Integration of remote sensing, GIS and expert knowledge in national knowledge-based crop recognition in Mediterranean environment

Yafit COHEN, Maxim SHOSHANY

Bar-Ilan University, Israel
coheny9@ashur.cc.biu.ac.il

KEY WORDS: Agriculture, Remote Sensing, GIS, Integration, Knowledge-based rules.

ABSTRACT
This paper describes a knowledge-based crop recognition system, integrating remote sensing analysis and geographical information in hierarchical way. Landsat TM and Spot pan images were merged to reduce heterogeneity by enhancing field boundaries. Multi-temporal NDVI maps generated from these images were classified into eight crop types using unsupervised classification algorithm. Using agricultural knowledge, the relations between natural vegetation and crop types, spectral and phenological properties and precipitation and soil types were engineered. These relations were used as binary rules in an experimental knowledge-based crop recognition system. The binary rules were activated by iterative "split-and-merge" mechanism of the mixed unsupervised classification clusters aimed at refining the map products given by the application of unsupervised classification algorithm alone. Experiments in a wide region of Israel using ground data from the Israeli Agriculture Ministry have shown that the use of knowledge-based "split-and-merge" rules gives improvement of approximately 9% compared to the unsupervised classification alone.

1 INTRODUCTION

Landuse policy in Israel attracted wide and profound attention (see for example Vitkon, 1991; Shlain & Feitelson, 1996; Efrat, 1998) due to its severe and unreversible environmental implications and due to its potential effect on life quality for the years to come. The rapid population and urbanization growth (Shoshany & Goldshlager, 1998) result strong pressure for urban development on the account of agricultural and open lands. Cultivated areas in Israel have shown a mean annual reduction of 2% in the last 15 years. Despite limited extent of agricultural lands, Israeli agricultural production satisfies the local fresh consumption (except for cereals and other crops, which demand wide areas and massive irrigation) due to a continuous increase in land productivity (Maor Eli, 1999). However, lack of a solid long-term national land-use invites inadequate land utilization and a continuous decrease of cultivated areas as a result of spontaneous processes and local economic interests (Niv, 1999).

The realization of national landuse policy aiming at achieving regulatory framework for food, agriculture, open areas and environmental quality requires adequate sources of information providing the capability of detailed landuse mapping. The development of a reliable national crop monitoring system is representing principal step of achieving this goal.
Remote sensing techniques have been shown to be cost efficient and useful for crop identification in wide regions of the world. Remotely sensed image analysis for land cover recognition is commonly applied through standard ML (maximum-likelihood) classification algorithm (Srinivasan & Richards, 1990). The synergy of remote sensing, environmental sciences and engineering has emerged from the challenge to improve present monitoring capabilities (Wilkinson, 1996) and to monitor heterogeneous environments at different scales (Peddle, 1995). The knowledge-based expert system approach has been increasingly adopted over the years (Matsuyama, 1987; Nicollin & Gabler, 1987; Ton et al., 1991; Kartikeyan et al., 1995; Pigeon et al., 1999; Soh & Tsatsoulis, 1999) in the view of the need for image analysis procedures integrating both numerical and logical information (Hinton, 1996).

Image classification knowledge-based systems can be defined as computer programs designated for matching a land cover type to a pixel in areas with high degree of complexity, which requires wide domain knowledge for acieving satisfactory recognition. In general, a knowledge-based system is composed of two main elements
(Frost, 1986). Adjustments of these two elements for land cover classification are as follows:

1. A knowledge base: A set of simple facts composed of imagery and environmental data and information and a set of rules, which describe relations between facts and ways for reliable pixel labeling from them.
2. A problem-solving mechanism: Finding a recognition path from a specific set of facts describing a pixel to one (or more) land cover through relevant production rules.

These systems enables handling uncertainty, by attaching confidence values to each production rule according to evaluation of expert knowledge (Alty & Coombs, 1984). By using this evidential approach the knowledge-based system separates the knowledge required for the recognition from the recognition mechanism (Kontoes et al., 1993) as the its algorithm passes over rules according to their confidence values, searching for the most suitable land cover (Lee et al., 1987).

This approach have been applied and studied for crop recognition in several researches (Janssen & Middelkoop , 1992; Kontoes et al., 1993; Kontoes & Rokos, 1996; Adinarayana & Krishna, 1996). In these studies knowledge-based system was utilized for the refinement of conventional image-based classification methods and thus was used in a post-classification process. Nevertheless, high spatial-temporal heterogeneity might be responsible for the limited utilization of this approach in Mediterranean regions (Shoshany, 2000). This paper examines the potential of the incorporation of remotely sensed data, GIS information and expert knowledge for crop recognition applied through knowledge-based classification rules in a post-classification manner in a wide heterogeneous Mediterranean region.

2 Knowledge-Based Crop Classification Method

Presented in this paper a knowledge-based crop recognition refinement of an unsupervised classification. Wheat, four types of legumes, cotton, sunflower and orchards were to identify in this framework presenting 70% of the overall crops in Israel. The heterogeneous characteristic of the study area invited two kinds of recognition difficulties. On one hand this variability entailed several spectral and phenological sub-features for a single crop type and on the other hand several crops had common sub-features resulting in significant confusion. The ability to unite all sub-features of a single crop and to overcome confusion between crops demands knowledge about crop spectral/ temporal feature variety and ways environmental characteristics and agricultural treatments affect them.

2.1 Knowledge acquisition and engineering

Acquisition of these kinds of knowledge was achieved through data structuring and a comprehensive learning course of the inter-relations existing between the data sources in relation to different crop types. Data structuring included integration of imagery properties and digitized soil and rainfall maps (figure 1). The relations were studied by domain literature, interviews with experts and GIS overlay analysis applied on data sources utilizing empirical samples of crop plots chosen from ground reference maps. Engineering of retrieved relations involved rules generalization and formalization and designated for rules base construction.

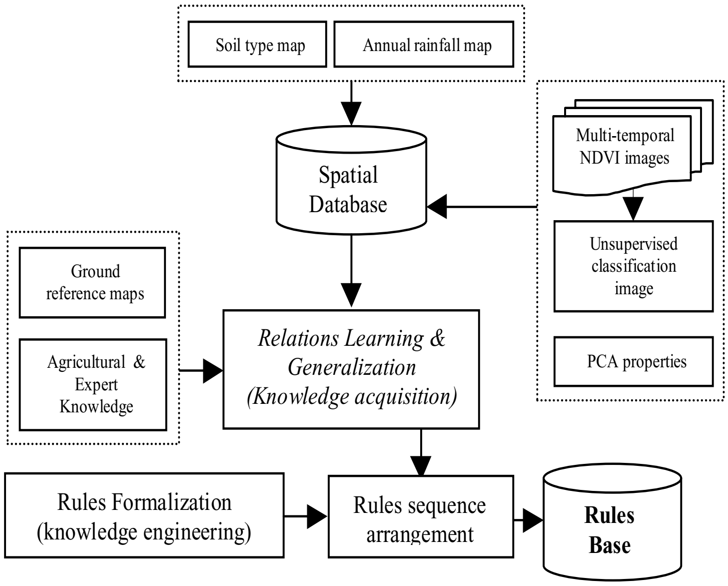

Figure 1. Rule-base construction procedure

2.1.1 Data structuring

Second-ordered imagery information of NDVI images, multi-spectral PCA and basic classification layer integrated with digitized rainfall and soil maps composed the database. These components together with agricultural knowledge and GIS analysis used for extracting the rule-base.

<u>*NDVI maps*</u>

Multi-temporal NDVI maps covering the central and southern parts of Israel were generated from five Landsat TM images acquired during 96'-97' growing season in clear sky conditions. A Spot Panchromatic image from June 96' was combined with the NDVI images in two stages. Firstly, the NDVI maps were re-sampled to the Spot spatial resolution (10m). Secondly, Spot edge pixels, were enhanced, coded and reduced from NDVI images in order to reduce overall image heterogeneity with minimal information loss. The analyzed multi-temporal NDVI images were used as a basis for traditional unsupervised classifier and as additional data in the knowledge acquisition process for the rule-based classification.

<u>*Isodata unsupervised classified images*</u>

Basic categorization of the multi-temporal NDVI maps was conducted by detailed Isodata (Iterative Self-Organizing Data Analysis) unsupervised classification algorithm. In heterogeneous areas the recognition flexibility of the unsupervised method makes it an efficient starting point since it enables recognition of sufficient amount of crop sub-features in the

images (Groome, 1998). Supervised grouping of phenological sub-classes into meaningful land covers followed the Isodata clustering. This process was based on training plots of cultivated fields from ground reference maps supplied by surveyors from the Israeli Agriculture Ministry and representing 1% of the total fields in the study area. Phenological features and geographical context consideration such as location and shape were also utilized in this grouping process.
All plots were then compared with the classified image. The comparison process involved classification accuracy estimation and locating "mixed clusters" i.e. classification clusters, which represent more than one crop or other land cover. In addition, possible reasons were attached as comment to plots, which classified incorrectly. The coments related to phenological behavior, plot size, plot texture and location. They were organized in tables and then used for further learning and defining phenological behaviors as reflected from the NDVI images and their relations with image spatial considerations. Examination of Isodata classification image shows recognition of wheat, vetch, chickpea, pea, cotton, sunflower, orchards, high dense shrublands, other natural vegetation, non-vegetated area.

Spectral PCA

Comprehensive examination of "mixed clusters" had shown that phenological attributes are not sufficient for distinguishing between the orchards and other natural tree formations (NTF). PCA was applied separately on winter (April) and summer (June) spectral images. Second and third PC's were visually identified as the most discriminative for this purpose. Utilizing the NDVI format, difference enhancement of the two PCs was calculated following amplitude enhancement of each PC channel. These values were successfully utilized in order to differentiate between tree formations.

Average annual rainfall map

Steep climatic gradients exist both North-to-South and West-to-East across the country. These climatic variations cause high variability in crop seeding and harvesting periods. Climatic index could mask climatic regions and limit specific phenological features to specific crop. Thus, identical features of two crops could be divided by the climatic mask and different sub-features of the same crop would be identified and united on the basis of their climatic-dependence differences. In this experimental recognition system an averaged annual rainfall map (1: 500000) from the Israeli Meteorological Center was combined as a representing climatic indicator. The analog map was scanned and converted into a co-registered raster form in a vector-to-raster process.

Soil type map

Soils have a strong influence on land use suitability and crop cultivation and on photosynthetic activity. Therefore, soil types map might improve recognition quality where for example similar phenological sub- features exist for both crop and natural vegetation types under similar climatic conditions. In the described experiment the soil contribution was examined on a small region from the study area, as both soil types heterogeneity in the study area is high and the connections between cultivated areas and soil types are complex.

2.1.2 "Split-and-merge" rules-base generation procedure

Rules-base generation involved an inductive-learning of existing inter-relations, among database sources; relation generalization process, rules formalization and combined sequence-dependent rules-base construction (figure 1). Empirical samples chosen from ground reference maps provided examples of the concept to be learned. The learning course involved domain expert's expertise to include from empirical samples some generalized relations, that can be used to classify the remaining data (Huang & Jensen, 1997). In this study retrieved generalized relations were used as splitting criteria of unsupervised classification "mixed-clusters" in a "split-and-merge" classification refinement mechanism. Figure 2 describes splitting criteria extraction stages. Firstly, relationship typology was

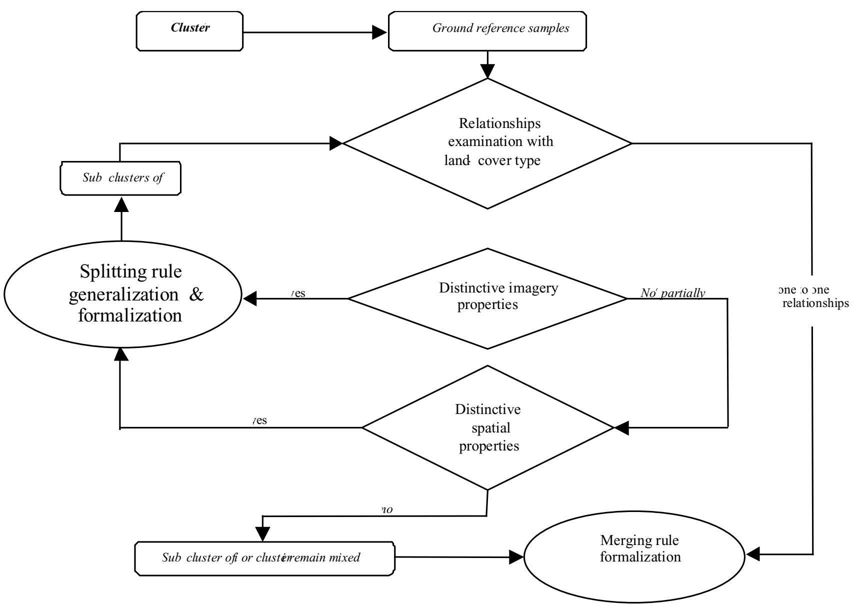

Figure 2. Splitting criteria extraction procedure

checked and defined between certain cluster and land-cover. If a checked cluster was found distinctive, i.e. relates to one land cover type, (one-to-one relationships) a merging rule was formalized. If a checked cluster was found mixed, i.e. relates to more than one land cover type (complex relationships), a hierarchic splitting rule generation process was applied. Primarily, splitting potential, which lies in imagery properties was examined. Splitting criteria that could be explained by environmental or agricultural effect were formalized for further classification. Only if splitting criteria based on imagery data wasn't sufficient were spatial properties examined. Actually, most splitting criteria combined imagery and spatial constraints. By this mechanism each "mixed-cluster" was splitted into more distinguished sub-clusters, which were then merged respectively with suitable land cover classes. Knowledge engineering was involved in order to make retrieved relations concept a computer-usable knowledge. They were translated into formal binary rules. Basic rule format used here is as follows:

IF <condition: C_1 and C_2 and …C_n> Then <conclusion: L_i>

where C_i are logical constrains on one or more database features and L_i are land cover types. Rule condition referred firstly to a mixed cluster and then to a set of constraints. In the conclusion part each pixel was labeled with certain land cover type whenever a set of conditions was true. Actually, the condition and the conclusion parts applied the splitting and merging actions respectively. Most knowledge-based systems in remote sensing analysis utilize evidential reasoning approach through Dempster-Shafer theory of evidence (for example: Lee et al., 1987; Srinivasan, 1990; Kontoes et al., 1993; Peddle, 1995). However, in this framework no evidential values were attached to rules and as a result no separation was made between the knowledge base and the inference mechanism and no confidence calculations were made to evaluate each accepted recognition conclusion. Since there were more than one rule that their set of conditions were fulfilled in a single pixel and lead to different land-cover, the rules were manually and iteratively put in changing order till best recognition was achieved. Figures 3a,b describe evidential versus definite knowledge-based iterative classification mechanisms. Primarily, definitive approach was applied in order to examine the potential of multi-source integration approach for crop recognition in Mediterranean region. Progressive process will include evaluation of attached evidential values for each rule, and the application of an inference engine separated from the knowledge base.

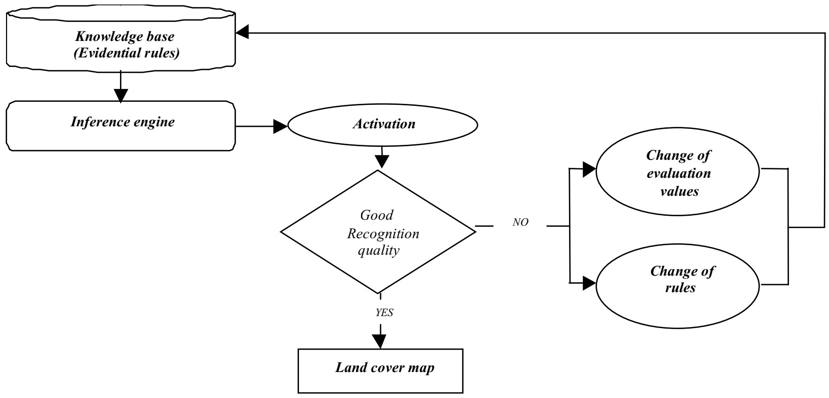

Figure 3a. Evidential knowledge-based iterative classification mechanism

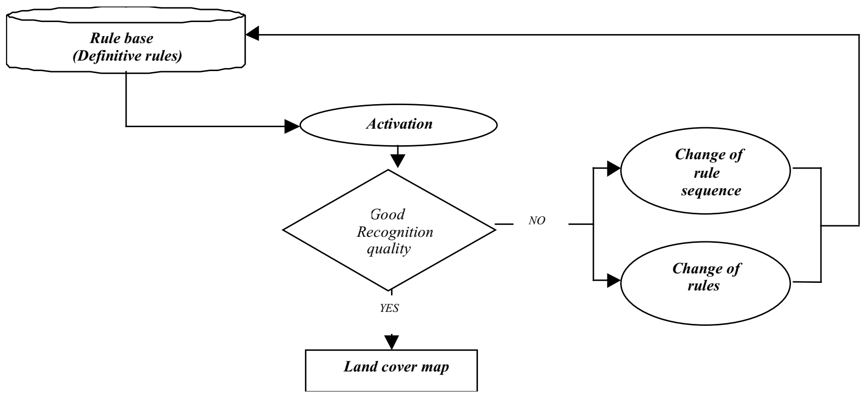

Figure 3b. Definite knowledge-based iterative classification mechanism

2.1.3 Examples of "split-and-merge" criteria extraction

Winter crops

Most winter crops are hardly irrigated as their development considerably depends on rainfall amounts and temporal distribution. As a result their phenologies are similar and achieving high differentiation between them is not simple. However, it seems that NDVI temporal values reflect real minor modifications between different winter crops' phenologies that are not recognized in unsupervised classification. In order to trace these changes, statistical analysis of mixed clusters was applied with ground reference samples and maximum and minimum values were found for sub-clusters. Nevertheless, only modifications that reflect real differences between crops were then formulated into splitting rules.

Wheat & Herbaceous

Among mixed clusters two were composed mainly of wheat and herbaceous with similar phenology. The potential ability to distinguish between them was checked through knowledge concerns wheat/herbaceous phenology variants and their dependence on environmental conditions. Following that, existence of these relations in imagery data was inspected by overlay analysis utilizing empirical samples of wheat and herbaceous from the mixed clusters. The analysis had shown that most wheat samples were found under semi-arid conditions and most herbaceous samples were found under more humid areas. Rainfall data was used then as a mask and combined with NDVI values constrains the mixed cluster was splitted and the two revealed sub-clusters were then merged into wheat and herbaceous "super" classes respectively.

Orchards and Natural tree formations

Major classification confusion was found between orchards and natural tree formations, both characterized by relatively stable photosynthetic activity with minor seasonal fluctuations. Comprehensive examination of "mixed clusters" had shown that phenological attributes are not sufficient for distinguishing between the two classes. Threshold enhanced PCA values for orchards and natural tree formations were determined based on empirical data and spectral-canopy connections. Initial examination of enhanced PC values combination had shown that orchards mislabeled as natural tree formations clusters have significantly higher values and natural tree formations mislabeled as orchards have significantly lower values. These results had proven a consistency in the behavior of the two land cover types, which was translated to several classification rules.

In addition, soil types map was used as splitting criteria between orchards and NTF on a limited area. Mixed clusters of TF were efficiently splitted by soil types. Furthermore, adjacent plots of orchards on different soils, extracted relatively distinctive phenologies. This fact shows that NDVI values are sensitive to dofferences caused by spatial variations.

3 RESULTS

In general, the integrated approach applied through knowledge-based classification rules improved recognition accuracy of crop types.

	wheat	**vetch**	**chick-pea**	**pea**	**clover**	**cotton**	**sunflower**	**Other crops**	**orchards**	**Natural veg**	**other**	
wheat	**82**	9		14	29			21	1	4		
vetch	5	**73**			14					4		
chickpea		9	**100**					7				
pea	3	9		**79**				3				
clover					**43**							
cotton						**95**	5	10	2			
sunflower						2	**95**	14	1			
other crops	2				7			**38**	1	10		
orchards						2		7	**71**	10		
Natural veg	6			7	7				19	**62**		
other	2					2			6	10	**100**	
Plots #	127	11	10	14	14	55	22	29	126	29	10	**75.5%**

Table 1. Confusion Matrix of unsupervised classification

	wheat	**vetch**	**chick-pea**	**pea**	**clover**	**cotton**	**sunflower**	**Other crops**	**orchards**	**Natural veg**	**other**	
wheat	**94**	9			7			21	1	17		
vetch	1	**91**			22							
chickpea			**100**					7				
pea	2			**93**				4				
clover					**64**							
cotton						**96**		14	1			
sunflower						2	**100**	14	1			
other crops	1				7			**38**	1	10		
orchards						2			**83**	10		
Natural veg	2			7				3	7	**53**		
other	0								6	10	**100**	
Plots #	127	11	10	14	14	55	22	29	126	29	10	**83.7%**

Table 2. Confusion Matrix of classification rules

Table 1 and 2 present confusion matrices of both unsupervised and knowledge-based classification. By using knowledge-based classification rules high classification accuracy was achieved for all crops, except for the clover. Despite significant improvement achieved for clover (>20%) it still wasn't recognized sufficiently. High confusion was found between clover and vetch. From experts and literature it was found that both legume types are often grown together in the same field. In addition, clover is hardly grown in Israel recently and most of its plots are very small. For these reasons it might be right to unite clover and vetch. A united class of them has high accuracy and reliability.

Main improvements were achieved through imagery splitting criteria generalized by agricultural knowledge. Rainfall data was utilized mainly to overcome confusion between legume types, wheat and legume types and wheat and herbaceous. Most splitting criteria based on rainfall data were combined with NDVI imagery data. Confusion between orchards and NTF was considerably decreased (12%) with the usage of enhanced values of spectral PCA alone. Soil-based splitting criteria achieved an improvement of 8% in orchards recognition in comparison to unsupervised classification in a limited area.

4 CONCLUSIONS

The integration of imagery data, environmental properties, agricultural and expert knowledge through "split-and-merge" rules has high potential for high crop recognition in Mediterranean areas. The classification process included detailed unsupervised classification for the recognition of crop sub-features, inductive learning of relations exist between spatial data and classification refinement by knowledge-based rules. It seems that each phase in the process has a considerable contribution for high recognition quality.

In this paper splitting criteria determination was made in hierarchical way. Imagery data got priority on environmental properties external to the image, in order to maximize image-based recognition and to utilize other information sources for refinements. In that way most rules were based on imagery data (spectral and temporal) and generalized with the incorporation of agricultural and expert knowledge. Environmental properties were shown to be very efficient in winter crops recognition and in distinguishing between crops and natural vegetation formations confusion.
Fully established knowledge-based system is under development in order to utilize these findings and improving recognition quality.

ACKNOWLWDGEMENT

We wish to thank the Ministry of Agriculture of Israel for the Farmer organization of Israel for financing this project.

5 REFERENCES

Adinarayana, J. & Rama-Krishna, N., (1996).Integration of multi-seasonal remotely-sensed images for improved landuse classification of hilly watershed using geographical information systems. Int. J. of Remote Sensing 17:1679-1688

Alty, J. L. & Coombs, M. J., (1984). Expert systems – concepts and examples. NCC publication, Manchester, England.

Efrat, E., (1998). Towards land policy changes in Israel. Ofakim in Geography 48-49: 147-155 (in Hebrew)

Frost, R., (1986). Introduction to knowledge base systems. Chapter 1 & Chapter 8. Macmillan, New York.

Groom, B. G., Fuller, R. M. & Jones, A. R., (1996). Contextual correction: techniques for improving land cover from remotely sensed images. Int. J. Remote Sensing 17:69-89.

Hinton, J. C., (1996). GIS and remote sensing integration for environmental applications. Int. J. GIS 10: 877-890.

Huang, X. & Jensen, J. R., (1997). A machine-learning Approach to automateed knowledge-base building for remote sensing analysis with GIS data. PE & RS 63: 1185-1194.

Janssen, L. L. F. & Middelkoop, H., (1992). Knowledge-based crop classification of a landsat TM image. Int. J. Remote Sensing 13: 2827-2837.

Kartikeyan, B., Majumder, K. L. & Dasgupta, A. R., (1995). An expert system for land cover classification. IEEE Trans Geoscience & RS 33: 58-66.

Kontoes,C. C., Wilkinson, G. G., Burrill, A., Goffredo, S. & Mégier, J., (1993). Int. J. GIS 7: 247-262.

Kontoes, C. C. & Rokos, D., (1996). The integration of spatial context information in an experimental knowledge-based system and the supervised relaxation algorithm- two successful approaches to improving SPOT-XS classification. Int. J. Remote Sensing 17: 3093-3106.

Lee, J., Richards, J. A. & Swain, P. H., (1987). Probabilistic and evidential approaches for multi-source data analysis. IEEE Trans Geoscience & RS GE-25: 283-293.

Ma'or, E., (1999). The agriculture in Israel in the Millenium – Towards what? Hed-Hamahar 18: 5-7.

Matsuyama, T., (1987). Knowledge-based aerial image understanding systems and expert systems for image processing. IEEE Trans Geoscience & RS GE-25: 305-316.

Nicollin, B. & Gabler, R., (1987). A knowledge-based system for the analysis of aerial images. IEEE Trans Geoscience & RS GE-25: 317-329.

Niv, A., (1999). National land policy is required. Soil 46: 5-8. (in Hebrew).

Peddle, D. R., (1995). Knowledge formulation for supervised evidential classification. PE & RS 61: 409-417.

Pigeon, L., Solaiman, B., Thompson, P. B. K. & Toutin, T., (1999). Remotely sensed images fusion for linear planimetric features extraction. Information fusion 99 conference, San-Francisco.

Shla'in, E. & Feitelson, E., (1996). Preservation policy of agricultural land – Its formation, institutionalization and undermining. Flousheimer institute for policy researches, Jerusalem. (in Hebrew).

Shoshany, M., (2000). Satellite remote sensing of natural Mediterranean vegetation: a review within an ecological context. Progress in Physical Geography 24: 153-177.

Shoshany, M. & Goldshlager, N., (1998). Growth rates in developed areas for the years 1950-1990: comparative examination of development trends in Israel in general and in the Carmel region in particular. Ofakim in Geography 48-49: 131-145.

Soh, L-K. & Tsatsoulis, C., (1999). Multi-source data and knowledge fusion for intelligent SAR sea ice classification. Proceedings, IGARSS99' - IEEE Int. Geoscience and RS Symposium, Hamburg, Germany.

Srinivasan, A. & Richards, J. A., (1990). Knowledge-based techniques for multi-source classification. Int. J. Remote Sensing 11: 505-525.

Ton, j., Sticklen, J. & Jain, A. K., (1991). Knowledge-based segmentation of Landsat Images. IEEE Trans Geoscience & RS 29: 222-232.

Vitkon, G., (1991). Land policy in the rural sector. Chapter 1. Jerusalem (in Hebrew)

Wilkinson, G. G., (1996). A review of current issues in the integration of GIS and remote sensing data. Int. J. GIS 10:85-101.