

**THEMATIC MAPPING FROM IMAGERY:  
AN ASPECT OF AUTOMATED MAP GENERALIZATION**

Minhua Wang<sup>1</sup>, Peng Gong<sup>2</sup> and Philip J. Howarth<sup>1</sup>  
Earth-Observations Laboratory,  
Institute for Space and Terrestrial Science  
at

<sup>1</sup>Department of Geography, University of Waterloo  
Waterloo, Ontario, Canada N2L 3G1 and

<sup>2</sup>4850 Keele Street, North York, Ontario, Canada M3J 3K1

**ABSTRACT**

Map generalization, as a means of portraying the complex real world, has to date been confined to the manipulation of map data. With the advent of new data acquisition techniques, particularly remote sensing, data sources for spatial analysis have greatly increased. However, map generalization from image data is a challenging problem. In this paper, the conceptual and technical problems in generalizing cartographic objects from remote sensing imagery are addressed. A two-stage generalization framework is proposed for thematic mapping from imagery. Specific interest is focused on mapping land use from SPOT satellite imagery.

**INTRODUCTION**

Thematic mapping is a major activity in both cartography and remote sensing. However, due to independent developments in remote sensing and cartography, the theoretical bases for thematic mapping are considerably different. In cartography, thematic mapping is considered as a process of generalization in which the spatial context and attributes of objects from a source map are transformed into a target map. This is done according to a scale change through generalization operators such as selection, simplification, symbolization and classification (Robinson 1984; Shea and McMaster 1989). In remote sensing, however, thematic mapping is considered to be a process of pattern recognition in which the spectral responses of pixels are grouped into a number of defined classes using statistical modeling techniques. The process is also called image classification or pixel allocation (Burrough 1986). A problem with this process is that because of the complexity of the real world, spectral responses for a high resolution image show great spatial variability (Woodcock and Strahler 1988). From such heterogeneous data, it is difficult to directly generate homogeneous polygons, such as those presented in conventional maps. It is suggested that the concept of map generalization can be introduced to solve this problem.

It is well known that map generalization involves transformations in both the spatial domain and the thematic domain (McMaster 1989). In the spatial domain, map generalization refers to the transformation of points, line and polygons; in the thematic domain, it refers to attribute transformation. Traditional numerical

generalization has focused on spatial transformation, specifically line generalization (Douglas and Peucker 1973; Shea and McMaster 1989; Muller 1990). This primarily involves a scale change.

In thematic mapping from remote sensing images, the key issue is the change in thematic representation. This may not involve procedures used in traditional map generalization, e.g., simplification. However, it could include other functions associated with raster representation, e.g., feature selection and feature smoothing (Monmonier 1983). At the same time, the process may not involve a scale change. We wish to argue that the classes obtained from image classification may not correspond to certain cartographic objects, as their spatial appearances are usually heterogeneous and their class membership may be uncertain (Robinson and Frank 1985). In this paper, we use the term 'entity' to describe the classes obtained from image classification. Map generalization, therefore, is concerned with the transformation of entities to cartographic objects.

In this paper, the aim is to extend the traditional map generalization concept into land use mapping from remote sensing imagery. To do this, a new procedure for mapping land use from satellite imagery has been devised. In this procedure it is assumed that land uses are highly generalized objects. As a result, they cannot be generalized directly from remote sensing imagery. The procedure has to be undertaken in two steps: image-to-entity generalization and entity-to-land-use generalization.

## **A CONCEPTUAL FRAMEWORK FOR MAPPING FROM IMAGERY**

Map generalization is a complex mental process involving perception, cognition and other intellectual functions. "It focuses on the extraction of the general, crucial elements of reality" (Brassel and Weibel 1988, p.230). It is usually related to the functions of selection, simplification, emphasis, classification, etc., by which observed reality is structured into a number of individual entities; then important entities are selected and represented on the map. Brassel and Weibel (1988) proposed a five-step conceptual framework for map generalization:

- *Structure Recognition* aims at the identification of objects, understanding their spatial relations and the establishment of measures of relative importance. It is the basic understanding of the essential structures of the spatial information available in the original database.
- *Process Recognition* is to establish the relationships between source objects and generalized objects (e.g., linguistic relations, spatial relations and statistical relations), based on the structure of the original database and the control parameters (e.g., objective).
- *Process Modeling* can be considered as a compilation of rules and procedures derived from a process library and the pre-setting of the process parameters that were established during process recognition.
- *Process Execution* and *Data Display* are operational procedures (e.g., classification, simplification and symbolization) which convert database and information structures into the target and generalized databases. These

procedures have been addressed in many existing approaches (McMaster 1989; Steward 1974).

Mapping from imagery is a generalization, which represents a process of transformation from the digital (spectral) domain to thematic and spatial domains. This process, however, is different from conventional cartographic generalization because the spatial units for generalization are not cartographic objects. In other words, they are not points, lines and polygons, but rather they are pixels which do not have any thematic meaning.

A two-stage procedure for mapping from imagery can be structured: statistical generalization and cartographic generalization. In statistical generalization, the original imagery is divided into a number of entities derived under statistical control. This represents processes of data reduction and transformation. The result is not a map but an entity image which shows basic spatial structures and thematic components of the remote sensing imagery. In cartographic generalization, the object is highly generalized, selective and subjective. Relationships between information entities and cartographic objects are modeled to produce a smooth, uniform map. A conceptual framework for mapping from remote sensing imagery is shown in Figure 1.

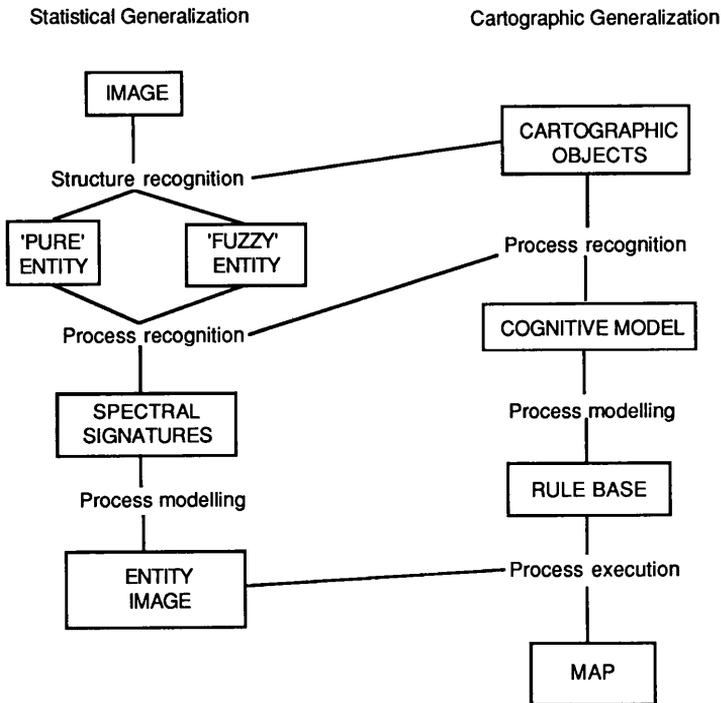


Figure 1 A conceptual framework for mapping from imagery.

In statistical generalization, the imagery should first be examined according to the mapping objectives. From this examination, a list of entities should be prepared. There are two types of entity: 'pure' entity and 'fuzzy' entity. A 'pure' entity is one which is homogeneous and can be clearly identified on the image, while a 'fuzzy' entity consists of mixtures of 'pure' entities and is ambiguous when observed on the image. The original image can be transformed into an entity image containing the two types of entity. These entities constitute the basic information for a cognitive model in cartographic generalization. In the process modeling stage, sampling procedures can be employed to extract spectral signatures in order to link the spectral values of the image with entities. In the process execution stage, a classifier can be used to assign each image pixel to an entity label, based on the results of the sampling.

In cartographic generalization, the relationships between entities and cartographic objects (such as logical relation, spatial relation and statistical relation) need to be identified in order to develop a cognitive model. Based on this model, a rule base for the generalization can be designed. The rule base may be constructed as a simple logical operation or as a more sophisticated expert system, depending on the complexity of the mapping task. In the final stage, process execution, a map is generalized from the input entity image.

## **A CASE STUDY**

Based on the conceptual framework presented above, a case study of land-use mapping from SPOT imagery was carried out. The land-use mapping procedure was divided into two steps: entity extraction and land-use map generation.

### **Study Area and Data Description**

The test site selected for the study is part of the city of Scarborough, one of the fastest-growing municipalities in Metropolitan Toronto, Canada. The study area is dominated by residential areas at different stages of development; industrial and commercial land uses are also prevalent. The image used for study was a subscene (256 x 256 pixels) from a multispectral SPOT image with 20 m x 20 m spatial resolution pixels. It was acquired on June 4, 1987 (Figure 2).

### **Entities and Land Uses Identifiable from the Image**

Based on the generalization concept, two types of entity in the image, the 'pure' entity and the 'fuzzy' entity, were identified. 'Pure' entities have a distinct spectral appearance on the image and have relatively narrow spectral distributions (i.e., the digital values have relatively low standard deviations). A 'fuzzy' entity is not defined precisely, but it has a relatively wide range of spectral values. In Table 1, eight entities which were recognized on the image are listed. Asphalt surface, concrete surface, bare surface, and the two types of trees are 'pure' entities; the other three are 'fuzzy' entities.

The objective of the case study was to map land use from the image. Six land uses were categorized (Table 2). As can be seen, each land use is composed of several land-cover entities (Campbell 1983). It should also be noted that the relationships between land-use types and entities are quite complicated and that many land uses have similar entity components.



—| = 600 Metres

Figure 2 A SPOT image of northern Scarborough, Ontario, Canada recorded on June 4, 1987.

Modeling Generalization from Entity Image to Land-Use Map

Modeling the generalization from entity image to land-use map involves establishing relationships between entities and cartographic

Table 1. Entities and Their Descriptions

Code	Entity	Description	Gray Level
1	Asphalt surface	roads, house roof, parking lot	1
2	Concrete surface	building, warehouse, parking lot	2
3	Bare surface	land cleared for construction	3
4	Soil surface	wasteland, non-cultivated surface	4
5	Deciduous trees	deciduous trees	5
6	Coniferous trees	coniferous trees	6
7	Low-density grass	grassland, lawn	7
8	High-density grass	grassland, lawn	8

Table 2. Land Uses and Entities Contributing to Them

Code	Land Use	Composition*	Gray Level
A	Old residential	1, 2, 4, 5, 6, 7, 8	1
B	New residential	1, 2, 4, 7, 8	2
C	Industrial/Commercial	1, 2, 4, 5, 6, 7, 8	3
D	Land under construction	1, 2, 3, 4, 7, 8	4
E	Open space	4, 7, 8	5
F	Woodland	5, 6	6

\*Numbers represent codes in Table 1

objects (i.e., building a cognitive model). However, the relationships between entities and land uses are complex; there is no one-to-one correspondence between them. It is, therefore, important to select the key factors or parameters to model the generalization. A simple arithmetic aggregation of entities into a specific land-use type, such as land use A = entity 1 + entity 3 + entity 5, is insufficient to accomplish the generalization process from entities to land use. Fortunately, the proportional distributions of different entities vary from one land use to another. Therefore, it is possible to model the generalization process from entity-image to land-use map using the spatial frequency of each entity as a parameter (Gong 1990). For example, high-density grass (entity 8) has contributions to both residential (land use A and land use B) and open space (land use E), but the spatial frequency of high-density grass in open space is much higher than in residential areas. Based on the different spatial frequencies and entity compositions, the various land uses can be distinguished.

#### Entity-Image Generation

The procedure used to produce the entity image was a supervised maximum-likelihood classification. First, spectral signatures of the eight entities were obtained using a supervised training algorithm. The entire image was then classified according to these spectral signatures using a maximum-likelihood classifier. Figure 3 shows the entity image obtained using this method; gray levels are listed in Table 1.

#### Generalization from Entity Image to Land-Use Map

Two procedures were employed to carry out the generalization from entity image to land-use map. The first procedure, based on differentiating entity frequencies, was used to derive old residential (land use A), new residential (land use B), industrial and commercial (land use C) and land under construction (land use D). The second procedure, arithmetic aggregation, was used to extract open space (land use E) and woodlot (land use F).

In the entity-frequency-based procedure, a pixel window (9 x 9) was first moved over the entity image to extract an entity-frequency



— = 600 Metres

 Asphalt surf	 Decid. trees
 Concrete surf.	 Conif. trees
 Bare surface	 L-dens. grass
 Soil surface	 H-dens. grass

Figure 3 An entity image generalized from the SPOT image.

vector  $F(i,j) = \{f_1(i,j), f_2(i,j), \dots, f_8(i,j)\}^T$ .  $F(i,j)$  was associated with the center pixel of each pixel window at row  $i$  column  $j$  on the image.  $0 \leq f_k(i,j) \leq 81$  ( $k = 1, 2, \dots, 8$ ) denotes the occurrence frequency of entity  $k$  in a pixel window. To determine whether pixel  $(i,j)$  belonged to one of the land-use types A - D, a city-block distance measure was used:

$$d_m(i, j) = \sum_{k=1}^8 |f_k(i, j) - c_{mk}|$$

where  $d_m(i,j)$  is the distance from the entity frequencies at pixel  $(i,j)$  to the average entity frequency  $c_m = \{c_{m1}, c_{m2}, \dots, c_{m8}\}^T$  for land use  $m$  ( $m = A, B, C, D$ );  $c_m$  was obtained from supervised training on the entity image.

Once  $d_m(i,j)$  was obtained, it was compared with a threshold  $\beta$  ( $0 < \beta < 81$ ). If  $d_m(i,j) < \beta$ , pixel  $(i,j)$  was a candidate for land use  $m$ . Otherwise, pixel  $(i,j)$  was rejected from land use  $m$ . If more than one land use was a candidate, pixel  $(i,j)$  belonged to the land-use type for which the distance was the shortest. A detailed description of entity-frequency extraction and land-use identification based on entity

frequencies can be found in Wharton (1982), Zhang *et al.* (1988) and Gong (1990).

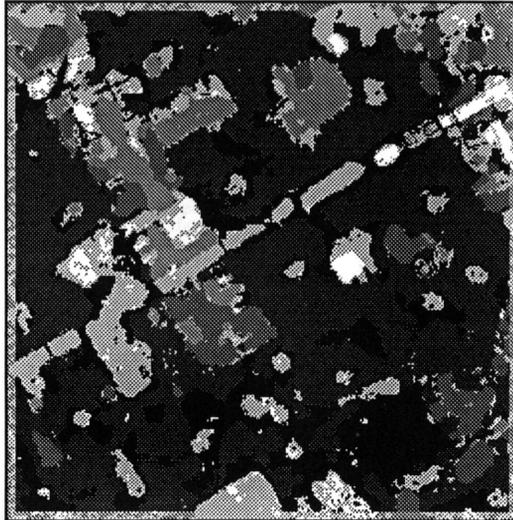
In arithmetic aggregation, two aggregation rules were used:

land use E = entity 4 + entity 7 + entity 8

land use F = entity 5 + entity 6.

However, entities 4, 5, 6, 7, and 8 are also components of land uses A - D. Therefore, a conflict will arise when an entity label at pixel  $(i,j)$  belongs to one of the land uses A - D and one of the land uses E - F. Under such circumstance, pixel  $(i,j)$  was assigned to one of the land uses A - D. This is reasonable because in the first procedure both entity information from pixel  $(i,j)$  and neighborhood entity information from a pixel window were included in the identification.

After the two procedures, a number of pixels remained unlabeled. These unlabeled pixels were relabeled using the entity-frequency-based method, but without thresholding the distances  $d_m(i,j)$ .



—| = 600 Metres

- |              |               |
|--------------|---------------|
| ■ Old resid. | ■ Land-const. |
| ■ New resid. | ■ Open space  |
| ■ Ind./com.  | □ Woodland    |

Figure 4 A land-use map generalized from the entity image.

### Results and Discussion

The map in Figure 4 is a result from this preliminary study. It shows homogeneity in polygon distribution, which is consistent with a conventional cartographic product. By visual comparison, most land uses on the map have corresponding locations on the image. The

results show the potential for mapping from remote sensing imagery using the generalization concept. However, there are still some problems to be overcome:

- The old residential land use is confused with open space, because there is high spatial frequency of grass cover in the old residential area.
- There are still some "salt-and-pepper" patterns on the map; smoothing is required.
- The selection of land uses is restricted. Some land uses, which can be identified by visual interpretation (e.g., recreational land use), are not generalized.

These problems result from modeling during the generalization. In modeling, only component factors were considered, while in human perception, spatial features such as shape, size, linearity and spatial adjacency are also important. It may not be possible to represent these factors by a statistical model; a fuzzy model or a logical model may be more appropriate. It is apparent, however, that more sophisticated models are required in the entity-to-map generalization procedure, such as an expert system. In future studies of entity-to-map generalization, a fuzzy-set-theory approach and a knowledge-based approach should also be considered.

## CONCLUSIONS

It is concluded that thematic mapping from remote sensing data is a challenging issue in numerical map generalization. It involves a process of thematic information extraction and entity-to-cartographic-object generalization, during which a scale change may not be involved. Research on this topic is limited. In this paper, we proposed a conceptual framework for mapping from imagery, based on the generalization concept. The aim of this approach was to extend the conventional cartographic generalization concept to remote sensing data, and to rethink the theoretical foundations for mapping from remote sensing. A case study of land-use mapping was undertaken to verify the theoretical model. A homogeneous land-use map was presented as a preliminary result for this procedure. Although the result is still not as good as that in human perception, it demonstrates the potential of the new methodology for mapping from remote sensing imagery. Further work involving more sophisticated models is justified.

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge the assistance of SPOT Image Corporation of France and the Canada Centre for Remote Sensing in supplying the SPOT data used in this study. This research is funded by a Centre of Excellence grant from the Province of Ontario to the Institute for Space and Terrestrial Science, and NSERC Operating Grant A0766 awarded to Dr. P. J. Howarth. Thanks to Denis Gratton for assistance with preparing the figures.

## REFERENCES

- Brassel, K. E. and R. Weibel 1988, A review and conceptual framework of automated map generalization, International Journal of Geographic Information Systems, Vol. 2, No. 3, pp. 229-244.
- Burrough, P. A. 1986, Principles of Geographical Information Systems for Land Resources Assessment, Clarendon Press, Oxford.
- Campbell, J. B. 1983, Mapping The Land -- Aerial Imagery for Land Use Information, Resource Publications in Geography, The Association of American Geographers, Washington, D. C.
- Douglas, D. H. and T. K. Peucker 1973, Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, The Canadian Cartographer, Vol. 10, No. 2, pp. 112-123.
- Gong, P. 1990, Improving Accuracies in Land-Use Classification with High Spatial Resolution Satellite Data: a Contextual Classification Approach, Unpubl. Ph.D Thesis, Department of Geography, University of Waterloo.
- McMaster, R. B. 1989, Introduction to 'Numerical Generalization in Cartography', Cartographica, Vol. 26, No. 1, pp. 1-6.
- Monmonier, M. S. 1983, Raster-mode area generalization for land use and land cover maps, Cartographica, Vol. 20, No. 4, pp. 65-91.
- Muller, J. C. 1990, The removal of spatial conflicts in line generalization, Cartography and Geographic Information Systems, Vol. 17, No. 2, pp. 141-149.
- Robinson, A. H. 1984, Elements of Cartography, John Wiley and Sons, Inc., New York.
- Robinson, V. B. and A. U. Frank 1985, About different kinds of uncertainty in collections of spatial data, AutoCarto 5, Washington, D.C. pp 440-449.
- Shea, K. S. and R. B. McMaster 1989, Cartographic generalization in a digital environment: when and how to generalize, AutoCarto 9, Baltimore, MD, pp. 56-67.
- Steward, H. J. 1974, Cartographic generalization: some concepts and explanation, Cartographica Monograph, No. 10.
- Wharton, S. W. 1982, A context-based land-use classification algorithm for high-resolution remotely sensed data, Journal of Applied Photographic Engineering, Vol. 8, No. 1, pp. 46-50.
- Woodcock, C. E. and A. H. Strahler 1987, The factor of scale in remote sensing, Remote Sensing of Environment, Vol. 21, pp. 311-332.
- Zhang, Z., H. Shimoda, K. Fukue, and T. Sakata 1988, A new spatial classification algorithm for high ground resolution images, Proceedings of IGARSS'88, Edinburgh, Scotland, pp. 509-512.