Robustness of Shape Descriptors to Incomplete Contour Representations

Material based on:

Anarta Ghosh and Nicolai Petkov – *"Robustness of Shape Descriptors to Incomplete Contour Representations"*, IEEE Transactions on Pattern Analysis and Machine Intelligence Vol. 27, No. 11, pp. 1793-1804, Nov. 2005.

Anarta Ghosh and Nicolai Petkov – *"A Cognitive Evaluation Procedure for Contour Based Shape Descriptors"*, International Journal of Hybrid Intelligent Systems, Special Issue. Vol. 2, No. 4, pp. 237-252, 2006.

Chapter 2

Incompleteness and Contour Based Shape Descriptors

Abstract

With inspiration from psychophysical researches of the human visual system we propose a novel aspect and a method for performance evaluation of contour based shape recognition algorithms regarding their robustness to incompleteness of contours. We use complete contour representations of objects as a reference (training) set. Incomplete contour representations of the same objects are used as a test set. The performance of an algorithm is reported using the recognition rate as a function of the percentage of contour retained. We call this evaluation procedure the ICR test. We consider three types of contour incompleteness, viz., segment-wise contour deletion, occlusion and random pixel depletion. As an illustration, the robustness of two shape recognition algorithms to contour incompleteness is evaluated. These algorithms use a shape context and a distance multiset as local shape descriptors. Qualitatively, both algorithms mimic human visual perception in the sense that recognition performance monotonously increases with the degree of completeness and that they perform best in the case of random depletion and worst in the case of occluded contours. The distance multiset method performs better than the shape context method in this test framework.

2.1 Introduction

If we look at the objects in the Figure 2.1 we can instantly recognize birds, even though 50% of the contour is removed segment-wise in the left image, the right half of the contour is not visible in the middle image, and 80% of the contour points have been removed (randomly) in the right image. This ability of human beings to recognize objects with incomplete contours was studied by the psychologist E. S. Gollin

(a) (b) (c)

Figure 2.1: *A bird can be recognized even though (a) 50% of its contour has been removed segment-wise, (b) its tail part is not visible (occluded), (c) 80% of the contour pixels have been randomly removed.*

(Gollin 1960). His objective was to investigate the performance of humans in recognizing objects with incomplete contours as a function of developmental characteristics, such as mental and chronological age and intelligence quotient. The subjects of his experiments were children of different age groups and a group of adults. Gollin used sets of contour[1] images with different degrees of incompleteness (Figure 2.2) and addressed the following questions: (1) In order to be recognized, how complete the contours of common objects need to be? (2) How does training affect the recognition performance in case of incomplete representations? The main conclusions drawn by him through his experiments were: human ability to recognize objects with incomplete contours (a) depends on intelligence quotient and (b) is improved by training.

In the context of processing visual information using computers this aspect of recognition of objects with incomplete contours is also very important. Figure 2.3 shows a natural image and two edge images, obtained from it. The middle image was obtained by applying a bank of Gabor energy filters. It contains the contours of the object of interest, viz., a rhinoceros, but it also contains a large number of texture edges in the background that are not related in any way to the shape of the rhinoceros. These texture edges will have a devastating effect on the performance of all currently known contour based shape recognition algorithms. Advanced contour detection methods based on surround suppression (Grigorescu et al. 2004), (Grigorescu et al. 2003) succeed in separating the essential object contours from the

[1]By "contours" in the following we refer to both occluding boundaries and inner edges that are defined by boundaries of parts of an object or perceptually important color or texture regions.

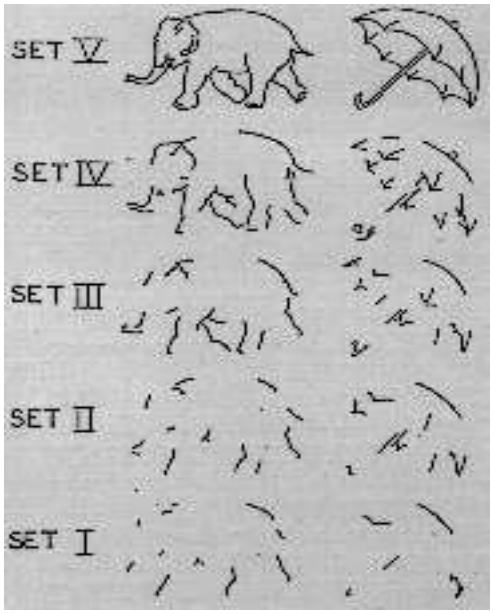


Figure 2.2: *Example of image sets used in Gollin's original test (Gollin 1960). The images in set V are complete contour representations and the other sets are derived from set V by removing segment-wise an increasing fraction of the contour.(Reproduced with permission of author and publisher from: Gollin E. S. Developmental studies of visual recognition of incomplete objects.* Perceptual and Motor Skills*, 1960, 11, 289-298. ©Southern University Press 1960.)*

texture edges, as illustrated by the right-most image in Figure 2.3, but at the same time these methods have a certain negative side effect of depleting the contours of the objects of interest. Hence, the robustness of shape recognition methods to contour incompleteness is an issue of practical importance.

With inspiration from Gollin's work we propose a novel attribute, viz., *robustness to incomplete contour representations*, that any contour based object recognition system/algorithm should have. The objective of this study is to show how the performance of recognition systems/algorithms can be investigated in an idealized situation where: (a) complete contour representations of the objects to be recognized form the reference (training) set or "memory" of the system/algorithm, (b) incomplete contour representations of the same objects are derived from the afore mentioned complete representations and are used as a test set, (c) the performance of the system/algorithm in recognizing the objects from these incomplete representations is evaluated. The main reason behind choosing such an ideal situation is the ratio-

nal logic that in order to perform well in a real world scenario (natural images) any recognition system should first perform well in such idealized (simple) situations. In his study (Gollin 1960) Gollin also worked in a similar idealized situation where it was assumed that the subjects are familiar with the complete representations of the objects to be recognized. To be precise, if any of the subjects could not recognize the complete representations of any of the objects then his data were discarded from the study.

We study the robustness of contour based shape recognition methods by comparing an object represented by incomplete contours with all objects in a reference set represented by complete contours and determining the nearest neighbor. If the nearest neighbor is the object from which the incomplete contour representation is derived we consider the recognition to be correct, otherwise incorrect. We also propose possible extensions and generalizations of this basic framework.

In addition to Gollin's method of segment-wise contour deletion (like set I to set IV of Figure 2.2), we also consider other types of incompleteness, viz., occlusion and random pixel depletion. We name the corresponding studies segment-wise deletion test, occlusion test, and depletion test. Collectively we call these tests in short *Incomplete Contour Representations (ICR) tests*.

The choice of the shape recognition methods we study is limited by the condition that they use contour information. Unless necessary modifications are done, methods which use other type of information fall outside the scope of this study. For instance, Gavrila (Gavrila 1998) proposes a method based on the distance transform in which every point of a binary object is characterized by its distance to the object's border. In our study objects are represented by their contour points only and hence, the distance transform is not informative. Due to the same reason Goshtaby's shape matrix (Goshtaby 1985) cannot be directly assessed in this framework either. Some other methods which do not use boundary information are the medial axis transform approach described in (Davis 1986, Blum 1967, Peleg and Rosenfeld 1981), and the moment based approach dealt with in (Belkasim et al. 1991, Prokop and Reeves 1992). Latecki and Lakämper's polygonal shape descriptor (Latecki and Lakämper 2000) inherently assumes that an object is represented by a closed curve, and therefore this method first needs some modification before it can be applied to objects represented by incomplete contours. Mokhtarian and Mackworth's curvature scale space method (Mokhtarian and Mackworth 1992), which plays an important role in the MPEG-7 standard, computes the curvature at every point of a closed curve (at different scales) to represent the shape of an object. Hence, this method also needs some modification (e.g., estimation of the missing portions of the contours) in order to be evaluated in the ICR test framework. For further aspects of and references to shape analysis and object recogni-

(a) (b) (c)

Figure 2.3: *(a) Image of a rhinoceros in its natural habitat. (b) Result of edge detection with a bank of Gabor energy filters. (c) Result of contour detection by a bank of Gabor energy filters augmented with a biologically motivated surround suppression of texture edges. The contours of the object of interest are better visible in the latter image but the suppression of texture edges has resulted in a partial contour depletion. (Both algorithms are available as web applets at www.cs.rug.nl/∼petkov)*

tion methods see e.g., (Grigorescu and Petkov 2003, Loncaric 1998, Veltkamp and Hagedoorn 1999, Nagy 2000, Pavlidis 1980, Aloimonos 1988, Basri et al. 1998).

We study the *shape context* method described in (Belongie et al. 2002) and the *distance multiset* method described in (Grigorescu and Petkov 2003) with respect to their robustness to contour incompleteness of different types. In Section 2.2 we briefly describe these methods. In Section 2.3 we present the basic experimental design and the achieved results. We discuss some further aspects and possible test extensions in Section 2.4. A summary and conclusions are presented in Section 2.5.

2.2 Contour Based Shape Recognition Methods

In both methods studied below the recognition of objects is done by computing dissimilarity between the contour representations of two objects by using a point correspondence paradigm. The point correspondences are found using shape descriptors associated with the points.

2.2.1 Shape Context

A shape descriptor, called the *shape context* (Belongie et al. 2002), of a point p belonging to the contour of an object is a bi-variate histogram in a log-polar coordinate

system that gives the distribution of contour points in the surroundings of p. Let an object $\mathcal{O}$ be represented by a set of contour points, $\mathcal{O} \equiv \{p_1 \dots p_N\}$. Formally, the authors of this method define the shape context of a point $p \in \mathcal{O}$ as a vector in the following way:

$$H_K^{\mathcal{O}}(p) = (h_1(p), h_2(p), \dots, h_K(p)) \tag{2.1}$$

where

$$h_k(p) = card\{q \neq p | q \in \mathcal{O}, (q-p) \in bin(k)\} \tag{2.2}$$

is the number of contour points in the k^{th} bin $bin(k)$ and K is the total number of histogram bins. The bins are constructed by dividing the image plane into K partitions (in a log-polar coordinate system) with p as the origin. In this study we use 5 intervals for the log distance r, and 12 intervals for the polar angle θ, so $K = 60$. As radius of the surroundings on which the shape context is computed (the upper bound of the radial distance r) we choose the diagonal of the image. In this way this radius is constant for all experiments. As suggested in (Belongie et al. 2002), we randomly choose 100 points (if available) from the contour of an object and calculate their shape contexts. The shape of the object is described using the set of shape contexts associated with the contour points in the following way:

$$S_{\mathcal{O}}^{SC} \equiv \{H_K^{\mathcal{O}}(p) | p \in \mathcal{O}\}. \tag{2.3}$$

The cost of matching a point p_i that belongs to the contour of an object $\mathcal{O}_1$ of M points, to a point q_j from the contour of an object $\mathcal{O}_2$ of N points is defined as follows:

$$c_{i,j}^{SC} \equiv \frac{1}{2} \sum_{k=1}^{K} \frac{[h_k(p_i) - h_k(q_j)]^2}{h_k(p_i) + h_k(q_j)} \tag{2.4}$$

An $M \times N$ cost matrix of point-wise dissimilarities is constructed according to relation (2.4). Next we compute the dissimilarity between the shapes $S_{\mathcal{O}_1}^{SC}$ and $S_{\mathcal{O}_2}^{SC}$ of the objects in the following way:

$$d^{SC}(S_{\mathcal{O}_1}^{SC}, S_{\mathcal{O}_2}^{SC}) \equiv \sum_{i=1}^{M} min\{c_{i,j}^{SC} | j = 1, \dots, N\}. \tag{2.5}$$

The authors of the shape context approach (Belongie et al. 2002) (and also the authors of the distance multiset approach (Grigorescu and Petkov 2003)) use a different method to compute the dissimilarity of two shapes from the point-wise dissimilarity matrix. More specifically they use the Hungarian algorithm (Papadimitroiou and Stieglitz 1982) of bipartite graph matching to solve the optimal assignment problem. In our experiments we found that the simple method according to relation (2.5)

gives sufficient[2] results. Further aspects of the shape context method as presented in (Belongie et al. 2002), such as a thin plate spline transform and certain sampling considerations are not deployed here for simplicity and in order to make the two algorithms used comparable regarding the number and complexity of processing steps.

2.2.2 Distance Multiset

For a point p in the contour of an object $\mathcal{O}$ of N points, the *distance multiset*[3] is formally defined as follows (Grigorescu and Petkov 2003):

$$D_N^{\mathcal{O}}(p) = \{\ln(d_1(p)), \ln(d_2(p)), \ldots, \ln(d_{N-1}(p))\} \tag{2.6}$$

where $d_j(p)$ is the Euclidean distance between p and its j^{th} nearest neighbor in $\mathcal{O}$ and $\ln$ denotes the natural logarithm. In this approach the shape of an object $\mathcal{O}$ $\equiv \{p_1 \ldots p_N\}$ defined by a set of contour points is described by the set of distance multisets in the following way:

$$S_{\mathcal{O}}^{DM} \equiv \{D_N^{\mathcal{O}}(p) | p \in \mathcal{O}\}. \tag{2.7}$$

Next, a cost $c(X, Y)$ of matching two distance multisets X and Y is defined, see Appendix A.

Let $c_{i,j}^{DM}$ be the cost of matching a point p_i in an object $\mathcal{O}_1$ represented by M contour points to a point q_j in an object $\mathcal{O}_2$ represented by N contour points, $M \leq N$:

$$c_{i,j}^{DM} \equiv c(D_N^{\mathcal{O}_1}(p_i), D_M^{\mathcal{O}_2}(q_j)) \tag{2.8}$$

Similar to relation (2.5) the dissimilarity between the shapes $S_{\mathcal{O}_1}^{DM}$ and $S_{\mathcal{O}_2}^{DM}$ is defined as follows:

$$d^{DM}(S_{\mathcal{O}_1}^{DM}, S_{\mathcal{O}_2}^{DM}) \equiv \sum_{i=1}^{M} min\{c_{i,j}^{DM} | j = 1 \ldots N\}. \tag{2.9}$$

Further aspects of the distance multiset method as presented in (Grigorescu and Petkov 2003), such as the use of multiple features in a data structure called the labelled distance set are not deployed here for the reasons mentioned at the end of Section 2.2.1.

[2] By sufficient we mean that the recognition rate is good enough to illustrate the conceptual aspects of the ICR test framework. We use the same simple method to compute the dissimilarity of two shapes for both algorithms that are studied.

[3] In (Grigorescu and Petkov 2003) the term distance *set* is used which is not always correct, since this data structure might contain repeating elements, and should therefore be called *multiset* or *bag*.

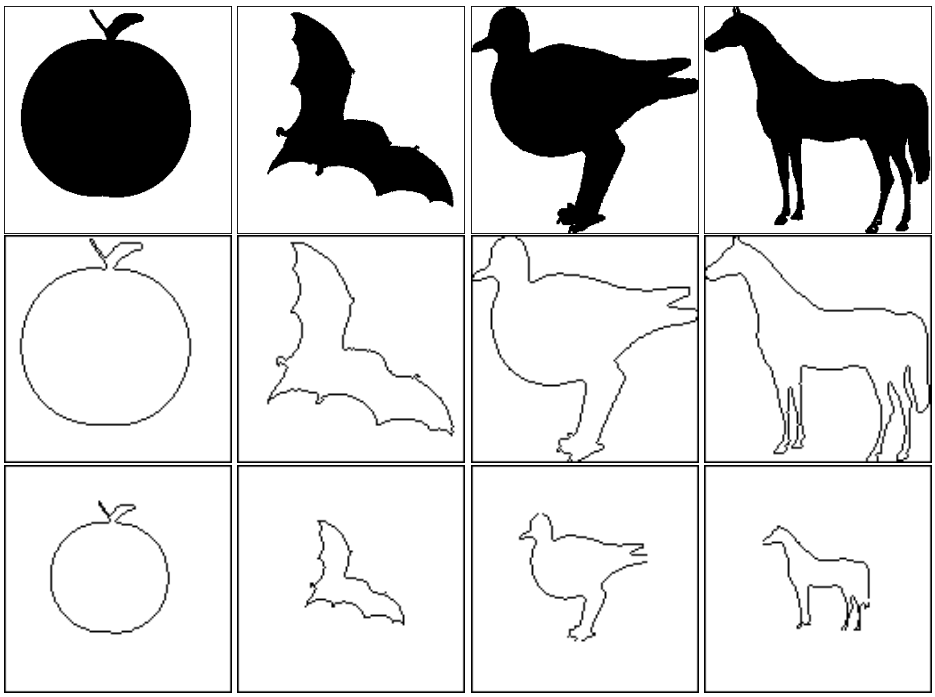


Figure 2.4: *Row 1: Sample of the MPEG-7 dataset (Latecki et al. 1998) images used. Row 2: Contour images extracted from the images in row 1. Row 3: Rescaled contour images; these images are considered as complete representations that comprise the memory of the recognition system.*

2.3 Experiments and Results

2.3.1 Dataset

As a data set we choose images from the MPEG-7 dataset (Latecki et al. 1998). It contains 1400 images divided in 70 classes, each of 20 similar objects (e.g., apple, bird, bat, etc). We choose one object from each class (Figure 2.4, row 1) and extract the contours of the object using Gabor filters (Grigorescu et al. 2003) (Figure 2.4, row 2). The resulting 70 contour images are rescaled in such a way that the diameter (maximum Euclidean distance between two contour pixels) is approximately the same (76 pixels) for all objects, c.f., row 3 of Figure 2.4. These 70 rescaled contour images are used as reference images in our experiments. The set of these images corresponds to the complete representations, set V of Figure 2.2, used in Gollin's original study and form the "memory" of the recognition system.

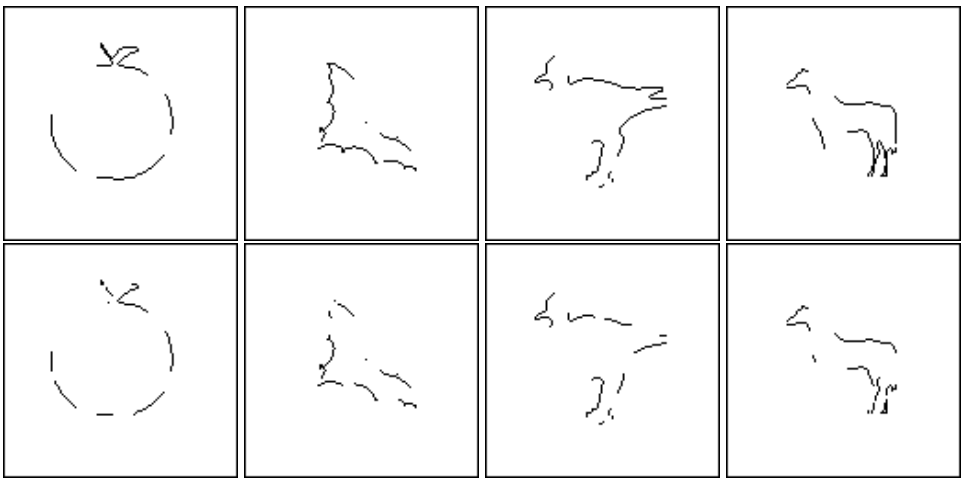


Figure 2.5: *Segment-wise deleted contour representations of objects (they correspond to the incomplete representations of Gollin's original study, set I to IV of Figure 2.2). The images are obtained by retaining 70% (row 1) and 50% (row 2) of the contour pixels (compare with row 3 of Figure 2.4).*

For the *segment-wise deletion test* incomplete representations (Figure 2.5) are constructed by randomly removing continuous segments of the contours and retaining a given percentage of contour pixels from the above mentioned complete contour representations.

For the *occlusion test* incomplete representations are created by removing a given percentage of consecutive contour pixels starting from the leftmost (row 1 of Figure 2.6) or the rightmost pixel (row 2 of Figure 2.6) of an object. The choice of left and right occlusion in our study is driven by the fact that in case of natural images the object of interest is most commonly occluded either from the left or from the right.

For the *depletion test* the incomplete representations (Figure 2.7) are obtained by randomly removing a given percentage of pixels from the contours of the complete contour representations.

In our experiments the percentages of retained pixels are chosen in the following way: from 2% to 4% in steps of 1%, from 5% to 85% in steps of 5%, and 100% for the depletion test; from 5% to 85% in steps of 5%, and 100% for the segment-wise deletion and the occlusion tests. For each type (segment-wise deletion, occlusion and depletion) and degree of contour degradation we create 70 test images from the corresponding reference images. All complete contour images and incomplete contour images obtained with different types and percentages of incompleteness are available in the web-site *www.cs.rug.nl/∼petkov*.

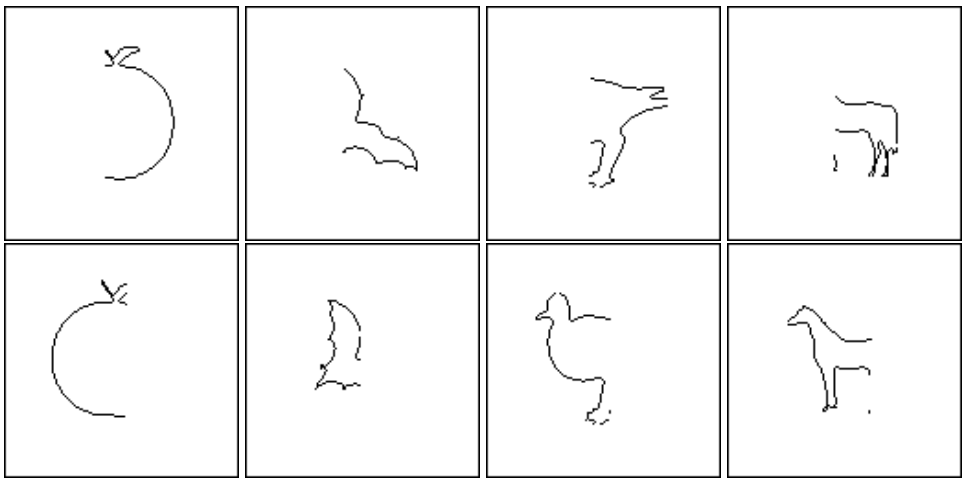


Figure 2.6: *Occluded contour representations of objects. The images are obtained by removing 40% of the contour from left (row 1) and right (row 2).*

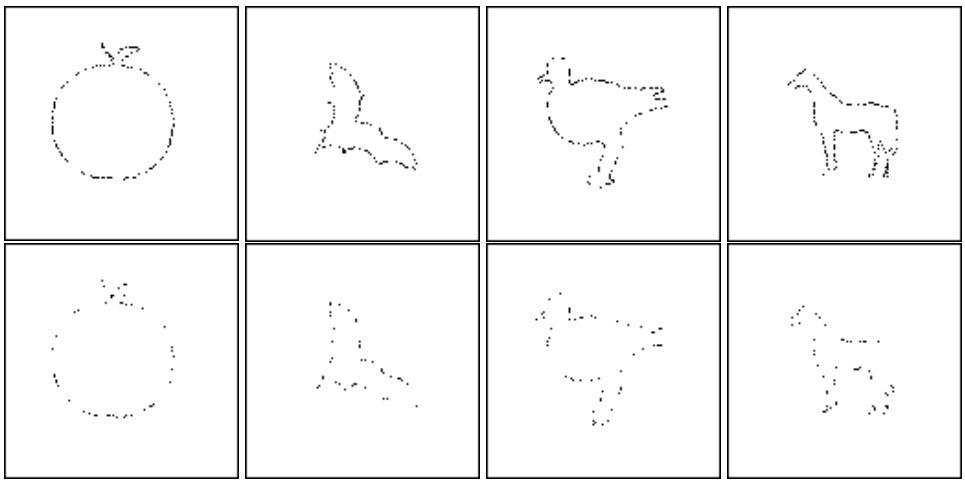


Figure 2.7: *Depleted contour representations of objects which are constructed by randomly removing 50% (row 1) and 80% (row 2) of the contour pixels.*

2.3.2 Method

An incomplete representation (segment-wise deleted or depleted or occluded contour image) obtained from one of the 70 reference images is compared with all 70 reference images and a decision is taken about which reference image the degraded image is most similar to (nearest neighbor search). The comparison is based on a

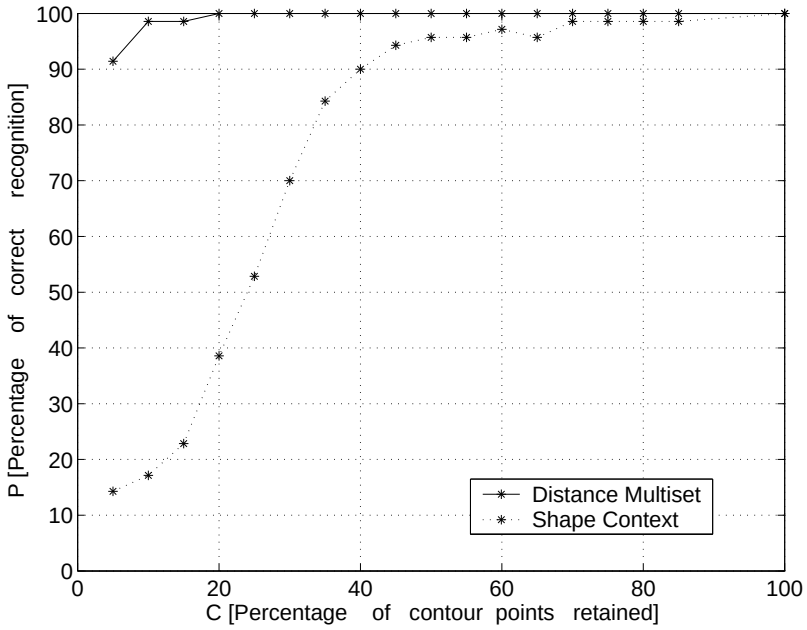


Figure 2.8: *Segment-wise deletion test: The distance multiset method gives 100% recognition rate when more than 20% of the contour is retained. The shape context method performs reasonably well (more than 90% recognition rate) when more than 40% of the contour points are retained. The results are for the MPEG-7 dataset.*

shape dissimilarity computed using a given shape comparison algorithm, described in Section 2.2. If the nearest neighbor is the reference image from which the degraded image was obtained, the recognition is considered correct, otherwise incorrect. If the nearest neighbor is found to be not unique then the recognition is also considered incorrect. For each of the three tests (segment-wise deletion, occlusion, depletion) and for each degree of contour image degradation, the corresponding 70 test images are compared with each of the 70 reference images and the percentage of correct recognition is determined. The percentage of correct recognition P is observed as a function $P(c)$ of the percentage c of retained contour pixels. In the case of occlusion test the percentage of correct recognition is calculated by averaging the correct recognition rates with left and right occluded images for a given percentage of retained contour.

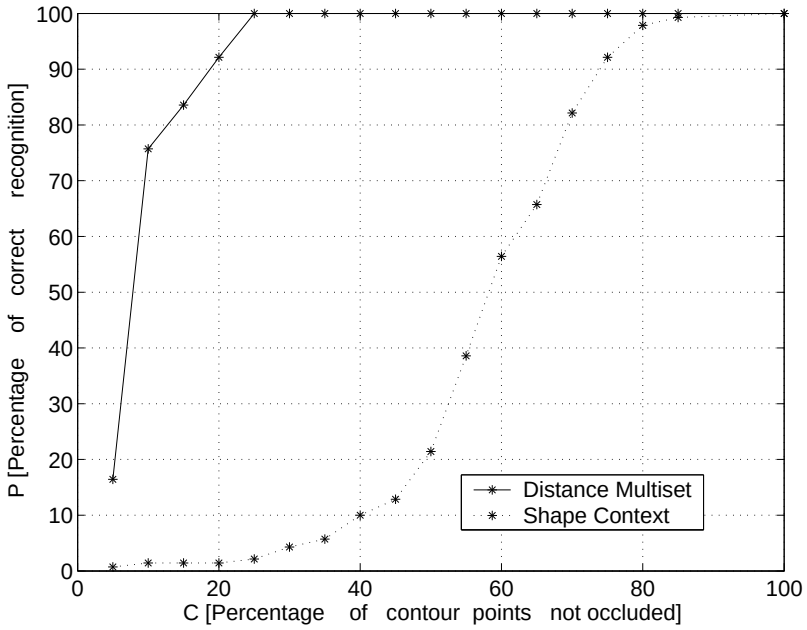


Figure 2.9: *Occlusion test: The shape context method is particularly affected by occlusion because the shape context descriptors of contour points near the occlusion boundary are radically different from the shape contexts of the points in the complete contours. The results are for the MPEG-7 dataset.*

2.3.3 Results

Figure 2.8, 2.9, 2.10 show the results of our experiments. In all three tests and for both shape comparison algorithms, the recognition rate is a monotonously increasing function of the percentage of contour retainment. In this respect the considered algorithms resemble the human visual system (Chihman et al. 2004), (Shelepin et al. 2004), (Foreman and Hemmings 1987). Both methods perform worst in the occlusion test and best in the depletion test, which also conforms with the recognition performance of humans, as occluded contour images carry the least amount of shape information and depleted contour images carry maximum shape information in the context of human visual perception (c.f., Figure 2.11).

In the case of the segment-wise deletion test (Figure 2.8) and the occlusion test (Figure 2.9), the performance of the distance multiset method is appreciably better than that of the shape context method for any percentage of retained contour pixels. From the results of the depletion test (Figure 2.10) we see that both the shape context method and the distance multiset method perform very well in recognizing objects

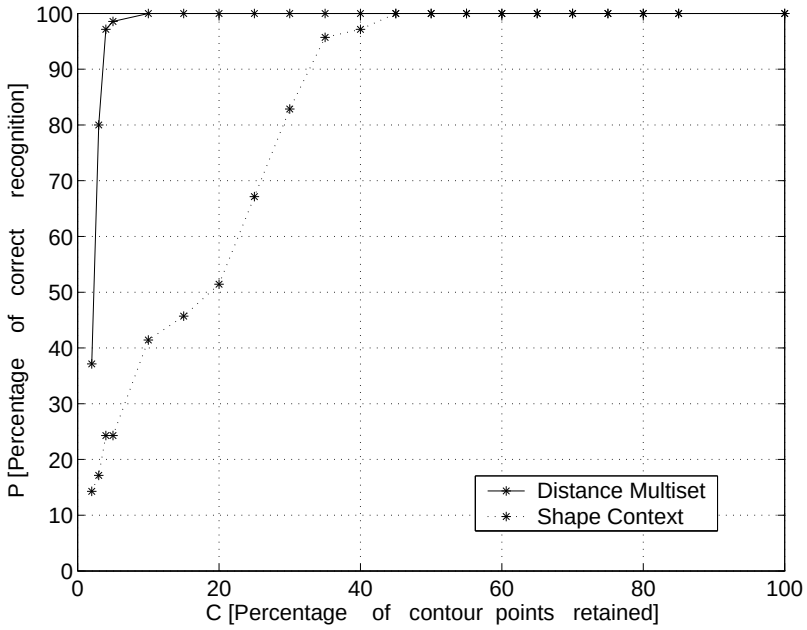


Figure 2.10: *Depletion test: The distance multiset method and the shape context method perform very well for more than 5% and 40% retainment of contour pixels. The results are for the MPEG-7 dataset.*

Figure 2.11: *Though amounts of incompleteness are same (80%) for three types of incomplete representation yet the randomly depleted contour (left most) is most easily and the occluded contour (right most) is least easily recognizable by humans as a butterfly.*

with depleted contour representations, if more than 40% and 5%, respectively, of the contour points are retained. The distance multiset method outperforms the shape context method when the degree of depletion is very high, i.e., a very low percentage (less than 40%) of the pixels are retained.

For both methods the results of the occlusion test are worse than the results in other tests. This is more evident for the shape context method and can be explained as follows: a contour point near the occlusion boundary has radically different shape context from the same contour point in the reference (unoccluded) object contour, since all the contour points on one side of the occlusion boundary are missing. Hence, such a point will have a large contribution to the dissimilarity between occluded and unoccluded contours.

In general, the better performance of the distance multiset method can be explained by the fact that the proposed ICR tests give advantage to algorithms which yield zero dissimilarity in a comparison of two objects represented by two sets of points where one is a subset of the other. This property of the distance multiset algorithm is explained in more detail below. (Another method with the same property is based on the non-symmetric Hausdorff distance.) Let us consider two sets $A, B, \subset \mathbf{R}^2$ such that

$$B = \{f(\mathbf{x}) : \mathbf{x} \in A\}, \tag{2.10}$$

where $f : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ is defined as follows:

$$f(\mathbf{x}) = \mathbf{L}\mathbf{x} + \mathbf{t}, \forall \mathbf{x} \in \mathbf{R}^2, \tag{2.11}$$

$\mathbf{L}$ being a 2×2 orthogonal matrix ($|det(\mathbf{L})| = 1$) and $\mathbf{t} \in \mathbf{R}^2$. Note that the transformation described by (2.11) preserves the Euclidean distance between points (isometry). The following special forms of f are of particular interest:

(a) $\mathbf{L} = I$ (the identity matrix) and $\mathbf{t} = \mathbf{0}$: identity transformation, $B = A$.

(b) $\mathbf{L} = I$ and $\mathbf{t} \neq \mathbf{0}$: pure translation, B is a translated version of A.

(c) $det(\mathbf{L}) = 1$, $\mathbf{t} = \mathbf{0}$: pure rotation, B is a rotated version of A.

(d) $det(\mathbf{L}) = -1$, $\mathbf{t} = \mathbf{0}$: pure reflection, the elements of B are obtained by reflecting elements of A across a straight line.

So if A is the set of contour points of an object $\mathcal{O}_1$ then B is the set of contour points of an object $\mathcal{O}_2$ that is derived from $\mathcal{O}_1$ through any of the transformations described in (a) through (d) or any combination there of.

Lemma

Let B be obtained from A according to (2.10 - 2.11) and let C be a subset of B,

$$C \subset B, \text{ card}(C) \geq 2. \tag{2.12}$$

It holds

$$d^{DM}(S_C^{DM}, S_A^{DM}) = 0 \tag{2.13}$$

where S_C^{DM} and S_A^{DM} are the shapes, described by distance multisets, corresponding to C and A, respectively.

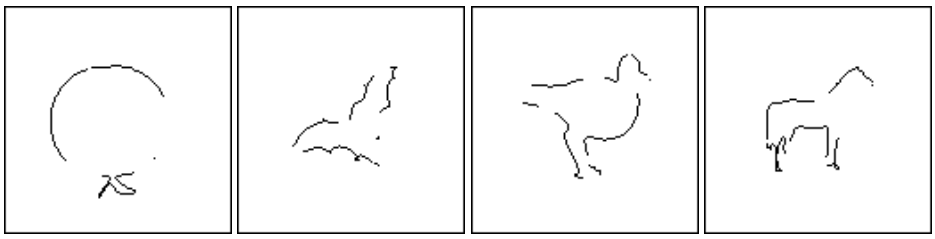


Figure 2.12: *Sample of incomplete contour representations of affine transformed objects used in a segment-wise deletion test (compare with rows of Figure 2.4).*

The proof of this lemma is included in Appendix B at the end of the chapter. In our study A corresponds to the set of contour points of a reference object, f is the identity transformation (i.e., $B = A$) and C is the set of contour points of an incomplete representation.

The implication of the lemma is two-fold:

(1) In the case of the distance multiset method, the recognition will be incorrect only when the nearest neighbor of a test object in the reference set is not unique.

(2) The distance multiset method should perform exactly the same way when f is not the identity transformation, that is, the incomplete representations of the objects are derived not directly from the reference objects but from affine transformed versions of them. This emphasizes the use of a distance multiset in our study instead of just matching pixels to calculate the dissimilarity between the shapes of objects.

To illustrate the latter implication (2) we consider the incomplete representations depicted in Figure 2.12. Here the incomplete representations are not directly derived from the objects depicted in row 3 of Figure 2.4 but from affine transformed versions of them. The affine transformations are chosen to be either rotation (by angle $n\frac{\pi}{2}$ to avoid discretization effects) or reflection across a line. We performed the segment-wise deletion test on the distance multiset method using the incomplete representations such as those shown in Figure 2.12. The results of this experiment are shown in Figure 2.13. If we compare the original segment-wise deletion test results for the distance multiset method (Figure 2.8) with these results we do not find any qualitative difference. A quantitative difference might arise due to the randomness involved in the construction of incomplete representations.

The above lemma does not hold for the shape context method, but this method can be modified in such a way that the relation (2.13) can be approximately fulfilled. Specifically, we normalize the shape context $H_K^{\mathcal{O}}(p)$ by dividing its elements by the total number of points $card(\mathcal{O})$ in the corresponding object $\mathcal{O}$. If $\mathcal{O}^{'} \subset \mathcal{O}$ is an incomplete representation derived from $\mathcal{O}$ and $H_K^{\mathcal{O}^{'}}(p)$ is the normalized (by $card(\mathcal{O}^{'})$) shape context of a point p ($p \in \mathcal{O}^{'}$) in this incomplete representation, the

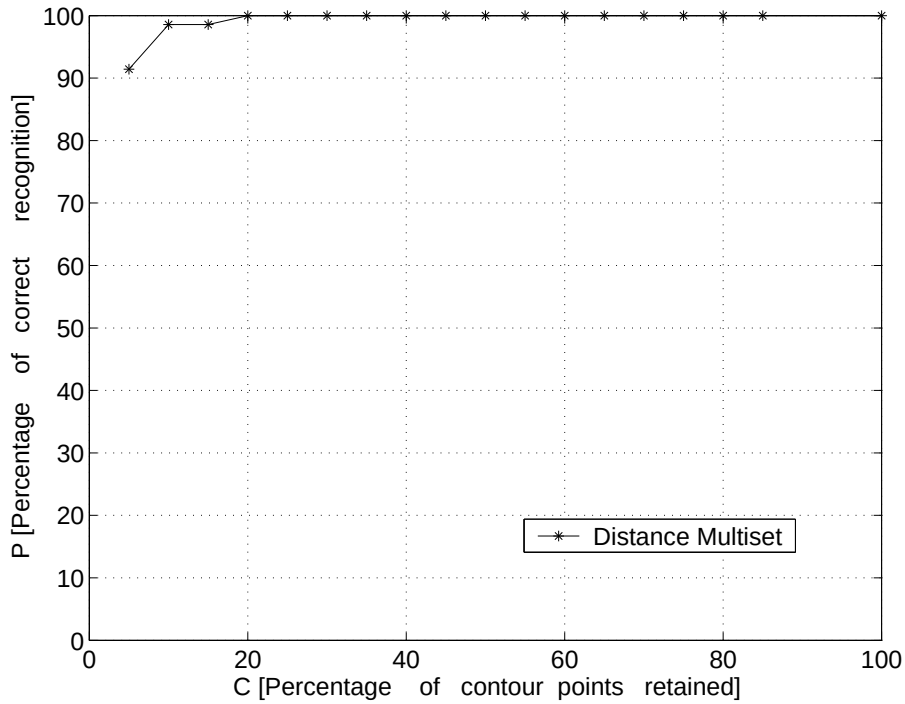
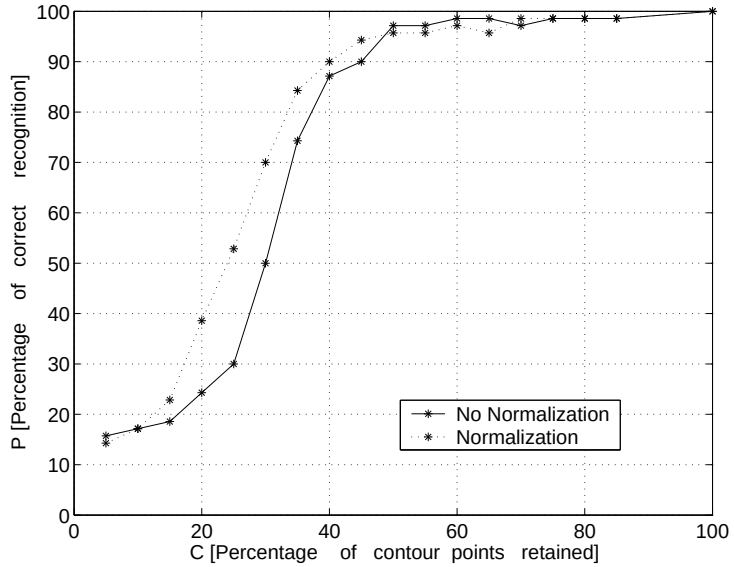


Figure 2.13: *Performance of the distance multiset algorithm in the segment-wise deletion test with incomplete representations obtained from affine transformed versions of the reference objects (Figure 2.12). Affine transformations have no substantial effect on the performance of the distance multiset method (compare with Figure 2.8).*

relation $H_K^{\mathcal{O}'}(p) \approx H_K^{\mathcal{O}}(p)$ will hold for modest degrees of contour deletion because the ratio of the number of contour points in each bin to the total number of points will be approximately the same for the complete and the incomplete contour representations. Hence, $d^{SC}(S_{\mathcal{O}}^{SC}, S_{\mathcal{O}'}^{SC}) \approx 0$ for the normalized shape contexts.

We performed experiments with and without the above mentioned normalization of the shape context, Figure 2.14, 2.15, 2.16. There is a significant performance improvement in the segment-wise deletion and the depletion tests due to the normalization procedure. This justifies the use of this procedure in the experiments whose results are shown in Figure 2.8, 2.9 and 2.10.



Segment-wise deletion ICR test.

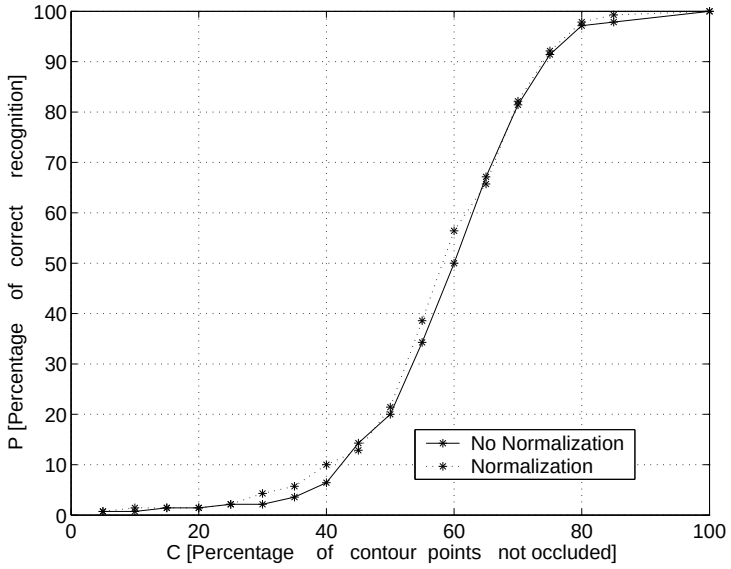
Figure 2.14: *The performance improvement of the shape context method due to normalization is significant in the segment-wise deletion test. The results are for the MPEG-7 dataset.*

2.4 Further Aspects

2.4.1 Choice of Dataset

The experiments presented in Section 2.3 above were carried out in the well known MPEG-7 dataset that is a de facto standard for comparison of shape recognition algorithms which use complete representations. As the scope of this study is to introduce a new test, it is important to check if the conclusions drawn from the ICR test are consistent across datasets. Furthermore, the MPEG-7 dataset has a certain restriction: the contours that define the objects are only outer contours. In contrast, the contour images used in the original psychophysical test of Gollin (Figure 2.2) as well as the contour images provided by computer algorithms (Figure 2.3) include inner edges next to the occluding boundaries of the objects.

For these reasons we carried out experiments in a second dataset, the Columbia University Image Library (COIL-20) dataset. It contains 1440 different images divided in 20 classes, each of 72 similar objects.



Occlusion ICR test.

Figure 2.15: *The performance improvement of the shape context method due to normalization is slight in the occlusion test. The results are for the MPEG-7 dataset.*

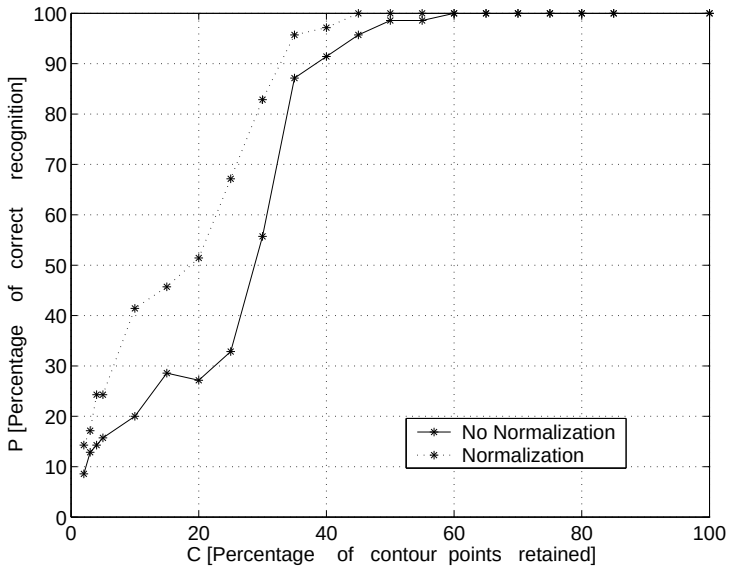
We choose one object from each class (Figure 2.17, row 1) and extract the contours. The resulting 20 contour images are rescaled to a diameter of 76 pixel units, c.f., row 2 of Figure 2.17. These 20 rescaled contour images are considered as the complete representations and are used as reference images in our experiments.

To prepare data for the ICR test we follow the same procedure as in the case of the MPEG-7 dataset, c.f., Figure 2.18, 2.19, 2.20.

We performed ICR test with these images and the results are shown in Figure 2.21. Comparing these results with the results in Figure 2.8 2.9 and 2.10 we see that there is no qualitative difference in the performances of the algorithms across the datasets. The positive effect of the aforementioned normalization of shape context is also observed with the COIL-20 dataset (Figure 2.22, 2.23, 2.24).

2.4.2 Combination of Variability Types

Object recognition methods can be evaluated for their robustness in various respects, viz., affine transformation, variation of shape, presence of noise etc. In this



Depletion ICR test.

Figure 2.16: *The performance improvement of the shape context method due to normalization is large in the depletion test. The results are for the MPEG-7 dataset.*

study we put forward a new attribute, namely robustness to incomplete contour representations. Since this is the focus of the research, the dataset used in the experiments includes only variability regarding contour incompleteness. This type of variability is not combined with other types, e.g., variation of shape or size. The rationale behind this decision is that first the algorithms should be characterized by their robustness to one type of variability at a time, e.g., shape changes, incompleteness of contours, rotation, size etc. Only after such characterization it makes sense to combine different types of variation, e.g., contour incompleteness with variation in shape or incompleteness with variation in size or any combination of these or other types of variation.

In this contest we note that good performance in the original ICR test does not guarantee good performance in other respects, e.g., robustness to shape or size variation. Hence, a good performance in the ICR test should be considered as a *necessary condition* for object recognition methods to perform well in a real world scenario but not as a sufficient one. We are not aware of any evaluation procedure for shape recognition methods which is sufficient in such respect. Once an algorithm is tested

Figure 2.17: *Row 1: Sample of the COIL-20 database images used. Row 2: Rescaled contour images that are considered as complete representations.*

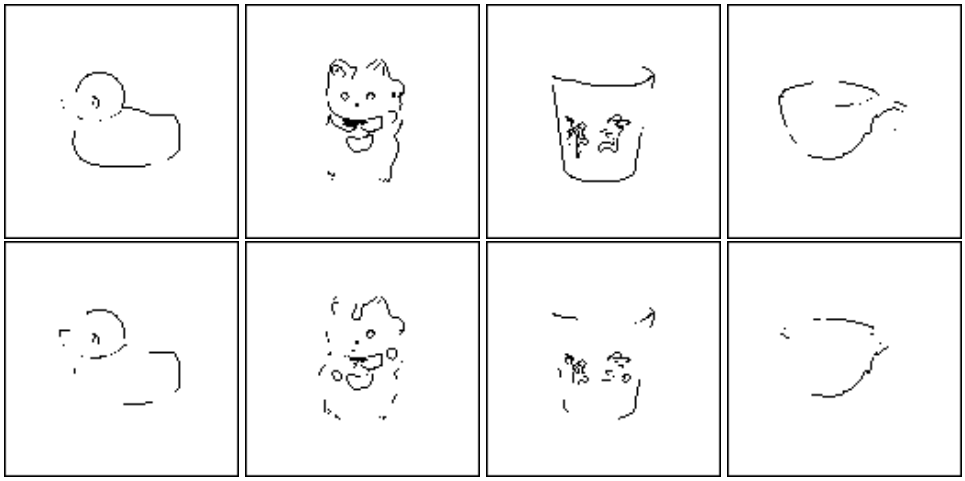


Figure 2.18: *COIL-20 Dataset: Segment-wise deleted contour representations of objects (they correspond to the incomplete representations of Gollin's original study, set I to IV of Figure 2.2). The images are obtained by retaining 70% (row 1) and 50% (row 2) of the contour pixels (compare with row 2 of Figure 2.17).*

for its robustness to shape variation (e.g., by the MPEG-7 bull's eye test (Latecki et al. 1998), (Grigorescu and Petkov 2003)), incompleteness (e.g., by the proposed ICR test) and other types of simple variation, more elaborate tests that combine different types of variation can be applied. Bellow we give an example.

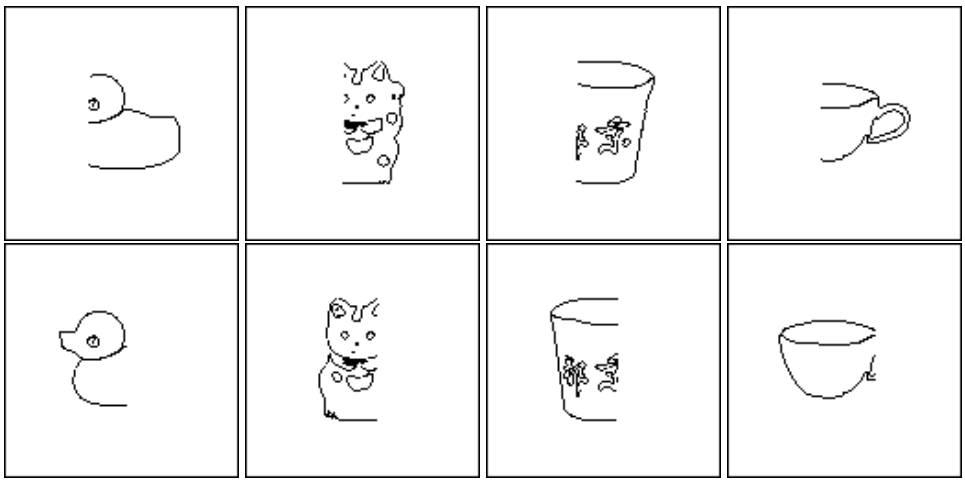


Figure 2.19: *COIL-20 Dataset: Occluded contour representations of objects. The images are obtained by removing 30% of the contour from left (row 1) and right (row 2).*

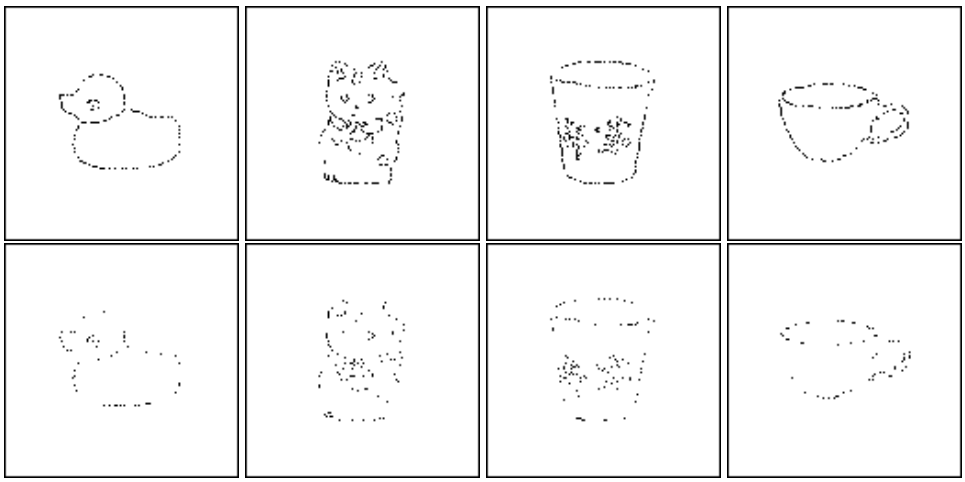
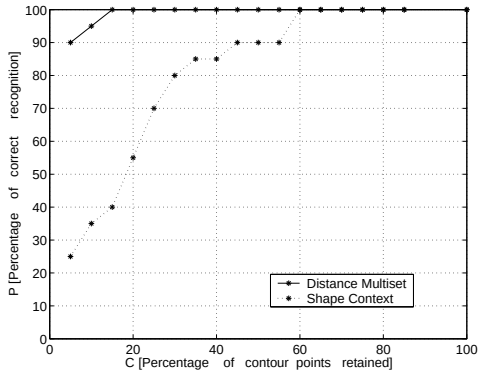
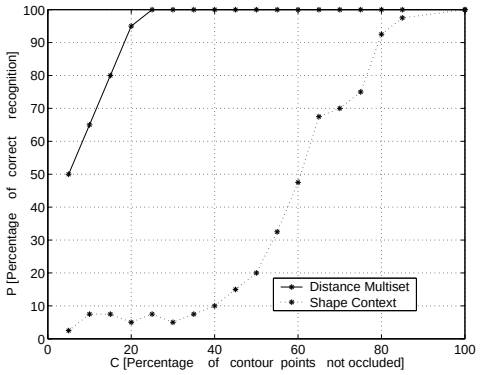


Figure 2.20: *COIL-20 Dataset: Depleted contour representations of objects which are constructed by randomly removing 50% (row 1) and 80% (row 2) of the contour pixels.*

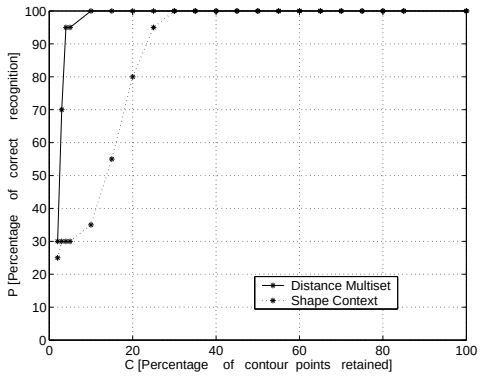
Shape variability and incompleteness: To assess robustness to incomplete representations along with robustness to variation in shape we propose a *bull's eye* ICR test using the MPEG-7 dataset. Unlike the basic ICR test proposed above, where we choose one complete contour image from each class of objects, in the bull's eye ICR test we choose more than one complete contour images, say n (out of 20 available in the MPEG-7 dataset), to represent a class. For a given percentage of contour retain-



(a) Segment-wise deletion ICR test (COIL-20)



(b) Occlusion ICR test (COIL-20)



(c) Depletion ICR test (COIL-20)

Figure 2.21: *Results of the ICR tests with a subset of the COIL-20 dataset. There is no qualitative difference in performance of the algorithms in the COIL-20 dataset compared to the MPEG-7 dataset (Figure 2.8 2.9 and 2.10).*

ment c, each of the $70 \times n$ incomplete contour images constructed by the methods described in Section 2.3 is compared with all $70 \times n$ reference (complete contour) objects and the corresponding n nearest neighbors are found[4]. Let ϵ_i ($\epsilon_i \leq n$) be the number of nearest neighbors that belong to the class of object i. The percentage of

[4]In the original bull's eye test (Latecki et al. 1998) a fixed number of 40 nearest neighbors is used. Here we use a modification of the test in which the number of nearest neighbors is taken to be equal to the number n of objects in one class.

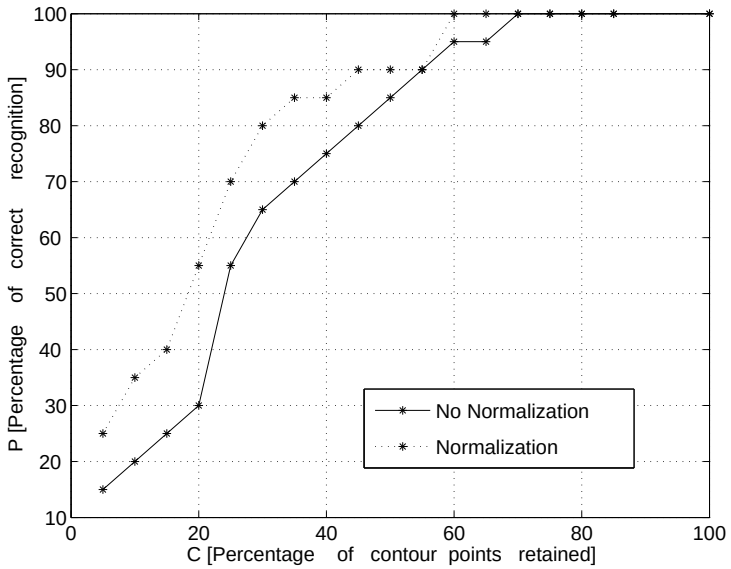


Figure 2.22: *Results of the segment-wise deletion test applied to the shape context method with and without the normalization procedure described in the text of the chapter with the COIL-20 dataset. The improvement in the performance of the shape context method due the normalization procedure is significant in the segment-wise deletion test.*

correct recognition $P(c)$ is calculated in the following way:

$$P(c) = \frac{\sum_{i=1}^{70\times n} \epsilon_i}{70 \times n^2} \times 100. \tag{2.14}$$

In Figure 2.25 we present the results of a bull's eye depletion test with $n = 2$. Increasing the value of n we can introduce more shape variation in this test procedure. An extreme case would be to consider the full MPEG-7 dataset ($n = 20$).

2.4.3 Object Size

The object size can have effect on the results of an ICR test through (a) the resolution of the reference objects and (b) a possible mismatch between the size of a reference object and a test object.

Regarding the resolution of the reference objects, in our experiments we found that for a given percentage of contour degradation (by any method) the performance of the algorithms grows with the diameter of the reference objects. This is in agreement with the results of psychophysical studies on humans where performance increases with the visual angle at which the objects are presented (Chihman

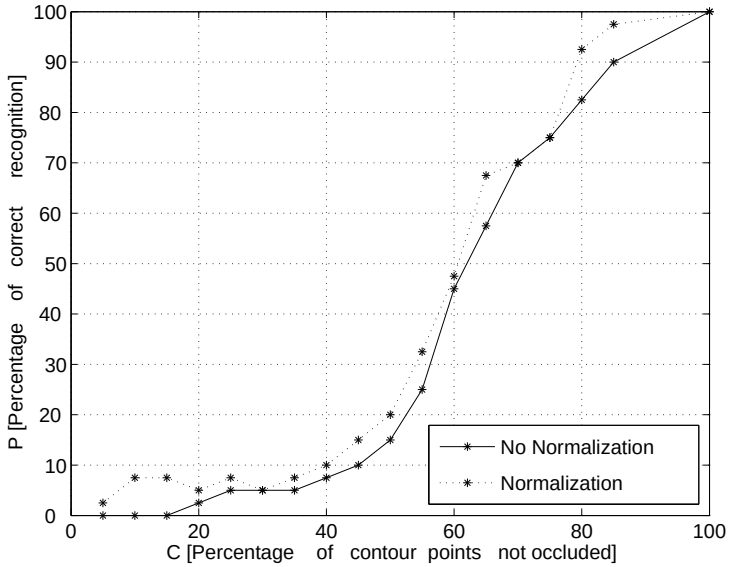


Figure 2.23: *Performance of the shape context method in the occlusion test with and without the normalization procedure with the COIL-20 dataset. In the occlusion test the performance of the shape context method is slightly improved due to the normalization procedure.*

et al. 2004). In Figure 2.26 we illustrate the performance of the shape context method in the depletion test for two different sizes of the reference objects. From this figure we see that for a low percentage (less than 45%) of retained contour the performance of the method is appreciably better in the case when the reference object size is bigger. To eliminate this effect and to standardize the test procedure we rescaled the reference contour images to a fixed diameter (76 pixel units).

The problem of a possible mismatch between the sizes of reference and test objects is not specific for the type of test objects (incomplete contour representations) that we use in this study. The problem is rather related to the way in which shape recognition algorithms deal or do not deal with size variation. Both algorithms (shape context and distance multiset) used here to illustrate the ICR test are not intrinsically scale invariant. In (Belongie et al. 2002) the authors of the shape context method suggest normalization of all radial distances by the mean distance between all point pairs in a contour in order to make the shape context descriptor scale invariant. To achieve the same goal, the authors of the distance multisets method (Grigorescu and Petkov 2003) prescribe dividing all distances by the diameter of the object under consideration.

In the ICR test the reference contour images are rescaled to have the same di-

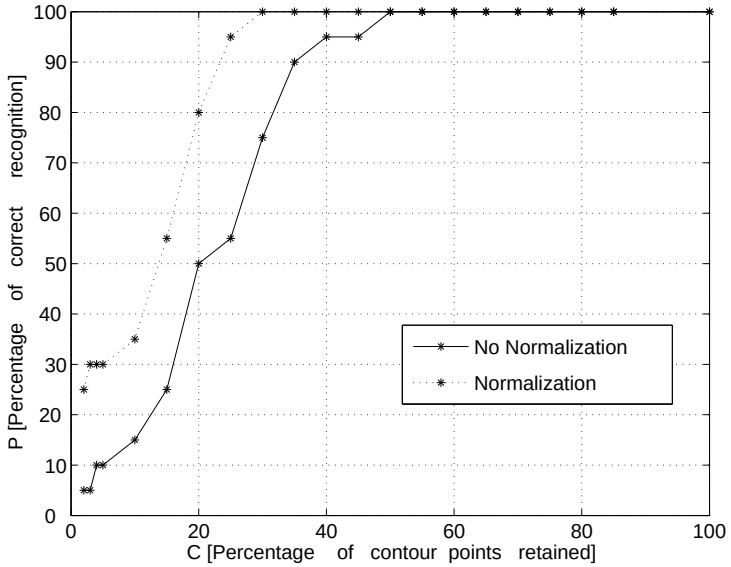


Figure 2.24: *Results of the depletion test in the case of the shape context method with and without the normalization with the COIL-20 dataset. A significant performance improvement of the shape context method is observed in the depletion test due to the normalization procedure.*

ameter (76 pixel units) and the incomplete representations are constructed from these rescaled versions of the reference contour images by removing contour points. Hence, the actual distances between retained points do not change and the algorithms are provided with test images having the same point-to-point distance as the reference images. The distance multiset algorithm benefits slightly more from this aspect of the construction of incomplete representations because the distance multiset of a point from an incomplete contour is a subset of the distance multiset of the same point in the corresponding complete contour. As pointed out above, the dissimilarity computed for pairs of such points will be zero for the distance multiset algorithm. Zero dissimilarity cannot be guaranteed for the shape context method.

The above choice of a procedure for the construction of incomplete representations was made deliberately: we want to quantify the success or failure of an algorithm in coping with incompleteness and for this purpose we want to minimize the effects of other aspects, e.g., size variation. In a real world situation such as the one illustrated by Figure 2.3, however, there is no guarantee that the size of a test object with an incomplete contour will match exactly the size of the corresponding reference object. Under such circumstances it is important that a method can deter-

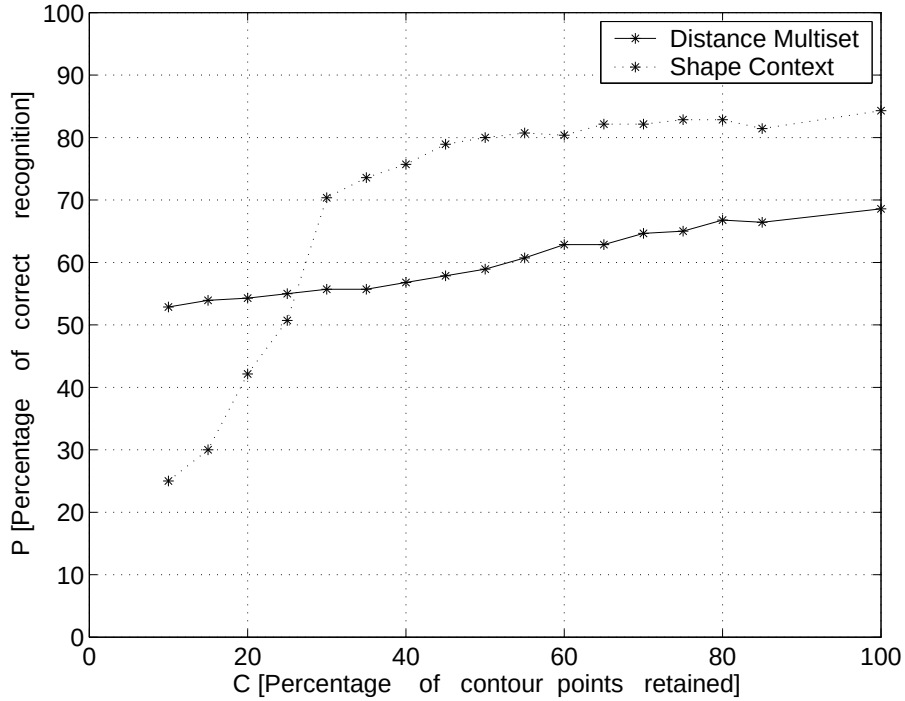


Figure 2.25: *Results of a bull's eye depletion test, using a subset of* 70×2 *images of the MPEG-7 dataset. For high percentage of pixels retained the shape context method performs better than the distance multiset method.*

mine automatically the appropriate size. Our experiments can easily be modified for this purpose by rescaling all test images (incomplete contour representations) to the same object diameter (of 76 pixel units) as the reference objects. Figure 2.27 shows the results of ICR tests with incomplete representations that have been obtained in this way. A comparison with Figure 2.8, 2.9 and 2.10 shows that performance degrades due to the fact that the sizes of incomplete representations do not exactly match the sizes of the corresponding complete representations. When an incomplete representation is constructed from a complete representation, one of the points for which the diameter of the object is measured can be removed. The consequence is that the diameter of the incomplete representation will decrease. After rescaling this diameter to the standard size (of 76 pixel units) all pair-wise distances in an incomplete representation will increase and become larger than their counterpart distances in the corresponding complete representation. The smaller the percentage of contour retainment, the larger this effect and the larger the performance degra-

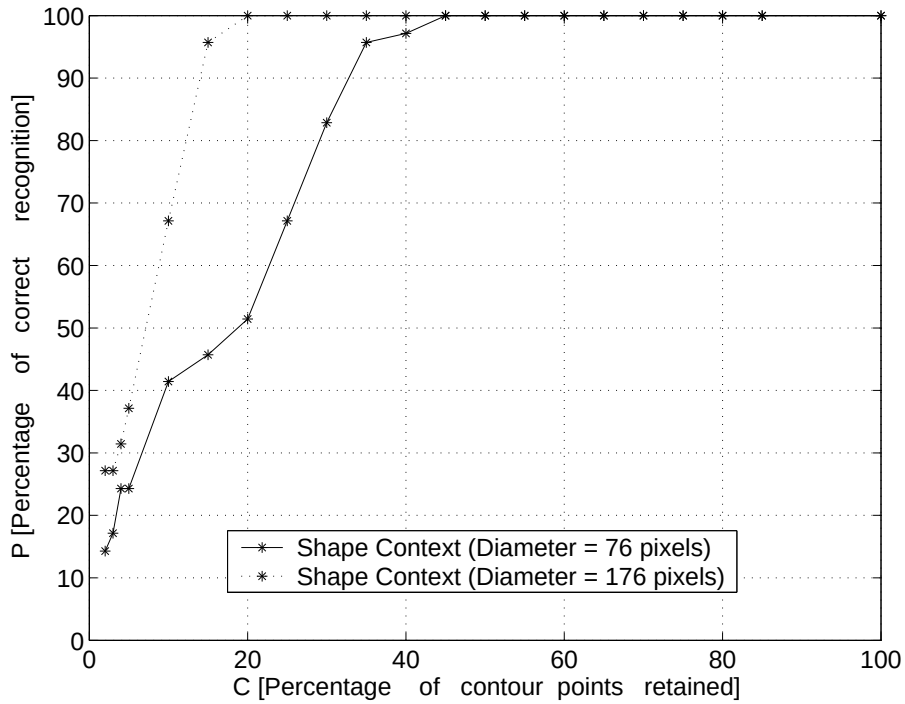
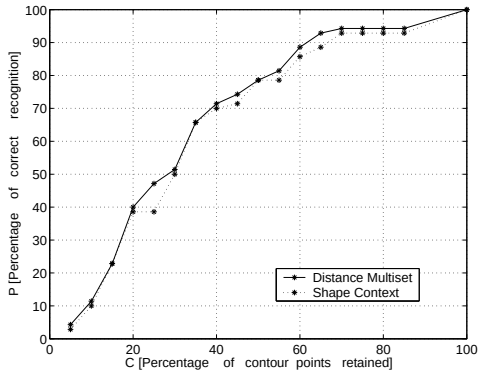


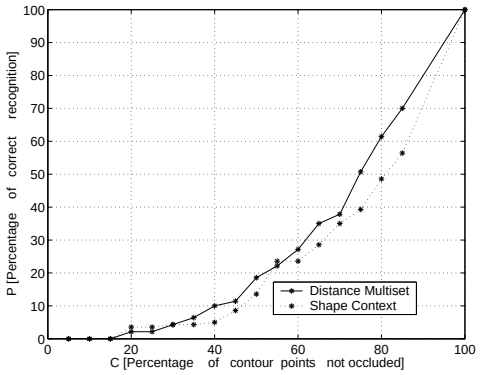
Figure 2.26: *The performance of the shape context method in the depletion test with two different sizes of the reference objects - the larger the diameter, the better is the performance. The results are for the MPEG-7 dataset.*

dation.

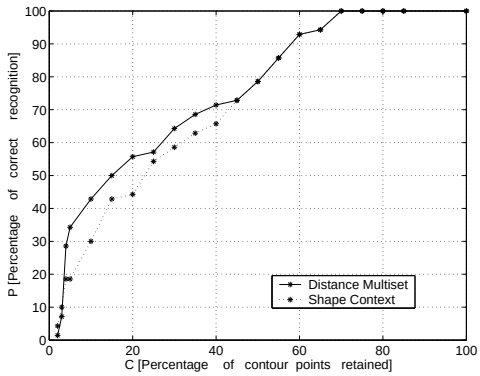
This method of rescaling has a devastating effect on the performance of the algorithms with incomplete representations obtained through occlusion (Figure 2.27(b)). For this case and also for the general case of objects that are not segmented from their background one should adopt a different, multiscale approach. In real world situations, such as the one illustrated by Figure 2.3, an object is not segmented from its background. In contrast, the very purpose of using a shape descriptor in such a situation is to test whether a given object is present in a complex scene and to separate it from the background. Under such circumstances and without any prior knowledge about the appropriate scale to be used, one can take a multiscale approach: shape descriptors are computed independently at multiple resolutions and the descriptors computed at each scale are compared with the reference descriptors. The multiscale approach has been advocated for both in biological (Koenderink and van Doorn 1978) and computer (Burt 1988) vision.



(a) Segment-wise deletion test



(b) Occlusion test.



(c) Depletion test.

Figure 2.27: *Results of the ICR tests with incomplete contour representations that are rescaled to a constant diameter of 76 pixel units. The performances of the algorithms are worse than in the case when the distances in the complete and incomplete representations are equal (compare with Figure 2.8 2.9 and 2.10). The effect is particularly strong for the distance multiset algorithm. The results are for the MPEG-7 dataset.*

2.4.4 Criterion for Acceptable Performance

The performance curves obtained in ICR tests can be used to compare algorithms as illustrated in Figure 2.8, 2.9, 2.10. It would be interesting to define a criterion for acceptable performance of an algorithm without having to compare it with another

algorithm. One possible way of achieving this is to use as a reference the performance of humans in a similar experimental setup. As a matter of fact similar studies exist in psychophysics (Chihman et al. 2004, Shelepin et al. 2004, Foreman and Hemmings 1987). For instance in (Chihman et al. 2004) the performance of humans in a test that is similar to the segment-wise deletion ICR test is studied whereby identical gaps between fragments of equal length are used. Similar to computer algorithms, the performance of humans depends on the size of the objects. To make a comparison possible, objects need, therefore, to be presented at a certain standard size (visual angle) that is related to the standard object size (in pixel units) deployed in computer algorithms. To establish such a relation one should use visual acuity data (minimum visual angle between two distinguishable points).

2.5 Summary and Conclusion

Object recognition methods that employ shape descriptors have been evaluated and compared using various characteristics like invariance, uniqueness and stability (Mokhtarian and Mackworth 1992). Marr and Nishihara (Marr and Nishihara 1978) proposed three criteria for judging the effectiveness of a shape descriptor, viz., accessibility, scope and uniqueness, stability and sensitivity. Brady (Brady 1983) put forward a set of criteria for representation of shape, viz., rich local support, smooth extension and propagation. A detailed survey and comparison of shape analysis techniques on the basis of some of the above mentioned criteria can be found in (Loncaric 1998). In the current work, motivated by characteristics of the human visual system (Gollin 1960), we propose an additional new criterion, viz., robustness to contour incompleteness to compare and characterize contour based shape recognition algorithms using their performance in recognizing objects with incomplete contours. We are not aware of any such comparison and characterization in the present literature.

We put forward the following procedure which we call the ICR test:

1. Take a set of images of objects and extract contours. Rescale all contour images to the same object diameter.
2. Train the recognition system with these complete contour representations.
3. Construct different sets of incomplete representations from the complete contour representations, quantifying the level of incompleteness using the percentage of contour pixels retained.
4. Using the incomplete representations as a test set, evaluate the recognition rate as a function of the percentage of contour pixels retained.

We distinguish three different types of incomplete contour representations according to the method used to remove parts of the contour: segment-wise deletion, occlusion and random pixel depletion. We created test datasets of such incomplete contour representations derived from images from the MPEG-7 and COIL-20 datasets and made them publicly available at *www.cs.rug.nl/∼petkov/*.

We illustrated the test framework with two shape recognition methods based on the shape context and the distance multiset. We should note that other shape recognition methods such as those based on Hausdorff measure (Huttenlocher et al. 1993, Huttenlocher et al. 1999), wavelet descriptors (Chuang and Kuo 1996), dynamic programming (Petrakis et al. 2002), graph matching (Cross and Hancock 1998), curve alignment (Sebastian et al. 2003), Fourier descriptors (Bartolini et al. 2005) can also be studied in this framework. As the main objective of the research presented in this chapter is to introduce a new test framework an exhaustive comparison of different methods under this framework is beyond the scope of this study. The two methods tested were chosen merely for illustrative purposes and we did not aim to prove superiority of any method. A complete comparative study of the two methods is out of the scope of this work. In our illustrative experiments we found that: (1) The distance multiset shape recognition method outperforms the shape context method regarding robustness to contour incompleteness, especially for high levels of incompleteness. (2) Both methods perform similar to the human visual system in the sense that their performances are increasing functions of the degree of contour completeness and are best in the case of the depletion test and worst in the case of the occlusion test.

Our main conclusions are as follows: The robustness of contour based shape recognition methods to incompleteness of contour representations is an important aspect of any contour based objects recognition system. The ICR test as defined and proposed in this study is an adequate framework for assessing the above mentioned performance and can be used as a standard test procedure for any contour based object recognition system/algorithm.

Appendix

A. Cost of Matching Two Distance Multisets

Consider the multisets

$$X = \{x_1, x_2, \dots x_M\}, \tag{2.15}$$

$$Y = \{y_1, y_2, \dots, y_N\}, \tag{2.16}$$

where $M \leq N$. Let π be a one-to-one mapping from the set $\{1, \dots, M\}$ to the set $\{1, \dots, N\}$ and let Π be the set of all such mappings. The mapping π defines an assignment of an unique element $y_{\pi(i)} \in Y$ to each element $x_i \in X$. The cost $c_\pi(X, Y)$

of a mapping/assignment $\pi \in \Pi$ is defined as follows:

$$c_\pi(X, Y) = \sum_{i=1}^{M} |x_i - y_{\pi(i)}|. \tag{2.17}$$

Let c be the minimum of the costs of all such possible mappings:

$$c(X, Y) = min\{c_\pi(X, Y) | \pi \in \Pi\} \tag{2.18}$$

Note that X and Y are sorted in ascending order by the definition of a distance multiset. To compute $c(X, Y)$ efficiently, we use the algorithm described in (Petkov 2003) which has complexity $O(M(N - M))$.

B. Proof of the Lemma

Claim 1 : The distance multisets of A and B are identical.

By definition f is an isometry, which implies that distance multisets of A and B are identical, that is, for every $p \in A$ $\exists$ a $q \in B$, $q = f(p)$ such that $D_N^A(p) = D_N^B(q)$, assuming that card(A) = card(B) = N.

Claim 2 : $d^{DM}(S_C^{DM}, S_B^{DM}) = 0$.

The definition of distance multiset along with relation (2.18) implies that for every $p_i \in C$, $\exists$ a $q_j \in B$, such that $c_{i,j}^{DM} = 0$. Hence the minimum of every row of the cost-matrix of point-wise dissimilarities is 0, which implies by relation (2.9), $d^{DM}(S_C^{DM}, S_B^{DM}) = 0$.

Claim 2 and the invariance of distance multisets in claim 1 imply that

$$d^{DM}(S_C^{DM}, S_A^{DM}) = 0 \ . \ \square$$