Reprinted from

Tenth International Symposium

Machine Processing of

Remotely Sensed Data

with special emphasis on

Thematic Mapper Data and

Geographic Information Systems

June 12 - 14, 1984

Proceedings

Purdue University
The Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907 USA

SCENE SEGMENTATION THROUGH REGION GROWING

R.S. LATTY

University of Maryland
Department of Civil Engineering
College Park, Maryland

I. ABSTRACT

Early work performed with simualted Thematic Mapper (TM) data, and later work with actual TM data, have demonstrated the inadequacy of per-point processing strategies for surface feature identification in many applications. The single pixel provides an increasingly inadequate representation of the surface feature as the spatial resolution of the imaging system increases. It is assumed here, with some substantiation from previous work (Wacker, 1972), that the accuracy of surface feature identification and description increases as the number of observations on which to infer the desired information increases. An algorithm for segmenting multispectral images into areas constituting surface features is presented. The scene segmentation employs a region growing approach which uses edge elements and their orientation as region delimiters. Those edge elements which most logically serve to delimit each particular region are determined individually for each potential region through an iterative search process in an expanding neighborhood about a set of arbitrarily initially positioned seeds. Adjacent regions which are not separated by edges are linked to form larger regions. Region growing is considered successfull if each region occurs in a single surface feature. Region linking is considered successful if each region provides a sufficient number of pixels with which to accurately identify the surface feature which the region represents, and the linked regions are all of the same feature. The region growing portion of the algorithm appears to work reasonably well, as all but a very small fraction of the regions appear to occur in a single feature. The region linking also appears to perform adequately, as approximately 80% of all of the test image pixels form regions of 100 pixels or more.

II. INTRODUCTION

The Thematic Mapper (TM) aboard Landsat 4 and 5 has provided the earth sciences with a data form which the community is not yet in a position to fully utilize with currently available computer-based image processing technology. The high spatial resolution and outstanding radiometric qualities result in images more appropriately processed with the well developed human visual system. As the spatial resolution, sampling frequency, optical transfer properties, and responsivity of a sensing system exceeds the predominant spatial frequencies of changing surface features in the scene being imaged, the image approaches that of the retinal image. That is, spatially speaking, the object observed through the imaging system does not differ dramatically from the object observed directly. To some, this has connoted an 'improvement' in the imaging system. If, however, the image processor (information extractor) component of the imaging system is not designed in accordance with the image properties as provided by the sensing component, the performance of the system may actually decrease. While this has led to bewilderment and frustration among the engineering community responsible for sensor design, the implications are extremely positive. That is, through the appropriate modifications and adaptations to the image processing component tremendous gains in the ability to provide highly detailed and accurate information can be achieved. These gains can be attained at costs far below those associated with additional upgrades to the sensor component.

The consequences to the imaging system performance of the TM data processed with Landsat MSS era processing technology is chronicled in detail with the work in aircraft scanner simulations (Kahn

& Ball, 1976; Sadowski & Sarno, 1977; Sadowski, et al., 1977; Morgenstern, et al., 1977; Landgrebe et al., 1977; Latty & Hoffer, 1981; Latty, 1981; Markham & Townshend, 1981). These studies conferred that increases in spatial resolution result in reduced classification accuracy when per-point type classifiers are used. These results are further corroborated by recent results obtained with actual TM data, simulated MSS data, and actual MSS data (Williams et al., 1983; Williams et al., 1984). The conclusion afforded by these studies, expressed in the simplest of terms, is that a single pixel provides an inadequate representation of surface features for the purpose of identifying the feature and distinguishing each from all other features present in a scene. This inadequacy increases with decreases in the area represented by the pixel.

The improvements in the information extraction capabilities with TM data will depend on the development of algorithms which are designed with the properties of TM data in mind. This will require formal models or concepts of scene structure and how this structure is manifest in the data. That is, the spectral and spatial properties of the image data in relation to the physical nature and arrangement of surface features will have to be identified and understood. In spite of the paucity of work conducted toward this end, particularly regarding spatial properties, numerous advances have been made in algorithm design relative to the properties of higher spatial resolution data. These advances, however, appear to be mostly heuristic in nature. The design of image processing algorithms has been based largely on the failures and shortcomings of predecessor algorithms, rather than an explicit and exacting formalization of the scene, the image, and the properties of the image which afford the extraction of detailed and accurate information regarding the scene. Contributions in the areas of scene formalization and its direct exploitation in algorithm design has been primarily from the artificial intelligence community, more specifically in robotics, where the variability of the imaged scene is very low and the <u>a priori</u> knowledge about the scene is <u>very high</u>, compared to scenes of interest to the earth sciences. The tremendous variation in the earth surface has discouraged the development of formal models, but has done little to diminish their need.

III. APPROACH

A. SOME PROPERTIES OF THE SCENE

While no formal model of the general scene is provided here, some properties of the scene are noted which serve as the basis for the design of the algorithm presented. The imaged segment of the earth surface is regarded as a two dimensional object or plane. This plane is comprised of areas, ai's (comprising the surface cover), which are discrete (more or less) and are definite (but unknown) in number. The property of being discrete signifies that the intersection of two areas (ai $\cap$ aj) is an empty set (while this may accurately depict most man induced landscapes, intersections may be a fairly accurate description of ecological transition zones). Transition zones in the discrete approach are assumed to be : a) sufficiently similar to aj as opposed to ai to be regarded as aj, b) sufficently similar to ai as opposed to aj to be regarded as ai, c) sufficiently dissimilar to ai and aj to be regarded as a distinct area ak. The dimensions of each area are assumed to be large compared to the width of any transition zones which fail to satisfy these conditions. The union of all ai's completely defines the plane or imaged segment. Each area ai has a set of physical properties or attributes which distinguish it from all other areas and serves to identify it as ai. These properties serve not only as the basis on which the surface features are distinguished, but constitute the basis for any interest in these features.

Regarding the scene as an assemblage of discrete areas logically leads to approaches which segment or partition the scene into discrete areas prior to attempting to conduct any identification or description. Initial segmentation of the scene allows the use of multiple observations over a feature in the attempt to identify it. Since, for an area of fixed size and dimension, the number of observations (i.e., pixels) representing an area increases approximately as a square of the increase in spatial resolution. The motivation for segmenting the scene, therefore, increases with the spatial resolution of the imaging system.

B. APPROACH TO SEGMENTATION

Localized iterative region growing is employed as the means for segmenting the scene into discrete ares. The algorithm is defined at a conceptual level in the flow diagram of Figure 1. Basically, the original multispectral image is used to

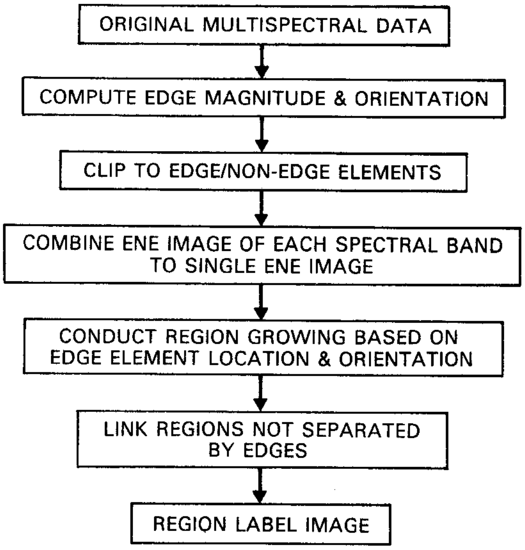


Figure 1. Conceptual level flow diagram for scene segmentation.

compute an image of edge magnitudes and orientations. This is performed through convolving a series of matricies with each corresponding window in the image. Each matrix is comprised of constants valued and arranged to represent a contrast of a particular compass orientation (see Latty, 1984 for details). In this case twelve matricies were used to represent 15 degree increments. The edge magnitude and orientation (EMO) image of each spectral band is used to compute an image of edge/non-edge elements through iterative clipping based on the location and orientation of the local edge magnitude maxima. The edge/non-edge (ENE) images for all spectral bands are then combined into a single ENE image. Contradictions as to location and orientation of edge elements are resolved based on the location and orientation of local edge magnitude maxima between spectral bands. These procedures are detailed elsewhere (Latty, 1984). The final ENE image is used to direct the iterative region growing through specifying the region limits.

Region Growing. A grid of 'seed' elements is superimposed on the ENE image. Each seed address is examined in the ENE image to determine whether the seed occurs on an edge element (i.e., non-zero value). Seeds which occur on an edge element are shifted half the distance between seeds in the direction perpendicular to the orientation of the edge element on which they occur. Any seeds remaining on edges subsequent to being shifted are ignored during region growing.

This region growing alorithm involves two steps - a region limit determination step, and a region membership determination step. The first step is to determine that set of edge elements which occurs in the neighborhood of each seed which defines the limits of the region in which the seed occurs. The second step is to determine all of the pixels which occur in the region defined by all of the egde elements found in the neighborhood. The domain of the neighborhood is defined by the location and orientation of previously encountered edge elements and the width of the search space. The width of the search space should, at a minimum, be equal to the spacing between seed elements. Since it is possible for edge elements which define the limits of a region to occur between other proximal seeds, the width of the search space should be somewhat larger than the distance between adjacent seeds. The accuracy of region definition should increase with increases in the width of the search space, however, processing time dramatically increases with the dimension of the search space.

Each of the above steps involves two phases - an expansion phase and a testing phase. The expansion phase simply increments a pointer to the next pixel address to be tested for a given seed. The testing phase executes a series of tests on the pixel of the current address. These tests depend on whether the process is in the limits determination step or the membership determination step. All of the following tests are executed in the region limit finding step, and all but the last test is conducted for the region membership finding step. The tests are as follows :

1) Pixel located in subimage resident in memory ?
2) Pixel located on seed side of edge for each edge element previously encountered for current seed ?
3) Pixel not an edge element ?

For each seed, the expansion and testing are conducted until the expansion phase has addressed every pixel in the current layer about the seed (see Figure 2 for the addressing sequence and the arrangement of layers about the seed). The cycle is then performed for the next layer of the next seed. The number of times this entire set of seeds is iterated through the layer expansion process is equal to the width of the search space, or until no pixel of a given layer passes all three of the above tests for all seeds. Any seed for which a layer fails to have at least one pixel pass all of the above tests, is regarded as fully determined, and is, therefore, skipped in subsequent

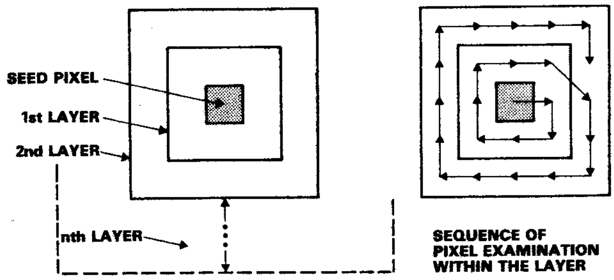


Figure 2. Schematic representation of the address sequence followed in the expansion phase and the arrangement of layers about the seed.

layer expansion iterations. Figure 3 provides a flow diagram summarizing the region growing algorithm.

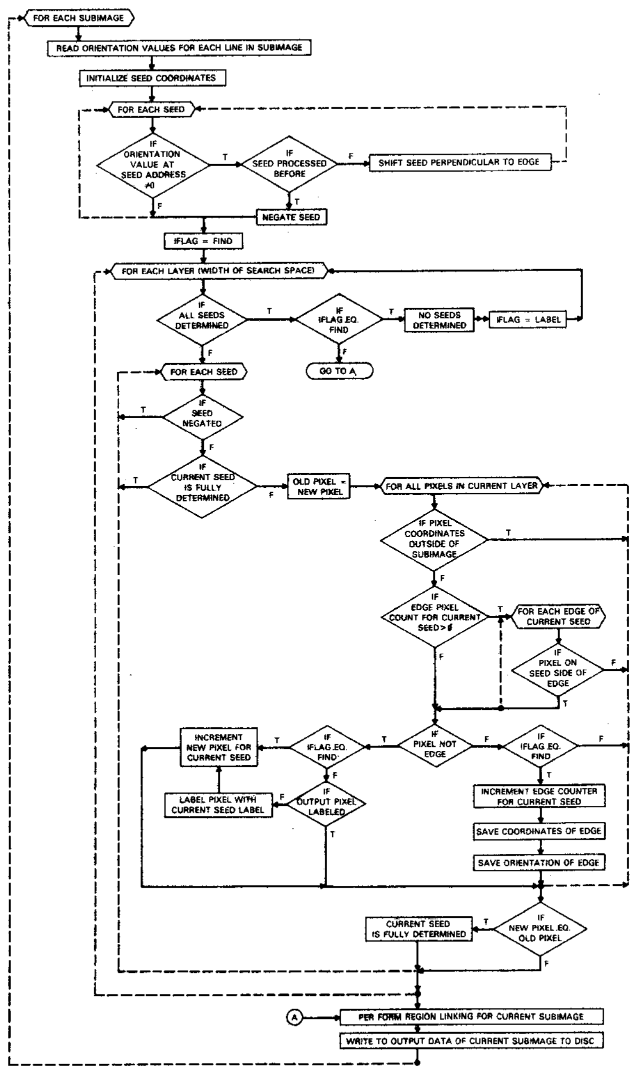


Figure 3. Detailed flow diagram for region growing.

The second test is the part of the algorithm which embodies the computational definition of a region - any two pixels belong to the same region if they occur on the same side of all edge elements which delimit the region. The regions are defined by taking an arbitrary pixel (the seed) and evaluating every pixel in a concentric, expanding neighborhood about the seed as to its location relative to the determined edge elements. Figure 4 illustrates how this test for 'seed-sidedness' is conducted. If the distance between the seed pixel and the current pixel is less than the distance between the seed pixel and the intersection of the line defined by the seed and the current pixel ($\overline{SC}$) and the line defined by the edge element and its orientation ($\overline{E\Theta}$), then the current pixel is on the seed side of the edge element. The values for the coordinates of the seed pixel (Xs,Ys), the current pixel (Xc,Yc), and the edge element (Xe,Ye), and the edge orientation (Θe), are known. The coordinates of the intersection (Xi,Yi) of $\overline{SC}$ and $\overline{E\Theta}$ are unknown. Solving for Xi and Yi involes a linear system of two equations :

$$\tan\Theta = (Ye-Yi)/(Xe-Xi) \quad (1)$$
$$(Yi-Ys)/(Xi-Xs) = (Yc-Ys)/(Xc-Xs) \quad (2)$$

The fact that the edge orientation equals the slope of the line defined by the intersection and the edge element provides Eq. 1. Since the three points (Xs,Ys), (Xc,Yc), and (Xi,Yi) lie on the same line, they define the same slope, which provides Eq. 2. Rearranging Eq. 2,

$$Yi = (Xi-Xs)(Yc-Ys)/(Xc-Xs)+Ys \quad (3)$$

Substituting Eq. 3 into Eq. 1 and rearranging,

let $a = (Yc-Ys)/(Xc-Xs)$, then

$$Xi = (Ye+Ys+aXs-Xe*\tan\Theta)/(a-\tan\Theta) \quad (4)$$

After solving for Xi,Yi, the distances between (Xs,Ys) and (Xc,Yc), and between (Xs,Ys) and (Xi,Yi) are computed and compared. If line segment $\overline{SC}$ is shorter than $\overline{SI}$, then the current pixel is in the same region as that of the seed, relative to that particular edge.

Regions which are not separated by edge elements are regarded as regions restricted in size by the seed spacing rather than the size of the feature in which they occur. Therefore, regions are tested for possible linking after they are fully grown.

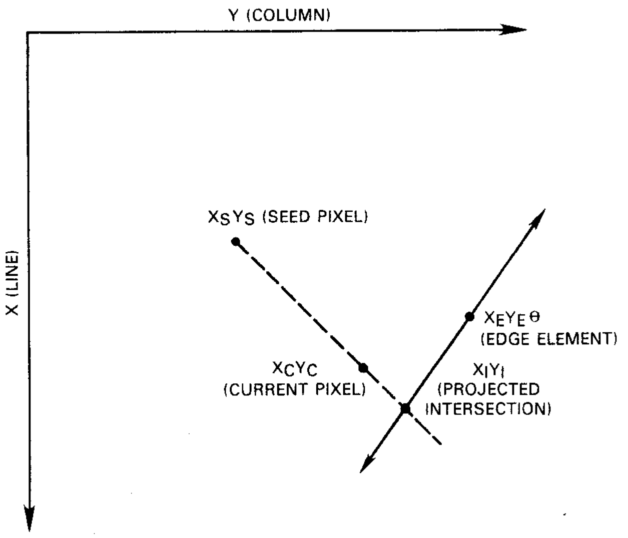


Figure 4. Geometry of test for 'seed-sidedness'.

Region Linking. Two adjacent regions are linked if, for all edge elements associated with each of the two regions, no edge element intersects the line segment defined by the two region seeds. This test is performed by allowing one of the region seeds to serve as a 'seed' and the other to serve as a 'current' pixel. The tests for seed-sidedness are performed, as in the region growing. The roles are then reversed, and the tests are repeated.

The link candidates are the south and east neighbors. The process moves first east, then south. Hence, duplication is avoided yet all four neighbors are tested for suitable links. No explicit tests are performed for the four diagonal neighbors (i.e., northeast, northwest,...), since no vector can intersect a diagonal of a quadralateral without intersecting at least two of the edges. Consequently, the diagonal neighbors are linked through the linking of cardinal elements.

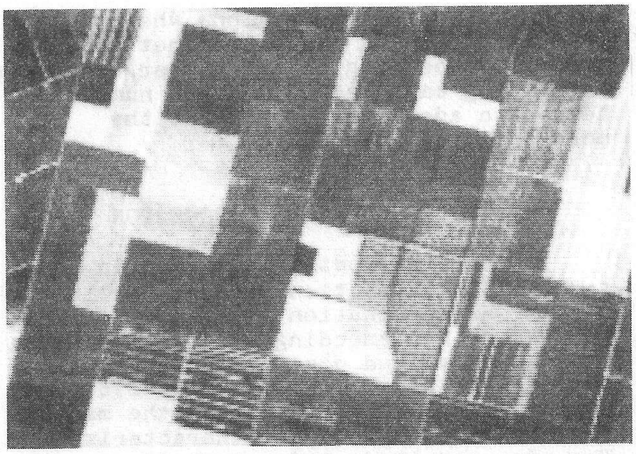
The continuity of the linking can be disrupted by the occurrence of edge spurs or false edges. The testing for and storage of forks can be used to overcome the link breaks caused by spurs and false edges. A fork is a condition which arises when linking tests are passed for a multiple of neighbors subsequent to a link test failure for two regions in the vacinity. The storage and processing of all multiple links would be wasteful and unnecessary, since not all multiple links result in forks. If a set of multiple links which result in a fork is not detected, recorded, and appropriately processed, it will result in a break of the linking series. A break in the linking series generates two regions where only one exists. This may or may not cause a problem, depending on the number of pixels comprising each region and the number needed to adequately represent the corresponding feature.

IV. RESULTS AND DISCUSSION

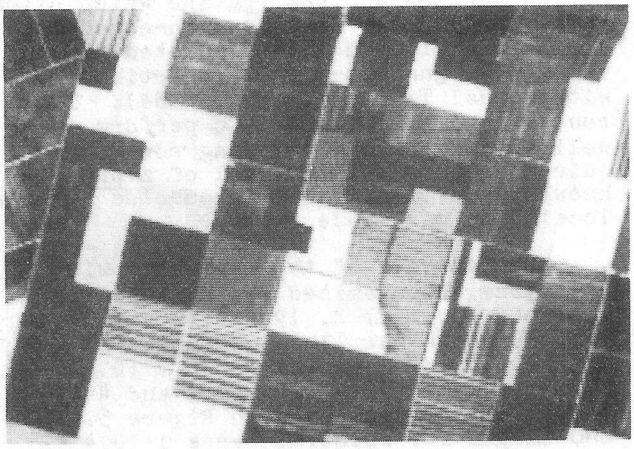
As region growing is a step in the information extraction process, full testing and evaluation will have to await the design and encoding of routines for : 1)spectrally (and possibly spatially) characterizing the regions, and 2)discriminant functions which employ the manner in which the regions are characterized. The edge magnitude and orientation computation routine and the routine which clips this EMO image to the edge/non-edge image has been evaluated with simulated data of variable signal-to-noise properties and with actual TM data (Latty, 1984). These routines have been found to perform quite well. However, the clipping routine requires an undesirable level of a priori knowledge of reasonable thresholds for the local edge magnitude maxima.

Performance of the region growing algorithm was examined for TM data obtained November 2, 1982 over northeastern North Carolina. Images of the P-tape data of spectral bands 2 (0.52 - 0.60 um), 3 (0.63 - 0.69 um), and 4 (0.76 - 0.90 um) are provided in Figure 5. The EMO images for spectral bands 1 (0.45 - 0.52 um), 2, 3, 4, and 7 (2.08 - 2.35 um) were generated. The EMO images for spectral bands 2, 3, and 4 are shown in Figure 6. The brightness of points in Figure 6 is proportional to the average directional contrast of the areas on either side of the point. The EMO images of spectral bands 1, 2, 3, 4, and 7 were used to generate the ENE images for each of these spectral bands. The ENE images of bands 2, 3, and 4 are shown in Figure 7. The ENE images of each of the five bands were used to form the combined ENE image (Figure 8). The combination of the ENE images of different spectral bands is required since earth surface features exhibit contrast in some spectral bands but not in others. The band combination approach employed here is only one of many possible means of combining the different spectral sources of information. Another approach is performing some spectral transform which results in dimension reduction (e.g., principle components transformation, Tassle-Cap transform,...) on the original spectral data prior to edge detection.

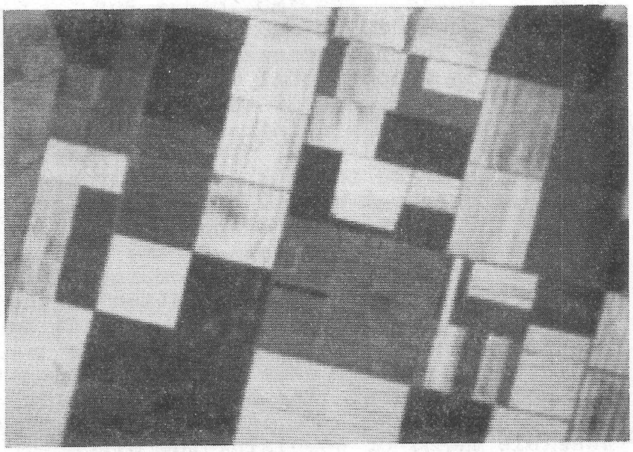
The region growing is considered successful if each region occupies a



(a)

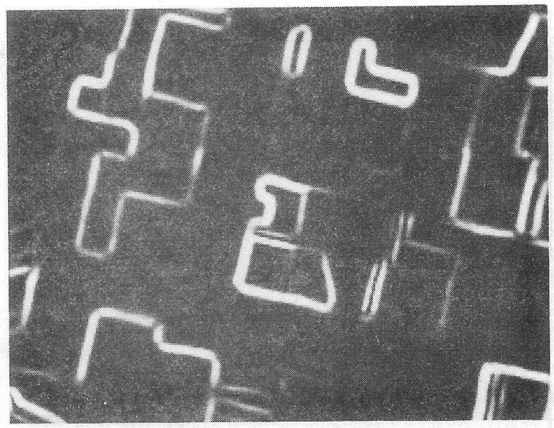


(b)

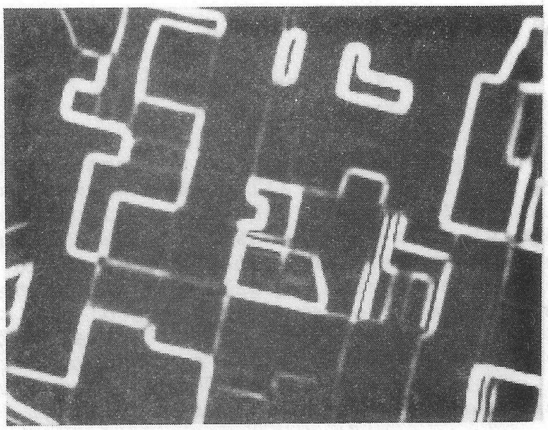


(c)

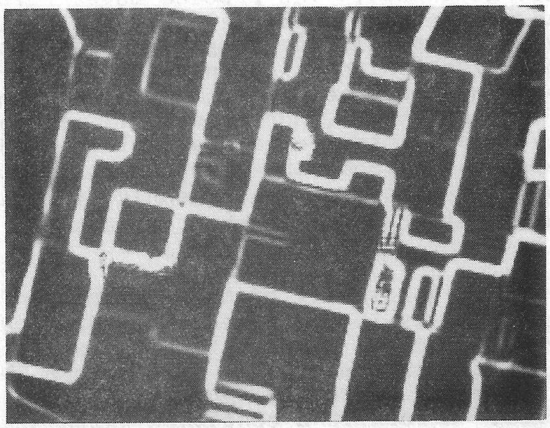
Figure 5. Thematic Mapper data obtained over northeastern North Carolina, November 2,1982: a) band 2 (0.52-0.60 um), b) band 3 (0.63-0.69 um), c) band 3 (0.76-0.90 um).



(a)



(b)

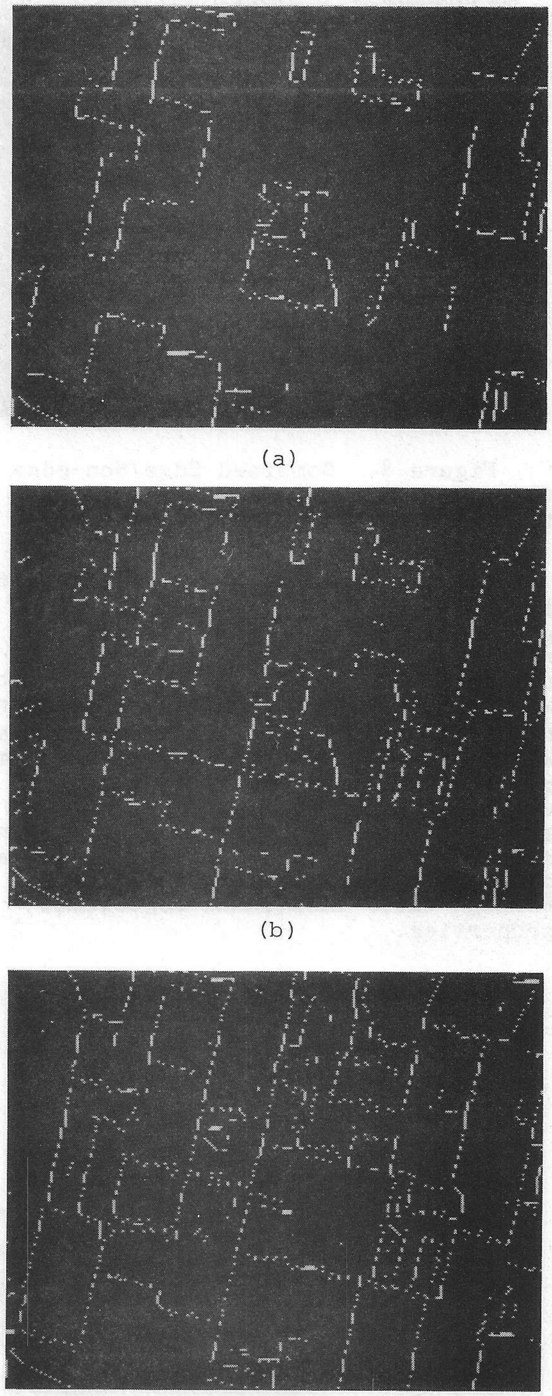


(c)

Figure 6. Edge magnitude images generated from images shown in Figure 5 using 7-by-7 matricies of 15 degree incremental rotations: a) band 2, b) band 3, c) band 4.

single surface feature. In addition to the above criterion, the success of region linking requires that each linked region contain a sufficient number of pure pixels to accurately characterize the surface feature in which it occurs. The performance of the region growing algorithm is dependent on the accuracy of the combined ENE image. If there are edges in the spectral image which the EMO and subsequent ENE generation processes fail to accurately represent in the combined ENE image, then regions will grow into neighboring features. Consequently, a single region will occupy more than one feature. The region growing does not introduce factors or parameters which must be accurately estimated by the analyst for each particular scene, nor which vary within the processes of the routine. The region growing routine employs only the data of the combined ENE and performs with no direct reference to the spectral data. The region growing process makes certain assumptions about the combined ENE image in relation to the arrangement of surface features, and performs in a set and consistent manner based on these assumptions. Figure 8 illustrates the linked region image generated from the ENE image of Figure 7. Comparing Figure 8 with the original spectral data of Figure 5 provides a qualitative idea as to the performance of the region growing. In general, the region growing works very well. It can be seen that the road in the upper-left area, and the narrow linear features in the center right are poorly represented. It appears from the EMO images of Figure 5 and the ENE images of Figure 6 that the edges of these features are poorly represented and subsequently suppressed. Falsely detected edges also occur and result in abreviated regions. The generation of more than one region per feature will not upset the information extraction process as long as the number of pixels in each region is sufficient to accurately represent the feature in which they occur.

The design of the region growing algorithm does not require that the edges in the combined ENE image be continuous (without gaps). The frequency of detected edge elements along a true edge need only be high enough such that the distance between detected edge elements is less than the width of the search space. This is an important performance property of this region growing approach, since the omission of edges along an actual edge has long been a hinderance to scene segmentation approaches based on edge detection (Gupta and Wintz, 1973; Robertson, et al., 1973; Gupta, et al., 1973).



(a)

(b)

(c)

Figure 7. Edge/Non-edge images generated from EMO images shown in Figure 6 using an iterative 4-by-4 saltatory window for determining local maxima: a) band 2, b) band 3, c) band 4.

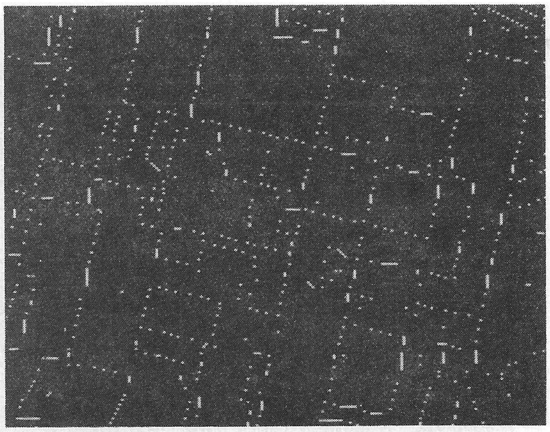


Figure 8. Combined Edge/Non-edge image generated through combining the ENE images of TM bands 1,2,3,4, and 7.

Table 1 shows the percentage of the image represented by regions equal to or greater than a series of sizes. With nearly 80% of the image in regions of 100 pixels or more, the regions should provide well defined spectral characterizations of the features which the regions represent. Wacker (1972) showed that with 40 or more pixels, features which differ only in co-variance (i.e., equal means) could be accurately distinguished. In addition, many of the remaining 20% of the image pixels can be accurately identified through their individual contextual properties.

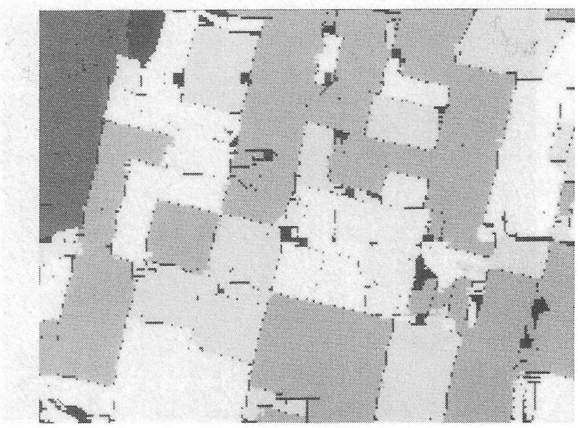


Figure 9. Linked region image generated through region growing and subsequent linking on the combined ENE image of Figure 8.

Table 1. Profile of percentage of image represented by regions of various sizes.

Number of Pixels in a Region	Percent of Image	Number of Regions
1000	45.5 %	10
500	58.5 %	18
250	70.0 %	32
100	79.5 %	58
60	84.1 %	185

V. SUMMARY AND CONCLUSION

The region growing algorithm presented demonstrates itself as a promising tool for information extraction from TM data and, potentially, other high resolution systems. The small instantaneous--field-of-view (IFOV), sampling frequency, optical transfer properties, and responsivity of TM results in images which possess very clear and distinct edges between adjacent surface features. These edges and their orientations can be readily extracted from the spectral data. Although the occurrence of planar, linear, and point sources of contrast in the same image results in some confusion in edge identification, a sufficient amount of work has been conducted in edge detection, line detecion, and point detection to provide design initiatives for estimating pixel properties relative to all three simultaneously. This should greatly improve edge detection and orientation estimation. Once the edges are accurately represented the region growing performs very well. Whether, or to what degree, the number of pixels in each region is adequate for accurately identifying or describing each region is not presently known. This will have to await the development of appropriate spectral and spatial descriptors and discriminant functions. This work is underway currently.

Besides offering potentially higher identification accuracies at higher levels of detail than are currently achieved with per-point approaches, there are two additional advantages in the scene segmentation appoaches. One of these advantages is that by identifying a single pixel within a region, all region members have been identified. This can provide a large number of observations for training the discriminant functions with very little effort. This is particularly valuable to industries or efforts which have available a database of points or small areas of known geographical location which are closely monitored. These points could be used in a highly automated fashion for rendering information about the rest of the scene. The second advantage is that once an area has been identified as to the type of surface cover, detailed analysis on a pixel-by-pixel basis can be performed for extracting higher level information regarding the properties of the region or segments (e.g., areas known to be forests can be examined for crown closure estimates, species mix levels, etc.).

Lastly, with the advent and availability of readily programmable parallel processors and super-computers, and the steady decline in hardware costs, the imagination is becoming less inhibited by processing time concerns.

VI. ACKNOWLEDGEMENTS

The author would like to express his sincere thanks to Malcom Tarlton for preparing the graphs and flow diagrams, and to NASA for supporting this work under Contract No. NAG9-5.

VII. REFERENCES

1. Gupta, J.N., and P.A. Wintz, "Closed Boundary Finding Feature Selection and Classification Approach to Multi-Image Modeling", LARS Information Note 062773, 1973. 31 p.

2. Gupta, J.N., R.L. Kettig, D.A. Landgrebe, and P.A. Wintz, "Machine Boundary Finding and Sample Classification of Remotely Sensed Agricultural Data", Machine Proc. of Remotely Sensed Data Sym., 1973. pp. 4B25-35.

3. Kan, E.P.F. and D.A. Ball, "Data Resolution Versus Forestry Classification', Lockheed Electronics Company Cntract NAS 9-12200, 1974. 32 p.

4. Landgrebe, D.A., L. Beihl, and W. Simmons, "An Empirical Study of Scanner System Parameters", IEEE Trans. on Geoscience Electronics, Vol. GE-15, No. 3, July 1977. pp.120-130.

5. Latty, R.S., "Computer-based Forest Cover Classification Using Multispectral Scanner Data of Different Spatial Resolutions", M.S. Thesis, Purdue University, West Lafayette, IN. 1981. 187 p.

6. Latty, R.S., "Minimun Bias Edge Orienation Estimators for Remotely Sensed Digital Images", in press, 1984.

7. Latty, R.S. and R.M. Hoffer, "Computer--based Classification Accuracy due to Spatial Resolution Using Per-Point Versus Per-Field Classification Techniques", Proc. of the 1981 Machine Proc. of Remotely Sensed Data Sym., 1981. pp. 384-393.

8. Markham, B.L. and J.R.G. Townshend, "Land Cover Classification Accuracy as a Function of Sensor Spatial Resolution", Proc. of the 15-th Int. Sym. on Remote Sensing of Env., 1981. pp. 1075-1090.

9. Morgenstern, J.P., R.F. Nalepka, and J.D. Erickson, "Investigation of Thematic Mapper Spatial, Radiometric, and Spectral Resolution", Proc. of the 11-th Int. Sym. on Remote Sensing of Env., 1977. pp. 1279-1288.

10. Robertson, T.V., K.S. Fu, and P.H. Swain, "Multispectral Image Partitioning", LARS Information Note 071373, 1973. 90 p.

11. Sadowski, F. and J. Sarno, "Forest Classification Accuracy as Influenced by Multispectral Scanner Spatial Resolution", NASA CR ERIM 109600-71, 1976. 116 p.

12. Sadowski, F., W.A. Malila, J.E. Sarno, and R.F. Nalepka, "The Influence of Multispectral Scanner Resolution on Forest Feature Classification", Proc. of the 11-th Int. Symp. on Remote Sensing of Env., 1977. pp.1279-1288.

13. Wacker, A.G., "Minimum Distance Approach to Classification", Ph.D. Dissertation, Dept. Elec. Eng., Purdue University, West Lafayette, IN, 1972. 345 p.

14. Williams, D.L., J.R. Irons, B.L. Markham, R.F. Nelson, D.L. Toll, R.S. Latty, and M.L. Stauffer, "Impact of Thematic Mapper Sensor Characteristics on Classification Accuracy", IEEE Proc. of the 1983 IGAR Symp.

15. Williams, D.L., J.R. Irons, B.L. Markham, R.F. Nelson, D.L. Toll, R.S. Latty, and M.L. Stauffer, "A Statistical Evaluation of the Advantages of Landsat Thematic Mapper Data in Comparison to Multispectral Scanner Data", in press 1984.

Richard S. Latty received his B.S. degree in 1976 at the University of Florida, and his M.S. degree in 1981 at Purdue University. He is presently with the University of Maryland, Hydrology Remote Sensing Laboratory, as Assistant Faculty Researcher, under support from NASA/Goddard. He has conducted work in Thematic Mapper simulation, scan angle effects on measured radiance, and information extraction algorithm design. His primary interests are in algorithm development for processing high spatial resolution data and image syntax.