

DETERMINATION OF LAND COVER/LAND USE USING SPOT 7 DATA WITH SUPERVISED CLASSIFICATION METHODS

F. Bektas Balcik^a, A. Karakacan Kuzucu^{a,b}

^a ITU Civil Engineering Faculty, Geomatics Engineering, 34469 Maslak Istanbul, Turkey- bektasfi@itu.edu.tr

^bBuyukcekmece Municipality, 34535, Buyukcekmece, Istanbul, Turkey- aysin.kuzucu@bckmece.bel.tr

ISPRS WG II/2

KEY WORDS: Land Cover / Land Use, Classification, Support Vector Machine, Maximum Likelihood Classification, Istanbul, SPOT 7

ABSTRACT:

Land use/ land cover (LULC) classification is a key research field in remote sensing. With recent developments of high-spatial-resolution sensors, Earth-observation technology offers a viable solution for land use/land cover identification and management in the rural part of the cities. There is a strong need to produce accurate, reliable, and up-to-date land use/land cover maps for sustainable monitoring and management. In this study, SPOT 7 imagery was used to test the potential of the data for land cover/land use mapping. Catalca is selected region located in the north west of the Istanbul in Turkey, which is mostly covered with agricultural fields and forest lands. The potentials of two classification algorithms maximum likelihood, and support vector machine, were tested, and accuracy assessment of the land cover maps was performed through error matrix and Kappa statistics. The results indicated that both of the selected classifiers were highly useful (over 83% accuracy) in the mapping of land use/cover in the study region. The support vector machine classification approach slightly outperformed the maximum likelihood classification in both overall accuracy and Kappa statistics.

1. INTRODUCTION

Classification of satellite images which have different spatial, spectral and temporal resolution obtained by remote sensing technology with high accuracy is quite necessary to access accurate and reliable information about the earth surface in many applications such as eco-environment research, urban planning, vegetation condition study, agriculture activities and soil management. However, land use/cover classification is a difficult process with high accuracy. Especially, the classification of heterogeneous areas that have rural and urban together is still faced with difficulties and problems.

Recent research efforts to improve the reliability and accuracy of image classification have led to the introduction of the Support Vector Machines (SVM). A variety of machine learning and pattern recognition techniques have been developed to improve land cover mapping in heterogeneous areas, and support vector machines (SVM) is found to outperform most of the conventional classifiers (Huang et al., 2002; Keuchel et al., 2003; Kavzoglu and Colkesen, 2009; Su and Huang, 2009; Shi and Yang, 2012). SVM is a new generation of supervised learning method based on the principle of statistical learning theory, which is designed to decrease uncertainty in the model structure and the fitness of data (Vapnik, 1995). SVM can work with large dimensional data sets and integrate with auxiliary data set.

In this study, maximum likelihood classification (MLC) method was applied as very common traditional classification method. MLC classification is based on Bayesian probability theory. It is one of the most widely used algorithm (Perumal and Bhaskaran, 2010; Ahmad and Quegan, 2012). The maximum likelihood classifier basically develops a probability function based on inputs from a dataset collected for training. After that the method considers

each individual pixel in an image, compares it with known pixels and assigns unknown pixels to a class based on similarity and highest probability to belong to one of the already known classes (Jensen, 2005; Kwesi, 2012). Maximum likelihood classification method involves the estimation of class mean variance and covariance matrices using training patterns chosen from known pixels of each particular class.

The study aims to classify heterogeneous land use/cover and agricultural pattern belong to Catalca district, Istanbul using two supervised classification methods such as MLC and SVM Classification methods to compare performances of both classification methods on a new generation SPOT 7 image. The support vector machine classification approach slightly outperformed the maximum likelihood classification in both overall accuracy and Kappa statistics.

2. STUDY AREA& DATA

For this study Catalca district has been selected as the study area. Catalca is located in Kocaeli-Catalca part of Marmara region, Istanbul (Figure 1). It is an urban center where established on western boundary of Istanbul's European side and the largest district of Istanbul with total area of 1115.50 square kilometers. The population of the region is 67.329 according to the population census of 2015 (Turkey Statistical Institute). In the region a majority of the surface area is covered with forests. The other land cover/use categories include artificial surfaces, water surfaces, wetlands, agricultural fields, pasture lands, forest lands and mining areas. The most important labor sector in Catalca district is agriculture. According to Istanbul provincial agriculture master plan, it has the most agricultural land distribution, agricultural production and productivity. It has

partially Mediterranean climate zone, partially Marmara transition type of climate and partially Black sea climate zone.

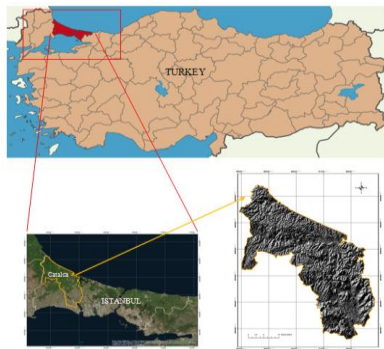


Figure 1. Study area

In this study SPOT 7 multispectral satellite image was used (<http://www.intelligence-airbusds.com>). The data acquired in August 2015 with four spectral bands (blue, green, red, near-infrared), and spatial resolution of the data is 6 m. (Table 1). SPOT 7 satellite was launched on June 30, 2014. It has four multispectral bands: blue (455 to 525 nm), green (530 to 590 nm), red (625 to 695 nm), and near-infrared (760 to 890 nm). It also has a panchromatic (450 to 745 nm) band. Images of the panchromatic band can reach 1.5 m resolution and images of multispectral bands obtain 6 m resolution. SPOT 7 satellite has daily revisits everywhere and its imaging swath is 60 km at nadir.

In the study, fieldwork was conducted for the classification and classification accuracy steps. Handhold GPS was used to collect different land cover/use categories location information. Due to the basic characteristics of the study area was mainly carried out field studies to act with determination of sample area of the agricultural product type. Considering the phenological development process of agricultural products fieldwork was carried out between June and August. Determination of the sample area is considered to be one of the most important process steps. This is because the spatial reference data determined is fundamental to characterize the spectral properties of consisting of different land cover / use and agricultural product types, to select sample area in the image classification stage as well as in the accuracy assessment after the classification. For this purpose, field studies were made with hands-GPS measuring, and 1:5000 scaled topographic maps and google earth images were used. Sample areas was grouped according to their properties and the photos of each sample were stored.

Product Attribute	Description
Multispectral Imagery (4 bands)	Blue (0.455 μm - 0.525 μm) Green (0.530 μm - 0.590 μm) Red (0.625 μm - 0.695 μm) Near-Infrared (0.760 μm - 0.890 μm)
Resolution (GSD)	6.0 m (B, G, R, NIR)
Geometrical	ORTHO
Processing Level	
Radiometrical	Linear_strech
Processing Level	
Coordinate System	WGS_1984_UTM_Zone_35 N
Acquisition Date	16.08.2015

Table 1. Properties of SPOT 7

3. METHODOLOGY

3.1. Image Pre – Processing

The methodology of the study includes mainly three steps which are pre-processing of satellite image (radiometric, atmospheric and geometric correction), classification of satellite image with MLC and SVM, evaluating and comparing classification accuracies of both classification techniques.

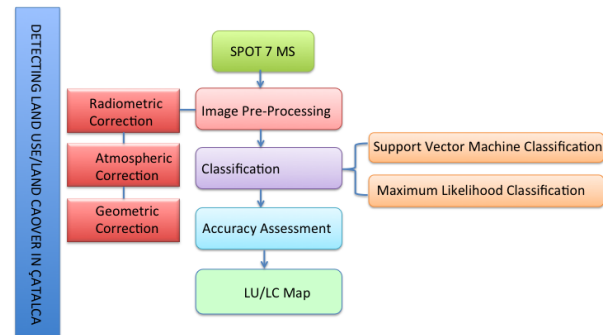


Figure 2. Flowchart of the study

In the image pre- processing stage radiometric and atmospheric correction were applied. Digital number values for the SPOT 7 satellite image were converted to radiance values and FLAASH Atmospheric Correction Model was used for atmospheric correction. Scene center location (lat/lon), sensor type, sensor altitude (km), ground elevation (km), pixel size (m), flight date, flight time GMT (HH:MM:SS), atmospheric model, aerosol model, initial visibility (km) were used as input parameters in the model. SPOT 7 image was geometrically corrected by image-to-map registration using 1:5000 scaled topographic maps and orthophotos. It took into consideration to be homogeneously distributed on the image of the selected control points. Connection between the image coordinate system and the ground coordinate system was accomplished by polynomial transformation. As a result of the conversion, it was obtained under the root mean square error of 0.5 pixels. Nearest neighbor method was used as resampling method.

3.2. Classification Methods

3.2.1 Maximum Likelihood Classification

Supervised classification is a technique that is based on the statistics of training areas representing different ground objects selected subjectively by users on the basis of their own knowledge or experience. In this study, maximum likelihood classification (MLC), which is the most common classification method in remote sensing, was used to derive land-use/cover categories of selected study area. In this method, the pixel is assigned to the class for which the probability of the pixel belonging is highest. MLC is based on Bayes' Theorem (Jensen, 2005).

3.2.2 Support Vector Machine Classification

The support vector machine (SVM) method, which was originated in the late 1970s by Vapnik (1995), is one of the machine-learning algorithms. It is based on statistical learning theory and has recently been extensively used in

remote sensing for pattern recognition and classification. The method was originally designed for binary classification, and it allows the use of optimal algorithms to locate the best boundaries separating the binary classes in the feature's space. The boundary is called the optimum separating hyperplane, which is aimed at maximizing the margin width.

The main advantage of SVM is the ability to make good generalizations of high-dimensional data with few training samples. SVM works with pixels in the boundaries of classes, which are called support vectors, and thus it is possible to get accurate classifications with small training sets.

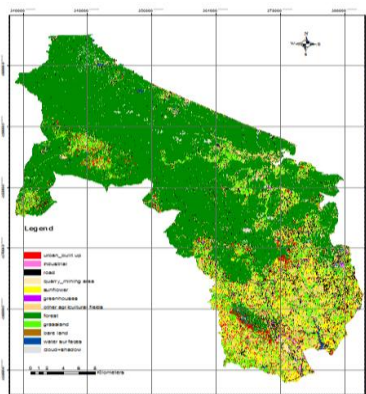
Related parameters vary according to used kernel function such as Linear kernel, Polynomial kernel, Radial Basis Function kernel, Sigmoid kernel (Yang, 2011; Shi and Yang, 2012). In this study radial basis function kernel was used and optimal parameter values for the selected kernel function determined based on literature review. The most challenging and important factor for the SVM classifications was suitable choice of the kernel types and the optimum parameters. The parameters for the RBF kernel were set to 0,250 and 100 for γ (gamma= 1/number of the band) and C (error penalty), respectively, for the SVM classification. The pyramid parameter was set to a value of 0 to process the satellite data at full resolution (6 m).

4. RESULTS & CONCLUSIONS

In this study, two different supervised classification methods were employed using SPOT 7 data to analyze the potential of mapping land use/land cover in the Catalca, Istanbul. In the classification, at the first stage, training and test data was prepared to classify satellite image that is covering the study area. The same training data was used for all classifications problems in this study.

As a result of the classification methods, twelve land-use categories were distinguished in the selected region: urban-built up, industrial, road, quarry/mining area, sunflower, greenhouses, other agricultural fields, forest, grassland, bare land, water surfaces and cloud+shadow (Figure 3).

(a)



(b)



Figure 3. Classification results for: (a) MLC, and (b) SVM

The output of the classified images without any error or bias is the accurate thematic map. There are a number of equations that can show the level of error statistically, such as producer accuracy, user accuracy, overall accuracy, and Kappa, which can be calculated using the error matrix (Foody, 2002). An accuracy assessment was performed using collected ground-truth data for all two types of classification using a standard error matrix. Kappa statistics and overall accuracy were used to compared to determine the performance of the selected methods for the selected heterogeneous region (Table 2).

Classification Method	Overall Accuracy	Kappa Coefficient
Maximum Likelihood Classification	83.0486%	0.7703
Support Vector Machine Algorithm	85.6109%	0.7981

Table 2. Classifier accuracies based on test data for MLC and SVM

The results showed that Spot 7 MS images could be used to produce land use/cover maps and statistics. Based on accuracy assessment results thematic maps were produced with overall accuracy and kappa statistic values of 83.05 % and 0,77 for MLC, 85.61% and 0.80 for SVM supervised classification methods.

The evaluation results demonstrated that the SVM algorithm with an overall accuracy and a kappa coefficient has a higher accuracy in comparison with the MLC algorithm in land use mapping.

This algorithm has been suggested as an optimal classifier for the extraction of land use/cover map because of its higher accuracy and better consistency with the study area.

ACKNOWLEDGEMENT

The authors would like to thank ITU Research and Application Center for Satellite Communications and Remote Sensing (CSCRS) for providing remotely sensed data.

7. REFERENCES

- Ahmad, A., Quegan, S., 2012. Analysis of Maximum Likelihood Classification on Multispectral Data. *Applied Mathematical Sciences*, 6 (129), pp. 6425 - 6436.
- Foody, G. M., 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80, pp. 185–201.
- <http://www.intelligence-airbusds.com> Received date: 15 Sept. 2016.
- Huang, C. Davis, L.S., Townshend, J.R.G., 2002. An Assessment of Support Vector Machines for Land Cover Classification. *International Journal of Remote Sensing*, 23(4), pp. 725–749.
- Jensen, R. J., 2005. Introductory Digital Image Processing, 3rd edition. New York: Prentice Hall.
- Kavzoglu, T.; Colkesen, I., 2009. A Kernel Functions Analysis for Support Vector Machines for Land Cover Classification. *International Journal of Applied Earth Observation and Geoinformation*, 11, pp. 352-359.
- Keuchel, J., Naumann, S., Heiler, M., Siegmund, A., 2003. Automatic Land Cover Analysis for Tenerife by Supervised Classification Using Remotely Sensed Data. *Remote Sensing of Environment*, 86 (4), pp. 530-541.
- Kwesi, N. I., 2012. Oil Palm Mapping Using Support Vector Machine With Landsat ETM+Data. Enschede, The Netherlands.
- Perumal, K., Bhaskaran, R., 2010. Supervised Classification Performance of Multispectral Images. *Journal of Computing*, 2 (2), ISSN 2151-9617.
- Shi, D., Yang, X., 2012. Support Vector Machines for Landscape Mapping from Remote Sensor Imagery. Proceedings - AutoCarto. Columbus, Ohio, USA.
- Su, L., Huang, Y., 2009. Support Vector Machine (SVM) Classification: Comparison of Linkage Techniques Using a Clustering- Based Method for Training Data Selection. *GIScience & Remote Sensing*, 46 (4), pp. 411-423.
- Turkey Statistics Institute, 2015.
- Vapnik, V.N., 1995. The Nature of Statistical Learning Theory, New York: Springer-Verlag.
- Yang, X., 2011. Parameterizing Support Vector Machines for Land Cover Classification. *Photogrammetric Engineering and Remote Sensing*, 77 (1), pp. 27-37.