Department of Computer Science
University of British Columbia
Vancouver B.C., Canada

Interpretation-Directed Segmentation of Erts Images

DALE W. STARR
and ALAN K. MACKWORTH

Technical Report 76-3

BEST-PRINTER CO. LTD.
VANCOUVER, B.C.

In Proc. ACM/CIPS Pacific Regional Symp.

```
 MMM
 MMMM       MMM
   MM      M MM
   M      M
  M      M     MMMMMMMMM
 MM     MM    MMMM    MMM
MMM     MM      MM    MMM
MMM    MMM      MM   MMM
MMMMMMMMMM       MMMMMMM                MM
 MMMMMMM MMMM    MMM   MM             MMMMM
          MMM    MM    MMM           M    MM
                 M     MMM          M     M
                M  MM  MMM         MM
            MMMM   MMMMMM         MMM
            MMM                   MMM
                                  MMM
                                   MMM     M
                                    MMMMM
```

```
*********************************************************
*                                                       *
*   INTERPRETATION-DIRECTED SEGMENTATION OF ERTS IMAGES *
*                                                       *
*********************************************************
```

by

Dale W. Starr and Alan K. Mackworth

Technical Report 76-3

June 1976

Department of Computer Science
University of British Columbia
Vancouver, B. C.

INTERPRETATION DIRECTED SEGMENTATION OF ERTS IMAGES

D.W. Starr and A.K. Mackworth

Department of Computer Science
University of British Columbia

Abstract

Automatic interpretation of images from Earth Resources Technology Satellite 1 (ERTS-1) can be used in a variety of applications with considerable accuracy. Most systems however, classify strictly on a point by point basis, with no use of any spatial knowledge. Standard photo-interpretation techniques are combined with some techniques from Artificial Intelligence to produce an increase in accuracy over a point-by-point classification method. Traditional classification methods are used to obtain an initial segmentation of the image. Then, a controlled region merging process allows the regions with unambiguous interpretations to influence the interpretation of neighbouring regions, thereby introducing considerable context sensitivity into the interpretation process. Results are given of an experiment to interpret areas of different forest cover.

Key Words and Phrases

Multi-spectral scanner, ERTS, maximum likelihood function, Artificial Intelligence, scene analysis, region merging, resource inventory.

CR Category 3.63

Introduction

In 1972 ERTS-1 was launched by NASA as part of a program to demonstrate the feasibility and practicality of remote sensing from space. This satellite offers repetitive coverage of most of the earth's surface every eighteen days, thereby allowing easy detection of changes in a particular area of interest.

The basic component of the satellite is a multi-spectral scanner system with wavelengths in the .5 to 1.1 micrometer range, divided into four spectral bands. The scanner has a field of view of 185 kilometers which is transmitted to the earth bound receiving stations as a stream of six bit words, each word or picture element (pixel) covering an area of approximately 4600 square metres.

The objective of automatic interpretation is then to produce a useful partitioning of the scene presented, placing each pixel in one of a number of classes, depending on the application of the study. For example, in the periodic tracing of the growth of a city it may be sufficient to have only three categories for a gross classification, say residential, industrial-commercial and other, while a more in-depth study may delve into more specific sub-classes of these larger groupings.

Because of the low resolution, there are certain limits beyond which the interpretation techniques can not go. In 1971 a conference on Land Use Information and Classification (7) provided some classification objectives. The participants divided their classes into two levels, the first consisting of such broad groupings as urban areas, water, farming etc. Level two categories were sub-classes of those level 1 categories for which this would be appropriate. For example, the urban and farming classes would have many sub-classes which are important in various applications. The aim was to be able to achieve level 1 through the use of satellite imagery and level 2 through a combination of satellite and air-photo techniques. It seems that the aims were slightly conservative though, since success has been achieved in identifying most of the level 2 classes from satellite imagery alone, at much less cost than if aircraft had been used. This study attempts to categorize some second-level features in forest lands.

Past Attempts

ERTS classification systems can generally be divided into two basic classes: supervised and non-supervised systems. Non-supervised classification usually employs cluster analysis (or sometimes factor analysis) to automatically group a given data set into the most spectrally separable clusters using several wavelength bands. Different features of the earth's surface would ideally have their own unique spectral response, thereby creating distinct clusters of points belonging to each feature in the spectral space. Unfortunately this is seldom the case and the clusters usually overlap to some extent, thereby causing errors in the classifier.

The supervised approach requires some ground-truth data as a training set so that the various statistical features can be calculated from the multi-spectral data for each class. The means and the covariance matrix are calculated for each feature under consideration and then a maximum likelihood approach is taken. The classifier assigns a data point to that class which gives it the highest probability of membership.

Many studies have used these approaches with considerable success. Ellefsen et al. (3) and Todd and Baumgardner (12) carried out studies in urban land-use mapping from satellite data. Both studies, using cluster analysis, reported about 87% correctness, although Todd and Baumgardner improved upon this figure when areal information was used. This allowed them to clear up such areas of misclassification as differentiating between grassy rural areas and older residential neighborhoods.

Robertson (8) used an approach that differed from most other systems. He partitioned an image into blocks of image points such that each block ideally contained only points from a single class. He then classified the blocks as a whole, instead of individual points, using texture and other spatial characteristics. His increased accuracy over the point by point method was about 2.5% (with an average accuracy of 82%), at a cost of about ten times the computing time. He did tests on both airphotos and satellite imagery but for some unspecified reason his results from aircraft photography were comparable to those obtained from satellite imagery whereas in most cases, airphotos resulted in greater accuracy because of the much smaller area covered by each pixel.

Gupta et al. (6) also felt that point by point classification was less than optimal. They therefore used a boundary finder which produced regions which were homogeneous in nature. The images under consideration were those of agricultural fields: obviously well suited to the method used since one would expect the boundarys to be relatively straight. Once their closed areas are found the classification is done using a maximum likelihood classifier and a minimum distance classifier. Their results were extremely good, being above 95% correct for the simple point by point classifier. As a result of this high accuracy they could hardly show a significant improvement with the boundary method although some improvement was shown. One must bear in mind that they were using airphotos so the excellent accuracy is not all that unexpected.

Bajcsy and Tavakoli (1) carried out an interesting study to identify bridges, islands, rivers and lakes from satellite pictures. Their initial segmentation of the scene was fairly simple, just dividing it into water and non-water, although detecting bridges requires a sensitive test since they are narrower than the pixel size, but they do significantly darken a pixel in a watery area.

A world model was used to describe the objects of interest. Their program then went through the picture, successively refining the interpretation. It would find hypothetical bridges for example, and then test the hypothesis against the world model, rejecting any which did not fit. After all this, they were able to find all the bridges they hoped they would, plus some they did not expect to find because of their smaller width.

Region Merging

Brice and Fennema (2) described a method of scene analysis involving the use of regions and region merging to reduce the number of small regions in a picture into larger, more meaningful pieces. (A region is simply a connected set of pixels.) Initially they partitioned their image into regions of equal gray value, proceeding to merge regions which had common boundarys and conformed to some heuristics.

Feldman and Yakimovsky (4,13) have done much work on a semantic based region analyzer. While their aims were different, the idea seemed to be readily applicable to ERTS imagery, where areas of uncertain classification can be merged into bordering areas whose classification is more certain.

They employed operators such as shape, position in space and colour and proceeded to give tentative classifications depending on the values of these measurements. A region grower is then used to produce a final partitioning of the picture, using a sophisticated algorithm with much semantic information, which is too complex to outline here. Their results were very impressive; hopefully ERTS image interpretation can be taken to such elegant heights.

Tenenbaum and Weyl (11) employed techniques from both the above to analyze everyday scenes. Merges involving ambiguous regions are deferred until the ambiguity is hopefully removed as a result of other, more reliable regions.

Classification Method

Our aim was to show that integration of some scene analysis methods from Artificial Intelligence could be applied to the problem of classifying an ERTS photo with an increase in accuracy over a straight point by point method. We obtained a detailed ground-truth map of a forested area on Vancouver Island so this became the test area, with the aim of classifying regions of old growth, second growth, recent logging and water.

This ground-truth data provided the training areas for which the statistics of these four classes could be determined. This was necessary since a maximum likelihood classifier was employed for the initial classification.

Assuming the data conforms to a multi-normal distribution about the class mean, a point is assigned to a particular class on the basis of a multinormal probability function. There are four such functions, one for each class. (see Steiner,9) A point is assigned to that class whose function results in the highest value. This basic technique was augmented in the implementation however. If p1, p2, p3 and p4 are the probabilities received by a point from the four probability functions, the greatest of them, pmax, must be greater than some threshold percentage of the total. More precisely, if $pmax \geq k(p1+p2+p3+p4)/100$ (k is the threshold value), the point is placed in the class which produced pmax. If not, it is assigned to a class which depends on pmax and the second highest probability. These new classes (called ambiguous classes) are used when the classifier is not sure about the exact class in which to place a point. For example, if a point received its highest probability from class 1, but also received a fairly close one from class 2, it would be assigned to a class which contains all points which may belong to class 1 or 2, about which there is some doubt.

This method increases the number of classes from the basic four to ten; specifically type 1, type 2, type 3, type 4, type 1 or 2, type 1 or 3 etc. In practice though one need only include those combinations for which some statistical overlap occurs. In our case water was immediately separable from all other classes and it was not necessary to include a class which could be water or old growth for example, since this seldom occurs.

This procedure is then applied to all points in the area being studied, producing a preliminary classification map which is then given to the region finder. That program simply delineates the boundarys of contiguous regions which are to be passed to the region merger.

The merging program's aim is to merge the ambiguous regions into those strong regions bordering them which they most resemble and in this way reduce the number of regions and increase the accuracy of classification. It makes several passes through the list of regions, beginning with a fairly rigid criterion for merging and relaxing this in subsequent passes.

The first pass will only merge an ambiguous region with a strong one if

1) the two have a certain percentage of their border in

common,

2) the ambiguous region receives its highest probability from the same class as the strong region and

3) the average of the feature vectors is "close" to those of the strong region.

This procedure is repeated as long as there is merging done, at which point a second pass is begun.

This pass allows merging of a strong region with an ambiguous one as long as one of its two possible interpretations is the same as the strong one, their common boundary is a certain percentage of the whole and again, the average data vectors are close.

Finally, in the third pass any ambiguous regions remaining are simply assigned to the class corresponding to the higher probability. These regions could have been split into smaller regions and then attempt to merge again, but this would probably not produce any better results but rather serve only to produce more small regions which tend to clutter up a visual display of the final partition.

Results

When a simple point by point classification method was used, placing each point in one of the four classes (no ambiguous classes), a success rate of 70% was achieved. This would seem a bit lower than might be expected, probably as a result of considerable statistical overlap among the three types of forest cover. Nevertheless, the merging algorithm increased accuracy to 79%, which represents a 30% decrease in the error rate.

Another positive result was the simplification of the final picture over the initial partition. Typically about 150 regions were found when a point by point technique was used in a picture of 2500 pixels as seen in Fig. 1, placing each point in one of the four original groups. Fig.2 shows the results when the ambiguous regions are included, resulting in over 350 regions. The final partition after the merging process resulted in about 70 regions, shown in Fig.3.

The system is a number of programs written mostly in Algol W with a few parts in Fortran. The point by point classification for 2500 pixels required about 10 sec. CPU time on an IBM 370/168. The merging program on the other hand took about 35 sec. in total. (including the initial classification)

Conclusions

Even with the minimal semantic knowledge employed, a significant increase in accuracy over the point by point method was achieved by the region merger. One could only hope for a greater improvement in areas where there is more semantic knowledge available. For example, parks within a city would be classified, context-free, as agricultural areas in all likelihood, while the mere fact of their position precludes this possibility. Or a highway may become blended into surrounding vegetation as it becomes more narrow than the area covered by one pixel, but if the program knew that a road should be continuing in the area, such points could be correctly identified.

Of course, the amount of semantic knowledge one is able to include is heavily dependent on the picture domain being studied as well as the intended application, but one can only hope to inprove accuracy with such techniques.

Future LANDSAT satellites should provide even better performance since the resolution will be finer, with each pixel covering less area. As this occurs, there is less chance of a pixel hitting two distinct features, such as occurs when a road becomes narrower than the pixel size. Of course, with smaller pixels there will be much more data per unit area of the earth's surface and a corresponding increase in computing time. However it would not always be necessary to look at individual pixels but rather they could be grouped together in homogeneous areas, only looking at increased pixel resolution as accuracy demands.

Summary

Results are given of an experiment to combine region growing methods with standard maximum likelihood classification techniques in automatic interpretation of ERTS images. A 30% decrease in the error rate compared to that of a point by point classification technique was observed.

Acknowledgement

We would like to express our appreciation to Dr. Murtha of the Faculty of Forestry at UBC for providing ground-truth data which was so essential to this study.

Bibliography

1. Bajcsy, R., and Tavakoli, M. A computer recognition of bridges islands, rivers and lakes from satellite pictures. Machine Processing of Remotely Sensed Data, Purdue Univ. (1973), 2a54-2a68.

2. Brice, C.R., and Fenema, C.L. Scene analysis using regions. Artificial Intelligence 1, 3 (1970), 205-226.

3. Ellefsen, R., Swain P.H., and Wray, J.R. Urban land-use mapping by machine processing of ERTS-1 multispectral data: A San Francisco Bay area example. Machine Processing of Remotely Sensed Data, Purdue Univ. (1973), 2a7-2a22.

4. Feldman, J.A., and Yakimovsky, Y. Decision theory and Artificial Intelligence: I. A semantics-based region analyzer. Artificial Intelligence 5 (1974), 349-371.

5. Gower, J.F.R., and Daniel, I. The use of ERTS-1 computer compatible tapes with respect to marine research. Unpublished Manuscript. Marine Sciences Directorate, Pacific Region, Victoria B.C. (1974).

6. Gupta, J.N., Kettig, R.L., Landgrebe, D.A., and Wintz, P.A. Machine boundary finding and sample classification of remotely sensed agricultural data. Machine Processing of Remotely Sensed

Data, Purdue Univ. (1973), 4b25-4b35.

7. Hardy, E., and Anderson, J.R. A land use classification system for use with remote-sensor data. Machine Processing of Remotely Sensed Data, Purdue Univ. (1973), 2a1-2a6.

8. Robertson, T.V. Extraction and classification of objects in multispectral images. Machine Processing of Remotely Sensed Data, Purdue Univ. (1973), 3b27-3b34.

9. Steiner, R. A methodology for the automated photo-interpretation of rural land use types. Automatic Interpretation and Classification of Images, Grasseli, A. (Ed.) Academic Press, New York, 1969.

10. Strome, W.M. Canadian ERTS data handling system. Canadian Computer Conference Session 72. (1972) 423101-423118

11. Tenenbaum, J.M. and Weyl, S. A region-analysis subsystem for interactive scene analysis. Stanford Research Institute Technical Note 104 (1975).

12. Todd, W., Mausel, P., and Baumgardner, M.F. An analysis of Milwaukee County land use by machine-processing of ERTS data. Laboratory for Application of Remote Sensing (LARS) Information Note 022773, Purdue Univ. (1973).

13. Yakimovsky, Y., Scene analysis using a semantic base for region growing. Stanford Artificial Intelligence Laboratory Memo AIM-209, Stanford Univ. (1973).

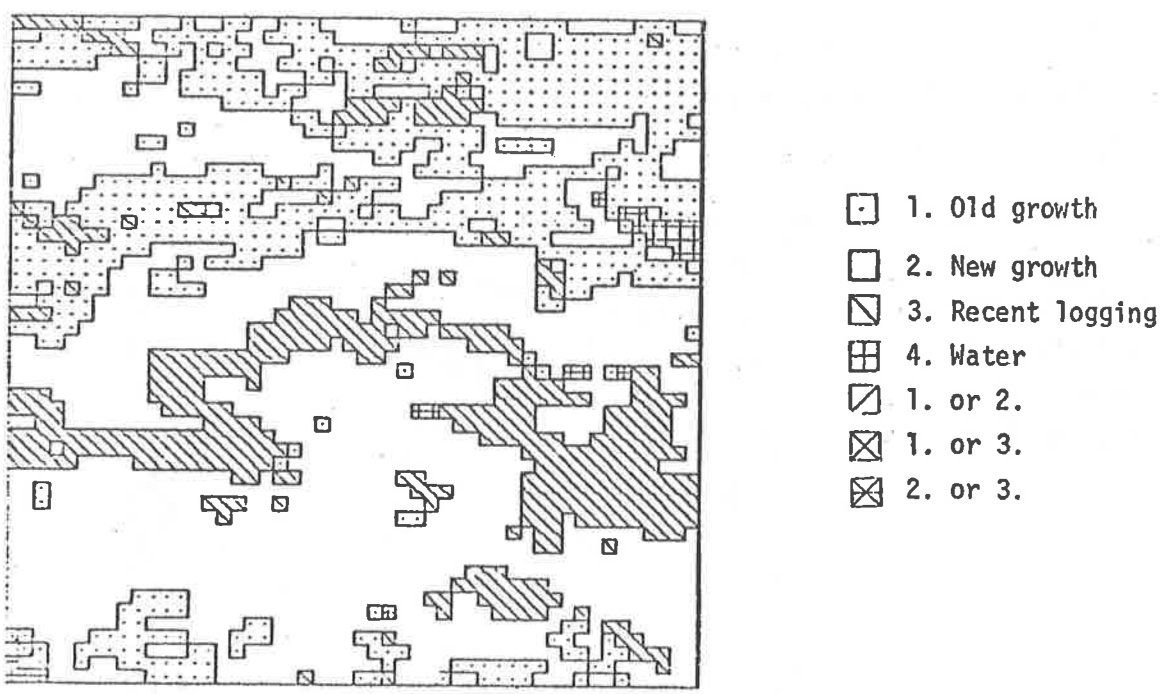

Fig. 1

Initial segmentation on a point by point basis with no ambiguous regions.

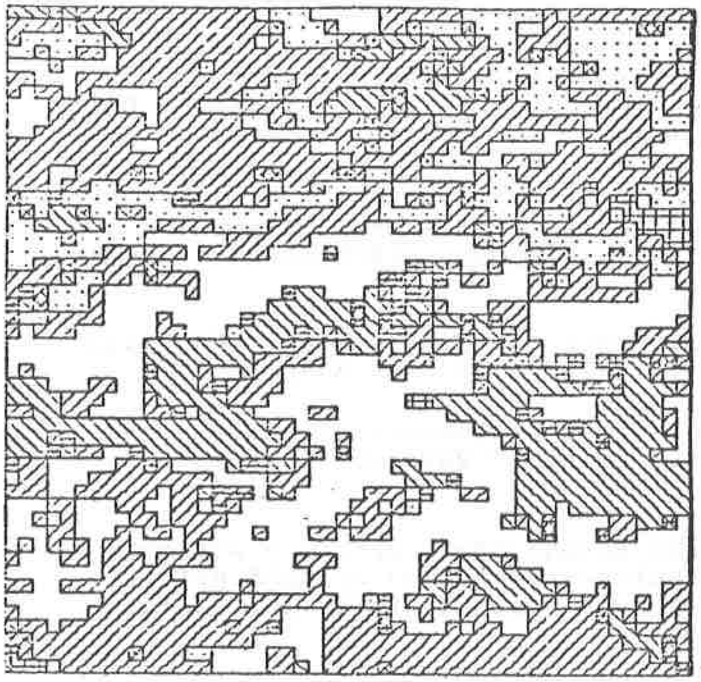

Fig. 2
Initial segmentation when ambiguous regions are included.

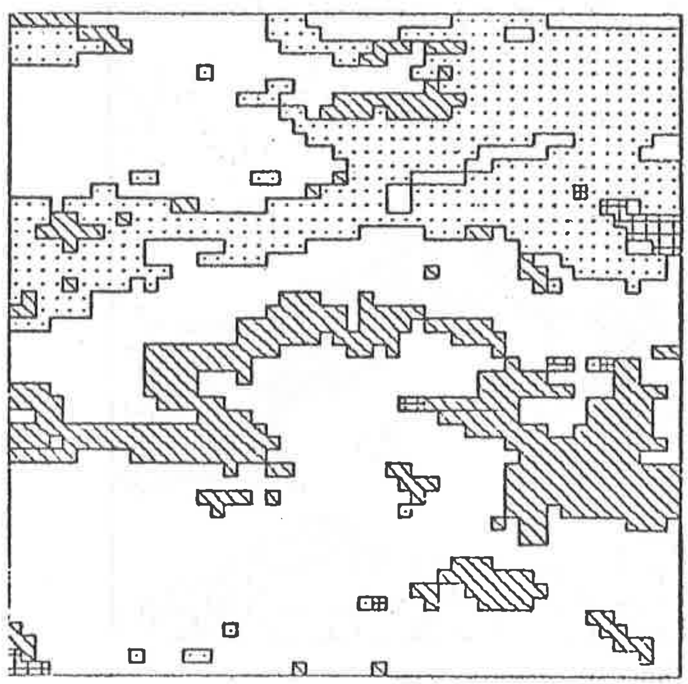

Fig. 3
Final picture after region merging is completed.