GIS-DRIVEN ANALYSES OF REMOTELY SENSED DATA FOR QUALITY ASSESSMENT OF EXISTING LAND COVER CLASSIFICATION

Ammatzia Peled, Michael D. Gilichinsky
University of Haifa, Department of Geography, Haifa, 31905 Israel
peled@geo.haifa.ac.il, michaelg@geo.haifa.ac.il

Commission II, IC WG II/IV

KEY WORDS: GIS, Quality Assessment, Land Cover, Classification, Updating, Multispectral

ABSTRACT:

Automatization of processes for revision and updating existing GIS information is essential for the modern maintenance of spatial databases. The integration of remotely sensed multi-spectral data into the process of database revision is affected here by the implementation of GIS-driven analyses. The adoption of the GIS-driven principles, provide also an accurate geographical basis for a future supervised classification of the spectral data. The goal of the present research was to define and develop an automatic quality assessment method for the Land Cover classification layer of the Israeli National GIS database. During the experiments on multi-spectral remotely sensed data, effort was carried out in attempt to define "typical" spectral ranges as statistical maximum-likelihood criteria for the classification of each of the land cover phenomenon. These ranges were envisaged to characterize each of the land cover classification groups and to provide quantitative criteria for the definition of various groups of land cover type-classes. The definition of a typical-spectral-variance was executed on the basis of visual, multi-spectral and index bands of remotely sensed data. The decision whether existing GIS classification match the new image reality was made by statistical criteria of maximum likelihood for each investigated land cover type, according to the results of each and every spectral band. The study was based on multi-spectral data of the CASI airborne scanner and the space borne IKONOS data. Acquisition of the various types of information (spectral and spatial) was done according to the GIS-Driven approach of spectral data analyses that also permit the treatment of any spectral phenomena in a local coordinate system. The proposed method was developed on the basis of a study area (~50 sq. km.) and was tested on a larger control area. Image reality and field verification (both by land and air) proved the method of GIS-Driven quality assessment to be a promising solution for a revising process of large core spatial Data Bases.

1. INTRODUCTION

Spatial information revision and updating is the main concern and production effort of maintaining the ever growing GIS systems and spatial databases. Developing easily effected automatic updating and quality assessment methods of spatial information become the key to the successful maintenance of the large GIS databases established by many mapping agencies all over the world [Peled, 1993a,b]. Data acquisition for the Israeli National GIS database was based on photogrammetric mapping from 1:40,000-scale air photographs [Peled, 1994a]. The existing process of quality assessment of the National GIS database is executed by a number of sub-steps that include interpretation of remotely sensed information, field surveying, data completion, etc. The fidelity of geographic information systems (GIS) is dependent on accurate and up-to-date datasets. The manual revision of GIS data is manpower-, cost-, and time consuming [Peled, 1994b]. “Shortening the updating cycles of GIS databases is only possible if the manual part of the updating process can be at least partly automated” [Walter, 1999]. Existing GIS database can be used to provide image analysis algorithms with "a priori" information which is used (e.g.) to restrict the search space or impose constraints [Baltsavias and Hahn, 2000]. Having assumed that the number of wrongly captured GIS objects (or classification types) and the number of changes in the real world 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, 1998]. Thus, a significant part of modern process of updating and revision of existing GIS databases are performed manually by human operator through investigation of up-to-date remotely sensed data [Baltsavias and Hahn, 1999].

The objective of the GIS-Driven [Peled, 1993, 1994] approach was to build a correlation system (geographic and radiometric) between any given digital image with an existing GIS database. Color orthophotos may serve as a source of non-calibrated spectral information through a relatively narrow range of visual light. Unlike color orthophotos, multi-spectral data (e.g. IKONOS or CASI imagery) provide better and more detailed characterization of relevant ground phenomena. These data are more expensive and within relatively spatial-limited areas. On the other hand, the availability of the orthophoto data and the country-wide cover allows digital processing of the data for quality assessment of land use in National GIS database [Peled, 1998]. The principal issue here should be the adequate requirement for different types of spectral data in correlation with the different object phenomena. Thus, even limited in terms of their inherent spectral information the widely spread and accessible color aerial high-resolution information may be used as a principal spectral source for revision, updating and quality assessment of many land use types within GIS databases.

2. RATIONALE OF THE STUDY

Each ground feature (e.g. land use phenomenon) has the unique spectral signature which may be represented by the reflection magnitude and distribution of electromagnetic radiation that is acquired by the detectors of a remote sensing system, at the different spectral bands. Spectral classification is based on choosing training data to characterize the pattern of spectral signatures for each feature type. Traditionally, this operation requires field surveys for data acquisition or digitization of existing maps. Existing maps are usually out of date and need some revision. The objective of this study was to build the training data sets using the GIS-Driven classification methodology [Peled, 1994]. GIS-Driven algorithms enable automatic processing without the necessity of collecting ground data. Characterization ability of ground feature through remote sensing tools is one of the basic assumptions of present study effort. Accordingly, one may assume that different ground features have different spectral characteristics that may lead to their identification. Although some ambiguous semantic definitions of land use types may exist, they present much accurate geometric borders for every elementary part (single land use polygon labeled with a unique ID). Thus, it is possible to use existing geometric (spatial)

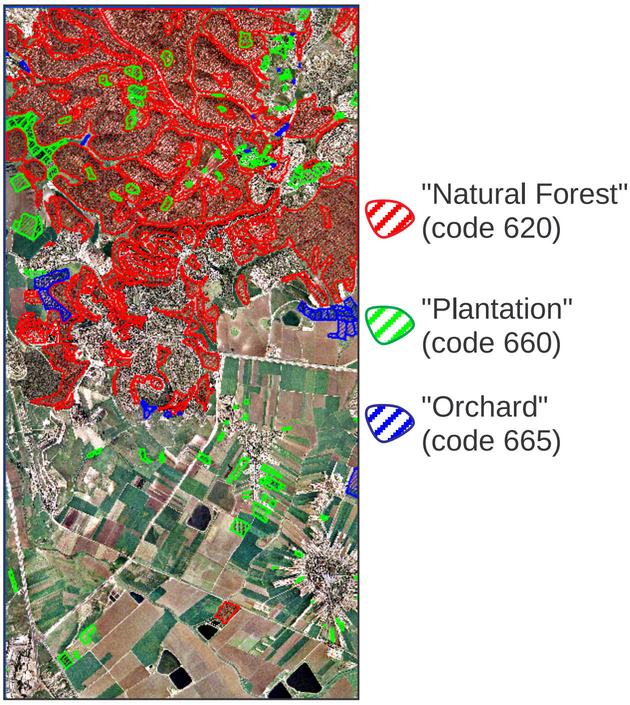


Figure 1. Spatial distribution of land use types within the study area.

definitions of land use features for GIS-driven collecting of spectral signatures. The existence of many features of the same type-group allows to collect a significant amount of spectral signatures for most of the various land use types (see Figure 1). This manifests with statistically significant unique signature.

The land use information stored in existing spatial databases is usually subjected to an intensive development and change processes that diminish their relevance. These changes impact the quality of stored thematic spatial information. The National Israeli GIS database consists of 10 thematic layers (including the land use and land cover). The original mapping and the digital photogrammetric updating are based on 1:40,000-scale air photographs. The massive amount of changes requires frequent updating and revision of the classification and of the geometric parameters of relevant features. Updating the spatial database by the traditional processes photogrammetric mapping is expensive and exhaustive. Automatization of the updating process, seemed to be the solution to the problem both in terms of standardization accuracy and quality and in terms of experts' labor and cost. In addition, the adoption of automatic and somewhat autonomous processes may be manifested by the introduction of additional multi-spectral data sources.

3. QC PROCESS STUDY

Traditional detection and estimation of changes in the land use or land cover is based on a human decision without a defined quantitative criterion. Therefore, one may detect discrepancies between the "world reality" and the "database reality" as these realities are strongly correlated to the expert perception. Required is a reliable process of quality assessment that will define the fidelity of the classifications in the national GIS database. The GIS-Driven method [Peled, 1994] was proposed as this process introduces a spectral based criteria for quality assessment operations for continuous maintenance of land use classification. Three data sources were used in order to evaluate the compatibility of the proposed method. These were: (1) multi-spectral IKONOS data; (2) hyper-spectral CASI data; and (3) Color orthophotos. Three well known vegetation indices: NDVI, SAVI and RVI were tested in the case of the CASI and IKONOS. Three bands of visual light (blue, green and red) were tested for the Color orthophoto. During the quality assessment process, each classification type of land use was characterized through a criterion of typical spectral variance (signature) that was determined according to a statistically significant amount of land use polygons. The QC process was executed for each GIS class-type, separately. All polygons were tested for their grey level population distribution within predefined ranges of 8, 16 or 32 grey levels. This was normalized by percentage to allow for variations in area. For each such range, a typical grey level average and thresholds were calculated. These values served to define the conformallity for each polygon per each band or index (see Fig. 2). These spectral signatures may be used also for the detection of land use polygons that were not defined in the existing GIS database, under the QC evaluation.

4. RESULTS AND DISCUSSION

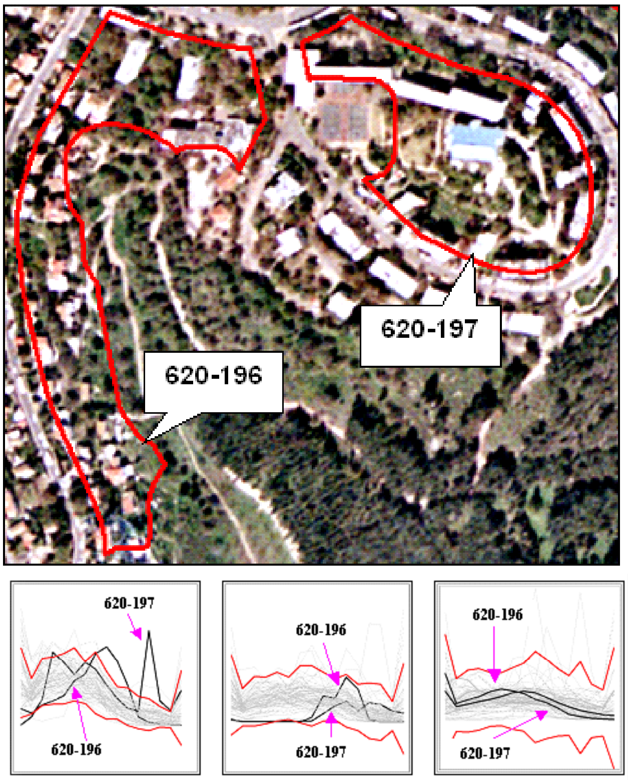
The implementation of the proposed method for detecting classification errors provided significant results for the investigated land use types and actually proved to be also a reclassification process. The quantitative parameters for determining of discrepant polygons (in terms of their GIS class-type) was based on a statistical likelihood of gray level values of the image pixels. The measure of statistical likelihood (defined by thresholds of standard deviation) is a criterion of structural homogeneity for the investigated land use types. Structural homogeneity of land use type is a decisive criterion for successful decision making and quantifying mismatches within the evaluated GIS database. In addition there is the aspect of some equivocal semantic descriptions of land use types in existing GIS database. This does not allow for a common (universal) spectral basis for defining a particular group or sub-set of a certain land use type. In such cases (e.g. “playgrounds”) many of the investigated polygons were found as not matching the typical spectral signature of their class-type. Yet, within the discrepant polygons, new spectral commonality was found. This proved that the land cover type could be detected and automatically classified but not the land use which was defined semantically.

Results of the QC process for three land cover class-types are partly presented in Table 1. The proposed method was executed based on the “image reality” of the three spectral data sources mentioned above: Color orthophoto, CASI and IKONOS. The results are based on automatically collected spectral signatures (for each source, separately). It shows the percentage of polygons that were found consistent or discrepant relatively to their GIS class-type. Note the significant amount of polygons that were found to be inconsistent. This high value was mainly due to the rapid changes and development processes that the study area went through since the time of the original data acquisition for the GIS database. This high level of change is an additional incentive to develop the proposed QC process as a tool to expedite the updating cycle. In some cases were found also errors inherent in the original classification (quite rare) and in original errors of mapping the polygons.

Table 1. Consistency results of existing land use classification within the GIS database

type Code	land use	relation to existing GIS	Color Orthophoto	CASI	IKONOS
620	natural forest	consistent	51 (61.4%)	28 (75.7%)	65 (78.3%)
		discrepant	32 (38.6%)	9 (24.3%)	18 (21.7%)
660	plantation	consistent	35 (61.4%)	20 (66.7%)	45 (78.9%)
		Discrepant	22 (38.6%)	10 (33.3%)	12 (21.1%)
665	orchard	consistent	12 (52.2%)	6 (75.0%)	18 (78.3%)
		discrepant	11 (47.8%)	2 (25.0%)	5 (21.7%)

The following examples (see Figures 2,3 and 4) show polygons that were found to be discrepant, overlaid with an image obtained from the three spectral data sources. Figure 2 includes also a graphic visualization of the proposed GIS-Driven spectral signature acquisition and QC process.



Blue Band Green Band Red Band

Figure 2. Results for two discrepant polygons of “natural forest” and their spectral signatures for the RGB bands of a Color Orthophoto.

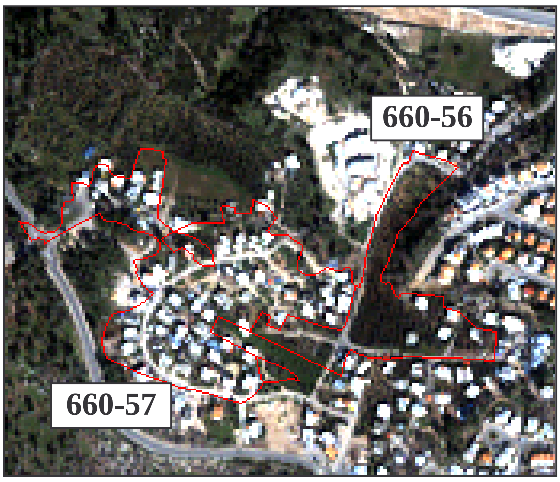


Figure 3. Discrepant polygons of “orchard” (Spectral data source: IKONOS)

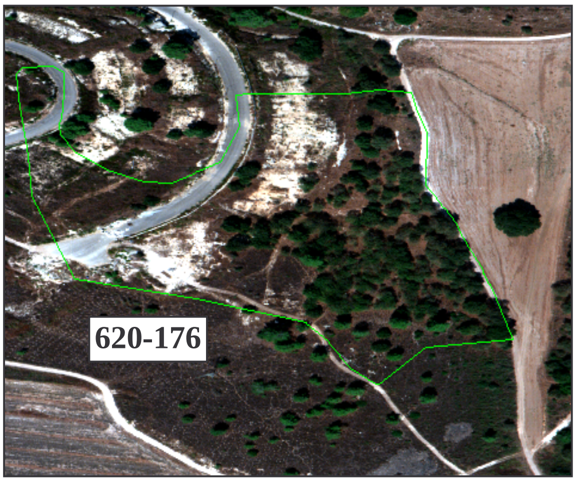


Figure 4. Discrepant polygons of “natural forest” (Spectral data source: CASI)

In order to evaluate the validity of the proposed quality assessment method, two sets of quality control were designed. All the land use polygons were tested visually, as per their virtual (spectral) reality or by field trips both on land or by a reconnaissance flying over the test area. By this QC step, all the automatic definitions related to consistency or inconsistency with the original GIS data base classifications, made by the newly developed system, were checked. These results, for the same three land use class-types mentioned above, are presented in Table 2.

Table 2. Evaluation of the proposed automatic land use QC method

type code	land use	Color Orthophoto		CASI		IKONOS	
		discrepant	consistent	discrepant	consistent	discrepant	consistent
620	natural forest	0.92	0.93	0.67	0.86	0.89	0.94
660	plantation	0.63	0.89	0.70	0.75	0.83	0.87
665	orchard	0.60	0.83	1.0	0.83	1.0	0.89

The first observation, viewing table 2 and additional such results, is that the consistency definition is of higher level of accuracy than the discrepancy definition. This is due to the motivation to minimize “false-true” cases. That is, one would rather re-check polygons that were defined as inconsistent rather than to accept polygons that were labeled as “consistent” were in reality, they are not. Also, one may notice that the “natural forest” class was better defined spectrally in general and of course relatively to the somewhat less accurate the “plantation” class-type. This is of course to point out that plantation fields are less homogeneous and are a meta-class were the specific type of plants should be classified in order to achieve better separation and higher fidelity. Another obvious observation is that the multi-spectral sources (CASI and IKONOS) served better to achieve high fidelity in labeling discrepant polygons. This was quite true for the less homogeneous classes. In the case of highly homogeneous classes, such as “natural forest”, the basic RGB bands were sufficient and even produced better fidelity. In all tested land use phenomena, the considerable amount of discrepancies apparently indicates the presence of sub-classes throughout of existing land use classification. Possible factors for wrong discrepancies recognition using the orthophoto is the excessive generalization of spectral phenomena due to un-calibrated radiometric information. The outstanding reduced correlation to image reality for discrepant polygons of "plantation" and "orchard" (see Table 2) was probably caused due to "coarse" image normalization to a default gray level scale. Thus, refinement of this normalization parameter might improve the accuracy and fidelity in the recognition pf discrepant polygons.

5. SUMMARY AND CONCLUSIONS

Statistical analyses of spectral information of digital color orthophotos are quite limited by three un-calibrated spectral bands of visual light. The absence of useful RGB-based spectral indices leads to extended use of raw data. In fact, better-calibrated multi-spectral data of IKONOS and CASI enabled to obtain detailed characterization of the studied land use patterns. The absence of country-wide multi-spectral information, motivated the implementation of the "GIS-Driven" method. The proposed system is based on statistical interpretation of spectral signatures of the investigated land use features within the red, green and blue bands. This provides an initial tool for defining the consistency or discrepancy of all type-class polygons within the analyzed GIS database. The statistical definition of typical spectral variance is based on a considerable amount of samples, for each class-type, separately. Due to the diverse nature of the studied land use phenomena, the statistical conditioning provides an approximate characterization of structural pattern for most land cover class-types. Therefore, the main component of the successful discrepancy recognition by the proposed method is the structural homogeneity of the investigated land use class-type based on statistical similarity of the pixels population of the investigated polygons. Additional aspect of the successful discrepancy recognition are the initial definitions of phenomena in the National GIS database that are frequently based on semantic human perception of the land use. The significant success of the proposed method for all of the investigated phenomena is reinforcing the proposed GIS-Driven approach to land use characterization even by using color orthophotos. According to the study results, the uncalibrated RGB data of color orthophotos achieved reliable recognition of discrepancies between existing GIS database and the image reality. Yet, the question of the exact definition of the discrepancy factor for each specific land use phenomenon is still open. Implementing the proposed method, using multi-spectral data, provided an opportunity for advanced processing (e.g. indexing, radiometric transformation, etc). This improved the results of the proposed method, as tool for fast quality assessment of existing GIS databases.

6. REFERENCES

Baltsavias, E., and Hahn, M., 1999. Integration of image analysis and GIS. *International Archives of Photogrammetry and Remote Sensing*, Vol. 32, Part 7-4-3 W6, 3-4 June, Valladolid, Spain.

Baltsavias, E., and Hahn, M., 2000, Integrating spatial information and image analysis - One Plus One Makes Ten. *19th ISPRS Congress, The International Archives of Photogrammetry and Remote Sensing*, Vol. III, Part B2, July 2000, Amsterdam , pp. 63-74.

Peled, A., 1993a. Change Detection: First Step toward Automatic Updating, *ACSM-ASPRS Annual Convention & Exposition Technical Papers*, USA, Vol. 3, pp. 281- 286.

Peled, A., 1993b. Remote Sensing in Israel - From Change Detection to GIS Generation. *International Symposium on Operationalization of Remote Sensing*, Enschede, the Netherlands, 19-23 April 1993. Vol. 6 - Remote Sensing and GeoInformatics, pp. 117-126.

Peled, A., 1994a. Remapping a Country - The Israel National GIS. *Conf. Proceedings, ACSM/ASPRS 1994 Annual Convention*, Reno, April 1994, Vol. 1 (ASPRS), NV. USA, pp. 245-249.

Peled, A., 1994b. Revision of Digital Maps and GIS Databases. *ISPRS Proceedings Symposium, Mapping and Geographic Information Systems*, Vol. 30, Part 4, Athens, Georgia, USA, pp. 268-272.

Peled, A., and Haj-Yehia, B., 1998. Toward automatic updating of the Israeli National GIS – Phase II. *The International Archives of Photogrammetry and Remote Sensing*, Vol. II, Part 4, ISPRS Commission IV – GIS Between Visions and Applications, Stuttgart, Germany, pp. 467–472.

Walter, V., 1998. Automatic classification of remote sensing data for GIS database revision. *The International Archives of Photogrammetry and Remote Sensing (ISPRS)*, Vol. 32, Part 4, Stuttgart, Germany, pp. 641 – 648.

Walter, V., 1999. Automated GIS data collection and update. In: *Photogrammetric Week `99*, D. Fritsch and R.H. Spiller, eds., Wichmann, Heidelberg, pp. 267–280.