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

REGION-BASED MODELING ALGORITHMS FOR REMOTELY-SENSED DATA

M. GOLDBERG, M.L. IMHOFF

National Aeronautics and Space Administration/
Goddard Space Flight Center
Greenbelt, Maryland

E. DADDIO

Science Applications Research
Riverdale, Maryland

I. ABSTRACT

In many Earth science investigations, a "region-based" approach to spatial modeling is needed. This paper reports on algorithms being developed for performing region-based modeling operations on classified remotely-sensed images.

Most spatial modeling functions performed by geographic information systems on classified remotely-sensed images are "pixel-based". That is, the largest uniquely identifiable unit on which they operate is the pixel. In "region-based" modeling, however, contiguous geographic areas on the Earth's surface such as lakes, forests, and fields are treated as distinct spatial units. Thus, the operands in region-based models are not individual pixels; they are contiguous groups of pixels.

Region-based modeling has long been possible for vector-formatted data, such as digitized maps, where polygons (as well as lines and points) are uniquely labelled and topological information is often explicitly stored. But for remotely-sensed images, which are grid-formatted, region-based modeling functions are difficult to implement. The key to the development of the algorithms described in this paper was the concurrent development of "topological grid structure". Topological grid structure is a spatial data structure which maintains the simplicity and transportability of standard grid structure while providing the essential capability to treat contiguous areas as unique spatial entities.

Initially, five region-based algorithms are being developed. The first converts images from standard grid structure into topological grid structure. The four remaining functions act upon topologically grid-structured images to perform region-based relabelling, overlaying, distance searching, and neighborhood scanning operations.

II. INTRODUCTION

A "region-based" approach is needed for Earth science modeling involving classified remotely-sensed data.

Most spatial modeling operations performed by geographic information systems on classified remotely-sensed data are "pixel-based". That is, the largest uniquely identifiable unit on which they operate is the individual pixel. Spatial modeling operations can be grouped into four categories: relabelling (or "reclassification) functions, overlay functions, distance functions, and neighborhood functions (Tomlin and Berry, 1979, pp. 269-284). Let us consider the standard pixel-based versions of these functions. Relabelling functions create a new image by reassigning the classes of an existing image. A typical relabelling function might change all "deciduous forest" and "coniferous forest" pixels on a classified image into "forest" pixels. Overlay functions define a new image based on the interrelationship of two or more existing images. Thus, an overlay function might define a pixel to be of "high erosion potential" if, at the same location on a classified image and an image derived from ground observations, it found a "sandy soil" and a "steep slope" pixel, respectively. Distance functions define a new image as a function of the distance between pixels on an existing image. A distance function might be used to find the distance from every pixel on an image to the nearest "water" pixel. Neighborhood functions create a new image based on the aggregated characteristic of neighboring points within an existing image. A neighborhood function might label any pixel completely surrounded by "forest" pixels to also be a "forest" pixel. It is clear that the pixel-based versions of these four basic types of functions are useful, but they lack one quality that is becoming increasingly important to Earth scientists: the ability to directly manipulate identifiable features in images, not just individual pixels, as distinct entities.

In "region-based" modeling, contiguous geographic areas on the Earth's surface such as lakes, forests, and fields are treated as distinct spatial units. Thus, the operands in region-based models are not individual pixels; they are groups of contiguous pixels. Based on discussions with Earth scientists at the Goddard Space Flight Center, three basic applications of region-based modeling have been defined. These applications would be very awkward to implement using pixel-based methods. They are: (1) performing feature-by-feature analysis and change detection, such as measuring the decrease in area of a particular forest stand due to logging operations; (2) deriving information on the location of regions relative to neighboring regions, such as determining which barren fields border lakes (for soil run-off studies); and (3) enhancing classification, such as relabelling all waterbodies below a certain size as ponds and all waterbodies above that size as lakes.

III. THE DATA STRUCTURE

Region-based modeling has long been possible for vector-formatted data, where polygons (as well as lines and points) are uniquely labelled and topological information is often explicitly stored. However, for remotely-sensed data, which are grid-formatted, region-based functions are difficult to implement. The key to the development of the algorithms described in this paper was the concurrent development of "topological grid structure". The development of the concept of topological grid structure was motivated in part by the arguments in favor of the addition of topological information into cartographic data structures (Peuker and Chrisman, 1975, pp. 55-69).

Topological grid structure is a spatial data structure which maintains the simplicity and transportability of standard grid structure while providing the essential capability, previously limited to vector structures, to treat contiguous areas as unique spatial entities.

Standard grid structure is the most common format for representing maps and images in geographic information systems and image processing systems. In standard grid structure, a map or image is simply represented by a two-dimensional array of numbers. Each number corresponds to a uniform-sized retangular cell on the original map or image. The storage address for the number implicitly defines the geographic location of the cell, based on the known position of control points and the assumed geometry of the map or image. The value of the number corresponds to the attribute or "class" associated with the cell.

Despite its advantages, however, standard grid structure lacks one feature which is essential for region-based modeling: the identification of regions as distinct entities. For example, consider a standard grid-structured classified landcover image. Each cell of the image has been assigned a numerical value which tells whether it represents "water", "bare soil", "forested land", or other landcover type. But, the digital representation of the map contains no explicit information which identifies a given cell as being part of a particular waterbody, barren field, or forest. Topological grid structure contains this information.

A map or image represented in topological grid structure consists of two equal-sized standard grid-structured data planes and an ancillary file. Thus, any geographic information system or image processing system that can handle multiple band images can easily be adapted to topological grid structure. The first data plane is identical to the original map or image. Each cell in the second data plane contains a numeric identifier which uniquely defines the contiguous region to which its corresponding cell in the first data plane belongs. Contiguity is defined in terms of "4-connectivity", i.e., a cell with attribute X is contiguous with its North, South, East, and West neighbors which also have attribute X. The ancillary file contains attribute identifiers arranged in increasing numeric order. Within each attribute, contiguous area identifiers are arranged in order of decreasing size (measured by cell count).

IV. THE ALGORITHMS

This section contains descriptions of the five region-based algorithms that are under development. These functions will provide a means to carry out advanced Earth science modeling applications, such as those defined by scientists at the Goddard Space Flight Center. The first function, BIGTOP, converts images from standard grid structure into topological grid structure. The four remaining functions: TOPSIZE, VENN, FINDNEAR, and REGADJ act upon topologically grid-structured images to perform region-based versions of the standard pixel-based relabelling, overlay, distance, and neighborhood spatial modeling functions described in the introduction. These functions are being implemented in parallel to create an advanced geographic information system modeling capability as part of the Interactive Digital Image Manipulation System on a Hewlett-Packard 3000 computer and as part of the Land Analysis System on a Digital Equipment Corporation VAX 11/780 computer.

A. BIGTOP

BIGTOP takes an image in standard grid structure and creates a corresponding image in topological grid structure. BIGTOP computes contiguous area identifiers as follows. For each line of the input image, the BIGTOP function compares each pixel value with the one above it, to the left of it, and to the right of it. A connectivity table keeps track of similarities. After all lines have been compared, the function then reads the lines in from bottom to top and compares each pixel with the one below it. Whenever there is a match between pixels (whether scanning top-to-bottom or bottom-to-top) the test pixel is assigned the area identifier of the first pixel it matches. When there is no match, the next available unused area identifier is assigned. The ancillary file is created from the connectivity table. BIGTOP sorts the ancillary file and computes and inserts contiguous region sizes.

B. TOPSIZE

TOPSIZE is the region-based version of a relabelling function. It performs relabelling based on size. Specifically, it enables a user to relabel to the attribute values of contiguous areas that fall within a user-specified size range. The algorithm works as follows. The user specifies the current value, the size range, and the desired new value. ancillary file of the topologically grid-structured input image is then searched to find all the contiguous area identifiers which have the specified characteristics. When these have been found, TOPSIZE simply scans through the input image, makes the requested changes, and writes out the output image (also in topological grid structure). Note that no changes are made to the contiguous area identifiers.

C. VENN

VENN is the region-based version of an overlay function. The purpose of VENN is to compute intersections and unions between contiguous areas on two classified satellite images which are for the same geographic location, but which were acquired at different times. Further, VENN seeks to compute net area loss and net area gain for each contiguous area on the first input image. Unfortunately, it has been extremely difficult to devise rules for determining which areas on the second input image correspond to the areas on the first input image, even when assuming the complete equality and correctness of the classification in both cases.

D. FINDNEAR

FINDNEAR is the region-based version of a distance function. FINDNEAR identifies all regions of a given target class that are within a specified distance of regions of another given classification. The distance refers to the minimum distance between any points within the two types of region. FINDNEAR works as follows. A pixel-based Euclidean search is conducted for the specified distance around regions of the class that is being searched from. Then, a pixel-based intersection function finds all regions of the target class that intersect the search contour. The target class regions that are found are assigned an indicative new class value.

E. REGADJ

REGADJ is the region-based version of a neighborhood function. REGADJ scans all regions that are adjacent to a specified class of region and replaces the specified class value with a new class value based on some aggregated characteristic of the surrounding regions. REGADJ finds adjacent regions in the same way that FINDNEAR finds regions at a distance of one pixel.

V. CONCLUSIONS

Assuming that classifications of remotely-sensed satellite imagery are reasonably accurate representations of the Earth's surface at a given time, region-based algorithms for processing these images provide a practical new tool for spatial analysis and modeling using geographic information systems. While they do not provide an entirely new capability (due to the previous existence of contiguous region-finding algorithms), the use of pre-computed topological information (through the use of topological grid structure) makes region-based algorithms highly accessible to Earth scientists.

REFERENCES

Peuker, T.K., and N. Chrisman. "Cartographic data structures." The American Cartographer. 2,1 (April 1975), pp. 55-69.

Tomlin, C.D., and J.K. Berry. "A mathematical structure for cartographic modeling in environmental analysis." Proceedings of the American Congress on Surveying and Mapping, 39th Annual Meeting. (March 1979), pp. 269-284.

AUTHOR BIOGRAPHICAL DATA

Michael Goldberg received a B.S. degree in Mathematics from the University of Maryland in 1973 and an M.S. degree in Computer Science from the Johns Hopkins University in 1983. Since 1978, he has worked as a mathematician in NASA/Goddard Space Flight Center's Information Extraction Division, where he specializes in geographic information system development and spatial data structure research.

Marc Lee Imhoff received a B.S. degree in Physical Geography in 1978 and an M.S. degree in Agronomy in 1980, both from the Pennsylvania State University. At NASA/Goddard Space Flight Center since 1980, he has served as a program manager of satellite Earth resources applications projects with the Eastern Regional Remote Sensing Applications Center and performed image geometric quality assessment for the LANDSAT 4 multispectral scanner. Mr. Imhoff presently works in the Information Extraction Division, where he specializes in remote sensing research and environmental modeling. He is currently an investigator and science team member on the Shuttle Imaging Radar-B experiment.

Ernest Daddio received a B.A. degree in Meteorology and Oceanography from New York University in 1972 and an M.S. degree in Physical Oceanography from the Louisiana State University in 1976. He has performed oceanographic research and has been involved in numerical hydrodynamic modeling. Since 1982, he has been employed by System and Applied Sciences Corporation and its affiliate, Science Applications Research. His current interests are in algorithm development and computer applications in geographic information systems, time series analysis, and spatial data analysis.