \mathcal{B} \text{ s.t. } d_i(\mathcal{S}, \mathcal{S}') \leq \delta\},$$

where $\#\cdot$ denotes the cardinality of a finite set. For every $\delta > 0$, $H_i(\delta)$ is an empirical estimate of the probability that a (database) shape lies at a distance d_i less than δ from the query shape $\mathcal{S}$, with respect to the i-th group of size functions.

The background model that permits to mix informations from the three groups is built on the following assumption: $\{\mathcal{S}' \in \mathcal{B} \mapsto d_i(\mathcal{S}, \mathcal{S}'), i = 1, 2, 3\}$ is a set of mutually independent random variables. With this assumption at hand, the probability that, for every $i \in \{1, 2, 3\}$, there exists a shape at a distance d_i less than δ_i is simply given by the product

$$H_1(\delta_1) \cdot H_2(\delta_2) \cdot H_3(\delta_3).$$

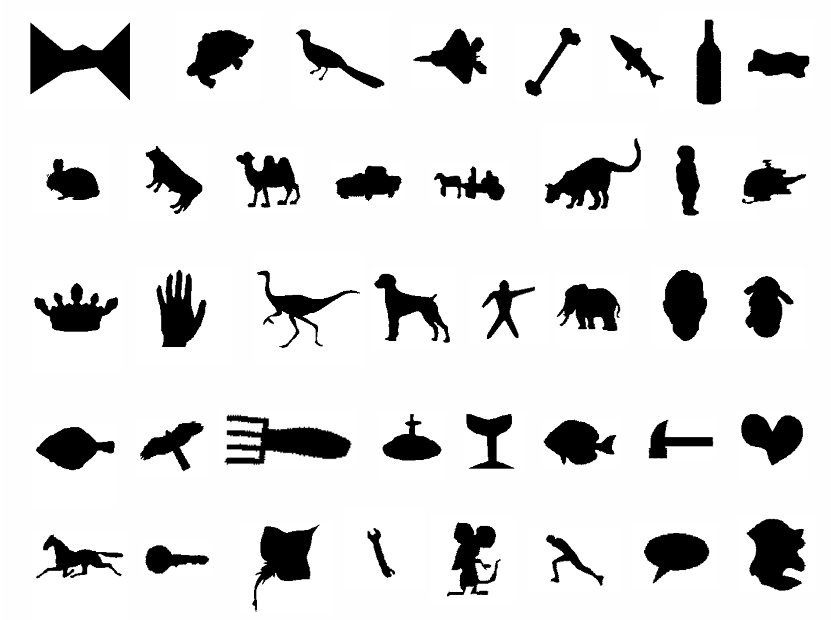

Fig. 2. Sample images from the shape database

This product is called *Probability of False Alarm* (PFA) since, with the supplementary assumption that the sought shape $\mathcal{S}$ is unusual (i.e. has no correct match) among the database shapes, it corresponds to the probability that a shape is at a distance $(\delta_1, \delta_2, \delta_3)$ from $\mathcal{S}$ although it is not a true correct match. Let us moreover note that such an "independence trick" is a standard data analysis technique and intervenes for instance in the naive Bayes classifier.

Since a probability has little meaning per se, we introduce the *number of false alarms* associated to a shape $\mathcal{S}$ at distances $(\delta_1, \delta_2, \delta_3)$, defined by multiplying the PFA by the number of shapes that are compared to the query shape (i.e. the cardinality of the database). This leads to

$$\text{NFA}\,(\mathcal{S}, (\delta_1, \delta_2, \delta_3)) = N \cdot H_1(\delta_1) \cdot H_2(\delta_2) \cdot H_3(\delta_3). \tag{1}$$

We also define the number of false alarms of a match between a query shape $\mathcal{S}$ and a database shape $\mathcal{S}'$ to be

$$\text{NFA}\,(\mathcal{S}, \mathcal{S}') = N \cdot H_1(d_1(\mathcal{S}, \mathcal{S}')) \cdot H_2(d_2(\mathcal{S}, \mathcal{S}')) \cdot H_3(d_3(\mathcal{S}, \mathcal{S}')). \tag{2}$$

Following these definitions, a match between the query $\mathcal{S}$ and a database shape $\mathcal{S}'$ is said to be ε-meaningful if $\text{NFA}(\mathcal{S}, \mathcal{S}') \leq \varepsilon$. It simply springs from the definitions that the number of false alarms is an upper bound of the expected number

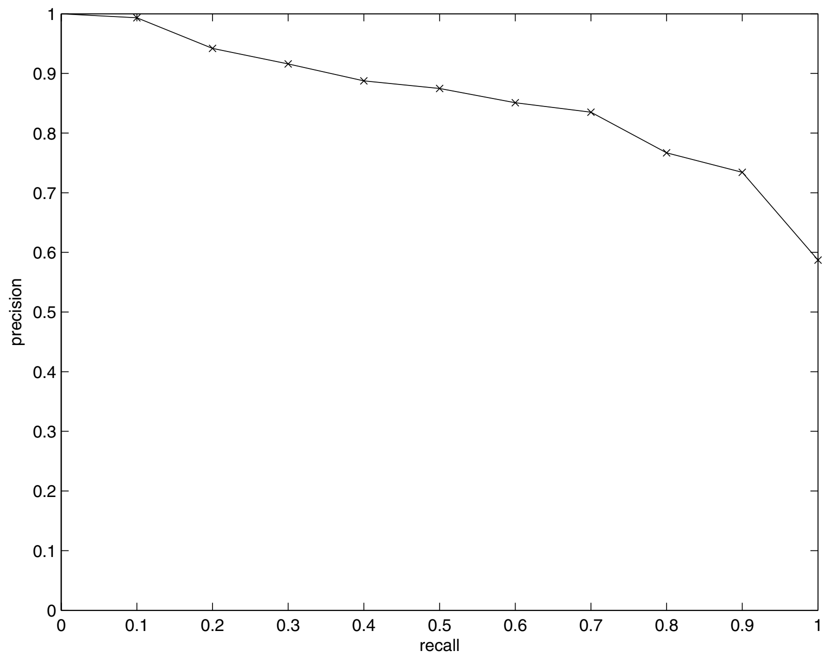

Fig. 3. Average precision-recall graph for fifty randomly selected query shapes from our database. The proposed algorithm shows very convincing performances over the tested database.

of database shapes whose distances (d_1, d_2, d_3) to $\mathcal{S}$ satisfy $d_i < d_i(\mathcal{S}, \mathcal{S}')$ for $i \in \{1, 2, 3\}$, when it is assumed that $\mathcal{B}$ obeys the background model. Consequently, the smaller ε is, the more significant the ε-meaningful matches are.

Let us conclude this section with a remark concerning the lack of completeness of shape representation via size functions. We have pointed out that two unrelated shapes $\mathcal{S}$ and $\mathcal{S}'$ may share the same representation with respect to a SF family. The proposed NFA enables to get rid of this issue: if a single distance (say $d_1(\mathcal{S}, \mathcal{S}')$) is low while the other ones ($d_2(\mathcal{S}, \mathcal{S}')$ and $d_3(\mathcal{S}, \mathcal{S}')$) are quite high, then the corresponding NFA should be large enough, so that $\mathcal{S}'$ is not a meaningful match of $\mathcal{S}$.

4 Experiments and Discussion

In order to assess the validity of the proposed approach, we tested it on the Kimia shape database which consists of 1065 shapes belonging to 40 different categories (1032 of these shapes have been used in [18]). Figure 2 shows sample images from this database.

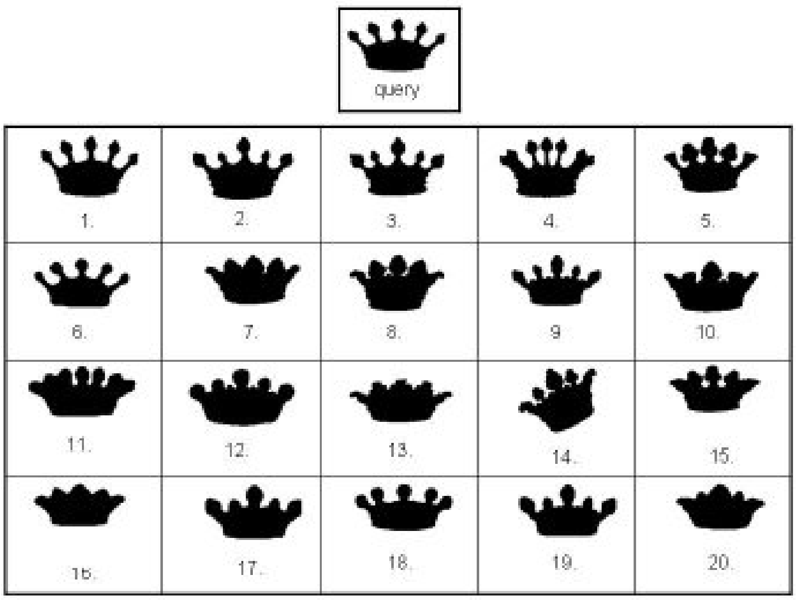

Fig. 4. Looking for a crown in the database: retrieved shapes with a NFA lower than 0.1, sorted along their NFA. The 20 most significantly matching shapes all belong to the correct class. The NFA of the most significantly matching shape (Nr 2, since Nr 1 is the query shape itself) is 5.10^{-5} while the NFA of the 20th match is 8.10^{-2}. The NFA of the 21st (false) match is 0.5 (not shown).

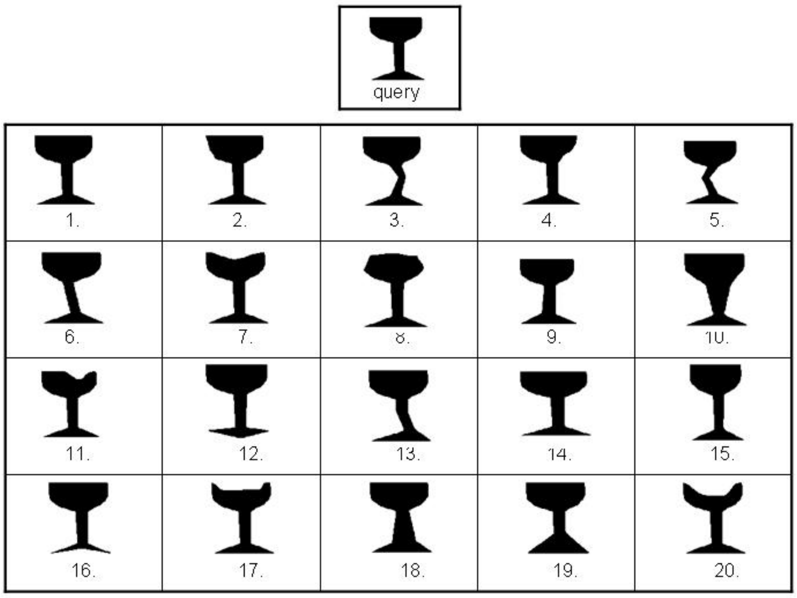

Fig. 5. Looking for a glass in the database: retrieved shapes with a NFA lower than 0.1, sorted along their NFA. The 20 most significantly matching shapes all belong to the correct class. The NFA of the most significantly matching shape is 7.10^{-6} while the NFA of the 20th match is 8.10^{-3}. The NFA of the 21st (false) match is 0.2.

Fifty randomly selected shapes were submitted as queries and compared with all the other shapes in the database. Figure 3 shows the average precision-recall graph (see for example [19]). We remind that the *precision* and *recall* descriptors attempt to measure the effectiveness of the retrieval method by measuring the ability of the system to retrieve relevant objects while discarding non relevant ones. Explicitely,

$$P(k) = \frac{NR(k)}{k} \qquad R(k) = \frac{NR(k)}{N_{rel}},$$

where N_{rel} is the total number of relevant images for a given query and $NR(k)$ is the number of relevant items among the first k retrieved.

Figure 4 (resp. Fig. 5) shows the retrieved shapes when searching for a shape belonging to the *crown* class (resp. to the *glass* class). In both cases, ranking the matching shapes according to their NFA allows to retrieve in the very first positions all the 20 shapes from the correct class. One can see a gap between the NFA of the 20th retrieval (which is generally much lower than 0.1) and the NFA of the 21st retrieval (which is larger than 0.5). Let us recall that the NFA of a match is an upper bound of the expected number of fortuitous occurences of such a match in the database. A match with a NFA larger than 0.5 implies poor confidence since it is too likely that it could appear "by chance"

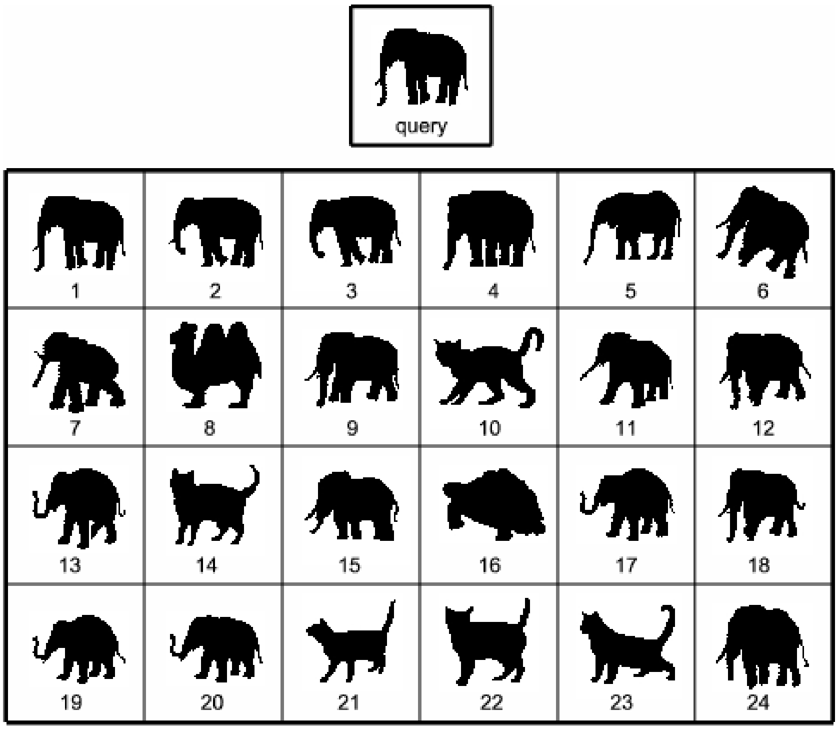

Fig. 6. Looking for an elephant in the database: retrieved shapes with a NFA lower than 0.1, sorted along their NFA. False matches can be seen because for example tails and snoots may be mixed in the size function representation of the perceptual information.

almost once. The experiment was led for all fifty randomly selected queries and gave similar results. Figure 6 shows the 24 retrieved shapes with a NFA lower than 0.1 when searching for an elephant. Some "false alarms" can be seen with a quite low NFA; the camel's NFA (nr 8) is for example 2.10^{-3}. Some elephants are missed and have a NFA larger than 0.1. However, this class of shapes is quite difficult to discriminate and it is not separated enough from other animal classes (roughly speaking, a tail may be coded as a trunk through a SF.)

This experimental assessment is quite promising: shape classes show strong enough variations that make the retrieval task difficult for methods dealing with exact similarities (such as normalized shapes or high order moments comparison). These very first results show 1) that size functions are handy shape descriptors and permit to capture a convenient amount of discriminative perceptual information, and 2) that an a contrario model successfully mixes information coming from different size functions families. A single detection threshold over the NFA replaces three thresholds over each SF family. Figure 6 shows that the perceptual information is not fully represented by size functions, leading to false matches. We are currently working on other groups of size functions that would capture supplementary perceptual information.

Acknowledgements. This work is supported by Galileo project between École Normale Supérieure de Cachan (France) and University of Bologna (Italy). We are grateful to Patrizio Frosini and Jean-Michel Morel for fruitful discussions.

References

1. Zahn, C., Roskies, R.: Fourier descriptors for plane closed curves. IEEE Transactions on Computers **C-21**(3) (1972) 269–281
2. Dudani, S., Breeding, K., McGhee, R.: Aircraft identification by moment invariants. IEEE Transactions on Computers **26**(1) (1977) 39–46
3. Monasse, P.: Contrast invariant image registration. In: Proceedings of International Conference on Acoustics, Speech and Signal Processing. Volume 6., Phoenix, Arizona, USA (1999) 3221–3224
4. Hu, M.: Visual pattern recognition by moments invariants. IRE Transactions on Information Theory **8** (1962) 179–187
5. Khalil, M., Bayoumi, M.: A dyadic wavelet affine invariant function for 2d shape recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **23**(10) (2001)
6. Mokhtarian, F., Mackworth, A.: A theory of multiscale, curvature-based shape representation for planar curves. IEEE Transactions on Pattern Analysis and Machine Intelligence **14**(8) (1992) 789–805
7. Alvarez, L., Mazorra, L., Santana, F.: Geometric invariant shape representations using morphological multiscale analysis and applications to shape representation. Journal of Mathematical Imaging and Vision **18**(2) (2002) 145–168
8. Sclaroff, S., Pentland, A.: Modal matching for correspondence and recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **17**(6) (1995) 545–561
9. Verri, A., Uras, C., Frosini, P., Ferri, M.: On the use of size functions for shape analysis. Biological Cybernetics **70** (1993) 99–107
10. Cerri, A., Ferri, M., Giorgi, D.: A new framework for trademark retrieval based on size functions. In: Proceedings of International Conference on Vision, Video and Graphics, Heriot Watt University, Edinburgh (2005) 167–172
11. Frosini, P., Landi, C.: Size functions and formal series. Applicable Algebra in Engineering Communication and Computing **12** (2001) 327–349
12. Verri, A., Uras, C.: Invariant size functions. In Mundy, J., Zisserman, A., Forsyth, D., eds.: Application of Invariance in Computer Vision. Springer (1994) 215–234
13. Desolneux, A., Moisan, L., Morel, J.M.: Computational Gestalt Theory. To appear in Lecture Notes in Mathematics, Springer (2006)
14. Lowe, D.: Perceptual Organization and Visual Recognition. Kluwer Academic Publisher (1985)
15. Moisan, L., Stival, B.: A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix. International Journal of Computer Vision **57**(3) (2004) 201–218
16. Desolneux, A., Moisan, L., Morel, J.M.: Edge detection by Helmholtz principle. Journal of Mathematical Imaging and Vision **14**(3) (2001) 271–284

17. Musé, P., Sur, F., Cao, F., Gousseau, Y., Morel, J.M.: An a contrario decision method for shape element recognition. International Journal of Computer Vision (2006) To appear.
18. Sebastian, T., Klein, P., Kimia, B.: Shock-based indexing into large shape databases. In: Proceedings of the European Conference on Computer Vision (ECCV 2002). Volume LNCS 2352., Copenhagen, Denmark (2002) 731–746
19. Salton, G., McGill, M.: Introduction to modern information retrieval. McGraw Hill (1983)