LAND USE/LAND COVER CLASSIFICATION OF THE TEYNA RIVER BASIN USING AUTOMATED FEATURE EXTRACTION (AFE) ALGORITHMS

Lachezar Filchev

Space and Solar-Terrestrial Research Institute - Bulgarian Academy of Sciences,
e-mail: lachezarhf@gmail.com

Key words*: Automated Feature Extraction, remote sensing, GIS, land use classification, accuracy assessment*

Abstract*: Land use information for regional and local scale studies is indispensible in policy- and decision-making on environment protection. During the recent decades, based on the growing input of remote sensing data, several land-use classification schemes have been created. Concurrently, numerous image classifications for digital imagery have been developed to provide fast and accurate results for land use. The Automated Feature Extraction (AFE) Algorithms are new approaches to image classification taking into account images' texture and other ancillary imagery data. The present study represents an example of AFE application to land-use/land-cover classification of a very high resolution satellite image on the territory of the Teyna river basin. The data used is a QuickBird satellite image, high resolution Digital Elevation Model (DEM), image texture and vector layers, such as hydrological network and the Teyna river basin boundary. The supervised AFE classification of land use/land cover was performed in Feature Analyst 4.2 (shareware license). The resulting 9 land use classes were assessed for accuracy achieving 93.52% overall accuracy. Among the best thematically discriminated classes are those of natural vegetation and infrastructure with almost 100% accuracy, whereas the grasslands, meadows and bare ground tend to be mixed with one another, due to the spectral inseparability and similar texture of the classes.*

КЛАСИФИКАЦИЯ НА ЗЕМЕПОЛЗВАНЕТО НА ВОДОСБОРНИЯ БАСЕЙН НА РЕКА ТАЙНА С ИЗПОЛЗВАНЕТО НА АВТОМАТИЗИРАНИ АЛГОРИТМИ ЗА РАЗПОЗНАВАНЕ НА ОБЕКТИ

Лъчезар Филчев

Институт за космически и слънчево-земни изследвания - Българска академия на науките
e-mail: lachezarhf@gmail.com

Ключови думи*: автоматизирани алгоритми за разпознаване на обекти, дистанционни изследвания, ГИС, класификация на земеползването, оценка на точността*

Резюме*: Информацията за земеползването и земното покритие на регионално и местно ниво е незаменима при вземането на решения свързани с опазването на околната среда. През последните десетилетия на основата на нарастващият обем от входни данни от дистанционните изследвания бяха създадени няколко класификационни схеми на земеползването и земното покритие. Едновремнно с това са създадени многожество класификации на цифрови изображения, за да бъдат достъпни точни и навременни резултати за земното покритие и земеползванието. Автоматизираните алгоритми за разпознаване на обекти са един нов подход за класификация на изображения използващ текстурата на изображението и допълнителна входна информация. Настоящото изследване представя пример на приложение на тези алгоритми за класификация на земеползването и земното покритие на територията на водосборния басейн на р. Тайна с използване на данни със свръхвисока пространствена разделителна способност. Използваните данни са: сателитно изображение на QuickBird, цифров модел на релефа с висока разделителна способност, текстура на изображението и векторни слоеве - хидроложка мрежа и граница на водосбора на р. Тайна. Контролираната класификация на земеползването и земното покритие е направена във Feature Analyst 4.2. Получените 9 класа за земеползването и земното покритие са оценени на 93.52% обща точност. Между най-добре тематично отделените класове са тези на естествената растителност и инфраструктурата, с около 100% тематична точност, докато класовете на ливадите, пасищата и голата почва се смесват един с друг поради спектралната неразделимост и подобната текстурата на класовете.*

Introduction

Land use information for regional and local scale studies is an indispensible for policy and decision making and environmental protection. Several land-use/land-cover (LU/LC) classification schemes based on the growing input of remote sensing data have been created. Concurrently, in order to provide fast and accurate results for land use numerous image classifications for digital imagery have been developed. The Automated Feature Extraction (AFE) Algorithms are new approaches to image classification by segmenting the image to distinct features taking into account images' texture and other ancillary imagery data.

The first attempts to develop routines for image segmentation have already been introduced in the mid 1970's (Neubert & Meinel 2003). It took more than two decades till segmentation algorithms were established as key functional features of contemporary software packages such as eCognition (Baatz & Schäpe 2000). Some of the RS software packages which use the imaging segmentation techniques are: eCognition 2.1 resp. 3.0 (Definiens Imaging GmbH, München, Germany); Data Dissection Tools (INCITE, Stirling University, UK); CAESAR 3.1 (N.A. Software Ltd., Liverpool, UK); InfoPACK 1.0 (InfoSAR Ltd., Liverpool, UK); Image segmentation for ERDAS Imagine (USDA Forest Service, Remote Sensing Applications Center, Salt Lake City, USA); Minimum Entropy Approach to Adaptive Image Polygonization (University of Bonn, Institute of computer science, Bonn, Germany); Feature Analyst 4.2 for ArcGIS 9.x ™ (Overwatch Geospatial LTD. a subsidiary of TEXTRON Systems Inc.), ERDAS IMAGINE™, GeoMedia™, SOCET SET™, and RemoteView ™ (Visual Learning Systems, Inc.).

Land-use/land-cover classifications are based on established LU/LC classification schemes, which aim at producing comparable results for LU/LC from the classifications on global, regional and local level. Such classification schemes are those of USGS, FAO, CORINE Land Cover, GLOBCOVER, GLC2000 etc. In present study USGS and CORINE classification schemes were adopted and combined into a hybrid one for the purpose of large-scale LU/LC classification. There are numerous applications of object-oriented image classifications of LU/LC and AFE in particular using one of the abovementioned software packages (Weih Jr & Riggan Jr ; Blaschke *et al.* 2000; Burnett & Blaschke 2003; van der Sande *et al.* 2003).

The main objective of present study is to classify the LU/LC of the *Teyna river* basin using novel approaches of the AFE Algorithms. It is achieved through accomplishment of the following steps:

1). To collect, create and manipulate geospatial dataset for the territory of *Teyna river* basin;

2). To classify the LU/LC using very high-resolution (VHR) satellite image from QuickBird by applying the AFE algorithms;

3). To assess the accuracy of the produced LU/LC map of the study area of *Teyna river* basin.

Study area

The catchment of *Teyna river* is located in the north part of Sofia kettle in the footsteps of *Sofijska Mala Planina* Mountain. The river is a small tributary of the *Iskar* River – the longest river in Republic of Bulgaria. The total area of the study area is – 4.775 km^2. The altitude of the catchment ranges from 500 m.a.s.l., at the basin's outlet at *Iskar* River gorge, to 964 m.a.s.l. on the topmost part.

Climatic conditions of the river basin are Temperate to transitional to Semi-Mediterranean and also local mountainous valley winds occur on a daily basis. *Geology* (lithology) – The bedrocks are a diverse mixture of Neogene-Quaternary argillite, Ordovician-Silurian argillite, alevrolite, schist, sandstone, breccias. All these rocks are more or less loose in their structure, so that they facilitate the manifestation of erosion. Their outcrops are observed on the slopes of Iskar River gorge.

The lithology of the rest part of the region is composed mainly of flish: argillite, silicite, sandstone with Palaeozoic and Cainozoic age, where the Palaeozoic rocks prevail. The human impact on geology of the study area is marked by the excavations and embankments of the former decommissioned *Kutina* coal mine and *Brezi Vrah* uranium-ore extraction site.

The *vegetation* of the river basin is represented mainly of deciduous oaks (*Quercus frainetto* Ten., *Q. cerris* L., and *Q. pubescens* Willd.) which are the primary vegetation of the river basin. European Beech is interspersed with planted conifers to prevent erosion such as Scots Pine, Macedonian Pine, Silver Fir and Norway Spruce mostly on the mountain slopes from 800 to 1200 m.a.s.l. The majority of vegetation is artificially afforested with durable to pollution forest types. Some of these species found on the study area are: Scots pine (*Pinus sylvestris* L.) and European Black Pine (*Pinus nigra* L.). The main *soil types*, which play major role in the LU/LC structure on the territory, are Chromic Luvisols with 45.8 % of the river basin's area. These soils are mainly located on the lower parts of the slopes, whereas the next prevailing soil type, i.e. Cambisols, with its 36.37 % are located on a higher altitude. The rest 17.4 % are covered with bare soils or Antroposols, which is connected to the human impact on soils in the region (Ninov, 2002; Soil Atlas of Europe, 2005).

Materials and methods

The input data for present study consists of Panchromatic and Multispectral (MS) satellite image from QuickBird satellite acquired in May 2008; Digital Elevation Model (DEM) with 5m accuracy derived from topographic maps with scale 1:5 000 and elevation points, Slope, Aspect and river basin's derivatives from the DEM; vector layers such as hydrological network and water bodies in the river basin, and semi-automatically generated hydrological network from the DEM. The software used is ArcGIS/ArcInfo 9.2 (Academic license), with the Feature Analyst 4.2 extension (shareware license), and ERDAS Imaging 9.1 for accuracy assessment of the results. The main dataset for the study was organized into a geodatabase in ArcInfo 9.2/ArcCatalog 9.2. The feature layers were stored in feature classes and feature datasets in UTM projection, WGS 84 reference ellipsoid, Zone 34 N.

The Automated Feature Extraction (AFE) Algorithms employs for any single image most of the major interpretation keys in remote sensing: Shape, Size, Colour, Texture, Pattern, Shadow, and Association. Each class of the LU/LC is classified by collecting a set of training test sites; whilst as an input for the image classification is used also image texture and DEM.

Results and discussions

Data collection and data manipulation are usually regarded as one of the most important stages in terms of time and resources in each earth-science research. These stages include data preparation, i.e. georeferencing, rectification, digitization, attributive data entry, geodatabase management of data layers. The main layers used as input parameters for the LU/LC classification are: QuickBird Pan and MS channels acquired in May 2008, DEM and relief's main derivatives: slope and aspect, and image texture. All this data was organized, maintained and manipulated in a personal geodatabase. The *Teyna river* basin border area was used to mask out the outer parts of the catchment from the classification in order to speed up the image classification and to enclose the classification within the study area only.

The main set of parameters for initialization of the AFE algorithms are usually chosen in advance, and for the current work the best option, after several trials with different set of parameters was: Multi-Class approach, which uses, two or more grouped training sets into a multi-class learning set using the Prepare Multi-class Input Layer feature function and the Wall-to-wall option being put on (Feature Analyst 4.2 for ArcMap, Reference Manual).

The image resolution and the size and type of the classified object are of crucial importance when setting up the learning parameters. Setting up the initial parameters is also important, because some basic pre-processing steps are taken at that stage, such as histogram stretching and bilinear interpolation and resampling of the input images. The texture pattern or the moving kernel used were generally Manhattan and Bulls-eye textures with ranging neighbours from 3 to 7-8 for better discrimination of the forest border. After running the AFE with this set of parameters, following clean-up of the meandering contours of the LU/LC types was done using the Smooth Features Tool and aggregating the output areas.

Accuracy assessment of the classification results was performed in ERDAS Imaging 9.1 software using stratified random sampling and ground truth collection during the summer of 2010. There are several sampling schemes used to perform accuracy assessment of image classification results, such as: simple random sampling to a scheme called stratified, cluster sampling, systematic and unaligned sampling. Among those the stratified random sampling is considered the most appropriate for accuracy assessment of thematic maps with classification results (Congalton, 2009). In present study the stratified random sampling was used by randomly sowing 108 sampling points proportionally to area percentage of the 9 resulting LU/LC classes. After running the AFE supervised classification the resulting LU/LC classes' percentages are represented on **Figure 1** and **Table 1** and the results from the accuracy assessment are presented on **Table 2** below.

Table 1. Percentage of LU/LC classes for *Teyna river* basin

No	Classes	Area (km^2)	Percent
1	Asphalt roads	0.036063	0.00%
2	Bare ground	0.648927	7.00%
3	Coniferous forest	1.906767	21.00%
4	Grasslands	0.198135	2.00%
5	Meadows	0.68886	8.00%
6	Mixed deciduous forest	1.562715	47.00%
7	Quarry	0.049689	1.00%
8	Sparse vegetation	1.221408	14.00%
9	Water bodies	0.051471	1.00%

Table 2. Accuracy assessment of *Teyna river* basin LU/LC classification

№	Class name	Asphalt roads	Bare ground	Coniferous forest	Grasslands	Meadows	Mixed deciduous forest	Quarry	Sparse vegetation	Water bodies	Total possible	Omissions	Commissions	Mapping accuracy
1	**Asphalt roads**	3	0	0	0	0	0	0	0	0	**3**	**0.00%**	**0.00%**	**100.00%**
2	**Bare ground**	0	5	0	0	0	0	2		0	**7**	**28.57%**	**0.00%**	**71.43%**
3	**Coniferous forest**	0	0	19	0	0	0	0	2	0	**21**	**9.52%**	**9.52%**	**82.61%**
4	**Grasslands**	0	0	0	1	1	0	0	0	0	**2**	**50.00%**	**50.00%**	**50.00%**
5	**Meadows**	0	0	0	0	8	0	0	0	0	**8**	**0.00%**	**12.50%**	**88.89%**
6	**Mixed deciduous forest**	0	0	0	0	0	47	0	0	0	**47**	**0.00%**	**0.00%**	**100.00%**
7	**Quarry**	0	0	0	0	0	0	3	0	0	**3**	**0.00%**	**66.67%**	**60.00%**
8	**Sparse vegetation**	0	0	2	0	0	0	0	12	0	**14**	**14.29%**	**14.29%**	**75.00%**
9	**Water bodies**	0	0	0	0	0	0	0	0	3	**3**	**0.00%**	**0.00%**	**100.00%**
		3	**5**	**21**	**1**	**9**	**47**	**5**	**14**	**3**	**108**	**Overall accuracy**		**93.52%**

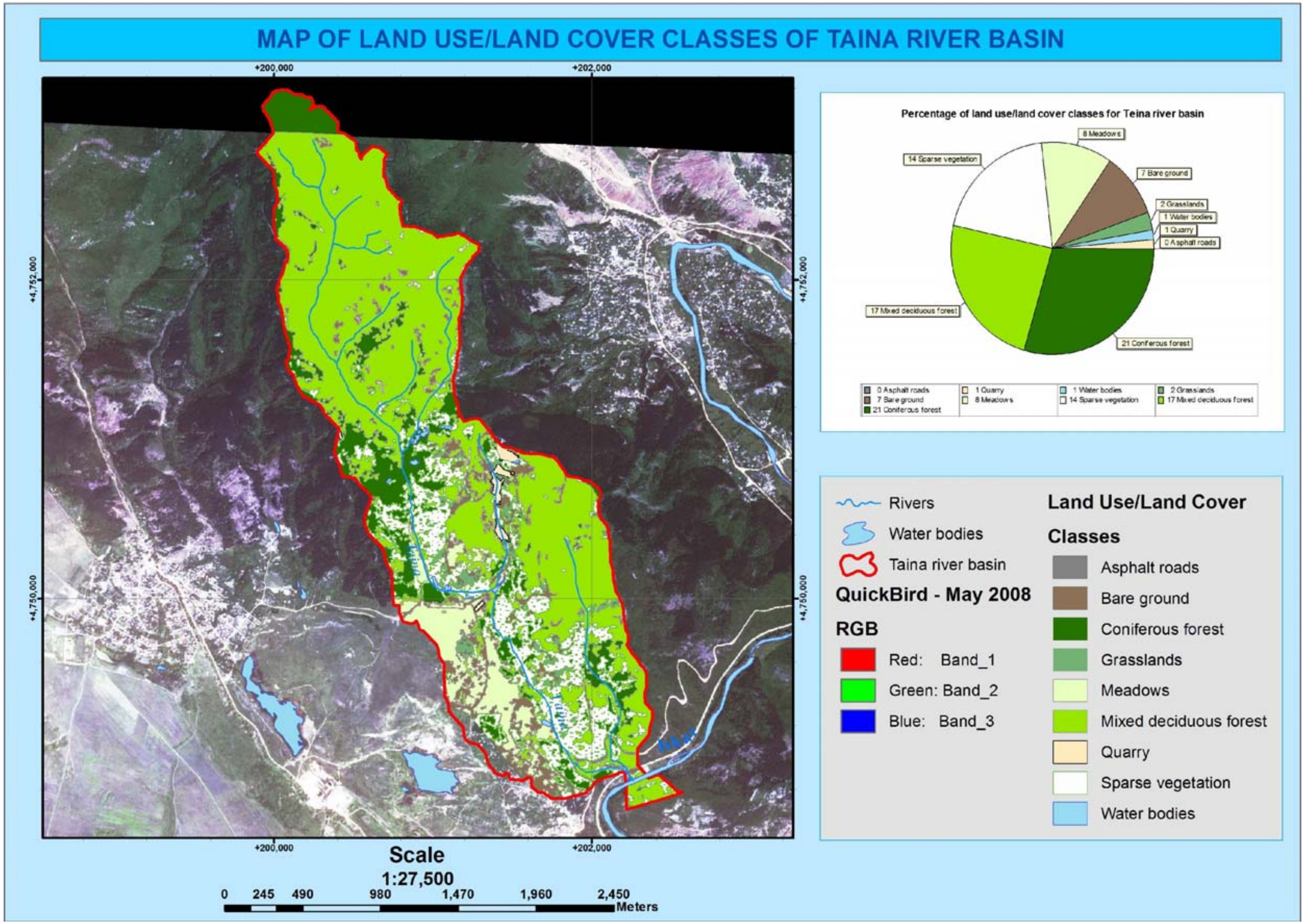

Figure 1. Map of LU/LC Classification of Teyna river basin

Conclusions

From the predefined patterns some of them preset for the specific LU/LC type, used for training over the image set, the Manhattan and Bulls-eye texture types were found to be more suitable for LU/LC classification of the of *Teyna river* basin. That is mainly because of the high resolution of the QuickBird image and the patchy mosaics of the landscape. This conclusion is also applicable for the texture image, because it was derived from the panchromatic channel of QuickBird, which does not differ significantly from the MS texture. Some key assumptions which could be drawn from the accuracy assessment of the LU/LC results are:

- The overall mapping accuracy is above the lower limit for LU/LC classifications – 93.52%, which is satisfactory for large scale LU/LC purposes.
- Thematic accuracy of the LU/LC classification of the *Teyna river* basin using the means of AFE algorithms revealed that the lowest accuracy is matched for the classes of Grasslands – 50%, Quarries – 60% and Bare ground – 71.43%. This could be addressed mainly to poor spectral discrimination in the spectral space between classes of QuickBird image, and their similar texture.
- The highest accuracy of about 100% was achieved in the LU/LC classes such as: Asphalt roads, Mixed deciduous forest and Water bodies.

The low thematic accuracy of some LU/LC classes could be due to the omission of the AFE classification algorithm or due to interference of other input layers such as texture and DEM. The highest thematic accuracy for abovementioned classes could be due to better separability in spectral space and clear borderlines of the natural features on the satellite image. Lower accuracy in Coniferous forests, which is typically better discriminated using other hard and soft image classifiers, could be due to mixing of shade, texture and DEM's values with cliffs, and other shadows cast by natural or manmade features.

Acknowledgements

Present study was kindly supported with data and software facilities from Project Nz-No.1507/05 *Development of Methodical Fundamentals of Landscape-Ecological Planning Using Geoinformation Technologies* between SRI-BAS and the Scientific Research Fund at the Ministry of Education and Science of Bulgaria. The work is also supported by the Framework of the Operational Program-Human Resources Development of the Bulgarian Ministry of Education, Youth and Science, Funded by the European Social Fund (ESF), CONTRACT No BG051PO001/07/3.3-02/63/170608. The ArcGIS 9.2 software used for the presented study was granted by ESRI Inc. in 2007 under the ESRI Conservation Program and SCGIS Scholarship Program. Feature Analyst 4.2 extension shareware license for this study ArcGIS 9.x was granted by Overwatch Geospatial LTD., a subsidiary of Textron Systems Inc. Special acknowledgements have to be given to the staff and the head of Department of Remote Sensing and GIS and especially to Assoc. Prof. Dr Eugenia Roumenina, Ph.D.

References:

1. Baatz, M. & A. Sch, (2000) Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. *Journal of Photogrammetry and Remote Sensing* 58(3-4): pp. 12-23.
2. Blaschke, T., S. Lang, E. Lorup, J. Strobl & P. Zeil, (2000) Object-oriented image processing in an integrated GIS/remote sensing environment and perspectives for environmental applications. *Environmental information for planning, politics and the public 2*: pp. 555–570.
3. Burnett, C. & T. Blaschke, (2003) A multi-scale segmentation/object relationship modelling methodology for landscape analysis. Ecological Modelling 168(3): pp. 233-249.
4. Congalton, R. & K. Green, (2009) Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. 2nd edition, CRC Press, New York.
5. Neubert, M. & G. Meinel, (2003) Evaluation of segmentation programs for high resolution remote sensing applications. In: Proceedings of the ISPRS Joint Workshop *High Resolution Mapping from Space*, pp. 6–8.
6. Ninov, N. Taxonomic List of Soils in Bulgaria According the Fao World Soil Classification." Teaching in Geography, No. 5 (2005). URL: http://www.geography.iit.bas.bg/2005/5-05/pp4-20.html [date accessed 10.01.2010].
7. Van der Sande, C. J., S. M. de Jong, & A. P. J. de Roo, (2003) A segmentation and classification approach of IKONOS-2 imagery for land cover mapping to assist flood risk and flood damage assessment. *International Journal of Applied Earth Observation and Geoinformation* 4(3): pp. 217-229.
8. Feature Analyst 4.2 for ArcMap Reference Manual, Visual Learning Systems, Inc., 2008.
9. European Soil Bureau Network. "Soil Atlas of Europe." 128 European Commission, 2005.