University „POLITEHNICA" of BUCHAREST
ETTI-B DOCTORAL SCHOOL

Summary of the
DOCTORAL THESIS

METHODS AND ALGORITHMS FOR INFORMATION RETRIEVAL AND SEARCH IN SATELLITE IMAGE TIME SERIES

PhD. Student: **Cpt. Eng. Alexandru-Cosmin Grivei**
Advisor: **Prof. Dr. Eng. Mihai Datcu**

BUCHAREST 2021

Table of Contents

Chapter 1 Introduction

The first satellite-based Earth observation missions date back to the late 1950s, with the launch of the first artificial satellite, Sputnik-1, which studied the ionosphere [1]. Over time, the interest of many countries and organizations in Earth Observation has increased. As a result, the amount of data collected over time has increased considerably, often exceeding processing capacity.

1.1 Presentation of the field of the doctoral thesis

As is the case with other big data domains[2], those aiming to process large volumes of data from satellites will face problems with storing and organizing data, manipulating it, extracting useful information, and presenting it.

The data resulted from Earth Observation goes through a series of stages to be capitalized by the end-user. These steps are the following: creation, storage, processing, and presentation. In the case of data produced by remote sensing missions, the users that want to process them face the problem of 5V (Volume, Velocity, Variety, Veracity, and Value) [3].

1.2 The goals of the doctoral thesis

Acquisitions covering the same geographical area form time series. Within them, different trends can be observed or changes that have occurred in an area in a given time interval can be quantified.

The thesis deals with the issues raised by the processing of satellite images time series (SITS), in terms of information extraction and data manipulation.

The first approach of the thesis consists of the development of algorithms, methods, and software systems capable of processing large volumes of data to extract useful information, all in a relatively short time. Considering the types of latent information in the remote sensing data, two approaches were taken:

- The first approach consists in the individual analysis of the acquisitions that form a time series, and the presentation of the results in the form of graphs that contain quantitative or statistical time evolutions.

- The second approach consists in the analysis of SITS as a unified data structure in which the aim is to identify the areas where changes occur (known or not) of the Earth's surface.

To enhance and speed up the analysis of large volumes of data, the thesis also deals with: indexing methods for remote sensing acquisitions which can help the user to quickly generate time series for a geographical area of interest, as well as to allow content-based search functionalities.

1.3 The content of the doctoral thesis

The thesis presents a series of software implementations for information extraction from satellite images time series, and which optimize the search process in remote sensing data archives, and results presentation, being structured in six chapters:

Chapter 2 presents the characteristics of remote sensing data used in developed software systems. The information presented is grouped, according to the type of acquisition used by the onboard satellite sensors, into multi-spectral, respectively radar.

In chapter 3 the methods and algorithms developed and used to extract information from remote sensing data, which have been integrated into developed software systems are described.

Capitolul 4 presents four software systems that facilitate the fast and accurate search and retrieval of latent information from individual remote sensing acquisitions.

Chapter 5 presents the developed software methods and systems that allow the manipulation of large volumes of data, the indexing of remote sensing data, the rapid generation of time series, the extraction of information, and the presentation of results in an interactive manner using a system based on web services.

The last chapter, 6, presents the conclusions for the software systems developed during the thesis, the obtained original contributions, the list of published articles, as well as perspectives for further development.

Chapter 2 Information regarding the remote sensing data used

This chapter presents the characteristics of each sensor whose acquisitions were used during the thesis to demonstrate the functionality and performance of the software systems developed for extracting information.

2.1 Principles of remote sensing

Depending on the acquisition method, the satellite images are acquired, the sensors used are divided into two categories: multi-spectral (passive), and radar (active). The multi-spectral sensors record the electromagnetic radiation reflected by the ground after the illumination made by the Sun. Active methods, such as Synthetic Aperture Radar (SAR), illuminate the ground using the instruments on board of the satellite and record the signals reflected by the Earth's surface.

The characteristics of the passive and active sensors, which gathered the data used in the developed software systems are presented further on.

2.2 Remote sensing instruments

The pixel is the smallest data structure in a satellite image. The information stored in each pixel depends on the spatial, spectral, and radiometric resolution of the sensor.

In the case of multi-spectral instruments, spatial resolution is a characteristic of the sensor and is defined by the smallest structures which can be distinguished in Earth Observation data. They usually fall into the following categories [4]: coarse resolution (> 1000 m), medium resolution (100-1000 m), fine resolution (5-100 m), very fine resolution (< 5 m).

The spectral resolution is defined by the number and spectral width of the spectral bands used for data acquisition. Depending on the number of bands, each sensor falls into one of the following categories: panchromatic (one band), multi-spectral (3 bands or more), or hyper-spectral (tens or hundreds of bands).

Radiometric resolution refers to the dynamic range of the sensor or the number of levels of electromagnetic radiation detected by the sensor. In digital format, these levels are given by the number of bits used to represent the data.

The temporal resolution is defined by the time elapsed between two successive visits of the same surface of the Earth.

2.3 Satellite images time series

Satellite Images Time Series represents a set of images gathered over the same geographical area at different times. To obtain denser series, for a relatively short period, SITS can also consist of acquisitions from different satellite missions. Sensors with high spatial and temporal resolutions allow more accurate observation of existing structures and patterns in dynamic scenes.

2.4 Conclusions

In terms of temporal sampling, SAR-acquisitions are not affected by weather conditions or scene lighting levels, making this type of data much more feasible for monitoring rapid changes that may occur in the event of natural disasters such as hurricanes, floods, vegetation fires, etc. Unfortunately, this type of data requires a higher level of processing than multi-spectral acquisitions, requiring advanced methods to reduce image noise.

Multi-spectral acquisitions allow better identification of ground cover classes, in the case of acquisitions made in optimal lighting conditions and the absence of meteorological phenomena and atmospheric disturbances. Also, the acquisition of data with improved spectral resolution allows faster data processing for different applications.

Chapter 3 Methods and algorithms used to extract information from remote sensing data

This chapter presents the feature extraction and classification algorithms implemented in the developed software tools, as solutions to the issues raised by the manipulation of data from Earth observation and the extraction of the contained information.

3.1 Existing concepts for extracting information from remote sensing data

A software system capable of automating the annotation of remote sensing data and the generation of accurate information from any type of data, for any possible scenario will always have a high level of complexity. In the development of such a system, it is very important to optimize the software component, flexibility and scalability being key points when tackling remote sensing data, due to the problems associated with 5V. For this reason, it is preferable to use a modular architecture [5], which makes it easier to integrate new functionalities and to use new data types. It is also recommended to use layered software architectures, as suggested in [6] and [7], with well-defined functionalities and logic.

3.2 Used methods for extracting information from remote sensing data

A first step in the process of extracting information from satellite images is to describe the existing content using mathematical representations based on spectral features, texture, or shape. These representations are used in the classification process to create models that can associate semantic labels to the mathematical representations. Classification algorithms are divided into supervised, in which the user must provide examples of elements to be associated with a semantic class, and unsupervised, in which the association of elements is done automatically depending on their mathematical representation and the used optimization function.

3.3 Methods for classifying data obtained from remote sensing

Classification algorithms allow the identification of similar structures in remote sensing data and represent an important means of extracting information. They are largely divided into unsupervised and supervised algorithms. The first category tries to automatically identify, representative elements (centroids) that can be associated with a semantic class. The number of centroids is suggested by the user and plays an important role in the generated results for most unsupervised algorithms. Supervised algorithms need examples to determine how data is distributed in semantic classes.

3.4 Metrics used to determine the performance of information retrieval methods

During the thesis, I used the following metrics to evaluate the performance of the implemented classification algorithms and software tools for the search of information in EO data: accuracy, precision, recall, F-score, and specificity. These metrics are often used to determine the performance level of algorithms and methods for information retrieval, classification, learning, pattern recognition, etc.

3.5 Conclusions

In this chapter, I presented a series of existing concepts dedicated to the extraction of latent information in remote sensing data. They solve various problems related to the automation of information extraction in the field of Earth observation.

The extraction of information from remote sensing data is performed using feature extraction and classification algorithms. Some of these algorithms have been integrated into the developed software systems, presenting good performance in the various scenarios presented in Chapters 4 and 5. The performance obtained was evaluated using metrics, such as accuracy, precision, recall, specificity, and F-score.

Chapter 4 Methods and software systems developed for the extraction of quantitative information from satellite images time series

This chapter presents a series of software tools and methods dedicated to processing and extracting information from Earth Observation data.

The first method uses representations based on spectral indices to automatically and quickly extract quantitative information on the coverage of different semantic classes in a given geographical area of interest.

In the second part of the chapter, a plug-in software tool is presented that facilitates the extraction of information from Sentinel-2 data, being available to the community dedicated to Earth Observation. It integrates feature extraction, classification algorithms, and visualization methods.

To accelerate and improve the process of extracting information from remote sensing data, I have developed a software system that allows the automatic saving and reuse of parameters used in classification scenarios.

At the end of the chapter, a complex software system is presented which is optimized for the processing and extraction of information from TerraSAR-X data. It contains several functionalities for managing very large databases, visualizing data, searching in datasets using content-based image retrieval methods, and semantic annotations.

4.1 Proposed method for the automatic analysis of satellite images time series

To automate the process of extracting quantitative information on the areas occupied by the existing coverage classes in the time series of satellite images, I have developed a fast and accurate method based on spectral indices and geographical references to reduce differences between data from multiple sensors. To demonstrate the method I presented the evolution of semantic classes in the geographical area of Bucharest, Romania, over 33 years [8].

The method used for the automatic extraction of the semantic coverage classes from SITS is represented in *Figure 4.1*. The first step is to query the satellite imagery database. The returned data goes through a process of atmospheric correction, after which only the bands required to compute the spectral indices are kept. The processed data is then sectioned according to the area of interest.

In the case of no ground-level measurements, atmospheric corrections are made with the Dark Object Subtraction 1 (DOS1) method [9]. This method assumes that, in the absence of atmospheric disturbances, the areas that are usually darker in the remote sensing images (areas covered by water) should be as close to black as possible. To apply the correction, the lowest value from each band is identified and the value is subtracted from each pixel of the band.

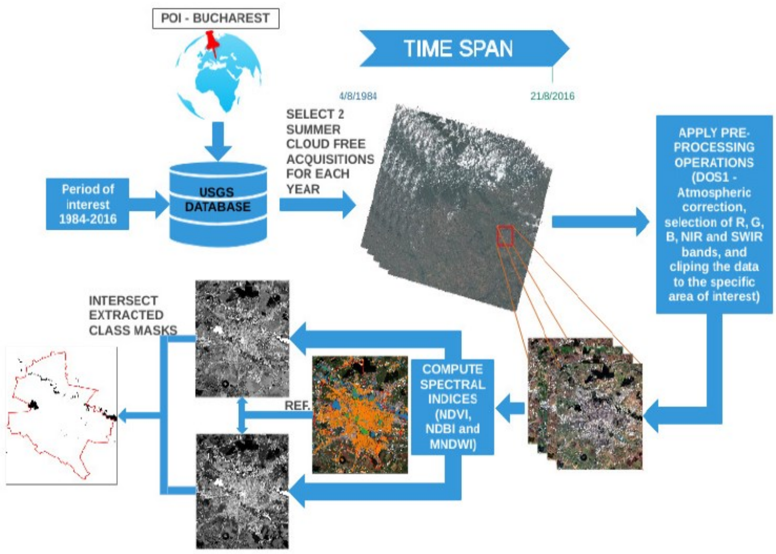

Figure 4.1 *Logic scheme for the automatic extraction of semantic classes from remote sensing acquisitions.*

Because the data comes from several sensors, to eliminate the differences between the obtained spectral indices values, several reference areas are used for each semantic class, in the method. The last step of the method is to identify all the pixels of the masks obtained for each semantic land cover class. These values are multiplied by the area covered by each pixel (900 m^2 for Landsat data, respectively 100 m^2 for Sentinel-2 data).

The method used for the analysis of time series proved to be viable for sets consisting of data from multiple satellite missions and sensors, as long as there is a reference unit for the monitored semantic classes.

4.2 Software system for extracting information from Sentinel-2 data

In this chapter, I present a software tool for extracting information from Sentinel-2 data, in the form of a plug-in for the Sentinel Application Platform (SNAP) [10]. SNAP is an open-source software platform developed by ESA for the exploitation of ERS-ENVISAT, Sentinel-1, Sentinel-2, Sentinel-3, and several other satellite missions.

The system integrates two distinct processing chains: a visual exploratory analysis of data, and semantic representation of satellite data content. They can be used both separately and chained, the semantic annotation process being able to succeed in the exploratory analysis one, through which the integrated classification algorithms can be optimally trained. In *Figure 4.2* green arrows are used to indicate computational flow and dependencies, while red arrows are used to highlight the user's purpose and mode of operation of the two processing chains.

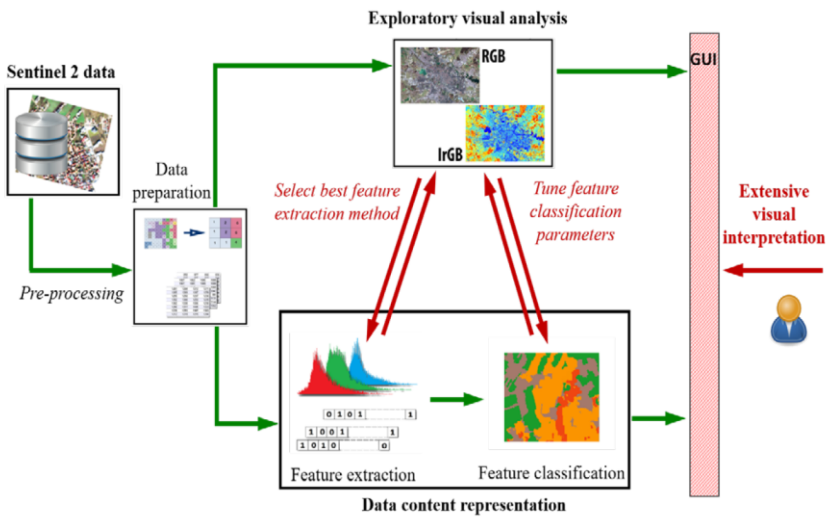

Figure 4.2 *Methodology for the exploitation and analysis of Sentinel-2 data.*

The system was developed using a modular logic and cand be integrated into the SNAP platform in the form of a plug-in. It uses the multitude of capabilities provided by the platform, such as the functionalities for visualization and exploitation of satellite images, layer management, arithmetic operations with spectral bands, etc.. All this makes SNAP a flexible platform that can serve several types of data and users.

The architecture of the plug-in follows the methodology presented in *Figure 4.2* and consists of two distinct modules: *Exploratory Visual Analysis* and *Semantic*

Annotation. The latter is divided into two sub-modules: *Feature extraction* and *Feature classification*.

The methodology presented can accommodate a wide range of algorithms for data visualization and content representation, but from previous experiences [11], I have identified several algorithms that have been integrated into the developed plugin.

To present the capabilities of the system, we chose two scenarios that use Sentinel-2 acquisitions. In the first scenario we performed a performance analysis of all algorithms integrated into the plug-in, using a scene for which there is a reference map of the semantic classes of coverage, while in the second scenario, the goal was to identify the coverage of a semantic class for which there is no reference map.

The proposed system [10], facilitates the analysis and exploitation of Sentinel-2 data through a platform known to the Earth Observation community, SNAP. The system integrates methods for searching for information using visual augmentation and semantic annotation methods based on methods and algorithms for feature extraction and classification.

4.3 Software system for improving the classification results of remote sensing data

To obtain optimal results in the supervised classification process, it has been shown that the user must provide both positive examples for the searched class and negative examples. [12]. I have developed a software system to allow the user to perform high-accuracy classifications for datasets gathered by remote sensing missions [13], and which is also capable of reusing the classification models on other data sets.

The functionalities of the system can be divided into five components represented in *Figure 4.3* (the information flow is from left to right): data annotation, image feature extraction, classification, loop for relevance feedback, and manipulation of results.

The system contains a set of functional components, which serve well-defined roles, such as graphical manipulation of images, extraction of image features, classification of data vectors, optimization of the training set, and manipulation of classification results. In the development of the system, I used the Model View Controller (MVC) programming paradigm [14] which separates the software content into three major components: model, view, and controller. These integrate five software packages used for remote sensing data representation, image data processing, classification, interactive training, and graphical interface.

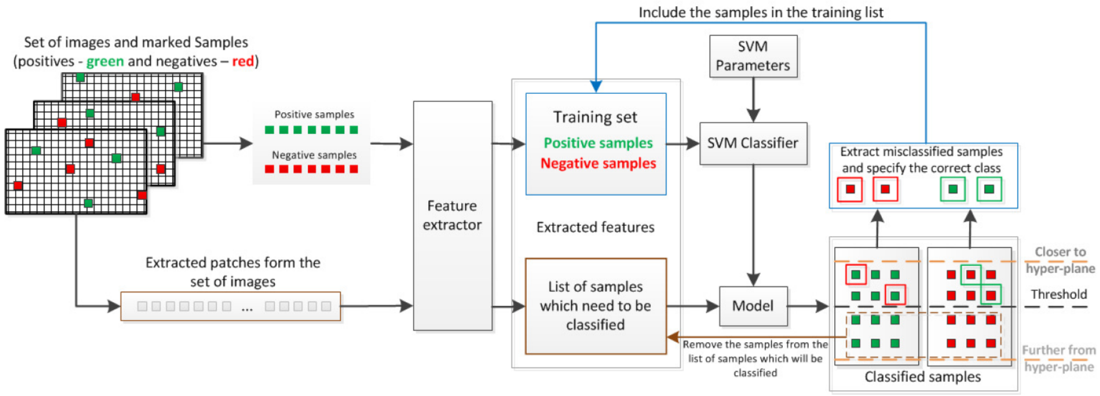

Figure 4.3 *Functional scheme of the system used to improve the results of the classification.*

To demonstrate the functionalities of the system and its compatibility with several types of remote sensing data, I built three scenarios in which I analyzed the performances obtainable with the developed system. In the first scenario, two data sets consisting of medium-resolution remote sensing acquisitions are compared, Landsat 8 OLI / TIRS (30 m spatial resolution) and Sentinel-2 MSI (10 m spatial resolution). In Scenarios 2 and 3, I analyzed data sets from GeoEye missions and aerial image acquisitions.

The interactive learning mode based on a small number of examples is effective from the user's point of view (he can indicate what he is looking for in a scene) and the computational effort. The system is able, due to the multiple types of integrated SVM algorithms, to perform both binary classifications (one-versus-all) and multi-class classifications.

The system can automatically save, after each iteration, the classification results, the SVM model, and the scenario parameters (examples used for training, grid size, features, etc.). The last two components can be reloaded into the system to continue the training process for the same data set or a different data set, which has similar characteristics. To help the Earth observation community, the application's source code is available on the github.com platform [15].

4.4 Software system for the extraction of information from very large volumes of remote sensing data

In this chapter, I present a software system, developed on a modular architecture, which offers innovative capabilities for extracting information from TerraSAR-X and remote sensing data [16]. The system consists of several modules that offer functionalities such as: adding data, extracting features from SAR data, extracting metadata, defining semantic tags for the content of image archives, advanced methods of querying databases based on content, meta-data, and semantic categories, as well as various data visualization functionalities.

To test the platform, from the TerraSAR-X data archive, I built a data set, covering urban and industrial areas, from different parts of the Earth. It consists of 1100 scenes selected according to availability, content, diversity of semantic classes, and acquisition parameters. To store the data, each purchase was segmented into sections of 200 x 200 and 160 x 160 pixels. Feature vectors were extracted from each of the sections and stored in a database. The database contains approximately 9.24 million vectors for each type of descriptor.

The developed system offers multiple functionalities for generating a data model, searching for information using 3D data visualization methods, various querying methods, and the creation of semantic class dictionaries. The system is capable of interpreting TerraSAR-X data and the accompanying meta-data to improve the entire processing chain that includes data calibration, feature extraction, semantic annotation, and the creation of digital product catalogs.

Figure 4.4 presents the main modules of the platform, together with the existing interfaces within the Payload Ground Segment, the user application, as well as the data, flows between the modules. The main components of the platform are Data Model Generation, Data-Mining Database, Query Engine, Content-based Image Retrieval, Visual Data Mining, Knowledge Discovery in Databases, Epitome Generation, and System Evaluation.

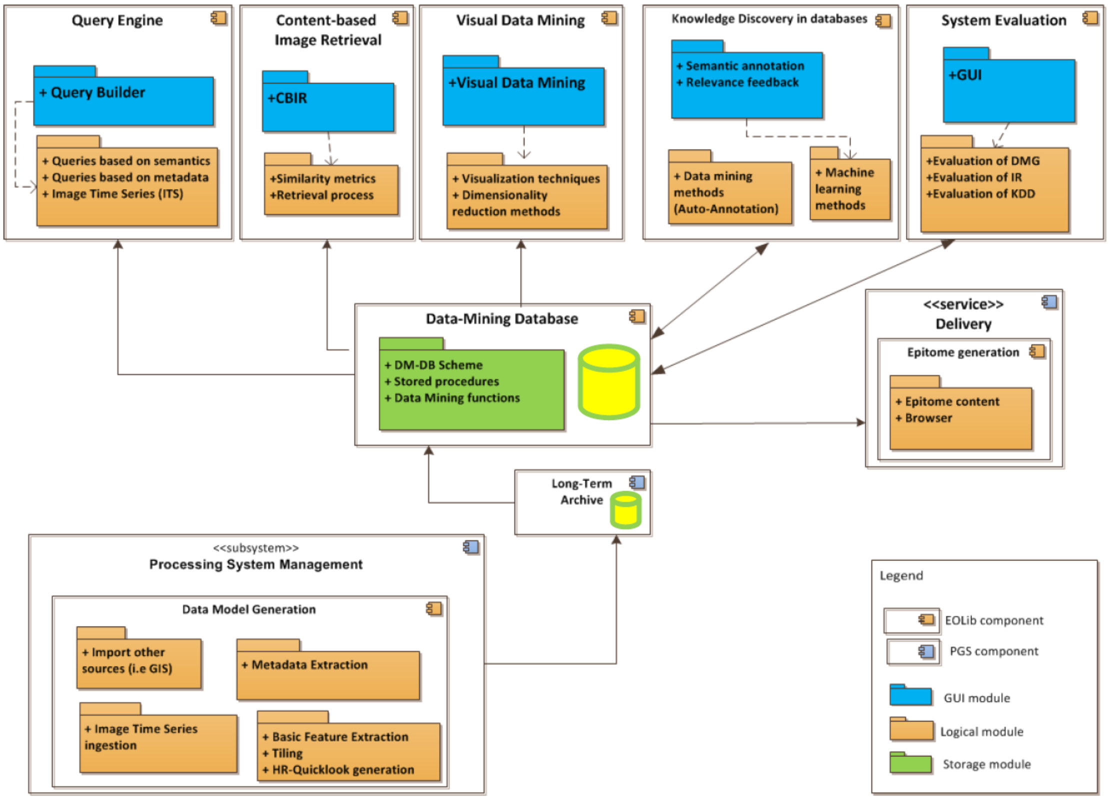

Figure 4.4 *The main components of the system, general functionalities, and dependencies.*

After adding data to the platform, users can perform the following operations: search for information, create new categories / semantic classes by exploring the database, tag and enter new taxonomies into the database, search for images as queries on the database of images given as examples, meta-data, classes or semantic categories.

4.5 Chapter conclusions

In this chapter, a series of methods and software applications for extracting information from remote sensing acquisitions were presented. All proposed methods integrate artificial intelligence algorithms and feature extraction that are adapted for remote sensing image processing. All the proposed systems were validated using usage scenarios, proving notable results for the field of Earth observation.

Chapter 5 Software methods and systems developed for the identification and quantification of changes in satellite images time series

In this chapter, a series of methods and software systems are proposed as solutions for the following issues: efficient indexing of databases with remote sensing acquisitions, generation of time series based on user-defined parameters, identification of areas where semantic changes take place, and interactive and selective presentation of information generated from time series analysis.

5.1 Indexing and information extraction method based on hashing algorithms

To optimize how the remote sensing data is indexed, I have implemented a method based on hash codes [17]. This facilitates the search for information based on content in very large Earth Observation databases. The main purpose of the implementation is to reduce the dimensionality of the feature vectors extracted from the images and to reduce the time needed for information extraction from very large databases. The logic behind the system contains two steps presented in *Figure 5.1*.

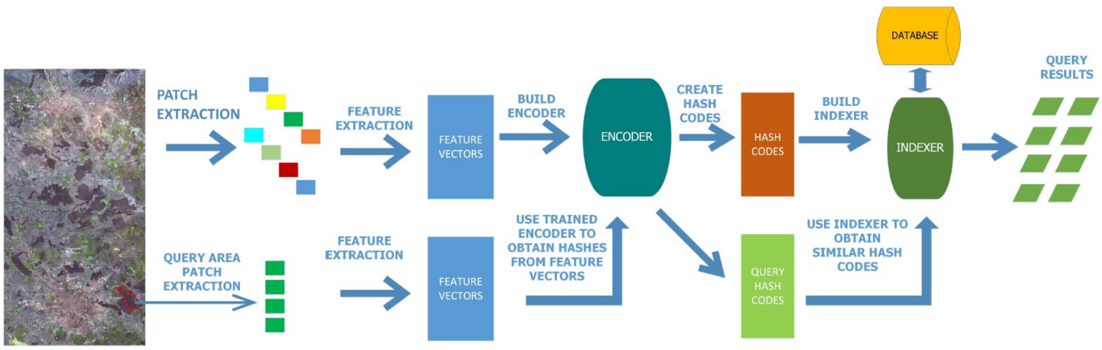

Figure 5.1 *Logical diagram representing the information flow of the encoder used to index the database (at the top), respectively, the information flow of the search in the database based on images given as an example.*

The method is based on a combination of feature extraction, hashing algorithms, and indexing methods. With the help of this system, the complexity and size of the database are much reduced. Also, the response speed for 20 million stored values (binary codes) is only 0.4 s.

To test the performance of the proposed method I built a database consisting of five Sentinel-2 images from all around Europe. The data set consists of 4 images, each covering an area of 109.8 km x 109.8 km and an image with an area of 40 km x 78 km. To perform the tests, each image was divided into sections of 5 x 5 pixels (or 50 m x 50 m). There are 20,537,664 sections in total, and the hash codes have been ingested into the database.

The system has very good scalability and allows the generation of results in a very short time, to the detriment of the longer computation times needed by the computation of hash codes.

5.2 Software system for the construction and temporal analysis of satellite images time series

To facilitate the manipulation of very large volumes of remote sensing data, I have developed a software platform that allows the user to create their time series on which they can apply different visualization methods to facilitate the observation of various developments in SITS, or download data given the potential for further processing [18].

The software system was developed using the Java programming language and has a layered and modular structure. The functionalities implemented by the platform are the following: visualization of geo-referenced data, indexing in the database, and non-perishable storage search and extraction of data, extraction of features, and classification. All modules are divided into abstraction levels: storage, modeling/representation, control/processing, and visualization. The created architecture is modular, flexible, allowing further development of the system.

The module associated with the database is developed based on the MySQL management system and uses the binary tree indexing mechanism applied to the tables. The indexing was performed for the satellite, sensor, band, and data fields for the “images” table to facilitating the extraction of information from the database with 830,000 records.

The graphical interfaces developed are components of the visualization level and represent the main way of interaction of users with the functionalities of the system. They consist of a series of windows, such as the one, presented in *Figure 5.2*.

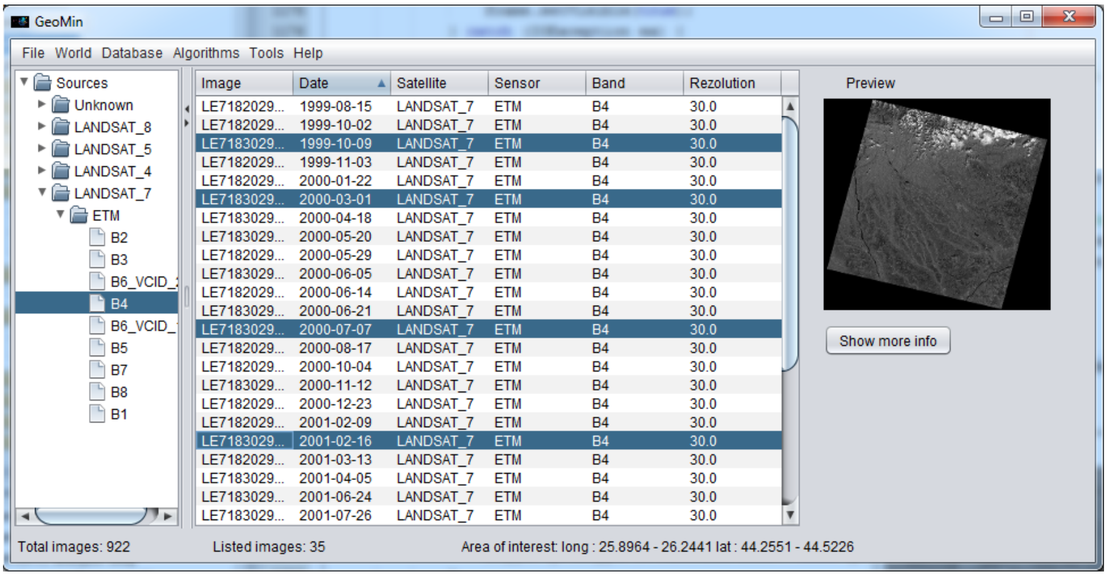

Figure 5.2 *The main application window.*

The interaction with the existing data can be done directly from the main window shown in *Figure 5.2* using the menu on the left, in which the data is filtered through a tree-like structure. It contains the following filters: satellite mission, sensor, and available bands.

The main source for the data used in the application is represented by the USGS platform. The system contains a total of 121 acquisitions, made by the satellite missions: Landsat 4, 5, 6, and 8. The data covers the southeastern part of Romania and was made in the period 1989-2015 [19].

The developed system has proven its usefulness in manipulating Spatio-temporal data and facilitates the search for information in data sets. This allows the creation of SITS for further processing such as feature extraction and classification.

5.3 Software system for extracting information from satellite images time series

In the case of systems based on supervised classification algorithms, functionalities are required for: loading and visualizing data, annotation, using feature extraction algorithms, and visualizing results.

All the functionalities mentioned above have been integrated into the software system developed for the processing and extraction of information from SITS, presented in this chapter.

To validate the developed system I used a time series consisting of 9 acquisitions distributed evenly over 9 years, on which I performed a series of tests to identify changes in the covered areas.

To be able to observe the SITS used as training sets, I set the format of the list (from the bottom center), from horizontal to vertical (see *Figure 5.3*). This allows better use of the available space and facilitates the observation of changes in each time series.

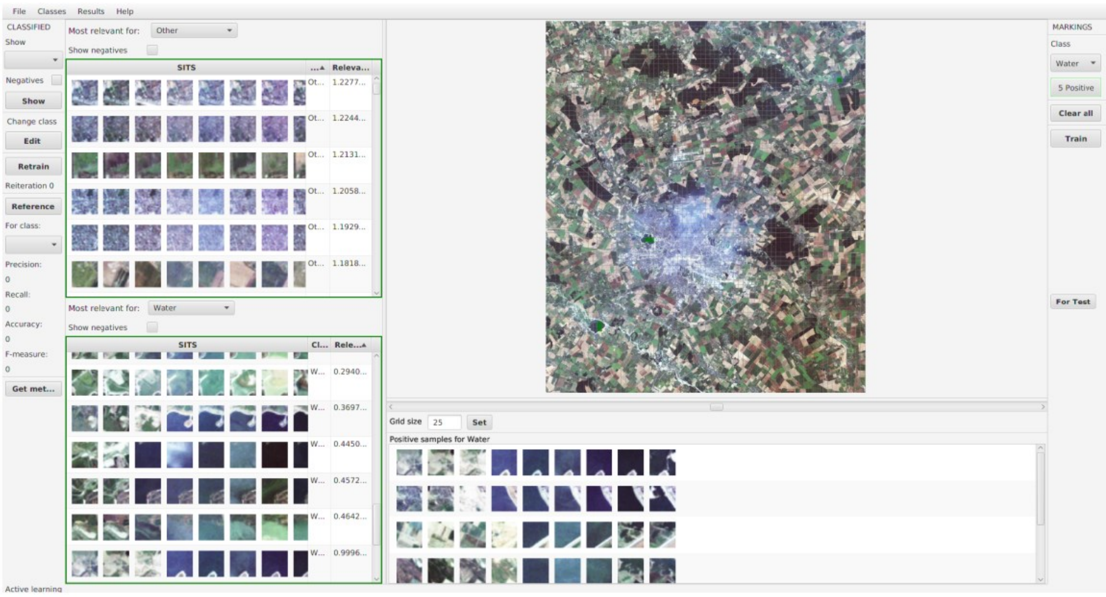

***Figure 5.3** The graphical interface of the system capable of processing SITS.*

The display of the results in the two tables on the left is set in a vertical format, and each entry contains all the sub-images that make up the time series. In this way, the user can adjust the training sets to be able to help the classifier in understanding the sought semantic categories for evolution. By increasing the complexity of the analyzed data, the probability of procedural errors also increases, and thus the user has better control over the entire process.

The data set used in this scenario consists of three Landsat 4 TM acquisitions and six Landsat 5 TM acquisitions, over Bucharest, Romania, during 1984-1993. To analyze the changes in the analyzed SITS, I created a pixel-level reference map for three evolution classes. These include landscaping, remodeling of urban areas, and fragmentation of agricultural land.

In the tests performed, I analyzed the results obtained by segmenting the data set into sections of 10 x 10 pixels (resulting in a total of 33490 sections), and 25 x 25 pixels (resulting in a total of 5372 sections).

I have shown that the system can obtain very good results, using a very small number of training examples (5-25), compared to the total number of elements analyzed (5372 and 33490 section). By using the assisted learning mechanism, the results improve, on average, by 5-35%, depending on the descriptor used to represent the data and the size of the sections that make up SITS.

5.4 Web platform for presenting information extracted from satellite images time series

By monitoring urban areas, one can identify and describe the processes by which land changes occur. With this in mind, I have developed a web platform for monitoring areas vulnerable to soil movements. It uses a series of satellite remote sensing data, together with measurements made on the ground to provide information about the level differences that may occur in urban areas, as a result of ground movement, or as a result of seismic movements.

The developed software system implements a web architecture, built using client-server technologies. The architecture is modular, the graphical interface of the application is developed using Java Server Faces (JSF) [20], while the back-end part of the application uses a MySQL database controlled using (Java Persistence API) JPA. Geo-referenced data (raster and vector) is stored and managed using GeoServer [21], which provides the data in a format optimized for web applications. These are displayed in the layered client interface using the OpenLayers3 [22], Javascript library.

The sources of the information displayed as products on the platform are PS interferometry (in the form of persistent scatterers maps, and point deformation profiles), change detection (in the form of binary change maps, multi-class change maps), and geological analyzes (in the form of geological maps, and textual descriptions of observations made at ground level).

The advantage of using a web architecture is that the user does not have to install anything, allowing quick and easy access for any user, regardless of the degree of expertise, geographical location, or resources available on the slide held.

SAR data is used for PS Layer type products that display terrain surface deformation information. These data come from the ERS1, ERS2, ASAR, and Sentinel-1 missions. The products available on the platform fall into the following categories: deformation maps, deformation evolutions (see Figure 5.4), geological maps, areas with geological risk, changes in coverage classes, statistics, and other products.

In addition to the products available in the graphical interface of the map type, on the web platform, there are also other types of information available that can be found on the secondary pages. These include the *Info* page which describes various terminologies related to layers with geological risk, or lithology. Also, if the user is authenticated, by accessing the *Download* page, he has at his disposal a series of documents with various additional information related to the areas of interest.

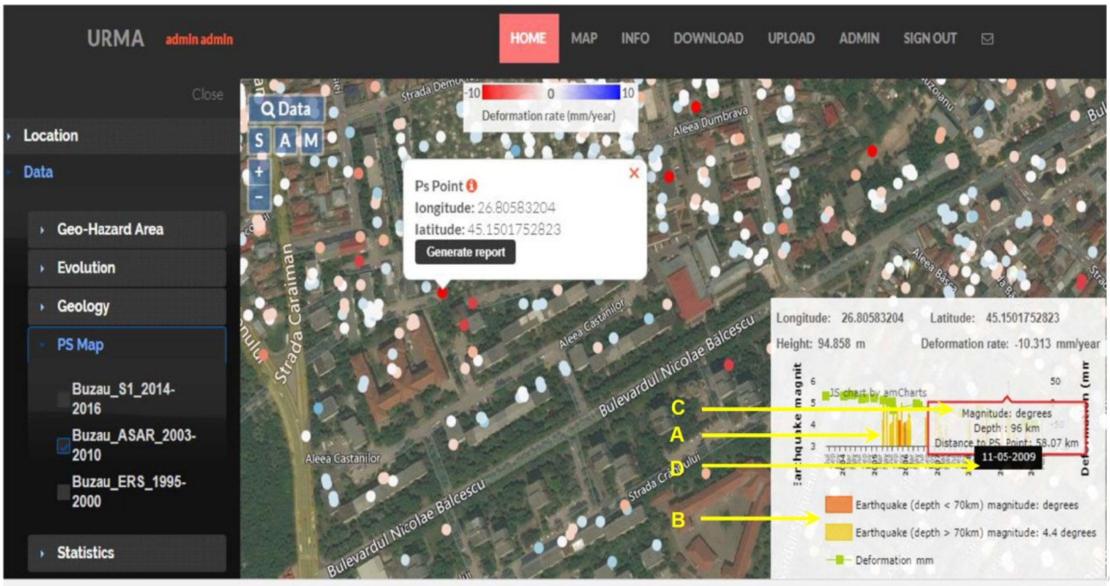

Figure 5.4 *Graphic representation of earthquake information.*

The software system was developed to facilitate the integration of several levels of information, providing the following services: visualization of deformation rates for various points in urban areas (building components), visualization of urban dynamics in the form of coverage class developments, identification, measurement and visualization areas with geological risk, displaying geological information (soil type, tectonic plates, earthquakes), and viewing various statistics for various fields (agriculture, demography, residence, etc.). The current version of the platform can be accessed at http://urma.ceospacetech.pub.ro.

5.5 Chapter conclusions

Extracting useful information from SITS data also raises issues related to their storage and indexing, searching, processing, and presenting the results obtained. In this chapter, a series of solutions have been presented that address one or more of the problems listed above in the case of processing satellite images time series.

Chapter 6 Conclusions

The thesis presents potential solutions to the problems raised when processing remote sensing data. The two directions followed by the thesis focus on the types of information that can be extracted from remote sensing data, quantitative (total areas at the level of individual purchases) or evolutionary (change maps). Through the solutions presented and demonstrated, several problems related to the processing of remote sensing data were treated.

To facilitate the extraction of quantitative information, resulted from the processing of individual acquisitions in a time series, I identified different sets of algorithms, for feature extraction and classification, which can be applied to multi-spectral data, respectively SAR data. These have been integrated into software systems dedicated to satellite image processing, and the proposed concepts were made available to the Earth Observation community.

The second direction generated solutions for the manipulation of data from a big data perspective. An optimized method for efficient storage of remote sensing data and fast search in databases with a very high volume of entries was presented, together with a software system for fast and customized generation of time series. The generation of evolution maps was treated within a software concept that facilitates the development of optimized classifiers that identify changes in remote sensing data.

The presentation of information extracted from satellite images time series was made within the web-based client-server developed system. It integrates several functionalities for the aggregation of several methods of presenting information which helps in a better understanding of the observed phenomena in satellite data.

6.1 Original contributions

The original contributions developed during the thesis, and the correlations with the articles in which they were presented (listed in Chapter 6.2), are as follows:

1. A method based on spectral indices that can be used to analyze time-series consisting of multi-spectral acquisitions from multiple sensors. The proposed solution was applied to a time series of multi-spectral images, covering 33 years [C7].

2. A plug-in architecture that facilitates the processing of remote sensing data by the community involved in Earth observation. It integrates multiple feature extraction and classification algorithms optimized for Sentinel-2 data processing using the SNAP platform. Several scenarios were tested and presented in [J1], [C1], [C3], and [C5].
3. A concept for optimizing the classification of large volumes of data using a small number of examples, based on a loop for relevance feedback. It also integrates mechanisms for applying the trained classifier on a separate data set from the one on which the training was performed. The proposed solution was validated by several scenarios presented in articles [C2], and [C12].
4. Optimization of the concept from point 3 for the analysis of time series as an individual data structure. The developed concept was validated by articles [C9] and [C10].
5. A software concept for extracting information from TerraSAR-X data. It integrates database and indexing solutions optimized for the type of data used, feature extraction and classification algorithms, and content-based search methods [J2], [C6].
6. An indexing solution for very large volumes of data based on hashing algorithms, which allows content-based searches [C4].
7. A system for manipulating very large volumes of remote sensing data from multiple sensors, which can be used to generate time series, and display evolution patterns present within them [C14].
8. A software concept for presenting the information obtained from remote sensing data, improving the perspective on the observed physical phenomena. The system has functionalities for the automatic generation of reports for geographical points of interest. This system integrates web solutions to display developments [C8].
9. Several state-of-the-art concepts were analyzed to determine potential future research directions [C11], [C13], and [C15].

6.2 List of original works

During the doctoral thesis, the proposed concepts were validated by the scientific community through 2 journal articles and 15 conference articles. Also, most of the software concepts were implemented in 4 research projects, in the field of Earth observation, which was funded by ESA (see Annex A.1).

6.2.1 Journal articles

J1. **A.-C. Grivei**, I. Neagoe, F.A. Georgescu, A. Griparis, C. Vaduva, Z. Bartalis, and M. Datcu, *Multispectral data analysis for semantic assessment – A SNAP framework for Sentinel-2 use case scenarios,* IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 4429-4442, 2020, DOI: 10.1109/JSTARS.2020.3013091.

J2. M. Datcu, **A.-C. Grivei**, D. Espinoza-Molina, C. O. Dumitru, C. Reck, V. Manilici, and G. Schwarz. *The Digital Earth Observation Librarian: A Data Mining Approach for Large Satellite Images Archives,* Journal: Big Earth Data (TBED), 2020, DOI: 10.1080/20964471.2020.1738196.

6.2.2 Conference articles

C1. C. Vaduva, F.A. Georgescu, **A.-C. Grivei**, I. Coca (Neagoe), A. Griparis, Z. Bartalis, and M. Datcu, Sentinel-2 data analysis based on explainable features and exploratory visual analysis, ESA EO Φ-WEEK 2020, 28 September – 02 October 2020, Virtual Event.

C2. **A.-C. Grivei**, *SVMIRE - An Open Source Svm Image Retrieval With Relevance Feedback System For Earth Observation Data Classification,* 2020 13th International Conference on Communications (COMM), Bucharest, 18-20 June 2020, pp. 171-176, DOI: 10.1109/COMM48946.2020.9141975.

C3. **A.-C. Grivei**, C. Văduva, and M. Datcu, *Assessment Of Burned Area Mapping Methods For Smoke Covered Sentinel-2 Data,* 2020 13th International Conference on Communications (COMM), Bucharest, 18-20 June 2020, pp. 189-192, DOI: 10.1109/COMM48946.2020.9141999.

C4. **A.-C. Grivei**, C. Văduva, and M. Datcu, *Improved Earth Observation Data Retrieval Through Hashing Algorithms*, IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, 28 July - 2 August 2019, pp. 5909-5912, doi: 10.1109/IGARSS.2019.8898911.

C5. C. Văduva, F. A. Georgescu, A. Griparis, I. Neagoe, **A.-C. Grivei**, and Mihai Datcu, *Exploratory Search Methodology for Sentinel 2 Data: A Prospect of*

Both Visual and Latent Characteristics, IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, 28 July - 2 August 2019, pp. 10067-10070, DOI: 10.1109/IGARSS.2019.8900349.

C6. C. O. Dumitru, G. Schwarz, **A. Grivei**, R. Hugues, J. Lorenzo, M. Datcu, *Artificial Intelligence for EO Sensor Fusion*, Living Planet Symposium, Milan, 13-17 May 2019.

C7. **A.-C. Grivei**, and M. Datcu, *Analysis of Bucharest's Land Cover Evolution Over a Period of 33 Years Using Multi-Sensor Data*, IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, 23-27 July 2018, pp. 1680-1683, DOI: 10.1109/IGARSS.2018.8518615.

C8. **A.-C. Grivei**, D. D. Taralunga , A. Popescu, V. Poncos, S.-A. Toma, D. Teleaga, A.-M. Vîjdea, A. Baltres and I. Cociuba, *Multi-Layer Geohazards Information System Concept for Urban Areas − URMA Urban Mapping*, EURAS 2018; 12th European Conference on Synthetic Aperture Radar, Aachen, 4-7 June 2018, ISBN: 978-3-8007-4636-1.

C9. **A.-C. Grivei**, A. Rădoi and M. Datcu, *Land Cover Change Detection in Satellite Image Time Series Using an Active Learning Method*, 2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), Bruges, 27-29 June 2017, pp. 1-4, DOI: 10.1109/Multi-Temp.2017.8035213.

C10. **A.-C. Grivei**, A. Rădoi, C. Văduva, and M. Datcu, *Land Cover Change Detection Using a Relevance Feedback System*, WorldCover 2017 Conference, Frascati, 14-16 March 2017.

C11. **A.-C. Grivei** and M. Datcu, *Data Analytics for Spatio-Temporal Patterns in Satellite Image Time Series: Methods and Architectures*, 2nd BigSkyEarth Workshop – Big Data processing and management concepts for new platforms, Sopron, 23-24 February 2017.

C12. **A.-C. Grivei**, A. Rădoi, C. Văduva, and M. Datcu, *An Active-Learning Approach to Query by Example Retrieval in Remote Sensing Images*, 2016 International Conference on Communications (COMM), Bucharest, 9-11 June 2016, pp. 377-380, DOI: 10.1109/ICComm.2016.7528339.

C13. **A.-C. Grivei**, and A.-M. Ghimeș , *Modern Considerations for Data Mining and Big Data*, 2016 International Conference on Communications

(COMM), Bucharest, 9-11 June 2016, pp. 385-390, doi: 10.1109/ICComm.2016.7528265.

C14. **A.-C. Grivei**, and M. Datcu, *HiperMINE – An Earh Observation Spatio-Temporal Data Mining System*, 2015 23rd Telecomunications Forum TELFOR (TELFOR), Belgrade, 24-26 November 2015, pp. 906-909, DOI: 10.1109/TELFOR.2015.7377612.

C15. C. Văduva, A. Rădoi, **A.-C. Grivei**, G. Schwarz, and M. Datcu, *Urban Mapping using Satellite Time Series*, Mapping Urban Areas from Space – MUAS 2015, Frascati, 4-5 November 2015.

6.3 Prospects for further development

The functionalities of the presented software tools can be extended with the help of algorithms based on neural networks, which produce results with a high level of accuracy.

Integration of all the concepts developed in the thesis, for indexing and displaying data, extracting information from time series of satellite images and presentation in a complete system, based on cloud services, to which users have access based on a web interface, is another direction for further developments.

Another direction for further development, which can increase the processing speed of remote sensing data is the implementation of the proposed algorithms on graphics cards with a large number of processing cores.

Annexes

A.1 Contributions to ESA-funded projects

The software solutions developed during the thesis were implemented in a series of software systems, desktop and web platforms that were the result of 4 projects funded by ESA.

P1. **Open Source Image Retrieval – Integration of Developed Tools (OSIRIDE)**

OSIRIDE was won in the competition "AO/1-7966/14/I/CS", and has been funded by the European Space Agency, General Support Technology Program (GSTP 6 El. 1), and conducted in the period 2015-2017. The objective of the project was to develop and implement tools for information search in remote sensing images based on the CIBC concept. The developed algorithms have been integrated into an open-source platform, which has expanded the processing capabilities of very large volumes of data obtained from Earth observation. The modular architecture includes a series of algorithms and methods for searching and extracting latent information, data visualization, supervised learning of the correspondences between the physical characteristics of the data and the associated semantic labels, through user interaction with the web platform.

The personal contribution to this project consisted of the development and demonstration of the concept of supervised learning based on relevance feedback, form the user, on the data returned by the system, and the optimization of classification algorithms. I also performed a series of optimizations to reduce the computational effort in the process of extracting features and classification.

P2. **Data Mining For Analysis and Exploitation of Next Generation of Time Series (DAMATS)**

The DAMATS project was won through the competition "AO/1-7967/14/I/CS", and has been funded by the European Space Agency, General Support Technology Program (GSTP 6 El.1), and developed within the timespan: 2015-2017. The project aimed to develop a software system that integrates different analytical methods for the exploitation of the information contained in SITS. The extracted information is presented in the form of categories of the evolution of the land cover classes observable in data resulted from Earth Observation. The interaction with the user is performed with the help of a web platform that gives him access to databases with remote sensing data, and to computational resources, to use the algorithms.

In this project I implemented and demonstrated the performance of state-of-the-art algorithms that were integrated into the system, I developed a demonstrator for an SVM-based method with relevance feedback that can be applied to extract evolution semantic classes from time series.

P3. **The Multi-layer Geohazards Information System Concept for Urban Areas (URMA - urban risk mapping)**

The project was won through an ESA competition, "Call for Outline Proposals under the Romanian Industry Incentive Scheme", and has been carried out in the timespan of 2016-2018.

Within this project, I have developed a web-based software system for the interactive exploitation of information on the dynamics of urban areas and which allows the user to create correlations between available geological, seismic, and PSI data. The platform provides users with features for the automatic generation of reports containing information on the evolution of a particular point of interest.

P4. **Multispectral Data Analysis Toolbox for SNAP – ESA's SentiNel Application Platform (DAS-Tool)**

The project was won following the competition "AO/1-8545/16/NL/Cbi Romanian Industry Incentive Scheme", being funded by the European Space Agency, and carried out in the period: 2017-2020. The objectives of the project were to develop dedicated algorithms for the description and semantic analysis of Sentinel-2 data. To achieve this, plug-in software architecture has been developed that interfaces with ESA's SNAP platform.

My contributions in this project consisted of implementing and testing the feature extraction and classification algorithms used for the semantic annotation by regions of Sentinel-2 data. Another contribution is the development of software architecture and graphical interfaces for extracting information from remote sensing data at the patch level.

Bibliography

[1] V. Kuznetsova, V. Sinelnikova și S. Alpertb, „Yakov Alpert: Sputnik-1 and the first satellite ionospheric experiment,” *Advances in Space Research,* vol. 55, nr. 12, pp. 2833-2839, 2015.

[2] A.-C. Grivei și A.-M. Ghimeș, „Modem Considerations for Data Mining and Big Data,” în *11th International Conference on Communications - COMM2016*, București, 2016.

[3] Y. Demchenko, P. Grosso, C. de Laat și P. Membrey, „Addressing big data issues in Scientific Data Infrastructure,” în *2013 International Conference on Collaboration Technologies and Systems*, 2013.

[4] S. Liang, X. Li și W. Jindi, Advanced Remote Sensing, Elsevier, 2012.

[5] D. Espinoza-Molina și M. Datcu, „Architecture Concept for Earth,” în *IGARSS 2013*, 2013.

[6] M. M. U. Rathore, A. Paul, A. Ahmad, B.-W. Chen, B. Huang și W. Ji, „Real-Time Big Data Analytical Architecture for Remote Sensing Application,” *IEEE Journal of Selected Topics in Applied,* 2015.

[7] P. Mazzetti și S. Nativi, „Multidisciplinary Interoperability for Earth Observation: Some Architectural Issues,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,,* vol. 5, nr. 3, pp. 1054-1059, 2012.

[8] A.-C. Grivei și M. Datcu, „ANALYSIS OF BUCHAREST’S LAND COVER EVOLUTION OVER A PERIOD OF 33 YEARS USING MULTI-SENSOR DATA,” în *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia, 2018.

[9] M. Moran, R. Jackson, P. Slater și P. Teillet, „Evaluation of simplified procedures for retrieval of land surface reflectance factors from satellite sensor output,” *Remote Sensing of Environment,* vol. 41, nr. 5, p. 69–184, 1992.

[10] A. Grivei, I. Neagoe, F. Georgescu, A. Griparis, C. Vaduva, Z. Bartalis și M. Datcu, „Multispectral data analysis for semantic assessment – A SNAP framework for Sentinel 2 use case scenarios,” *J-STARS,* 2020.

[11] C. Vaduva, F. Georgescu, A. Griparis, I. Neagoe, A. Grivei și M. Datcu, „EXPLORATORY SEARCH METHODOLOGY FOR SENTINEL 2 DATA: A PROSPECT OF BOTH VISUAL AND LATENT CHARACTERISTICS,” în *IGARSS 2019*, Yokohama, 2019.

[12] M. Babaee, S. Tsoukalas, G. Rigoll și M. Datcu, „Visualization-based active learning for the annotation of sar images,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 8, nr. 10, p. 4687–4698, 2015.

[13] A.-C. Grivei, „SVMIRE - AN OPEN SOURCE SVM IMAGE RETRIEVAL WITH RELEVANCE FEEDBACK SYSTEM FOR EARTH OBSERVATION,” în *The 13th International Conference on Communications*, Bucharest, 2020.

[14] F. Buschmann, R. Meunier, H. Rohnert, P. Sommerlad și M. Stal, Pattern-Oriented Software Architecture - Volume 1: A System of Patterns, Wiley Publishing, 1996.

[15] A. Grivei, „GitHub,” [Interactiv]. Available: https://github.com/acgrivei/svmire. [Accesat 22 07 2020].

[16] M. Datcu, A.-C. Grivei, D. Espinoza-Molina, C. Dumitru, C. Reck, V. Manilici și G. Schwarz, „The Digital Earth Observation Librarian: A Data Mining Approach for Large Satellite Images Archives,” *Big Earth Data,* pp. 1-30, 2020.

[17] A.-C. Grivei, C. Văduva și M. Datcu, „Improved Earth Observation Data Retrieval through Hashing Algorithms,” în *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, Yokohama, 2019.

[18] A.-C. Grivei și M. Datcu, „HyperMINE – An Earth Observation Spatio-Temporal Data Mining System,” în *23rd Telecommunications forum TELFOR 2015*, Serbia, 2015.

[19] C. Văduva, A. Rădoi, A.-C. Grivei, G. Schwarz și M. Datcu, „Urban Mapping using Satellite Time Series,” în *Mapping Urban Areas from Space Conference*, Frascati, 2015.

[20] „Java Server Faces,” [Interactiv]. Available: https://javaee.github.io/javaserverfaces-spec/.

[21] „GeoServer,” 14 06 2020. [Interactiv]. Available: http://geoserver.org/.

[22] „Open Layers,” 14 06 2020. [Interactiv]. Available: https://openlayers.org/.

[23] T. Costăchioiu, R. Constantinescu și M. Datcu, „Multitemporal Satellite Image Time Series analysis of urban development in Bucharest and Ilfov areas,” în *10th International Conference on Communications (COMM)*, Bucharest, 2014.

[24] C. Dumitru, G. Schwarz, A. Grivei, R. Hugues, J. Lorenzo și M. Datcu, „Artificial Intelligence for EO Sensor Fusion,” în *Living Planet Symposium*, Milan, 2019.

[25] USGS, „USGS Earh Explorer,” USGS, [Interactiv]. Available: http://earthexplorer.