Robust Normalization of Shapes

Javier Cortadellas[1], Josep Amat[2], and Manel Frigola[3]

[1]Departament d'Electrònica, Enginyeria La Salle, Universitat Ramon Llull
Pso. Bonanova 8, 08022- Barcelona, Spain
xavicort@salleURL.edu
[2]IRI - Institut de Robòtica e Informàtica, Universitat Politècnica de Catalunya
Llorens Artigas 4-6, 08028 - Barcelona, Spain
amat@iri.upc.es
[3]Departament d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial
Universitat Politècnica de Catalunya
C. Pau Gargallo 5, 08028 Barcelona, Spain
frigola@esaii.upc.es

Abstract. The normalization of a binary shape is a necessary step in many image processing tasks based on image domain operations. When one must deal with deformable shapes (due to the projection of non-rigid objects onto the image plane or small changes in the position of the view point), the traditional approaches doesn't perform well. This paper presents a new method for shape normalization based on robust statistics techniques, which allows to keep the location and orientation of shapes constant independent of the possible deformations they can suffer. A numerical comparison of the sensitivity of both methods is used as a measure to validate the proposed technique, together with a ratio of areas between the non-overlapping regions and the overlapping regions of the normalized shapes. The results presented, involving synthetic and real shapes, show that the new normalization approach is much more reliable and robust that the traditional one.

1   Introduction

The normalization of a binary shape is a necessary step in many image processing tasks based on image domain operations. It is even a crucial step when the main goal of a computer vision system is the discrimination of objects depending on their visual appearance. In such systems, the way an object is represented arises as an important question to be solved. Although other appearance properties such as color or texture can be exploited, shape can be used as a powerful tool for describing and recognizing objects, if a reliable image segmentation method is available.

The shape of an object refers to its profile and physical structure. In this paper, shapes are considered binary images that come from the projection of 2D or 3D objects onto a 2D image plane (they are also usually called silhouettes [14]). There are basically four ways of representing this important characteristic: by means of boundaries, regions, moments or structural representations. Among them, the description based on the region that an arbitrary shape occupies is the simplest one. The benefits of this simple description are that it has a straight meaning for human visual perception systems, it avoids the computation of shape descriptors [3,4,5], which can be ambiguous, and finally, as the description lies on the image domain, it is

possible to use specialized hardware architectures like [13] that operate over the image domain at very high speeds.

However, the main drawback for this kind of representation is that, when used in shape recognition tasks, it needs to be normalized with respect to similarity transformations (translation, rotation and scaling) in order to grant a unique representation of the projection of an object onto an image plane. Traditionally, the centroid (also called center of gravity) and the angle of the axis of the least moment of inertia have been used to normalize the localization and orientation of the shapes [2,3,4,5]. A standard method consists of rotating the shape around the centroid so that the axis previously mentioned, called principal axis, has some predefined orientation.

This normalization does not present difficulties when dealing with shapes of rigid objects that have always been projected onto the image plane from the same point of view. However, problems arise when one must deal with deformable shapes, produced by small changes of perspective or alterations of the physical structure of non-rigid objects, which produces significant changes of the periphery of their shape.

In many applications such as shape recognition or registration, it would be desirable that two different shapes of a same object class share to a certain extent (depending on the degree of deformation) the same location and orientation, because this fact would ease further processing tasks, but the classical normalization approaches doesn't perform well when dealing with such deformed shapes.

The main contribution of this paper is the development of a reliable technique for accurately estimating the robust centroid and the robust principal axis of deformable shapes, making use of robust statistics [1,7,8]. Size normalization is not treated in this paper because it is assumed that objects with different scale must be discriminated. The structure of the paper is as follows: after this introduction, the traditional approach for normalizing shapes, based on the centroid and the principal axis, is briefly described. After pointing out some of the existing problems of this approach, a new robust normalization method is developed. Then, by means of a similarity measure based on the discrepancy of overlapping areas, the proposed method is compared with the traditional normalization technique. Some results involving synthetic and real shapes are also presented and finally, the conclusions are summarized.

2 Normalization of Shapes. Classical Method

There are a number of different methods that have been used to determine the orientation of a shape [3], like the maximum Feret's diameter – the line between the two points on the periphery that are furthest apart - or the major axis of an ellipse fitted to the contour of the shape silhouette. Among these methods, those that take into account all the pixels of the shape have often worked better than those based on contour or boundary representations, because they are less influenced by the presence or absence of a single pixel around the periphery. One of the most widely used methods among those that consider the whole shape is related to the computation of the principal axis.

It is important to mention that the main goal of this work is not to find a meaningful axis in terms of visual perception but to find an axis insensitive to alterations of the shape due to admissible deformations. This axis would allow different deformed shapes of a same object to be all aligned in a fixed direction.

2.1 Principal Axis Based Orientation Method

In this method, the centroid is used to normalize the location and the orientation is defined as the angle of the least moment of inertia axis [5]. This axis corresponds to the line about which it takes the least amount of energy to spin an object of like shape. It is called *principal axis*, and can be regarded as the line that "best fits" the shape. It is obtained by minimizing, with respect to the angle θ, the sum of the squared distances of each point of the shape to an axis that, passing through the centroid of the shape, has a slope with value *tan θ*.
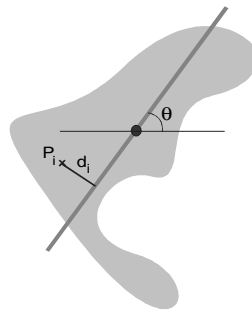

Fig. 1. Arbitrary shape with its principal axis, which is the line that minimizes the sum of the distances (d_i) of all the points of the shape (P_i). Assuming discrete images, this points are the coordinates of the pixels of the shape.

  The result of this minimization is [3,5]:

$$\theta = \frac{1}{2}\tan^{-1}\left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}}\right) \tag{1}$$

where μ_{jk} are the central moments of the shape. Being I a binary-valued picture, and $S = \{(x,y) \mid I(x,y) = 1\}$ the set of pixels representing a two-dimensional shape, for each pair of nonnegative integers (j,k), the above central moments of S are given by [5]

$$\mu_{jk} = \sum_{(x,y)\in S}(x - \bar{x})^j(y - \bar{y})^k \tag{2}$$

where $(\bar{x}, \bar{y})$ are the coordinates of the centroid of the shape, also used for location normalization. Moments can be given a physical interpretation by regarding S as an area composed of a set of point masses located at the points (x,y), and thus providing useful information about the spatial arrangement of the points of S. The standard method for normalizing S with respect to rotation is to rotate it in such a way that its principal axis has some standard orientation, say vertical.

  The main problem of this principal axis approach is that, although it is quite insensitive to small variations on the boundary of the shape (i.e. due to discretization noise), it becomes very sensitive when the shape varies significantly (i.e. due to structural changes of the object it represents). This fact can be seen in Fig. 2.
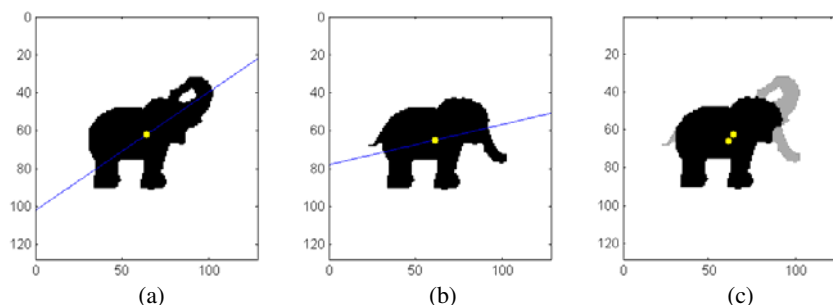
Fig. 2. Principal Axis estimation of a deformable object: (a) θ = 30.47º, centroid =(64,61), (b) θ = 13.92º, centroid=(61,65). (c) Overlapping of the two previous images. Notice the difference between the two shapes and its centroids.

This exemple shows two instances (a and b) of the same class of object "elephant". Despite of the fact that both shapes have a very similar structure, - and thus it would be desirable that they share, to a certain extent, the same orientation and centroid -, there is a difference of more than 16 degrees between both principal axis and of 7 pixels between the centroids (using the chessboard metric). These normalization results would be unacceptable in a template matching framework for instance, because the missalignment would cause an important decrease of the similarity measure.

The reason for these differences is that a region on the periphery of the shape has changed (in this case, the position of the elephant's trunk). This change has biased the centroid and orientation estimation.

3   Robust Normalization of Shapes

Considering the problem stated above, the objective now is to reduce the sensitivity of the centroid and principal axis estimation to shape deformations. To achieve this goal, it is assumed that not all the regions of a shape are equally important when computing the principal axis. This means that pixels that are far away from the tendency of the bulk of the pixels or that are in regions easily subjected to changes should be downweighted, instead of letting each one vote equally in the estimation. This weighting can be dealt by means of robust statistics.

Robust estimation has been successfully used to solve many computer vision, signal processing and statistical problems [1,6,9,10]. There are many robust procedures to tackle these problems, such as Least Median Squares, Ransac, M-estimator, etc (see [7,10] for a review). In this paper, an M-estimator is chosen since the solution can be found with a continuous derivative process, which is very appealing.

In order to apply robust statistics to the principal axis estimation, the problem has been reformulated. It can be shown [3] that the orientation computed in (1) coincides with the direction of the eigenvector corresponding to the larger eigenvalue of this matrix:

$$\begin{pmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{pmatrix} \qquad (3)$$

Following a similar formulation as in [15], this eigenvector estimation problem can be related to the minimization of the following energy function, based on a quadratic norm:

$$E\{\vec{v}, c_i\} = \sum_{i=1}^{N} \left\| \vec{d}_i - \vec{m} - \vec{v} \cdot c_i \right\|_2^2 \qquad (4)$$

where $\vec{d}_i = (x_i, y_i)^T$ are the coordinates of the N pixels of the shape, $\vec{m} = (\bar{x}, \bar{y})^T$ is the centroid, c_i are the linear coefficients that minimize the projection error and $\vec{v}$ is the eigenvector associated to the largest eigenvalue of Eq. 3.

It is well known that in a least squares framework, this energy function minimization would not be robust to outliers, which can be interpreted as the pixels that are not close to the principal axis. Changes in the position of these pixels would cause significant variations in the orientation estimation. For this reason, the quadratic function in the energy function for the least square estimation is replaced by a *p*-function, which assigns smaller weights for constraints with larger residues. Two *p* - functions commonly used in computer vision are the Lorentzion function and the Geman & McClure's function given as follows [6,7,9]:

$$\rho_{LO}(x, \sigma) = \log\left(1 + \frac{x^2}{2\sigma^2}\right), \quad \rho_{GM}(x, \sigma) = \frac{x^2}{\sigma^2 + x^2} \qquad (5)$$

where x is the residue of the data constraint and σ is the scale parameter.

When using the *p*-function for robust estimation in an error minimization framework (for example, line fitting), the influence of each data constraint to the solution is characterized by an influence function, which is the derivative of the *p*-function. If the derivatives of the above two *p*-functions are taken, it is straightforward to see that the influence functions decrease as the magnitude of the residue increases [1]. For the least-square estimation, the influence function is linearly increasing as the magnitude of the residue increases. Therefore the least-square estimation is more sensitive to outliers than the robust estimation.

Incorporating the *p*-function into Eq. 4 yields the following new energy function

$$E\{\vec{v}, c_i\} = \sum_{i=1}^{N} \rho\left(\vec{e}_i^T \cdot \vec{e}_i, \sigma\right) \qquad (6)$$

where $\vec{e}_i = \vec{d}_i - \vec{m} - \vec{v} \cdot c_i$ is the error vector and σ is the scale parameter. In this paper, the Geman-McClure *p*-function has been used. It must also be remarked that unlike [6], intra-sample outliers have not been taken into account, because it has no meaning that only the *x* or the *y* coordinate of a distant pixel were an outlier.

In order to minimize (6), the relation between robust statistics and IRLS (Iterative Reweighted Least-Squares) has been exploited [6], reformulating the previous robust estimation problem in terms of Weighted Least Squares:

$$E\{\vec{v},c_i\}=\sum_{i=1}^{N} w_i \left(\vec{d}_i - \vec{m} - \vec{v}\cdot c_i\right)^T \left(\vec{d}_i - \vec{m} - \vec{v}\cdot c_i\right) \qquad (7)$$

with weights w_i related to the derivative of the p-function by

$$w_i = \frac{\dfrac{d\rho\,(\vec{e}_i^{\,T}\cdot\vec{e}_i,\,\sigma)}{d\,(\vec{e}_i^{\,T}\cdot\vec{e}_i)}}{\vec{e}_i^{\,T}\cdot\vec{e}_i} \qquad (8)$$

An efficient way of computing the minimization of Eq. (8) consists of applying Alternated Least Squares. The iterative scheme, which can be verified taking the partial derivatives with respect to $\vec{v}$ and c_i, would be:

$$\vec{m}(k+1) = \frac{\sum_{i=1}^{N} w_i(k)\cdot\vec{d}_i(k)}{\sum_{i=1}^{n} 1} \qquad (9)$$

$$\vec{v}(k+1) = \frac{\sum_{i=1}^{N} w_i(k)\cdot c_i(k)\cdot\left(\vec{d}_i(k) - \vec{m}(k)\right)}{\sum_{i=1}^{n} w_i(k)\cdot c_i^{\,2}(k)} \qquad (10)$$

$$c_i(k+1) = \frac{\vec{v}(k)^T \cdot \left(\vec{d}_i(k) - \vec{m}(k)\right)}{\vec{v}(k)^T \cdot \vec{v}(k)} \qquad (11)$$

It is important to notice that the robust centroid of the shape (9) is also iteratively computed taking into account the weights w_i associated to each location of pixels of the shape. It is also worth mentioning here that at every iteration k the weights are updated following Eq. (8).

A solution can be usually reached in less than 10 iterations, and the computational cost of each iteration is just $O(n)$. The scale parameter σ can be automatically estimated using the Median Absolute Deviation (MAD) [7], which can be viewed as a robust statistical estimate of the standard deviation.

3.1   Shape Dependent Weighting Function

As it has been shown in the previous section, it is necessary to weight the contribution of the pixels to the principal axis computation. These weights should penalize those pixels that are far away from the tendency of the bulk of the pixels or that are more likely to vary due to shape deformations. However, Equation (8) does not consider the later dependence on the shape, as it can be seen in Fig. 3.
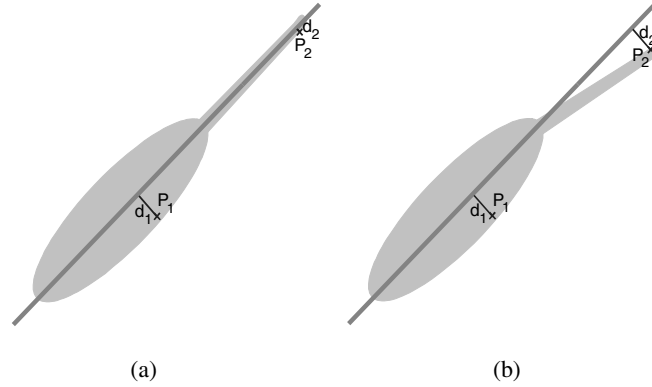
Fig. 3. Two instances of a deformable shape, with the desired principal axis and the distances of two pixels to this axis.

This example shows two instances of a deformable shape. It would be desirable that both have the same principal axis, regardless of small changes of the position of the tail. In Fig. 3(a) the pixel P_2 is closer to the axis than P_1, so according to Eq. (8), its influence on the final solution is bigger. However, in Fig. 3(b) pixel P_2 and pixel P_1 are at the same distance, so both have now the same influence on the final solution. A small change in the position of the tail can causes significant differences in its associated weights, affecting the centroid and orientation estimation. It is clear that the tail of this shape is more likely to vary, and that the influence of P_2 in comparison with the influence of P_1 should be downweighted in both figures.

In order to cope with these situations and to add more meaning to the weights, the original weights w_i are modified by multiplying them by a shape dependent weighting function $M(x_i, y_i)^\beta$, whose value depends on the location of each pixel i of the shape. These new weights Ω_i are:

$$\Omega_i = w_i \cdot \left(M(x_i, y_i) \right)^\beta \tag{11}$$

where β is a constant factor that controls the influence of the periphery on the robust principal axis estimation.

The shape dependent weighting function function $M(x_i, y_i)^\beta$ must penalize those regions of the shape which are more likely to vary due to deformations or changes of the shape. As it is assumed that the major alterations occur on the periphery of the shapes, a measure of the distance between the pixels of the shape and its boundary is a good candidate for being $M(x_i, y_i)$. In this paper, the chessboard distance metric [3,11] has been used to compute $M(x_i, y_i)$. It is defined as follows:

$$d_{ch}(p_i, p_j) = \max \left\{ \left| x_i - x_j \right|, \left| y_i - y_j \right| \right\} \tag{12}$$

for pixels $p_i = (x_i, y_i) \in S$ and $p_j = (x_j, y_j) \in S_B$ where S_B denotes the set of shape pixels that are in the boundary of the shape. According to this, $M(x_i, y_i)$ is defined for each pixel of the shape as the distance d_{ch} from that pixel to the nearest pixel on the boundary. There are efficient algorithms for computing distance maps based on the

chessboard distance metric that can be used to generate $M(x_i, y_i)$ [2,3]. Figure 4 shows two examples of shape dependent weighting functions $M(x_i, y_i)$ for a deformable shape.

The β factor depends on the dimension of the image, i.e. the size of the shape. The bigger the β factor is, the more down weighted are the outer pixels of the shape in relation to the inner pixels. Consequently, this factor also downplays the effect of spurious edges and noise due to binarization on the boundary of the shape. In the experiments, a value between 0.5 and 2 worked well, being 1.4 the standard value used for images or 128x128 pixels.
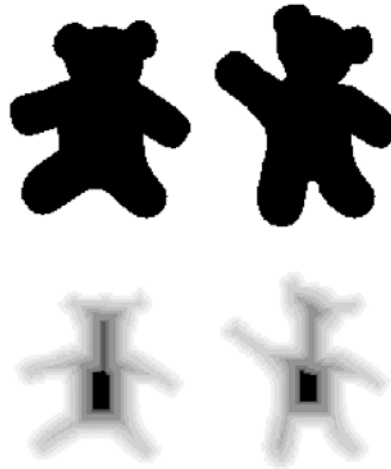

Fig. 4. Above, two shapes of the same deformable object class "teddy bear". Below, their corresponding shape dependent weighting functions $M(x_i, y_i)$ showing that the parts more likely to vary (arms and legs) have smaller associated weights.

4   Evaluation Method

In order to measure the effectiveness of the proposed robust orientation and localization method in comparison with the traditional principal axis estimation, three parameters are carried out. First, the difference between the centroid estimation of a a shape (COG_{DIFF}) that is subjected to a deformation, and has been computed using the chessboard metric as $COG_{DIFF} = |x_1 - x_2| + |y_1 - y_2|$. The second is the change in the orientation estimation (θ_{DIFF}) when a shape has been deformed and has been computed for both the standard method (T) and the robust method (R) as $\theta_{DIFF} = \theta_1 - \theta_2$. Finally, a similarity measure (δ) is carried out for pairs of shapes once they have been aligned using the previous estimations.

This third parameter has been introduced to validate the use of this normalization procedure in a deformable object recognition framework based on shapes, and it is closely related to the work of P. Rosin [12].

Given the areas shown in Figure 5: R the difference between the original shape and the modified shape, D the difference between the modified and the original shape, and T the overlapped region common to both shapes, then the similarity measure δ is

$$\delta = 1 - \frac{R+D}{T} \qquad (13)$$

This parameter measures the normalized discrepancies between the areas of the aligned shapes and has a maximum equal to 1 in case that both the original and the deformed shape are the same.

In contrast to the simple sum of the overlapping area, which would be equivalent to a template matching approach, this measure is valid even when due to the deformation, one shape encloses the other.

The main difference between this measure and those from [12] is that here the objective is not to measure shape properties but to recognize them. Hence, the errors have been normalized by the intersection of areas instead of the whole area of the reference shape to provide better selectivity when it comes to recognize shapes.
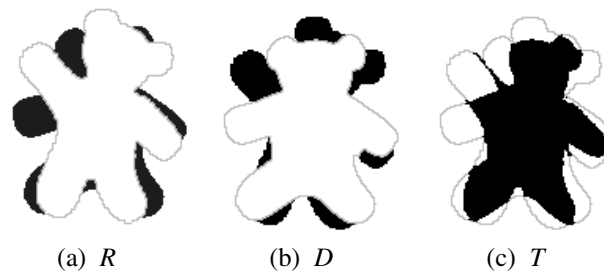

(a) *R*   (b) *D*   (c) *T*

Fig. 5. Areas (in black) used in the similarity measure based on the discrepancy method, applied to the deformable shape shown in Fig. 4.

5   Results

The robust and the traditional normalization method have been applied to a set of deformable shapes related to recognizable objects. Some of these shapes are shown in Fig. 6. The procedure has been as follows: for each shape, both the standard method based on moments and the robust normalization method have been carried out. This results in two estimates, the centroid and the orientation for both methods. Then each shape has been rotated around its centroid in order to align the principal axis vertically and it is shifted to place the centroid of the shape centered in the image.

Table 1 condenses some results about the reliability of the robust method. The left column indicates the pair of shapes whouse parameters have been compared. This table clearly suggests that robust normalization (R) is much less sensitive to changes of the shape than the traditional approach (T). The maximum difference between the orientation estimation of a shape that has been deformed is of less than 3°, in comparison with the 29° of the classical method.

There is also a significant improvement in the stability of the centroid estimation, as can be seen in the comparison between shapes h1 and h2. The similarity measure has also verified that normalizing shapes with respect to robust estimates results in a better resemblance measurement between shapes, although the selectivity was not as high as expected. This is due to the fact that great part of the pixels of two silhouettes

of the same class of object (those sorrounding the centroid) overlap even when the difference of orientation is significant.

These results also show that when two shapes that belong to different object classes are compared, the similarity measure is quite low, thus it is nonsense to compare in this cases the robust and traditional estimation methods.

Table 1. Comparison between the standard (T) and robust (R) normalization of shapes of deformable objects.

	θ_{DIFF}		COG_{DIFF}		δ	
	T	R	T	R	T	R
a1/a2	16.64°	2.61°	7	1	0.61	0.88
b1/b2	12.54°	1.26°	2	1	0.58	0.84
c1/c2	14.1°	2.97°	6	1	0.59	0.77
d1/d2	6.25°	0.42°	7	2	0.72	0.90
e1/e2	29.05°	0.01°	4	0	0.50	0.91
f1/f2	2.21°	0.19°	2	1	0.76	0.80
g1/g2	7.56°	2.88°	9	3	0.53	0.63
h1/h2	9.98°	0.27°	10	1	0.61	0.72
i1/i2	6.59°	1.47°	6	1	0.42	0.74
a1/f2	-	-	-	-	0.47	0.42
e1/b2	-	-	-	-	0.5	0.52
f1/i2	-	-	-	-	0.39	0.48
i1/d2	-	-	-	-	0.33	0.28

6   Conclusions and Further Work

The classical location and orientation normalization method, based on the centroid and the least moment of inertia axis, is not appropriate when dealing with deformable shapes, because it is based on a least squares minimization framework that is not robust to outliers. They can be interpretated as pixels that are far away from the tendency of the bulk of the data, and have a strong influence on the estimation, so minor changes of their position have a great impact on the final solution.

By means of a reformulation in terms of an energy function, it has been possible to incorporate robust statistics and make the estimations much less sensitive to variations of the shape. This goal has been attained by using a shape dependent weighting function, which penalizes those pixels of the shape that are more likely to be subjected to changes due to deformations, i.e. the periphery of the shape. Finally, to prove the validity of the proposed method, a similarity measure has been introduced. It is based on the discrepancy between the areas of two shapes, once they have been aligned with both the traditional and the robust normalization methods.
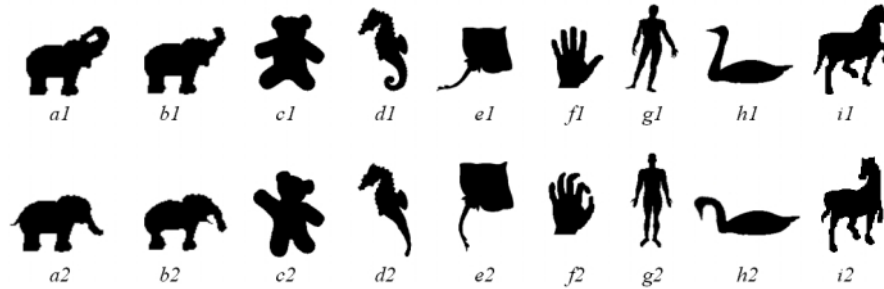
Fig. 6. The set of shapes used for evaluation purposes

The experimental results have demonstrated the superior performance of the robust approach with respect to the traditional one, even when shapes suffer important structural changes. Although the similarity measure is able to differentiate among the deformable shapes and validates the proposed normalization method, it turned out to be not as selective as it would be desirable.

It is important to remark that the method tolerates deformations of the shapes to a certain extent, over which it can be considered that the shape is scarcely resembling to the original one and thus, the normalization procedure results in erroneous estimations.

This paper does not deal with size normalization. It has been assumed that different instances of the same object are available, so changes of size are only due to deformations, which have been treated. Future research will focus on the scale normalization.

References

[1]   M. J. Black and P. Anandan. "The robust estimation of multiple motions: parametric and piecewise-smooth flow-fields". Computer Vision Image Understanding, Vol. 63, No. 1, pp. 75-104, 1996.
[2]   J. C. Russ. The Image Processing handbook, 2nd. edition. CRC Press, 1995.
[3]   A. Rosenfeld, A. C. Kak. Digital Picture Processing, Vols. 1 & 2. Academic Press, 1982.
[4]   A. K. Jain, Fundamentals of Digital Image Processing. Prentice Hall, 1989.
[5]   R. M. Haralick, L. G. Shapiro. Computer and Robot Vision, Vols. I&II. Addison-Wesley, 1993.
[6]   F. de la Torre, M. Black. "A Framework for Robust Subspace Learning". Accepted for Int. Journal of Computer Vision, 2002.
[7]   F. Hampel, E. Ronchetti, P. Rousseeuw, W. Stahel. Robust Statistics: The Approach based on Influence Functions. Wiley, 1986.
[8]   P. J. Huber. Robust Statistics. Wiley, 1981
[9]   S. Geman, D. McClure. "Statistical methods for tomo-graphic image reconstruction". Bulletin of the Inter-national Statistical Institute. Vol. LII, pp. 4-5, 1987.
[10]   D. Mintz, P. Meer. "Robust Estimators in Computer Vision: an Introduction to Least Median of Squares Regression". Artificial Intelligence and Computer vision, Y.A. Feldman, A. Bruckstein (Eds.). Elsevier, 1991.
[11]   G. Borgefors. "Distance transformations in arbitrary dimensions". CVGIP, Vol.27, pp.321-345, 1984

[12] P.L. Rosin. "Measuring Shape: Ellipticity, Rectangu-larity, and Triangularity". 15th Int. Conf. Pattern Recognition, Barcelona, Spain, vol. 1, pp. 952-955, 2000.

[13] J. Cortadellas, J.Amat, "Image Associative Memory". 15th Int. Conf. Pattern Recognition, Barcelona, Spain, vol. 3, pp. 638-641, 2000.

[14] V. Bruce, P.R. Green, M.A. Georgeson. Visual Perception. Physiology, Psychology and Ecology. 3rd. Ed., Psychology Press, 1996.

[15] L. Xu, A. Yuille, "Robust Principal Component Analysis by Self-Organizing Rules Based on Statistical Physics Approach", IEEE Trans. on Neural Networks, Vol. 6, Nº 1, January 1995