Scale-Space Local Thresholding and Segmentation

Xue-Dong Yang
Department of Computer Science
University of Regina
Regina, Sask.
CANADA S4S 0A2

Xiaoxing Chen and Mingyuan Chen
Faculty of Engineering
University of Regina
Regina, Sask.
CANADA S4S 0A2

Abstract

Previous techniques for adaptive local thresholding generally divide an image into a set of rectangular subimages. A problem with them is that no systematic consideration is given to the size of subimages and no study is done about the effect of size on the results. In addition, they tend to create artifacts at the artificial boundaries between subimages. Some efforts have been made to reduce such artifacts by using interpolations, which, however, do not completely solve the problem. This paper presents a novel scale-space local thresholding technique. A variable sized window is used to scan across an image, and an accurate threshold is to be computed for the center pixel at each window stop. By varying the size of a window, a threshold is computed at multiple scales and, consequently, the image can be segmented at multiple scales. The size of the window is allowed to vary continuously for threshold selection. In the obtained segmentations, the contours of segmented regions vary continuously in the scale space. This provides a new structure for multiple resolution segmentation.

1 Introduction

Image segmentation is a process of partitioning an image into a set of non-overlapping regions. Its purpose is to decompose an image into parts that are meaningful with respect to a particular application. A variety of segmentation techniques have been developed for different applications [15]. However, each has its limitations. The existing segmentation techniques can be broadly classified into three categories: discontinuity-based techniques, region growing techniques, and thresholding techniques. The work reported in this paper belongs to thresholding techniques. Thresholding techniques are computationally simple and never fails to define disjoint regions with closed boundaries.

Image characteristics that are used for defining a partition can change over a broad range of intensity distribution across an image. In the case of thresholding techniques, a threshold value may work well in one region in an image, but may perform poorly in another. This makes it difficult to select thresholds based on global information. This consideration leads naturally to the investigations of adaptive thresholding techniques [15] [2][3][11][13].

By adaptive thresholding, a threshold is selected according to local characteristics in a subimage. Fernando and Monro [3] suggested a local thresholding technique for X-ray angiograms. The histogram of such images are unimodal with narrow peaks, and so most global thresholding techniques produce unsatisfactory results. According to this method, the image is partitioned into 16 nonoverlapping subimages and the entropic thresholding technique of Pun [13] is applied to determine the threshold value for each of these subimages. This method may yield a thresholded image with graylevel discontinuities at the boundaries of two different subimages. A low-pass filter is then used to reduce these discontinuities.

Chow and Kaneko [2] proposed a variable thresholding method. Their idea is to test for bimodality. If the histogram for a window is bimodal, then a threshold is computed. If the histogram for a window is unimodal, then no threshold can be directly computed. In this case, the threshold will be defined by a value interpolated from the thresholds found in neighboring bimodal windows. Nakagawa and Rosenfeld [11] extended this method to trimodal cases and found an improvement over bimodal cases.

Weszka, Nagel and Rosenfeld [17] suggested determining a histogram for only those pixels having high Laplacian magnitude. They reason that there will be a shoulder of the gray level intensity function at each side of the boundary. The valley-seeking method for threshold selection has a chance of working on those local histograms.

Though the above methods have been successful in certain applications, they have several problems. First, there is no systematic consideration for the size of a window. We believe that the size of the window should be related to the scale of features under consideration in an image. We argue that the size of the window must be selected appropriately so that it generates reliably a unimodal histogram when it is inside a homogeneous region, and a bimodal histogram when it crosses the boundary of two distinguished regions. More specifically, let us consider the situation where a window is located near a boundary between two regions. In this case, if the window is too small, it may not cover enough regions on both sides of the boundary to produce a well-defined bimodal histogram for reliable threshold selection. On the other hand, if the window is much larger than the scale of the features, it may cover too large an area such that a multimodal histogram is produced, making threshold selection difficult. Therefore, the size of the window should not be arbitrarily chosen, instead, it has to be determined with a consideration to the scale of the features under consideration in the image. Furthermore, the scale of the features may even vary across an image. In such a case, the size of the window should not be fixed, but, can vary across an image according to the scale variation of the features.

Second, because of the subdivision of the image into independent windows, there are chances that the threshold values for each window might change abruptly, creating unnecessary artifacts at the artificial boundaries between the subimages. Lower-pass filtering technique has been used to reduce these artifacts, but the problem is not completely solved. Ideally, a window should scan each pixel across an image. But it will be very expensive in terms of computational cost. A compromise is to divide the image into subregions and then have thresholds interpolated between regions as presented in [2][11]. However, it can perform very poorly in situations where the object is between two adjacent windows.

Motivated by the above considerations, a novel scale-space thresholding technique, by using continuously varied window, is studied. It should be noticed that multiple resolution segmentation problem has been studied by a few researchers previously. For example, Bouman and Liu [1] studied this problem using pyramid images. Montanvert, Meer and Rosenfeld [10] studied this problem also using a pyramid structure, however with irregular tessellations. The scale-space technique developed in this work is different from these techniques which construct pyramids and do coarse-to-fine analysis through the hierarchy of the pyramids. Instead, our analysis is based directly on the input image with an operator whose scale-parameter can vary continuously. This results in a new structure with a finer analysis which bridges the gaps appeared between the adjacent levels of pyramidal hierarchies.

The idea of using scale-space technique for signal analysis was introduced by Witkin [18]. It has been later successfully applied to multiple resolution edge detection problems [8][4]. The previous scale-space techniques are mostly based on the convolution of an image with a Gaussian distribution or its derivatives. The standard deviation σ in Gaussian distribution is defined as the scale parameter. Multiple resolution edge detection can also be achieved through wavelet transforms [9], which has been shown to be equivalent to the previous scale-space edge detection techniques.

The scale-space technique, introduced for the purpose of threshold determination in this paper shares a similar philosophy and objective with the previous scale-space techniques. It, however, has a different definition due to the nature of threshold selection problem. In this work, the size of the window which is used for determining local threshold is allowed to vary continuously. Consequently, an image is segmented by the thresholds selected at multiple scales. In the obtained segmentations, the contours of partitioned regions vary continuously in the scale space. By tracing out those contours, a new structure is obtained for multiple resolution segmentation.

The paper is organized as follows. Section 2 discusses the methods to determine the modality of a local histogram, and selection of local threshold. Section 3 presents a new scale-space adaptive thresholding method. Section 4 discusses construction of contours of segmented region boundaries in scale-space. Finally, some interesting problems for future research along the direction initiated by this preliminary work are briefly discussed.

2 Local Bimodality and Threshold Selection

In local thresholding, there are two cases for the histogram of the subimage under the scanning window. When the window is mostly inside a homogeneous region, the histogram will be unimodal. When the window is across the boundary of two regions, the histogram will be bimodal. Therefore, the first computation in local thresholding is to determine the modality of a local histogram. Second, a local threshold is to be computed in the case of bimodal histogram.

The modality of a local histogram can be determined based on the variance of the histogram. Given a histogram $h(g)$, $g = 0, 1, ..., L-1$, where L is the number of gray levels, the variance of the histogram is defined as:

$$s = \frac{1}{M} \sum_{g=0}^{L-1} (h(g) - \mu)^2, \qquad (1)$$

where M is the total number of pixels in the image:

$$M = \sum_{g=0}^{L-1} h(g) \qquad (2)$$

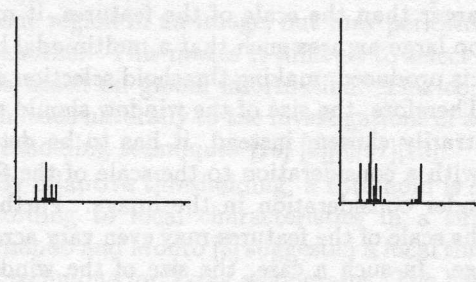
and μ is the mean gray level:

$$\mu = \frac{1}{M} \sum_{g=0}^{L-1} g h(g) \qquad (3)$$

A prespecified threshold variance value s_T can be used to determine modality. If $s \geq s_T$, the histogram is considered to be bimodal. Otherwise, it is unimodal.

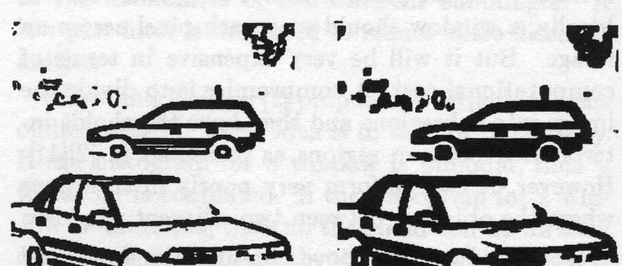
After the modality is determined, the next task is to compute a threshold for partitioning. Many threshold finding methods have been studied in the past and can be applied here. The simplest method used in earlier time is to find the valley point between the two peaks corresponding to two principal modes in the bimodal histogram [5][6]. Some more sophisticated later techniques are based on finding a threshold that minimizes a criterion function. For example, Otsu [12] suggested a method that minimizes the weighted sum of group variances. The weights are the probabilities of the respective groups. Kittler and Illinggworth [7] suggested a different criterion from Otsu's. They assume that the observations come from a mixture of two Gaussian distributions. They determine the threshold that minimizes the Kullback (1959) directed divergence from the observed histogram to the unknown mixture distribution. In our situation, since a bimodal histogram is resulted from a window located across the boundary of two regions, two peaks in the histogram are generally well separated and roughly even. Thus, the simple threshold finding technique performs reasonably well. It should be pointed out that the local histogram could be noisy, making the determination of valley point nontrivial. Therefore,

Figure 1. Car image.



(a) (b)

Figure 2. (a) A unimodal histogram; (b) a bimodal histogram.



(a) (b)

Figure 3. Modality classification under (a) 5X5 window and (b) 7X7 window.

a smoothing operation on the histogram is necessary before the search operation for the valley point is performed.

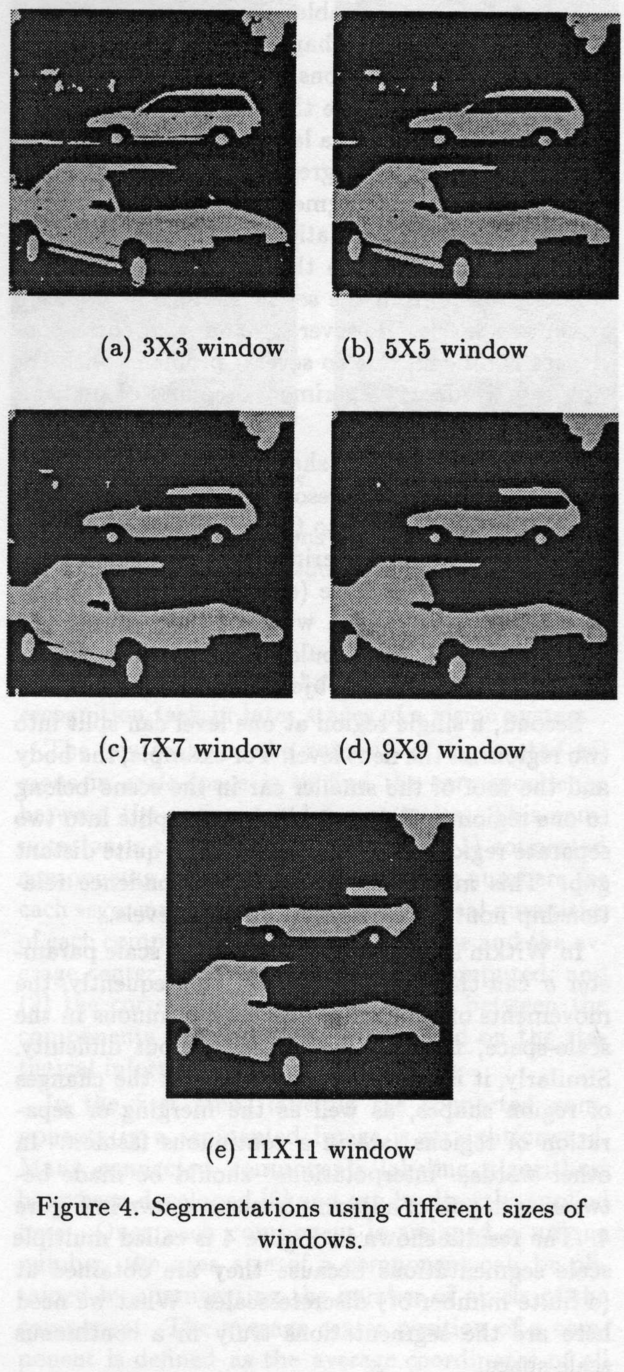
Figure 1 shows a test image. Figure 2.(a) shows a unimodal local histogram at the pixel (143, 97) with a 5X5 window. The local histogram becomes bimodal if a 7X7 window is used at the same location (Figure 2.(b)). Figure 3 shows the results of modality classification. The white regions correspond to pixels with a unimodal local histogram, and the black regions a bimodal histogram. Figure 3.(a) was obtained by using a 5X5 window, and Figure 3.(b) by a 7X7 window. This tells us clearly that the modality of a local histogram is only relative, and is strongly related to the size a spatial range is based on. However, there is no rule from previous literature for selecting an appropriate window size for finding local thresholds. The fact is that there is, perhaps, no single optimal window size which will be suitable for all kinds of images, or even for different regions in a same image. In a very similar spirit to Witkin's scale-space filtering, we studied the local thresholding problem at multiple scales, and provide a segmentation structure in scale-space which will allow a reliable segmentation to be derived. We want the new technique to be suitable for general situations of different types of images. This leads to the central work of this paper discussed in the next section.

3 Scale-Space Local Thresholding

Our investigation of scale-space local thresholding starts with a set of simple experiments. We begin with a 3X3 window which scan across an image to determine a local threshold at each pixel location and perform thresholding. Then, the window is expanded by one pixel at all four sides, and do the local thresholding again. This window expansion is repeated for many times. A sequence of segmentations, obtained by using 3X3, 5X5, 7X7, 9X9, and 11X11 windows respectively, are shown in Figure 4.

In this set of partitions, a multiple resolution segmentation of the image is illustrated. At the smallest scale, the image is partitioned into many regions, some are corresponding to small features in the image. As the scale increases, small regions gradually disappear and remaining regions have simpler shapes. When the window size reached 11X11, only a few larger regions stay, which correspond to the abstract shapes of those large objects in the scene.

Segmentation is usually a preprocessing stage in



(a) 3X3 window (b) 5X5 window

(c) 7X7 window (d) 9X9 window

(e) 11X11 window

Figure 4. Segmentations using different sizes of windows.

 Vision Interface '95

a vision system. Although this work itself is not a study of a complete vision system, we should keep in mind clearly the role of the segmentation in a vision system, and the objective of the segmentation stage. The multiple resolution segmentations obtained above may facilitate some object recognition tasks in the later stages of a vision system. The segmentation at a suitable large scale may provide a simple and abstract shape of an object. As compared with segmentations at smaller scales, some small features and noise that are shown at smaller scales are eliminated at a larger scale. A description of an object can be progressively refined by finding finer features in the segmentations at lower scales. To perform this computation, it is necessary to find correspondence between the regions that represent a same object from the segmentations at adjacent resolution levels. However, to find such correspondences is difficult due to several problems with the above preliminary experimental results of multiple scale segmentations.

First, the boundary shape of a region changes significantly from one resolution to another. This phenomenon is similar to the movement of contour points in scale-space filtering [18]. Although regions obtained at a large scale (e.g. two cars in the segmentation by an 11X11 window) have simple and abstract shapes, they could largely deviate the actual boundaries of the objects in the scene.

Second, a single region at one level can split into two regions at the next level. For example, the body and the roof of the smaller car in the scene belong to one region in Figure 4.(d); but, it splits into two separate regions in Figure 4.(e) with a quite distant gap. This makes finding the correspondence relationship nontrivial between different levels.

In Witkin's scale-space filtering, the scale parameter σ can change continuously. Consequently, the movements of contour points are continuous in the scale-space, thus can be traced without difficulty. Similarly, it is desired in our case that the changes of region shapes, as well as the merging or separation of regions, are in a continuous fashion. In other words, "interpolations" should be made between each pair of adjacent levels as shown in Figure 4. The results shown in Figure 4 is called multiple scale segmentations because they are obtained at (a finite number of) discrete scales. What we need here are the segmentations truly in a continuous scale-space.

Direct interpolation between regions from adjacent levels is not possible because it requires the correspondence between a pair of regions from two different levels to be known first. To obtain a scale-space segmentation, we investigated the problem of local thresholding by using a window with continuously varied sizes.

A digital image f[x, y] is a set of equally spaced discrete samples of a continuous light-intensity function f(x, y), arranged in the form of two-dimensional array. In conventional window based operations, the grids of a window are aligned with the grids in a digital image. In order to allow an arbitrarily sized window to be used, what we need is the reconstruction (or an approximation) of the original continuous function f(x,y), such that this continuous function can be resampled at an arbitrary interval. Thus, in the following an approximation of f(x, y) is first defined, and then the local thresholding with arbitrary window size is introduced.

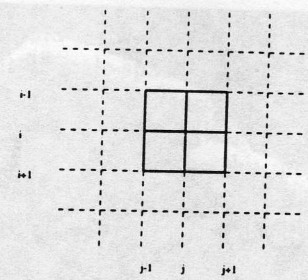
Definition: .ft I Given a digital image f[x, y], for x, y = 0, 1, 2, ..., N-1, a continuous approximation of f[x, y], denoted by f(x, y), is defined as the bi-linear interpolation of f[x, y]. That is, for any real point (x, y), $0 \leq x, y \leq N-1$, f(x, y) is defined as the bi-linear interpolation from the its four nearest grid points: $f[\lfloor x\rfloor, \lfloor y\rfloor]$, $f[\lfloor x\rfloor+1, \lfloor y\rfloor]$, $f[\lfloor x\rfloor, \lfloor y\rfloor]+1$, and $f[\lfloor x\rfloor+1, \lfloor y\rfloor]+1$. .ft R

Once the continuous approximation function f(x, y) is defined, we can resample f(x, y) at any real point (x, y). Suppose a $D \times D$ window, where D is a real number, is centered at an arbitrary pixel in the digital image, we can digitize the window into equally spaced grids and use resampled values on these grids to perform window operation. It is necessary to know what is an appropriate choice of interval d for this digitization. By Whittaker-Shannon sampling theorem [5], the interval d should not be greater than 1. Otherwise, there will be some information loss. On the other hand, there is no need to use a very small interval because oversampling will produce redundant information. Thus, it is justified to choose the largest d, for $d \leq 1$, such that d completely divides the dimension D. This leads to the following definition for selecting d:

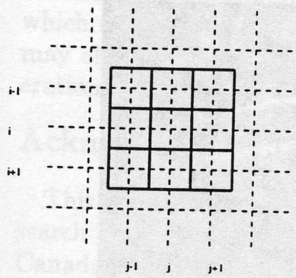
$$d = \begin{cases} 1 & \text{if } D - \lfloor D\rfloor = 0 \\ D/(\lfloor D\rfloor + 1) & \text{otherwise} \end{cases} \tag{4}$$

It is easy to see that $d \leq 1$.

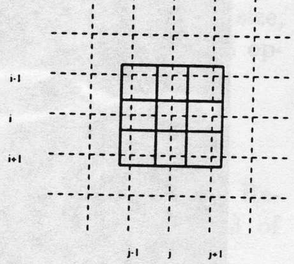
Figure 5 shows several examples of digitized windows of various sizes. If D is an odd integer (e.g. a 3X3 window as shown in Figure 5.(a)), the layout of the digitized window over the digital image is exactly the same as the traditional window. If D is an even integer (e.g. a 4X4 window), the layout of the digitized window over the digital image is shown in Figure 5.(b). When D is a non integer value (e.g.



(a) 3X3



(b) 4X4



(c) 3.5X3.5

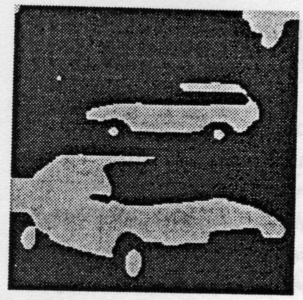
Figure 5. Digitization of arbitrarily sized windows

a 3.5X3.5 window), the layout of the digitized window over the digital image is shown in Figure 5.(c). In the last case, the interval d between grids is less than 1.

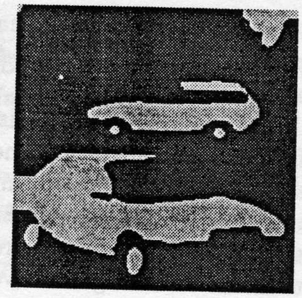
Finally, the computations described in Section 2 can be applied to the window we defined here to determine the modality of a local histogram and perform thresholding. Figure 6 shows a sequence of segmentations using windows of sizes 9.25X9.25, 9.5X9.5, 9.75X9.75, and 10X10 respectively, to illustrate the gradual separation between the roof and body of the smaller car in the scene. Any number of levels can be inserted between a given interval in the scale-space. Therefore, a scale-space local segmentation technique is completely defined.

4 Contour Tracing in Scale-Space

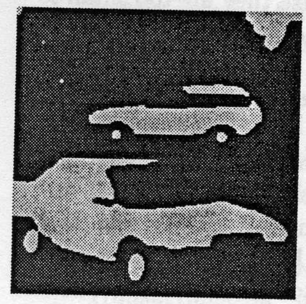
With the scale-space segmentation obtained above, the task of traing the contours of corresponding regions from different scales becomes relatively easy. The purpose of contour tracing described in this paper is the same as that of Witkin's scale-space filtering, which is to provide a qualitative description about the segmentation in scale-space. The qualitative description, as an output from the seg-



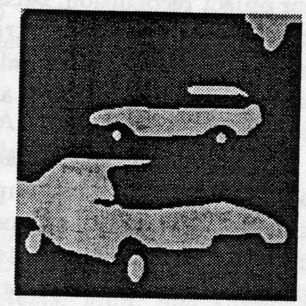
(a) 9.25X9.25 window



(b) 9.5X9.5 window



(c) 9.75X9.75 window



(d) 10X10 window

Figure 6. Segmentations using different sizes of windows.

mentation process, can better facilitate the object recognition task in later stages of a vision system.

The key to trace the contours of segmented regions in scale-space is to find the correspondence between the regions at different scales. This computation is carried out in two steps: (1) connected components are labeled by identifying numbers for each segmented image, and the statistical quantities of each component, such as the area size and the average center of the component, are computed; and (2) the correspondence is established between the components from adjacent levels based on the statistical information of the components.

In the first step, labeling the connected components in a segmented image is straightforward. Many connected components labeling algorithms have been developed [6] and can be directly applied here. Once each component is assigned a unique number, the area size of a component can be obtained by enumerating the number of pixels of the component. The average center position of a component is defined as the average coordinates of all pixels of a component. We like to point out that the computations of the area size and the average center of a component can be carried out in separate steps, or integrated into the connected components labeling algorithm. For each segmentation, a components list is constructed in sorted order by their sizes.

Before the second step is performed, a preprocessing on the labeled components may be introduced to simplify the segmentation results. The preprocessing is to eliminate very small components. Those small components are present usually due to the noise in the image. However, those small components may also correspond to small features in the image. Nevertheless, they generally cannot be used for object recognition, because they are too small to be reliable. Therefore, a criterion may be set for the area size to eliminate small components. The eliminated small components are merged into neighboring components.

In the second step, each component from one segmentation level is compared with the candidate components from an adjacent segmentation level. A component is considered to be a candidate for comparison if its size is similar to the size of the component being compared. If the center location of the candidate component is also similar to that of the component being compared, a corresponding relationship is established between the two components. This corresponding relationship is represented by a symbolic link in the internal data structure. Figure 7 shows a set of contours corresponding to the smaller car in the scene found by the tracing operation described above. It is a stack of contours found at different scales shown in 3D views. The brightest layer corresponds to the smallest scale (i.e. 3X3 window), and the darkest the largest scale (i.e. 11X11 window). The continuous variation of contours is clearly illustrated.

With this contour structure, the object recognition task may become easier. For example, the existence of an object may be found more easily and reliably at a relatively large scale. The bottom level of the contour trace provides a more accurate shape of the object which allows a more conclusive match. In particular, the separate roof and body at large scale can be found to belong to the same object by tracing them out to small scale.

5 Discussions

This paper presents a novel scale-space local thresholding technique. A window with continuously varied size is used to scan an entire image, and an accurate threshold is to be computed for the center pixel at each window stop. By varying the size of a window, the thresholds are computed at multiple scales and, consequently, the image is segmented at

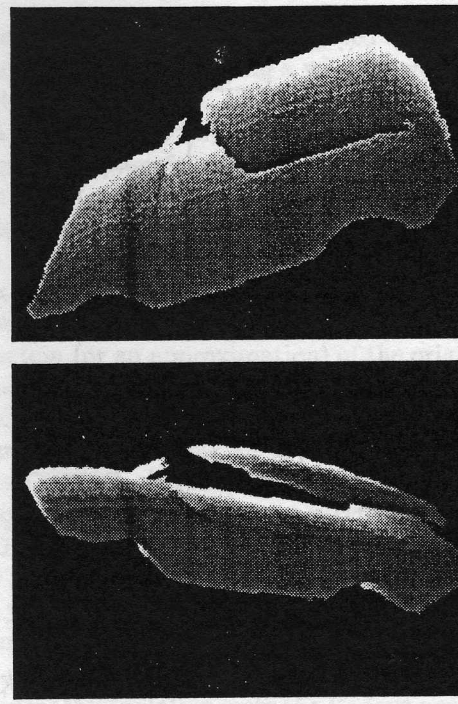


Figure 7. Two different views of contour surfaces.

multiple resolutions. In the obtained segmentation results, the contours of segmented regions vary continuously in the scale space. By tracing out the contours of the corresponding regions, a qualitative structure of segmentation in scale-space can be derived, which will facilitate object recognition task in vision systems.

The computational cost of the current technique is relatively expensive because several local thresholds are computed at several different window sizes for each pixel location. The computing for a 3X3 window is arroud 30 seconds. 7X7 and 11X11 are 40 seconds and 48 seconds respectively, on Silicon Graphics Personal IRIS workstation. Some efforts have been made to improve the efficiency. For example, consider the case where the dimension D of the window is an integer. When the window is shifted by one pixel distance, there is a large overlapping among the previous grid locations and the new grid locations of the window. Therefore, the variance and mean of the local histogram at the new location do not need to be recomputed completely. Instead, they can be evaluated incrementally from the variance and mean from the previous window location. However, this improvement is not applicable if D has a non-integer value. In this case, the computing time required for a window is around a few minutes. Further improvement for these computations should be studied in the future.

The result of scale-space segmentation can serve as a basis to derive a hierarchical graph structure for the segmented regions at different levels. A node at a higher level in the hierarchy may represent a coarse region found in the segmentation at a large scale. The children of this node may correspond to finer regions available in the segmentations at smaller scales. To find a general such technique seems to be quite challenging.

Finally, the technique developed in this work, which uses a window with continuously varied size, may also be applicable to other window based operations.

Acknowledgments

This work is supported in part by NSERC Research Grants OGP105708 and OGP0121863 of Canada, X.D. Yang and M. Chen, respectively.

References

[1] Bouman, C. and B. Liu. "Multiple resolution segmentation of textured images." *IEEE Tran Pattern Analysis and Machine intelligence 13(2)*, 1991, pp.99-113.

[2] Chow, C.K. and T. Kaneko. "Automatic boundary detection of left ventricle form cineangiograms." *Computer Biomedical Research 5*, 1972, 338-410.

[3] Fernando, S.M.X. and D.M. Monro. "Variable thresholding applied to angiography." *Proceedings, 6th International Conference on Pattern Recognition*, 1982.

[4] Fleck, M.M. "Multiple widths yield reliable finite differences." *IEEE Tans Pattern Analysis and Machine Intelligence, 14(4)*, 1992, pp.412-429.

[5] Gonzalez, R.C. and P. Wintz. *Digital Image Processing.* 2nd edition. Addison-Wesley, Reading, MA, 1987.

[6] Haralick, R.M. and L.G. Shapiro. *Computer and Robot Vision, Vol. I*, Addison-Wesley, Reading, MA, 1992.

[7] Kittler, J. and J. Illingworth. "On threshold selection using clustering criteria." *IEEE Trans Systems, Man, and Cybernetics, 15*, 1985, pp.652-655.

[8] Lu, Y. and R.C. Jain. "Reasoning about edges in scale space." *IEEE Trans Pattern Analysis and Machine Intelligence, 14(4)*, 1992, pp.450-468.

[9] Mallat, S. and S. Zhong. "Characterization of signals from multiscale edges." *IEEE Trans Pattern Analysis and Machine intelligence, 14(7)*, 1992, pp.710-732.

[10] Montanvert, A., P. Meer and A. Rosenfeld. "Hierarchical image analysis using irregular tesselations." *IEEE Tans Pattern Analysis and Macjine Intelligence, 13(4)*, 1991, pp.307-316.

[11] Nakagawa, Y. and A. Rosenfeld. "Some Experiments on Variable Thresholding." *Pattern Recognition. Vol. 11*, 1979, 191-204.

[12] Otsu, Nobuyuki. "A Threshold Selection Method from Gray-Level Histograms." *IEEE Transactions on Systems, Man, and Cybernetics. Vol. SMC-9, No. 1*, 1979, 62-66.

[13] Pun, T. "A new method for gray level picture thresholding using the entropy of the histogram." *Signal Processing 2*, 1980, pp.223-237.

[14] Reddi, S.S, S.F. Rudin, and H.R. Keshavan. " An Optimal Threshold Selection Scheme for Image Segmentation." *IEEE Transactions on System, Man, and Cybernetics. Vol. SMC-14, No.4*, July/August 1984, 661-665.

[15] Sahoo, P.K., S. Soltani, A.K.C. Wong, and Y.C. Chen. "A Survey of Thresholding Techniques." *Computer Vision, Graphics, and Image Processing 41*, 1988, 233-260.

[16] Wang, N.Q., X.D. Yang, and L.R. Symes. "Scale-Space Filtering and Threshold Hierarchy for Image Segmentation." *Proceedings of Canadian Conference on Electrical and Computer Engineering*, Sept. 25-27, 1991, 20.4.1-4.

[17] Weaszka, J.S., N. Nagel and A. Rosenfeld. "A threshold selection technique." *IEEE Trans Computers, Vol. C-23*, 1974, pp.1322-1326.

[18] Witkin, A.P. "Scale space filtering." *Proc. IJCAI*, Karlsruhe, W. Germany, 1983, 1019-1023.

[19] Yang X.D. and Vipin Gupta. "An Improved Threshold Selection Method for Image Segmentation." *Proceedings of Canadian Conference on Electrical and Computer Engineering*, 1993, 531-534.