October 1989

Scene Segmentation in the Framework of Active Perception

Ruzena Bajcsy
University of Pennsylvania

Alok Gupta
University of Pennsylvania

Helen Anderson
University of Pennsylvania

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-89-69.

Scene Segmentation in the Framework of Active Perception

Abstract

It has been widely acknowledged in the Machine Perception community that the Scene Segmentation problem is ill defined, and hence difficult! To make our primitives adequately explain our data, we perform feedback on processed sensory data to explore the scene. This is Active Perception, the modeling and control strategies for perception.

Comments

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-89-69.

Scene Segmentation In The Framework Of Active Perception

MS-CIS-89-69
GRASP LAB 194

Ruzena Bajcsy
Alok Gupta
Helen Anderson

Department of Computer and Information Science
School of Engineering and Applied Science
University of Pennsylvania
Philadelphia, PA 19104-6389

October 1989

Acknowledgements:
This work was in part supported by Airforce grant AFOSR 88-0966, Army/DAAG-2984-K-0061, NSF-CER/DCR82-19196 A02, NASA NAG5-1045, ONR SB-35923-0, NIH 1-R01-NS-23636-01, NSF INT85-14199, ARPA N00014-88-K-0630, NATO grant 0024/85, DuPont Corp. by Sandia 75 1055, Post Office, IBM Corp. and Lord Corp, We would also like to thank Patricia Yannuzzi for her editorial assistance on this document.

Scene Segmentation in the Framework of Active Perception

Ruzena Bajcsy
Alok Gupta
Helen Anderson

Computer & Information Science Department
University of Pennsylvania
Philadelphia, PA 19104 *

October 18, 1989

1 Introduction

It has been widely acknowledged in the Machine Perception community that the Scene Segmentation problem is ill defined, and hence difficult! To make our primitives adequately explain our data, we perform feedback on processed sensory data to explore the scene. This is Active Perception, the modelling and control strategies for perception.

Definition of the problem:
Segmentation process is a data reduction and requantization of the sensory measurements. The key question is what are the new units, primitives, into which we wish to requantize the data. Unless we define what these primitives are we cannot measure the performance and completion of the segmentation process. This is why we tie the segmentation process to the part primitives [baj,sol,gup88].

What should the primitive be?
In general this depends on the task and the nature of the measurements. In order to make progress we shall limit ourselves to only visual, non-contact measurements, 2D or 3D. Throughout this work we are not assuming that any higher level knowledge is available! One consequence of this limitation is that movable and removable parts will not be recognized. This is because these parts cannot be recognized without manipulation [baj,tsik89]. We also assume that the objects are static, the illumination is fixed in relationship to the camera, but the observer is mobile, and can

*Acknowledgement: This work was in part supported by: Airforce grant AFOSR 88 0244, AFOSR 88-0966, Army/DAAG-29-84-K-0061, NSF-CER/DCR82-19196 Ao2, NASA NAG5-1045, ONR SB-35923-0, NIH 1-RO1-NS-23636-01,NSF INT85-14199, ARPA N0014-88-K-0630, NATO grant No.0224/85, DuPont Corp. by Sandia 75 1055, Post Office, IBM Corp. and LORD Corp. We would also like to thank Patricia Yannuzzi for her editorial assistance on this document.

take and control the data acquisition process, hence the Active Perception paradigm. The goal of 3D segmentation is to divide and cluster range measurements into solids of primitive shapes and primitive surfaces that correspond to (at least in appearance) to one physical material and primitive curves that correspond to physical boundaries. The goal of 2D segmentation is to divide and group intensity measurements into regions with some determined characteristics and primitive 2D curves. We postulate that the problem of 3D segmentation is better defined and hence easier than the 2D segmentation. This is because the projection of a 3D shape into a 2D shape is a nonlinear transformation. Therefore, there are many possible 3D interpretations of a 2D shape.

2 Segmentation problem - a brief essay

As stated in the introduction, the question is: **what primitives should we choose?**

There are two extreme approaches:

1. Simple and one only primitive, such as:
   - for 3D volume: a cube or a sphere
   - for 3D surface: a plane
   - for 3D boundary: a straight line segment
   - for 2D region: homogeneous, constant gray/color value
   - for 2D boundary: a straight line segment
2. Multiple primitives, as many as the data requires for the best fit:
   - for 3D volume: n dimensional parametric volume
   - for 3D surface: nth order surface description (n-order polynomial as an example)
   - for 3D boundary: nth order curve description
   - for 2D region: arbitrary surface description of gray scale and/or color
   - for 2D boundary: arbitrary curve description

The advantage of the first approach is the simplicity of detection of these primitives. The disadvantage is that the data is poorly approximated and typically either oversegmented, or undersegmented, or both. The advantage of the second approach is that the segmentation process will result in a natural best fit approximation to the data. The disadvantage is that it is very difficult and expensive to compute the fit and it does not always give unique results.

Example demonstrating the first approach: Consider a circular segment fit into a straight line or a straight line fit into a circle.

Example demonstrating the second approach: Consider an undulated surface 3 times as large as one finds on sandumes, or waves in the ocean. This surface will be fitted by a 10th order polynomial.

We hope that the reader sees the point that neither of these extreme approaches are desirable. Hence, one is seeking a compromise. However, every compromise will cause some problems. In this paper we shall make a design decision and choose:

- for volume: superquadric primitive with deformation parameters of bending and tapering along the major axis, as introduced by Solina [sol87].
- for 3D surface: up to second order surfaces
- for 3D boundary: up to second order curves
- for 2D regions: up to second order surface fit to the signal measurement (either brightness or saturation)
- for 2D boundary: up to second order curved segments

The advantage of this choice is that it covers a larger class of possible geometric objects (more than just one primitive) and yet it is easier to compute than the general nth order polynomial fit. The disadvantage is that invariably we will have scenes/objects oversegmented and/or undersegmented. The goal is however to recognize both cases, i.e. the oversegmentation and undersegmentation and make the appropriate corrections. It is in the process of correction where the Active Perception comes into play. The above primitives provide a vocabulary in which the final description of the scene will appear.

3 Segmentation of 3D data

This section is based on work of Gupta [gup89]. The assumption here is that we have a mobile 3D range camera available, as constructed by Tsikos [tsik87], and as shown in Figure 1. The physical properties of this camera are such that one does not obtain any shading measurements, though shadows play a role which does need to be corrected by different views of the camera (this is just one example why) the mobile camera/observer is important.
The goal of the segmentation process is to describe the scene/objects in terms of volumetric parts, surface and boundary details. These three components form the complete representation of the object, yet clearly representing different granularity/resolution of the description. The evaluation of the segmentation process is done in terms of the magnitude of the residuals between the models (primitives) and the data. In the case of oversegmentation, chunking of segments into larger entities will be attempted. For example, if an undulated surface is segmented into consecutive second order patches then they maybe grouped together and denoted as one surface. Notice, this grouping is done on the symbolic level and one needs to verify it by going back to the signal level, perhaps invoking higher order primitive than just second order (this is acceptable since it is on a different level in the hierarchy of representations).
Following Gupta's proposal, we begin with a volumetric fit, see Figure 2. The residual is measured two ways simultaneously: one is the difference between the occluded contour of the model and the data, the other is the difference between the surface points of the model and the data. If both residuals are smaller than a threshold (obtained from signal/noise ratio), then there is only one

verification step left to check. That is why we need to verify the assumption that this object/part is symmetric. This calls for moving the camera 180 degrees, scanning the object and repeating the volumetric fit. If the surface residual is the only one which is bigger than the threshold then the implication is that the surface is undersegmented and we apply local fit up to second order patches. An example of this case is the vase in Figure 3 where the first approximation is a tapered cylinder, and it is only with the surface analysis that we get the second approximation, i.e., the undulated surface. Undulated surface is a name for composition of consecutive second order convex and concave patches. If one needs to verify whether the boundaries between the patches are continuous or if they are true surface boundaries, one would have to perform further tests, such as fit the data to higher order surface.

Another example of this sort can be a case of one cavity in otherwise convex object. This happens if the camera is looking perpendicularly on the cavity. The cavity can be modeled two ways: one as a negative volumetric part, the other as a combination of bending and rotating, an example of the two cases is presented in Figure 4a and b. The difference between the two is in the magnitude of the rest of the data. What we mean is this: if the remaining data is bulky then it seems more natural to explain the cavity as a negative volume (bulk implies volume). An example of this type is the block with a cut-out half circular cylinder. On the other hand, if the remaining data is more like a disk (two dimensions are much bigger than the remaining third one) then the bending plus a rotating operation seems more appropriate (case of a bowl). In order to distinguish between the two cases again one needs to consider at least one other view.

Recognition of a hole comes from a combination of the surface residual and the contour of the hole must be closed .

Analysis of significant residual of the occluded contour.

Here again two cases are:

- overestimation, where the model covers more than what data is, i.e., missing data.
- underestimation where the model covers less than where the data is, i.e., excess data.

The overestimation case:

The first test is the magnitude by which the model exceeds the data, both at the contour and the surface level. If this magnitude is small then the description is adequate. If it is bigger than a threshold then search in a radial fashion for the nearest concavity of the contour. Since we know a priori that the volumetric primitives are convex objects, we use this fact to follow the contour of the object until the next concavity on the contour. That is the breaking point; There must be at least two convex points between the two concave points so that the segmentation can lead to primitives. In the case where there are two concave points following each other, the occluding segment is pushed in perpendicular direction until the next boundary is found. These heuristics follow again from the assumption about convex and symmetric primitives. It could be the case that the next boundary will not be symmetric. This can happen either because the true boundary is not symmetric or because of missing data due to the angle between the laser and the camera which receives the reflected laser stripe (the shadow problem). In order to decide which is which we must invoke the camera, and scan the object from an angle that will confirm one case against the other. If after the new data acquisition, the boundary is still asymmetric, there is no choice but to segment the data into two or more convex and symmetric parts. Using this strategy, recursively remove one by one the segmented data and fit it individually to proper superquadric primitives.

After every removal of the segmented data, the remaining data is refit to a new superquadric model. Check if there is any underestimation. If there is none, then one can apply the segmentation fit procedure to the part without refitting the remaining data. However, if there is some data which is underestimated then one must include it into the recursive process, refitting that portion as well. When all the data fits in the models, the process terminates.
In summary, the basic strategy is first to examine the undersegmented data with continuous rechecking and fitting of the not explained data.

4 Two dimensional segmentation process

Gray scale segmentation of a scene as a problem of Computer Vision has existed for about 20 years. So the question can be raised, another paper on segmentation - what can be new? The problem as we see it, is that so far most of the criteria of what is considered a "good segmentation" are subjective, based on people's perception and interpretation. However if Machine Vision is going to be a module that delivers its output to another mechanical device and/or module, such as a manipulator, or an autonomous vehicle, then the output of the segmentation process must be well defined, parametrized, quantified and measurable. So, the purpose of this section is an attempt to develop a theory for the 2D segmentation process.
For the discussion here we make the following assumptions:

1. We assume a stationary observer and non moving scene
2. We assume known illumination (diffused or point source with known direction)
3. We assume known distance between the observer and the observed scene
4. We are limiting ourselves to 2D segmentation only, in that boundaries must correspond to some change in intensity
5. We assume that all the conditions above are constant during the time of observation

Given the above assumptions, the goal of this part of image processing is to produce a segmented image, where segments correspond to visually coherent, monotonically changing surfaces. The reflectance could be piecewise constant or linear, but no texture and geometrically meaningful units, meaning that the regions are enclosed. The segmentation process as defined above can be stated also as finding the partitioning of the data into equivalence classes, where the equivalence relationship is the homogeneity measure together with the constraints given by the external parameters. We propose to model segmentation process by the flow diagram in Figure 5.
In order to be able to evaluate the segmentation process one must have a model, or a form of data decomposition. For us the ingredients of the model are a set of spatial scales . The values of scale are discrete and vary by powers of two. This part is built on the basis of the Wavelet Representation developed by Mallat [mal88], who has shown that the one dimensional signal as well as the two dimensional image are completely represented by a sequence of Wavelet representations. Using the Wavelet representation, Mallat has derived another type of representation of the band-pass filtered images based on their zero-crossings and the signal between the consecutive zero-crossings, see [tre,mal,baj89], parametrized with respect to a set of different orientations. The type of signal

between the consecutive zero-crossings can be: constant, linear and quadratic. This is consistent with our 3D surface primitives.
When edge information is associated across different scales, it is possible to separate texture from border edges. The texture edges have high spatial frequency content but not low spatial frequency content. Shading has low frequency content but not high. Borders have frequency content at all scales.
In summary, we shall have the following parameters: scales, orientations, the signal type and its magnitude. One external parameter to the theory is the signal to noise ratio, which comes from the camera characteristics. This last parameter will determine the detection threshold. The lowest frequency, the largest scale is given of course by magnification of the optics of the camera and the nyquist characteristics. The highest frequency, the smallest scale, is determined by the spatial resolution of the CCD chip. If from the task one can obtain more accurate bounds for the largest and smallest scale, it is desirable to do so, because of saving of processing time.
Finally, we have only those external parameters which come from the devices, i.e. the camera noise, camera magnification and spatial resolution.
In the past most work in image segmentation has been using either edge based methods or region growing methods [har,shap85]. We have recognized for some time that these two processes are not independent and should be considered together [and 88]. However we as others have still aplied them independently. The new approach does not separate the region growing from edges i.e., the signal between two zero-crossing. Rather the edges are used as markings of discontinuity on the signal. The considerations of all scales provide a natural data driven selection for different granularity of the segmentation process. This is shown in Figure 6 on one dimensional signal. The work on two dimensional signals is in progress.

5 Conclusion

Scene segmentation is still an art rather than science. We have tried in this paper to introduce some analytic methodology into segmentation. Firstly, we claim that unless one commits to some primitives, i.e the vocabulary of the segmented signal, one has no chance of evaluating the performance of the segmentation process. Of course we recognize that by doing so, that is committing ourselves to some primitives, we will have errors, that is oversegementation and/or undersegmentation. We argue, however, that this is not so bad, providing that one recognizes these two situations and acts upon them. Secondly, we assert that geometric primitives are well justified in 3D but not for 2D objects because the projective transformation that takes the 3D shape and maps it to many 2D possible shapes. Based on this argument we pursue on 2D signal reflectance/color primitives and 2D shape comes into play only as descriptors of the boundary but not of the shape of the region. Thirdly, in the spirit of Active Perception, we sustain that segmentation is an active process, that is the process is driven by the task. The task determines at what level of details and accuracy the segmentation may stop.

References

[baj,tsik87] Bajcsy, R., & Tsikos, C. (1987). Perception via Manipulation. Proceedings of the 4th International Robotics Symposium on Robotics Research, Santa Cruz, CA. MIT Press, pp. 199-206.

[baj,sol,gup88] Bajcsy, R., Solina, F. & Gupta, A. (1988). Segmentation versus Object Representation - are they separable? Technical Report MS-CS-88-58. University of Pennsylvania, Philadelphia, PA 19104.

[gup89] Gupta, A. (1989). Part Description and Segmentation using Contour, Surface and Volumetric Primitives. Technical Report MS-CS-CS-89-33. University of Pennsylvania, Philadelphia, PA 19104.

[and,baj,min88] Anderson, H., Bajcsy, R. & Mintz, M. (1988). Adaptive Image Segmentation. Technical Report MS-CS-88-26. University of Pennsylvania, Philadelphia, PA 19104.

[mal88] Mallat, S. (1988). Multiresolution Representations and Wavelets. Ph.D Thesis. University of Pennsylvania, Department of Computer and Information Science, Philadelphia, PA 19104.

[har,shap85] Haralick,R. & Shapiro, L.G. (1985). Survey of Image Segmentation Techniques. Computer Vision, Graphics, and Image Processing, Vol. 29, pp. 100-132.

[tre,mal,baj89] Treil, N., Mallat, S. & Bajcsy, R. (1989). Image Wavelet Decomposition and Applications. Technical Report MS-CIS-89-22. University of Pennsylvania, Philadelphia, PA 19104.

[tsik87] Tsikos, C. (1987). Segmentation of 3D scene using Multi-Modal Interaction between Machine Vision and Programmable Mechanical scene manipulation. Ph.D Dissertation, Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA 19104.

[sol87] Solina, F. (1987). Shape Recovery and Segmentation with Deformable Part Models. Ph.D dissertation, University of Pennsylvania, Philadelphia, PA 19104.

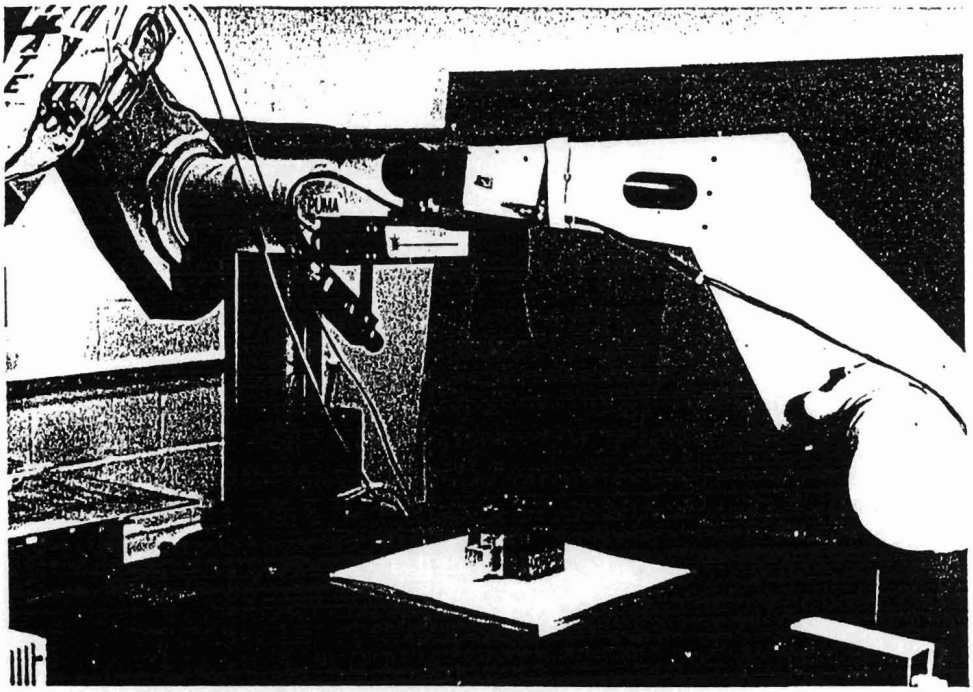


Figure 1: Laser Range Scanner

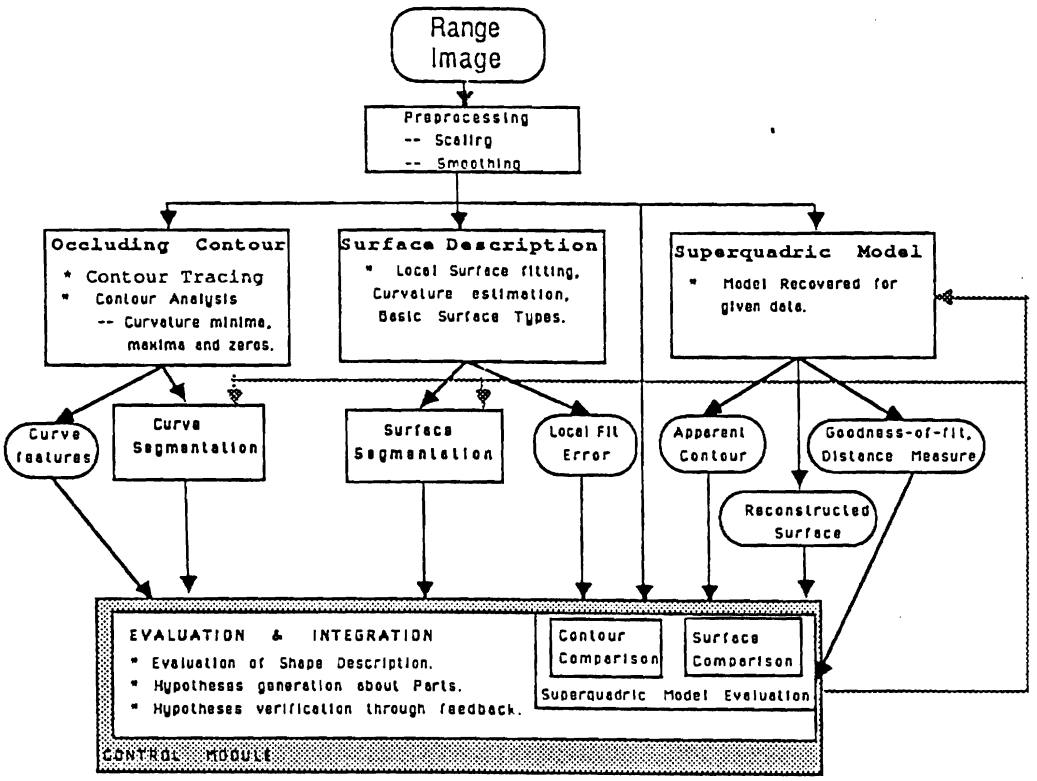


Figure 2: Detailed Block Diagram of our Approach for 3-D Segmentation

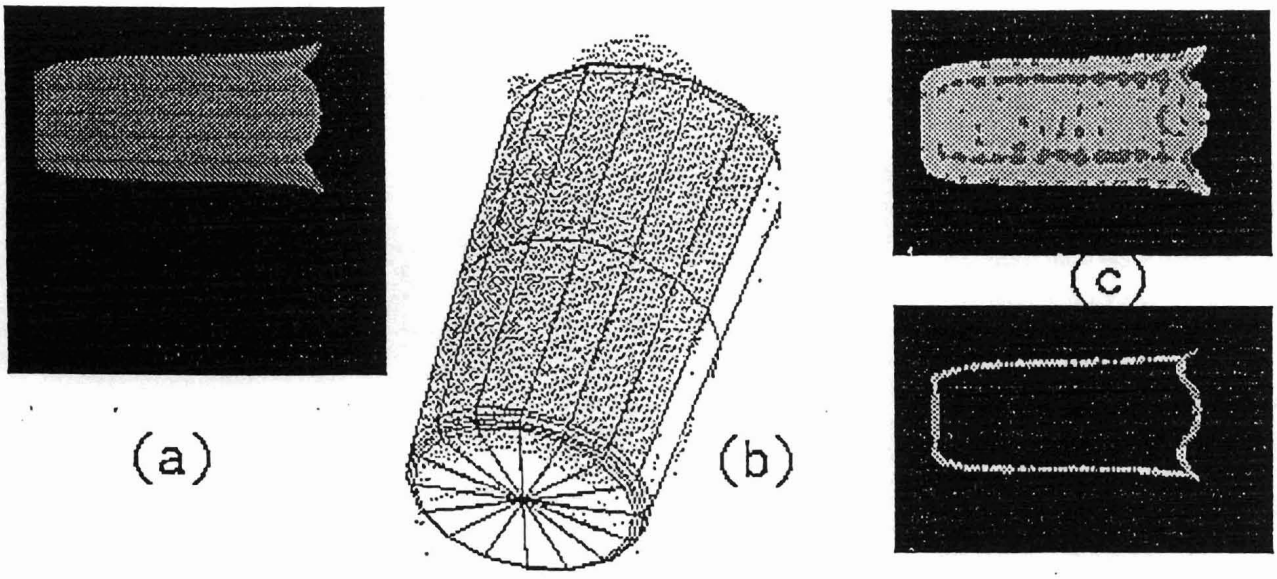


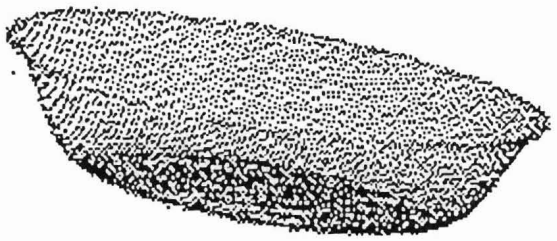
Figure 3: Analysis of a Vase:

a) original range image

b) superquadric model recovered for the vase. A tapered cylinder gives acceptable volumetric approximation of the vase.

c) sign of the Gaussian (bottom) and mean (top) curvature: mean curvature map indicates presence of 3 convex regions separated by two concave regions (boundaries of the convex patches). Zero Gaussian curvature on the vase shows that the patches are cylindrical. Three second-order surfaces can be used to describe the convex patches at the surface level.

(a)

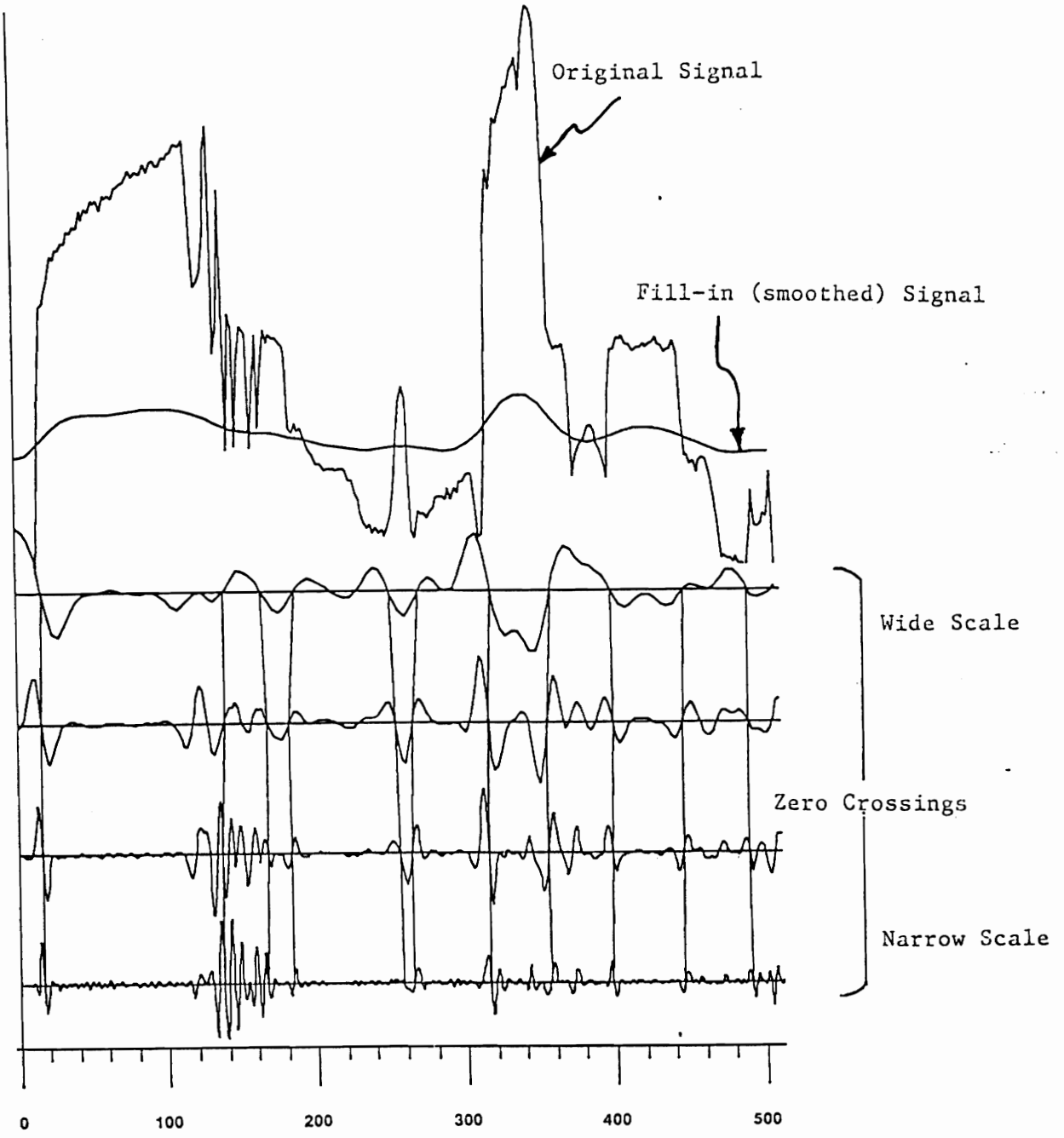


(b)

Figure 4a & 4b

a) Range points of an arch

b) Range points of a bowl



Multi-scale Decomposition of 1D Signal

Figure 6

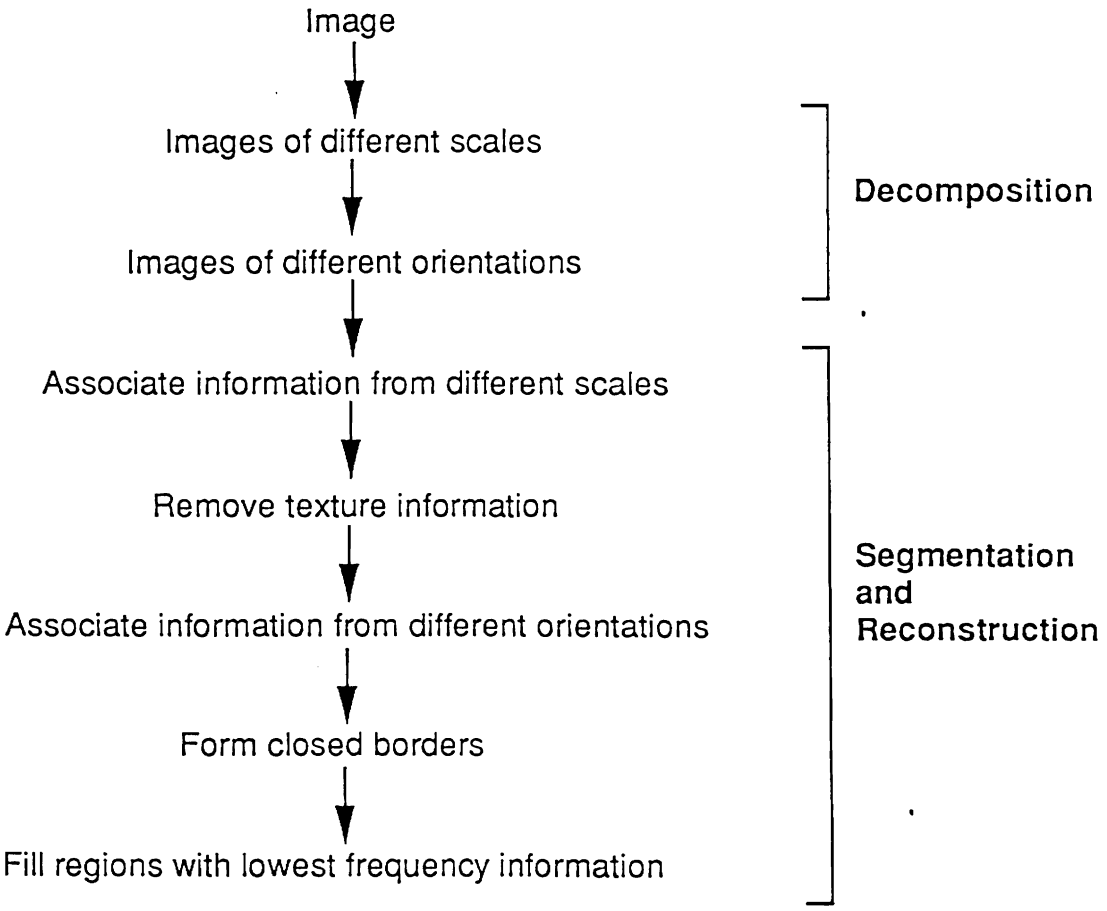


Figure 5: Signal decomposition using wavelets