Integrating Topographic Data with Remote Sensing for Land-Cover Classification

Lucas L. F. Janssen, Marijke N. Jaarsma, and *Erik T. M. van der Linden*
The Winand Staring Centre for Integrated Land, Soil and Water Research, P. O. Box 125, 6700 AC, Wageningen, The Netherlands

ABSTRACT: Accurate land-cover data are required for environmental studies on a regional scale. The use of spectral classification of remote sensing images alone is insufficient to meet these needs. Therefore, topographical data from a Geographical Information System (GIS) was added to improve the classification accuracy. A digital topographical map was modified to serve as input for an object classification. An object has been defined as an area where only one land-cover type is expected. The results of a per pixel classification were used per object to determine the land-cover type of that object. This result was fed back to the GIS. The object classification improved the overall accuracy of two agricultural regions in The Netherlands by 12 percent and 20 percent.

INTRODUCTION

LAND-COVER MAPS are used for many purposes. In land consolidation projects and in environmental and hydrological studies, accurate, up to date information about land cover on a regional scale is often required (e.g., Thunnissen *et al.*, 1990). Knowledge of changes in land cover is becoming increasingly important from both the ecological and economical point of view. In the Netherlands, although there is actual statistical data about land cover, this information is not available in a geographical form. In the near future, supernational organizations such as the European Community will be involved in producing accurate geographical data. A research study has commenced to test the application of space remote sensing technologies for generating improved agricultural statistics for incorporation within current agricultural information systems (MARS, 1989). Moreover, a project dealing with an inventory of land cover in all the member states of the European Community by means of satellite remote sensing (Heymann, 1987) is already underway.

There is an increase in the use of remote sensing as a technique for collecting various types of information. Land-cover information can be obtained by classifying air- and spaceborne remote sensing images. Typically this is performed by the spectral analysis of individual pixels. The results of per pixel classification depend largely on the type of area, land-cover type, and the image acquisition date. Previous studies of land-cover classification of satellite data on a regional scale have shown that accuracies of 50 to 90 percent could be achieved (e.g., Hill and Megier, 1988; Shimoda *et al.*, 1988). However, for specific applications the overall accuracy ought to be 80 percent or higher.

Classification results are affected by spectral confusion of land-cover types and mixed pixels (Ioka and Koda, 1986). Mixed pixels are present at the boundary of two or more classes, and their spectral reflectance is the mixture of different characteristic reflectances. Classifications based solely on spectral observations are often not sufficiently accurate for regional studies. The solution would be to extend the classification procedure using data and/or knowledge. These methods can be subdivided according to their function in the classification process: pre- and post-classification techniques and classifier operations (Hutchinson, 1984). In the development of GIS, the number of classification procedures using geographical data was increased. The typical information used includes digital thematic maps, elevation models, and topographical maps. The advantage of the integrated use of geographical data and remote sensing data is becoming relatively commonplace (Catlow *et al.*, 1984; Van der Laan, 1988; Kenk *et al.*, 1988).

The first objective of this study was to derive a more accurate land-cover classification using geographical data from a GIS. The second objective was to enable feed back of remote sensing derived information to a GIS. Both objectives were met by means of an object classification. The object classification was tested for two different agricultural regions in The Netherlands.

MATERIALS AND METHODS

GENERAL APPROACH

In the object classification geographical information is added to improve classification accuracy. The misclassifications caused by the mixed pixels, and to a lesser extent by spectral confusion, are corrected by providing spatial context. In this case the spatial context is the geometry of objects, the definition of an object being an area in which only one land-cover type is expected. The two assumptions for this approach were that

- the object boundaries are stored in a GIS and
- the majority of the pixels within an object have been correctly classified in a per pixel classification.

Based on these assumptions, the incorrectly classified pixels within an object could be corrected by

- performing a per pixel classification; and, for all given objects,
- determining the pixels that are within an object;
- determining the label with the largest frequency by means of a frequency table;
- assigning this label to the object and also to all the pixels that are within the given object.

If integrated vector and raster processing were possible, the object classification could be executed as described above. Because an integrated approach was not possible using the systems available, the objects were gridded using the unique object identifier as grid item. All raster elements with the same number belong to the same object: per pixel object identification.

The program OBJCLASS was developed to effect the object classification. This program creates a frequency table using two raster files: the output of the per pixel classification and the per pixel object identification (Figure 1). A frequency table was established to determine the label of each object. The output were a raster file and an ASCII file with the statistics (label and frequency) for each object. The latter was used to enable feedback to the GIS. An example of a frequency table is given in Table 1. Objects 1012, 1013, and 1014 are given the label with the highest frequency: class 2, 1, and 6, respectively. In the output raster file, all pixels within object 1014 are given label 6. In fact, the frequency table offers even more information: if classes 1 and

PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING,
Vol. 56, No. 11, November 1990, pp. 1503–1506.

0099-1112/90/5611–1503$03.00/0

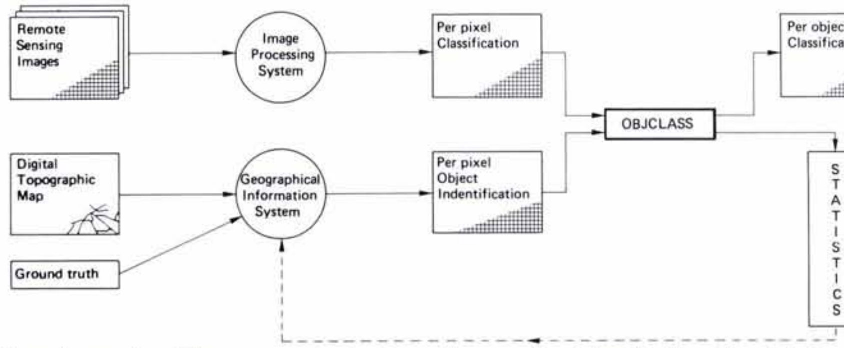

FIG. 1. Flow chart of the followed procedure. The program OBJCLASS enables an object classification. The statistics created in this program serve as feedback to the GIS.

TABLE 1. THE FREQUENCY TABLE FOR OBJECT NUMBERS 1012-1014. FOR EACH OBJECT, THE RELATIVE FREQUENCY OF LABELS 1 TO 6 IS LISTED

	Frequency (%)					
Object number	1	2	3	4	5	6
1012	36	41	14	0	9	0
1013	85	1	0	14	0	0
1014	0	2	0	2	0	96

2 are spectrally easy to discriminate, there is strong evidence that object 1012 is not a single object according to the definition and that the geometrical data should be updated. (The updating is not included in this article.)

The unique object number enabled feedback to the GIS. The statistics file (Figure 1) was used to add the new found label to the database of the GIS. Most geographical information systems offer specific analysis possibilities which are not operational on image processing systems. Another advantage is that the more extensive possibilities of a GIS is useful for graphical output.

TEST REGIONS

The object classification was tested on two regions: Ulvenhout and Biddinghuizen (Figure 2). Both areas represent a typical agricultural region in The Netherlands. Ulvenhout is a small-scale region. The agricultural fields are small and irregular. Furthermore, patches of forest and small townships are scattered throughout the region. Biddinghuizen is a modern agricultural area in Oostelijk Flevoland, one of the polders in the former Lake IJssel. The agricultural fields are large and rectangular. Some characteristics of the regions are listed in Table 2. The elevation differences in the Ulvenhout and Biddinghuizen region are very small (within 10 metres).

DATA

Landsat Thematic Mapper (TM) images were applied. Both images were obtained under good atmospheric conditions. Bands 3 (0.63 to 0.69 μm), 4 (0.76 to 0.90 μm), and 5 (1.55 to 1.75 μm) were used for the classification. The size of Landsat TM pixels is approximately 30 by 30 meters. The acquisition dates for the Ulvenhout and Biddinghuizen regions were 3 August 1986 and 5 July 1987, respectively.

Land-cover data for the Ulvenhout region were obtained by interpreting false color photographs taken at an altitude of 2000 m, resulting in a scale of 1:13 200, and field information for a subarea of 400 ha. For the Biddinghuizen region, land-cover data were obtained from land-cover maps (scale 1:5000) of a regional administrative institute. Until 1988 these data had been gathered yearly for planning purposes. These data were used to digitize the object boundaries required for the object classification. The actual land-cover classes were also stored for validation purposes.

DATA PROCESSING

The Landsat TM raster data were processed using the remote sensing image processing system ERDAS while the vector data were stored and processed by means of the geographical information system ARC/INFO.

To facilitate integration of the geographical and remote sensing data, the remote sensing data were georeferenced to the National Triangulation System. The images were geometrically corrected using a first order affine transformation. The root mean square (RMS) error for this transformation was 0.6 pixel (18 m) and 0.7 pixel (21 m) for the Ulvenhout and Biddinghuizen regions, respectively. The pixels were resampled to 30 metres using the nearest neighbor resampling method.

For the Ulvenhout region, training fields were selected for six different land-cover classes: water, built-up area, bare soil, grass, maize, and forest. In addition to these six land-cover classes, 7 percent of the area was occupied by land-cover classes (e.g., horticulture) which could not be discriminated by their spectral reflectance. In the Biddinghuizen region, seven different classes were distinguished: grass, potatoes, cereals, sugar beets, beans, peas, and onions. For every class, 125 to 460 pixels were used to determine the mean reflectance and covariance matrix. The per pixel classification was performed with a maximum likelihood classifier with equal prior probabilities for each class.

All relevant boundaries from the ground truth data were transposed to a standard 1:10,000-scale topographical map (projection: National Triangulation System). Object boundaries as well as actual land-cover classes were digitized and stored as polygons in a GIS. The digitized object boundaries for both regions are provided in Figures 2A and 2B. Additional information regarding the number and size of the objects is given in Table 2. The boundary index provides the total length of vectors for a square kilometre and is therefore a measure of the complexity of a region.

Two raster files were extracted from the vector database of both regions:

- a file with the per pixel object identification to enable the object classification (see General Approach) and
- a file with the ground truth for validation purposes.

The size of the raster elements was equal to the size of the Landsat TM pixels: 30 m by 30 m.

The object classification was performed subsequent to the maximum likelihood classification for both regions. The results were fed back to the GIS by means of the output ASCII file.

The raster files with the ground truth were used to validate the results of both classifications by calculating confusion matrices

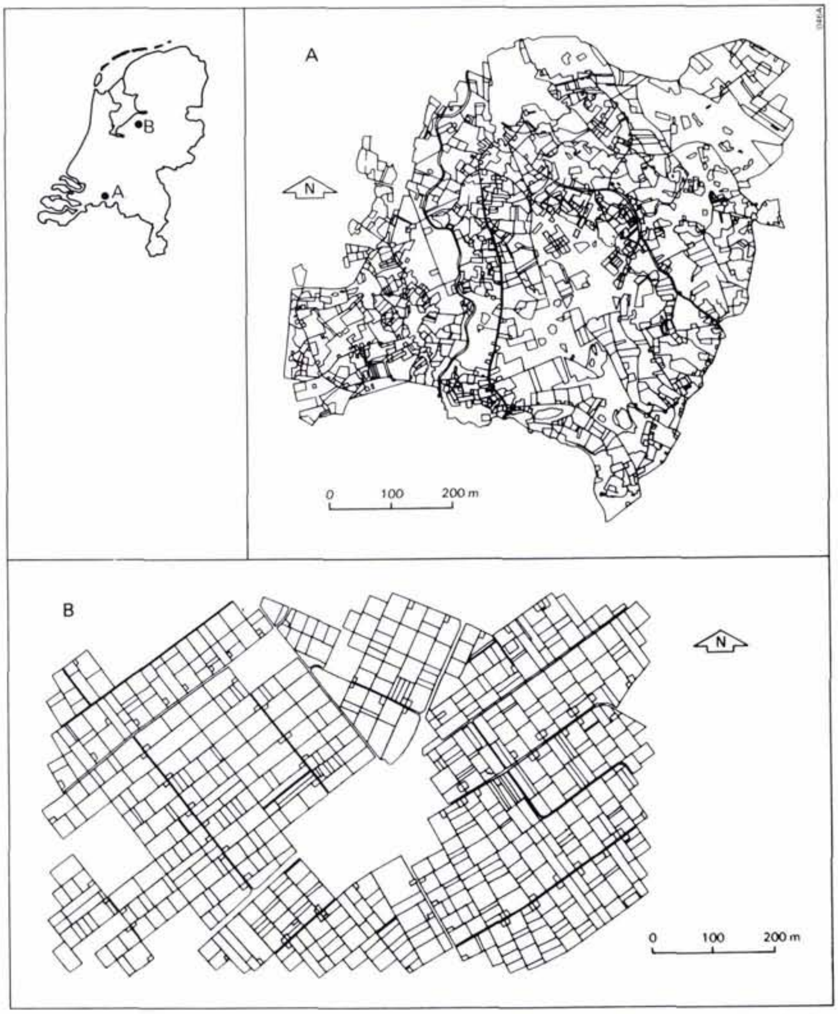

FIG. 2. Location of both test regions in The Netherlands. (A) Object boundaries in the Ulvenhout region. (B) Object boundaries in the Biddinghuizen region.

TABLE 2. MAIN CHARACTERISTICS OF BOTH TEST REGIONS

Test region	Area (ha)	No of objects	Average size of object (ha)	Boundary index ($km \cdot km^{-2}$)
Ulvenhout	3933	1360	2.9	13.2
Biddinghuizen	3754	542	6.9	10.0

TABLE 3. CLASSIFICATION ACCURACIES FOR THE PER PIXEL AND THE OBJECT CLASSIFICATION

	Overall accuracy (%)		
Test region	per pixel class.	object class.	Difference (%)
Ulvenhout	72	84	+12
Biddinghuizen	76	96	+20

and the overall accuracy. It should be noted that, for both regions, ground truth was available for the total area. Therefore, a small part (approximately 5 percent) was also used for training purposes. The biased validation caused by using the same areas for training and validation was ignored as this was of little consequence.

RESULTS

The results of the classification accuracies for both test regions are presented in Table 3. The accuracies of the per pixel classifications are indicative of land-cover classification performance in The Netherlands using satellite images. The object classification resulted in a significant rise in classification accuracy for both test regions.

The increase in overall accuracy in the Ulvenhout region was less than that of the Biddinghuizen area. The objects in the Ulvenhout region were smaller and more irregular (see Table 2), enlarging the fraction of mixed pixels per object. Furthermore, there was more spectral confusion between the classes of the Ulvenhout region than of the Biddinghuizen region.

From the principle of the object classification, not individual pixels but objects are incorrectly classified. For the Biddinghuizen region the correctly and incorrectly classified objects have been further analyzed. The mean frequency of the class that determined the label of the object was 77 percent in the case of the correctly classified objects and 62 percent for the incorrectly classified objects. It was concluded that the incorrectly classified objects were caused by spectral confusion rather than by mixed pixels. Figure 3 shows the objects of the Biddinghuizen region that were incorrectly classified. This plot demonstrates the comparison of actual and remote sensing derived labels in the GIS.

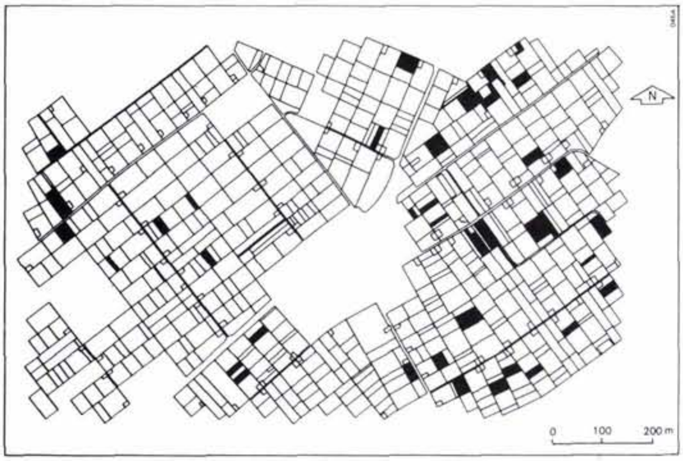

FIG. 3. Plot of the incorrectly classified objects in the Biddinghuizen test region.

DISCUSSION AND CONCLUSIONS

At present an increasing amount of geographical data are being stored in geographical information systems. The data from a GIS could prove useful in the processing of remote sensing images. In addition, remote sensing images may be applied to store and update data in a GIS.

The object classification developed should be considered as an attempt to integrate the processing of geographical data and remote sensing imagery. A prerequisite of the object classification is that the boundaries of all objects are required to be stored. It is therefore suitable for small areas (less than 10 000 ha); the higher classification accuracy would outweigh the effort of digitizing. The ongoing digitizing of maps and the advances in digital image processing will in the near future provide more digital topographical data. The possibilities of updating locational data using remote sensing data are being studied (Lemmens and Verheij, 1988; Swann *et al.*, 1988).

An advantage of the object classification is the availability per object of land-cover data. Consequently, this is much more appropriate from the point of view of data storage and feed back to a GIS. The integration of vector and raster data was made possible by gridding the vector data and processing the data with the image processing system. In an inventorying of geographical information systems, it was found that there is a large growth in systems that can handle both raster and vector structures (Parker, 1989). However, most of these systems only support a graphical integration. Database integration is also required for methods such as the object classification. This calls for much more of the concepts of an integrated GIS (Ehlers *et al.*, 1989).

The presented study illustrates that geometrical data can improve remote sensing classification accuracy. The geometric data stored in a GIS were successfully used to improve land-cover classification by means of an object classification. Objects are defined as areas (polygons) in which one land-cover type is expected. This spatial context enabled the correction of the incorrectly classified (mixed) pixels to a large extent. In the tests presented the overall accuracy increased by 12 percent and 20 percent, respectively.

Future research dealing with land-cover classification will be focused on the use of GIS thematic data (soil type or former land-cover type) in a knowledge-based remote sensing classification procedure.

REFERENCES

Catlow, D. R., R. J. Parsell, and B. K. Wyatt, 1984. The integrated use of digital cartographic data and remotely sensed imagery. *Proceedings Integrated Approaches in Remote Sensing*. Guildford, UK, ESA SP-214, pp. 41–45.

Ehlers, M., G. Edwards, and Y. Bedard, 1989. Integration of remote sensing with geographic information systems: a necessary evolution. *Photogrammetric Engineering & Remote Sensing* 55, 11:1619–1627.

Heymann, Y., 1987. Corine project – land cover programme. *Statistical Assessment of Land Use: The Impact of Remote Sensing and other Recent Developments on Methodology*. Eurostat 3E, Luxembourg, pp. 31–46.

Hill, J., and J. Megier, 1988. Regional land cover and agricultural area statistics and mapping in The Department Ardeche, France, by use of Thematic Mapper data. *International Journal of Remote Sensing* 9,10/11:1573–1595.

Hutchinson, C. F., 1984. Techniques for combining Landsat and ancillary data for digital classification improvement. *Photogrammetric Engineering & Remote Sensing* 48,1:123–130.

Ioka, M., and K. Masato, 1986. Performance of Landsat-5 TM data in land cover classification. *International Journal of Remote Sensing* 7,12: 1715–1728.

Kenk, E., M. Sondheim, and B. Yee, 1988. Methods for improving accuracy of thematic mapper ground cover classification. *Canadian Journal of Remote Sensing* 14,1:17–31.

Lemmens, M. J. P. M., and K. F. Verheij, 1988. Automatically GIS updating from classified satellite images using GIS knowledge. *Proceedings 16th Intern. Congress ISPRS*, Kyoto, Japan, Vol. 27, Part B4: pp. 198–206.

MARS, 1989. *Monitoring Agriculture with Remote Sensing. Tele-Agri-News 1*. Joint Research Centre, ISPRA, Italy. 13 p.

Shimoda, H., K. Fukue, R. Yamguchi, Zinjue Zhang, and T. Sakata, 1988. Accuracy of landcover classification of TM and SPOT data. *Proceedings IGARS Symposium*, Edinburgh, Scotland, ESA SP-284: pp. 529–535.

Swann, R., D. Hawkins, A. Westwell-Roper, and W. Johnstone, 1988. The potential for automated mapping from geocoded digital image data. *Photogrammetric Engineering & Remote Sensing* 54,2:187–193.

Thunnissen, H. A. M., M. N. Jaarsma, and O. F. Schoumans, 1990. Land cover inventory in The Netherlands using remote sensing. *International Journal of Remote Sensing* (submitted).

Van der Laan, F. B., 1988. Improvement of classification results from a satellite image using context information from a topographic map. *Proceedings 16th Intern. Congress ISPRS*, Kyoto, Japan, Vol. 27, Part B8: VII, pp. 98–108.

(Received 7 December 1989; revised and accepted 5 May 1990)