Multiscale Methods and the Segmentation of Medical Images

TR88-051

November 1988

Stephen M. Pizer†

The University of North Carolina at Chapel Hill
Department of Computer Science
CB#3175, Sitterson Hall
Chapel Hill, NC 27599-3175

†Departments of Computer Science, Radiology, and Radiation Oncology, UNC

MULTISCALE METHODS
AND THE SEGMENTATION OF MEDICAL IMAGES

Stephen M. Pizer

Depts. of Computer Science, Radiology, and Radiation Oncology
University of North Carolina
Chapel Hill, NC 27599-3175, USA

ABSTRACT

Multiscale methods analyze an image via relationships between its properties at many different levels of spatial scale. Details and noise appear largely at small scale, and global properties of image objects appear at large scale. The segmentation of images into objects or coherent regions is therefore aided by viewing the image in multiscale terms. The methods and approaches that have been suggested for multiscale image analysis and for segmentation based on this analysis are summarized.

1. INTRODUCTION

Multiscale approaches for dividing an image into objects (segmentation) are based on the idea that the identifying features of an object or any coherent image region exist simultaneously at many levels of spatial scale. Global properties of the object are dominant at large scale, and at smaller scales details of size appropriate to that scale are principally represented. Thus, an oak tree consists of a trunk and a treetop at large scale, of limbs at a smaller scale, of branches and leaves at a yet smaller scale, and of leaf veins and indentations at a quite small scale. Defining the part of an image corresponding to an object therefore requires examining the image simultaneously at many levels of spatial scale. There is evidence that the human visual system operates in such a multiscale manner (Young, 1986).

Moreover, an image viewed at large scale is much simpler than one viewed at smaller scales. Therefore, for efficiency object recognition should occur in a sort of top-down manner, finding things tentatively at large scale where the image is simple, and then verifying and refining the recognition at smaller scales where details are better represented.

The computation involved in multiresolution segmentation thus requires a means of successive simplification of the image by increasing in scale, examining image features at various scales, and actually defining the objects in terms of the features found at the respective scales. The following chapters treat these matters in turn.

2. MEANS OF IMAGE SIMPLIFICATION

2.1. Pyramid methods.

Historically the first multiresolution methods were pyramid approaches. These depend on forming new larger scale pixels each of which combines the information from image pixels into square groups of m X m, thus creating a summary of the image with fewer pixels (by $1/m^2$ -- as shown in Fig. 1). This larger scale image is in turn simplified by combining its pixels into groups of m X m, creating a yet larger scale image. This process is repeated to produce what can be viewed as a pyramid of image representations if the successive summaries are piled one on top of the other.

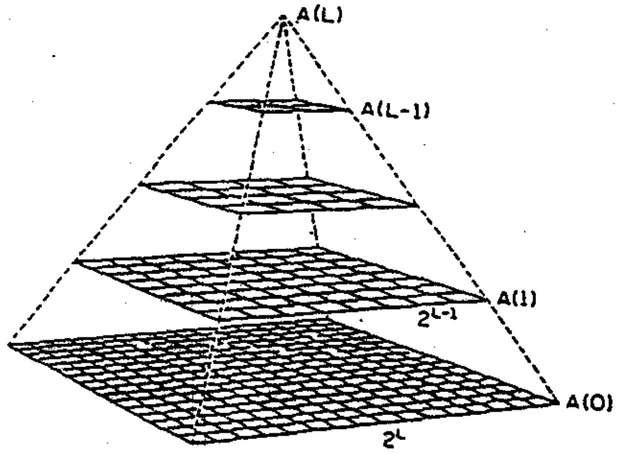

Figure 1: Pyramid. Compliments of Lawrence Lifshitz (1987).

Different pyramid methods vary by the means of combining image pixels into groups, i.e., by the information that is recorded in the parent pixels. Two general categories of pyramid methods can be defined: feature following and summary, and wavelet decomposition.

In feature following and summary (Burt, 1981; Rosenfeld, 1987) a parent pixel summarizes the information about some feature in the pixels that are its children at the next lower level of scale. For example, if the feature of interest was image intensity, the parent might record the average intensity of its children; or if the feature of interest was a line segment, segments described in some of the child pixels would be summarized into a global segment in the parent pixel.

In wavelet decomposition (Jaffard, 1986; Mallat, 1987) the information at smaller scale gives only the difference between the image analyzed at that scale and the image analyzed at the larger scale. This provides a space-efficient representation. By an elegant development, the information at each scale is represented by orthogonal functions that are band-limited in both the space and frequency domains.

2.2. Blurring methods.

Another way to simplify the image is to blur it by convolution with a blurring kernel. A number of authors (Yuille, 1983; Witkin, 1983; Koenderink, 1984) have shown that the kernel that to the greatest degree limits the creation of new structure under blurring is the Gaussian; this kernel also has many advantages of compatibility with the measured receptive fields of the human visual system, being simultaneously isotropic and separable by coordinates, and holding its form under cascaded operation (Koenderink, 1989). As a result there has been considerable attention (Pizer, 1988b; Koenderink, 1984; Bergholm, 1987; Stiehl, 1988) to image segmentation by following features through Gaussian blurring.

It has been suggested that non-isotropic or non-stationary Gaussian blurring can allow the blurring to adjust to the orientations of and spacings between objects. This adds a considerable complication to the analysis and leads to iterative methods where tentative segmentation suggests the parameters of reblurring. The potential strengths of such iteration, as suggested by the human visual system, has led to initial consideration of such an approach (Lifshitz, 1988) and some trials indicating some success of this idea (Hsieh, 1988b). A related approach with some promise creates non-Gaussian summaries over areas grouped according to consistent geometric properties such as intensity gradient direction (Colchester, 1988).

Attention to edges suggests following the Laplacian of the image through Gaussian blurring. This idea can be combined with resampling with fewer pixels to produce the so-called Laplacian pyramid (Burt, 1984). Since D*(B*Image) = (D*B)*Image for any differential operator D and blurring kernel B, the Laplacian of a blurred image can be found by applying to the image the operator that is the Laplacian of a Gaussian. The Laplacian of a Gaussian is approximately equal to the difference of that Gaussian (DOG) with a Gaussian with standard deviation approximately 1.6 times as large. If one uses $\sqrt{2}$ in place of 1.6, one can efficiently produce a multiresolution image sequence from a series of DOG applications: G_i * Image - G_{i+1}* Image, where $G_{i+1} = G_i * G_i$ (Crowley, 1984). Crowley has used this idea to produce a series of images resulting from successive DOG filters, each with the same energy but covering a successively lower frequency passband.

2.3. Discretization and finiteness.

In actual computation both the space and scale will be discrete, so the spatial discretization (pixel spacing) and scale discretization (spacing between levels of blurring: for Gaussian blurring, the value of s) must be decided. Sampling theory argues that the pixel spacing should be proportional to the scale, but a successive decrease in number of pixels covering an image forces the decision of how to map locations at higher scale to the new scale. Since the image at a larger scale will be simplifed, the representation of that image can frequently be made efficient of computer memory without increasing the pixel spacing.

Computational ease has led to many pyramid methods increasing the pixel spacing by 2 in each coordinate dimension at each scale step. This is the scale increase used in the wavelet decomposition. Crowley has shown how taking advantage of diagonal distances can allow efficient computing of the Laplacian pyramid with a scale increase of $\sqrt{2}$ at each scale step.

Koenderink (1984) and Pizer (Lifshitz, 1987) have suggested that the amount of blurring between successive discrete scales be determined by limiting the change in intensity of some basis function as it is blurred between successive discrete scales. They limit the intensity change to an amount proportional to the accuracy of the underlying floating-point intensities undergoing blurring.

Such analysis concludes that increasing the scale by a constant factor between steps is correct, but a factor of 2 or even $\sqrt{2}$ produces too much change in the basis function intensity between steps.

In contrast, some authors have recognized that the essential issue in scale discreteness is how it supports the following of image features that is involved in multiresolution methods. Lifshitz (1988) has suggested that if one is following features such as intensity extrema through scale space, with the intention that annihilations of one feature into another are to be discovered, large successive blurrings that are related to the spacing between the features should be used. Similarly, Bergholm (1987) has noted that if one is following a feature such as an edge from an optimal scale for recognition through successively smaller scales to its location in the in the original image, it is reasonable to limit the successive blurrings so that the feature movement is limited to a fixed distance. Bergholm assumes a fixed minimum corner angle of edges and then chooses his blurring to limit edge movement to 1 pixel.

The method for following features through scales can equivalently be thought of as the need to identify a multiscale form, e.g. (as shown in Fig. 2), an extremal path (a point extended through a part of the scale dimension) (Koenderink, 1984; Lifshitz, 1988) or a Laplacian zero crossing surface (a curve extended through part of the scale dimension). This form can be determined, either directly or by relaxation, after the corresponding single-scale form (e.g., extremum of intensity or zero of Laplacian, respectively) has been identified at each scale, but artifacts arising due to different discrete approximations at each scale can make difficult the determination of cross-scale coherences. Instead one can treat the coherence across scales as part of the identification process. In this approach, inspired by the "Snakes" approach of Terzopoulos, Witkin, and Kass (1987), a multidimensional form, such as a surface, with one of its dimensions being scale, is fit to the family of multiscale images in such a way as to maximize a combination of coherence of the form and fit to whatever image properties correspond to the feature being followed (e.g., the derivative properties corresponding to extrema of intensity or zero-crossing of the Laplacian, respectively) (Gauch, 1988b).

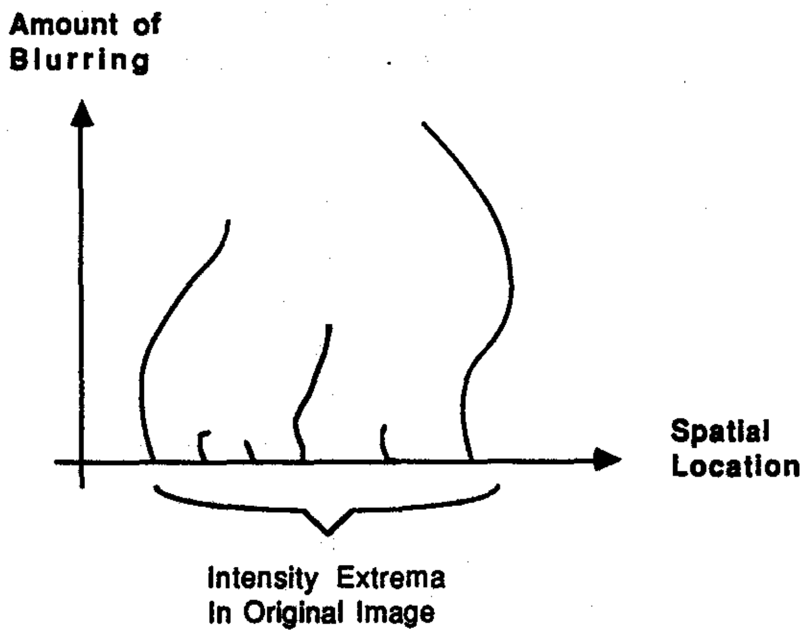

Figure 2: Extremal paths through scale. Compliments of Lawrence Lifshitz (1987).

In addition to discreteness, the finiteness of the image can cause problems when scale increase is accomplished by Gaussian blurring. The problem is to continue the finite image across infinite space so as to achieve desired behavior under scale reduction. Reflection across boundaries or the implicit wraparound continuation of frequency domain analysis does not allow the image to simplify fully as scale is increased without bound. Toet (1984) suggested that seeing Gaussian blurring as applying the diffusion equation leads to the solution. Based on this idea, Lifshitz (1987) modifies the image by subtracting a solution to the diffusion equation that at all levels of scale agrees with the original image at its boundary. Applying multiscale analysis to the result, extended everywhere outside the image by zeroes, can now go ahead.

3. USING VARIOUS SCALES

There are a number of strategies for using the images at multiple scales. Many authors (e.g., Coggins, 1986a; Neumann, 1988; Burt, 1981) try to find the optimal scale for a particular decision. For example, Laplacian zero-crossings can be used to find edges at an optimal resolution, even in 3D (Bomans, 1987). Similarly, in stereo matching, a scale that sets the stereo disparity at a given location to a fixed number of pixels is correct for that location (Hoff, 1988). In segmentation, the decision that a region is a sensible segment is best made at the scale of that segment. In these methods the decisions are frequently taken in decreasing order of scale, but it is also possible to go in reverse order, choosing the optimal scale when one comes to a scale where certain conditions first fail.

Another strategy for using images at multiple scales is to use each larger scale to summarize at each image location the information from the children at lower scales (Rosenfeld, 1987; Meer, 1986; Sher, 1987). In the simplest form of this method, intensity itself might be summarized by averaging over pixels that at the next lowest level of scale have been associated with the parent pixel being computed. The various methods (Burt, 1981) differ by the way in which these associations are determined. For example, one might associate a pixel with nearby potential children that are closer in intensity to its present intensity than any other potential parent. After these associations are determined, the parents' intensities are recomputed and the associations redetermined, and this process repeats until no associations change. After these associations have been set, a region can be determined as all the descendants of a pixel that is not close enough in intensity to any of its potential parents.

Instead of summarizing a feature as simple as intensity, one might follow a geometric feature such as an edge curve or an intensity trough, or a texture through scale space (Sher, 1987). Thus, for example, two approximately collinear line segments might be summarized into the larger line segment formed by their combination. In this case, also, there must be a means of choosing the pixels at the next lower scale that the present pixel is to summarize. Even given this, the approach frequently has the difficulty of finding a means of summarizing the information from the next smaller scale without the storage for a pixel increasing unmanageably with increasing scale.

Another class of methods takes advantage of the fact that as the image simplifies under increase in scale, component features can be expected to disappear into the more "important" features of which they are a part (if the features are correctly chosen to characterize image structure of interest).

The order of disappearance and the way in which image pixels collapse into structure-defining geometric forms can be used to define regions and impose a hierarchy on them (Koenderink, 1984; Pizer, 1988b). For example (as shown in Fig. 3), Lifshitz (1988) and Koenderink have followed intensity extrema to annihilation while following intensity values as they collapse into extrema, while Gauch (1988a,b) and Blom (1988a) are following to annihilation ridges and troughs defined by iso-intensity curve vertices.

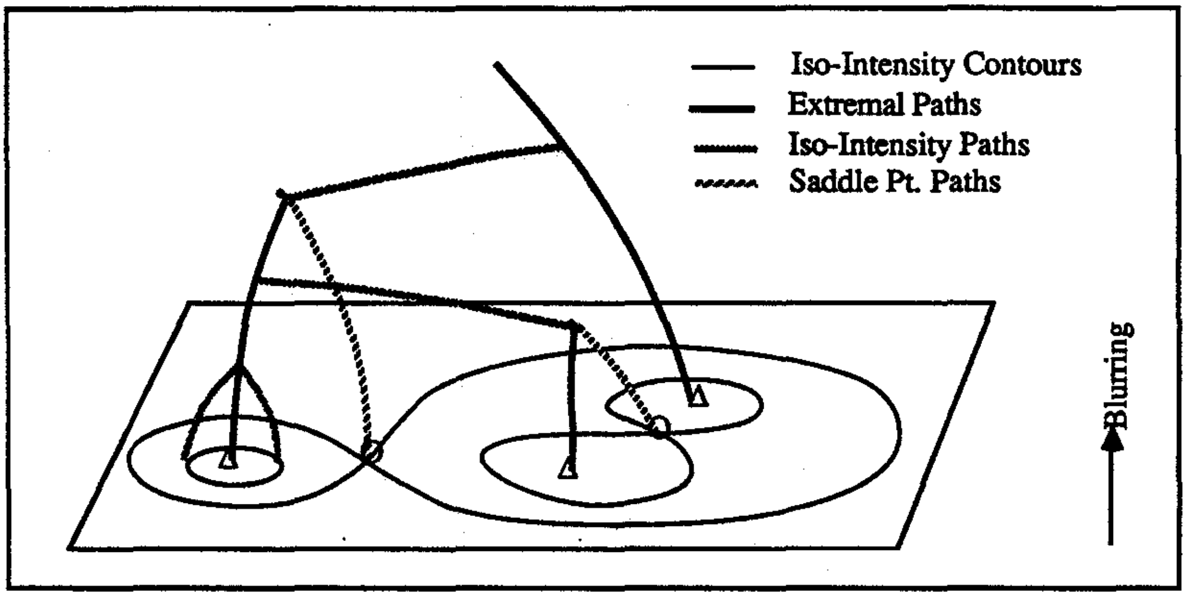

Figure 3: Extremal paths and iso-intensity paths through scale space. Note that maxima (Δ), and saddle points (O) move together and annihilate. The resulting non-extremal point (•) is then linked via an iso-intensity path to another extremal path.

The final member of our list of approaches to use multiple resolutions attempts to identify image features, such as edges or ridges, by their local behavior through the full family of resolutions. For example, a pixel on a linear edge can be identified by its sequence of values of intensity and derivatives as the image is blurred (Korn, 1985; Back, 1988; Blom, 1988a), or the occurrence of a particular geometric shape centered at a pixel can be identified by its response under a particular filtering as the scale parameter of that filter goes over some range (Coggins, 1986b). Furthermore, information about the extent of an edge, its curvature, and its distance from other edges can be obtained by the way in which this sequence of values varies from the sequence that would correspond to an infinite linear edge.

4. GEOMETRIC FEATURES IN OBJECTS AND INTENSITY AND TIME FAMILIES

The success of a multiresolution method depends not only on the strategy of using the multiple resolutions but also on the geometric forms that are examined at each resolution. The forms are chosen to capture image structure of interest while at the same time having appropriate behavior as resolution is lowered: e.g., capable of being summarized, or smoothly moving to annihilation. The choice of these forms requires a good understanding of the mathematical discipline of geometry (Koenderink, 1989). This geometry may well involve not only the two or three spatial dimensions of the image and possibly the dimension of time in a time-series of images, but also the dimensions of intensity and scale. The two-dimensional methods discussed below generalize to three spatial dimensions, though that generalization has in many cases not been worked out or tried out.

4.1. Intensity extrema.

Bright and dark spots are features of importance in an image. Furthermore, Morse theory tells us that under Gaussian blurring these intensity extrema undergo a regular behavior: each extremum moves smoothly, with maxima decreasing in intensity and minima increasing in intensity as scale is increased. This behavior continues until the extremum annihilates with a saddle point. Unfortunately, it is also possible, though uncommon, for extremum-saddle point pairs to be created at some level of blurring (Lifshitz, 1987).

It seems attractive to use extrema as seeds for extremal regions, locally bright or dark image areas. Computation of extremal regions requires a means of associating pixels with nearby extrema. Furthermore, it seems desirable to use extremal annihilation to induce a hierarchy, by seeing one extremum as disappearing under scale increase into the hillside or pitside of another extremum. To this end a means of association of annihilating extrema as being part of the extremal region of another extremum must be found. Koenderink (1984) and Lifshitz (1988) have dealt with this problem by following constant intensity levels through scale space. They show that each such iso-intensity path must run into an extremum, thereby associating either the pixel or the just annihilated extremum at the source of the path with the extremal region that contains it. There are two difficulties with this means of association. First, it leads to iso-intensity curves as region boundaries. Image features such as edges and intensity ridges and troughs play no part. Second, all the pixels inside an iso-intensity curve that forms a partial boundary are not necessarily associated with the extremum in question (Lifshitz, 1988).

4.2. Edges.

Object edges capture more geometry than extrema. Edge information is found in the derivatives of the image. Generally methods of edge calculation can be divided into those based on the gradient and those based on the Laplacian, though Koenderink and Blom have argued that it is useful to include information from derivatives of higher order than the second to determine edges (Koenderink, 1988b).

The gradient-based methods compute an edge strength from the magnitude of the gradient and a normal direction to the edge from the direction of the gradient. These values can be followed through Gaussian blurring (Koenderink, 1987). Hsieh (1988a) has combined the connectionist ideas of Grossberg (1985) for producing closed edges with a scheme that lets edges or edge continuations at one resolution support those at other resolutions, producing closed edge contours that frequently match what we see. Thus, in a multiscale fashion image locations cooperate and compete to select strong edges (and corners) with consistent edge directions, and continue edges across gaps (at multiple scales). Neumann (1988) adds line and channel features to edges in the Grossberg scheme, and attempts to identify these at optimal scale. Korn (1985) and Back (1988) attempt to find edges according to gradient-of-Gaussian edge strength by looking at the family of such edge strength measures across many scales. Bergholm 1987] is following the commonly used Canny edges, from a scale at which detection is straightforward but the edges are displaced, through decreasing scale to their actual location in the original image. Canny edges (1983) are locations with gradient magnitude above some threshold and of maximum gradient magnitude in the gradient direction. Pyramid methods have also been used with edge strengths.

Marr (1982), among others, has focused on the Laplacian as a direction-free indicator of edge strength. Near edges the Laplacian goes from a high positive value through zero to a high magnitude negative value. Marr has particularly focused on the closed contours given by Laplacian zero crossings, and it has been suggested to follow these contours through scale space to select important edges (Marr, 1980). While this frequently finds edges well, Blom (1987) has shown that all edges of interest do not correspond to a Laplacian zero crossing.

Edges, i.e. region boundaries, can be used in a multiscale way to describe objects, after these boundaries have been found. For example, Richards' (1985) codons can be followed through decrease in scale as one annihilates into the next (Gauch, 1988a). This type of analysis involves describing objects by deformation, the subject of section 4.4.

4.3 Central Axes.

While edges provide important object information, that information is rather local. The axis down the center of an object, together with the behavior of the width of the object about its axis as a function of axis position, gives more global information about the object. When put in a form that applies to intensity-varying images, such axes can summarize information about edges, spatial shape, and the shape of variations in the intensity dimension. Two different categories of central axes have received attention. The first involves actual axes of symmetry, and the second involves intensity ridges or troughs, which tend to run down the center of objects.

Three symmetry axes of objects have been suggested (as shown in Fig. 4); all are related to the the family of circles tangent at two locations on the object boundary. The locus of the circle centers forms the *symmetric* or *medial* axis (Blum, 1978). The internal medial axis, made from the centers of tangent circles entirely within the object, has the advantage, for objects without holes, of being a tree which when divided into its branches subdivides the object into bulges. Each branch endpoint is the center of a circle which touches the object in a second degree way at one point, a point of maximum positive curvature of the object boundary. Similarly, the object's external medial axis, defined as the internal medial axis of the object's complement, is the center of a circle touching the object at a point of minimum negative curvature. A 3D object's axial surface, defined in terms of tangent spheres, has much the same properties listed above for the medial axis of a 2D object (Nackman, 1985; Bloomberg, 1988).

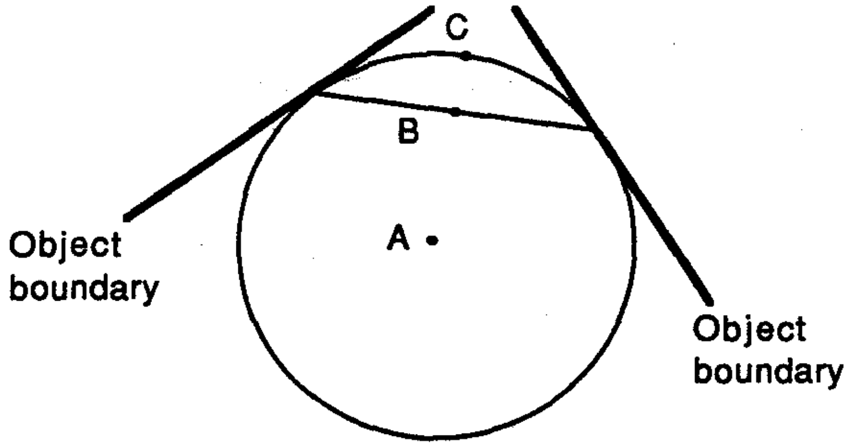

Figure 4: Points on axes of symmetry. Medial axis is locus of points of type A. Smooth local symmetry axis is locus of points of type B. Process-inferred symmetric axis is locus of points of type C.

The second axis, the locus of the centers of chords between the two tangent locations, forms the smooth local symmetry axis (Brady, 1984). It is not necessarily connected, even when it is restricted to internal circles. Brady has suggested combining this axis with the behavior of the object boundary curvature as it is followed through scale increase to produce object descriptions.

The third axis is the locus of center points of the smaller of the parts of the tangent circle between the two tangent points. Leyton (1988) calls it the *process-inferred symmetric axis* (PISA), because it touches the object on the side of the object where a deforming force can be thought to have been applied to make the object from a circle (or sphere, in 3D). We will return to the PISA in section 4.4, but here we will focus on the medial axis.

The problem is to find a form of the medial axis that applies to intensity-varying images, rather than to objects whose boundaries have already been determined (Gauch, 1988a,b; Pizer, 1988b). One can view the image as an intensity surface one dimension above the n-dimensional image. However, because intensity is incommensurate with the spatial dimensions, one cannot use a surface of symmetry in n+1 dimensions. Rather, think of the image as a single parameter family of intensity level curves, as in a terrain map. For each level curve, the medial axis can be computed and the result placed a its corresponding height. The resulting medial axis pile, that we will call the *intensity axis of symmetry* or *IAS*, can be shown to be made of branching sheets (as shown in Fig. 5), in some cases with loop branches. The IAS depends on the way the image was divided into an intensity family; in the above it was sliced along iso-intensity levels. An open area for study is the relative advantages of various means of slicing the intensity terrain.

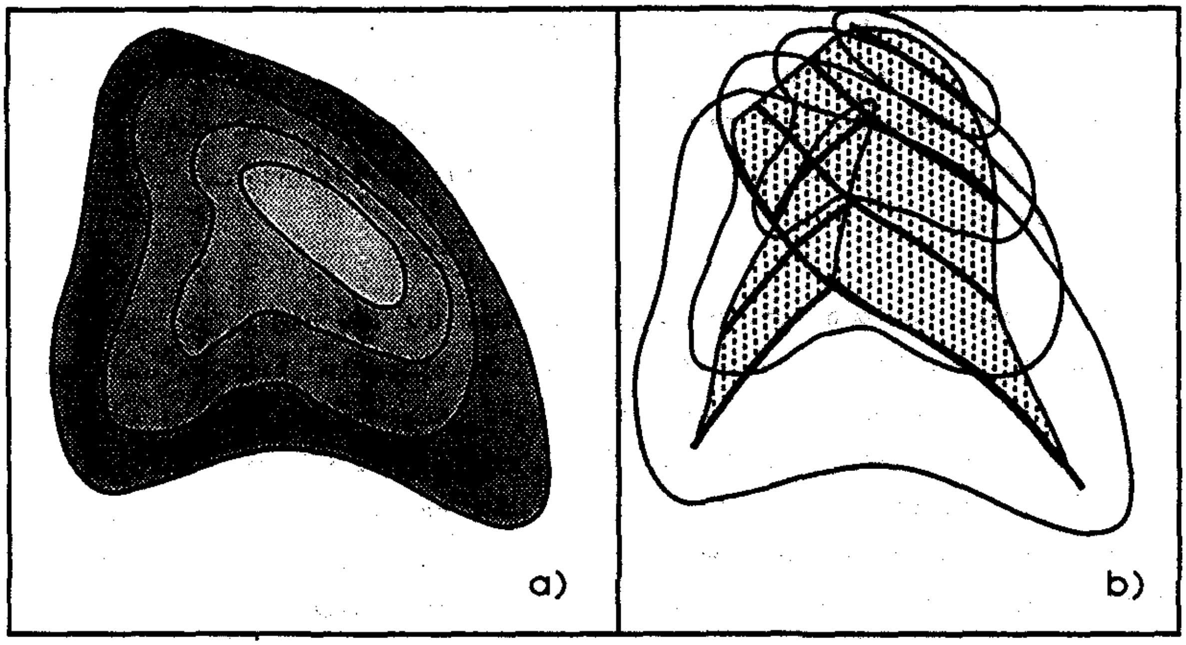

Figure 5: A simple intensity varying image and its IAS. a) 4 level curves of intensity; b) level curves and corresponding IAS.

Under Gaussian blurring, medial axis sheets annihilate into other sheets until a set of separate, simple (unbranching) sheets remains. Thus, the blurring process induces a set of hierarchies. Furthermore, associated with each subtree of sheets in this hierarchy is a subimage: at each pixel the intensity of the subimage is the maximum of the intensities of the IAS disks covering that pixel, where each IAS sheet point forms the center of a disk at the intensity of the point. That disk is the maximal disk corresponding to the medial axis point in question (for the particular level curve at that intensity). The way in which image regions should be formed by associating pixels absolutely or probabilistically with a sheet has yet to be determined. Presumably the association should take place via these subimages or geometric features such as surrounding intensity ridges or troughs It will not be surprising if the answer itself involves the multiscale following, back to the original scale, of a feature found near the annihilation scale of the IAS sheet determining the region.

Ridges and courses on the image terrain are features that not only tend to run down the middle of image objects but also have visual power. In the following, the many definitions of the general notion of ridge will be given, with the understanding that for each there is a corresponding notion of the complementary geometric form, the course. Full image analysis must lead to the possibility of light figures on a dark ground or the reverse, and simultaneous analysis in both polarities must be a part of any successful general segmentation process.

Crowley 1984] focuses on locations of high positive (or low negative) values of an energy-normalized Laplacian to find a sort of ridge in these values. He follows these ridges and peaks through scale space. Scale-space maxima in the magnitude of these peaks and pits designate the scale of a feature. While this method has a good intuitive basis, the mathematical behavior of the normalized Laplacians in scale space has not been worked out. Other methods with good intuitive basis also depend on cross-correlation of the image with some template function (Neumann, 1988).

Ridges can also be computed as watersheds, but the identification of watersheds is unfortunately nonlocal. That is, a change in the image at some distance away from a pixel can move a watershed from or to that pixel.

Other definitions of a ridge apply to any surface, independent of its orientation, and thus have the probably desirable property of being independent of the value of intensity or intensity slope. The best of these take advantage of the knowledge that the intensity surface forms a function, of space, i.e., is single-valued. One of particular interest is the locus of intensity level curve maxima of curvature, or *vertex curve,* because of its relation to the IAS: each vertex curve point is the touching point of the maximal circle corresponding to the endpoint of the medial axis of the intensity level curve at that intensity (as shown in Fig. 6). That is, there is a 1-1 relationship between vertex curves and IAS sheets. As a result, vertex curves can be followed through scale space to annihilation, a much simpler task computationally than following IAS sheets. The connectedness of the image is established by the branching of the IAS, which needs only to be computed for the original image, but the parent:child relationships are established by the order in which the vertex curves annihilate together with the relationship of the vertex curves to the IAS branches.

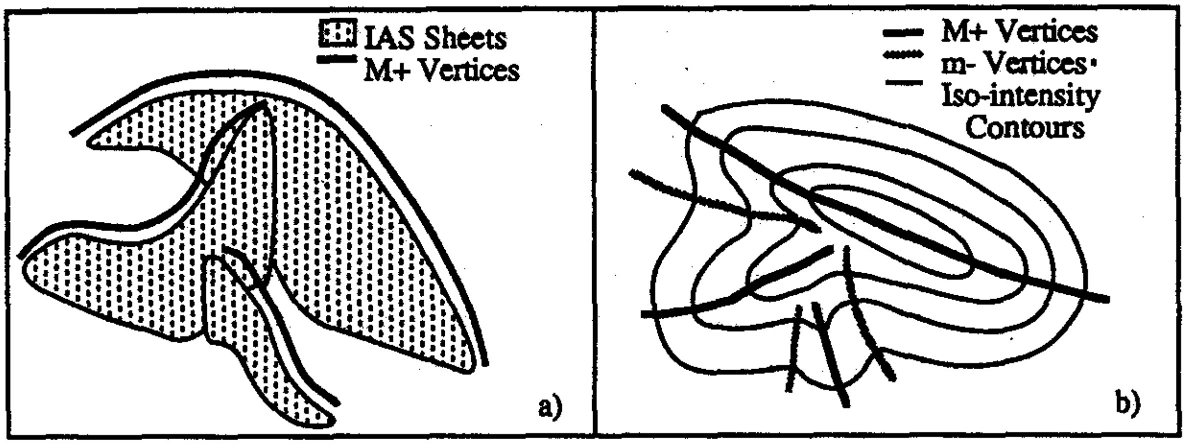

Figure 6: a) The relationship between the IAS and vertex curves; b) the same vertex curves shown superimposed on the intensity level curves. Vertices corresponding to positive curvature maxima (ridges) are indicated by M+, and vertices corresponding to negative curvature minima (courses) are indicated by m-.

4.4. Deformation.

Time series of images make it necessary to follow the moving images, and their component objects, through scale space. In general, the motion involves non-rigid deformation, rotation, and translation of the image objects, and this compound motion can be used to help define these objects.

The same issues of multiple scales that are important in the spatial dimensions are also important in the time dimension. However, time is not simply another spatial dimension; it is asymmetric (the past affects the future); it is incommensurate with space; yet space and time are interrelated, by pixel velocities. The means for space-time multiscale analysis using these ideas has not been adequately developed. Certain approximations may nevertheless be useful.

For example, analyzing the time course as if it were independent of space can yield useful information. Koenderink (1988a] has treated the question of how time should be increased in scale to analyze a time series, e.g. at a single pixel, in a multiresolution fashion. He concludes that a present moment, t_0, should be chosen as a parameter and the transformation from the asymmetric variable $t \leq t_0$ to the symmetric variable $s = \log(t_0 - t)$ be carried out. Then increase in the time scale by Gaussian blurring in s has the desired properties of not violating temporal causality and treating temporal intervals in proportion to their time in the past. This transformation and blurring in s can be combined with the ordinary Gaussian blurring in the spatial variables to produce a form of spatio-temporal scale increase, with independent scale parameters in the time and spatial dimensions. With this scale change one is analyzing the time series of images as a single parameter family of images and looking at geometric structures such as bulges (ridges) and indentations (courses) in this "pile" of images.

With such a scheme, if regions are segmented to produce objects, the deformation of these objects can be described in terms of closest distances from an object boundary (or some other special object feature) at one point in time to that boundary (or feature) as time is increased. However, this does not lead to a description of the deformation of the whole object.

The description of this deformation can itself be described in multiscale terms if the object boundaries at successive times have already been found. The description is produced by minimizing a measure of energy to deform the object at between the successive points in the time series. If an appropriate energy measure is used (Bookstein, 1988; Terzopoulos, 1983), a multiscale deformation description is produced. Bookstein has shown how thin-plate spline deformation, generalized to be applied independently *within* each spatial co-ordinate, provides an energy function that is a quadratic form. The simplicity of the quadratic form makes the energy minimization mathematically tractable. Furthermore, the eigenvectors of the matrix involved in the quadratic form define principal warps, each concentrated in a particular region whose size is inversely related to the magnitude of the corresponding eigenvalues. Since the final deformation is made up of a linear combination of these basis principal warps, the deformation can be thought to be decomposed into multiscale components.

Oliver and Bookstein (1988) have suggested how this energy can be combined with a translational energy term to give a form whose minimization will specify, for all points in an object, the deformation between the object at two times, given the object boundaries at the two times. Oliver suggests that matching the medial axis endpoints which survive after increase in scale can provide a useful initial boundary match for the iterative minimization of energy that determines the deformation.

This approach, and the one to follow, depends on knowing the object boundaries. Its generalization to the situation in which one starts with intensity-varying images has yet to be investigated.

Deformation can also be used as a means of object description. The idea is that an object is described by the way it is obtained by deformation from a primal form, normally a circle or avoid (Koenderink, 1986; Leyton, 1988). Leyton has shown how the PISA gives locations of maximum positive or minimum curvature at which these deformational forces are to be applied, and gives the direction of each force, as well. Lee (1988) is investigating how a hierarchy can be imposed on this set of forces, by multiscale analysis with Gaussian blurring of the object boundary, i.e., convolving a Gaussian in arc-length s with the boundary function (x(s), y(s)). This approach might be applicable to grey-scale images by examining the PISA IAS, that is, the intensity family of PISA axes on intensity level curves.

Koenderink (1986) has investigated multiscale shape description by deformation with three-dimensional objects and blurring of the object's characteristic function rather than the boundary. This analysis focuses on the genesis, under decrease in scale (deblurring), of surface regions of specified curvature: convex elliptic, concave elliptic, hyperbolic (saddle-shaped), and parabolic (curves of flatness separating elliptic and hyperbolic regions). This idea of object analysis by treating the object layout in space as an intensity-varying image given by a characteristic function has much power. This approach can be used to describe models of images, but there is little experience in the use of such model descriptions for image segmentation, so characteristic function blurring will not be discussed further here.

5. OBJECT DEFINITION APPROACHES

The approaches described in chapters 1-3 produce image descriptions by multiscale analysis. In this section means are discussed of using these descriptions to determine objects or object probabilities for each pixel. This object definition must involve not only these image descriptions based only on image structure, but also semantics of the image, i.e., knowledge of what objects and object groupings appear in the real world. The two basic approaches to object definition, automatic and user-driven, differ in the way this semantic information is provided. In automatic object definition the information is provided in models or prototypes stored in the computer, whereas in user-driven object definition the user utilizes his own knowledge of semantics in creating an object definition from the computer-generated image description.

Automatic recognition of objects from multiscale image description can be divided into two subcategories: statistical and structural. These two methodologies need not be mutually exclusive.

5.1. Object definition by statistical pattern recognition.

In an application of classical methods of statistical pattern recognition (Duda & Hart, 1973; Jain and Dubes, 1988), an m-vector of values for various local features is measured at each location, and objects of various types are associated with different regions of the m-space. This vector can include features computed at different scales, or a feature that is the optimal scale at that location, so this method can be implemented as a multiscale object definition strategy.

An example of methods of this type can be found in the work of Coggins 1985, 1986b]. Convolution with a sequence of m multiscale filters computes an m-vector for each pixel composed of the intensity levels of the m filtered images at corresponding locations in each filtered image. The pattern of responses to the filters at each pixel describes the pixel's relationship to its neighborhood. Scale extrema in the m-vector as well as spatial extrema in each filtered image provide useful information about the image's content. Some simple objects can be identified and measured directly from these patterns without an explicit image segmentation.

Coggins Coggins, 1986a; Packard, 1986] has shown not only the effectiveness of this idea, but also its generalization to maxima across other dimensions than scale, such as orientation. If the filter decomposition is nonspecific, more filters may be needed to obtain the same accuracy as with filters designed and tuned to the most critical aspects of the image. Also, the inference mechanism may need to be more complex since the object may be identified by the appearance of features at different scales and, in particular, relative spatial relationships to each other.

5.2. Object definition by structural pattern recognition.

Structural methods represent objects by a graph whose nodes describe object substructures and whose arcs describe relationships between these substructures. The substructures may in turn be described by graphs, but at some stage they consist of feature measurements or coded representations of primitive structures. Structural methods recognize objects by matching a graph describing a prototype object with the graph obtained from the image. Thus, object recognition maps into a variant of the graph isomorphism problem (Read, 1977). Syntactic pattern recognition (Fu, 1982) is a variant in which the graph is reduced to a sequence and the prototype is represented as a grammar. This approach maps object definition into the problem of parsing a language.

Structural approaches often founder either by being excessively time-consuming (graph isomorphism is an NP-complete problem) or by being error-prone because of the required explicit representation of structural aspects of the object. Structural methods are likely to fail if some relevant structural feature is omitted from the model, or if the placement of structures does not quite conform to that expected in the model, or if the image introduces visual structure (e.g., shadows) that is not part of the model.

Multiscale descriptions for producing the graph or string can help with these problems. First, if matching proceeds top-down by scale, coarse matches can not only be efficient themselves but can guide matches at lower scales, limiting the computational complexity of the matching. Second, since important items in the graph can be expected to appear at large scale, minor details that may not match correctly can affect the matching at small scale, where their low importance can be weighted lightly in determining the match.

There is a problem with this scheme when the hierarchical descriptions have the property that a region corresponding to a parent node does not simply represent the union of the subregions that are its children, but also an additional part (as shown in Fig. 7). This is a property of the hierarchies produced by following geometric structures to annihilation. With such hierarchies a small change in an object in an image can exchange the role of two components, one being part of a larger scale component and the other being a subcomponent (as shown in Fig. 7). A solution to this problem, somehow allowing both components to be seen as subcomponents, at least with some probability, must be found by future research.

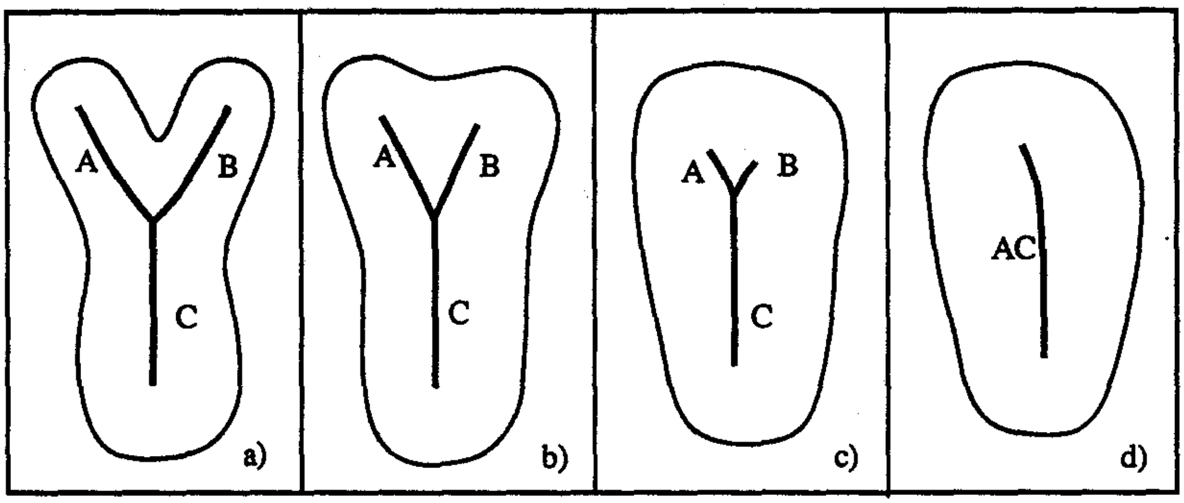

Figure 7: Effect of scale increase on figure 7a through stages b, c, and d illustrates the creation of a hierarchy with parent AC with child B. If region B is made slightly longer, scale increase will cause the branch A to annihilate sooner, so region A will appear in the description, and region B will not.

5.3. Object definition by a user via hierarchical descriptions.

There seems to be more hope, at least in the short term, of letting the human user define objects than having the computer do it automatically. For user-based object definition to be quick, the image description should be to a reasonable degree in human terms, thus allowing the human easily to manipulate the components of the description into objects that match the user's understanding of semantics. The hope of multiresolution image descriptions is that they will generate such descriptions.

Therefore, a multiscale image description should generate both image regions that are visually coherent to the user and an organization of those regions into a hierarchy or simple graph. If such a description were fully successful, the user would be able to pick semantically meaningful regions simply by pointing to a pixel in them and then, if the region is a subregion, asking for the next larger containing region an appropriate number of times.

It is reasonable to hope for a multiresolution image description largely to meet these needs, but it is probably realistic to expect that it would at least have the following kinds of faults:

1. Regions might have small numbers of extra or missing pixels at their boundaries.

2. A semantically sensible region might not exist in the description but rather be joined, e.g. by a narrow isthmus, with another semantically sensible region (that may or may not be in the description) to form a region that does appear in the description.

3. A semantically sensible region does not appear as a region in the description but rather is a union of regions appearing at various positions of the hierarchy. The need to combine these regions probably reflects an error in the hierarchy.

Lifshitz (Lifshitz, 1988; Pizer, 1988a) has shown in prototype and Coggins et al 1988] are developing further an interactive display tool with the following functions that can allow the fast definition of objects, even from descriptions that have these faults.

1. Given a region or set of regions selected from the image description, it displays that region and its relation to the the original image data.

2. Given a pixel (or voxel) pointed to by the user, it displays the smallest region in the image description containing that pixel; given a displayed region, under command it displays the parent (containing) region in the description hierarchy.

3. Given two or more selected regions, it displays the union or difference of the regions, and under command modifies the hierarchy to reflect this logical operation.

4. It allows hand editing of regions, and the splitting of a region into two, given the painting out of joining pixels.

It remains to be demonstrated whether such a tool together with multiresolution image descriptions discussed in this paper can provide convenient 2D and 3D segmentation of medical images.

6. SUMMARY

A variety of types of multiscale methods for image segmentation have been reviewed. While these methods have shown great promise and deserve attention because of their attractive conceptual properties and basis in approaches of the human visual system, intensive research on these methods has a history of only a few years. The ultimate promise of these methods and the bases of choice among them remain to be brought out by further research.

ACKNOWLEDGEMENTS

I gratefully acknowledge the helpful comments of Dr. James Coggins, Mr. John Gauch, Mr. Cheng-Hong Hsieh, and Dr. William Oliver and the continuing collaboration with Prof. Jan Koenderink. I thank John Gauch, Carolyn Din, Sharon Walters, and Sharon Laney for help in manuscript preparation. This paper was written with the partial support of NIH grant number P01 CA47982.

REFERENCES

Back, S "Konturextraktion für Roentgen-Computertomogramme nach dem Korn'schen Multi-scale-Ansatz," Diplomarbeit, Informatics Dept., Technical Univ. of Berlin, 1988.

Ballard, DH and Brown, CM *Computer Vision,* Prentice-Hall, Englewood Cliffs, NJ, 1982.

Bergholm, F "Edge Focusing," *IEEE Trans-PAMI, 9(6)*:726-741, 1987.

Blom, H Personal communication. Department of Medical and Physiological Physics, State University of Utrecht, 1988a.

Blom, H "Geometrical Description With a Jet Space Upon a Multiresolution Base," Internal Report, Department of Medical and Physiological Physics, State University of Utrecht, 1988b.

Blom, H Personal communication. Department of Medical and Physiological Physics, State University of Utrecht, 1987.

Bloomberg, S "Three-Dimensional Shape Description Using The Multiresolution Symmetric Axis Transform." Ph.D. Dissertation, University of North Carolina, Department of Computer Science, 1989.

Blum, H, and Nagel, RN "Shape Description Using Weighted Symmetric Axis Features." *Pattern Recognition 10*:167-180, 1978.

Bomans, M, Riemer, M, Tiede, U, Höhne, KH, "3D-Segmentation von Kernspin-Tomogrammen", *Informatik-Fachberichte 149*: 231-235, 1987.

Bookstein, F "Principal Warps: Thin-Plate Splines and the Decomposition of Deformations." Accepted for publication in *IEEE Trans-PAMI*, 1988.

Brady, M and Asada, H "Smoothed Local Symmetries and Their Implementation." Tech Report, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1984.

Burt, P, Hong, TH, Rosenfeld, A "Segmentation and Estimation of Image Region Properties Through Cooperative Hierarchical Computation." *IEEE Trans SMC 11*:802-809, 1981.

Burt, P "The Pyramid as a Structure for Efficient Computation." *Multiresolution Image Processing and Analysis*:6-35, Springer-Verlag, Berlin 1984.

Canny, JF "Finding Edges and Lines in Images." Artificial Intell. Lab., Massachusetts Inst. Technol., Cambridge, MA, Tech Report #720, 1983.

Coggins, J, Cullip, T, Pizer, S "A Data Structure for Image Region Hierarchies." Internal Report, Department of Computer Science, University of North Carolina, 1988.

Coggins, JM, Poole, JT "Printed Character Recognition Using an Artificial Visual System." *Proc of the 1986 IEEE International Conference on Systems, Man, and Cybernetics*, Atlanta, 1986a.

Coggins, JM, Fogarty, KE, Fay, FS "Development and Application of a Three-Dimensional Artificial Visual System." *Journal of Computer Methods and Programs in Biomedicine*, special issue of papers from the Ninth SCMAC, 22:69-77, 1986b.

Coggins, JM, Jain, AK "A Filtering Approach to Texture Analysis." *Pattern Recognition Letters 3*:195-203, 1985.

Colchester, A. et al: "Detection and Characterization of Blood Vessels on Angiograms Using Maximum-Gradient Profiles," this volume, 1988.

Crowley, J & Parker, A "A Representation for Shape Based on Peaks & Ridges in the Difference of Low-Pass Transform," *IEEE Trans. PAMI, 2(2)*:156-169, 1984.

Duda, R and Hart, P *Pattern Classification and Scene Analysis*, Wiley-Interscience Publication, John Wiley & Sons, 1973.

Fu, KS *Syntactic Pattern Recognition and Applications*, Englewood Cliffs, NJ: Prentice Hall, 1982.

Gauch, J, Oliver, W, Pizer, S, "Multiresolution Shape Descriptions and Their Applications in Medical Imaging." *Information Processing in Medical Imaging* (IPMI X, June 1987): 131-150, Plenum, New York,1988a.

Gauch, J and Pizer S, "Image Description via the Multiresolution Intensity Axis of Symmetry," *Proc. 2nd Int. Conf. on Comp. Vis.*, IEEE, 1988b.

Grossberg, S and Mingolla, E "Neural Dynamics of Perceptual Grouping: Textures, Boundaries, and Emergent Segmentations." *Perception and Psychophysics 38(2)*:141-171, 1985.

Hoff, W "Multi-Resolution Techniques in Stereo Vision." To appear in *IEEE Trans-PAMI*, 1988.

Hsieh, CH "A Connectionist Algorithm for Image Segmentation", Ph.D. Dissertation, University of North Carolina, Department of Computer Science, 1988a.

Hsieh, CH "A Connectionist Model for Corner Detection," Internal Report, University of North Carolina, Department of Computer Science, 1988b.

Jain, AK, Dubes, RC *Algorithms for Clustering Data*, Englewood Cliffs, NJ: Prentice Hall, 1988.

Jolion, J, Meer, P, Rosenfeld, A "Border Delineation in Image Pyramids By Concurrent Tree Growing." Tech Report #349, Center for Automation Research, University of Maryland, 1988.

Jaffard, S, Lemarie, PG, Mallat, S, & Meyer, Y "Multiscale Analysis," Internal Report, Centre Math., Ecole Polytechnique, Paris, 1986.

Koenderink, J.J., *Solid Shape*, MIT Press, Cambridge,1989.

Koenderink, JJ "Scale-Time," *Biological Cybernetics, 58*:159-62, January 1988a.

Koenderink, JJ "Operational Significance of Receptive Field Assemblies," *Biological Cybernetics, 58*:163-71, 1988b.

Koenderink, J & van Doorn, AJ "Representation of Local Geometry in the Visual System," *Biological Cybernetics, 55*:367-375, 1987.

Koenderink, J & van Doorn, AJ "Dynamic Shape," *Biological Cybernetics, 53*:383-396, 1986.

Koenderink, J "The Structure of Images," *Biological Cybernetics, 50* 363-370, 1984.

Korn, A "Combination of Different Space Frequency Filters for Modeling Edges and Surfaces in Gray-value Picctures." *SPIE 595*:22-30, *Computer Vision for Robots*, 1985.

Lee, SJ Personal communicaton. University of North Carolina, Department of Computer Science, 1988

Leyton, M "A Process-Grammar for Shape." *Artificial Intelligence 34* 213-247, 1988.

Lifshitz, L & Pizer, S "A Multiresolution Hierarchical Approach to Image Segmentation Based on Intensity Extrema," *Information Processing In Medical Imaging* (IPMI X, June 1987):107-30, Plenum, 1988.

Lifshitz, LM *Image Segmentation Using Global Knowledge and A Priori Information*, Ph.D. Dissertation, UNC Chapel Hill Technical Report 87-012, 1987.

Mallat, S "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," Tech Report MS-CIS-87-22, University of Pennsylvania, May 1987.

Marr, D *Vision*, Freeman, San Francisco, 1982.

Marr, D & Hildreth, EC "Theory of Edge Detection." *Proc R Soc Lond*, B 207:187-217, 1980.

Meer, P, Baugher, S, Rosenfeld, A "Hierarchical Processing of Multiscale Planar Curves." Tech Report #245, Center for Automation Research, University of Maryland, 1986.

Nackman, LR, Pizer, SM "Three-Dimensional Shape Description Using the Symmetric Transform I: Theory." *IEEE Trans-PAMI 7(2)*:187-202, 1985.

Neumann, H "Extraction of Image Domain Primitives with a Network of Competitive/Cooperative Processes." To appear in *Proc German Workshop on Artificial Intelligence*, W. Hoeppner, ed., Springer-Verlag, 1988.

Oliver, W & Bookstein, F Personal communication. University of North Carolina, Department of Computer Science, 1988.

Packard, C "Recognition of Neural Impulse Trains." MS Thesis, Computer Science Department, Worcester Polytechnic Institute, 1986.

Pizer, S, Gauch, J, Lifshitz, L "Interactive 2D and 3D Object Definition in Medical Images Based on Multiresolution Image Description," *SPIE 914*:438-444, 1988a.

Pizer, S, Gauch, J, Lifshitz, L, Oliver, W "Image Description via Annihilation of Essential Structures." UNC Chapel Hill Technical Report 88-001, 1988b.

Read, RC, Cornell, DG "The Graph Isomorphism Disease." *Journal of Graph Theory 1*:339-363, 1977.

Richards, W, Hoffman, DD "Codon Constraints on Closed 2D Shapes." *Computer Vision, Graphics, and Image Processing 31*:265-281, 1985.

Rosenfeld, A "Pyramid Algorithms for Efficient Vision." Tech Report #299, Center for Automation Research, University of Maryland, 1987.

Rosenfeld, A *Multiresolution Image Processing and Analysis*, Springer-Verlag, Berlin 1984.

Sher, A, Rosenfeld, A "Detecting and Extracting Compact Textured Objects Using Pyramids." Tech Report #269, Center for Automation Research, University of Maryland, 1987.

Stiehl, HS Personal communication. University of Hamburg, 1988.

Terzopoulos, D, Witkin, A, & Kass, M "Symmetry-Seeking Models for 3D Object Reconstruction," *Proc 1st Int. Conf. on Comp. Vis.* (London, England), IEEE Catalog #87CH2465-3, 1987.

Terzopoulos, D "Multilevel Computational Processes for Visual Surface Reconstruction." *Computer Vision, Graphics, and Image Processing 24*:52-96, 1983.

Toet, A, Koenderink, JJ, Zuidema, P, Graaf, CN "Image Analysis - Topological Methods." *Information Processing in Medical Imaging* (Proc. IPMI VIII, 1983), Martinus Nijhoff, 1984.

Witkin, A "Scale-Space Filtering," *Proceedings of International Joint Conference on Artificial Intelligence*, 1983.

Young, RA "Simulation of Human Retinal Function With The Gaussian Derivative Model." *Proc. IEEE CVPR*, pp. 564-569, 1986.

Yuille, AL & Poggio, T "Scaling Theorems for Zero-Crossings," MIT A.I. Memo 722, 1983.