Symbolic Knowledge Representation for Remote Sensing

Gordon W. Plunkett and David G. Goodenough
Canada Centre for Remote Sensing
2464 Sheffield Road
Ottawa, Ontario, Canada
K1A 0Y7

email: plunkett@ccrs.cdn

Abstract

One of the more elusive goals in the field of remote sensing is to develop a completely automated feature extraction and identification system. The output of such a system is usually an annotated map, and thus such a system is often referred to as fully automated computer cartography system. To date only computer assisted cartographic systems have been developed. Human photointerpreters are still required for feature identification and verification. There are many challenging steps along the way to achieving this goal. A mix of digital image analysis techniques and expert systems technology has shown some successes in this area. However one of the biggest stumbling blocks at the moment, is how to represent the knowledge contained in a scene in a compact reliable structure that is compatible with expert systems. This paper outlines this knowledge representation problem, presents an analysis of various image segmentation techniques, discusses the segment attributes required for inferencing and discusses some of the problems associated with this technique.

Introduction

Remotely sensed images contain a wealth of information that can be readily processed by trained human experts, such as photointerpreters. Humans can almost instantaneously identify objects or make some other semantic judgments about an image. Some researchers are using algorithmic techniques to identify and classify cartographic features [Guindon 88]. Other researchers have shown that expert systems can be engineered to perform inferenced reasoning about simple human visual tasks (simple in the human visual system context). [Rao 88]. However, one of the problems that has plagued expert system researchers is how to represent image knowledge in an expert system. This problem is often referred to as the iconic to symbolic gap problem.

In essence the iconic to symbolic gap problem is : How does one convert the spatial, spectral or other knowledge that is stored in an image into a semantical form, that can be processed by an expert system? Typically expert systems store knowledge in a symbolic format. Computerized image processing systems typically store imagery as a large often multi-dimensional matrix of binary numbers. Thus an expert system requires the binary matrix to be converted to a symbolic format, for inference processing.

Many researchers have found that image segmentation has provided meaningful results. Some researchers have used contextual [Stansfield 86] segmenting techniques and others have used non-contextual segmenting techniques [Plunkett 86]. However, once the segments have been determined, then segment attribution can be performed relatively easily. The attributed segment information, can then be passed to the inference engine.

This paper provides an analysis of the concepts and some results of applying various image segmentation techniques to an image and attributing the segments for applying inference rules to perform photogrammetric analysis. The study scene is a geocoded Landsat Thematic Mapper scene of the Ottawa area. A subarea of this image was segmented using several gradient operators. Different thresholds were then used on the segmentation operation. Following segmentation, each segment was attributed with a number of attributes that relate to the segments spectral and spatial properties. These attributes were stored in a flat ASCII file as PROLOG statements and input to an Expert system for analysis.

Methodology

Some authors have indicated that there is a hazy or grey area developing between the image processing and knowledge based processing parts of computer vision systems [Straight 87]. Other authors [Rao 88] have described the differences as low-level vision is the image processing stage and the knowledge based phase is the high level vision stage. Other authors have performed the low-level vision stage manually by humans and then fed this knowledge into the inference engine [Pascucci 87]. Pascucci does however indicate that some research is ongoing to perform some of the low-level vision via computer.

This paper uses the paradigm that the low-level vision is performed by conventional image processing techniques and the high-level vision is performed by expert system techniques. However, the area on which this paper concentrates, is how the results of the low-level vision techniques can be computed, formatted, and stored for further processing.

One of the most common methods of performing image interpretation using artificial intelligence techniques is to preprocess the image data by generating various labeled image segments. This preprocessing stage allows the knowledge based portion of the system to make inferences on segments of the image, rather than on a pixel-by-pixel basis. The preprocessing stage is a very important portion of any vision expert system, because it can significantly improve the system performance, if appropriate segments are generated.

A typical method of performing this segmentation is to use the Sobel gradient operator [Stansfield 86] [Plunkett 86]. The resulting segments are however often not truly indicative of the original image. One finds that areas which are relatively constant in content (ie a river in a remotely sensed image), may be segmented into a number of smaller segments. This is due to the side-effects of the Sobel operator. The reason for this phenomenon is that, if there are small variations in the grey level of the pixels in these areas, the variations are amplified during the application of the gradient and are significant enough to be above the threshold used for the segmentation stage.

Another problem is that if the threshold of the gradient that is selected, is too high, then significant segments may not be computed. On the other hand, if the threshold is set too low, then a large number of segments are produced.

For robotic vision applications non-contextual segmentors are often sufficient, because these systems work under constant lighting and environmental conditions. Usually these system need only identify an object or not. Thus for robotic vision the same set of low-level vision operations may be performed on each scene. In remote sensing however, there are often many more parameters, that may vary. Also the type of target that the system is evaluating may reflect the type of low-level vision operations that may be applied.

Image Processing Methodology

The first step in any segmentation paradigm, is to apply an image gradient operator, and then segment the gradient image based on a selected threshold. A non-contextual segmentation paradigm is displayed in Figure 1. A contextual segmentation paradigm is shown in Figure 2.

Several gradient operators were used in this study. A brief description of the operators follows, which is based on the following conventions:

The elements of a 3 x 3 pixel window in channel k can be represented by:

$$X_{ijk} \rightarrow \begin{bmatrix} X_{11k} & X_{12k} & X_{13k} \\ X_{21k} & X_{22k} & X_{23k} \\ X_{31k} & X_{32k} & X_{33k} \end{bmatrix}$$

where X_{ijk} is the video value of line i, pixel j of channel k, and X_{22k} is the video value of the center element of channel k in a 3x3 matrix.

The elements of a 2 x 2 pixel window are similarly represented as:

$$X_{ijk} \rightarrow \begin{bmatrix} X_{11k} & X_{12k} \\ X_{21k} & X_{22k} \end{bmatrix}$$

Sum of Maximum Differences

The sum of the maximum differences in N channels over a 3x3 window gradient operator consists of computing the sum of the maximum absolute difference in N channels for each neighbor in a 3x3 window. Then the operator can be written as:

$$Y_{22k} = \sum_{i=1}^{3}\sum_{j=1}^{3} MAX_{k=1}^{3}\,[X_{ijk} - X_{22k}]$$

This means that for each of the eight directions, the gradient is taken and the largest absolute difference over the N channels is retained. Thus, for example, one big change in 1 channel will be more important than 4 small changes in 4 channels.

Sum of Differences

The sum of the differences in N channels over a 3x3 Window gradient, calculates the sum of the absolute differences in N channels for each neighbor in a 3x3 window.

$$Y_{22k} = \sum_{k=1}^{3}\sum_{i=1}^{3}\sum_{j=1}^{3} [X_{ijk} - X_{22k}]$$

This operator is smoother than the sum of maximum differences operator, since all channels have equal weight.

Maximum Sum of Differences

The maximum sum of differences in a 3x3 window over N channels gradient calculates the maximum of the sum of absolute differences for all neighbors in a 3x3 window over N channels.

$$Y_{22k} = MAX_{k=1}^{3}\sum_{i=1}^{3}\sum_{j=1}^{3} [X_{ijk} - X_{22k}]$$

Roberts Gradient Distance 1

The Roberts gradient cross operator of distance 1 in a 2x2 window [Gonzales 77] is computed as the sum of the absolute difference between the diagonal pixels in channel k.

The following weighting functions are used:

$$\begin{array}{|c|c|}\hline 1 & 0 \\ \hline 0 & -1 \\ \hline \end{array} \qquad \begin{array}{|c|c|}\hline 0 & 1 \\ \hline -1 & 0 \\ \hline \end{array}$$

Then the operator can be written as

$$Y_{21k} = |X_{11k} - X_{22k}| + |X_{12k} - X_{21k}|$$

Because of the size of the window, this is a sharp operator, but it is not symmetric because the resulting output will be placed at the lower left of the window. Therefore, there is a 0.5 pixel misregistration in the line and pixel directions.

Roberts Gradient Distance 2

The Roberts gradient cross operator of distance 2 in a 3x3 window is computed as the sum of the absolute differences between the opposite pixels in a vertical and horizontal direction in channel k [Gonzales 77].

The following weighting functions are used:

0	0	0
1	0	-1
0	0	0

0	1	0
0	0	0
0	-1	0

Then the operator can be written as:

$$Y_{22k} = |X_{21k} - X_{23k}| + |X_{12k} - X_{32k}|$$

This operation is smoother than the Roberts gradient distance 1 operator and it is symmetric. However, it does not take into account diagonal differences.

Sobel Gradient

The Sobel Gradient [Duda 71] [Levine 85] operator over a 3x3 window in channel k is computed using the following weighting functions:

Vertical		
1	0	-1
2	0	-2
1	0	-1

Horizontal		
1	2	1
0	0	0
-1	-2	-1

The Sobel operator can be written as:

$$V_k = X_{11k} + 2X_{21k} + X_{31k} - (X_{13k} + 2X_{23k} + X_{33k})$$

$$H_k = X_{11k} + 2X_{12k} + X_{13k} - (X_{31k} + 2X_{32k} + X_{33k})$$

$$Y_{22k} = S_k = \sqrt{V_k^2 + H_k^2}$$

Laplacian Operator

The Laplacian [Gonzales 77] operator over a 3x3 window in channel k is computed using the following weighting function:

0	1	0
1	-4	1
0	1	0

The Laplacian operator can be written as:

$$Y_{22k} = X_{12k} + X_{32k} + X_{21k} + X_{23k} - 4X_{22k}$$

Variance Operator

The variance operator for an M x N window [Sachs 82] for channel k is computed using the M and N parameters to vary the degree of smoothing in the gradient operator.

$$VAR_k = \frac{1}{n-1}\left[\sum_{i=1}^{M}\sum_{j=1}^{N}(X_{ijk})^2 - \frac{\left(\sum_{i=1}^{M}\sum_{j=1}^{N} X_{ijk}\right)^2}{n}\right]$$

where n is the size of the window or M x N. For this comparison, only a 3x3 window was used.

Expert System Methodology

The Canada Centre for Remote Sensing (CCRS) has developed an expert system shell, based on the MPROLOG language, called RESHELL [Goodenough 87]. The attributed segments must therefore be converted into a format that is readable by PROLOG. The shell supports a frame based structure based on the following syntax:

frame (object , attribute , facet , value).

In this case each segment is treated as an object and is given an appropriate unique name (ie segment-53), based on the segment label number assigned to that segment during the segmentation phase.

There are a large number of both spectral and spatial attributes that are currently being computed for each segment. These include :

- size - this attribute is the number of pixels within the segment.
- shape - this attribute is the shape of the segment as defined by:

$$segment\ shape = \frac{(segment\ perimeter)^2}{segment\ area}$$

- bounding rectangle (or window) - this attribute indicates the maximum and minimum pixel and line values. By definition, all pixels belonging to this segment are within the bounding rectangle.
- maximum grey level for each channel - this attribute specifies the highest pixel value found in this segment in each channel.
- minimum grey level for each channel - this attribute specifies the lowest pixel value found in this segment in each channel.
- mean grey level value for each channel - this attribute specifies the mean grey level value for this segment in each channel.
- grey level variance for each channel - this attribute specifies the variance about the mean of the grey levels for this segment for each channel.
- covariances for all channels - this attribute specifies the covariance of the grey levels for this segment between each pair of channels.
- pixel locations - this attribute specifies the exact pixel locations of the pixels in this segment.

The segmented image is converted into a PROLOG symbolic format file by a FORTRAN program. This file is then input to the knowledge based system.

Knowledge Representation Results

In order to interpret the segment knowledge passed to the expert system from the low-level vision system, it is important to get the best segments possible with the most meaningful attributes. For example, it is virtually impossible for an expert system to find water body segments if the segmentor has removed them because the threshold is set too high. On the other hand if the threshold is set too low then a large number of segments are produced, which defeats the purpose of this approach. It is therefore important to note all of the important variables when analyzing the segmentation problem.

A 256 pixel by 256 line by 7 channel Landsat Thematic Mapper image of the Ottawa area was selected as the test site. This image contains urban areas, agricultural areas, lakes and rivers and several conspicuous road networks. Channel 4 of this image is depicted in Figure 3.

Channels 4 and 5 of this image were segmented using all of the operators and some of the test results are in the following tables. Also channel 4 of this image was segmented with the Sobel operator and the effect of varying the segmentation threshold is also included in the following tables. The min, max, mean and standard deviation are for the gradient image.

The following table indicates the effect of varying some of the parameters of the low-level vision algorithms.

Figure 10 shows the segments derived from the PROLOG expert system for gradient and segmentation values for the sum of max differences given in Tables 1 and 2, row number 1.

Conclusions

The results of this research point out two major problems. The first problem is: which segmentation operator is the best for locating the segments of interest? The other problem is: once an operator is selected, what gradient and threshold values must be used in order to get the desired interpretation? Each of these issues will be discussed individually.

The selection of the best operator to use is dependent on the desired class. For instance the sum of maximum differences appears to give the best results when looking for any land/water boundaries. The Sobel operator seems to give the best overall general results. The Laplacian operator seems to give the best results for road edge detection. It is clear that more research is required into which operator gives the best results. But, it is also clear that no operator seems to be an obvious choice for all classes.

This leads one to believe that for remote sensing image analysis, a context dependent low-level vision system is required. Thus the expert system will first have to decide what class it is looking for and then select the operator that is best for locating that class. The expert system will then receive the results from the low-level operation and make its inferences based on this information. This approach also has the advantage that if the expert system selects and uses several operators for performing the interpretation, then it can assess the results based on several "views" of the scene. This should help improve the accuracy of the interpretation.

Once an operator has been selected, then the next question that must be answered is: what parameters should be varied in the gradient and segmentation operations? This again is class dependent. For instance land/water interfaces are best visualized on TM channel 4 or 5. Thus if the water class is desired, then only those channels need be processed. Certain vegetation is best determined from channels 2, 3, 4 and 5.

The selection of the segmentation threshold is again class dependent. Figure 6 shows the Sobel segments generated for a threshold that is near the mean of the gradient. These segments are not that significant. Better results are not obtained until the threshold is lowered as in Figures 7, 8 and 9.

Thus although this work indicates that more research is required into gradient, channel(s) and threshold selection; it is clear that such a contextual segmentation paradigm can produce meaningful image interpretations.

References

[Duda 71] R. O. Duda, and P. E. Hart; *"Pattern Classification and Scene Analysis"*; Wiley, New York; 1971.

[Gonzales 77] R. C. Gonzales, and P. Wintz; *"Digital Image Processing"*; Addison-Wesley Publishing Company, Inc.; Reading, Mass; 1977.

[Goodenough 87] D. G. Goodenough, M. Goldberg, G. Plunkett, J. Zelek; *"An Expert System for Remote Sensing"*; IEEE Transactions on Geoscience and Remote Sensing, Vol. GE-25, No. 3; pp 349-359; May 1987.

[Guindon 88] B. Guindon; *"Multi-Temporal Scene Analysis: A Tool to Aid in the Identification of Cartographically Significant Features on Satellite Imagery"*; submitted to: Canadian Journal of Remote Sensing, Canadian Aeronautics and Space Institute, Ottawa; 1988.

[Levine 85] M. D. Levine; *"Vision in Man and Machine"*; McGraw-Hill, New York; 1985.

[Pascucci 87] R. F. Pascucci; *"An Expert System For the Computer-Assisted Analysis of Radar Imagery"*; Proceedings of 1987 ASPRS Annual Convention, Baltimore; pp 355-363 March 29- April 3, 1987.

[Plunkett 86] G. W. Plunkett, D. G. Goodenough and M. Goldberg; *"Map/Image Congruency Evaluation Knowledge Based System"*; Graphics/Vision Interface 86; Vancouver, B.C.; pp 273-278; May 1986.

[Rao 88] A. R. Rao, and R. Jain; *"Knowledge Representation and Control in Computer Vision Systems"*; IEEE Expert, Vol. 3, No. 1; pp 64-79; Spring 1988.

[Sachs 82] L. Sachs; *"Applied Statistics"*; Springer-Verlag, New York; 1982.

[Stansfield 86] S. A. Stansfield; *"ANGY - A Rule Based System for Automatic Segmentation of Coronary Vessels from Digital Subtracted Angiograms"*; IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-8, No. 2; pp 188-199; Mar 1986.

[Straight 87] L. Straight, M. K. Menchel and A. Levine; *"Flexible Knowledge Representation for Application to Hybrid Image Processing/Artificial Intelligence Systems"*; Proceedings of 1987 ASPRS Annual Convention, Baltimore; pp 327-338; March 29- April 3, 1987.

Figure 1: A Non-contextual Segmentation Paradigm: The Expert System has no control over the iconic knowledge being passed from the low-level vision system.

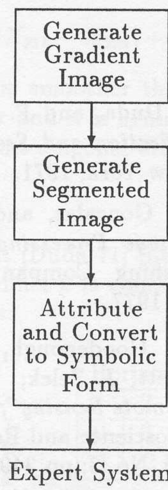

Figure 2: A Contextual Segmentation Paradigm: The Expert System controls all the steps of the low-level vision system and selects parameters for this processing based on the context of the feature of interest.

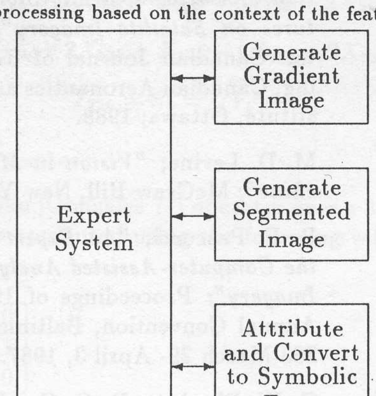

Figure 3: Image of channel 4 from Landsat Thematic Mapper. This test area is just west of Ottawa, Ontario.

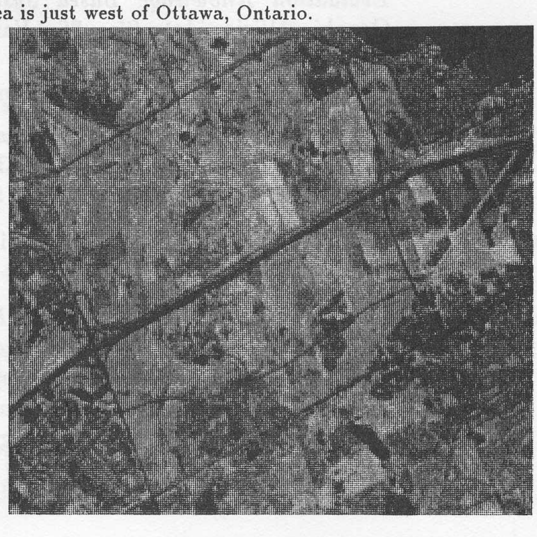

Figure 4: Edge image of Sum of Maximum Difference gradient generated segments with threshold set to 40 (Tables 1 and 2 entry number 1).

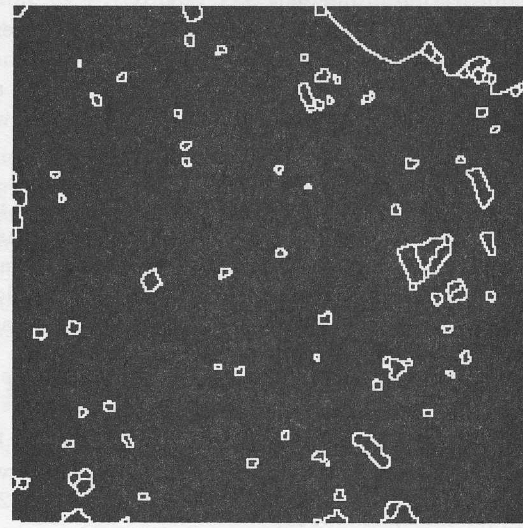

Figure 5: Edge image of Laplacian gradient generated segments with threshold set to 40 (Tables 1 and 2 entry number 7).

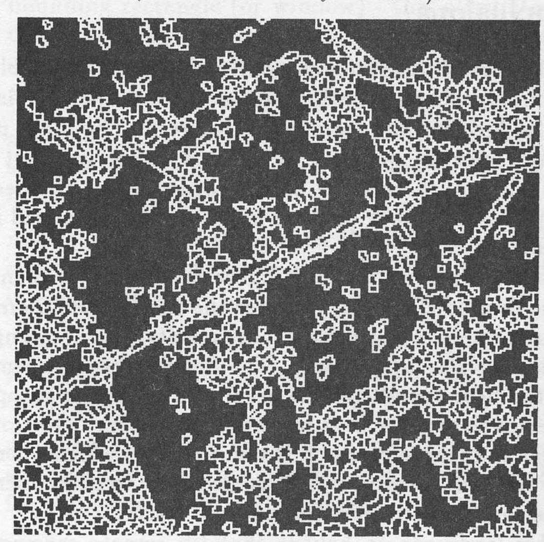

Figure 6: Edge image of Sobel gradient generated segments with threshold set to 60 (Tables 1 and 2 entry number 13).

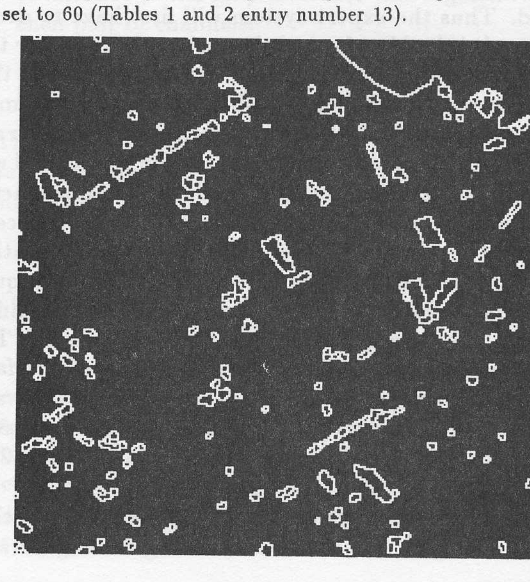

Table 1: Results of various gradient operators on the test image

No	Gradient Operator	TM Chan Numbers	Min Value	Max Value	Mean Value	Standard Deviation
1	Sum of Max Diff	4,5	2	786	86.54	46.57
2	Sum of Diff	4,5	2	1002	123.37	68.32
3	Max Sum of Diff	4,5	1	783	77.45	44.35
4	Roberts dist 1	4,5	0	309	34.51	23.52
5	Roberts dist 2	4,5	0	324	42.67	30.14
6	Sobel	4,5	0	825	114.46	79.77
7	Laplacian	4,5	-409	242	-0.0037	32.15
8	Variance	4,5	0	2348	87.89	112.91
9	Sobel	4	0	470	62.11	48.99
10	Sobel	4	0	470	62.11	48.99
11	Sobel	4	0	470	62.11	48.99
12	Sobel	4	0	470	62.11	48.99
13	Sobel	4	0	470	62.11	48.99
14	Sobel	4	0	470	62.11	48.99
15	Sobel	4	0	470	62.11	48.99
16	Sobel	4	0	470	62.11	48.99
17	Sobel	4	0	470	62.11	48.99

Table 2: Results of various segmentation thresholds based on Table 1 gradient images

No	Segment Threshold	Number of Segments	Comments
1	40	100	Figure 4
2	40	510	
3	40	96	
4	40	64	
5	40	188	
6	40	2087	
7	40	1653	Figure 5
8	40	435	
9	100	34	
10	90	57	
11	80	96	
12	70	171	
13	60	290	Figure 6
14	50	488	
15	40	927	Figure 7
16	30	1850	Figure 8
17	20	3114	Figure 9

Figure 7: Edge image of Sobel gradient generated segments with threshold set to 40 (Tables 1 and 2 entry number 15).

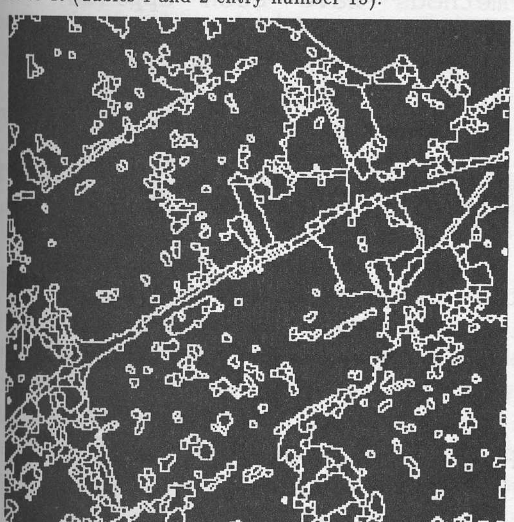

Figure 8: Edge image of Sobel gradient generated segments with threshold set to 30 (Tables 1 and 2 entry number 16).

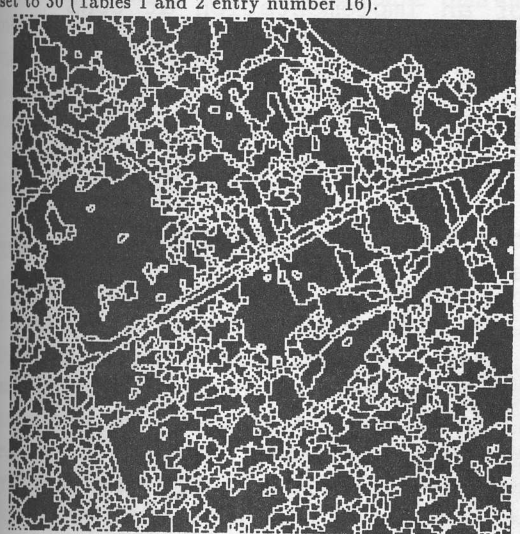

Figure 9: Edge image of Sobel gradient generated segments with threshold set to 20 (Tables 1 and 2 entry number 17).

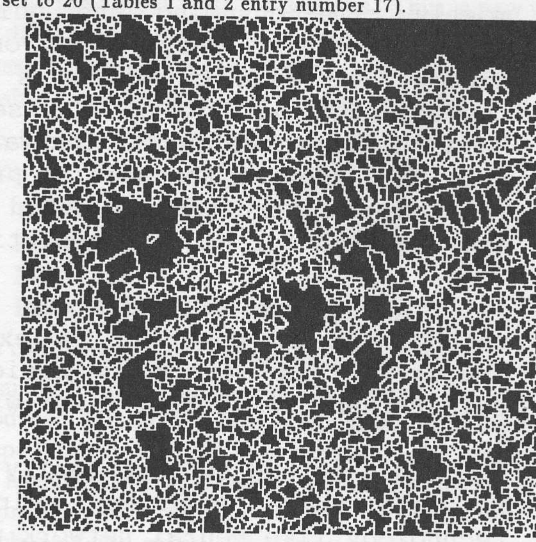

Figure 10: Coded image resulting from analyzing the segments of Figure 4 by a PROLOG expert system. The black areas indicate water bodies.