

THE EFFICIENCY OF RANDOM FOREST METHOD FOR SHORELINE EXTRACTION FROM LANDSAT-8 AND GOKTURK-2 IMAGERIES

B. Bayram ^{1,*}, F. Erdem ², B. Akpinar ¹, A. K. Ince ², S. Bozkurt ², H. Catal Reis ³, D. Z. Seker ⁴

¹ Yildiz Technical University, Civil Engineering Faculty, Dept. of Geomatic Engineering, Davutpasa Campus, 34220 Esenler-İstanbul, Turkey - (bayram,bakpinar)@yildiz.edu.tr

² Yildiz Technical University, Graduate School of Natural And Applied Sciences, Dept. of Geomatic Engineering, Davutpasa Campus, 34220 Esenler-İstanbul, Turkey – (f5016010, f5016002, f5016003) @std.yildiz.edu.tr

³ Gumushane University, Faculty of Engineering, Dept. of Geomatics, Gumushane-Turkey - hcatal@gumushane.edu.tr

⁴ ITU, Civil Engineering Faculty, 80626 Maslak Istanbul, Turkey - sekert@itu.edu.tr

KEY WORDS: Random Forest, Shoreline Extraction, Image segmentation, LANDSAT-8, GOKTURK-2

ABSTRACT:

Coastal monitoring plays a vital role in environmental planning and hazard management related issues. Since shorelines are fundamental data for environment management, disaster management, coastal erosion studies, modelling of sediment transport and coastal morphodynamics, various techniques have been developed to extract shorelines. Random Forest is one of these techniques which is used in this study for shoreline extraction.. This algorithm is a machine learning method based on decision trees. Decision trees analyse classes of training data creates rules for classification. In this study, Terkos region has been chosen for the proposed method within the scope of "TUBITAK Project (Project No: 115Y718) titled" Integration of Unmanned Aerial Vehicles for Sustainable Coastal Zone Monitoring Model – Three-Dimensional Automatic Coastline Extraction and Analysis: Istanbul-Terkos Example ". Random Forest algorithm has been implemented to extract the shoreline of the Black Sea where near the lake from LANDSAT-8 and GOKTURK-2 satellite imageries taken in 2015. The MATLAB environment was used for classification. To obtain land and water-body classes, the Random Forest method has been applied to NIR bands of LANDSAT-8 (5th band) and GOKTURK-2 (4th band) imageries. Each image has been digitized manually and shorelines obtained for accuracy assessment. According to accuracy assessment results, Random Forest method is efficient for both medium and high resolution images for shoreline extraction studies.

1. INTRODUCTION

Coastal areas are the most important settlement areas throughout human history. Due to increasing of population, urbanization, shorelines and eco-system are under the threat of human being (Bendell and Wan, 2011). According to the International Geographical Data Committee, coastal areas are one of the 27 important natural heritage on Earth (Li et al., 2001). Thus, rapid, up-to-date, and correct information is essential for coastal management. The determination of shoreline dynamics has primary importance for coastal managers. Therefore, shoreline extraction is the primary step for coastal management issue. Remote sensing and image processing techniques provide rapid shoreline extraction solutions compare to traditional methods (Bayram et al., 2017). The monitoring of shoreline changes is one of concerns of researchers. (Dornbusch et al., 2006; Marques, 2006). Therefore, temporal monitoring of shorelines has primary importance (Gens, 2010). LANDSAT-8 and GOKTURK-2 imageries are open data resources (Machado et al., 2014; Kalkan et al., 2015). Some of commonly used shoreline extraction methods are Unsupervised classification techniques (ISODATA-Iterative Self Organized Data Analysis) (Guariglia et al., 2006), normalized difference water index (NDWI) (Zheng et al., 2011), thresholding and morphological filtering techniques (Pardo Pascual et al., 2012), Wavelet transformation (Yu et al., 2013), active contour method (Shmittet al., 2015), genetic algorithm based methods (Yousef and Iftekharuddin, 2014), particle swarm optimization method (PSO) (Bayram-a et al., 2016), Mean-shift segmentation (Bayram-b, vd, 2016), object oriented fuzzy classification methods (Bayram et al., 2015; Bayram et al., 2013; Bayram et al. 2008), normalized cut approach (Ding and Li,

2014) . In this study, the shoreline of the Terkos/Istanbul has been extracted using Random Forest method (Breiman, 2001) from LANDSAT-8 and GOKTURK-2 imageries. Extraction results have been evaluated by DSAS and efficiency of Random forest method has been discussed.

2. STUDY AREA

Study area is consisting of 19 km shoreline of Black Sea which located in the north part of Lake Terkos/Istanbul. Terkos shoreline is under the threat of erosion due to increasing urbanisation. The study area has been given in Figure 1.



Figure 1. Study area.

3. MATERIAL AND METHOD

In the presented study, LANDSAT-8 (06.09.2015) and GOKTURK-2 (30.06.2015) images have been used. The 5th band of LANDSAT-8 and 4th band of GOKTURK-2 have been processed and shoreline of Terkos/Istanbul has been extracted.

* Corresponding author

The specifications of LANDSAT-8 and GOKTURK-2 have been given in Table 1 and Table 2 respectively.

Bands	Resolution (m)	Specifications	Spectral Range (nm)
1	30	Coastal/Aerosol	435-452
2		Blue	452-512
3		Green	533-590
4		Red	636-673
5		NIR	851-879
6	100	TIR-1	1060-1119
		TIR-2	1150-1251
7	30	SWIR-2	2107-2294
8	15	Pan	503-676
9	30	Cirrus	1363-1384

Table 1. Specifications of LANDSAT-8 (Landsat-8, 2017).

Bands	Resolution (m)	Specifications	Spectral Range (nm)
1	5	Blue	450-520
2		Green	520-600
3		Red	630-690
4		NIR	760-900
5	2.5	PAN	450-900

Table 2. Specifications of GOKTURK-2 (Kalkan et al., 2015).

Random Forest classification algorithm is based on decision trees and a pixel based machine learning method. By analysing of training data sets, rules are created and object classes are determined. The rules consist of several if-then conditions (Breiman, 2001). Random Forest algorithm requires two parameters which are number of trees and the number of random variables to be used for each node for creation of decision trees (Belgiu and Drăguț, 2016).

After completion of parameters, if any additional data set is not existing, 2/3 of training data set is used as learning data, 1/3 of is used as test data. Multiple CART-like trees are created by Random Forest in training step (Breiman et al., 1994). To determine a split for each node, bootstrapped technique is used and randomly selected subsets from input variables are searched (He et al., 2015; Gislason et al., 2006). CART algorithm uses GINI index to determine the best split (Gislason et al., 2006). GINI index measures the homogeneity of samples for each node. Algorithm calculates the GINI index for random selected variables for each node. The variable which has the minimum GINI index is selected and algorithm and calculations are repeated for next node. If the GINI index is resulted with zero, it means that related node is totally homogeneous and this node is defined as and of branching (Gislason et al., 2006). The out of bag samples (samples of remaining training set are not

bootstrapped for a particular tree) of each tree are cross validated. For each pixel, a classification vote is calculated according to weight of decision tree and the pixel is assigned to the majority voted class (Gislason et al., 2006).

In this study, the Random Forest algorithm has been realized by using MATLAB platform. Both LANDSAT-8 and GOKTURK-2 images have been classified and two classes which are land and water body, were created. TreeBagger function of MATLAB has been used. The number of trees and the number of random variable for both images have been selected as 50, 1 respectively. Predict function of MATLAB has been used to define the corresponding class for each image pixel. The segmentation results for LANDSAT-8 and GOKTURK-2 are given in Figure 3 and 4 respectively.



Figure 3. The segmentation results of LANDSAT-8.



Figure 4. The segmentation results of GOKTURK-2.

4. RESULTS

Shape data for both images have been obtained after applying raster to vector conversion process. Manually digitized shorelines each image has been compared with obtained shorelines separately. For this purpose, Digital Shoreline Analysis System (DSAS) (Thieler et al., 2009) has been used. DSAS is a plugin for ArcGIS commercial software to evaluate extracted shoreline with reference data (Jayson-Quashigah et al., 2013). The Net Shoreline Movement function of DSAS has been used for accuracy assessment which calculates perpendicular distances with defined spacing between input and reference data (Oyedotun, 2014). The spacing has been chosen for LANDSAT-8 and GOKTURK-2 as 5m. Length of transects have been defined for LANDSAT-8 and GOKTURK-2 as 300m and 100m respectively as given in Figure 5,6. Red and blue lines represent

photogrammetrically digitized and extracted shorelines respectively in Figure 5,6. The number of transects were 3925 for LANDSAT-8 and 3942 for GOKTURK-2.

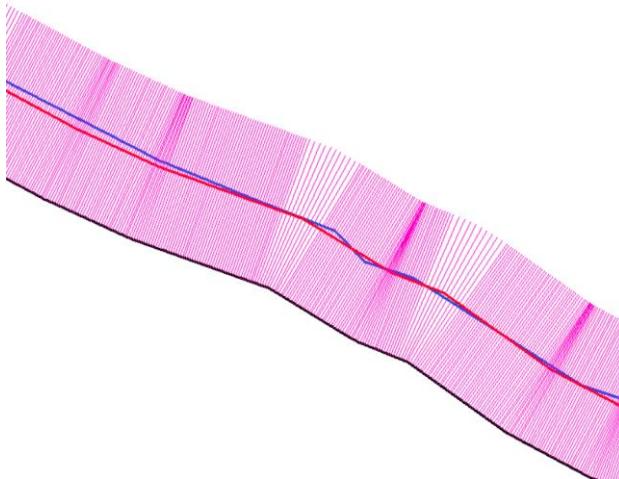


Figure 5. Transects for LANDSAT-8 image.

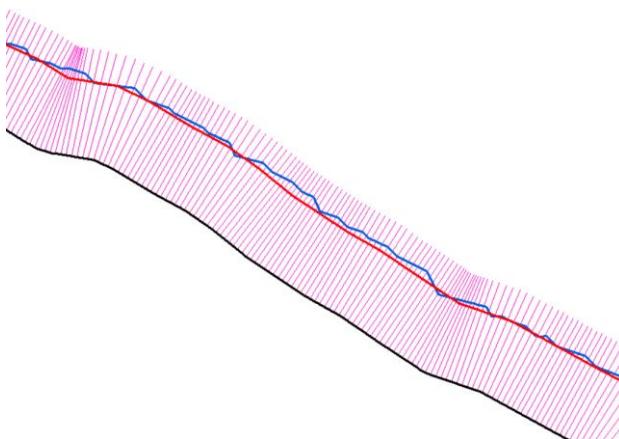


Figure 6. Transects for GOKTURK-2 image.

The results for LANDSAT-8 and GOKTURK-2 have been given in Table 3,4.

Calculated distance (m)	0-6	6-10	10-15	15-30	30-37
Number of transects	1432	914	816	728	35
Ratio (%)	36.48	23.29	20.79	18.55	0.89
Average (m)			9.238		
Stdev (m)			6.553		
RMS (m)			11.327		

Table 3. Accuracy assessment results for LANDSAT-8.

Calculated distance (m)	0-1	1-1.67	1.67-2.5	2.5-5	5-24
Number of transects	755	464	592	1313	818
Ratio (%)	19.15	11.77	15.02	33.31	20.75
Average (m)			3.248		
Stdev (m)			2.534		
RMS (m)			4.119		

Table 4. Accuracy assessment results for GOKTURK-2.

5. CONCLUSION

The average distances between manual digitized and extracted shoreline for LANDSAT-8 and GOKTURK-2 imageries are calculated as 11.327m and 3.248 m respectively. The calculated distance ratios in pixel size has been given in Table 5. 1/5, 1/2 and 1pixel size of for LANDSAT-8 image are calculated as 36.48, 80.56% and 99.11% respectively. According to Table 5 80.56% of the distances are calculated in 1/2 pixel size for LANDSAT-8 image. As it can be seen in Table 5, 79.25% of the distances are calculated in 1 pixel size for GOKTURK-2 image. One of the reason can be depended on the number of training pixels. The results show that number of training pixels should be increased for GOKTURK-2 image.

	0-1/5 pixel	0-1/3 pixel	0-1/2 pixel	0-1 pixel	> 1 pixel
LANDSAT-8 (%)	36.48	59.77	80.56	99.11	0.89
GOKTURK-2 (%)	19.15	30.92	45.94	79.25	20.75

Table 5. Calculated distance ratios in pixel size.

The success of Random Forest method is higher than GOKTURK-2 than LANDSAT-8. Since Random Forest method is a pixel based method, obtained results are not confusing. As many researchers mentioned, object-oriented methods can provide more successful results compare to pixel based methods and this was provided with achieved results. In the future studies, Support Vector Machine method and Random Forest method will be used to increased number of training pixels applied on GOKTURK-2 imageries and obtained results will be compared.

ACKNOWLEDGEMENTS

This study has been supported by "TUBITAK (The Scientific and Technological Research Council of Turkey) with project number 115Y718and entitled " Integration of Unmanned Aerial Vehicles for Sustainable Coastal Zone Monitoring Model – Three-Dimensional Automatic Coastline Extraction and Analysis: Istanbul-Terkos Example “.

REFERENCES

Bayram B., Acar U., Seker D. Z., Ari A., 2008. A Novel Algorithm for Coast Line Fitting Through A Case Study Over Bosphorus. *Journal of Coastal Research*, 24(4), pp. 983–991.

Bayram B., Avşar E. Ö., Şeker D. Z., Kayı A., Erdoğan M., Eker O., Janpaula I., Çatal R. H., 2017. The Role Of National And International Geospatial Data Sources In Coastal Zone

Management. *Fresenius Environmental Bulletin*, 26(1), pp. 383-391.

Bayram B., Demir N., Ogurlu M., Catal R. H., Seker D. Z., 2016. 3D Shoreline Extraction Using Orthophoto-Maps and LIDAR. In: *37th Asian Conference on Remote Sensing*, 17-21 October 2016, Sri Lanka, Colombo, pp.1-5.

Bayram B., Janpaul I., Avsar O., Ogurlu M., Bozkurt S., Catal Reis H., Seker, D.Z., 2015. Shoreline Extraction and Change Detection using 1:5000 Scale Orthophoto Maps: A Case Study of Latvia-Riga, *International Journal of Environment and Geoinformatics*, 2(3), pp. 1-6.

Bayram B., Seker D. Z., Acar U., Yuksel Y., Guner A. H. A., Cetin, I., 2013. An Integrated Approach to Temporal Monitoring of the Shoreline and Basin of Terkos Lake. *Journal of Coastal Research*, 29(6), pp. 1427-1435.

Bayram, B., Seker, D. Z., Acar, U., Yuksel, Y., Guner, A. A. H., 2013. An Integrated Approach to Temporal Monitoring of the Shoreline and Basin of Terkos Lake. *Journal of Coastal Research*, 29(6), pp. 1427–1435.

Bayram, B., Demir, N., Seker, D. Z., Ogurlu, M., Oy, S., Bozkurt, S., İnce, A., 2016. Shoreline extraction using Meanshift segmentation from LIDAR-Intensity data. In: *Abstract proceedings of World GIS days*, 24-25 November 2016 İstanbul, p.43 (Original is in Turkish)

Belgiu M., Dragut L., 2016. Random forest in remote sensing: A review of applications and future directions, *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, pp. 24–31.

Bendell, L. I., Wan, P. C., 2011. Application of aerial photography in combination with gis for coastal management at small spatial scales: a case study of shellfish aquaculture. *Journal of Coastal Conservation*, 15(4), pp. 417–431.

Breiman, L., 1994. Bagging predictors. Technical Report No. 421, Department of Statistics, University of California, Berkeley.

Breiman, L., 2001. Random forests. *Machine learning*, 45(1), pp. 5-32.

Ding, X., Li, X., 2014. Coastline Detection in SAR Images Using Multiscale Normalized Cut Segmentation. In: *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*, Quebec City, Canada pp. 4447-4449.

Dornbusch, U., Robinson, D. A., Moses, C. A., Williams, R. B. G., 2006. Chalk coast erosion and its contribution to the shingle budget in East Sussex, *Zeitschrift für Geomorphologie*, 144, pp. 215–230.

Erdem F., Nasırzadehdızajı R., Derinpinar M.A., D.Z. Seker, Bayram.B., 2017. Shoreline extraction using Random Forest Case study: Istanbul/Terkos, In: Proceeding of IXth TUFUAB Technical Symposium, 27-29 April, Afyon-Turkey, pp. 84-89.

Gens, R., 2010. Remote sensing of coastlines: detection, extraction and monitoring. *International Journal of Remote Sensing*, 31(7), pp. 1819–1836.

Gislason P. O., Benediktsson J. A., Sveinsson J. R., 2006. Random Forests for land cover classification, *Pattern Recognition Letters*, 27, pp. 294–300.

Guariglia, A., Buonamassa, A., Losurdo, A., Saladino, R., Trivigno, M. L., Zaccagnino, A., 2006. A multisource approach for coastline mapping & identification of the shoreline changes. *Annals of Geophysics*, 49(1), pp. 295–304.

He, J., Harris, J. R., Sawada, M., Behnia, P., 2015. A comparison of classification algorithms using Landsat-7 and Landsat-8 data for mapping lithology in Canada's Arctic. *International Journal of Remote Sensing*, 36(8), pp. 2252-2276.

Jayson-Quashigah, P. N., Addo, K. A., Kodzo, K. S., 2013. Medium resolution satellite imagery as a tool for monitoring shoreline change. Case study of the Eastern coast of Ghana, *Journal of Coastal Research*, 65(sp1), pp. 511-516.

Kalkan, K., Orhun, O., Filiz, B.G., Teke, M. , 2015. Vegetation Discrimination Analysis from Göktürk-2 Imagery, In: *Recent Advances in Space Technologies (RAST), 2015 7th International Conference* on 16-19 June 2015 Istanbul, Turkey, DOI: 10.1109/RAST.2015.7208364.

Landsat-8, <https://landsat.gsfc.nasa.gov/landsat-8/landsat-8-overview/> (Last Updated: April 3, 2017)

Li, R., K. Di., R. Ma., 2001. A Comparative Study of Shoreline Mapping Techniques. In: *4th International Symposium on Computer Mapping and GIS for Coastal Zone Management*, 18-20 June 2001, Halifax, Nova Scotia, Canada.

Machado C. A. S., Beltrame A. M. K., Shinohara E. J., Giannotti M. A., Durieux L., Nobrega T. M. Q., Quintanilha J. A., 2014. Identifying concentrated areas of trip generators from high spatial resolution satellite images using object-based classification techniques. *Applied Geography*, 53, pp. 271-283.

Marques, F. M. S. F., 2006. Rates, patterns, timing and magnitude-frequency of cliff retreat phenomena. A case study on the west coast of Portugal. *Zeitschrift für Geomorphologie-Supplementbände*, 144, pp. 231–257.

Oyedotun, T. D. T., 2014. Shoreline Geometry: DSAS as a Tool for Historical Trend Analysis. *Geomorphological Techniques*, Chap. 3, Sec. 2.2 (2014), 12.

Pardo-Pascual J. E., Almonacid-Caballer J., Ruiz L. A., Palomar-Vázquez J., 2012. Automatic extraction of shorelines from Landsat TM and ETM+ multi-temporal images with subpixel precision. *Remote Sensing of Environment*, 123, pp. 1–11.

Schmitt, M., Lingyun, W., Xiao X. Z., 2015. Automatic Coastline Detection in Non-locally Filtered TANDEM-X Data. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2015, IEEE Xplore. IGARSS 2015*, 26.-31. Jul. 2015, Mailand, Italien. pp. 1036-1039.

Thieler, E. R., Himmelstoss, E. A., Zichichi, J. L., Ergul, A., 2009. Digital Shoreline Analysis System (DSAS) version 4.0 - An ArcGIS extension for calculating shoreline change. *U.S. Geological Survey Open-File Report*. pp. 2008-1278.

Yousef A., Iftekharuddin K., 2014. Shoreline extraction from the fusion of LiDAR DEM data and aerial images using mutual information and genetic algorithms. In: *2014 International Joint Conference on Neural Networks (IJCNN)*. Beijing, China, pp. 1007–1014.

Yu S., Mou Y., Xu d., You X., Zhou L., Zeng W., 2013. A New Algorithm for Shoreline Extraction from Satellite Imagery with Non-Separable Wavelet and Level Set Method. *International Journal of Machine Learning and Computing*, 3(1), pp. 158-163.

Zheng G., Peng L., Tao G., Wang C., 2011. Remote sensing analysis of Bohai Bay West Coast shoreline changes. In: *Spatial Data Mining and Geographical Knowledge Services (ICSDM), 2011 IEEE International Conference on*, Fuzhou, China, pp. 549-552.