

EFFECT OF DIFFERENT SEGMENTATION METHODS USING OPTICAL SATELLITE IMAGERY TO ESTIMATE FUZZY CLUSTERING PARAMETERS FOR SENTINEL-1A SAR IMAGES

B. Bayram^{1,*}, N. Demir², B. Akpınar¹, Selen Oy², F. Erdem¹, T. Vögtle³, D. Z. Seker⁴

¹ Yildiz Technical University, Department of Geomatics Engineering, 34220 Esenler Istanbul, Turkey - (bayram, bakpınar@yildiz.edu.tr; firaterdem5@gmail.com)

² Akdeniz University, Space Science and Technologies, 07058 Antalya, Turkey- nusretdemir@akdeniz.edu.tr; selenoy171@gmail.com

³ Karlsruhe Institute of Technology (KIT), Institute of Photogrammetry and Remote Sensing, D-76128 Karlsruhe, Germany- thomas.voegtle@kit.edu;

⁴ Istanbul Technical University, Department of Geomatics Engineering, 80626 Maslak Istanbul, Turkey – seker@itu.edu.tr

Commission I, WG I/6

KEY WORDS: Mean-Shift, Random Forest, Whale Optimization, Image Segmentation, Fuzzy Clustering

ABSTRACT:

Optical and SAR data are efficient data sources for shoreline monitoring. The processing of SAR data such as feature extraction is not an easy task since the images have totally different structure than optical imagery. Determination of threshold value is a challenging task for SAR data. In this study, SENTINEL-2A optical data was used as ancillary data to predict fuzzy membership parameters for segmentation of SENTINEL-1A SAR data to extract shoreline. SENTINEL-2A and SENTINEL-1A satellite images used were taken in September 9, 2016 and September 13, 2016 respectively. Three different segmentation algorithms which are selected from object, learning and pixel-based methods. They have been exploited to obtain land and water classes which have been used as an input data for parameter estimation. Thus, the performance of different segmentation algorithm has been investigated and analysed. In the first step of the study, Mean-Shift, Random Forest and Whale Optimization algorithms have been employed to obtain water and land classes from the SENTINEL-2A image. Water and land classes derived from each algorithm – are used as input data, and then the required parameters for the fuzzy clustering of SENTINEL-1A SAR image, were calculated. Lake Constance, Germany has been chosen as the study area. In this study, additionally an interface plugin has been developed and integrated into the open source Quantum GIS software platform. The developed interface allows non-experts to process and extract the shorelines without using any parameters. But, this system requires pre-segmented data as input. Thus, the batch process calculates the required parameters.

1. INTRODUCTION

Shorelines have a dynamic characteristic, and changes with natural or human induced effects. Remote sensing and image processing techniques are useful and modern tools to use for shoreline monitoring and change detection (Kutser et al., 2012). Shoreline can be extracted using different algorithms which have been proposed for temporal monitoring of coastal regions. (Gens, 2010). Shoreline extraction problem is a popular topic and there is considerable amount of researches related to this issue. Some of these studies related to application of unsupervised methods (Guariglia et al., 2006), some are related to application of water based indices (Zheng et al., 2011) and some of them deal with the application of morphology (Pardo-Pascual et al., 2012). Active contours have also been applied by (Schmitt et al., 2015). Particle Swarm optimisation, Mean-Shift and object-oriented fuzzy classification approaches were also used in different studies (Bayram et al. 2016, Catal Reis, et al., 2016, Bayram et al., 2017).

On the other hand, SAR images are more advantageous with their capability to acquire data in all weather conditions than optical data. Fuzzy clustering based shoreline extraction from SAR images have been realized in several studies (Demir et al., 2016, 2017). The processing and classification of SAR images

required several parameters. Primarily results from multispectral images can be used as training data sets to estimate these parameters. In this study, different methods for shoreline extraction from multispectral image were investigated. These are Mean-Shift, Random Forest and Whale Optimisation. The results from these methods have been used as input data to estimate the fuzzy clustering parameters for SENTINEL-1A image classification. After this process, obtained shorelines were compared with manually digitized shoreline from SENTINEL-2 image.

2. USED DATA AND METHODOLOGY

Lake Constance which is located on the Rhine at the northern foot of the Alps was selected as study area (Figure 1). This lake is situated a transboundary lake by Germany, Switzerland, and Austria (Hammerl, 2006).

* Corresponding author

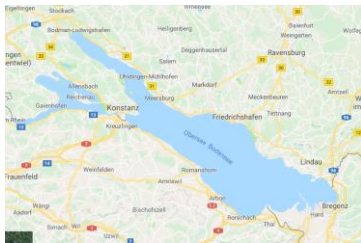


Figure 1. Lake Constance (Google maps)

The properties of SENTINEL-1A SAR and SENTINEL-2A data are given in Table 1 and Table 2 respectively.

Polarisation:	VH
Range and Azimuth Spacing:	10 m
Acquisition date	13.09.2016

Table 1. Properties of SENTINEL-1A SAR image (ESA, 2018a)

Spatial Resolution (m)	Band Number	S2A	
		Central Wavelength (nm)	Bandwidth (nm)
10	2	496.6	98
	3	560.0	46
	4	664.5	39
	8	835.1	133
20	5	703.9	20
	6	740.2	18
	7	782.5	28
	8a	864.8	32
	11	1613.7	141
60	12	2202.4	238
	1	443.9	45
	9	945.0	27
	10	1373.5	76

Table 2. Properties of SENTINEL-2A image, acquisition date 9 September 2016 (ESA, 2018b)

To obtain ancillary water and land classes, blue, red and near infrared bands of SENTINEL-2A image have been used as given in Figure 2.

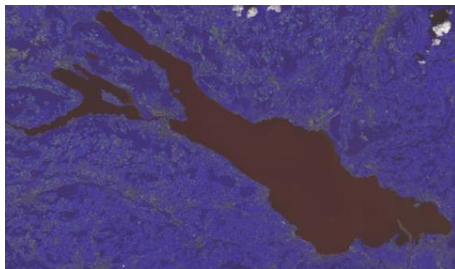


Figure 2. SENTINEL-2A image of Lake Constance

2.1 Shoreline segmentation by Mean-Shift Method

The Mean-Shift algorithm is an object-based method which has been proposed first by (Fukunaga and Hostetler., 1975) and modified by (Comaniciu and Meer, 2002). Mean shift is a nonparametric method for kernel density estimation and finds the maxima of a local neighbourhood. According to this method, feature space is considered as a probability density function (pdf). Regions with dense points in feature space corresponds to local maxima or modes (Qin, 2015). So, for each

data point, gradient ascent on the local estimated density until convergence was performed (Zhang, et al., 2012).

(Comaniciu and Meer, 2002) suggested RGB to L^*u^*v color conversion in the first step of Mean-Shift segmentation. But, (Chauhan and Shahabade, 2014) have proposed RGB to HSV transformation.

In this presented study, (Comaniciu and Meer, 2002)'s proposal has been considered. EmguCV library was used to implement Mean-Shift algorithm in .NET environment. The used parameters are given in Table 3.

Parameters	Value
Spatial Window Radius	10
Color Windows Radius	5
Maximum Iterations	100
Minimum Segment Size	25

Table 3. Used Mean-Shift parameters

After thresholding, Mean-Shift segments were created to obtain binary water and land classes. Threshold values have been defined as 0-255, 0-255 and 1-2 for blue, red and NIR bands respectively. Segmentation result is given in Figure 3, a and obtained binary water and land classes are given in Figure 3, b.

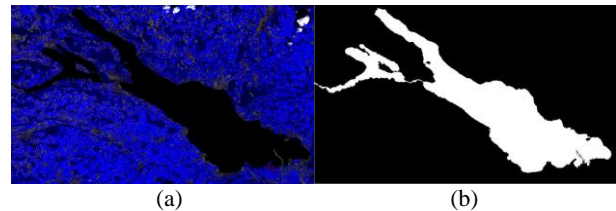


Figure 3. (a) Mean-Shift segmentation, (b) and thresholding results

2.2 Shoreline segmentation by Random Forest Method

Random Forest is a learning based classification algorithm which is based on decision trees. This method creates determined object classes by analysing of given training data sets (Breiman, 2001). Two parameters are required for this algorithm, number of trees and the number of random variables to be used for each node for creation of decision trees (Belgiu and Dragut, 2016). The algorithm creates Multiple CART-like trees in training step (Breiman, 2001). Bootstrapped technique is used for determination of split for each node (He et al., 2015). The best split is determined by GINI index according to CART algorithm. Thus, for each node, from randomly selected variables, the homogeneity of samples is measured. The variable which corresponds to minimum GINI index is selected and calculations are repeated for next node. GINI index is resulted with zero, related not is evaluated as totally homogeneous and defined as end-of-branching. The out of bag samples (test data) are cross validated for each tree. Classification vote is calculated by considering weight of decision tree and the related pixel is assigned to the majority voted class (Gislason et al., 2006).

Used SENTINEL-2A image consists of 4072 x 6951 pixels. In the training step, totally 1200000 pixels are selected for training. 600000 pixels have been selected for water and 600000 pixels of land classes. Statistics and Machine Learning Toolbox of MATLAB environment has been used for

implementation of the Random Forest algorithm. The used parameters are given in Table 4. Blob analysis has been utilized after classification for noise removal. Classification result and obtained water and land classes after noise removal have been given in Figure 4-a, and Figure 4-b.

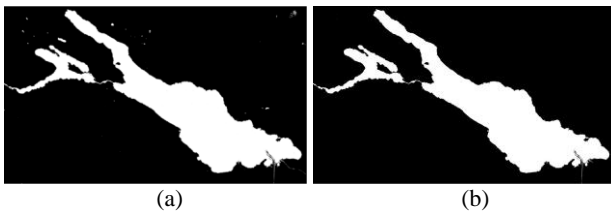


Figure 4. (a) Random Forest classification, (b) and noise removal

2.3 Shoreline segmentation by Whale Optimisation Algorithm

Whale Optimisation Algorithm is a multi-level thresholding technique which divides an image into multiple regions to perform segmentation. Automatic selection of optimum thresholds is one of the biggest challenges in image segmentation (Muangkote et al., 2016). Multilevel thresholds are generated using either the Otsu or the Kapur entropy function by the method (Bhandari et al., 2014). It is based on social behaviour of whales (Mirjalili and Lewis 2016). Determination of optimum threshold values is done by optimization of fitness function (Aziz et al., 2017). Each whale represents a set of solutions which consists of threshold values. First the threshold values are determined randomly between minimum and maximum values of the image histogram for each solution set. The whale which take the highest value according to the fitness function is set. The location of whales is updated according to their encircling and bubble-net behaviour. This process is iteratively maintained (Mirjalili ve Lewis, 2016). In this study, Whale Optimization Algorithm was utilized using MATLAB platform. The used parameters are given in Table 4.

Parameters	Value
Count of Whales	100
Maximum iteration	100
Number of thresholds	2

Table 4. Used Whale Optimization parameters

Otsu function has been used as the fitness function. Since two threshold values have been created for each band, totally 27 classes have been generated.

As there are 2 threshold values in each band, a total of 27 classes are obtained. Figure 4 shows the resulting segmented image of the Whale Optimization Algorithm.

Thresholding has been applied on the segmented image. Threshold values have been defined as 3-143, 2-145 and 4-14 for blue, red and NIR bands respectively. Blob analysis has been utilized after classification for noise removal. Thresholding result and obtained water and land classes after noise removal are given in Figure 5-a, and Figure 5-b.

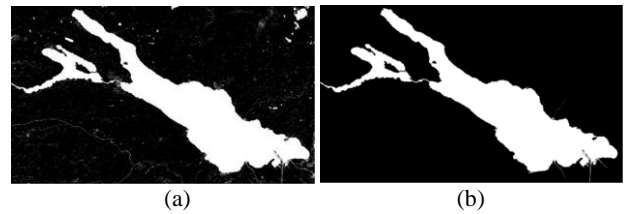


Figure 5. (a) Whale Optimisation, (b) noise removal

2.4 Shoreline extraction from SENTINEL-1A image

Lee filter has been applied on SENTINEL-1A image for speckle reduction as given in Figure 6.

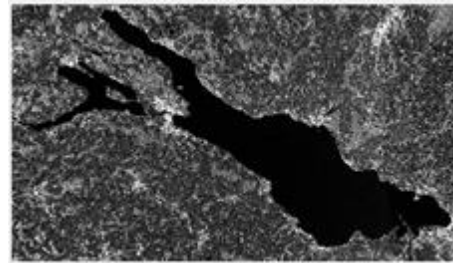


Figure 6. Used SAR data (after pre-processing)

Land and water memberships have been defined for clustering of SENTINEL-1A image. Because the mean-standard values between the classes expected as large, MS Large membership function has been selected. The membership equations for water has been defined in equation (1) as follows:

$$\mu(x) = 1 - (bs / (am + bs)) \text{ if } x < am \text{ else } \mu(x) = 0 \quad (1)$$

The membership for the land is given with the equation 2 as follows;

$$\mu(x) = (bs / (am + bs)) - 1 \text{ if } x > am \text{ else } \mu(x) = 0 \quad (2)$$

After definition of memberships, the centroid method was used to determine the threshold for defuzzification (Figure 7). The used parameters mean, standard deviation a and b are derived from the results of SENTINEL-2 classification.

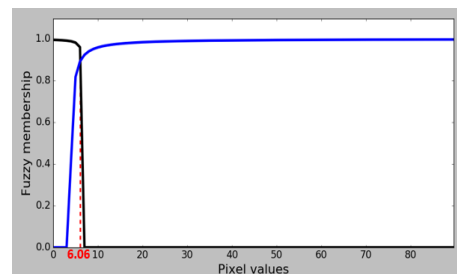


Figure 7. Centroid method for the defuzzification (Black: Water, Land: Blue)

Quality assessment was performed with comparison manually digitized shoreline from SENTINEL-2 data.

3. RESULTS

Water and land segmentation results from Mean-Shift, Random Forest and Whale Optimization methods shown were used to calculate the defuzzification parameters as shown in Table 5.

	Land			Water		
	Mean-Shift	Random Forest	Whale Opt.	Mean-Shift	Random Forest	Whale Opt.
Mean	27.21	27.13	27.19	3.34	3.27	3.49
Median	241.28	241.11	241.35	2.97	1.90	5.45
Standard Dev.	20.42	20.42	20.41	3.06	3.05	3.06
a	1.97	2.06	1.939	0.01	0.01	0.02
b	0.01	0.02	0.01	0.001	0.001	0.001

Table 5. Calculated parameters for fuzzy clustering

After calculation of the parameters, the defuzzification thresholds have been calculated from Mean-Shift segmentation, Random Forest and Whale Optimization methods as 6.06, 5.92 and 6.0 respectively as given in Figure 8-a, b, c. Segmented images and shoreline extraction results by Mean-Shift segmentation, Random Forest and Whale Optimization methods are given in Figure 9-a, b, c and Figure 10-a, b, c respectively.

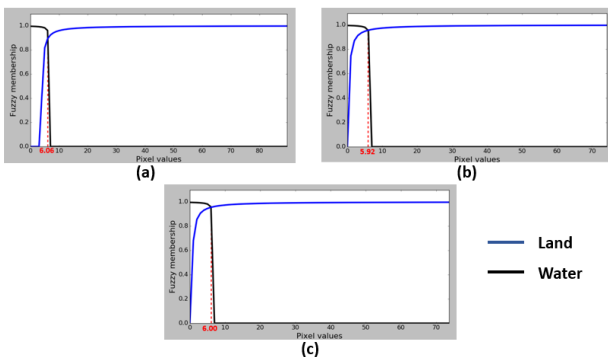


Figure 8. Membership of land and water, calculating the threshold with centroid defuzzification. (a) Using Mean-Shift segmentation results, (b) Using Random Forest segmentation results, (c) Using Whale Optimization segmentation results

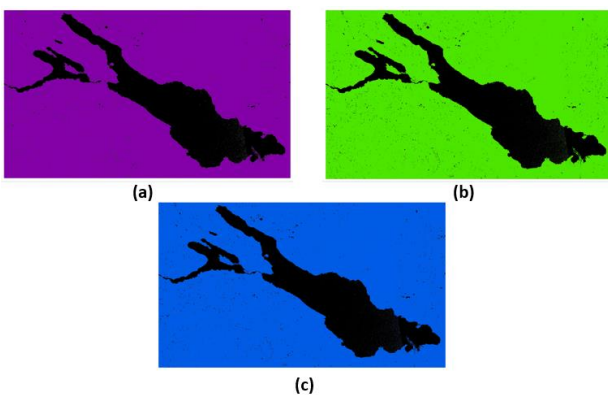


Figure 9. Segmentation results (a) Using Mean-Shift segmentation results, (b) Using Random Forest segmentation results, (c) Using Whale Optimization segmentation results

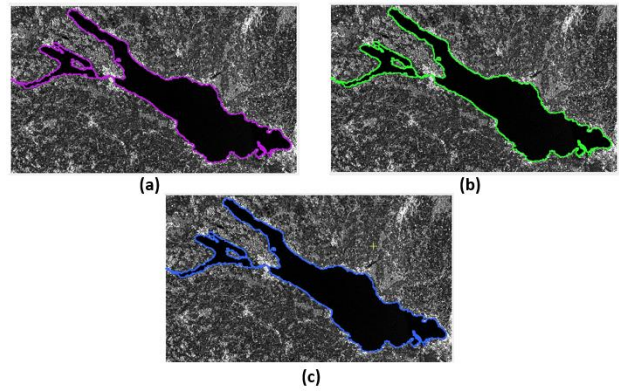


Figure 10. Shoreline extraction results. (a) Using Mean-Shift segmentation results, (b) Using Random Forest segmentation results, (c) Using Whale Optimization segmentation results

The derived results from each method were compared with manually digitized shorelines. The perpendicular distances between reference data and shoreline extraction results have been calculated using 10 m space length which is the spatial resolution of SENTINEL-1A SAR image. Statistics of the perpendicular distances to the reference shoreline were computed. The calculated values are listed in Table 6.

	Mean-Shift	Random Forest	Whale Optimisation
Mean	26.53 m	12.50 m	32.56 m
Std.Dev.	96.69 m	27.05 m	89.16 m
Median	6.31 m	6.05 m	16.03 m

Table 6. Quality assessment statistics of the shorelines derived from fuzzy clustering with use of estimated parameters.

4. CONCLUSIONS

In this study, shorelines were extracted from SENTINEL-1A SAR image with estimated parameters from SENTINEL-2A multispectral image classification results. As shown in the accuracy assessment, the random forest algorithm has the best performance among other methods for the estimation of the parameters used in SAR data fuzzy clustering. According to the outcomes of the study, it could be confirmed that the accuracy of the segmentation plays a key role for parameter estimation. Mean-Shift and Whale Optimization methods require definition of thresholds for post processing which is essential for these methods. Random Forest approach requires well collected training data which affects the results directly. The number of trees is another important parameters for this method.

ACKNOWLEDGEMENTS

This study has been supported by "TUBITAK (The Scientific and Technological Research Council of Turkey) with project number 115Y718 and entitled" Integration of Unmanned Aerial Vehicles for Sustainable Coastal Zone Monitoring Model – Three-Dimensional Automatic Coastline Extraction and Analysis: Istanbul-Terkos Example ".

REFERENCES

- Aziz, M. A., Ewees, A. A., Hassanien, A. E., 2017. Whale Optimization Algorithm and Moth-Flame Optimization for multilevel thresholding image segmentation. *Expert Systems with Applications*, 83, pp. 242-256.
- Bayram, B., Demir, N., Seker, D.Z., Ogurlu, M., Oy, S., Bozkurt, S., Ince, A., 2016. Shoreline extraction from LIDAR-Intensity data. In: *World GIS day*, pp. 43.
- Bayram, B., Avşar, Ö., Şeker, D. Z., Kayi, A., Erdogan, M., & Eker, O., 2017. The role of national and international geospatial data sources in coastal zone management. *Fresenius Environmental Bulletin (FEB)*, 26, pp. 383–391.
- Belgiu, M., Drăguț, 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.
- Bhandari, A. K., Singh, V. K., Kumar, A., Singh, G. K., 2014. Cuckoo search algorithm and wind driven optimization based study of satellite image segmentation for multilevel thresholding using Kapur's entropy. *Expert Systems with Applications*, 41(7), pp. 3538-3560.
- Breiman, L., 2001. Random forests. *Machine Learning*, 45(1), pp. 5-32.
- Chauhan, P., Shahabade, R. V., 2014. Mean shift segmentation on Enhanced Color image. *IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN*, Vol. 3, pp. 2278-0661.
- Comaniciu, D., Meer, P., 2002. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 24(5), pp. 603-619.
- Catal Reis H., Bayram B., 2016. Shoreline extraction using Particle Swarm Optimization Method: Case study of Lake Ercek, Van-Turkey. In: *VI. UZAL-CBS Symposium*, Adana, Turkey, pp.1-5.
- Demir, N., Kaynarca, M., Oy, S., 2016. Demir, N., Kaynarca, M., Oy, S., 2016. Extraction of coastlines with fuzzy approach using SENTINEL-1 SAR image. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XLI-B7, pp. 747-751.
- Demir, N., Oy, S., Erdem, F., Şeker, D. Z., & Bayram, B., 2017. Integrated shoreline extraction approach with use of RASAT MS and Sentinel-1A SAR images. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. IV-2/W4, pp. 445-449.
- European Space Agency (ESA), IW GRD Resolutions, 2018a. <https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-1-sar/products-algorithms/level-1-algorithms/ground-range-detected/iw> (Last access 01.08.2018).
- European Space Agency (ESA), Resolution and Swath, 2018b. <https://sentinel.esa.int/web/sentinel/missions/sentinel2/instrument-payload/resolution-and-swath> (Last access 01.08.2018).
- Fukunaga, K., Hostetler, L., 1975. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on information theory*, 21(1), pp. 32-40.
- 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(4), pp. 294-300.
- Guariglia, A., Buonamassa, A., Losurdo, A., Saladino, R., Trivigno, M.L., Zaccagnino, A., Colangelo, A., 2006. A multisource approach for coastline mapping and identification of shoreline changes. *Annals of geophysics*, 49, 295–304.
- Hammerl, M., 2006. Lake Constance. *Experience and Lessons Learned Brief*, pp. 145-157 http://www.worldlakes.org/uploads/10_lake_constance_27february2006.pdf (Last access: 10.07.2018).
- 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.
- Kutser, T., Paavel, B., Verpoorter, C., Kauer, T., Vahtmäe, E., 2012. Remote Sensing of Water Quality in Optically Complex Lakes. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXIX-B8, pp. 165-169.
- Mirjalili, S., Lewis, A., 2016. The whale optimization algorithm. *Advances in Engineering Software*, 95, pp. 51-67.
- Muangkote N., Sunat K., Chiewchanwattana S., 2016. Multilevel thresholding for satellite image segmentation with moth-flame based optimization. In: *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, Khon Kaen, Thailand, pp. 1-6.
- 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.
- Qin, R., 2015. A mean shift vector-based shape feature for classification of high spatial resolution remotely sensed imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(5), pp. 1974-1985.
- Schmitt, M., Wei, L., Zhu, X. X., 2015. Automatic coastline detection in non-locally filtered tandem-X data. In: *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Milan, Italy, pp. 1036-1039.
- Zhang, X., Cui, Y., Li, D., Liu, X., Zhang, F., 2012. An adaptive mean shift clustering algorithm based on locality-sensitive hashing. *Optik-International Journal for Light and Electron Optics*, 123(20), pp. 1891-1894.
- Zheng, G., Peng, L., Tao, G., Wang, C., 2011. Remote sensing analysis of Bohai Bay West Coast shoreline changes. In: *2011 IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM)*, Fuzhou, China, pp. 549-552.