

DETECTION OF DISCREPANCIES IN LAND-USE CLASSIFICATION USING MULTITEMPORAL IKONOS SATELLITE DATA

Michael Gilichinsky^a and Ammatzia Peled^b

^a Samaria and Jordan Rift R&D Center, Ariel, Israel, e-mail: michaelgi@ariel.ac.il

^b Department of Geography and Environmental Studies, University of Haifa, Haifa, Israel.
e-mail: peled@geo.haifa.ac.il

Commission IV, WG IV/4

KEY WORDS: GIS-Driven classification, discriminant analysis, discrepancy detection, land-use

ABSTRACT:

Geoinformation systems (GIS) and other spatial databases containing land-use data are usually subjected to intensive change processes that impact the quality of their inherent classification and diminish its relevance. Consequently, with time, databases accumulate various types of erroneous information. The combination of the satellite data with the thematic land-use data from a core national GIS, provides an excellent case for GIS-driven analysis of land-use changes. The aim of this research was to assess the land-use changes using time-series of optical Ikonos satellite data.

An area of ~35 km² in the north of Israel served as the study case of the research. Seven land-use classes were detected in the relevant National GIS spatial database layers updated in the year 2000 and further in the year 2009. These seven type-classes were: water bodies, residential areas, agricultural fields, badlands, natural forests, build-up areas, and plantations. The Iterative discriminant analysis (IDA) algorithm was applied on both GIS datasets using corresponding Ikonos images acquired in 2002 and 2010, respectively. The IDA process resulted with a re-classification of the initial land-use polygons. It was assessed by validating the classification of all the land-use polygons. Comparing with Ikonos image from the year 2002, the fraction of the polygons that were correctly detected as consistent with the corresponding GIS dataset (77.9%) was relatively close to the fraction of polygons correctly detected as discrepant (75.5%). Classification of Ikonos image from the year 2010 showed that 81.9% of the land-use polygons were correctly detected as consistent whereas the fraction of polygons that were correctly detected as discrepant was about (78.3%). The main advantage of the proposed GIS-driven methodology for detection of changes in land-use classification is its analytical simplicity that allows for straightforward employment of spectral and spatial data in the classification process.

1. INTRODUCTION

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.

Compared to more traditional mapping approaches such as terrestrial survey and basic aerial photo interpretation, GIS-driven classification of land cover has the advantages of lower cost, area-wide coverage, and a possibility for frequent updating. Consequently, additional geospatial 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 erroneous 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 supposed to improve the process of quality assessment by discrepancies detection and might lead towards automatic updating of existing GIS databases.

2. MATERIALS

2.1 Study area and land-use GIS data

A rectangle area of ~35 km² Near Kiryat-Tivon served as the study case of the research. The region is characterized by semi-arid Mediterranean climate (0.2 – 0.5 p/pet) where natural vegetation is adapted to the distinctive climatic regime of dry warm summers and cool moist winters. The landscape is mostly covered with mosaic Mediterranean natural vegetation, villages, and croplands. The most favored seasons for vegetative growth are winter and spring when the soil is moist.

The land-use in the area has been exhaustively classified as a part of the National GIS since mid-1990s with the revision periods of 5 – 7 years. The land-use classification of the Israeli National GIS database is stored in digital vector format and

contains geo-spatial definition of nearly 100 classes based on visual interpretation of 1:40,000-scale air photographs. The basic element of the land-use classification is a polygon assigned by an interpreter with the standard attributes like land-use type, topological properties, and revision date. The massive amount of land-use data of the Israeli National GIS database is regularly updated and manual revision of classification is carried out.

In our study, the vector subset of the land-use classification layer (revised in the year 2000 and 2009) was used to cover the extent of the Ikonos satellite time-series of images, acquired in the years 2002 and 2010. For analysis seven land-use classes were selected: water bodies, residential areas, cultivated fields, badlands, natural forests, impervious surface, and plantations (Table 1). The final selection comprised in a total of 714 polygons for the year 2002 and 627 for the year 2010. The spatial constrain of the selection was the condition for each polygon to have an area of ten Ikonos pixels (160 m^2) at least.

2.2 Satellite data

Two cloud-free Ikonos scenes were acquired over the study area in January 12, 2002 and December 10, 2010. Ikonos satellite data is provided in form of multispectral imagery, consisting of four spectral bands: blue (B1, $0.45 - 0.52\ \mu\text{m}$); green (B2, 0.51

$- 0.59\ \mu\text{m}$); red (B3, $0.63 - 0.70\ \mu\text{m}$) and NIR (B4, $0.76 - 0.85\ \mu\text{m}$). The instantaneous field of view (IFOV) of Ikonos is providing the ground resolution for as about $4\times 4\text{ m}$. The images from the both time periods were geo-referenced and radiometrically calibrated. The geometrical correction was applied to the UTM 36N coordinate system and was co-registered image-to-map to coincide with the vector layer of the land-use classification. The radiometric correction was performed to minimize the errors associated with solar elevation, the four multispectral Ikonos bands (blue, green, red, and near infrared) were corrected to at-sensor radiation using the calibration parameters provided in the metadata (Figure 1).

To provide the land-use classes with additional characteristics, the normalized difference vegetation index (NDVI) was calculated. Using four Ikonos spectral bands and NDVI, the reflectance was calculated in relative units as band-wise average of pixels for every polygon. The per-class averages of reflectance are given in Table 1.

Table 1. Statistics of the existing land-use classification. The class means for the selected spectral bands are given in relative reflectance units.

Class code	polygons Class description	in a class		B1		B2		B3		B4		NDVI	
		Year 2002	Year 2010	Year 2002	Year 2010	Year 2002	Year 2010	Year 2002	Year 2010	Year 2002	Year 2010	Year 2002	Year 2010
400	Water bodies	15	13	32.3	45.4	33.4	34.3	24.7	21.8	35.5	11.5	-0.21	-0.33
600	Residential areas	113	86	37.6	61.9	41.4	55.8	34.2	48.8	57.3	57.4	0.11	0.09
608	Agricultural fields	156	120	33.8	53.9	35.8	46.8	27.4	37.9	56.3	60.4	0.24	0.19
609	Badlands	187	174	36.4	58.2	39.6	51.6	31.8	43	58.7	60.9	0.19	0.17
620	Natural forest	120	133	34.4	54.3	36.1	46.2	27.8	37.7	55.7	54.4	0.25	0.18
650	Impervious surface	56	50	43.1	65.5	49.9	59.9	43.8	54.1	62.7	55.6	-0.05	0.03
660	Plantations	67	51	33.9	52.03	35.7	43.5	27.6	33.8	55.6	59.5	0.34	0.27

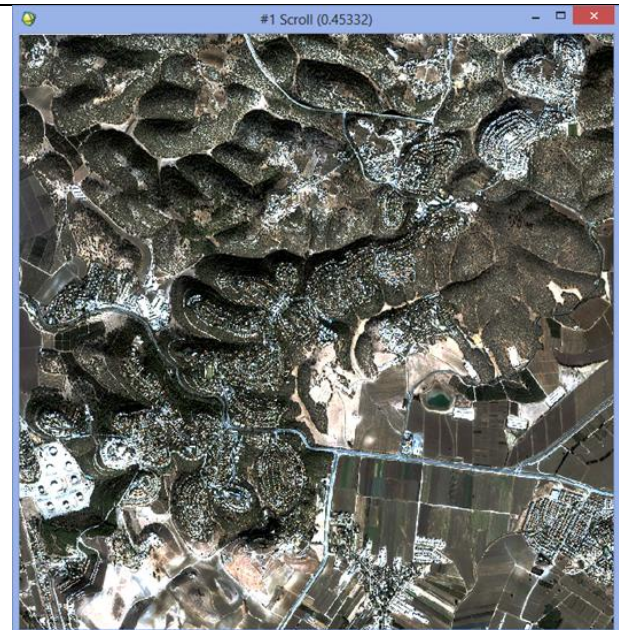
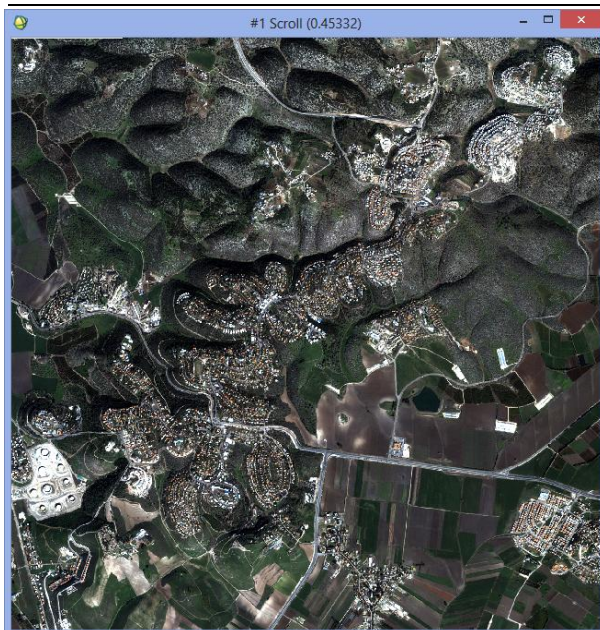


Figure 1. Ikonos images, acquired in the year 2002 (left) and 2010 (right)

3. METHODS

3.1. GIS-driven analysis of land-use

The heterogeneity of land-use usually results in high spectral variation within the same class in satellite imagery. The straightforward pixel-wise solely spectral-based methods (e.g. maximum-likelihood) cannot overcome the high spectral variation of intensity within the same class and retain its spatial distribution. In order to address land-use classification, one critical issue is utilization of information inherited in the existing land-use classification. The GIS-driven approach offers for satellite image analysis the spatial information about land-use that captured in the thematic classification. This may reduce the within-class spectral variation and improve the spatial proximity of classification results (Peled, 1994; Peled & Haj-Yehia, 1998). In order to provide the analysis of Ikonos time series with the meaningful image objects, we propose the utilization of land-use polygons from the existing land-use classification.

The underlying assumption is that the knowledge about land-use is that the majority of classification polygons is correctly represent the land-use reality of the image. In such a case the discrepant polygons have to be re-classified in respect to available land-use classes.

The following describes the steps involved in analysis of the time-series of Ikonos imagery.

3.2 Iterative discriminant analysis

The iterative discriminant analysis (IDA) implemented in Exelis ENVI/IDL 5 was applied to determine the spectral bands which separate on statistically significant level between the land-use classes. The spectral characteristics for the existing classification were captured for each polygon as average intensity values and then used in IDA processing. The detailed description of IDA algorithm implemented in signal recognition is given by Malmgrem (1997) and in bioinformatics by Mallios (1999). The method as adopted in the study generalizes standard linear discriminant analysis and attempts to use the spectral bands as independent variables to discriminate between the land-use classes through a series of iterations.

At the beginning of the IDA process, the discriminant functions are generated from polygons for which land-use classification is known when the classification existing in GIS database is accepted as initial classification. Then the functions applied iteratively to new classification cases with measurements on the same set of spectral variables. On each iteration step the combination of the statistically significant variables is determined and the classification is corrected as according to the posterior probabilities provided by current discriminant functions. The iterative process continues until no change is found between previous and current classification. This situation represents the best discrimination between the spectrally homogeneous land-use classes.

In order to define the spectral characteristics for the existing classification, the land-use classes were related to the reflectance. Then the stepwise linear discriminant analysis was used to determine the discriminant functions that distinguish the land-use classes on a statistically significant level (estimated with Wilk's lambda test). Discriminant function is a linear

combination of the discriminating variables (spectral bands) given for each spectral band in form of following equation:

$$F_{km} = u_1 X_{1km} + u_2 X_{2km} + \dots + u_n X_{nkm} \quad (1)$$

Where,

F_{km} = the score value for case m in land-use class k ;

X_{nkm} = the value on discriminant variable X_n for case m in land-use class k ; and

u_n = coefficients which produce the desired characteristics in the function.

The coefficients for the first discriminant function are derived so as to maximize the differences between the land-use class means. The coefficients for the second and following discriminant functions are derived to maximize the difference between the land-use class means, subject to the constraint that the values of the latter and former discriminant functions are not.

3.3 Detection of discrepancies

In our study the discrepancies are detected as according to the difference between the initial classification and the IDA classification, assuming that the best spectral discrimination between land-use classes represents the most correct land-use classification. Therefore, in naturally homogeneous land-use class, the amount of discrepant polygons is considered to be substantially less than the total number of polygons in the same class. The assumption was that the classes subjected to rapid land-use changes as "cultivated fields" or "residential areas" will comprise more discrepancies in comparing with classes of "water bodies" or "natural forests".

The resulting IDA classification was compared to the initial land-use classification in form of error matrix. The contingency matrix has summarized the distribution of polygons between the initial and the IDA classifications. The polygons re-classified to a different class were assigned as discrepant. Polygons that remained in the same land-use class were assigned as consistent. The accuracy of the detection was assessed in terms of false and true discrepancy/consistency using the validation dataset.

4. RESULTS

The stepwise discriminant analysis was applied for each Ikonos image in the time-series and allowed the selection of spectral bands for the classification. For the year 2002, the blue (B1), red (B3) and NDVI bands were found as the best combination that distinguish the land-use classes on statistically significant level ($p < 0.01$). Consequently, the IDA classification was processed using this combination of bands, resulting in final re-classification of polygons after 16 iterations. This is noticeable in Figure 1a that IDA was successful in reclassification of the polygons to spectrally homogeneous groups. Consequently, the within-class spectral variability of land-use classes has been reduced. As expected the least affected by IDA land-use class was the "water bodies". The classes of "natural forest" and "plantations" became clearly associated with the higher NDVI values in resulting IDA classification. Contrary, the classes of "impervious surface" and "residential areas" became more associated with large reflectance values and lower NDVI. For the year 2010, the red (B3), NIR (B4) and NDVI bands were found as the best combination that distinguish land-use

classes on a statistically significant level ($p < 0.01$). The final re-classification was accomplished after 18 iterations. Per-class distribution of land-use polygons in the domain of three Ikonos bands is demonstrated in Figure 1b. The IDA process was successful in re-classification of the polygons into spectrally homogenous groups that followed the reduction in within-class spectral variability of land-use classes. The least affected by the

IDA process was the “water bodies” class with the smallest number of reclassified polygons. Similarly to the year 2002, the classes of “natural forests” and “plantations” became positively associated with higher NDVI values whereas “impervious surface” and “residential areas” were found associated more with higher reflectance values and lower values of NDVI.

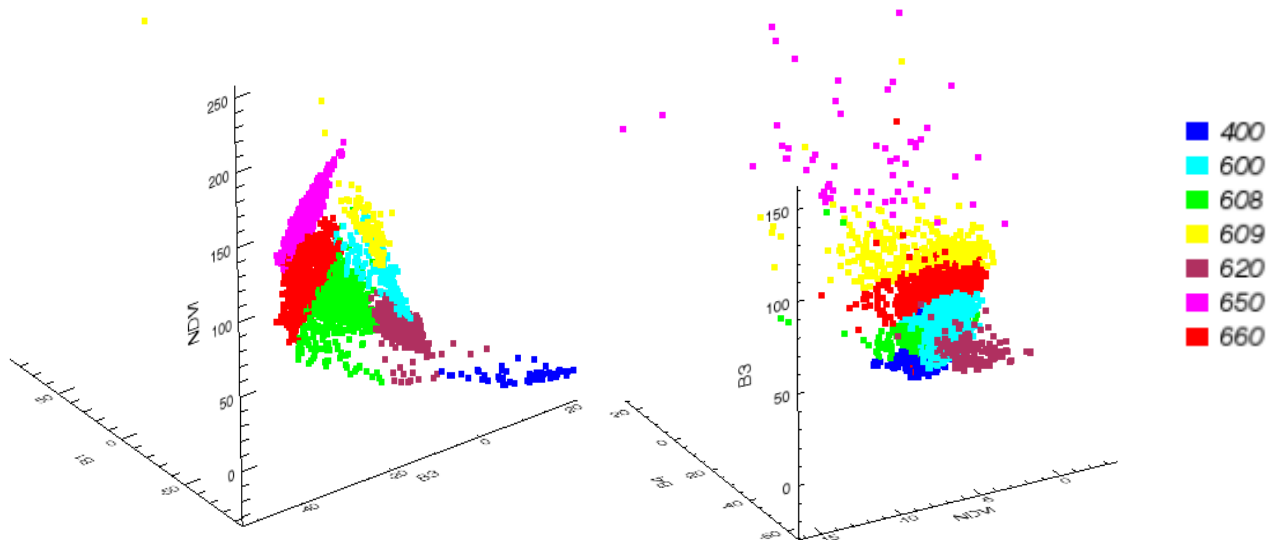


Figure 2. The final within-class distribution of land-use polygons in accordance to the Ikonos time-series. (Left). final IDA classification for the year 2002 and the year 2010 (Right). The class-codes are displayed according to the legend bar at the right hand side and correspond to the land-use class description given in Table 1.

The discrimination between the land-use classes was effected by using simple 3-linear parametric classification models. The classification models for the land-use classes that were calculated as a result of the IDA iteration process are given below in a form of non-intercept linear functions (Table 2).

Table 2. The classification models calculated by IDA whereas B1, B3, B4, and NDVI denote the best combination of bands for that discriminating the land-use classes (denoted by subscripts).

Year 2002	Year 2010
$C_{480} = B1 - 4.03 * B3 - 3.11 * NDVI$	$C_{480} = B3 - 0.61 * B4 - 19.63 * NDVI$
$C_{600} = B1 - .22 * B3 + 4.06 * NDVI$	$C_{600} = B3 + 0.82 * B4 - 7.42 * NDVI$
$C_{608} = B1 - .19 * B3 + 7.71 * NDVI$	$C_{608} = B3 - 0.49 * B4 + 22.69 * NDVI$
$C_{609} = B1 - .19 * B3 + 2.71 * NDVI$	$C_{609} = B3 + 0.81 * B4 - 31.60 * NDVI$
$C_{620} = B1 - 2.4 * B3 + 9.15 * NDVI$	$C_{620} = B3 - 0.82 * B4 + 19.28 * NDVI$
$C_{650} = B1 - 1.33 * B3 - 1.13 * NDVI$	$C_{650} = B3 + 0.84 * B4 - 32.07 * NDVI$
$C_{660} = B1 + 1.26 * B3 + 20.29 * NDVI$	$C_{660} = B3 - 0.36 * B4 + 6.22 * NDVI$

4.1 Detection of discrepancies

The resulting land-use classes of the IDA classification comprise the different number of polygons in comparison to the initial classification due to inter-class exchange. The contingency matrix was calculated to assess the inter-class difference between the initial and the IDA classifications (Table 3 and Table 4).

Table 3. Contingency matrix of the land-use classes obtained from original and IDA classifications for the year 2002. Note the total of the polygons that summarize the number of the detected consistent and discrepant polygons.

Table 4. Contingency matrix of the land-use classes obtained from original and IDA classifications for the year 2010. Note the total of the polygons that summarize the number of the detected consistent and discrepant polygons.

Table 3.

		IDA class							
		480	600	608	609	620	650	660	Discrepant
Land-use class	480	13	0	0	2	0	0	0	2
	600	0	61	0	31	0	21	0	52
	608	0	10	66	26	34	9	11	90
	609	0	25	22	49	29	49	13	138
	620	0	7	11	6	85	3	8	35
	650	0	10	0	10	0	36	0	20
	660	0	2	9	0	5	0	51	16
Consistent		13	61	66	49	85	36	51	Total: 714

Table 4.

		IDA class							
		480	600	608	609	620	650	660	Discrepant
Land-use class	480	11	0	2	0	0	0	0	2
	600	0	24	14	13	2	32	1	62
	608	0	3	52	43	4	17	1	68
	609	0	10	48	53	9	48	6	121
	620	0	23	45	33	17	12	3	116
	650	0	5	7	2	0	36	0	14
	660	0	3	8	17	3	4	16	35
Consistent		11	24	52	53	17	36	16	Total: 627

The "consistency" values here are calculated by dividing the number of the polygons classified by the IDA algorithm with the total number of polygons in a given class-type. For the year 2002, the fraction of land use polygons that were found consistent with the initial classification was estimated as 50.6% (361 polygons). The most consistent class was "water bodies" (86.6%) and "plantations" (76.1%). Similarly were found the most discrepant classes: "badlands" (73.7%) and "agricultural fields" (57.7%).

For the year 2010, the fraction of land use polygons that were found consistent with the initial classification was estimated as 33.3% (209 polygons). The most consistent class was "water bodies" (84.6%) and "impervious surface" (72.1%), estimated as relation of the number of consistent polygons to the total number of polygons in the corresponding class. Similarly were found the most discrepant classes: "residential areas" (72.1%)

and "natural forest" (87.2%). Thus, the majority of the polygons in these classes as well as in "agricultural fields", "badlands" and "plantations" was found spectrally discrepant by the IDA classification in comparison to the initial classification.

5. DISCUSSION

The GIS-driven approach has allowed for successful integration of geospatial data about land-use classes with the up-to-date Ikonos satellite data. The IDA process has resulted in the re-classification of the initial land-use polygons and which was assessed by the validated classification. The validation of the spectral-based discrepancy detection allowed for assessment of classification in terms of ground truth. The fraction of the polygons that were as consistent was similar to the fraction of polygons correctly detected as discrepant (50.6% vs. 33.3%, for the years 2002 and 2010 respectively). The suggested method

was found as performing well for the detection of discrepant and consistent classification. When interpreting the results, and in particular when studying the classification functions, it has to be noted that NDVI band gave a higher degree of explanation than any of the other bands alone. Therefore, the seasonal changes in vegetation cover may affect the discrepancy detection using spectral imagery acquired in different vegetative periods (dry summer vs. moist winter). This was the case with low consistency fraction in the image from 2010 that has been characterized by abnormally low precipitations that caused the changes in vegetation reflectance.

The distribution of the polygons of the initial class between different classes in the IDA classification indicates the within-class spectral heterogeneity that can be interpreted as discrepancy caused by the temporal changes in land-use (Table 2). The efficiency of the IDA for discrepancy detection also suggests its suitability for the quality assessment of newly built land-use databases.

It must be remembered that in the GIS-driven approach, the quality of the initial classification is important. Object-based thematic accuracy assessment only accounts for the accuracy of the class labels assigned to the image-derived segments and not how well the spatial characteristics of objects are represented (Stow et al., 2008). The spatial delineation of the polygons in GIS databases might be subjectively captured by the human interpreter following the semantic description, property lines, natural boundaries or management units. In fact, within our study area the polygons of the class “residential areas” contain various homogeneous features as buildings, roads, trees along the roads and so on. As another example the class of “crop fields” includes green and already harvested fields, characterized by substantially different spectral properties. Thus, the main limitation of the discriminant analysis as suggested here is the parametrical nature of the method that assumes unimodal Gaussian within-class likelihoods. If the within-classes distributions are significantly non-Gaussian, especially in smaller land-use classes, IDA will not be able to effectively preserve the data structure which is needed for correct classification. However, in such a case, IDA is still effective for detection of land-use polygons comprising delineation and misregistration errors.

To overcome the difficulty in classification of spectrally heterogeneous land-use classes, additional spectral and geometrical parameters might be included in discrepancy detection, as well as formalization of semantic definitions by means of evidential reasoning (Shoshany and Cohen, 2007). These and other ancillary characteristics might be easily integrated into the suggested method.

The main advantage of the proposed detection of discrepancies by iterative discriminant analysis is the analytical simplicity that allows for straightforward employment of additional data sources into classification process. Application of the presented method on large-scale or countrywide extent might be considered and further developed methodologically. 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 GIS-driven discrepancy detection suggests a possible way to maintain the integrity of land-use data and provides tools towards automatic revision/updates of land-use databases. Currently, a major obstacle for operational use of high resolution satellite data (e.g. Ikonos) for quality assessment of

land-use classification on national-wide level is the relatively high cost of the data. That difficulty may however be reduced with the planned Sentinel 2 satellites having a swath width of 290 km in combination with new satellites such as Landsat 8.

REFERENCES

- Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen, M., 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58, pp. 239–258.
- Mallios, R. R., 1999. Class II MHC quantitative binding motifs derived from a large molecular database with a versatile iterative stepwise discriminant analysis meta-algorithm, *Bioinformatics*, 15, pp. 432–439.
- Malmgrem, T., G. M., 1997. An iterative nonlinear discriminant analysis program: IDA 1.0. *Computer Physics Communications*, 106, pp. 230–236.
- Merchant, J. W., 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, pp. 257–268.
- Peled, A., 1994. Revision of Digital Maps and GIS Databases. *ISPRS, Proceedings Symposium, Mapping and Geographic Information Systems, Athens, Georgia, USA*, In: *ISPRS International Archives of Photogrammetry Vol. 30, Part 4*, pp. 268–272.
- Peled, A., Haj-Yehia, B., 1998. Toward automatic updating of the Israeli National GIS Phase II,
- D. Fritsch, M. English, M. Sester (Eds.), *International Archives of Photogrammetry & Remote Sensing, Volume XXXII, Part 4*, pp. 467–472.
- Wiedemann, C., 2003. External evaluation of road networks. In: *International Archives of Photogrammetry and Remote Sensing, Volume 34(3W/8)*, pp 93–98.
- Willrich, F., 2002. Quality control and updating of road data by GIS-driven road extraction from imagery. *Joint International Symposium on 'Geospatial Theory, Processing, and Applications'*, July 8-12 2002, Ottawa, Canada, In: *Proceedings of International Archives of Photogrammetry and Remote Sensing Vol. 34*, pp. 761–767.
- Shoshany, M., Cohen, Y., 2007. Inference versus evidence in reasoning remote sensing recognition, with an information foraging perspective, *International Journal of Remote Sensing*, 28:11, pp. 2613–2634.
- Stow, D., 2008. Monitoring shrubland habitat changes through object-based change identification with airborne multispectral imagery. *Remote Sensing of Environment*, 112, pp. 1051–1061.