Monitoring of Coastal Environments Using Data Mining

Corneliu Octavian Dumitru, Gottfried Schwarz, and Mihai Datcu
Remote Sensing Technology Institute, German Aerospace Center (DLR),
82234 Wessling, Germany
email: corneliu.dumitru@dlr.de, gottfried.schwarz@dlr.de, mihai.datcu@dlr.de

***Abstract*—Current satellite images provide us with detailed information about the state of our planet, as well as about our technical infrastructure and human activities. A range of already existing commercial and scientific applications try to analyze the physical content and meaning of satellite images by exploiting the data of individual, multiple or temporal sequences of images. However, what we still need today are advanced tools to automatically analyze the image data in order to extract and understand their full content. In this paper, we propose a highly automated approach for application-adapted image content exploration, targeting coastal environmental monitoring. For the selected coastal areas, different use cases can be considered such as: detection of wind turbines *vs.* boats, differences between beaches, tidal flats, and dams, and identification of fish cages/aquaculture. The average accuracy is ranging from 80% to 95% depending on the satellite images.**

Keywords- coastal monitring; data mining; Sentinel-1; Sentinel-2; TerraSAR-X.

I. INTRODUCTION

In Earth observation, a very popular satellite image analysis system is the one from Digital Globe, named Tomnod, or Google Earth together with its related tools, which are targeting general user topics. In the Earth observation (EO) domain, there are systems such as LandEX [1], which is a land cover analysis system, while GeoIRIS [2] is a system that allows the user to refine a given query by iteratively specifying a set of relevant, and a set of non-relevant images. A similar information retrieval system is IKONA [3], which is using relevance feedback in order to exploit very high resolution EO images. Further, the Knowledge-driven Information Mining (KIM) system [4] is an example of an active learning system providing semantic interpretation of image content. The KIM concept evolved into the TELEIOS prototype [5], complementing the scope of searching for EO images with additional geo-information and in-situ data integrated into an operational EO system [6] to interpret TerraSAR-X images. Similar concept with KIM concept is presented in [29] while in [30] a data mining approach for Big Data is described.

The proposed system is very fast compared with the existing systems and with only few examples can retrieve the desired category with higher accuracy. The diversity of applications that can be considered for such systems are rather broad and include, for instance, coastal environmental monitoring (sea level, tides and wave direction), land cover/use changes, disaster monitoring, forest management, ice monitoring, monitoring of active volcanoes, waste deposit site management, traffic monitoring, vegetation monitoring, urban sprawl, soil moisture dynamics, etc.

The paper is organized as follows. Section II describes the selected test areas. Section III presents our datasets. Section IV details the data mining methodology applied in this paper. Section V shows the results and we conclude the paper in Section VI. The acknowledgements close the paper.

II. SELECTION OF TEST AREAS, USE CASES, AND APPLICATIONS

We emphasize here three use cases for monitoring coastal environments. For these use cases, we selected for our investigation the Wadden Sea with the Dutch Delta (in the Netherlands), the Danube Delta (in Romania), and the Curonian Lagoon (in Lithuania and Russia) which are internationally recognized protected areas as UNESCO (United Nations Educational, Scientific and Cultural Organization) Natural Heritage sites.

A. The Wadden Sea, Netherlands

Site description: The Wadden Sea (Dutch: Waddenzee, German: Wattenmeer, Danish: Vadehavet) is an intertidal zone in the south-eastern part of the North Sea. It lies between the coast of N-W continental Europe and the range of Frisian Islands, forming a shallow body of water with tidal flats and wetlands [7], protected by a 450 km long chain of barrier islands, the Wadden Islands. The Wadden Sea region measures about 22,000 km^2, divided between land and sea. About 63% of the region lies in Germany, with about 30% in the Netherlands, and 7% in Denmark [8]. In 2009, the Dutch-German Wadden Sea was inscribed on the UNESCO World Heritage List and the Danish part was added later in 2014.

The landforms in the Wadden Sea region have essentially been created from a marine or tidal environment [9].

Typical for the Wadden Sea are large tidal flats, which are characterized by very high benthic biomass and productivity, dominated by molluscs and polychaetes.

State-of-the-art publications: In the research literature there are several studies treating the Wadden Sea area along the years. In order to understand the Wadden Sea dynamics, a number of recent publications [10]-[13] already used remote sensing images and addressed the issue of Synthetic Aperture Radar (SAR) satellite image classification and interpretation in these areas. At present, the option of data fusion from different sensor has not yet been fully exploited.

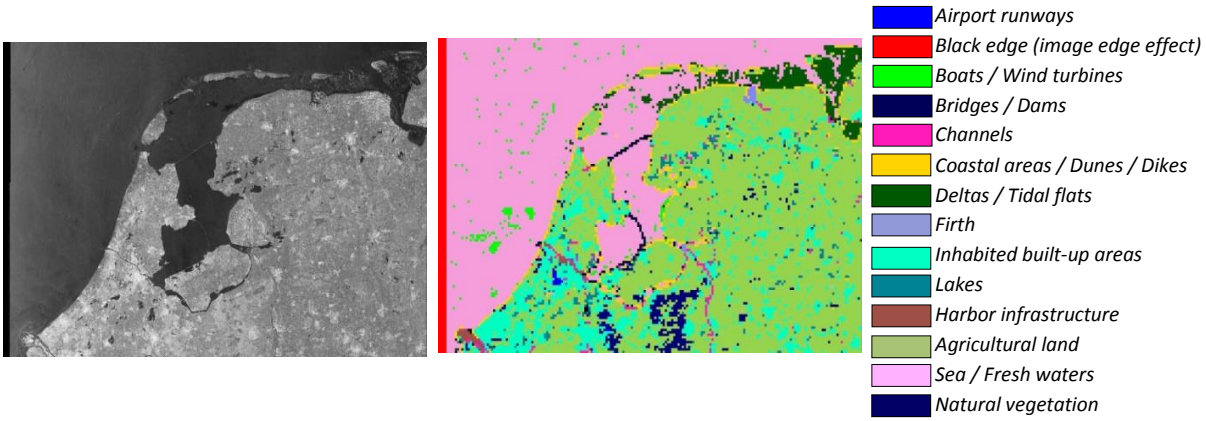

Figure 1. Sentinel-1A quick-look view (left) and classification map (right) for an image of the Wadden Sea, Lake IJssel, and Marker Lake, and the surrounding areas.

Image interpretation goal: The Wadden Sea area faces a strong economic impact due to recreation, fisheries and maritime traffic. The last impact is due to, e.g. the ports of Bremerhaven, Hamburg, and Rotterdam whereby the traffic runs through or nearby this area, which makes that monitoring of sand banks and any decrease of the water depth and the tide levels in this area is a critical topic for maritime security. A second important topic is the monitoring of biodiversity as described by [14].

Typical examples: The diversity of categories identified from a single image and a typical classification map of the Wadden Sea and its surrounding areas are shown in Figures 1 and 2.

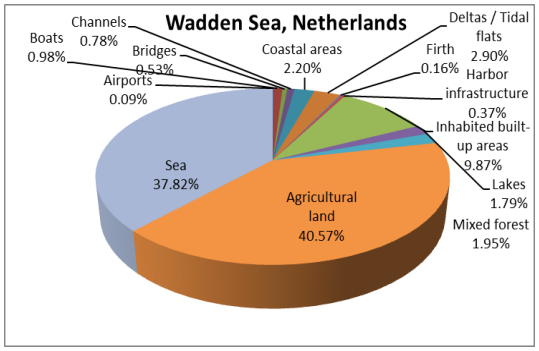

Figure 2. Diversity of categories identified from a single image of the Wadden Sea, the Netherlands.

B. *The Danube Delta, Romania*

Site description: The Danube Delta is the second largest river delta in Europe and is the best preserved one on the continent [15].

Formed over a period of more than 10,000 years, the Danube Delta continues to grow due to the 67 million tons of alluvia deposited every year by the Danube River [16]. The delta is an ideal test and validation area for vegetation monitoring as it is characterized by high biodiversity and various crops.

The Delta is formed around the three main channels of the Danube, named after their respective ports Chilia (in the north), Sulina (in the middle), and Sfantu Gheorghe (in the south).

The greater part of the Danube Delta lies in Romania (Tulcea County), while its northern part, on the left bank of the Chilia arm, is situated in Ukraine (Odessa Oblast). Its total surface is 4,152 km^2 of which 3,446 km^2 are in Romania. The waters of the Danube, which flow into the Black Sea, form the largest and best preserved delta in Europe. In 1991, the Danube Delta was inscribed on the UNESCO World Heritage List due do its biological uniqueness.

State-of-the-art publications: In the image processing literature there are not many studies treating the Danube Delta especially for SAR data [17]-[19]. However, the monitoring of biodiversity from in-situ measurements has attracted more interest [20].

Image interpretation goal: At the mouth of the Danube, the alluvial discharge decreases every year from 81 million tons in 1894, to 70 million tons in 7939, 58 million tons in 1982, and about 22 million tons in 2015. This makes it interesting to monitor the evolution of the alluvial discharge and to investigate its impact on the Danube Delta and the three channels together with their ports (Chilia, Sulina, and Sfantu Gheorghe) through the years.

The data can be combined with other types of information, such as the volume of water of each channel in order to prepare risk flood maps needed for the safety of the

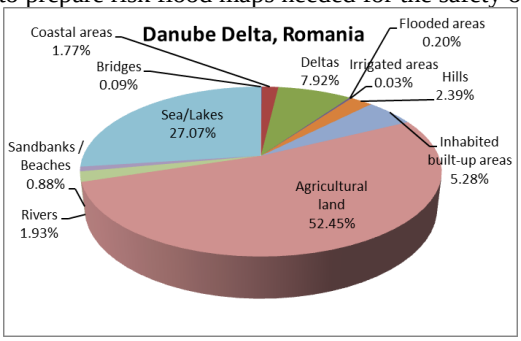

Figure 3. Diversity of categories identified from a single image of the Danube Delta.

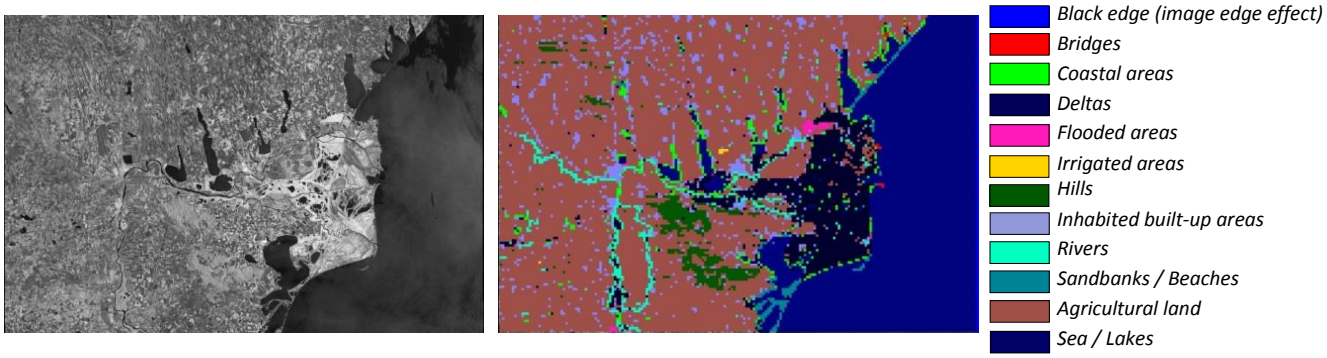

Figure 4. Sentinel-1A quick-look view (left) and classification map (right) for an image of the Danube Delta and the surrounding areas.

shipping traffic and also for the local authorities to protect the human settlements. Another image interpretation goal is vegetation monitoring, in particular, biodiversity issues and crop type analyses.

Typical examples: The diversity of categories identified from a single image and a typical classification map of the Danube Delta and its surrounding areas are shown in Figures 3 and 4.

C. The Curonian Lagoon, Lithuania and Russia

Site description: The Curonian Lagoon is the largest European lagoon. Situated in the southern part of the Baltic Sea with a total area of 1584 km^2, the lagoon receives water from the River Nemunas. The salinity of the water is higher and fluctuates between the northern and southern part of the lagoon [14]. The entire Lithuanian part of the Curonian Lagoon has been designated as a NATURA 2000 area and in 2000 the Curonian Spit cultural landscape was as well inscribed on the UNESCO World Heritage List.

State-of-the-art publications: In the remote sensing literature, there are not many studies treating the Curonian Lagoon especially for SAR data. However, the monitoring of biodiversity has attracted greater interest [21]-[23].

Image interpretation goal: We analyzed the effect of socio-economic activities of the area regarding: the ceasing commercial fisheries, the prohibition of the extraction of mineral resources, the agricultural sector, the hunting sector, the restriction of recreational use of the aquatic areas, and the oil drilling/pollution of the area.

Typical examples: The diversity of categories identified from a single image and a typical classification map of the Curonian Lagoon and its surrounding areas are shown in Figures 5 and 6.

III. DATASETS

An important aspect to be addressed is the creation of a reference dataset for test and validation of the different systems. We already possess an initial synthetic aperture radar dataset composed of 1000 TerraSAR-X images and 100 Sentinel-1 images covering target areas from around the world.

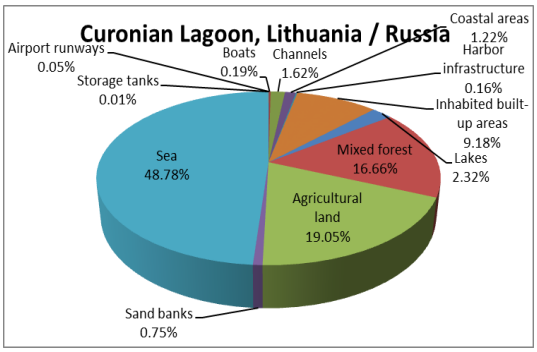

Figure 5. Diversity of categories identified from a single image of the Curonian Lagoon.

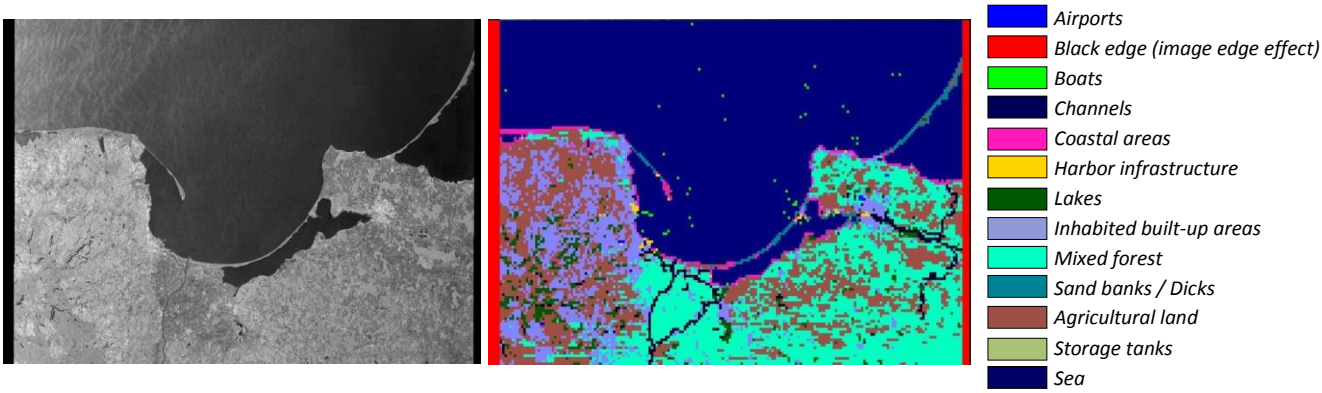

Figure 6. Sentinel-1A quick-look view (left) and classification map (right) for an image of the Curonian Lagoon and the surrounding areas.

From this database, about 295 TerraSAR-X and 25 Sentinel-1A images have already been annotated by a remote sensing expert using a semi-automatic semantic annotation system resulting in a semantic catalogue of hundreds of semantic labels grouped in a 3-level hierarchical scheme [24]. This annotated database mainly covers urban and industrial areas together with their infrastructure predominantly from Europe, and can be considered as our initial ground truth dataset [25].

Our latest dataset also contains optical satellite data with multi-spectral images (e.g., Sentinel-2A), and synthetic aperture radar images (e.g., TerraSAR-X and Sentinel-1A / 1B). These data cover 10 protected areas from Europe (national parks, mountains, arid and semi-arid areas, and coastal and marine ecosystems) [14].

IV. METHODOLOGY

The data mining system [6] (used in this paper) is composed of four main modules: Data Model Generation (DMG), Database Management System (DBMS), Knowledge Discovery in Databases (KDD), and Statistical Analytics (SA).

The DMG module transforms the original format of original Earth observation products into smaller and more compact product representations that include image descriptors, metadata, image patches, etc. The DBMS module is used for storing all the generated information and allows querying and retrieval of the available image data. The KDD module is in charge of finding patterns of interest from the processed data and presenting them to the user. Moreover, the KDD module allows annotating the image content by using machine learning algorithms and human interaction resulting in physical categories. The SA module provides classification maps of each dataset and distribution results of the retrieved categories in an image. These four modules are operated automatically and interactively with and without user interaction.

We summarize our data mining methodology as a pseudo-code segment in Table 1.

TABLE 1: THE PROPOSED METHODOLOGY.

Step 1: EO Dataset
Select and download typical EO images.
Step 2: Data Model Generation (DMG)
for each EO_i image (i=1…N) **do**
tile EO_i into $p_{i,j}$ patches (j=1…M), where the size of the patches depends on the image resolution
store all $p_{i,j}$ into the DBMS
for each $p_{i,j}$ patch **do**
extract an $f_{i,j}$ primitive feature vector from optical / SAR algorithms
//e.g., Gabor filters with 5 scales and 6 orientations and compute the means and standard deviations of the coefficients //
store all $f_{i,j}$ vectors into the DBMS
end
end

Step 3: Knowledge Discovery in Databases (KDD)
if r_k (k=1…K) ∄ **do** //if the patch reference label has not yet been generated//
for all $f_{i,j}$ primitive feature vectors **do**
group the $f_{i,j}$ into g_k clusters and group them into c_k categories using an SVM (Support Vector Machine)
for each c_k category **do**
select an r_k semantic annotation label //visual support via Google Earth //
store reference r_k labels into the DBMS
end
end
else // routine processing after label generation//
for all $f_{i,j}$ primitive feature vectors **do**
group the $f_{i,j}$ into g_l clusters (l=1…L) and group them into c_l categories using an SVM
store all g_l into the DBMS
for each c_l category **do**
select an a_l semantic annotation //visual support via Google Earth//
store a_l labels into the DBMS
end
end
end

Step 4: Statistical Analytics (SA)
for selected EO_i and its a_l **do**
generate classification maps
compare obtained a_l annotations with r_k //reference annotations (generated previously)// and *generate* change maps
compute characteristic metrics //e.g., precision/recall by comparing the results with the r_k //
end

V. RESULTS AND DISCUSSIONS

For the selected areas of interest, different use cases can be considered such as: *detection of wind turbines vs. boats*; *differences between beaches, tidal flats,* and *dams; fish cages/aquaculture;* etc.

For example, we selected the Wadden Sea area and we show the results for the d*etection of wind turbines vs. detection of boats*. The images were acquired in order to cover, as much as possible, the same area on the ground and/or the same acquisition date or a date closer between the acquisitions. The data set consists of different images acquired by three different satellites: a TerraSAR-X image acquired on May 13, 2015 with a resolution of 2.9 meters, a Sentinel-1A image acquired on May 15, 2015 with a resolution of 20 meters, and a Sentinel-2A single quadrant-image acquired on April 21, 2016 with a resolution of 10 meters (considering only the RGB bands). In Figure 7, we show the available data for the Wadden Sea protected area.

All these images were tiled into patches and from each patch a feature vector was extracted. We classified the images considering only two categories of interest, namely *Wind turbines* and *Boats* (see Figure 8). Based on the extracted features and the specific patterns of these categories we were able to separate them during classification. Figures 9, 10, and 11 illustrate the retrieved categories after the classification projected on the quick-look of each image product. For each image product the locations of *Wind turbines* and *Boats* are marked in green and blue, respectively.

The complete processing chain from ingestion to annotation was run on a desktop PC with software coded in Java 8 and Matlab R2105a. The PC used for our experiments had a processor clock rate of 2.40 GHz, and a RAM capacity of 8 GB. Typically, we obtain a CPU usage of less than 25% as we store all image files onto a disk and have to wait for the completion of all data transfers. The actual memory consumption of our PC configuration is less than 50 MByte per image. The classification and display of a new set of retrieved patches needs about 4 to 6 ms when we have a collection volume of 2 GByte of image data.

The accuracy of the results was computed for each sensor and for each retrieved category. For each image (EO_i) we compared the category al with its corresponding reference category r_k and we computed its classification accuracy. The attained average accuracy is 93%, ranging from 80% to 95% depending on the image type (e.g., TerraSAR-X, Sentinel-1A or Sentinel-2A). When we compare the different SAR sensors, we notice that the overall classification accuracy is higher for the high resolution instruments, for example for TerraSAR-X.

Figure 7. Locations of the Wadden Sea shown on OpenStreetMap; the TerraSAR-X footprints are in green, the Sentinel-1A footprint is in orange, and the Sentinel-2A footprint (all quadrants) is in blue.

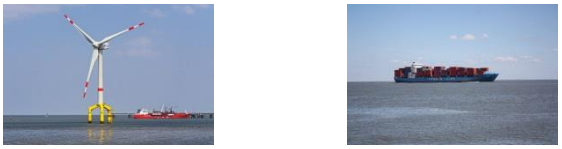

Figure 8. In-situ data: wind turbines *vs*. boats.

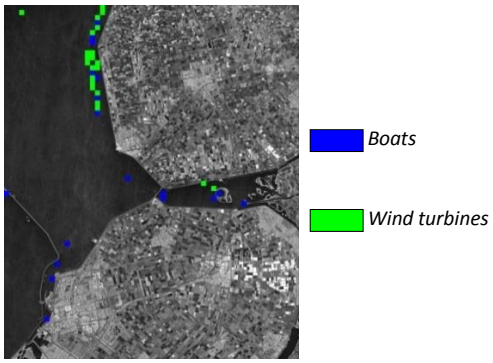

Figure 9. TerraSAR-X "patch-based" classification results projected on a SAR image of Flevoland, the Netherlands.

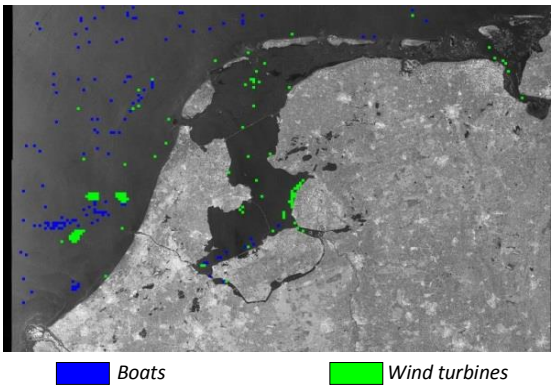

Figure 10. Sentinel-1A "patch-based" classification results projected on a SAR image of the Wadden Sea, Lake IJssel, and Marker Lake, and the surrounding areas in the Netherlands.

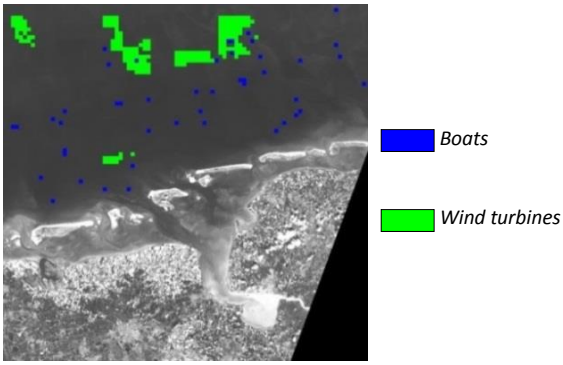

Figure 11. Sentinel-2A "patch-based" classification results projected on a (gray level) image of the German and Dutch Wadden Sea.

VI. Conclusions and Future Work

In this paper, we analyzed several protected areas all over Europe by a high- and a medium-resolution space-borne instrument (SAR and multi-spectral images).

By exploiting the specific imaging details and the retrievable semantic categories of these three image types (TerraSAR-X, Sentinel-1, and Sentinel-2), we can semantically fuse the image classification maps. In order to verify the classification results, we need to compare them with in-situ data.

Another example would be the *difference within beaches, tidal flats, and dams* that we can find from the images available in our dataset (e.g., the Danube Delta, the Curonian Lagoon, or the Wadden Sea). In this case, we obtained similar accuracy results with the case described in Section V.

For future evaluation, we plan to compare the classification accuracy of the wind turbines considering more parameters such as: the size of the pylon, the blade angles of the wind turbines, the rotation rate of the propeller, and the viewing direction and the resolution of the satellite image.

At this moment, there exist some studies about wind turbines [26]-[28] using SAR images but none of the existing papers analyzes all these parameters simultaneously. We will also compare the results from the point of view of accuracy between high-resolution *vs.* medium-resolution and between SAR sensors *vs.* multi-spectral sensors.

ACKNOWLEDGEMENTS

This work was supported by the H2020 ECOPOTENTIAL project. We thank the TerraSAR-X Science Service System for the provision of images (Proposals MTH-1118 and LAN-3156).

REFERENCES

[1] T. Stepinski, P. Netzel, and J. Jasiewicz, "LandEx-A GeoWeb Tool for Query and Retrieval of Spatial Patterns in Land Cover Datasets", IEEE JSTARS, 7(1), pp. 257-266, 2014.

[2] C.R. Shyu, M. Klaric, G. Scott, A. Barb, C. Davis, and K. Palaniappan, "GeoIRIS: Geospatial Information Retrieval and Indexing System – Content Mining, Semantics Modelling, and Complex Queries", IEEE TGRS, 45(4), pp. 839-852, 2007.

[3] N. Boujemaa, "Ikona: Interactive Specific and Generic Image Retrieval", in Proc. of MMCBIR, Glasgow, UK, pp. 1-4, 2001.

[4] M. Datcu et al., "Information Mining in Remote Sensing Image Archives: System Concepts", IEEE TGRS, 41(12), pp. 2923-2936, 2003.

[5] TELEIOS project [accessed March 2018]. Available: http://www.earthobservatory.eu/.

[6] EOLib project, [accessed March 2018]. Available: http://wiki.services.eoportal.org/tiki-index.php?page=EOLib.

[7] ECOSTRESS (Ecological Coastal Strategies and Tools for Resilient European Societies) project, [accessed March 2018]. Available: http://ecostress.eu/pilot-areas/dutch-german-wadden-sea/.

[8] Wadden Sea World Heritage, 2017. Available: http://www.waddensea-worldheritage.org/.

[9] K.S. Dijkema, J.H. Bossinade, P. Bouwsema, and R.J. de Glopper, "Salt Marshes in the Netherlands Wadden Sea: Rising High-Tide Levels and Accretion Enhancement", in "Expected Effects of Climatic Change on Marine Coastal Ecosystems", Kluwer Publishers, Dordrecht; pp 173-188, 1990.

[10] S. Brusch and S. Lehner, "Monitoring River Estuaries and Coastal Areas using TerraSAR-X", in Proc. of OCEANS, Bremen, Germany, pp. 1-4, 2009.

[11] S. Wiehle and S. Lehner, "Automated Waterline Detection in the Wadden Sea Using High-Resolution TerraSAR-X Images", Hindawi Journal of Sensors, pp. 1-6, 2015.

[12] G. Heygster, J. Dannenberg, and J. Notholt, "Topographic Mapping of the German Tidal Flats Analyzing SAR Images with the Waterline Method", IEEE TGRS, 48(3), pp. 1019-1030, 2010.

[13] M. Gade and S. Mechionna, "The Use of High-Resolution RADARSAT-2 and TerraSAR-X Imagery to Monitor Dry-Fallen Intertidal Flats", in Proc. of IGARSS, Quebec, Canada, pp. 1218-1221, 2014.

[14] ECOPOTENTIAL project, 2017. Available: http://www.ecopotential-project.eu/.

[15] Danube Delta, 2016. Available: http://romaniatourism.com/danube-delta.html.

[16] Danube Delta World Heritage, 2017. Available: http://whc.unesco.org/en/list/588.

[17] S. Niculescu, C. Lardeux, I. Grigoras, J. Hanganu, and L. David, "Synergy between LiDAR, RADARSAT-2, and Spot-5 Images for the Detection and Mapping of Wetland Vegetation in the Danube Delta", IEEE JSTARS, 9(8), pp. 3651-3666, 2016.

[18] M. Mierla, G. Romanescu, I. Nichersu, and I. Grigoras, "Hydrological Risk Map for the Danube Delta – A Case Study of Floods Within the Fluvial Delta", IEEE JSTARS, 8(1), pp.98-104, 2015.

[19] R. Tanase, A. Radoi, M. Datcu, and D. Raducanu, "Polarimetric SAR Data Feature Selection using Measures of Mutual Information," in Proc. of IGARSS, Milan, Italy, pp. 1140-1143, 2015.

[20] P. Gastescu, "The Danube Delta Biosphere Reserve. Geography, Biodiversity, Protection, Management", Romanian Journal of Geography, 53(2), pp. 139-152, 2009.

[21] D. Vaičiūtė, I. Olenina, R. Kavolytė, I. Dailidienė, and R. Pilkaitytė, "Validation of MERIS chlorophyll a products in the Lithuanian Baltic Sea case 2 coastal waters", in Proc. of IEEE/OES Baltic International Symposium (BALTIC), Riga, Latvia, pp.1-2 , 2010.

[22] G. Garnaga and Z. Stukova, "Contamination of the south-eastern Baltic Sea and the Curonian Lagoon with oil products," in Proc. of IEEE/OES US/EU-Baltic International Symposium, Tallinn, Estonia, pp. 1-8, 2008.

[23] S. Gulbinskas, E. Trimonis, and I. Mineviciute, "Sedimentary fluxes in the marine-lagoon (Baltic sea – Curonian Lagoon) connection," in Proc. of IEEE/OES Baltic International Symposium (BALTIC), Riga, Latvia, pp. 1-6, 2010.

[24] C.O. Dumitru, G. Schwarz, and M. Datcu, "Land Cover Semantic Annotation Derived from High Resolution SAR Images", IEEE JSTARS, 9(6), pp. 2215-2232, 2016.

[25] C. Dumitru, G. Schwarz, and M. Datcu, "SAR Image Land Cover Datasets for Classification Benchmarking of Temporal Changes", IEEE JSTARS, 11(5), pp.1-21, 2018.

[26] C. Clemente and J.J. Soraghan, "Analysis of the effect of wind turbines in SAR images," in Proc. of IET International Conference on Radar Systems (Radar 2012), Glasgow, UK, pp. 1-4, 2012.

[27] T. Cuong, "Radar cross section (RCS) simulation for wind turbines", Master Thesis, Naval Postgraduate School, California, 94 pages, 2013.

[28] M.B. Christiansen and Ch.B. Hasager, "Wake effects of large offshore wind farms identified from satellite SAR", Remote Sensing of Environment, 98(2-3), pp. 251-268, 2005.

[29] J. Zhang, W. Hsu, and M.L. Lee, "Image Mining: Trends and Developments", 2002. Available: http://www.comp.nus.edu.sg/~whsu/publication/2002/JIIS.pdf.

[30] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, "Data Mining with Big Data", IEEE TKDE, 26(1), pp. 97-107, 2014.