An Image Description for Object Definition, Based on Extremal Regions in the Stack

Technical Report 85-016

1985

Stephen M. Pizer, Jan J. Koenderink, Lawrence M. Lifshitz, Leendert Helmink, Antonie D. J. Kaasjager

The University of North Carolina at Chapel Hill
Department of Computer Science
New West Hall 035 A
Chapel Hill, N.C. 27514

TR85-016

AN IMAGE DESCRIPTION FOR OBJECT DEFINITION, BASED ON EXTREMAL REGIONS IN THE STACK

Stephen M. Pizer[*+] , *Jan J. Koenderink*[-] ,
Lawrence M. Lifshitz[*], *Leendert Helmink*[-] ,
Antonie D. J. Kaasjager[-]

The University of North Carolina at Chapel Hill
Department of Computer Science
New West Hall 035 A
Chapel Hill, N.C. 27514

Department of Computer Science[*] *and Department of Radiology*[+] *, University of North Carolina, Chapel Hill, NC, USA and Department of Medical and Physiological Physics*[-] *, Rijksuniversiteit Utrecht, Netherlands*

To appear in Inf. Proc. in Med. Imaging, 9th IPMI Conf., Washington, D. C., 1985.

AN IMAGE DESCRIPTION FOR OBJECT DEFINITION, BASED ON EXTREMAL REGIONS IN THE STACK

Stephen M. Pizer * +, Jan J. Koenderink ~ , Lawrence M. Lifshitz * ,
Leendert Helmink ~ , Antonie D. J. Kaasjager ~
Dept. of Computer Science* and Dept. of Radiology +,
University of North Carolina, Chapel Hill, NC, USA
and
Department of Medical and Physiological Physics ~ ,
Rijksuniversiteit Utrecht, Utrecht, Netherlands

ABSTRACT

A promising image description is produced by dividing an image into nested light spots and dark spots by considering the image simultaneously at many levels of resolution [Koenderink, 1984]. These spots each include an image extremum and are thus called *extremal regions*. The nesting can be specified by a tree indicating the containment relationships of the extremal regions. This tree description, with each region described by intensity information, size, shape, and most significantly a measure of the importance, or *scale*, of the spot, absolutely and relative to its containing spot, ought to be usable in finding meaningful image objects when it is used together with *a priori* information about the expected structure of the image or its objects. This paper will describe work in the development of a computer program to compute such a description and its application to the display and segmentation of images from x-ray computed tomography and nuclear medicine.

1. PATTERN RECOGNITION AND DISPLAY VIA IMAGE DESCRIPTION

1.1. Introduction

Pictorial pattern recognition involves finding and labeling image objects. User interaction with and appreciation of an image frequently requires finding and labeling image objects, for example to allow the measurement of properties of the object or to provide a display which enhances the object. Indeed 3D display by shaded graphics depends on first defining the object to be displayed.

Most of the techniques attempted for defining objects work locally, directly with pixel intensity values. For example, both edge following and conventional region growing are done pixel by pixel. These methods have achieved only limited success, because the pixel values are too local to be easily combined into objects that are defined to a significant degree by their global properties. The first stage of a more attractive approach is to produce an *image description* that is more global. The creation of such a description should use only information from the image and not semantic information (from expectations about the scene or the viewing task). The second stage would use semantic ("real world") *a priori* knowledge together with the image description to define meaningful objects.

This approach of object definition based on a precomputed image description can be used for either automatic or interactive scene analysis. In automatic object definition the real world *a priori* knowledge is provided by predefined structures which are matched to the image description. In interactive object definition information is provided in addition by the human observer about objects he or she sees so that the computer can "perceive" the same object. The idea is that the observer can specify global properties of the object that is seen on some display, e.g. its location, intensity, scale, or name, and since the image has been reduced to a description reflecting global properties, fast interactive selection of image objects that match the observer's indications can be accomplished. The result can be displayed to the observer, who can accept the definition, edit it, or modify the defining parameters. The resulting object can be used to provide measurements, e.g. of volume or shape, or to serve as the basis for display or display interactions.

An attractive image description in this spirit [Koenderink, 1981] focuses on decomposing the image into light and dark spots, each, except for the spot representing the whole image, contained in others. Thus a face might be described as a light spot containing a light spot (a reflection from the forehead) and three dark spots (the mouth and the regions of the two eyes). In turn the eye regions would be described as containing a dark spot (the eyebrow), a light spot (the eyelid), and a dark spot (the eye), with the latter containing a light spot (the eyeball) which itself contains a dark spot (the iris) which finally contains a yet darker spot (the pupil). We call these light and dark spots, at whatever scale, *extremal regions,* since they each include a local intensity maximum or minimum.

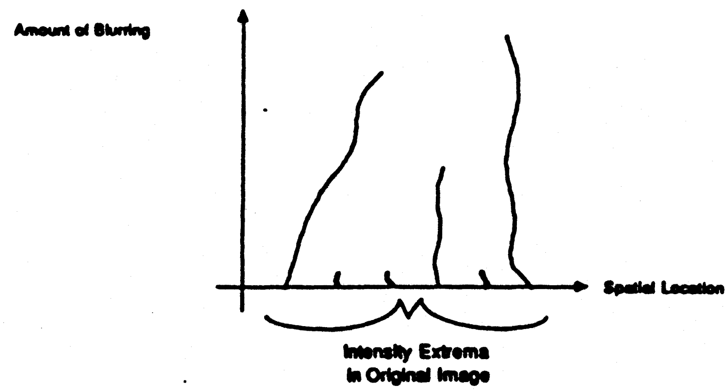


FIGURE 1. Extremal paths through the stack

1.2. Hierarchical descriptions from multiresolution processing

The image description in terms of extremal regions can be produced by following the paths of extrema in a stack of images in which each higher image is a slightly blurred version of the previous one. As illustrated in figure 1 and explained in Koenderink [1984], progressively blurring an image causes each extremum to move continuously, and eventually to annihilate as it blurs into its background. An *extremal path* is formed by following the locations of an extremum across the stack of images.

Intensity must be monotonic (increasing for dark spots and decreasing for light spots) as one moves along an extremal path from the original image towards images of increased blurring. As illustrated in figure 2, while following each extremal path

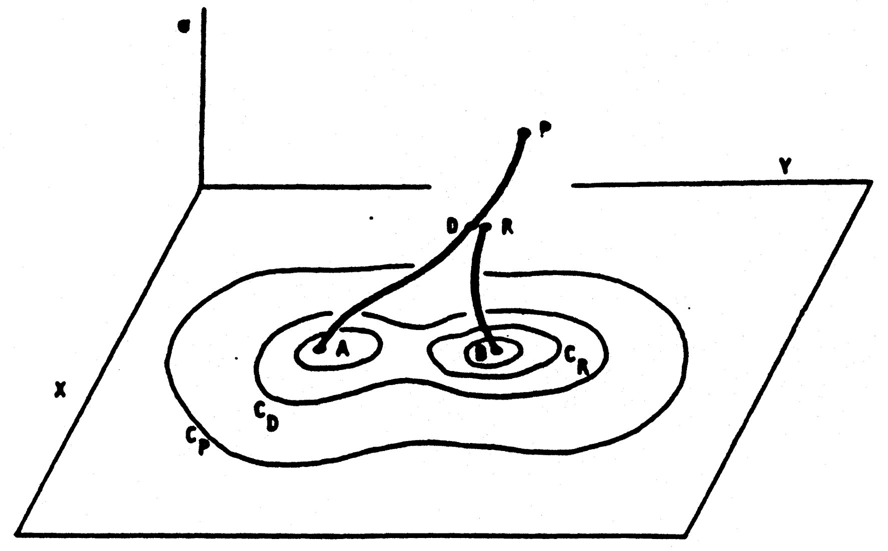


FIGURE 2. Extremal paths and associated iso-intensity contours

one can associate each path point with the iso-intensity contour that is at that point's intensity and that surrounds that extremum in the original image [Koenderink, 1984]. The points (pixels) in the original image on the contours thus associated with each extremal path then form an extremal region (see figure 3).

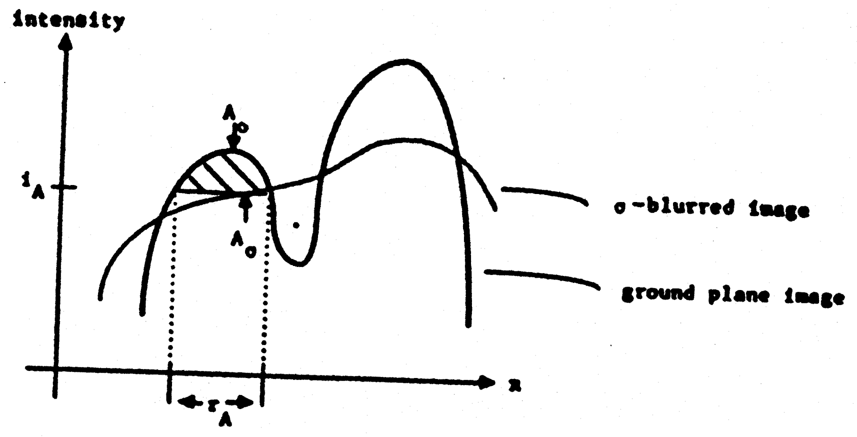
Each contour point in the original image can be associated with its extremal path by following the point to another with its intensity level at the next level in the stack, continuing through the levels until the extremal path is reached (see figure 4). This process defines an *iso-intensity path*.

It has been shown that all extremal paths must start in the original image if gaussian blurring is used. Extremal paths cannot be initiated at higher stack levels. However, as indicated above, extrema annihilate when the blurring is sufficient to make the light or dark spot blur into an enclosing region. The amount of blurring necessary for an extremum to annihilate is a measure of the importance or scale of the extremal region, including the subregions that it contains.

The intensity of the topmost point on an extremal path is its annihilation intensity. This is the intensity of the iso-intensity contour that forms the boundary of the associated extremal region. Remember that the annihilation intensity bounds from below (above) the intensities in the extremal region if the associated extremum is a maximum (minimum).

1.3. A tree of extremal regions for image description

As illustrated in figure 2, when an extremum annihilates at some annihilation intensity, another region's iso-intensity contour at that intensity encloses the region associated with the annihilating extremum [Koenderink, 1984]. Thus, a containment relation among extremal regions is induced by the process. This set of extremal regions together with their containment relations can be represented by an *image description tree* in which nodes represent extremal regions and a node is the child of another if the extremal region that it represents is immediately contained by the extremal region represented by the parent (see figure 5). The root of the image description tree represents the entire image.



If extremum at A_o has moved to A_σ for the degree of blurring σ at which the extremum annihilates, then r_A gives the extremal region and i_A its characteristic intensity.

FIGURE 3. Extremum annihilation by blurring and consequent extremal region

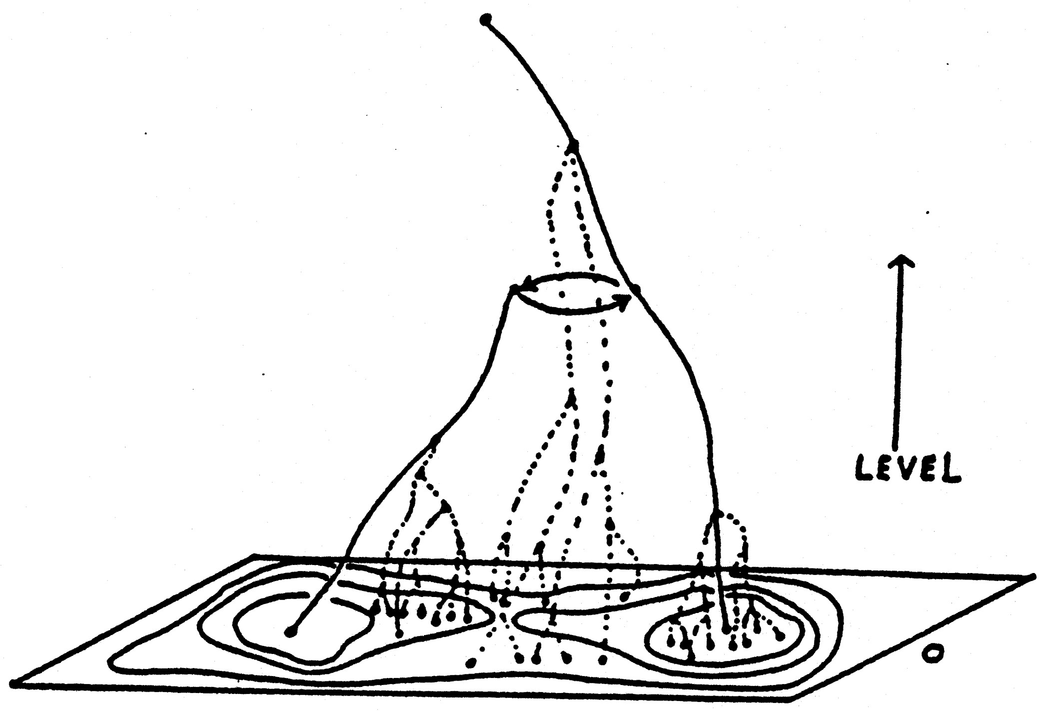


FIGURE 4. Extremal paths (indicated by solid lines) and iso-intensity paths (indicated by broken lines). Iso-intensity contours are indicated in the original image (level 0). The relation of an extremal region enclosing another region is indicated by arrows.

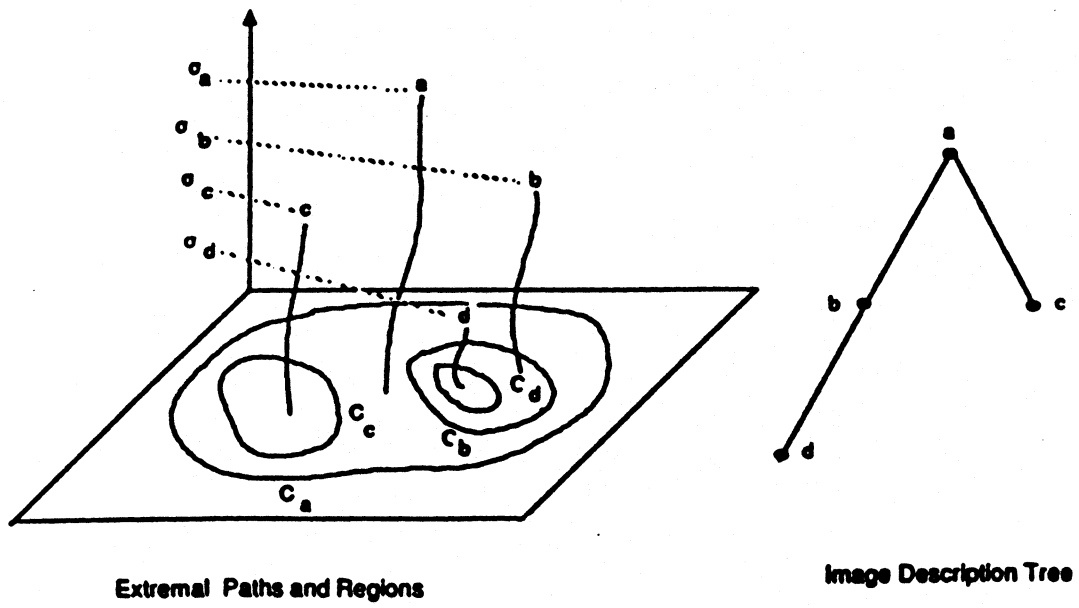


FIGURE 5. Extremal paths, with their regions and scales, and the associated image description tree

Each node in the image description tree can be labeled with its scale, i.e. the total amount of blurring necessary for its extremum to annihilate. Furthermore, each node can be labeled with the annihilation intensity of the associated extremum. Finally, the node can be labeled with the size, shape, orientation, location, or other spatial characteristics of its extremal region.

It is possible that the description process described above can be beneficially preceded by some preprocessing, e.g. to enhance contrast or edges. In fact, Crowley [1984] has developed a similar scheme of extremum following in a multiresolution pile of images that are edge-enhanced by a sort of unsharp masking, and he has applied it with promising results. We have tried such preprocessing a few times with some benefit but will not discuss it in greater detail in this paper. However, it is worth noting that a noncognitive component of human visual perception may possibly be well modeled by an edge-enhancing preprocessing followed by the production of a stack-based image description.

2. SAMPLING

With computer implementation both space and the amount of blurring become discrete. That is, space is divided into pixels and blurring is not done continuously but by convolution with a gaussian of a non-infinitesimal standard deviation, which may vary from blurring step to blurring step.

The amount of blurring in each step needs to be controlled so that confusion in following extremal paths and associated iso-intensity paths across stack levels is avoided, while limiting the number of steps so that reasonable efficiency is achieved. When there are many extremal paths, we have taken this criterion to imply an inter-level blurring that is just large enough to ensure that real changes dominate changes due to arithmetic error. When there are few extremal paths, an inter-level blurring is chosen that produces significant progress toward annihilation of one of the paths.

Progress toward annihilation depends on taking steps related to the distance between paths. We keep count of the number of remaining paths as we move up the stack, and when there are fewer extremal paths than some threshold (in present use, 6), we blur at each level by averaging over a region whose diameter is the distance between the closest paths at that level.

For levels below the point where efficiency considerations lead to the blurring just described, we need to interpret when "real changes dominate changes due to arithmetic error." Koenderink [1984] and Pizer [van Os, 1984] have both interpreted this to mean that the attenuation of the height of some basic function is a small integer multiple of the arithmetic error, but these two investigators have chosen a different basic function. Koenderink chose a sinusoid at the Nyquist frequency associated with the total amount of blurring at any given level, while Pizer chose a gaussian which was a spike in the scene (but not the original image, which already is a blurred version of the scene) on a flat background, where the ratio of the height of the spike to the background has some value chosen as a parameter.

Koenderink's choice leads to

$$\frac{\delta\sigma^2}{\sigma^2} = k\frac{\sqrt{2\rho}}{-\log_e \rho},$$

where σ^2 is the variance of the total blurring in the image at the present stack level, $\delta\sigma^2$ is the variance of the blurring to be applied to that image, ρ is the bound on the relative error in the computer representation of intensity, and k is a small integer. This has the attractive property that the amount of additional blurring ($\delta\sigma^2$) is proportional to the total blurring done to create this level.

Pizer's choice leads to

$$\frac{\delta\sigma^2}{\sigma^2} = \frac{k'\rho(1+\beta\frac{\sigma^2}{\sigma_0^2})}{1-k'\rho\beta\frac{\sigma^2}{\sigma_0^2}},$$

where ρ is as before, k' is a small integer, β is the ratio of background to peak height in the scene, and σ_0^2 is the blurring due to imaging. Eventually the peak height relative to the background becomes so small that no blurring can reduce it to the criterion degree. At this point the blurring at each step is chosen to allow a decrease in spatial sampling of 2 in each dimension, a common approach in multiresolution methods. To achieve this goal, $\delta\sigma^2$ must be proportional to σ^2 with a constant of proportionality of 3; the result is that the total blurring standard deviation increases by 2 at each step.

Studies by van Os and Pizer [1984] indicate that Pizer's choice leads to fewer levels of blurring with no major loss in the quality of the result, when the blurring used in the very first step of the two approaches is the same.

The sampling in space should, by normal sampling practice, decrease as you move up the stack, i.e. as the amount of blurring increases. More precisely, the inter-pixel distance should be proportional to the standard deviation of the total amount of blurring due to imaging together with the stack blurring. Using an argument based on the aliasing error at the Nyquist frequency, Pizer suggests a proportionality constant of approximately $\pi/\sqrt{1-\log_e \rho}$.

However, changing the spatial sampling as you move up the stack complicates the extremum and iso-intensity path following processes. As a result, in this early stage of our research, we have left the spatial sampling in all stack levels the same as in the original image. Part of the efficiency of resampling is achieved in our method, since only pixels that are on an iso-intensity or extremum path are followed to the next level.

3. AN ALGORITHM FOR PRODUCING A STACK-BASED IMAGE DESCRIPTION TREE

3.1. Algorithm description

It is our plan to use the image description region tree presented above for both automatic and interactive object definition. However, this object definition work is still in its early stages, and most of the remainder of this paper will focus on computing the image description and on properties of the resulting description.

Programs to compute the image description tree by successive blurring, extremal path following, and iso-intensity pixel linking have been written at Rijksuniversiteit Utrecht and The University of North Carolina. These programs differ slightly in the implementation details. Some of these differences will be presented later. The initial description will be of the UNC version. The programs are applicable in one, two, and three dimensions, in the last case not slice by slice but fully in 3D.

The tree described in section 1.3 represents the nesting structure of extremal regions. On the way to producing this structure it is useful to create an *intermediate description tree* which contains more detailed information about individual pixels. Each node in this tree corresponds to one pixel in the image stack, but only some pixels in the stack have a node in the tree. In particular, pixels that are either on an extremum path or an iso-intensity path make up the nodes in the tree.

A node corresponding to a non-extremum pixel is called a *normal node*, while one representing an extremum is an *extremum node*. Extremum nodes are linked by extremal path links to form a representation of an extremal path. Similarly, normal nodes are linked by normal path links to represent an iso-intensity path. A link to a normal node parent is called a normal path link, regardless of whether the child is a normal or extremum node, and an extremum node can be linked to an extremum parent by a normal path link if the connection is not part of an extremal path. Annihilation is represented by the connection of an extremum node via a normal path link to either a normal parent, or an extremum parent on a different extremal path.

An overview of the algorithm is as follows. The program links pixels in each level of the stack to pixels in the level above them. During this process it also creates nodes in the intermediate description tree. To accomplish this linking, the program works on the two adjacent images at the two stack levels being linked.

Each pixel in an image can be thought of as a potential node in the intermediate description tree. All pixels in the original image form leaf nodes of the tree; each will be either on an extremal path or an iso-intensity path. After acting on the two images as described below, the lower image is discarded (but any nodes created are kept in the tree); the upper image becomes the lower image; and the next slightly more blurred image becomes the new upper image. This process continues until only one extremal path remains. The remaining nodes are linked to this path, and the tree is then written to a file.

A skeleton of the procedure in pseudo-code is as follows:

```
find extrema in the original (currently "lower") image
while more than one extremum remains do
    blur "lower" image to create "upper" image
    find extrema in the upper image
    link pixels in the lower level to those in the upper level, creating nodes in the
        intermediate description tree as appropriate
    discard lower image; upper image becomes new lower image
link remaining paths to last extremal path. output tree to a file.
```

An explanation of the main elements of this algorithm follows.

3.1.1. Extrema identification. Each of the two active planes is examined separately to find the existence and location of extrema. A point in an image is considered a maximum (minimum) if it is of greater (lesser) intensity than all of its eight surrounding pixels. The entire image is examined, and each of the extrema are marked as such. All other pixels are called "normal" pixels. Once this is done for the upper plane (it having previously been done for the lower plane), the linking routine is invoked.

3.1.2. Linking. The linking routine examines each pixel in the lower plane that is on an extremum or iso-intensity (normal) path, i.e., that is a node in the tree. It finds an appropriate pixel in the upper plane to link to. The linking strategy for extremum pixels differs slightly from that for normal pixels.

The algorithm tries to link an extremum pixel in the lower image to a similar type (maximum or minimum) of extremum in the upper image. If it fails to be able to do so, then it invokes the linker for normal pixels, which has less stringent criteria for linking. The general strategy is to search for a parent in the small neighborhood surrounding the pixel directly above it in the upper image. The pixel in this neighborhood with intensity value closest to that of the child's in the lower image is picked as its parent. There is a maximum intensity difference allowed between a pixel in the lower plane and its parent. If no pixel in the selected upper plane neighborhood is close enough in intensity to the lower pixel, the neighborhood is enlarged slightly and the search continues. If no viable candidate is found then, the maximum intensity difference allowed is incremented and the neighborhood search is repeated. If this process fails to find a viable parent, the pixel is linked to the pixel directly above it.

Many extremum nodes can link to the same extremum father even though the theory for the continuous case does not allow an extremum to merge directly with another extremum, but instead requires it to merge with a saddle point, with both the saddle point and the extremum annihilating. If more than one extremum pixel is linked to the same extremum father, the extremum son with the closest intensity on the appropriate side of the father's intensity (above for a maximum, below for a minimum) is connected via an extremal path link. All of the others are connected by normal path links.

3.1.3. Node creation. Conceptually any pixel in the upper level which has at least one child becomes a node in the intermediate description tree. Space is saved by representing the frequently occurring chains of pixels linked to the pixel directly below them with a single node holding the range of levels of the chain. A node may have many children. Both normal and extremum nodes may have normal and extremum children.

3.1.4. Extremum annihilation and object definition. Extremum nodes lie on extremal paths. That is, they have at least one child which is an extremal node, and they usually have a father which is an extremal node. This extremal path is considered to have annihilated if the extremum node is not linked to an extremum father or if it is linked via a normal path link.

The extremal node which is has been identified as annihilating is at the root of a subtree in the intermediate description tree. All nodes in this subtree belong to the extremal region associated with its extremal path. The nodes in the original image form the extremal region associated with this annihilation. This region frequently includes other extremal regions; that is, some of the subtree's nodes are annihilating extrema in their own right.

3.1.5. Termination. When only one extremum remains in the upper plane, the program is near completion. Links are created from the lower to upper plane in the usual fashion. Following this, one additional node is created. This becomes the root node of the intermediate description tree, and all nodes from the upper image plane are forced to link to this root node, in the process creating new subregion to region connections. The tree data structure, now complete, is written out to a file which can be read in and traversed by a display or pattern recognition program at a later time.

3.2. Implementation complications

The theory behind the stack technique was developed in continuous space. This means that intensity quantization (floating point), extremum and iso-intensity path following across discrete levels with discrete pixels, and finiteness of image size are all aspects which are not explicitly addressed in the basic theory.

The most significant problem arises from the non-infinitesimal blurring between each level in the stack. The major complication that this introduces is an uncertainty as to which pixel in the upper plane a pixel in the lower plane should link. In the continuous case, there is always a pixel in the upper plane with precisely the same intensity as the pixel in the lower plane, and it is always "close" to the same position as the lower pixel (in fact the path taken by the pixel from level to level is an integral curve of a vector field [Koenderink, 1984]). In the discrete case no pixel in the upper plane will have exactly the same intensity as its child, and the distance to be traversed for the link may be several pixels. Decision criteria must be developed to determine which possible linking candidate for parent pixel is the one to be chosen.

Linking criteria developed at UNC and at Rijksuniversiteit Utrecht differ somewhat. The main difference is in the way extremal points are handled. As mentioned above, at UNC each of the two active images is examined separately to locate extremal points. The points in the lower plane are then linked to an extremal point in the upper plane if possible, first by checking a close neighborhood and then a larger one. If no extremum father is found, the point is linked to a normal point.

At Rijksuniversiteit Utrecht extrema are identified only if an extremum son links to it. As at UNC the algorithm attempts to link each extremum in the lower image to an extremal point in the neighborhood just above it. If a match is not found, then hill climbing (or pit sliding) is performed until a maximum or a minimum, as appropriate for the extremal path in question, is reached. The pixel is linked to this. An extremal path is said to annihilate if a second extremal path links to the same parent pixel, in which case a decision is made as to which of the two is considered to be annihilating based on geometric and intensity differences between the father and the extremum sons.

The UNC approach requires more searching for extrema than the Utrecht approach, and it allows new extremal paths to begin at levels above the bottom. However, for evaluation of the method it allows the user to see which extrema have been missed, and extrema which are in fact due to arithmetic error do not cause problems because they annihilate quickly. The hill climbing (or pit sliding) heuristic used in Utrecht is not guaranteed to find correct links, but it may infrequently find a link that is missed by the UNC method and thus avoid creating a false annihilation.

Some other complications include determination of appropriate boundary conditions for the finite image and creation of extrema due to quantization effects. These complications can be dealt with by standard approaches without much difficulty.

3.3. Implementation performance

The current implementations have not been optimized for speed of execution or minimization of space. Some indications of the speed and size of the current UNC algorithm are nevertheless in order. The program has been applied to between five and ten CT images of the upper abdomen. It has also analyzed several synthetic images of one, two, and three dimensions. The Utrecht program has been applied to over ten 2D and 3D images from scintigraphy, MRI, radiography, and normal photography; the third dimension in the 3D images has sometimes been spatial depth and sometimes time. Most of the images have been reduced to 64 by 64 images to reduce time and space requirements of the algorithm in its testing and initial evaluation stage. Running on a moderately loaded VAX 780, the UNC program takes approximately 45 seconds to 1 minute to create each level in the stack, together with all its associated structures (marking extrema, linking to the new level, etc.). The 64 by 64 images tend to need about 20 levels of blurring before only one extremum remains. Thus the program runs for about 20 minutes. The intermediate description tree created takes about 250 kbytes, of which about half is blank inter-entry separators. Considerable space and display time could be saved by reducing the intermediate description tree to the simpler image description tree.

All of the above numbers scale approximately linearly with image area. Of course, an image with a lot of noise or very many objects will tend to take longer and create a larger data structure and so forth. If it is known in advance that the structures of interest in an image are of small scale, the processing may be terminated before only one extrema remains, saving much time.

4. INTERACTIVE DISPLAY BASED ON AN IMAGE DESCRIPTION TREE

The subdivision into regions labeled by scale and intensity provides the opportunity for the user to explore the extremal regions in the tree and select ones that are clinically meaningful and of interest. With previously available methods, defining such objects frequently involves drawing the boundary point by point. This is time consuming in two dimensions. In three dimensions not only is it unacceptably time consuming, but the normal approach of working slice by 2D slice impedes the use of interslice relations in defining the boundary.

In the approach that we have investigated, the user specifies dynamically a range of intensities and a range of scales, as well as a spatial window, and all original image pixels in regions with intensity or scale labels in these ranges are displayed if they are in the appropriate spatial window. When the user sees an object that is meaningful, the selected pixels can be taken to define the object or the result can be edited by the user. Then display or measurement operations on that object can commence.

Whenever it is desired to view the image, the display program reads in the image description tree from a file. The user is then able interactively to control which extremal regions in the image are displayed. This is done by means of various A/D devices. Two sliders are used to specify a scale range for objects to be viewed. For example, only big or high contrast objects could be displayed. Similarly, two sliders specify the intensity range of objects to be displayed. This would be used, for example, to select bright objects (like the spinal column). Four knobs are used to control the spatial locations in the image that are to be displayed (maximum and minimum x and y dimensions).

Currently the display program takes approximately two or three seconds to update the image displayed. This number is highly variable depending upon the number of objects in the image, their inter-relationships, the system load, and the speed of writing to the display device.

Below is a series of images displayed on our system. The original image that was analyzed is shown in figure 6a. This is a CT image of the upper abdomen, scaled down to a size of 64 by 64 pixels. It is displayed by adjusting the interactive knobs and sliders so as to ask for objects of all possible resolution sizes, intensity ranges, and spatial positions. The light area in the center bottom of the image is the spinal column. The roundish objects, one on either side of the spinal column, are the kidneys. On the left, about halfway up is the liver, which seems to be merging with the chest wall in this image. On the right side, above the kidney is the jejunum, and above that the transverse colon. The very dark regions are gas.

In figure 6b we have adjusted the sliders to ask only for those objects of larger scale, thus eliminating the noise. Notice that the darker regions around the kidneys are displayed as objects, even though medically they are not organs. This is an example of the program finding something which is not semantically meaningful, even though it is understandable why it did so. The major organs are clearly visible.

In figure 6c the spinal column has been eliminated by changing the intensity sliders to specify that bright objects not be displayed. It should be emphasized that these regions have been eliminated because they are regions associated with extremal paths whose annihilation intensities did not fall within the specified range. The display program is *not* simply looking at individual pixel intensities in the image. Only *extremal regions* can be displayed or removed.

In figure 6d the scale range requested specifies only the biggest objects, and all intensities are selected, resulting in a selection of the liver.

In figure 6e we have specified objects of slightly smaller scale only and have limited intensity to a middle range, thus obtaining the kidneys and jejunum. The liver is now gone.

5. EFFECTIVENESS AND FUTURE DIRECTIONS

The image descriptions produced when this stack method is applied to medical

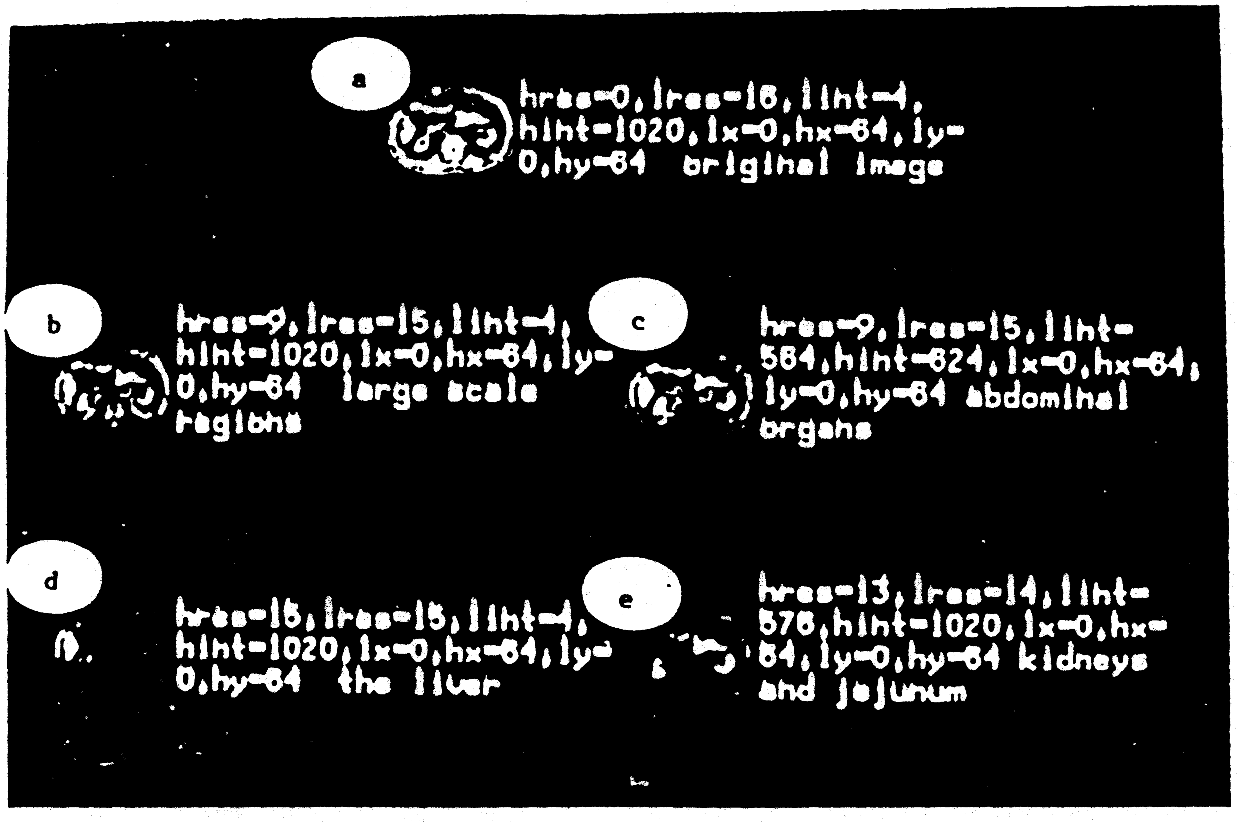


FIGURE 6. Interactive display based on image description of CT of the upper abdomen (at UNC)

images frequently contain regions that seem well related to meaningful image segments. This is true even in images with a low signal to noise ratio, such as those from nuclear medicine (see figure 7). There is therefore indication that segmentation methods based on this description will obtain far better segments than more common segmentation methods of edge detection or region growing. However, regions sometimes break up in semantically unnatural ways. For example, a blood vessel may break up into a number of pieces, and one piece may be a subregion of the organ in which it is contained, and another piece, say on the edge of the organ, may be a subregion of the region adjacent to the organ.

Structural pattern recognition techniques seem appropriate to bring semantic information to bear to correct this situation and at the same time label the objects, e.g. the blood vessel as such. These techniques operate by matching descriptions of known objects, e.g. an image description tree for a typical structure for a particular organ, to the description of our image, or a portion thereof. The multiresolution approaches have already shown themselves to be very well suited to this requirement [Rosenfeld, 1984; Crowley, 1984], as they allow one to operate at large scale (high in the description tree) first and to use matches there to guide matches at lower scale.

The results of such a matching process is the labeling of objects in the image description or the reorganization of the description tree to combine subobjects into previously undefined objects and then label the results. However, because the labeling

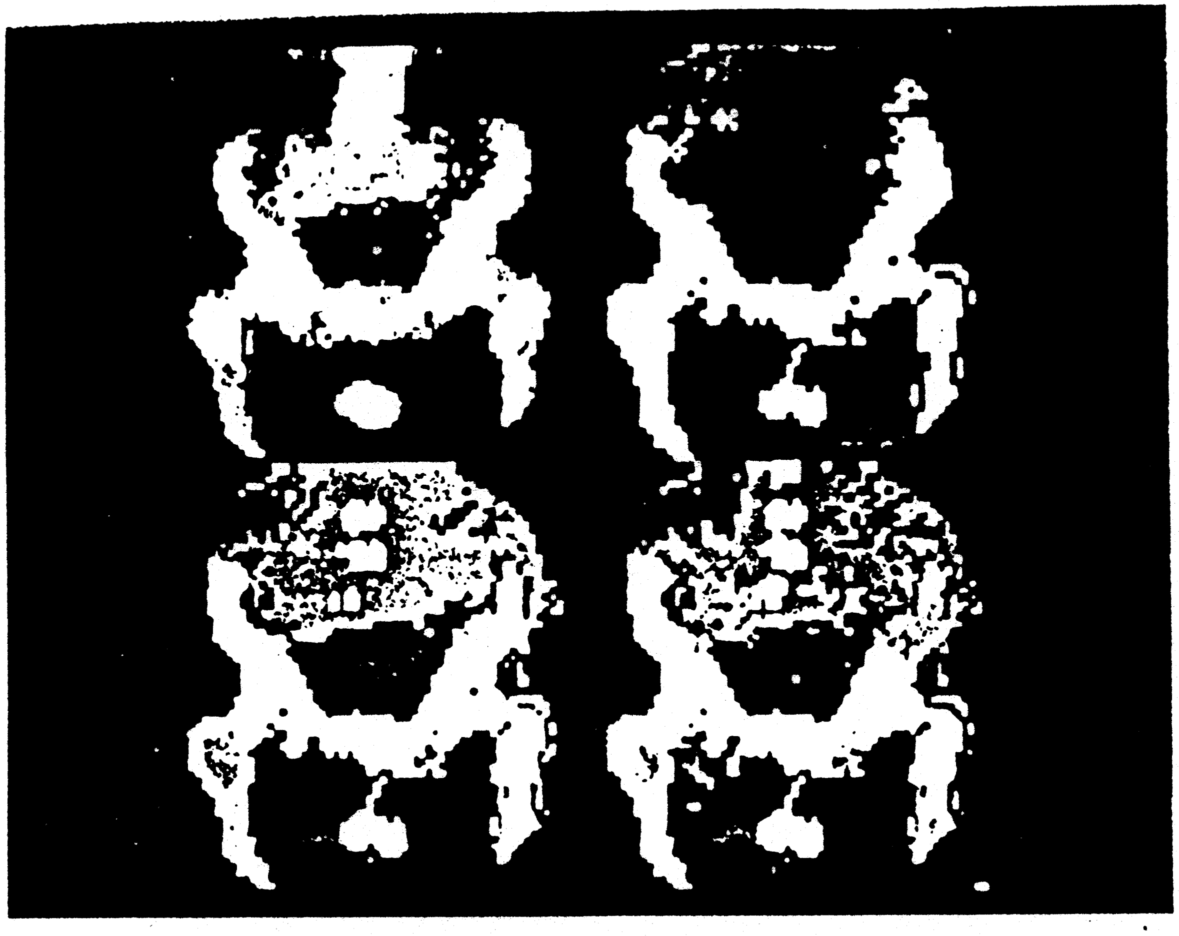


FIGURE 7. Extremal regions from image description of scintigram of the pelvis (at Rijksuniversiteit Utrecht). Upper left: original image. Other quadrants: each extremal region of scale smaller than some specified value is displayed with all its pixels at its average intensity.

depends on an image description that in the semantic sense is in error, it is likely that matching errors will result. We hypothesize that the labeling produced to date by image description followed by matching can be taken as tentative and used to produce an improved image description, which in turn could be used to produce an improved object definition and labeling. Therefore, we are presently working on

(1) creating improved image descriptions by letting the blurring used in producing the stack at any step depend on the previous tentative segmentation, and
(2) developing methods for matching the image description tree to *a priori* description trees for images or image objects to produce a labeled segmentation into image objects.

The modified blurring is nonstationary and nonisotropic, depending in scale in each direction on the scale and orientation of nearby objects that are at the stack level in question in the tentative segmentation.

We also anticipate using display based on the image description tree for defining objects in three dimensions. We plan to allow the user to select extremal regions by scale and intensity ranges, plus a 3D window, with the result displayed on the varifocal mirror or another self-luminous 3D display. When a meaningful object is seen, its surface can be directly calculated and used as the basis of a display that is more appropriate for therapy planning, such as a shaded graphics display.

ACKNOWLEDGEMENTS.

We thank Alexander Toet, Cornelis de Graaf, and Matthew Mauro for useful discussions and the latter two for providing images used in this study. We are indebted to Joan Savrock for manuscript preparation and to Bo Strain and Karen Curran for photography. The research reported here was done partially with the support of grant number 1-R01-CA39060 from the U.S. National Institutes of Health and partially with the support of the Dutch Institute for Pure Scientific Research (ZWO).

REFERENCES

1. Crowley, J.L. and A.C. Sanderson, "Multiple Resolution Representation and Probabilistic Matching of 2-D Gray-Scale Shape", Tech. Rept. No. CMU-RI-TR-85-2, Carnegie-Mellon Univ., 1984. Also see Crowley, J.L. and A.C. Parker, "A Representation for Shape Based on Peaks and Ridges in the Difference of Low-Pass Transform", *IEEE Trans. PAMI*, March, 1984.
2. Koenderink, J.J. and A. van Doorn, "A Description of the Structure of Visual Images in Terms of and Ordered Hierarchy of Light and Dark Blobs", *Proc. 2nd Int. Conf. on Vis. Psychophysics and Med. Imaging*, IEEE, Cat. No. 81CH1676-6, 1981.
3. Koenderink, J.J., "The Structure of Images", *Biol. Cybernetics, 50:* 363-370, 1984.
4. Rosenfeld, A., ed., *Multi-resolution Image Processing and Analysis,* Springer-Verlag, New York, 1984.
5. van Os, C.F.A., "The Inca-Pyramid Algorithm", Tech. Rept. No. VMFF 30-84, Institute of Medical Physics, Rijksuniversiteit Utrecht, Utrecht, Netherlands, 1984.