Multi-Scale Salience Distance Transforms

Paul L. Rosin & Geoff A.W. West
Cognitive Systems Group
Department of Computer Science, Curtin University of Technology
GPO U1987, Perth, 6000, Western Australia
email: {rosin,geoff}@cs.curtin.edu.au

Abstract

The distance transform has been proposed for use in computer vision for a number of applications such as matching and skeletonisation. This paper proposes two things: (1) a multi-scale distance transform to overcome the need to choose edge thresholds and scale and (2) the addition of various saliency factors such as edge strength, length and curvature to the basic distance transform to improve its effectiveness. Results are presented for applications of matching and snake fitting.

1 Introduction

Given an image containing a set of features (edgels, lines, points), the function of a distance transform (DT) is to assign to each pixel the distance to the nearest feature. One of its most common purposes is to evaluate hypothesised model poses by superimposing the model on the distance transform [1] which gives a measure of how well the model fits.

The majority of applications of distance transforms use the output of edge detection applied to an intensity image as the binary feature map from which to calculate distances. However, this leads to several problems involving the accuracy and reliability of the edge detection process itself. Most edge detectors produce mislocated edges due to various factors such as noise, smoothing, and the interaction of close edges, and spurious and missed edges (false positive and false negatives). Spurious edges will cause distances from pixels to "real" edges to be underestimated if the spurious edges are closer. Missing edges have the opposite effect and distances will be calculated to the nearest edge that is present, overestimating the distance if false positives are not considered.

Since the distance transform requires a binary feature map the edges must be thresholded. Choosing the threshold value involves a balance between rejecting false positives and false negatives. Too low a threshold produces excessive amounts of spurious edges, too high a threshold eliminates too many desirable edges. In addition to the edge strength, connectivity can be utilised. An old idea recently reported [2] is to threshold on the summed edge strength for each list of connected pixels. Although little attention has been paid to developing techniques which can automatically threshold an edge map, this and other techniques such as non-maximal suppression can produce perceptually good results. Automatic methods are obviously more desirable and can give good results [3]. However, the problem of defining what an edge is really depends on the context.

Even if a reasonable threshold value can be selected the results of the distance transform may still be inadequate for several reasons. First, no operator can be optimal over all scales. Therefore the output of the edge detector will not capture all the structures of interest in the image if they exist at different scales. Although several multi-scale edge detectors are available they invariably still require parameters relating to the scale. For instance, the edge focussing process [4] the coarsest scale must be specified. This determines which image features will be retained in the final output. Second, when attempting to evaluate a hypothesised model pose it is not simply the distance to the nearest edge that is relevant, but rather a combination of several image based factors. In general, given the task of matching two structures it is desirable to allow the sub-parts of each to have differing degrees of salience or importance. This extends for the case of distance transformations to allowing some image features to have more effect than others. This paper describes a technique for incorporating factors in addition to just the edge location into the distance transform. The transformed image will then represent the salience of each point rather than simply its distance from the nearest edge. The problem of which scale to use for the distance transform is discussed. A multi-scale salience distance transform is proposed that captures information at various scales without the need for thresholding. The usefulness of these techniques is discussed in the context of two applications namely template matching and the determination of the boundary of a surface of revolution using an active contour.

2 Adding Salience to the Distance Transform

When edges are detected in images many attributes may be extracted in addition to the edge location. Examples include the grey scale edge profile normal to the edge, orientation and magnitude. Connected edges which are linked to form curves have attributes such as length and curvature. These attributes may be used singly or in combination to calculate the salience of the edges. For instance, magnitude has regularly been used as an indication of edge salience. Others have used combinations of edge magnitude and length for the extraction of significant edges [5,2,3]. Recently, some work has been carried out in measuring the saliency of every point in an image based on edge length and curvature [6] and orientation and curvature [7].

Here we consider incorporating edge magnitude into the distance transform. At a fine scale most edge detectors produce a noisy edge map even in regions of supposedly constant grey level due to noise and quantisation effects which cause small variations in the grey level producing edges. The result of applying the chamfer 3-4 distance transform [1] is dependent on the edge map, and is therefore severely affected by the noise. The affect of the noise can be reduced by using a threshold on the edge data which is problematic. With no threshold, the boundary of a model superimposed almost anywhere in the image would have a low average distance. This can be alleviated to some extent by considering not only the distance to the nearest edge pixel but also the orientation of that pixel [9]. However, noisy or cluttered regions in the image are likely to give rise to edges at all orientations, and so the distance transform will still produce a considerable response. Since the edges produced by noise generally have low magnitudes, if we could weight the edges by their magnitudes then the effect of noise edges will be reduced.

We propose an algorithm based on the chamfer 3-4 algorithm [1] which provides a reasonable approximation to the Euclidean distance for a moderate computational cost. Computation of the distance transform can be either parallel or sequential. In this context parallel means that all the pixels in the image are processed at one iteration. Here we describe the iterative parallel version. It can be written thus:

$$d^i_{x,y} = \min_{a,b=-1\ldots 1}(d^{i-1}_{x+a,y+b} + w_{a,b})$$

where $d^i_{x,y}$ is the distance value at location (x, y) calculated at iteration i, and $w_{a,b}$ is the value of the mask at location (a, b) where the centre is $(0, 0)$. d^0 is the initial binary edge map.

The new algorithm generates a salience distance transform (SDT) and stores two images: the approximate Euclidean distance, and the edge magnitude. The edge magnitude relates to the edge that the distance has been measured from. At each iteration the neighbourhood around each pixel is examined. The local distances are calculated using the 3-4 mask. Each of the nine distances are weighted by their corresponding edge magnitudes to provide nine salience measures. The most salient (i.e. the minimum value) is selected and the corresponding distance and edge magnitude values at that location are stored in the two images. The process at pixel location (x, y) at iteration i can be written as:

$$a', b' = \arg\left(\min_{a,b=-1\ldots 1}\left(\frac{d^{i-1}_{x+a,y+b} + w_{a,b}}{m^{i-1}_{x+a,y+b}}\right)\right)$$

$$d^i_{x,y} = d^{i-1}_{x+a',y+b'} + w_{a',b'}$$

$$m^i_{x,y} = m^{i-1}_{x+a',y+b'}$$

and the final salience map at iteration f is calculated as:

$$\frac{d^f_{x,y}}{m^f_{x,y}}$$

A discussion of the performance of the salience distance transforn (SDT) is in the next section. It is straightforward to extend the above method to incorporate additional saliency factors such as curve length, shape, and image clutter. These additional salience measures further improve the distance transform [10].

3 Combining Distance Transforms Over Scale

There are a number of problems with the distance transform. The first is at which scale do you compute the distance transform? This is not surprising as the distance transform is heavily dependent on the edge detection stage. One solution for model based recognition is to use the same parameters for the edge detection in the two stages of (1) feature extraction for hypothesis generation and pose estimation, and (2) using the distance transform for matching and confirming/rejecting the hypothesis. This has disadvantages [2] in that the best scales for the two stages are not necessarily the same. In fact, a fine scale is good for feature extraction (ellipses, lines etc.) because of the requirement for accurate feature and hence pose estimation, whereas a coarse scale is good for the distance transform because it reduces the effect of clutter caused by spurious edges and texture. A good match may be obtained in a region of dense edges because it results in an area of low distance values. This problem is exacerbated by a poor estimate of the pose of the model. The hierarchical distance transform [11] has been proposed for iterative refinement of pose for model to image matches. This, to some extent overcomes the problem of choosing a suitable scale for the distance transform as a pyramid of images is used to successively refine the match. In fact, in this context the clutter is seen as having little effect whereas a missing feature can create problems [12].

The second problem is what threshold value do you use for the edge detector? Although a number of techniques exist for detecting edges at the required accuracy and reducing spurious edges (due to noise and 'unwanted' detail), the design of an optimal edge detector requires some idea of the context in which it is being used. The same argument holds for the distance transform. To overcome the problem of which threshold and which scale to use, two approaches are possible. The first is to generate the distance transform from edge maps at a number of scales and use them in some form of multi-scale object recognition scheme [13]. This would differ from that suggested by Borgefors [11] in that the distance transforms would be generated at the chosen scales for a number of edge images, not at the different levels of a pyramid generated from one fine scale edge map. The second approach considered here simply combines the distance transform images for the scales together to form one distance transform. The simplest way of doing this is to add the images together without taking into account the scale at which each distance transform was generated. This approach is predicated on the following ideas (1) without any context, all edges at all scales are equally important, and (2) an edge becomes more important if it is present over a number of scales.

Summing the DT images together causes similar distance values to be reinforced and dissimilar values to be weakened. The result will be a single DT image that has low distance values for edges that occur in a number of scales and increasing values for edges that only occur over some scales. Most clutter and texture will have high values of distance as these produce edges at a small number of scales which tend to be unstable. Summing edges over scale can be thought of as incorporating yet another saliency measure to pixels, namely scale lifetime. Although summing individual scales only provides an approximation to individual pixel lifetimes, it is more straightforward than applying techniques like the edge focussing procedure [4].

When considering the multi-scale approach there are a number of questions that have to be answered. These are: (1) what is the finest scale to use? (2) what is the coarsest scale to use? and (3) what are the intervals between scales to be combined? The values of σ used for the Marr-Hildreth edge detector [8] were 1, 1.4, 2, 2.8, 4, 5.7, 8, 11.3 and 16. These result in frequency bands at octave intervals. The minimum σ of 1 was chosen as this gave all the detail in the image to pixel accuracy. The maximum scale of 16 was chosen as this was the coarsest scale at which features were still recognisable in the image. Figure 3.1 shows a noisy image of the `noisy pear` on a complex background. Figures 3.2a&b show the DTs for scales 4 and 16, and figure 3.2c shows the result of adding the DTs at all scales together to form the Multi Scale Distance Transform (MSDT). As expected as the value of σ increases the noise is reduced leaving the more perceptually significant edges and distances. The result of combining the scales together is not perfect because of a property most edge detectors have, namely that of inaccurate edge position estimation in the proximity of other edges. For a pair of straight edges close together, the Marr-Hildreth edge detector detects them at increasing separation as the edges are brought closer together. At coarse scales this effect is most apparent. However this generally only affects small portions of the edges such as near corners. For most features of large objects there is a well defined distance. The same multi-scale approach can be applied to the Salience Distance Transform (SDT) described in section 2 to form the Multi Scale Salience Distance Transform (MSSDT). Figures 3.3a&b show the SDTs for the edge strength weighted algorithm and figure 3.3c shows the MSSDT. Incorporating edge strength into the distance transform gives perceptually better results, especially at the finer scales, at which the edges are generally weak for the fine detail. Results for the DT and SDT at the higher scales are similar because the smoothing has effectively removed the weak edges.

The MSDT and MSSDT show a good trade off between visibility of all of the object and the amount of clutter. The edge of the object is reasonably dark (good salience) and the majority of the rest of the image is reasonably light (poor salience). There are other strong edges giving high salience which is to be expected as there is no reason why the object should stand out more that other strong edges.

4 Using The Salience Distance Transform For Model Evaluation

An alternative to high-level feature matching is to match the model features to the edge pixels in the image. This allows any model feature to match edge data and remove the reliance on good segmentation. Borgefors [11] has proposed a hierarchical distance transform which is used to perform coarse to fine model alignment to the edge data. Mundy and Heller [9] have used the distance transform at one scale on an edge image obtained using the Canny edge detector. West and Rosin [2] have investigated the use of a different scale for the distance transform than that used for feature extraction for recognising surfaces of revolution. A coarse scale was used to reduce the effect of clutter caused by texture near the edges of the object of interest.

To obtain an idea of the performance of the different DTs proposed in this paper, a number of experiments were performed on simulated images. Template matching was performed using boundary models of 2D objects on the DT images. Summing the values of the DT for each model pixel should result in low values for good matches as the model will line up with the edges of the object in the image. It is not our purpose to describe efficient alignment and matching algorithms in this paper but to demonstrate that the proposed DTs can be used in such a scheme. The important requirement with template matching is for there to be a well defined minimum for the correct match. The error space should be reasonably smooth with few local minima. To determine the error space, the error is computed for a number of discrete model positions by perturbing the pose parameters about the position of correct match. Translation in the two orthogonal directions was considered. It can be reasonably assumed that varying the other parameters (rotation, scaling) will have similar effects.

For the first set of experiments, the DT and SDT for each of the edge detector scales were processed. Figure 4.1 shows the `noisy pear` model and figures 4.2a&b and 4.3a&b show the results at scales 4 and 16 for the two algorithms for a region of the error space 100×100 pixels in size approximately centred on the reference point of the model. Notice that in most there is a reasonably well defined minimum. The important point to note is that for the finer scales, there are many local minima while at the coarser scales there is usually one minima. Figures 4.2c and 4.3c show the results for the multi-scale algorithms which show well defined minima.

The same experiments were performed using the same `noisy pear` image but the model used was the `smooth pear` (see figure 4.1). The `smooth pear` model is the underlying shape of the `noisy pear` without the fine detail. It can be regarded as a coarse representation of the model or as a noise-free description. As expected, similar results for matching were obtained using the multi-scale distance transforms.

5 Using The Salience Distance Transform with Active Contours

We show here another application of the salience distance transform. Active contours (*aka* snakes) are a popular method for refining an initial boundary estimate. Early applications of snakes calculated the gradient by running an edge detector on the image [14]. However, techniques for determining the optimal deformation of the snake generally only use a local window around each contour point of the snake at each step. Therefore, since edge detectors produce localised gradient maps, the snake will not be attracted towards the edges unless it is very close. This would require the initial estimate of the snakes position to be very accurate. To overcome this problem, the edge magnitudes were blurred. Unfortunately, this distorts the position of the maximum gradient, requires a parameter to specify the degree of blurring, and still produces a finite width energy well.

More recently, the distance transform has been used instead to generate the image energy term. This has the advantage that it produces a continuous smooth gradient over the whole image, and does not require any parameters. However, the disadvantage is that it does not use the magnitude values of the gradient. Thus, a significant amount of important information is being discarded. If instead one of the SDTs is used, then the advantages of both techniques are available. A continuous gradient is produced which incorporates as much information as is available.

The application of the SDT is demonstrated in the context of bottom-up perceptual grouping [15]. Certain arrangements of ellipses in the image are hypothesised as being the projections of a surface of revolution (SOR) in the scene. This hypothesis is reinforced by searching for symmetric sets of lines or edges about the projected axis of revolution in the image. This enables the occluding boundary of the SOR to be estimated. However, due to clutter and occlusion this estimate is typically very crude. Here we use it as an initial estimate which is refined using snakes. The greedy algorithm [16] is used to calculate the deformation of the snake at each iteration. The initial contour in the examples contains 150 points, and the greedy algorithm was terminated in all cases when less than five points moved during one iteration. The original image is shown in figure 5.1a, and the initial estimate is shown superimposed in figure 5.1b. The blurred edge magnitude (log mapped) is shown in figure 5.2a. Even though the writing on the can is not significant its density causes the blurred edge map to contain a strong energy well. In turn, this causes the snake to move in the wrong direction, and the final snake (figure 5.2b) is mostly wrong except in a few sections in which there was little spurious clutter. The DT is generated from the original thresholded edge map as shown in figure 5.3a. Although not compounded by blurring, the clutter produces many distracting energy wells. This can be seen in the final snake (figure 5.3b) which in several instances (e.g. the bottom right hand rim) has latched on to adjacent background noise. The SDT shown in figure 5.4a does not suffer from either of the above defects, and the final snake follows what we would consider to be the significant contours.

The results of the DT and the edge blurring would be improved by using a larger scale for edge detection, although this may reduce accuracy and in any case, which scale should be used? Even at a single scale the SDT is much less sensitive to artefacts produces by noise and clutter. Using the MSSDT would allow a multi-scale snake algorithm to be used.

6 Conclusions

We have described a method for incorporating various additional factors into the standard chamfering algorithm for calculating the distance transform. This allows a more general salience distance transform to be generated, which is more useful than the standard distance transform for many applications. Due to the limitations of space we have concentrated on edge magnitude. but factors such as curve length and shape and image clutter can easily be added [10]. The accuracy of the various algorithms for calculating the standard DT can be easily analysed. However, saliency is a less well defined concept, and varies from application to application. Given the variability of the components of saliency, its usefulness can be assessed best by experimental demonstrations for each application. We have compared the performance of the SDT with the DT for model matching and an example involving active contours. In both cases, the SDT gave superior results compared to the basic DT.

None of the stages of the MSSDT require any parameters. This is essential if the technique is to be robust over a variety of data without user intervention. In particular, there is no requirement to threshold the edge map. As would be expected, different threshold values can give drastically different results. Moreover, thresholding removes potentially useful information. Although the multi-scale approach removes the necessity for a scale parameter it requires significantly more computation than a single scale SDT. In many cases the single-scale application of the SDT may be sufficient.

References

[1] Borgefors, G., "Distance transformations in digital images". *Computer Vision, Graphics and Image Processing*, **34**, 344-371, 1986.

[2] West, G.A.W. and P.L. Rosin. "Using symmetry, ellipses and perceptual groups for detecting generic surfaces of revolution in 2D images". in *Proc. SPIE Conf. Applications of Artificial Intelligence: XI: Machine Vision and Robotics*, Orlando, USA, **1964**, 369-379, 1993.

[3] Venkatesh, S. and P.L. Rosin. "Dynamic threshold determination by local and global edge evaluation". in *Proc. SPIE Conf. Applications of Artificial Intelligence: XI: Machine Vision and Robotics*, Orlando, USA, **1964**, 40-50, 1993.

[4] Bergholm, F., "Edge focusing." *IEEE Trans. PAMI*, **9**, 726-741, 1987.

[5] Lowe, D.G., "Three-dimensional object recognition from single two-dimensional images", *Artificial Intelligence*, **31**, 355-395, 1987.

[6] Sha'ashua, A. and S. Ullman. "Structural Saliency: the detection of globally salient structures using a locally connected network", in *Proc. 2nd ICCV*, 321-327, 1988.

[7] Guy, G. and G. Medioni. "Perceptual grouping using global saliency-enhancing operators", in *Proc. IAPR*, The Hague, Holland, 1992.

[8] Marr, D. and E. Hildreth, "Theory of edge detection". *Proc. Royal Soc. London B.*, **207**, 187-217, 1980.

[9] Mundy, J.L. and A.J. Heller. "The evolution and testing of model-based object recognition system". in *Proc. 3rd ICCV*, Osaka, Japan, 1990.

[10] Rosin, P.L. and G.A.W. West, "Salience distance transforms". Tech. Rep. 12, Dept. of Computer Science, Curtin University of Technology, Perth, Australia, 1993.

[11] Borgefors, G., "Hierarchical chamfer matching: a parametric edge matching algorithm". *IEEE Trans. PAMI*, **10**, 849-865, 1988.

[12] Borgefors, G., Private communication, 1993.

[13] Neveu, C., C. Dyer, and R. Chin, "Two-dimensional object recognition using multiresolution models". *Computer Vision, Graphics and Image Processing*, **34**, 52-65, 1986.

[14] Kass, M., A. Witkin, and D. Terzopoulos. "Snakes: active contour models". in *Proc. 1st ICCV*, London, UK, 1987.

[15] Rosin, P. and West G.A.W., "Detection and verification of surfaces of revolution by perceptual grouping". *Pattern Recognition Letters*, **13**, 453-461, 1992.

[16] Williams, D. and M. Shah. "A fast algorithm for active contours". in *Proc. 3rd ICCV*, Osaka, Japan, 1990.

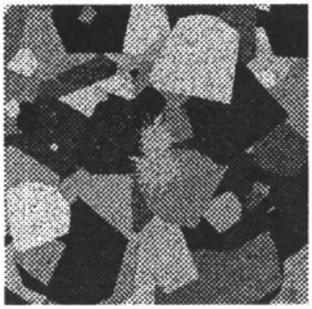

Figure 3.1

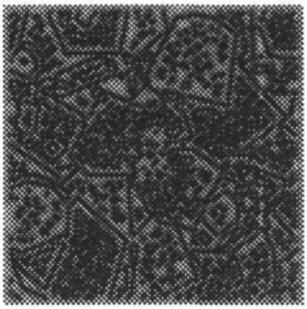

Figure 3.2a

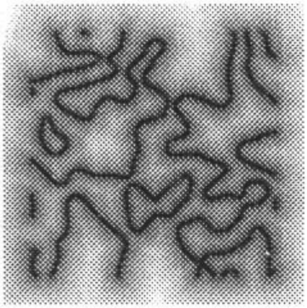

Figure 3.2b

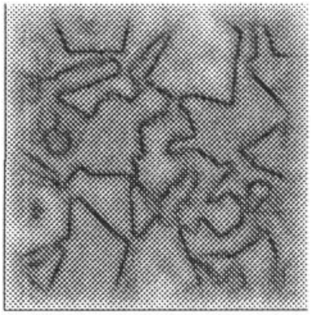

Figure 3.2c

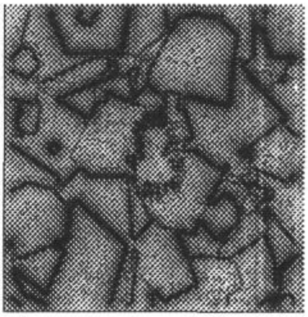

Figure 3.3a

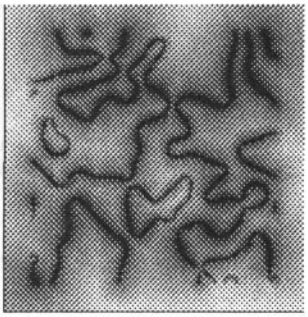

Figure 3.3b

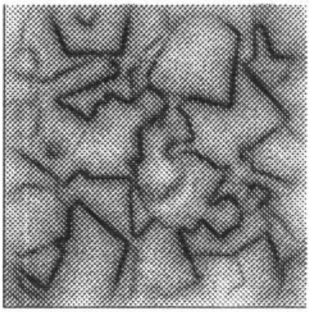

Figure 3.3c

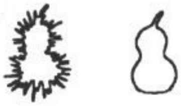

Figure 4.1

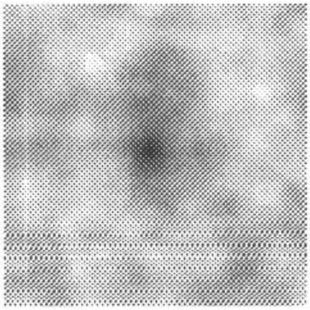

Figure 4.2a

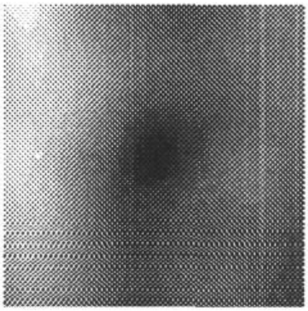

Figure 4.2b

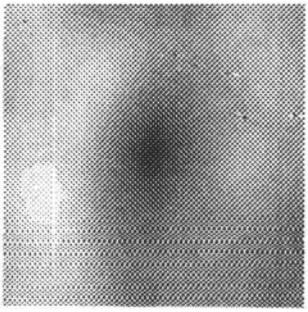

Figure 4.2c

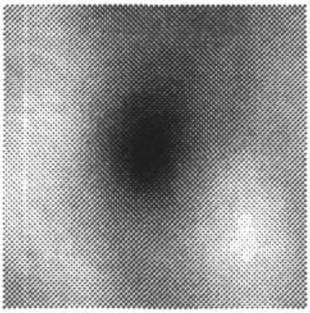

Figure 4.3a

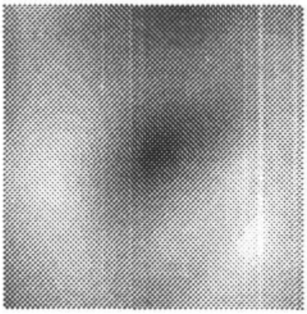

Figure 4.3b

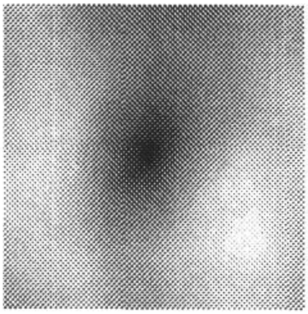

Figure 4.3c

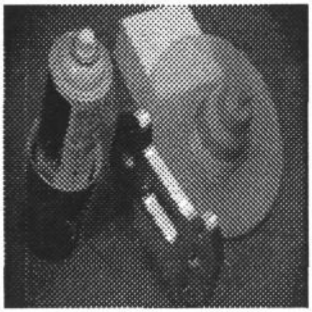

Figure 5.1a

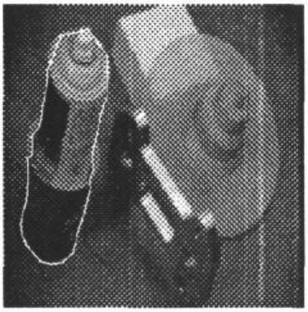

Figure 5.1b

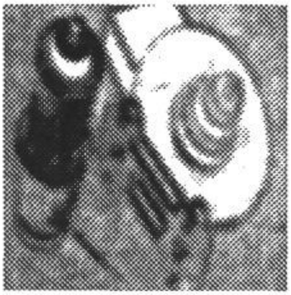

Figure 5.2a

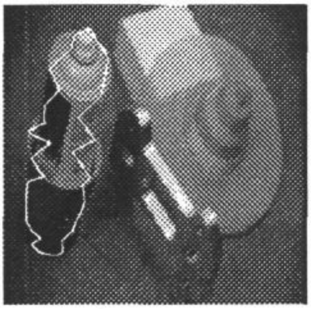

Figure 5.2b

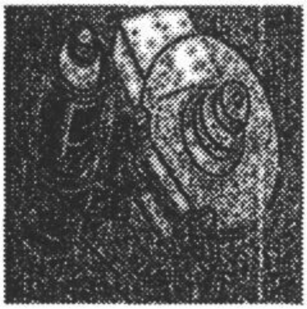

Figure 5.3a

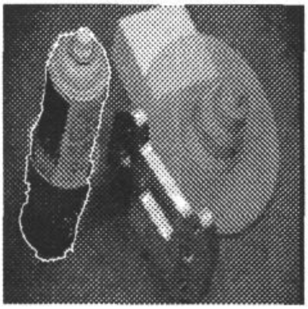

Figure 5.3b

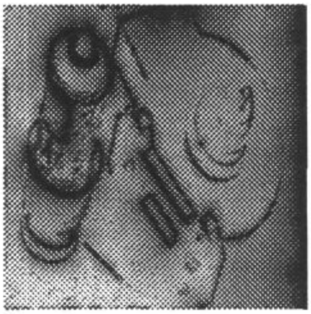

Figure 5.4a

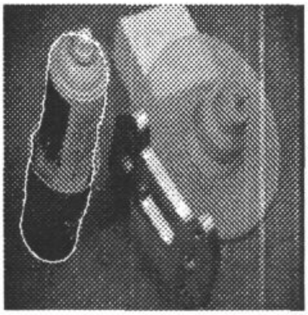

Figure 5.4b