IMAGE SEGMENTATION SURVEY

Robert M. Haralick
Virginia Polytechnic Institute and State University, Dept. of Electrical Engineering and Dept. of Computer Science, Blacksburg, VA 24061 USA

Abstract. There is no complete theory for image segmentation although there are a variety of techniques for segmenting images. This paper discusses the major ideas behind the measurement space clustering, single linkage, hybrid linkage, region growing, spatial clustering, and split and merge techniques.

INTRODUCTION

What should a good image segmentation be? Regions of an image segmentation should be uniform and homogeneous with respect to some characteristic such as gray tone or texture. Regions interiors should be simple and without many small holes. Adjacent regions of a segmentation should have significantly different values with respect to the characteristic on which they are uniform. Boundaries of each segment should be simple, not ragged, and must be spatially accurate.

Achieving all these desired properties is difficult because strictly uniform and homogeneous regions are typically full of small holes and have ragged boundaries. Insisting that adjacent regions have large differences in values can cause regions to merge and boundaries to be lost.

There is no theory of image segmentation. Image segmentation techniques are basically ad-hoc and differ precisely in the way they emphasize one or more of the desired properties and in the way they balance and compromise one desired property against another. Image segmentation techniques can be classified as: measurement space clustering, single linkage schemes, hybrid linkage schemes, region growing/centroid linkage schemes, spatial clustering schemes, and split and merge schemes.

The remainder of the paper briefly describes the main ideas behind the major image segmentation techniques. Additional image segmentation surveys can be found in Zucker (1976), Riseman and Arbib (1977), Kanade (1980), and Fu and Mui (1981).

MEASUREMENT SPACE CLUSTERING

This technique for image segmentation uses the measurement space clustering process to define a partition in measurement space. Then each pixel is assigned the label of the cell in the measurement space partition to which it belongs. The segments are defined as the connected components of the pixels having the same label. Because of the large number of pixels in an image, clustering using the pixel as a unit and comparing each pixel value with every other pixel value can require excessively large computation time. Iterative partition rearrangement schemes, such as ISODATA, have to go through the image data set many times and if done so without sampling can also take excessive computation time. Histogram mode seeking, because it requires only one pass through the data, probably involves the least computation time of the measurement space clustering techniques.

Histogram mode seeking is a measurement space clustering process in which it is assumed that homogeneous objects on the image manifest themselves as the clusters in measurement space. Image segmentation is accomplished by mapping the clusters back to the image domain where the maximal connected components of the mapped back clusters constitute the image segments. For images which are single band images, calculation of this histogram in an array is direct. The measurement space clustering can be accomplished by determining the valleys in this histogram and declaring the clusters to be the interval of values between valleys. A pixel whose value is in the i^{th} interval is labeled with index i and the segment it belongs to is one of the connected components of all pixels whose label is i.

Ohlander (1975) recursively uses this idea. He begins by defining a mask selecting all pixels on the image. Given any mask, a histogram of each band of the masked image is computed. Measurement space clustering enables the separation of one mode of the histogram set from another mode. Pixels on the image are then identified with the mode to which they belong. Then each connected component of all pixels with the same mode is, in turn, used to generate a mask which during successive iterations selects pixels in the histogram computation process.

For ordinary color images, Ohta, Kanade, and Sakai (1980) suggest that histograms not be computed with the red, green, and blue (R,G, and B), but with a set of variables closer to what the Karhunen Loeve transform would suggest. They suggest $(R + G + B)/3$, $(R - B)/2$ and $(2G - R - B)/4$.

Weszka and Rosenfeld (1978) describe one way for segmenting white blobs against a dark background by a threshold selection based on busyness. Panda and Rosenfeld (1978) suggest another approach for segmenting the white blob against a dark background. Pixels having high gradients and gray levels in the valley between the two histogram modes are likely to be edge pixels. Non edge pixels are those with low gradient values and either high or low gray levels. The segments are the connected components of the non edge pixels.

Watanabe (1974) suggests choosing a threshold value which maximizes the sum of gradients taken over all pixels whose gray level equals the threshold value. A survey of threshold techniques can be found in Weszka (1978)

For multiband images such as LANDSAT, determining the histogram in a multi-dimensional array is not feasible. For example, in a six band image where each band has intensities between 0 and 99, the array would have to have $100^6 = 10^{12}$ locations. A large image might be 10,000 pixels per row by 10,000 rows. This only constitutes 10^8 pixels, a sample too small to estimate probabilities in a space of 10^{12} values were it not for some constraints of reality: (1) there is typically a high correlation between the band to band pixel values and (2) there is a large amount of spatial redundancy in image data. Both these factors create a situation in which the 10^8 pixels can be expected to contain only between 10^4 and 10^5 distinct 6-tuples. Based on this fact, the counting required for the histogram is easily done by hashing the 6-tuple into an array.

Clustering using the multidimensional histogram is more difficult than univariate histogram clustering. Goldberg and Shlien (1977, 1978) threshold the multidimensional histogram to select all N-tuples situated on the most prominent modes. Then they perform a measurement space connected components on these N-tuples to collect together all the N-tuples in the top of the most prominent modes. These measurement space connected sets form the cluster cores. The clusters are defined as the set of all N-tuples closest to each cluster core.

An alternate possibility (Narendra and Goldberg, 1977) is to locate peaks in the multi-dimensional measurement space and region grow around it constantly descending from the peak. The region growing includes all successive neighboring N-tuples whose probability is no higher than the N-tuple from which it is growing. Adjacent mountains meet in their common valleys.

SINGLE LINKAGE IMAGE SEGMENTATION

Single linkage image segmentation schemes regard each pixel as a node in a graph. Neighboring pixels whose properties are similar enough are joined by an arc. The image segments are maximal sets of pixels all belonging to the same connected component. Single linkage image segmentation schemes are attractive for their simplicity. They do, however, have a problem with chaining, because it takes only one arc leaking from one region to a neighboring one to cause the regions to merge.

The simplest single linkage scheme defines similar enough by pixel difference. Two neighboring pixels are similar enough if the absolute value of the difference between their gray tone intensity values is small enough. Bryant (1979) defines similar enough by reference to the quantity (square root of 2) times the root mean square value of neighboring pixel distances taken over the entire image.

For pixels having vector values, the obvious generalization is to use a vector norm of the pixel difference vector. Instead of using a Euclidean distance, Asano and Yokoya (1981) suggest that two pixels be joined together if this absolute value of their difference is small enough compared to the average absolute value of the center pixel minus neighbor pixel for each of the neighborhoods the pixels belong to. Haralick and Dinstein (1975), however, do report some success using the simpler Euclidean distance on LANDSAT data. They, as did Perkins (1980), region grew the edge pixels in order to close gaps. The ease with which unwanted region chaining can occur with this technique limits its potential on complex or noisy data.

Hybrid single linkage techniques are more powerful than the simple single linkage technique. The hybrid techniques seek to assign a property vector to each pixel where the property vector depends on the KxK neighborhood of the pixel. Pixels which are similar, are similar because their neighborhoods in some special sense are similar. Similarity is thus established as a function of neighboring pixel values and this makes the technique better behaved on noisy data.

One hybrid single linkage scheme relies on an edge operator to establish whether two pixels are joined with an arc. Here an edge operator is applied to the image labeling each pixel as edge or non-edge. Neighboring pixels, neither of which are edges, are joined by an arc. The initial segments are the connected components of the non-edge labeled pixels. The edge pixels can either be left assigned edges and be considered as background or they can be assigned to the spatially nearest region having a label.

The quality of this technique is highly dependent on the edge operator used. Simple operators such as the Roberts and Sobel operator may provide too much region linkage, for a region cannot be declared as a segment unless it is completely surrounded by edge pixels. Yakimovsky (1976) uses a maximum likelihood test to determine edges. Edges are declared to exist between pairs of regions if the hypothesis that their means are equal and their variances are equal has to be rejected.

Haralick (1980) suggests fitting a plane to the neighborhood around the pixel, and testing the hypothesis that the slope of the plane is zero. To determine roof or V-shaped edge, Haralick suggests fitting a plane to the neighborhoods on either side of the pixel and testing the hypothesis that the coefficients of fit are identical. Haralick (1982) discusses a very sensitive zero-crossing of second directional derivative edge operator. In this technique, each neighborhood is fitted by least squares with a cubic polynomial in two variables. The first and second partial derivatives are easily determined from the polynomial. The first partial derivatives at the center pixel determine the gradient direction. With the direction fixed to be the gradient direction, the second partials determine the second directional derivative. If in the gradient direction, the second directional derivative has a zero-crossing inside the pixel, then an edge is declared in the neighborhood's center pixel.

Another hybrid technique first used by Levine and Leemet (1976) is based on the Jarvis and Patrick (1973) shared nearest neighbor idea. Using any kind of reasonable notion for similarity, each pixel examines its KxK neighborhood and makes a list of the N pixels in the neighborhood most similar to it. Call this list the similar neighbor list, where we understand neighbor to be any pixel in the KxK neighborhood. An arc joins any pair of immediately neighboring pixels if each is in each other's shared neighbor list and if there are enough pixels common to their shared neighbor lists; that is, if the number of shared neighbors is high enough.

REGION GROWING / CENTROID LINKAGE

In region growing, as contrasted to single linkage, pairs of neighboring pixels are not compared for similarity. Rather, the image is scanned in some predetermined manner such as left-right top bottom. A pixel's value is compared to the mean of an already existing but not necessarily completed segment. If the values are close enough, then the pixel is added to the segment and the segment's mean is updated. If no neighboring region has its mean close enough, then a new segment is established having the given pixel's value as its first member.

Pavlidis (1972) suggests a more general version of this idea. Given an initial segmentation where the regions are approximated by some functional fit guarranteed to have a small enough error, pairs of neighboring regions can be merged if for each region the sum of the squares of the differences between the fitted coefficients for each region and the corresponding averaged coefficients, averaged over both regions, is small enough. Pavlidis gets his initial segmentation by finding the best way to divide each row of the image into segments with a sufficiently good fit. He also describes a combinatorial tree search algorithm to accomplish the merging which guarantees a best result.

Gupta, Kettig, Landgrebe, and Wintz (1973) suggest using a t-test based on the absolute value of the difference between the pixel and the region near as the measure of dis-similarity. Kettig and Landgrebe (1975) discuss the multi-band situation leading to the F-test and report good success with LANDSAT data.

Nagy (1972) just examines $|y-\bar{X}|$. If this distance is small enough pixel y is added to the region. If there is more than one region, then y is added to that region with smallest distance.

The Levine and Shaheen scheme (1981) is similar. The difference is that Levine and Shaheen attempt to keep regions more homogeneous and try to keep the region scatter from getting too high. They do this by requiring the differences to be more significant before a merge takes place if the region scatter is high.

Brice and Fennema (1970) accomplish the region growing by partitioning the image into initial segments of pixels having identical intensity. They then sequentially merge all pairs of adjacent regions if a significant fraction of their common border has a small enough intensity difference across it. Muerle and Allen (1968) suggest a Kolmogorov-Smirnov test for merging one region with another.

Simple single pass approaches which scan the image in a left right top down manner are, of course, unable to make the left and right sides of a V-shaped region belong to the same segment. To be more effective, the single pass must be followed by some kind of connected components algorithm in which pairs of neighboring regions having means which are close enough are combined into the same segment.

One minor problem with region growing schemes is their inherent dependence on the order in which pixels and regions are examined. A left right top down scan does not yield the same initial regions as a right left bottom up scan or for that matter a column major scan. Usually, however, differences caused by scan order are minor.

SPATIAL CLUSTERING

It is possible to determine the image segments by combining clustering in measurement space with a spatial region growing. Such techniques are called spatial clustering. In essence, spatial clustering schemes combine the histogram mode seeking technique with a region growing or a spatial linkage technique.

Haralick and Kelly (1969) suggest it be done by locating, in turn, all the peaks in measurement space. Then determine all pixel locations having a measurement on the peak. Beginning with a pixel corresponding to the highest peak not yet processed, simultaneously perform a spatial and measurement space region growing in the following manner. Initially, each segment is the pixel whose value is on the current peak. Consider for possible inclusion into this segment the neighbors of this pixel (in general, the neighbors of the pixel we are growing from) if the neighbor's N-tuple value is close enough in measurement space to the pixel's N-tuple value and if its probability is not larger than the probability of the pixel's value we are growing from. Matsumoto, Naka, and Yamamoto (1981) discuss a variation on this idea.

Milgram (1979) defines a segment for a single band image to be any connected component of pixels all of whose values lie in some interval I and whose border has a higher coincidence with the border created by an edge operator than for any other interval I. The technique has the advantage over the Haralick and Kelly technique in that it does not require the difficult measurement space exploring of climbing down a mountain. However, it does have to try many different intervals for each segment. Extending it to efficient computation in multiband images appears difficult. However, Milgram does report good results of segmenting white blobs against a black background. Milgram and Kahl (1979) discuss embedding this technique into the Ohlander (1975) recursive control structure.

Minor and Sklansky (1981) make more active use of the gradient edge image than Milgram but restrict themselves to the more constrained situation of small convex-like segments. They begin with an edge image in which each edge pixel contains the direction of the edge. The orientation is so that the higher valued gray tone is to the right of the edge. Then each edge sends out for a limited distance a message to nearby pixels and in a direction orthogonal to the edge direction. The message indicates what is the sender's edge direction. Pixels which pick up these messages from enough different directions must be interior to a segment.

The spoke filter of Minor and Sklansky counts the number of distinct directions appearing in each 3x3 neighborhood. If the count is high enough they mark the center pixel as belonging to an interior of a region. Then the connected components of all marked pixels is obtained. The gradient guided segmentation is then completed by performing a region growing of the components. The region growing must stop at the high gradient pixels, thereby assuring that no undesired boundary placements are made.

Burt, Hong, and Rosenfeld (1981) describe a spatial clustering scheme which is a spatial pyramid constrained ISODATA kind of clustering. The bottom layer of the pyramid is the original image. Each successive high layer of the pyramid is an image having half the number of pixels per row and half the number of rows of the image below it. Initial links between layers are established by linking each father pixel to the spatially corresponding 4x4 block of son pixels. Each pair of adjacent father pixels has 8 son pixels in common. Each son pixel is linked to a 2x2 block of father pixels. The iterations proceed by assigning to each father pixel the average of his son pixels. Then each son pixel compares his value with each of his father's values and links himself to his closest father. Each father's new value is the average of the sons to which he is linked etc. The iterations converge reasonably quickly and for the same reason the ISODATA iterations converge. If the top layer of the pyramid is a 2x2 block of great grandfathers, then these are at most 4 segments which are the respective great grandson of these 4 great grandfathers. Pietikainen and Rosenfeld (1981) extend this technique to segment an image using textural features.

SPLIT AND MERGE

The split method for segmentation begins with the entire image as the initial segment. Then it successively splits each of its current segments into quarters if the segment is not homogeneous enough. Homogeneity can be easily established by determining if the difference between the largest and smallest gray tone intensities is small enough. Algorithms of this type were first suggested by Roberston (1973) and Klinger (1973). Kettig and Landgrebe (1975) try to split all non-uniform 2x2 neighborhoods before beginning the region merging. Fukada (1980) suggests successively splitting a region into quarters until the sample variance is small enough.

Because segments are successively divided into quarters, the boundaries produced by the split technique tend to be squareish and slightly artificial. Sometimes adjacent quarters coming from adjacent split segments need to be joined rather than remain separate. Horowitz and Pavlidis (1976) suggest a split and merge strategy to take care of this problem.

Chen and Pavlidis (1980) suggest using statistical tests for uniformity rather than a simple examination of the difference between the largest and smallest gray tone intensities in the region under consideration for splitting. The uniformity test requires that there be no significant difference between the mean of the region and each of its quarters. The Chen and Pavlidis tests assume that the variances are equal and known.

The data structures required to do a split and merge on images larger than 512x512 are extremely large. Execution of the algorithm on virtual memory computers results in so much paging that the dominant activity is paging rather than segmentation. Browning and

Tanimoto (1982) give a description of a split and merge scheme where the split and merge is first accomplished on mutually exclusive subimage blocks and the resulting segments are then merged between adjacent blocks to take care of the artificial block boundaries.

REFERENCES

T. Asano and N. Yokoya, "Image Segmentation Schema For Low-Level Computer Vision," Pattern Recognition, Vol. 14, No. 1, 1981, pp.267-273.

C. Brice and C. Fennema, "Scene Analysis Using Regions," Artificial Intelligence, Vol. 1, 1970, p205-226.

J.D. Browning and S.L. Tanimoto, "Segmentation of Pictures into Regions With a Tile by Tile Method," Pattern Recognition, Vol. 15, No 1, 1982, pp.1-10.

P.J. Burt, T.H. Hong, and A.Rosenfeld, "Segmentation and Estimation of Image Region Properties Through Cooperative Hierarchial Computation" IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-11, No. 12. December 1981

J. Bryant, "On the Clustering of Multidimensional Pictorial Data", Pattern Recognition, Vol. 11, 1979, p115-125.

P.C. Chen, and T. Pavlidis, "Image Segmentation as an Estimation Problem," Computer Graphics and Image Processing, Vol. 12, 1980, pp. 153-172.

K.S. Fu and J.K. Mui, "A Survey On Image Segmentation", Pattern Recognition, Vol 13, 1981, p3-16.

Y. Fukada, "Spatial Clustering Procedures For Region Analysis", Pattern Recognition, Vol 12, 1980, p395-403.

M. Goldberg and S. Shlien, "A Clustering Scheme For Multispectral Image," IEEE Systems, Man, and Cybernetics, Vol. SMC-8, No 2, 1978, pp.86-92.

M. Goldberg and S. Shlien, "A Four-Dimensional Histogram Approach to the Clustering of LANDSAT Data," Machine Processing of Remotely Sensed Data, IEEE CH 1218-7 MPRSD, Purdue University, West Lafayette, Indiana, June 21-23, 1977, pp.250-259.

J.N. Gupta, R.L. Kettig, D.A. Landgrebe, and P.A. Wintz, "Machine Boundary Finding and Sample Classification of Remotely Sensed Agricultural Data," Machine Processing of Remotely Sensed Data, IEEE 73 CHO 834-2GE, Purdue University, West Lafayette Indiana, October 16-18, 1973, p4B-25 to 4B-35.

R.M. Haralick and G.L. Kelly, "Pattern Recognition with Measurement Space and Spatial Clustering for Multiple Image," Proceedings of IEEE, Vol. 57, No. 4, April 1969, pp.654-665.

R.M. Haralick and I. Dinstein, "A Spatial Clustering Procedure for Multi-Image Data," IEEE Circuits and Systems, Vol. CAS 22, No. 2, May 1975, pp.440-450.

R.M. Haralick, "Edge and Region Analysis For Digital Image Data", Computer Graphics and Images Processing, Vol. 12, 1980, p60-73. Also reprinted in Image Modeling a. Rosenfeld (Ed.), Academic Press, 1981 New York, p171-184.

R.M. Haralick, "The Digital Step Edge," to be published in 1982.

S.L. Horowitz and T. Pavlidis, "Picture Segmentation by a Tree Traveral Algorithm", Journal of the Association for Computing Machinery, Vol 23, No 2, April 1976, pp.368-388.

R.A. Jarvis and E.A. Patrick, "Clustering Using A Similarity Measure Based on Shared Near Neighbors," IEEE Transactions on Computers, Vol. C-22, No. 11, 1973, pp.1025-1034.

T. Kanade, "Region Segmentation: Signal vs Semantics" Computer Graphics and Image Processing, Vol. 13, 1980, p279-297.

R.L. Kettig and D.A. Landgrebe, "Computer Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects". The Laboratory for Application of Remote Sensing LARS Information Note 050975, Purdue University, West Lafayette, Indiana 1975.

A. Klinger, "Data Structures and Pattern Recognition," 1st International Joint Conference on Pattern Recognition, Washington D.C., Oct 1973 pp.497-498.

M.D. Levine and J. Leemet, "A Method For Non-Purposive Picture Segmentation," 3rd International Joint Conference on Pattern Recognition, 1976.

M.D. Levine and S.I. Shaheen, "A Modular Computer Vision System For Picture Segmentation and Interpretation", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-3, No.5, Sept. 1981.

K. Matsumoto, M. Naka, and H. Yamanoto, "A New Clustering Method For LANDSAT Images Using Local Maximums of a Multi-Dimensional Histogram," Machine Processing of Remotely Sensed Data, IEEE CH 1637-8 MPRSD, Purdue University, West Lafayette, Indiana, June 23-26, 1981, p321-325.

D.L. Milgram "Region Extraction Using Convergent Evidence", Computer Graphics and Image Processing, Vol. 11, 1979, p1-12.

D.L. Milgram and D.J. Kahl "Recursive Region Extraction", Computer Graphic and Image Processing, Vol 9, 1979, p82-88.

L.G. Minor and J. Sklansky "The Detection and Segmentation of Blobs in Infrared Images" IEEE Transactions on Systems Man and Cybernetics, Vol. SMC-11, No 3, March 1981, p194-201.

J. Muerle and D. Allen, "Experimental Evaluation of Techniques for Automatic Segmentation of Objects In a Complex Scene", Pictorial Pattern Recognition, G. Cheng et. al Eds., Thompson Book Co., Washington D.C., 1968, p3-13.

G. Nagy and J. Tolaba, "Nonsupervised Crop Classification Through Airborne Multispectral Observations", IBM Journal of Research and Development, Vol 16, No.2, March 1972, P138-153.

P.M. Narendra and M. Goldberg, "A Non-parametic Clustering Scheme for LANDSAT", Pattern Recogniton, Vol. 9, 1977, p207-215.

P.M. Narendra and M. Goldberg "Image Segmentation with Directed Trees", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-2, No 2, March 1980, p185-191.

R. Ohlander, Analysis of Natural Scenes, Ph.D. dissertation, Carnegie-Mellon University, Pittsburgh, Penn, 1975.

Y. Ohta, T. Kanade, and T. Sakai, "Color Information For Region Segmentation", Computer Graphics and Image Processing, Vol. 13, 1980, p222-241.

D.P. Panda and A. Rosenfeld, "Image Segmentation by Pixel Classification in (Gray Level, Edge Value) Space", IEEE Transactions on Computers, Vol. C-27, No. 9, September 1978, p875-879.

T. Pavlidis, "Segmentation of Pictures and Maps Through Functional Approximation", Computer Graphics and Image Processing Vol. 1, 1972, p360-372.

W.A. Perkins, "Area Segmentation of Images Using Edge Points", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-2, No. 1, January 1, 1980, p8-15.

M. Pietikainen and A. Rosenfeld, "Image Segmentation by Texture Using Pyramid Node Linking" IEEE Transactions on Systems, Man, Cybernetics, Vol. SMC-11, No. 12, December 1981.

E. Riseman and M. Arbib, "Segmentation of Static Scenes", Computer Graphics and Image Processing, Vol. 6, 1977, p221-276.

T.V. Robertson, "Extraction and Classification of Objects In Multispectral Images," Machine Processing of Remotely Sensed Data, IEEE 73 CHO 837-2GE, Purdue University, West Lafayette Indiana, October 16-18, 1973, p3B-27 to 3B-34.

S. Watanabe and the CYBEST group, "An Automated Apparatus for Cancer Prescreening: CYBEST", Computer Graphics and Image Processing Vol. 3, 1974, p350-358.

J.S. Weszka, "A Survey of Threshold Selection Techniques", Computer Graphics and Image Processing Vol. 7, 1978, p259-265.

J.S. Weszka and A. Rosenfeld, "Threshold Evaluation Techniques", IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-8, No.8, August 1978, p622-629.

Y. Yakimovsky, "Boundary and Object Detection in Real World Image", Journal of the Association for computing Machinery., Vol 23, No. 4, October 1976, p599-618.

S. Zucker, "Region Growing: Childhood and Adolescence" Computer Graphics and Image Processing, Vol 5, 1976, p382-399.