KNOWLEDGE-BASED CLASSIFICATION OF LAND COVER FOR THE QUALITY ASSESSEMENT OF GIS DATABASE

Ammatzia Peled[a,*], Michael Gilichinsky[b]

[a]University of Haifa, Department of Geography and Environmental Studies, Haifa, 31905, Israel - peled@geo.haifa.ac.il

[b]Swedish University of Agricultural Sciences, Department of Forest Resource Management, SLU Umeå, SE-901 83, Sweden – michael.gilichinsky@srh.slu.se

Commission IV, ICWG IV/VIII

KEY WORDS: Knowledge-based classification, Land cover Classification, GIS updating, GIS quality assessment.

ABSTRACT:

Valid classification of remotely sensed data is one of the most studied issues in the geo-information science. In recent years knowledge-based approach to image analysis has been developing for assessment and improvement of traditional statistically-based image classification. Knowledge-based classification procedure integrates remote sensing imagery with ancillary geospatial information from GIS. Data about land cover stored in GIS database are usually subjected to an intensive change processes that diminish their relevance and include different types of discrepant information. Classification of land cover by up-to-date satellite imagery and automatic updating of GIS database allows revision of discrepant or erroroneous data. The knowledge-based classification doesn't require any assumptions regarding the data distribution and allows straightforward incorporation of ancillary data from GIS. Compared to traditional mapping approaches knowledge-based classification has the advantages of lower cost, area-wide coverage, and possibility to frequent updating. In perspective of large GIS maintenance the knowledge-based classification may contribute to detection of change and assist automatic updating of spatial databases.
The objective of this study was to perform knowledge-based classification of land cover using satellite remote sensing data and GIS ancillary data. The selected target groups of land cover from the Israeli National GIS have been characterized spectrally by multispectral IKONOS data and geometrically by GIS data. The formalized knowledge about the target groups was incorporated into classification of remote sensing data. By means of classification results the discrepant land cover polygons have been detected and suggested for revision. Discussed are classification results and the analysis of detected discrepancies. The classification results have provided an indication of the utility of formalized knowledge for classification of land cover. The proposed method could be one possible approach to quality assessment and may contribute to automatic updating of existing spatial databases.

* Corresponding author

1. INTRODUCTION

Valid classification of remotely sensed data is one of the most studied issues in the geo-information science. Statistical methods of image analysis in remote sensing have served as most common approach to classification of the investigated phenomena. Classification is, in general, an area of multivariate statistics, according to that deals with grouping a set of items by assigning them to several similar classes (Duda and Hart, 1973). The GIS-driven methodology was proposed by Peled (1994) for integration of spatial information from GIS databases for national-wide classification of remote sensing data. The GIS-Driven classification is done by training the system, about the spectral characteristics of GIS objects and phenomena. In a similar way, Hellwich et al (2005), postulates that without training the human interpreters in the detection of unfamiliar objects, it is nearly impossible to correctly classify images.

Fast development of advanced methods for image classification such as artificial neural networks, decision trees approaches or expert systems has led to the integration of imagery with ancillary non-image (geospatial) information in classification procedure (Lawrence and Wright, 2001; Mertikas and Zervakis, 2001). Such approaches do data analysis lend themselves well to the integration of imagery with non-image (geospatial) data because, unlike traditional classification algorithms, they do not require assumptions regarding the normality of the data distribution and they can readily accommodate heterogeneous data (discrete and continuous, numerical and categorical) that may have varying degrees of accuracy (Merchant and Narumalani, 2009).

Compared to more traditional mapping approaches such as terrestrial survey and basic aerial photo interpretation, knowledge-based classification of land cover has the advantages of lower cost, area-wide coverage, and possibility to frequent updating. Consequently, additional geo-spatial information products have become an essential tool in many operational programs involving land resource management (Sun, 2003).

Data about land cover stored in GIS database are usually subjected to an intensive change processes that diminish their relevance and include different types of discrepant information. Classification of land cover by up-to-date satellite imagery and automatic updating of GIS database allows revision of discrepant or erroroneous data. Having assumed that the number of wrongly captured GIS objects (or classification types) are substantially less than the number of all GIS objects of the data set, the training areas can be derived automatically from the already existing GIS data (Walter, 2000).

In this study we present the framework method for incorporation of knowledge about land cover types from the Israeli National GIS database into classification procedure of remote sensing data. This method is suppose to improve the process of quality assessment by discrepancies detection and might lead towards automatic updating of existing GIS databases.

2. STUDY DATA

2.1 The Israeli National GIS

The Israeli National GIS database is stored in digital vector format and corresponds to the 1:40,000-scale air photographs. All spatial phenomena in the system are classified into ten main thematic layers (including land cover). The information about land cover phenomena is highly useful for selecting training samples and supervised classification. As ancillary data, the Israeli National GIS data may contribute to improvement of the quality of automated land-use classification based on satellite imagery data. The massive amount of spatial data of the Israeli National GIS database (nearly 1000 types) is regularly updated and manual revision of classification and features geometry is carried out. In present study five target groups of land cover polygons from the Israeli National GIS (610 polygons in total) have been selected: natural forest; citrus orchards; fruit orchards; agricultural fields and water reservoirs.

2.2 Satellite data

The study area of ~50 km^2 is located in country side in north of Israel (Fig. 1). The area is covered by the geo-referenced and radiometrically calibrated IKONOS images (RGB+NIR) with 4 meter spatial resolution (January 2005). The NDVI image for the scene was generated for representing the differences between land cover types in further processing.

Figure 1. Location map of the study area (black rectangle).

3. CLASSIFICATION

The knowledge-based classification in our study was realized by four main steps:

I. Obtaining of spectral (NDVI) and geometric parameters (area and compactness) for selected land cover target groups (Table 1);
II. Segmentation of NDVI image accordingly to obtained spectral parameters and area of target groups. The standard deviation was used for setting of segmentation thresholds of variation ranges;
III. Classification of revealed segments accordingly to their compactness;
IV. Validation of classification results by ground truth data.

Table 1. Characterizations of land cover target groups.

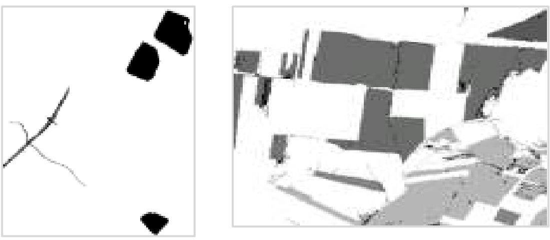
	NDVI		*Compactness*		*Area, m^2*	
	Mean	**St. Dev**	**Mean**	**St. Dev**	**Mean**	**St. Dev**
Natural forest	0.24	0.05	0.33	0.18	70015	200286
Citrus orchards	0.35	0.13	0.38	0.11	9372	12075
Fruit orchards	0.36	0.16	0.47	0.17	12200	17503
Agricultural fields	0.29	0.16	0.30	0.13	73303	145702
Water reservoirs	-0.13	0.07	0.51	0.11	24498	39333

The knowledge about land cover polygons from the target groups of the Israeli National GIS database was formalized by geometric parameter of compactness and mean polygon area. The parameter of compactness expresses the extent to which a shape is a circle. A circle's compactness is 1 and a long bar's compactness is close to 0. Spectrally all target groups were characterized by mean NDVI value.

Negative values of NDVI (values approaching -1) correspond to water. Values close to zero (-0.1 to 0.1) generally correspond to barren areas of rock, sand, or build-up areas. Low positive values represent shrub and grassland (approximately 0.2 to 0.4); high values approaching 1, indicate rich vegetation (tropical rainforests) [Earth Observatory, 2010]. In Mediterranean environment utility of NDVI for crop recognition through knowledge-based classification was shown by Cohen and Shoshany (2002; 2005).

Thus "typical" variation range for every selected land cover target group has been defined in form of spectral and geometric rules (step I). The regional segmentation of NDVI image into areas of connected pixels based on the pixel NDVI value (step II) was performed accordingly to NDVI and area variation ranges. This object-wise approach is more suitable rather than only spectral thresholding typical to single-pixel-based methodologies.

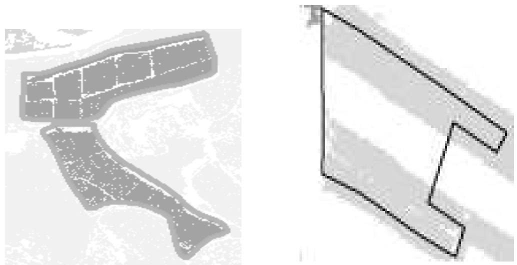
On the step III the rules obtained from step I were applied on segmentation results (step II) and revealed segments have been classified accordingly to the compactness of land cover target groups (Fig 2).



(a) (b)

Figure 2. Example of image segmentation. (a) – Land cover segments with similar range of NDVI values but different compactness (water reservoirs and roads); (b) – Land cover segments of agricultural fields (in dark grey) and citrus orchards that have the similar range of compactness but disparate range of NDVI values.

On the final stage the classified segments were matched to outlines of existing land cover polygons and decision about polygon's correctness has been made when more than 80% of the polygon's area lay within the segment of the same class (Fig 3a). Otherwise the polygon is defined as discrepant and suggested for revision in GIS database (Fig 3b).



(a) (b)

Figure 3. Example of discrepancy detection. (a) – Two correct polygons of citrus orchards, more than 80% of the polygon's area lay within the segment of citrus orchards (in dark grey). (b) Discrepant polygon of fruit orchard, less than 80% of the polygon's area lay within the segment of the same class.

4. RESULTS

The classification results of five land cover target groups (classified with overall accuracy of 83.6% and kappa=0.71) are presented in Table 2.

Table 2. Confusion matrix of classification results (reported in percents).

		Ground truth				
		Agricultural fields	*Natural forest*	*Citrus orchards*	*Fruit orchards*	*Water reservoirs*
Classification	*Agricultural fields*	62.0	14.5	4.5	8.4	2.4
	Natural forest	17.2	78.7	6.0	9.2	2.0
	Citrus orchards	11.3	2.7	75.1	0.6	0.5
	Fruit orchards	6.7	3.2	13.2	80.4	0.9
	Water reservoirs	2.8	0.9	1.2	1.5	94.2

The comparison of classification results with ground truth land cover information was carried out on polygon level by up-to-date manual air-photo interpretation and field survey. The misclassified (discrepant) polygons were suggested for revision in GIS database (Table 3).

Table 3. Polygon-wise discrepancy detection

	Total	*Detected as correct*	*Detected as discrepant*
Agricultural fields	256	159	97
Natural forest	179	141	38
Citrus orchards	84	63	21
Fruit orchards	51	41	10
Water reservoirs	40	38	2

The target group of water reservoirs has demonstrated highest accuracy (94.2%) among studied target groups with only two discrepant polygons. The target groups of citrus orchards and natural forest have demonstrated similarly moderate accuracy (75.1% and 78.7%, correspondingly). The classification accuracy of fruit orchards has been estimated on 80.4%. The lowest classification accuracy of 62.0% with 97 polygons suggested for revision has been achieved for target group of agricultural fields.

5. DISCUSSION

In the present study we propose the method for incorporation of ancillary land cover data into classification process of remote sensing imagery. The proposed method of knowledge-based classification is meant to improve the process of discrepancies detection and automatic updating of existing spatial databases. The method might be especially useful for land cover types that included in the Israeli National GIS database.

The main component of the successful discrepancy recognition by the proposed method is spectral and geometric similarity of polygons in land cover target groups. The homogeneous land covers with lesser deviation (i.e. water reservoirs) of spectral and geometric characteristics have demonstrated higher classification accuracy. In this perspective poor classification accuracy of most deviated target group of agricultural field is reasonable.

To overcome difficulty in classification of heterogeneous phenomena (i.e. urban land cover) the additional spectral and geometrical parameters might be included in knowledge-based classification, as well as formalization of semantic definitions by means of evidential reasoning (Shoshany and Cohen, 2007). Additional spectral (i.e. vegetation indices), geometric parameters (eccentricity, thickness etc) or ancillary soil-type information might be easily integrated into suggested classification process. Application of suggested knowledge-based classification framework on country-wide extent might be considered and further methodologically developed. In this perspective, consideration of seasonal changes in land cover by means of NDVI time series might be of interest.

The overall conclusion is that the proposed framework method of knowledge-based classification seems to be a possible way of quality assessment and might lead towards automatic revision/updating of land cover GIS databases.

REFERENCES

Cohen, Y. and Shoshany, M., 2002. A national knowledge-based crop recognition in Mediterranean environment, International Journal of Applied Earth Observation and Geoinformation, Vol. 4, Issue 1, pp. 75-87.

Cohen, Y. and Shoshany, M., 2005. Analysis of convergent evidence in an evidential reasoning knowledge-based classification. Remote Sensing of Environment, Vol. 96 , pp. 518-528.

Duda, R., and Hart, P. E., 1973. Pattern Classification and Scene Analysis, Wiley, New York, 482p.

Hellwich, O., Heipke, C. and Wessel, B., 2001. Sensor and data fusion contest: Information for mapping from airborne SAR and optical imagery. In: Proc. IEEE International Geosciences Remote Sensing Symposium. (IGARSS), Sydney, OEEPE Newsletter, no. 1, pp. 10-11.

Lawrence, R. L. and Wright, A., 2001. Rule-based classification systems using classification and regression tree (CART) analysis. Photogrammetric Engineering and Remote Sensing, Vol. 67, Issue 10, pp. 1137-1142.

Merchant, J. W. and Narumalani, S., 2009. Integrating Remote Sensing and Geographic Information Systems, In T. A. Warner, G. M. Foody, and M. Duane Nellis (Eds.), The SAGE Handbook of Remote Sensing, SAGE Publications, London, p. 263.

Mertikas, P. and Zervakis, M. E, 2001. Exemplifying the theory of evidence in remote sensing image classification, International Journal of Remote Sensing, Vol. 22, pp. 1081-1095.

Peled, A. 1994. Revision of Digital Maps and GIS Databases. ISPRS International Archives of Photogrammetry, Vol. 30, Part 4, pp. 268-272.

Shoshany, M., and Cohen, Y., 2007. Inference versus evidence in reasoning remote sensing recognition, with an information foraging perspective. International Journal of Remote Sensing, Vol. 28, pp. 2613-2634.

Sun, W., Heidt, V., Gong, P. and Xu, G., 2003. Information fusion for rural land-use classification with high-resolution satellite imagery, IEEE Transactions on Geocience and Remote Sensing, Volume 41, Issue 4, pp. 883-889

Walter, V., 2000. Using New Computer Aided Techniques: Can manual updating of GIS data become obsolete?, GeoInformatics No. 6, Vol. 3, pp. 38-41.