

## SEMI AUTOMATED LAND COVER LAYER UPDATING PROCESS UTILIZING SPECTRAL ANALYSIS AND GIS DATA FUSION

L. Cohen, E. Keinan, M. Yaniv, Y. Tal, Y. A. Felus, R. Regev

Survey of Israel, (liorc, eran, moshey, tal, felus, ronenr)@mapi.gov.il

Commission III, WG III/7

**KEY WORDS:** NTDB (National Topographic Data Base), Segmentation, Classification, nDSM, data fusion

### ABSTRACT:

Technological improvements made in recent years of mass data gathering and analyzing, influenced the traditional methods of updating and forming of the national topographic database. It has brought a significant increase in the number of use cases and detailed geo information demands. Processes which its purpose is to alternate traditional data collection methods developed in many National Mapping and Cadaster Agencies. There has been significant progress in semi-automated methodologies aiming to facilitate updating of a topographic national geodatabase. Implementation of those is expected to allow a considerable reduction of updating costs and operation times. Our previous activity has focused on building automatic extraction (Keinan, Zilberstein et al, 2015). Before semi-automatic updating method, it was common that interpreter identification has to be as detailed as possible to hold most reliable database eventually. When using semi-automatic updating methodologies, the ability to insert human insights based knowledge is limited. Therefore, our motivations were to reduce the created gap by allowing end-users to add their data inputs to the basic geometric database. In this article, we will present a simple Land cover database updating method which combines insights extracted from the analyzed image, and a given spatial data of vector layers. The main stages of the advanced practice are multispectral image segmentation and supervised classification together with given vector data geometric fusion while maintaining the principle of low shape editorial work to be done. All coding was done utilizing open source software components.

### 1. INTRODUCTION

#### 1.1 Motivation

Today Spatial DBs are the most common way of organizing and presenting insights in governance agencies. Those agencies mostly rely on external mapping layers such as building and roads layers, to facilitate managing activities and planning. The ability to acquire high-resolution satellite and areal images has made it possible to increase spatial data gathering demands and specification to support needs of different government offices.

The basic spatial data which aimed to support varied general needs include the following layers: built-up areas, roads, vegetated areas, agriculture fields, open areas (soil types) and different water bodies. An accurate and up-to-date description of those layers in national scale will promote managing abilities in different scales of all range, from national ministry to regional council.

In this work we have defined the hierarchical structure of spatial entities which can be described as Image classification and segmentation combined with external spatial data sources; (Those entities were chosen based on a previous LCLU standard, which defined the work of the interpreter. The hierarchical structure which derived from the LCLU standard is described in figure 1 chapter 1.2 and contain different layers which can be extracted and together consist a land cover updated layer.

#### 1.2 Background

Land cover layer purpose is to be the fundamental data source of usages in many fields, including Land management, climate, hydrology, agriculture, health and many more. Detection of large-scale trends has significant effects on governance activity and decision making.

Among the government ministries and agencies which use and rely on the land cover national layer there can be found:

1. Central statistics agency.
2. Ministry of agriculture.
3. National Planning agency.
4. 'Keren Kayemet Le'Israel' (national forestry organization).
5. Ministry and units of public security.
6. Ministry of national infrastructures.

All are considered costumers of the Survey of Israel spatial layers and expect the national scales layers. Since the variety of needs of customer mentioned above are spread on many aspects as explained in chapter 1.1 and to keep high relevance of the product, we have defined the output layer entities to match the basic needs of as many costumers possible. One of the main principals in our methodology was the ability also to receive data from each end-user, and in the end to be able to present more diverse and new data types which could not form in a way other than spatial integration.

In order to form the land cover layer we have defined the following data structure to be obtained, based on GIS and remote sensing principals. In figure 1a-1d are presented the defined end-members, according to detailing levels:

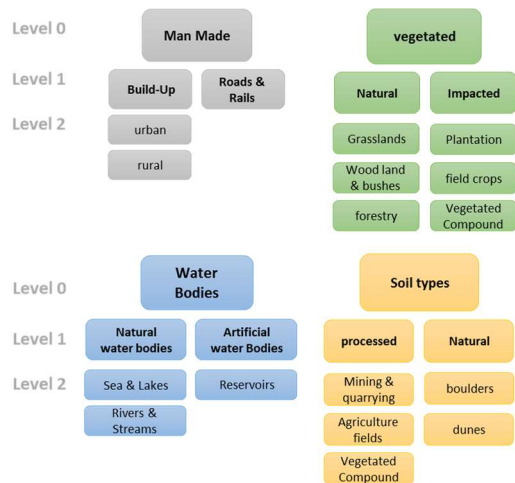


Figure 1. Land cover Layer classes and entities types:

The data structure is divided hierarchically to four classes:

1. Impervious: the land cover will contain residence and non-residence built-up areas, and roads and rails. A significant source of impervious data is the NTDB which holds the roads, rails and complexes layers.
2. Vegetated Areas: our approach was to distinguish between natural and non-natural vegetated areas, when natural vegetated areas are divided according to height, and non-natural vegetated areas can be acquired as agricultural or man-made compounds.
3. Water Bodies: same as in vegetated areas, we were guided by natural or artificial object types: sea and lakes, pools and reservoirs.
4. Soil types: in this category there is lots of interpretation when not involving external data sources. Therefore there remained the option of classifying natural soils as type 'other'.

The data structure described is in general not similar to common land cover layer standards. It may be seen that this structure contains more entities than needed, but we have mapped our needs (see above paragraph in this chapter) and matched it to the ability to extract the required data from the image automatically, the ability to combine external entirely reliable data sources. The hierarchical structure allows data to be still correctly classified even if data is missing. Besides, this data structure allowed to set priority for external data fusion, as will be displayed in next chapters.

As mentioned, in comparison to other national land cover layers we hold much-specified land cover layer. In addition to the level of details, we chose to deal with shape aspects and topology by not forcing squared objects. Meaning, there has been put particular noticing to boundaries of each data portion, and much effort was put on correctly segmenting image objects.

### 1.3 Case Studies

The usage of external national-scale confirmed data layers, and the need to apply one methodology on a variety of landscapes, and images describing them, regardless their primal conditions (areal or satellite image, number of spectral bands, spatial resolution etc.) has driven us to select very diverse case-studies

and aim for a very efficient and robust process. The selected areas are presented in table 1:

Area	Source	Spatial Resolution	No. of Bands
Metula	Pleiades Satellite	0.5 GSD	4 RGB NIR
Tirat ha'Carmel	Z/I DMC II	0.5 GSD	4 RGB NIR
Netanya	Z/I DMC II	0.5 GSD	4 RGB NIR
Center	Pleiades Satellite	0.5 GSD	4 RGB NIR
Sderot	Pleiades Satellite	0.5 GSD	4 RGB NIR
Hura	Z/I DMC II	0.5 GSD	4 RGB NIR

Table 1. Case studies areas and data

Areas of images tested covers between 25 km<sup>2</sup> in aerial images and 100 km<sup>2</sup> in satellite images. Each of the images used was Orthophoto prepared in advance. For each image there has been made a set of supporting layers which hold the relevant data: buildings and complexes, roads, streams, agricultural parcels, forests, water bodies, quarries.

## 2. METHODOLOGY

As mentioned in the previous chapter, our methodology of producing and updating the national land cover layer includes a combination of detailed data and image analysis. The main stages presented in figure 2:

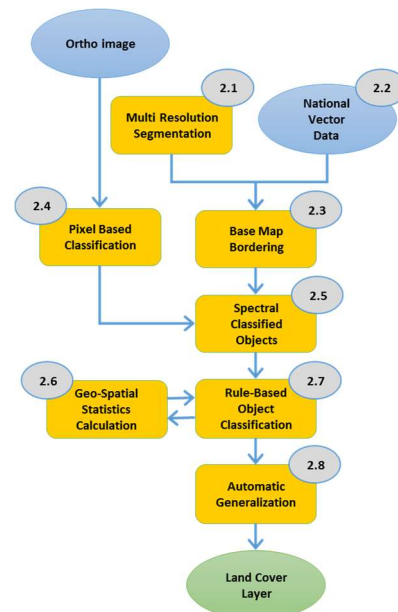


Figure 2. Land Cover layer production main stages

In this section we will introduce milestones and required processing of input data. For each stage will be given a short explanation of implementation with a possible example, if relevant. In order to start the layer processing we require: matching spatial resolutions between all raster inputs (orthophoto, nDSM) and Presence and correctness of external vector data.

## 2.1 Segmentation

Drawing the borders between the different cover types is one of the most significant tasks in the making of the land cover database. It is by far the most time-consuming one. Dividing the image into different regions concerning the various cover types is the skeleton of which the whole project will be built upon.

Segmentation intends to cluster individual pixels into homogenous regions. The usage of segmentation in suggested methodology is to create a coarse spatial division based on spectral data. We apply eCognition 'multiresolution segmentation' algorithm to achieve non-diverse spectral population and controlled segment size. Testing on other algorithms including watershed, mean-shift and some less known techniques did not come up with good segmentation results.

The 'multiresolution segmentation' then integrated with vector data to combine a fabric made of spectral based region merge with premade NTDB vector data regions (like roads, buildings and so).

## 2.2 National Vector Data

In order to allow proper usage of NTDB entities, there was a need to know in advance which features are most significant to the Land cover final layer. Then set those inputs to override segments created in the previous stage. Another benefit of this fusion is the usage of already existing validated data. To combine the data sources correctly, we have defined the hierarchy of integration layers. The primary purpose was the prevention of topology errors. For example, the roads and rails must override agriculture parcels.

We paid attention to variance in layers level of detailing, source (reliability) and spatial meaning. Furthermore, we have chosen out of each layer the meaningful geometrical data to be integrated. For example, we did not integrate all road types, but only Highways and intercity roads. Eventually, we have combined following layer ordering, as featured in figure 3:

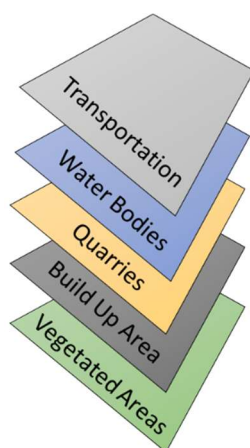


Figure 3. Set of layers to be fused. Upper layer overrides lower layers.

## 2.3 Base Map Bordering

Our methodology aimed to define to geometrical background of to be produced layer as solid as possible. Performing a segmentation process, fused with product of vector data, produces satisfying foundation for next stages application. Result of segmentation and hierarchical vector layers fusion is presented in figure 4:



Figure 4. Base map examples resulted of segmentation and vector data fusion (Tirat ha'Carmel area).

## 2.4 Classification

Supervised classification is an evolving field which efficiency and usefulness are increasing steadily. In our work, we have adopted classification libraries from open source image processing and machine learning framework. This tools provide a supervised pixel-wise classification chain from multiple images.

The classification stage was the most challenging. We had to upgrade our hardware to process a significant amount of data. We are looking all the time for new calcification algorithm and methods to improve the results and reduce the computer costs. We joined research with the 'Earth and Planetary Image Facility' ('Ben-Gurion' University) to create a model for classification and discrimination of objects resides in a geographic scene.

### 2.4.1 Preparing training data

Most of the classification success is based on high-quality training samples. In order to operate on a national scale, there is a need to collect massive data set. Samples which will represent the variety of physical image attributes in different areas of interest. Forming a national scaled training data samples is a major task, obligating to collect the samples, and monitor the samples database afterward. To collect the samples, we developed an open-source based method, gridding input image to 20X20 pixel polygons which hold statistics of common cover types. The entities representing the diversity of image contents were chosen because of its significant amount of appearance forms and variety, and because of they are expected to be found in almost all landscapes and areas. Training data had been derived from sourced layers. The selected cells have been analyzed and found corresponding to matching prior requirements declared for each basic category.

### 2.4.2 Training

The training process contains several classifiers. We value the results with a testing set (made up from 30% of the training set). To increase the reliability, we calculate the confusing matrices of every classifier and merging the results into one output, through the Dempster-Schiefer voting algorithm. Another significant step at this point is reducing heterogeneity of pixel level classification by voting based on pixel surrounding. Afterward and if available, we differentiate between on ground and 'lifted' pixels with nDSM map (normalized Digital Surface Model). The nDSM is calculated by the formula:

$$nDSM = DSM - DTM \quad (1)$$

Figure 5 present the pixel based classification result:



Figure 5. classification

### 2.5 Spectral classified objects

In this stage, we transform examination of data received from prior stages from pixel-based data into object-based. For each segment, we count the number of pixels classified to each main branch at level 0 (see figure 1).

### 2.6 Geo-Spatial Statistics calculation

In this stage, we first define the usage of the vector NTDB layers. For example, we would like to distinguish urban areas and village areas. In early stages, we understood that spatial building characteristics of villages could be seen very similar in cities and suburbs. The layers in use were selected due to their contribution to required final data structure and level of detailing (see figure 1). As in stage 2.2, we filter all non-contributing subtypes. Layers contributing to data structure are presented in table 2.

Layer Name	Source	Data Type
Roads	Survey of Israel	Polyline
Railways	Survey of Israel	Polyline
settlements	Survey of Israel	Polygon
Buildings & Complexes	Survey of Israel	Polygon
Agricultural Parcels	Ministry of Agriculture	Polygon
Quarries	Ministry of Energy	Polygon
Forest and Grove Areas	MAARAG (National Ecosystem Assessment Program)	Polygon

Table 2: Contributing layers

Every segment will hold the usage and the overlapping area of intersection with the corresponding NTDB layer.

In the next step we count for each segment the amount of classified pixels within its borders, and use image spatial resolution to calculate area covered, and cover percentage.

We suggest smart use of prior knowledge. For each type of vector data we can build the decision rule supporting products demands.

### 2.7 Rule-Based objects classification

As explained in previous stages, we based our methodology on an efficient combination of:

1. Accurate description pixel based supervised classification result.
2. Validated vector data on a national scale.

After examination of spectral classification results and vector layers in previous stages, which will be used to classify each of the segments, we have conducted a list of decision rules, based on two parameters:

1. Class's population of each main category, as described in stage 2.4.
2. Covered area and relative area by each vector layer.

The rules-based classification is a decision tree taking relevant attributes from each segment statistics. A classified segment is recognized by one of the destination classes (see figure 1).

Since each segment is tested against every rule, there was a need to prioritize rules aimed for classes representing some categories over others. Those were determined according to spatial assumptions. For example, we would classify a segment as 'urban area' when there is a relatively small amount of urban classified pixels in it even though the majority of pixels were classified as trees. In this case, prioritizing the built-up area categories allows spatial continuity of built-up areas that fits overall examination scale of a resulted layer.

### 2.8 Automatic Generalization

After matching class to each object and producing the land cover layer, there was a need to manage overall visualization. It was done using 'simplify polygon' tool, which reduced nodes while keeping the overall shape. The generalization process was followed by neighbor-border indexing of same-classified objects (eCognition environment). Definition of such spatial rule has dramatically reduced the total number of segments and approved overall visualization.

## 3. RESULTS

Our process is resulted in a continuous object vector layer, therefore, the accuracy assessment will be based on extraction of all contained pixels within selected objects bounds. The objects tested are used as a representative validation set, and were chosen randomly. For each class we extracted the amount of polygons to check according to standardization presented in ISO documents. In figures 8, 9 and in table 3 are presented accuracy & recall results.

### 3.1 Tirat haCarmel test area



Figure 6(a) orthophoto of Tirat ha'Carmel



Figure 6(b) 2014 official national database Land Cover layer



Figure 6(c) 2015 semi-automated Land Cover layer

### 3.2 Gush-Dan test area



Figure 7(a) Pleiades Orthophoto of Gush-Dan

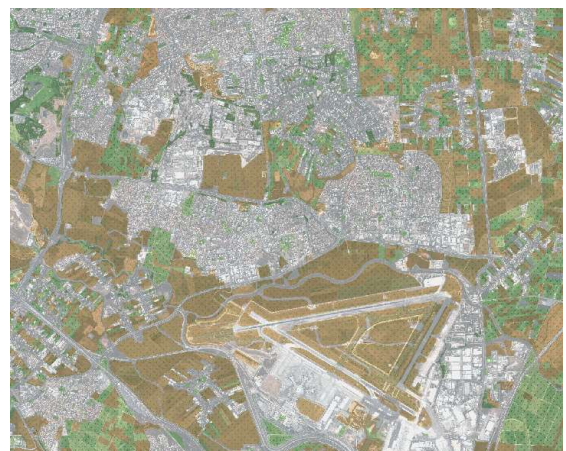


Figure 7(b) 2014 official national database Land Cover layer.



Figure 7(c) 2015 semi-automated Land Cover layer

### 3.3 Results Analysis

The values obtained in table 3 show that presented methodology is well defined, and manages dealing with relatively detailed data demanding while maintaining satisfying classification results.

		Results		
		User Accuracy % (Recall)	Overall Producer % Accuracy	Kappa %
AREA	Natanya	93	86.2	91
	Sderot	88	88.3	86
	Gush-Dan	89	91.4	87
	Hura	90	86.9	88
	Metula	94	93.1	90
	Tirat Ha'Carmel	92	92.8	95

Table (3) – precision, recall accuracy and Kappa values of tested areas

Accuracy and precision achieved in this paper indicate high sustainability of results to differences in traditional spectral image classification outputs. The use of inserted vector data allows applying hierarchical classification rules, contributing to overall quality of land cover layer.

Another big advantage in our approach is derived from referring to all pixels cluster within a segment to be classified homogeneously. Since the final classification is made on objects level, no noises or dissimilarity occurred, contributing to product's smoothness and future usage effectivity.

One of main testing aspects was usage of digital surface models as one of the main contributors to increasing accuracy and precision rates. The results shown in table 3 indicate that our assumption was incorrect, allowing high success rates expectations while not obligating having 3D data (which is very hard to process and hold in high resolution and large scales).

### SUMMERY

The methodology presented in this article is based on conceptual novel approach, which classifies objects, created by mixing similar and unrecognized spectral data clusters with geometrically founded vector data. Result is characterized with short runtime, low cost open-source based coding and high accuracy and precision rates.

The ability to maintain high detailing shown in this paper supports the assumption of creation of map to be basis to other spatial knowledge fields to be implemented in outcome of presented process. Distribution of well-defined comprehensive, standardized and useful GIS layer, will promote spatial based insights in the fields of housing, urban planning, education, wellness and more.

The overall 90% accuracy and precision rates based on usage of vector data allows multiple updating possibilities, which will reduce the need in frequent data bases update needs.

### ACKNOWLEDGEMENTS

The authors would like to thank Mr. Moshe Yaniv of Taldor ltd. and Mrs. Natali Trachtenberg of Opisoft ltd. Mr. Ishay Balash of Opisoft ltd. for their suggestions and useful implementations.

The authors would like to thank CNES® for developing and maintain the Orfeo toolbox framework.

### REFERENCES

- Keinan, E., Felus, Y. A., Zilberstien, O., Trachtenberg, N., 2015, Advanced mapping process for automatic updating National Topographic Data Base based on change detection. Israeli Land Surveyors Annual conference, March 2015.
- Baltsavias, E.P., 2004, Object extraction and revision by image analysis using existing geodata and knowledge: state-of-the-art and steps towards operational systems. ISPRS Journal of Photogrammetry & Remote Sensing 58, pp. 129 – 151.
- Airbus®. Pléiades Imagery User Guide. October 2012.
- Emilio D., Clément M.2014. Change Detection in High-Resolution Land Use/Land Cover Geodatabases (at Object Level). EuroSDR Official Publication No 64, April 2014.
- Alon Tairy, Shimrit Maman-Tirosh, Stanley R. Rodman, Dan G. Blumberg, in prep. Large Scale Machine Learning Methods Applied to Object Recognition in Multichannel and Multisensor Geospatial Imaging. Acceleration by Massively Parallel Processing (CPU + GPU).