

Received December 18, 2017, accepted January 19, 2018, date of publication February 1, 2018, date of current version March 16, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2801032

Understanding Heterogeneous EO Datasets: A Framework for Semantic Representations

CORINA VADUVA¹, (Member, IEEE), FLORIN A. GEORGESCU¹, (Member, IEEE),
AND MIHAI DATCU^{1,2}, (Fellow, IEEE)

¹Research Center for Spatial Information, University Politehnica of Bucharest, 061071 Bucharest, Romania

²Remote Sensing Technology Institute, German Aerospace Center, 82234 Oberpfaffenhofen, Germany

Corresponding author: Corina Vaduva (corina.vaduva@gmail.com)

ABSTRACT Earth observation (EO) has become a valuable source of comprehensive, reliable, and persistent information for a wide number of applications. However, dealing with the complexity of land cover is sometimes difficult, as the variety of EO sensors reflects in the multitude of details recorded in several types of image data. Their properties dictate the category and nature of the perceptible land structures. The data heterogeneity hampers proper understanding, preventing the definition of universal procedures for content exploitation. The main shortcomings are due to the different human and sensor perception on objects, as well as to the lack of coincidence between visual elements and similarities obtained by computation. In order to bridge these sensory and semantic gaps, the paper presents a compound framework for EO image information extraction. The proposed approach acts like a common ground between the user's understanding, who is visually shortsighted to the visible domain, and the machines numerical interpretation of a much wider information. A hierarchical data representation is considered. At first, basic elements are automatically computed. Then, users can enforce their judgement on the data processing results until semantic structures are revealed. This procedure completes a user-machine knowledge transfer. The interaction is formalized as a dialogue, where communication is determined by a set of parameters guiding the computational process at each level of representation. The purpose is to maintain the data-driven observable connected to the level of semantics and to human awareness. The proposed concept offers flexibility and interoperability to users, allowing them to generate those results that best fit their application scenario. The experiments performed on different satellite images demonstrate the ability to increase the performances in case of semantic annotation by adjusting a set of parameters to the particularities of the analyzed data.

INDEX TERMS Earth observation data understanding, human machine communication, image information mining, semantic gap, sensory gap, semantic representation.

I. INTRODUCTION

There is a close dependency between the dynamics of Earth surface and the technological evolution. Since the beginning of remote sensing, a permanent cycle between applications and sensors development was generated and maintained by the need of more accurate data acquisition. The need to understand and exploit the environment had generated a loop where social response must be in line with scientific and industrial breakthroughs. The Earth Observation (EO) field comes as a consequence of human activities to monitor the surrounding processes. A wide range of new applications emerged from the development of remote sensing technology as various imaging sensors have been designed and manufactured to measure precise aspects of the Earth surface. Spatial,

spectral and radiometric characteristics of the recorded data reveal important information to support a correct environment understanding in view of application scenario assessment and sustainable development. However, the capabilities to explore it are currently reduced, perhaps even missing in some situations. Human analysis for large and heterogeneous amounts of remotely sensed data is very hard to perform and the assistance of computers in the process is preferred. EO data mining systems must provide access to very large collection of data, including imagery, metadata or additional vector maps, with the main purpose of discovering hidden information, adding semantic labels and highlight those specific structures needed to support decision makers, large scale planning or thorough monitoring.

EO data structuring and analysis started with the development of content based image retrieval (CBIR) systems. The main idea in CBIR consists of searching semantic resemblance. To bridge the gap between the knowledge about the land cover and the data recorded can be a challenge due to different representations given by EO sensors. Moreover, the analysis of an image collection by a machine is only able to provide similarity by data processing. There are new discrepancies between machine and human interpretation, called semantic gap. The content is firstly described by means of its main characteristics (color, texture, shape) in order to extract objects and modeled as a feature vector using various algorithms. There is no standard descriptor, each aspect of the image content may be highlighted by a different algorithm. The same problem occurs when trying to group objects based on common patterns in order to express high level semantics. Based on a series of parameters, classification methods are always leading to different results. A user may also influence the results through the training process, where applicable. The behavior of each method and algorithm has been extensively studied [7], [8], [9] and explored in information mining and retrieval approaches.

One of such attempts in the EO field, KIM is a prototype of a knowledge-driven image information mining system developed for the exploration of large image archives [2], [3]. The SemQuery approach in [4] is trying to enhance the effectiveness of the process by means of heterogeneous texture, color and shape features embedded in the images. The system supports visual queries by semantics-based clustering and indexing approach. These two systems are valuable solutions for the retrieval of images containing certain regions, mostly based on their primitive features. The VisiMine system was designed to support interactive classification and image retrieval by extending content modelling from pixel level to regions and scene modelling [5]. Perhaps the most evolved and complex CBIR system for EO imagery that has been designed is introduced in [6]. The GeoIRIS System contains automatic feature extraction, visual content mining from large-scale image databases and high-dimensional database indexing for fast retrieval in EO image archives. A generous overview of the CBIR systems developed for the EO field is presented in [1]. More recent approaches of query engines expand data modeling with linked open data and ontology-based analytics. Common very high resolution image products are transformed into actionable intelligence information, including image descriptors, metadata, image tiles, and semantic labels. [10].

The literature refers to all these solutions as adaptable tools able to deal with rapid mapping, as well as with archive exploration [1]. The accent goes on dedicated techniques for a comprehensive analysis of the image content in the attempt to provide a universal tool that could be further optimized for fast searches in large databases. They all combine standard procedures, yet different algorithms, to enable functions such as data characterization, supervised content based annotation and indexing in the attempt of dealing with

EO data particularities and variability. While there are no changes in the CBIR principles and data flow, this research domain is continuously expanding, mostly due to the new type of data to be analyzed and information to be retrieved. Periodic reviews of state of the art ideas, influences and trends [11], [12], [13], [1] prove the increased interest of the research community on the image retrieval domain. There is a broad category of applications defined by constraints like the user intent, the data scope, the query category, the algorithmic process and the results visualization. The key elements that impact the search process in EO data collections are defined in [14]: template-based, attribute-based, metadata-based, semanteme-based and integrated retrieval. As a consequence, the process of browsing through collections of unstructured data is resumming at finding a series of patterns and defining interactions between them in order to explain concepts and create knowledge. Searching by association is the most interactive approach and is the easiest manner to tackle a database in pursuit of interesting objects [12]. In addition, there are applications focusing on finding categories of objects or just targeting specific information.

CBIR systems tailored to EO data are obviously an actual evolving topic trying to overcome issues concerning the particularities of the acquisition sensor, the emerging use applications and the final data interpretation. Each of the proposed solutions has its strengths, limitations and specific performances, without emphasizing a general solution, although there is a common framework. CBIR systems normally contain a set of functions to extract properties of image objects and combine them further in semantic levels. The human factor is acting like a bridge over the machine development and the domain issues. Active learning and relevance feedback functions are applied to discover that information similar to the knowledge shared by the user through his query. Only particular feature combination is targeted as a short list of available functions is limiting the system ability to distinguish between fine aspects of land cover structures. This shortcoming is even more underlined by the fact that no algorithm has been proven to maintain the same precision or recall when the type of data is changed. For instance, 2 images acquired over the same area with different sensor will illustrate different structures. The meaning of the scene is strongly depending by the resolutions (spatial, spectral or radiometric) of the acquisition sensor.

In order to overcome these limitations, we propose a compound framework addressing relevant information extraction that includes several algorithms for each of image processing levels: feature extraction and scene classification. Its configurable architecture represents the principal advantage, offering the user the opportunity to create specific representation based on the addressed application, or even to compare several results obtained with different configurations. In order to avoid misinterpretation and inappropriate selection of various parameters, a rate distortion theoretical approach will guide the user through the process. At the end, a validation module enables the comparison of the classified scene with

a reference data set, when available. A prototype system was developed in order to demonstrate the proposed framework.

Apart from providing a solution adapted to the data and application at hand, the system was designed to reduce the differences between machine interpretation and human understanding of the data and also to deal with the influence of the EO imagery on the human reasoning on the real environment. There is a growing public awareness of the great impact that semantic and sensory gap have in the EO domain, caused mainly by the remote sensing technological development. The problem was carefully considered, as the results of unsupervised CBIR techniques are not always relevant and satisfactory in terms of user perspective [20]. Ontologies have been built and methods to add signs to small objects and symbols to larger areas have been conceived in order to provide a common understanding and bridge these gaps. In this context, the authors present a human machine communication concept to ease knowledge transfer from the user to the computer.

The rest of the paper is structured such that each section to address a point in the work logic dedicated to the understanding of heterogeneous EO datasets towards semantic representation. Section II is introducing the sensory gap and explains the different environmental perception. Section III presents the proposed framework for image information mining that is trying to unify the results of human and computer analysis. Section IV explains the semantic gap between image features and the user understanding, while Section V presents a list of parameters whose adjustment will guide the generation of relevant information. A human machine communication procedure to enable automatic associations between human reasoning and the numeric representation is introduced in Section VI. The final workflow is detailed in Section VII and the experimental results are illustrated in Section VIII. The last section is dedicated to conclusions.

II. EARTH OBSERVATION DATA UNDERSTANDING

In some aspects, the Earth Observation domain generated a revolution in the field of image understanding, impacting also the image processing procedures. First and foremost, the perspective on the objects around us was changed. The human beholds the scene at the same level, the eye level, while the satellite is looking from above (FIGURE 1). But for the pattern recognition techniques this is not an issue, as the identification of numerical similarities follow the same procedures. The user though had to learn new rules to bound group of elements into a meaningful object that matches the structures in everyday life.

There is an important number of instruments to remotely capture broad information about the environment in a fast way. The technological progress encouraged the development of many sensors for Earth Observation in order to capture various particularities of the land surface that are not available at a glance, but through prior knowledge (i.e. object composition) or further analysis (i.e. a large structure out passing the eye view). The amount of new information



FIGURE 1. A commercial area in Bucharest observed from a horizontal angle (left) and from a vertical angle (right).

yields uncertainty in the knowledge regarding the state of the environment. The literature names it sensory gap and it refers to the discrepancies between the standard object and the information generated through computation by remote sensing instruments [12].

The list of causes behind the sensory gap is not being limited to the scene content, but also to the variety and particularities of EO sensors. Sometimes it is difficult even for humans to understand an object by itself. A contextual analysis is required, as different spatial relations among a number of elements will generate different comprehension over the scene. For instance, an area containing trees, a lake and residential area could represent either a park inside a city, or an urban area besides the forest. Clutter and occlusion will also have a great impact on scene analysis. Unwanted echoes received during the image acquisition process (coming from clouds, atmospheric particles or ground targets) will deepen scene understanding and further semantic analysis. Moreover, objects may obstruct each other, due to different heights and acquisition angle.

From the sensor's point of view, the list of discrepancies becomes longer. Their properties deteriorates the scene understanding. The most common factors behind the sensory gap are the type of sensor, the spatial resolution and the spectral resolution.

With a slightly smaller impact on the sensory gap, the radiometric resolution deepens the differences in terms of discernible details, as the EO sensors are able to spot 65536 levels of radiation intensity, while the human eye cannot distinguish more than 256 levels. This particularity is not decisive for the applications targeting general classes of land cover. However, it can become crucial when small elements or details must be highlighted. The analysis of EO imagery can provide an advantage over visual inspection.

Much more insight over the Earth surface can be obtained based on the ability of sensors to capture energy outside the visible portion of the electromagnetic spectrum. Images obtained by measuring microwave, ultraviolet or infrared radiation will provide information that the human eye is not able to perceive, increasing thus the gap between the knowledge about scene and its actual composition. Structures have a unique way of reflecting the radiation in each portion of the electromagnetic spectrum. Besides revealing properties about the elements constituting the objects, EO sensors

provide single spectral signatures that support a deeper land cover analysis where a semantic class can be divided into further semantic subclasses (i.e. inside an area of woodland we can differentiate several areas of forests, based on the trees species, soil types can be differentiated). Exploiting this particularity of EO data in the image analysis process usually complements the visual interpretation.

The variety of EO sensors is increasing the sensory gap. The available instruments so far provide a rather large number of different perspectives on the scene geometry and reflected radiation due to different acquisition manner. The same object can be described using different energy signatures: microwaves in the Synthetic Aperture Radar (SAR) images and visible, ultraviolet and infra-red electromagnetic radiation in the multi-spectral data. An example of three different representations, underpinning complementary information, is showed in FIGURE 2.



FIGURE 2. The House of Parliament, Bucharest: 3 different energy signatures measurements obtained using 3 different sensors: Sentinel 1 (left), Sentinel 2 (center), aerial (right).

Illumination, acquisition mode, acquisition angle are just some more factors influencing the sensory gap. However, their impact is minor. Therefore, the framework proposed in this paper will focus mainly on providing a solution to cope with the differences introduced by the variety of EO sensors and their characteristics (spatial, spectral, radiometric resolutions).

The development of several sensors targeting specific information comes as a consequence of current needs and applications. For this matter, either optical or SAR data, together with their properties, impact the data understanding and interpretation process, for both human and computer based analysis. Same objects will have different representations for different acquisition modes, spatial and radiometric resolutions of the remotely sensed images. FIGURE 3 presents a comparison between different satellite images with spatial resolutions ranging from 30m to 0.5m, covering the same area over a region in Constanta, Romania.

The lower resolution images reveal the information regarding the land cover, while the high resolution ones highlight specific elements, such as buildings and industry structures. Besides, the details regarding the same structure are not as expressive in Sentinel-1 SAR images as they are in Sentinel-2 optical images. Additionally, SAR data tends to emphasize structures with high reflectivity/backscatter.

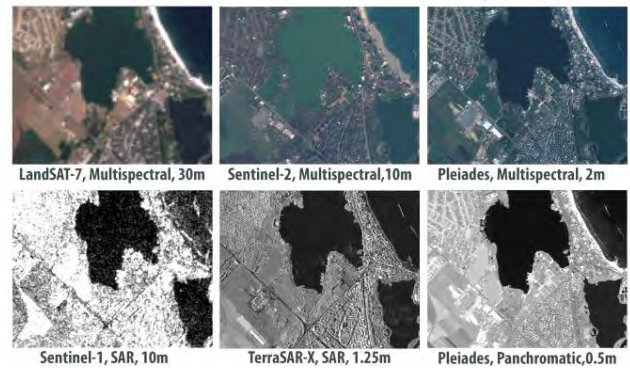


FIGURE 3. Comparison between various types of EO data, at different spatial resolutions and acquisition modes.

III. FRAMEWORK FOR IMAGE INFORMATION MINING AND KNOWLEDGE DISCOVERY

As presented in the previous section, the remotely sensed object can induce different opinions over its nature, compared to the actual structure as perceived on the ground. Moreover, each EO sensor provides distinct information about the objects within the sensed scene, as per its designated purpose (i.e. optical imaging, thermal imaging, radar imaging, etc), making it hard for the machine to obtain a description matching human understanding. In our attempt to overcome the problem of EO data variety in the data analysis process, we introduce a compound, yet simple, framework for image information mining (IIM) (FIGURE 4).

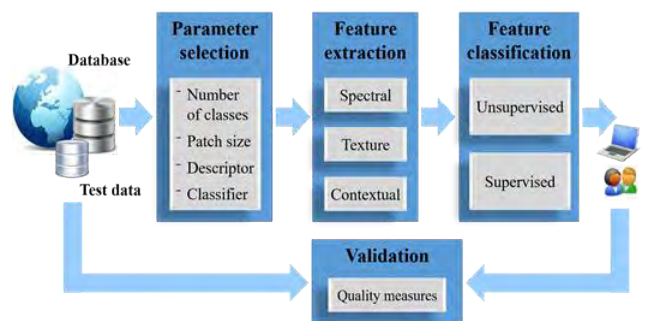


FIGURE 4. Image information mining framework.

The procedure is targeting EO data in general, optical and radar imagery, adapting to the particularities (mainly to spatial and spectral resolutions) of each acquisition sensor. The user's intervention is required in the beginning of the process. He must define the size of the grid used to cut the scene into patches. A single patch becomes the main semantic element composing the image. The data turns into a collection of basic elements expressing those local features serving the definition of the overall scene understanding. Further spatial dependencies will be defined through a data modeling process aiming at grouping similar neighboring areas and transposing the data into a semantic annotation map. As each type of EO image and application requires a dedicated process, the data model selection is sustained by a rate-distortion analysis. Its goal lies in outlining those hidden

relationships imposing the number and the geometric shape of the classes that best matches the user's perspective over the ground structures. As a result, the system presents to the user the appropriate number of distinct object categories best approximating the semantics of the selected scene given the statistical modeling. This information can be of great help for the user to understand the data and to support further selection of training set for scene classification. The chosen samples will serve two purposes: 1) to exploit the particularities of the data and transpose them into an individual procedure for feature extraction and classification; 2) to generate, based on the corresponding feature vectors, a data model that will be applied to the entire scene. The main innovation of this framework consists in the fact that it can create a distinctive procedure for each analyzed scene, tackling the issue of "sensory gap". We envisaged a list of several feature extraction and classification algorithms and methods for the content based analysis, aiming to highlight spectral, texture or contextual characteristics, in an unsupervised or supervised manner. For the image information mining procedure to be complete, the user is able to verify the obtained results against a reference data set.

Consequently, we propose a modular structure, where each of the six main component elements has a well-defined purpose:

- Repository – to store the data;
- Control unit –enabling the selection of parameters for further processing;
- Feature extraction module – includes several algorithms for image characteristics analysis;
- Classification module – integrates different machine learning techniques for data modeling;
- Validation module – where the results are compared against reference data;
- Graphical User Interface – enables human integration into the process for the active learning techniques, data validation and results visualization.

In the following sections, the authors unravel more details, as well as the methods and algorithms considered for the proposed framework.

A. REPOSITORY

The data repository refers to a structure designated to host large collections of heterogeneous EO images. It is desirable that the user can access this data in the minimum amount of time and retrieve basic metadata such as type of acquisition (multispectral or radar), quick preview and cloud coverage. This means that the general understanding of "storage" is not fulfilling the needs for the EO domain.

The proposed framework envisages a database to enable data partitioning and further ingestion of computed features, scene classification and reference datasets. A built-in project handler is responsible to store the main settings and intermediate results provided during the human machine interaction. This fact will ensure a fast transfer between the data and the processing unit. Also this will be of real importance in

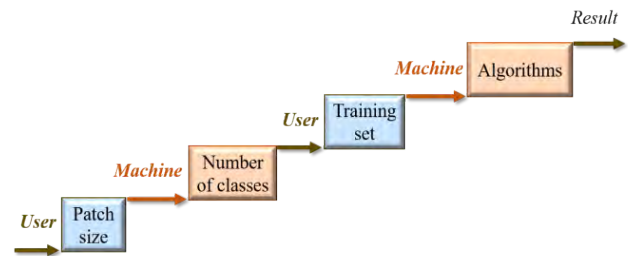


FIGURE 5. Flowchart presenting the parameter settings and the responsible for each action.

further analysis and assessments specific to each application scenario.

B. CONTROL UNIT

The variety and volume of EO imagery render significant information. Storage, management and access are mandatory, however, not enough to fully exploit that information. The list of separable patterns is expanding with every data analysis technique employed. The challenge consists of finding the method that is able to discover the most relevant content description with respect to the human understanding. It is known that the objects can be easier outlined from the rest of the scene if the image is analyzed at the pixel level. However, in order to reduce the influence of hazard and accidental grouping inside the scene, a region based analysis is more appropriate. The patches are the natural semantic units of a scene, in the sense that they contain a simple structure of objects, which can be labeled. Working with large areas of an image or with the whole scenes can lead to a generality lack.

This approach entails feature extraction and classification in the attempt to highlight those groupings carrying a semantic meaning for the user. At this point, the most important issue to be considered is the choice of the information extraction method to be employed. For this reason, the control unit is the core of the proposed framework. Its main role is to ensure that all the process is entirely lined up and correlated with the data content. Therefore, this unit will gather the user preferences and command the rest of the process. It demands the definition of five categories of parameters: the patch size, the number of semantic classes, a set of samples from each expected semantic class, the feature extraction algorithms and the classification methods. Setting appropriate values for these parameters is the main challenge of the information mining process and the scope of the control unit.

This paper introduces an innovative approach where the user must define only two of the five parameters, leaving the system to automatically suggest optimum values for the other three. The parameter selection can be considered as the baseline of the human-machine dialogue, where each participant is contributing towards the final decision regarding the image information extraction. The flowchart in FIGURE.5 illustrates that the user starts by choosing the patch size, with respect to his interest on the scene content.

We reach next a much debated topic, the number of distinguishable groups of elements able to express the semantics

inside an EO image. We believe that information theory provides a coherent solution for this problem. Each object on the Earth surface is responsible for the generation of a measurement (i.e. the pixel value sums up the radiation reflected by that object) as a form of uncertainty regarding the interpretation of reality. The aim is to transform these measurements into mutually exclusive areas and label them based on the similarities they share. The homogeneity of the areas is determined with respect to a set of discrete properties, such as spectral, texture or contextual features. All of these characteristics try to go beyond the perception of the human eye, as the recorded measurements do not always correspond to the visible electromagnetic radiation. The algorithmic process uses numeric resemblance to define groupings, propose classes of objects and express semantics. In order to align human and machine understanding and correlate them with the real land cover, the main challenge is to identify the amount of uniform areas for which one is able to generate classes of objects with semantic meaning and a unique symbolic representation. In the field of information theory, rate distortion theory is to provide the best source coding, eliminating redundancy [15]. The core of the processing lies in the minimization of the mutual information computed between the EO imagery and its semantic representation. When applied to our problem, the rate distortion function indicates the optimum number of groupings (classes of objects) that best estimates the Earth surface.

The control unit includes this analysis and hence, it automatically provides an estimation of the optimum number of semantic classes that machine can separate, given the image particularities. The user will take the result as a suggestion, a reference to consider when setting the parameter into the framework.

The next step in the parameter selection flowchart refers to the definition of a training data set that will be employed by the supervised classification methods we describe in the following sections. The user is responsible for this action and the main concern will be to extract the most relevant patches for each of the semantic classes. The control unit upholds the samples selection and storage into the repository for further use.

The last parameter to be defined enables the data processing itself. A variety of techniques for feature extraction and classification were introduced over the years, however the discussions did not converge to a conclusion regarding the best performances when applied to the EO imagery. As the available data is increasing and the CBIR approaches are evolving, the opportunity arises to learn from small amount of data and extend the knowledge to the full scene. In this regard, the training data set selected for scene classification, will also serve as reference data collection to help the identification of the most suitable data processing technique. The control unit will automatically demand the computation of all the “feature extraction algorithm – feature classification methods” combinations, together with a validation measure attached. The user will be informed on the performances

that the available techniques can deliver for the analyzed image and provide guidance for a tailored selection of the methods and algorithms to be used to obtain the final result. Each EO image will be the object of a specific processing chain underpinning the particularities of the data. However, the double role of the training dataset can be a drawback for the proposed framework because the user will be compelled to create a consistent collection in order to provide reliability to the validation measures computed.

C. FEATURE EXTRACTION

In most of the data representation approaches, specific classes in the image being analyzed are grouped according to some dominant characteristics like coarseness, contrast, color distribution or directionality. Considering the diversity of image discernible properties, feature extraction methods sensitive to spectral, texture, and shape information have been proposed. The data is thus divided based on its informational content into multi-dimensional feature vectors, a mathematical representation of the image properties.

Due to the fact that remote sensing data is measured within specific wavelength intervals, there is no general rule that can be applied to create a universal information retrieval procedure regardless of the data being analyzed. In most of the cases we have to use specific algorithms for specific types of data. Furthermore, in the context of image indexing, most of the methods are based on identification and classification of image texture, image intensity or by using statistical models.

1) SPECTRAL ANALYSIS

The properties derived from the spectral values of the image pixels are efficient and easy to compute, compared with other feature extraction methods, being used on a large scale in scene classification and CBIR applications. Their most important advantages lie in the simplicity of extracting color information about the structures in the scene and the ability to represent the visual elements.

Some of the common spectral features for remote sensing image analysis are color histograms [9] and color moments [24]. Other methods are based on color coherence vectors [25], color correlograms [26] and even on the dynamic color distribution entropy of neighborhoods [27]. Furthermore, in EO image analysis, specific feature extraction methods based on the spectral indexes [28] have emerged due to the high spectral resolutions provided by the multispectral remote sensing sensors.

In the frame of spectral analysis, the proposed approach includes algorithms to compute features based on Spectral Histogram and Spectral Indexes. Those features are efficient and easy to compute [18].

2) TEXTURE ANALYSIS

The purpose of texture analysis is to quantify intuitive qualities of an image described by terms such as rough, smooth, silky, or bumpy as a function of the spatial variation in pixel gray levels. Dedicated procedures can be helpful when

objects in an image are mainly characterized by their texture instead of intensity, and traditional thresholding techniques cannot be used effectively.

The utmost employed statistical approach for texture analysis is the grey level co-occurrence matrix (GLCM). Introduced in [22], it was the first approach to describe and classify the texture inside an image. Haralick's GLCM approach assumes that texture can be represented using pure statistical description. The work in [23] states that for a better description of texture it is advisable to combine geometrical structures with statistical ones, such as in the case of the Statistical Geometric Features method.

Recent experiments in the field of EO data processing have revealed good results for texture analysis based on Gabor filtering [18] and Weber Local Descriptors [19]. The two algorithms have been included in the feature extraction process.

3) CONTEXTUAL ANALYSIS

Currently evolving texture analysis and local feature extraction techniques have led the way to mixed feature methods, that are joining texture and spectral features in the same descriptor, and also to Bag of Words (BoW) based methods, which are relying on learning dictionaries of visual words. Even though BoW was initially used for video search, a lot of derivative methods that emerged from it could solve problems like image classification, image retrieval and object recognition.

In the remote sensing community this technique has been recently introduced for image annotation, object classification, target detection and land use classification, and it has already proven its discrimination power in image classification [29], by performing a vector quantization of the spectral descriptors in an image against a visual codebook. In the BoW framework, there are several ways to generate the visual codebook. Even though K-means is the most common clustering procedure used to produce a code list [30], there are some attempts in using random dictionaries [29]. Depending on the features involved in the codebook generation, different classification results may be obtained.

The proposed framework considers both approached for contextual analysis. Joint features such as Gabor-Histogram Descriptor and WLD-Histogram Descriptor may be computed to provide one of the top precision score in literature for EO image classification [18]. For the BoW based descriptors, the authors focused on Bag of Spectral Indexes, a particular descriptor that is providing enhanced features for EO multi-spectral image classification [42]. The last one cannot though be applied to SAR imagery.

D. FEATURE CLASSIFICATION

Classification is a process employed to assign a class label to a set of measurements [31]. If the desired output consists of one or more continuous variables, then this process is referred to as a regression [33]. Moreover, the applications in which the training data comprises examples of the input vectors with

their corresponding target vectors are known as supervised learning problems. For that particular situation when the training data consists of a set of input vectors that does not have any corresponding target labels, we are dealing with an unsupervised learning problem. This time, the purpose is to discover groups of similar examples within the data, whether they correspond or not to a visual, recognizable element.

Common classification procedures use, among others, minimum distance, maximum likelihood, Mahalanobis distance, parallelepiped algorithm and expectation-maximization. These algorithms are simple and can perform very fast. For remote sensing image classification, the use of classical approaches does not provide best results. New classification algorithms based on fuzzy logic have emerged. Experiments have showed that fuzzy algorithms can deal with mixed pixels and improve the accuracy of the classification [32]. Fuzzy equal relationship, fuzzy ISODATA, fuzzy synthesized judgement or fuzzy language are just some of the most popular techniques. Another special category of classification algorithms uses neural networks, such as multi-layer perceptron network (MLPN), radial basis function neural network or fuzzy self-organizing neural network. These classification algorithms are more complex, but provide enhanced accuracy of the classified data.

Given the fact that knowledge transfer from user to the machine is a central issue of the proposed framework, supervised learning actions are envisaged. Support Vector Machine (SVM) [35], [36] algorithms and k-Nearest Neighbors (kNN) [37], [38] classification algorithms were included. The aim is to model the EO images according to the training set provided by the user and a series of precomputed feature vectors describing the data content. To define the proper feature vectors is usually a challenge in the remote sensing application. Therefore, unsupervised classification methods are recommended to highlight data driven patterns that are naturally formed through statistical computation and guide further selection. Thereby, the authors included the k-means algorithm [34] to algorithmically outline those groupings to whom a semantic meaning can be attached.

E. VALIDATION MEASURES

The remote sensing process provides a set of measurements on the environmental condition as an additional information to the human knowledge. The data mining procedure aims at a joint exploitation that goes beyond the common knowledge and reveals the hidden information that is correlated to the human understanding. In other words, computer based processing is used to estimate facts about the land cover and surface transformations. Ambiguity is expected, therefore, validation metrics were developed to quantify the similarity degree between the estimated information and the perceivable situation. Generally, the uncertainty measured is influenced by the nature of the experiment. There is no universal metric that is widely recommended for computer simulations and human interpretation [39]. An evaluation of the most common validation metrics is presented in [40]. The type of data

and the user bias have a great influence on the predictive capabilities of data collections exploration process. Efficient EO image information mining requires a set of methods and algorithms that are fit for the image content they analyze. The metrics become a tool to assess the model quality and select from the list of available techniques. The interest is to obtain an accurate representation of the real world from the perspective of the data properties and its role in the EO applications. For the proposed framework, individual Precision - Recall curves are helpful in identifying among the potential groupings inside the data, the ones for which good matching to the users' interpretation is reached.

The proposed concept offers flexibility and interoperability to the user, generating those results that best fit the application scenario. The architecture of the proposed framework was designed such that new methods and algorithms can be added to the process, as long as a general data representation scheme is preserved. There must be an input / output correlation in order to avoid system and algorithmic errors. The feature extraction is computed independently of external factors. Nevertheless, data classification requires human intervention for feature selection and, sometimes, for training. The graphical user interface (GUI) will translate the user's perception and understanding of the scene into numeric indexes and statistical dependencies that will serve the information mining process. It assists the transfer between the human language and the machine representation, as defined in the next chapters. We can consider the interaction between user and the machine as a dialogue, where communication behavior guides the computational process through a set of elements. These elements are in fact the parameters in the control unit. By changing them, the human transfers his opinion to the machine in a way that is usable for it. The user integration in the process tries to reduce the "semantic gap", as described next.

IV. ADD SEMANTIC MEANING TO THE DATA

Most of the CBIR presented in the literature are facing serious issues regarding the semantic consistency of the information the machine considers relevant. The description that humans attach to an image is almost always related to visual elements that the eye can perceive and to the contextual relationships that the brain can comprehend. Although intensely debated, the lack of coincidence between the results of data processing and the user interpretation is yet a problem to be solved [12]. Periodic emergence of new types of data, due to the development of satellite missions, entails the definition of different pattern recognition approaches. The focus is to maintain the data-driven observable connected to the level of semantics and to human awareness.

The proposed concept tries to bridge the semantic gap by using a set of parameters to strategically guide the data modelling inside the information extraction procedure. A certain flexibility is assumed, turning the issue of unsupervised analysis into a user based parameter selection task. A common ground is thus defined between the mathematical handling

of the full EO data and human interpretation based on visual data recordings. This turns into a matching problem between numerical similarities, obtained through acknowledged pattern recognition techniques, and semantic groupings, resulted after the human's inspection on visual elements (FIGURE 6).

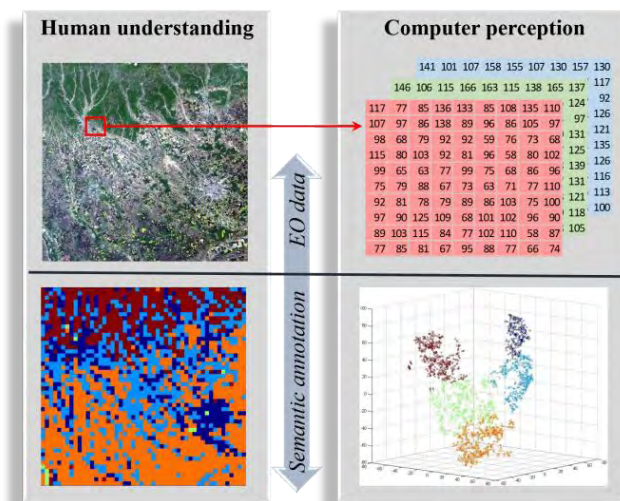


FIGURE 6. A comparison of EO data interpretation between human understanding and computer perception.

The general idea behind the presented framework envisages the knowledge transfer from the user to the computer based on hierarchical representation of the data content. There will be at first a data decomposition into a set of basic elements with no visual meaning. Considering these numerical patterns, the literature offers a full set of algorithms capable to highlight specific features which can be linked with perceivable characteristics, such as color, texture or shape. At this point, the mathematical associations start to be adaptable and the user can enforce his judgement on the data processing results. The development can continue, and complex structures can be revealed if the proper feature combinations are identified. Mathematical and cognitive similarities can be correlated, as observable structures, with semantic meaning, are targeted. FIGURE 7 illustrates the hierarchical content representation supporting the analogies between human and machine understanding of the EO data.

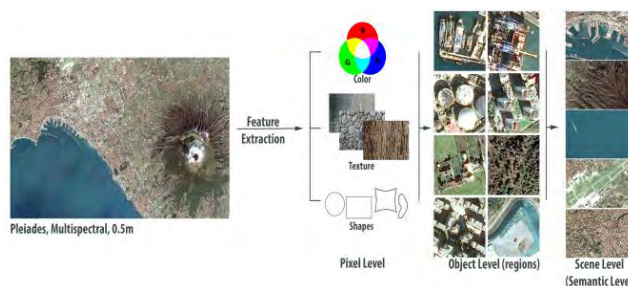


FIGURE 7. Hierarchical content representation for EO data understanding.

Along with the decomposition of the EO data, the set of points restraining the information extraction process is

underpinned. The corresponding parameters allow the user to interfere with the machine learning process and try to wrap the mathematical modeling around the perceivable elements. Usually, there are few situations when the visual aspect is entirely embodied into the mathematical model. The struggle is however to identify those algorithms and adjust the parameters that best fit the data content in terms of semantics.

Given these considerations, the environment created by the proposed framework supports the human machine interaction in the sense that the core data model is defined in an attempt to harmonize the visual similarities with the groupings obtained by data processing. For this reason, the first step is to perform an unsupervised classification, such that the user to gain an insight on the numeric characteristics of the data. Then, he will preserve only the most semantically relevant samples from each class and provide them as a training set for the machine learning algorithms to adjust the data model. New visual rules will complement the mathematical relationships in order to separate classes of elements in strict correlation with human understanding and data properties. There is thus the promise of results for scene classification that will combine similarities by computation with the real meaning of the scene.

V. PARAMETER SETTING

An overview of the article so far reveals a modular framework for EO image information extraction that aims to bridge the semantic and sensory gaps caused by the EO data variety and the human's lack of ability to perceive data as a whole, beyond the visible domain. Unlike other CBIR approaches, this concept integrates a complete set of feature extraction and classification algorithms. The control unit provides full access to the user, enabling the data process configuration at his choice. This is a four steps task, and the parameters to be defined are presented in Section III. B (FIGURE.5). On this line, a set of additional constraints will add the human understanding to the data-driven computation. Further, the article details the way the parameters impact on data modeling, with respect to semantic interpretation.

At first, we introduce the patch, a parameter that is directly responsible for laying the foundations of the semantic meaning concept. It narrows the size of the searchable patterns composing the data. It also withholds the human's attention on configurations bound by visual and spatial relationships. The patch size is correlated with the amount of details the user is interested in distinguishing inside the scene, but also with the spatial resolution provided by the acquisition sensor.

For instance, in the case of a 30 m resolution image, one can't expect to discriminate objects like trees, gardens, houses, or even neighborhoods. It is more likely to search for general category of elements, such as urban area, forest or agriculture. Given their usual coverage, small patches of 25-50 pixels can be regarded as an element purely describing such structures, particularly for a 30 m spatial resolution image. However, for a higher level of semantics, such as "harbor", "forest outside the city" or "rural area" the size

of the patch shall reach 100-120 pixels, in order to include a combination of two or three general categories of elements. As the patches get larger, the information is more uniformly distributed and separation between different classes becomes difficult [19]. We face a similar analysis for the case of very high resolution. Smaller patches tend to become a discriminant feature pattern for the discernible structures, while larger patches enclose a mix of elements whose scale and spatial interaction will determine the human understanding. From the semantic point of view, the size of the patches can have the same value as in the case of low resolution image data. Let us consider the case of a 2 m spatial resolution image. The user expects to distinguish precise objects, like buildings, construction sites, sport and leisure facilities, streets, water courses, types of crops and so on. A square area with a 50 to 100 m length side will capture the core features of these elements, while 200 to 240 m length side square area is large enough to outline groups of elements with semantic meaning.

In order to illustrate the impact that the patch selection has it on the understanding and analysis of EO images, FIGURE 8 presents various patches of different sizes, cut from images acquired over the same area with several sensors. If we consider the same patch size for all the images, we observe that the higher spatial resolution provides more details about the scene. On the other hand, larger patches tends to include groups of structures whose interpretation require contextual analysis.

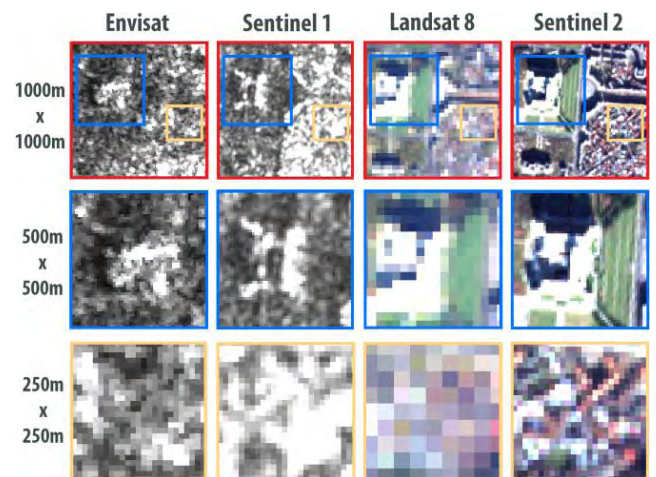


FIGURE 8. Understanding the EO image content with respect to the size of the patch: examples of patches as the main unit to represent its content. The first row illustrates wider areas, while the second (with blue contour) and the third (with yellow contour) one cover smaller areas, depicting well defined ground structures: the House of Parliament in Bucharest, Romania (patched with blue contour) and a residential area (patches with yellow contour).

The purpose of the patch size is to enhance the aspects of the data content that are related to the final CBIR application. The information extraction framework creates knowledge in response to the idea of an use case scenario. As the variety and complexity of data prevails in the EO domain, a content based analysis can always result in different interpretations

and concepts. There are two main levels of search: conceptual level and syntactic level [16].

The first approach requires a certain amount of knowledge over the scene. Usually, when searching for a particular category of elements, the user is familiar with the meaning of the target. The object is easy to spot through the visual analysis, but the computer is not able to provide the same performance. Human input and a smaller patch are appropriate when trying to draw boundaries on regular areas.

In the case of syntactic level, the user is aiming for the visual composition. The scene is not analyzed by singular objects but by the way they interact and form vicinities. These associations provide high level semantic to an image. Larger patches are opportune for the applications where contextual information is preferred over the homogeneity based characterization.

The second parameter controls the partitioning of the image content into classes of similar patterns. From the computers' point of view, these patterns could stand for feature vectors compressing statistical characteristics. However, the user may consider them visual elements that support the definition of semantic meaning. Extremely debated topic in the literature, the image partitioning aims at enhancing the content based information with respect to the application, and not to describe the image entirely. In the context of the proposed framework, the challenge is to identify the image content representation that best fit the user understanding. Ideally, the number of computed classes must equal the number of visual element categories. The rate distortion theory will provide an estimate of the best possible representation achievable. Such an estimation of the optimum number of classes for Sentinel 2 data classification was presented in [19]. For the current paper we consider the same approach. We perform a k-means classification using Gabor_Histogram [18] as descriptor. The number of classes will range between 2 and 100 and the mean square error will be computed between two image classification results obtained with consecutive numbers. The unsupervised clustering process will reveal the patterns that are naturally grouping by data processing. The resulted associations are entirely based on statistical similarities and they offer the machine's perspective on the data. The Gabor-Histogram descriptor provides a complex description of the content characteristics because it underpins spectral and texture features at the same time. Experiments have shown that data analysis approaches based on this descriptor perform well, guaranteeing a precision above the mean for several types of EO data [18].

The rest of the distortion could be adjusted further through a proper tailoring of the feature space representation for the data content. Spectral, texture or contextual characteristics are influencing the image representation, in both cases of data numerical processing and visual analysis. Each feature extraction algorithm will highlight a particularity of the image and each feature classification method will imprint on the way the identified patterns will cluster. Then, if the feature extraction is generating a low dimensional space with a lot of

points, classification methods such as k-NN tend to perform a better separation. Instead, for a high dimensional feature space with few points, SVM algorithms will outperform the rest of the classifiers [41]. Altogether, these methods will act upon encoding the information in a hierarchical structure, where the intermediate layers are connected such that they restore the relationships perceived visually by the human eye. Nevertheless, the relevant information is changing from an application to another and depends on the properties of the data. An important parameter refers to the selection of the techniques that estimate best the associations of visual perception.

Defining an algorithm able to automatically determine perceivable interactions is rather utopic in the context of EO data. It is more likely to learn the semantics of an object from their appearance. Pattern recognition has been successfully integrated into the classification process [17]. Therefore, the last parameter required to bridge the semantic gap is a training data set to illustrate the human vision with respect on the representative categories of elements that carry a meaning. The selected samples must contain observable signs and symbols whose interaction is able to generate relevant information for each of the semantic categories considered. We established that the proper number of distinguishable and understandable classes is estimated through a rate distortion analysis.

The modularity of the proposed framework reflects in a divided process, where each step has a clear focus and can be adjusted based on a well-defined parameter. This approach results in a hierarchical representation of content data. In order to accommodate with the mathematical processing, the human interaction with the computer will follow the same hierarchy and guide the semantic annotation layer by layer.

VI. HUMAN MACHINE COMMUNICATION

The human-machine communication (HMC) is performed by means of graphical user interface (GUI). This creates automat associations between human reasoning and the numeric representation. The semantic meaning will correspond to a high level of image abstraction. The HMC concept is amplifying the capabilities of the proposed IIM framework, as the user is passing its knowledge to the algorithms, acting similar to a high level semantic indexing procedure.

The dialogue between the user and the computer is based on a language with discrete entities approximating the semantic meaning of human language. These entities are defined in terms of feature vectors and they express content characteristics that the user is able to confirm visually. The implementation of such dialogues in CBIR approaches is difficult, but the user integration into the processing cycle will reduce the semantic gap between data representation and the actual meaning of the data. The development of a HMC mechanism is important, together with dedicated algorithms to define "signal – symbol – signal" analogies. The knowledge transfer will be provided and the correct perception of the

information will depend on its representation and not on its shape.

In the perspective of information theory, EO image is the result of a stochastic process and the pixels are a realization of a random variable that carries information in the form of uncertainty. The image information extraction implies the detection of the land cover structure defined by the recorded measurements. From the machine's point of view, this is a mathematical modelling problem. The human instead, will address the subject like a perceptual process, mainly based on visual interpretation and cognitive modeling. To this aim, it is imperative to define explanatory conventions for the content of an image, such that both mathematical and perceptual association are correctly expressed. We assume that a symbolic representation is the most appropriate way to express the common understanding.

There are four main categories of symbols to be employed: the patches (considered the smallest element of the images), the descriptors (feature vectors extracted from the patches), the classes of elements (groups of similar descriptors, patches sharing similar visual characteristics), semantic labels (classes of elements annotated by the user, in agreement to his perception). The machine is extracting step by step these categories of symbols from the data. The HMC is hierarchically performed. The human interaction is required at each level and it can be defined as a parameter setting. The information representation is shown in FIGURE 9.

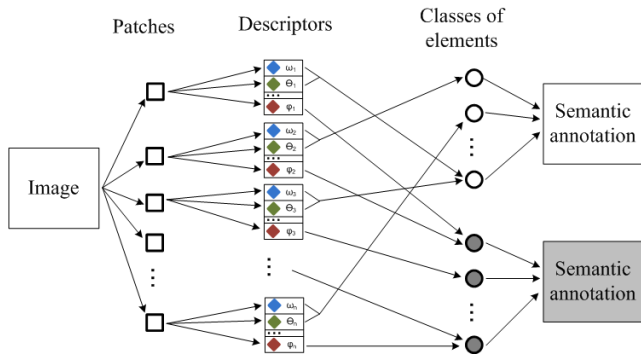


FIGURE 9. Information hierarchic representation based on the symbol categories.

The way that the information is compressed in symbolic representation resembles to a coding and decoding process. The image is resumed to a series of component elements that are further clustered by means of similarities. New content composing rules are discovered and the focus will be on restoring the image in the shape of a map trying to reproduce the semantic content of the original image. A communication channel may be imagined (FIGURE 10). The image represents the transmitter and the semantic annotation, the receiver. The information content is gradually coded, such that the correct message reach the destination. Any losses due to the symbolic approximation are considered to be a perturbation effect.

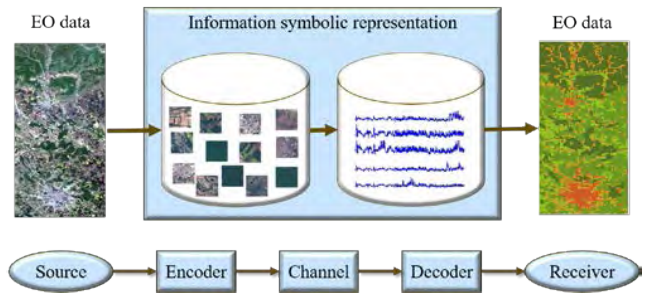


FIGURE 10. Image information process as a communication channel.

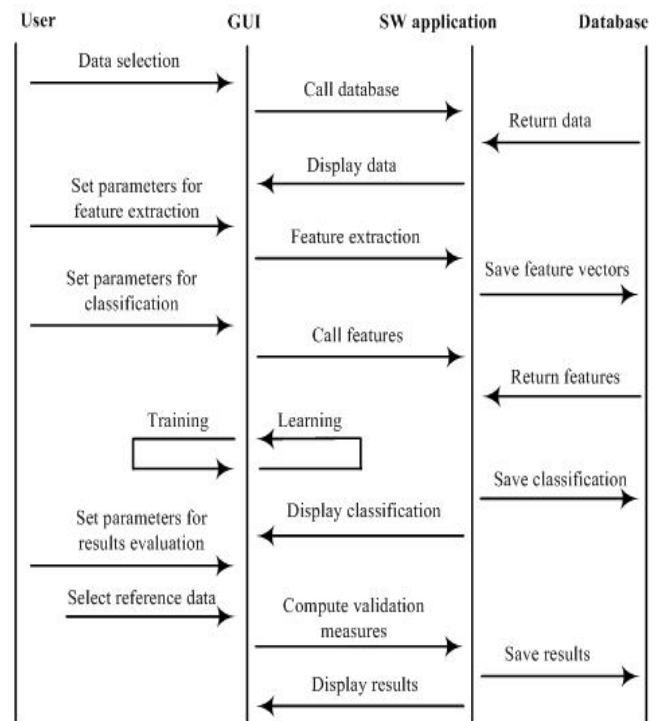


FIGURE 11. State diagram for human – computer dialogue.

The proposed image information mining framework is wrapped around such a communication channel where the computer tries to discover the relevant information about land cover hidden in the EO data. As expected, transmission is not perfect. Nevertheless, the perturbation effect is limited through the user interaction. Appropriate tools are provided by the GUI to the human in order to help him assist the machine during the process. A dialogue begins: the user is setting the parameters, while the computer provides adequate response. The values defined will be considered as input data. by the software modules. The processing is completed and the machine returns the results to the user through the GUI. During the dialogue, the database is accessed to read the data and save the feature vectors and semantic annotation. The GUI acts like a connection point, translating the user's vocabulary to the computer and vice versa. FIGURE 11 presents a state diagram for the human – computer interaction. Each step

TABLE 1. Communication methods and information representation modes.

Communication methods	Information representation	Properties
Dialogue (through GUI access)		User – System System – User
Visualization	Signals (images display)	Multi – scale
	Signals (images display)	Multi – sensor
Interactive learning	Signs(pointing patches/indexes)	Training samples
Knowledge transfer	Codes (descriptors)	Set parameters for feature extraction
	Symbols (semantic labels)	Set parameters for classification
	Signs	Training samples
	Symbols (semantic labels)	Reference data

requires that a certain type of information to be transmitted between two component elements. TABLE 1 outlines the main technique employed to transmit the information and the relevant category of symbolic representation for each step.

VII. PROPOSED WORKFLOW

After presenting all the details of a conceptual framework for EO image information extraction, the paper continues with the engineering approach and experimental results. For the implementation of a prototype software system, we used “.Net” environment and open source libraries like GDAL and OpenCV.

The proposed architecture follows the same logic as the conceptual framework in FIGURE 4, with the main difference that process is not structured based on types of analysis (i.e. feature extraction, feature classification), but on groups of activities controlled entirely by a specific parameter. The diagram illustrating the workflow (FIGURE 12) includes six main software modules. 1) Data reader module, 2) Patch selection module, 3) Module for detection of optimum number of classes, 4) Module for detection of the best feature extraction method suitable for the input dataset, 5) Feature extraction and classification module, 6) Graphical user interface. The order to set the parameters during the data processing determines the interaction between the software components.

In the beginning of the process, the machine must identify the image type (i.e. multispectral or radar image) and convert it to a specific data format in order to browse the metadata and plot the geospatial data. Further, the definition of the patch size will enable the software to display a grid over the image. The user will be able to assess the computer understanding on the image partitioning and decide what the best patch size for his interest is. Once this step is completed, the computer will enter an iterative process with the purpose of pointing the optimum number of classes with respect to the analyzed data. When displayed, the rate distortion function will indicate an interval used for choosing the best number of classes in the analyzed image. After deciding how many relevant classes

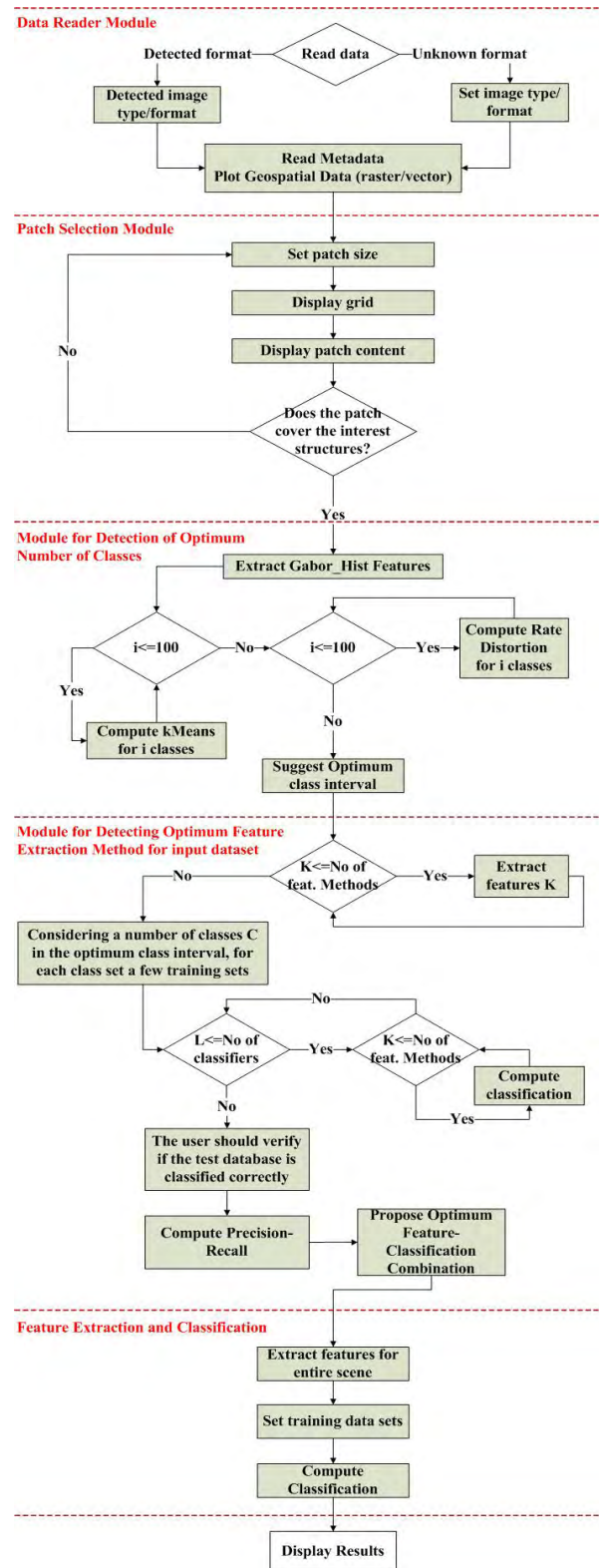


FIGURE 12. The diagram of the proposed workflow.

he is able to visualize inside the data, the user is compelled to build a dataset that includes samples from every class. Part of this selection will be employed as a training set for supervised

classification methods, while the entire collection of samples will stand for a reference data set. The feature extraction and feature classification algorithms will complete a puzzle that the computer must solve. The solution consists in the data analysis procedure that provides the higher percentages for precision and recall. Once identified, it will be extended to the full EO image. The final scene classification will be displayed to the user.

VIII. EXPERIMENTS AND DISCUSSION

In order to illustrate the capabilities of the described framework, we propose an assessment of a 2500 sq. km area picturing Bucharest, Romania, and surroundings. For this particular region of interest we considered 4 images, acquired with 4 different EO sensors: Envisat, Sentinel 1, Landsat 8 and Sentinel 2. The purpose of the experiment is to demonstrate the flexibility of the approach when dealing with heterogeneous data sets (both multispectral and synthetic aperture radar imagery).

Each image will be modeled using the same number of classes. We opt for a limited number of general land cover classes correlated to visual elements carrying a semantic meaning to the user: forest, agriculture, water, sparse urban and dense urban. The plan is to study the prototype system's behavior under real experimental conditions. For each EO image, the test requires the identification of the most efficient data processing technique. The obtained results are expected to be similar, but not identical. Much of the differences are due to the sensory gap (different sensors measure the Earth's surface differently). Moreover, each type of data will require a particular analysis technique, generating even more distortions between the final results.

Since the experiment envisages the retrieval of classes with general meaning, the patch is compelled to include only patterns representative for the selected class. They will be associated to visual symbols. Therefore, we will link the size of the patch to the size of a ground area and preserve the value for all the analyzed images. For this experiment, we consider a 60×60 square meters area as a relevant visual symbol.

A reference dataset of 630 patches was built for each image in part. The selection of patches was thus made simultaneously from all the images. The user selected the visually relevant samples for each class by looking at a single image. The machine withhold the geographical coordinates and cut the same area from the other images. Human perception may be influenced by the results of the imaging technique provided by different EO sensors. For the sake of the experiment, the purpose was to avoid supplementary distortions and the computer to learn about the data and model the scene according to the same amount of knowledge. The training set includes 10 samples of each class, while the entire user selection goes up to 140 patches for individual classes picturing forest, agriculture, sparse urban and dense urban. The poor water coverage identified in the area resulted in a smaller set, of only 70 patches comprising that class.

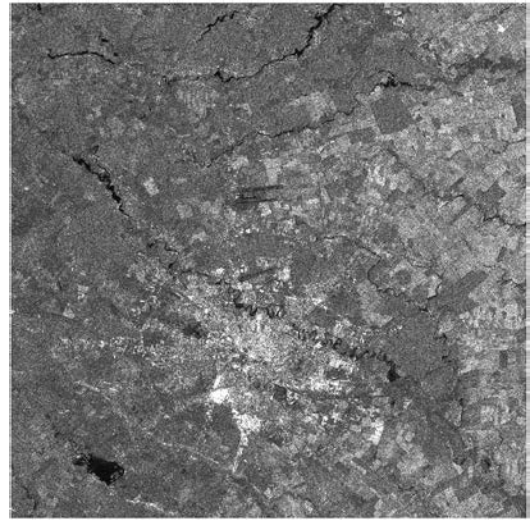


FIGURE 13. ENVISAT image, Bucharest area and surroundings.

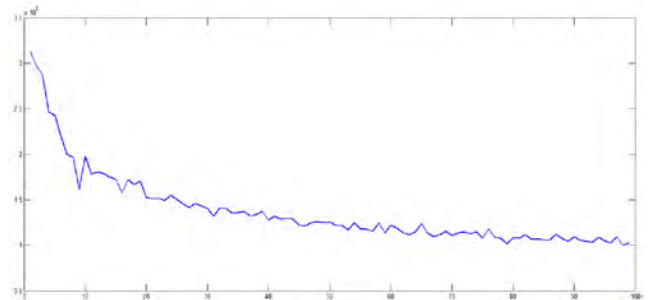


FIGURE 14. The rate distortion function for the Envisat image content classification. The optimum class number lies between 5 and 15.

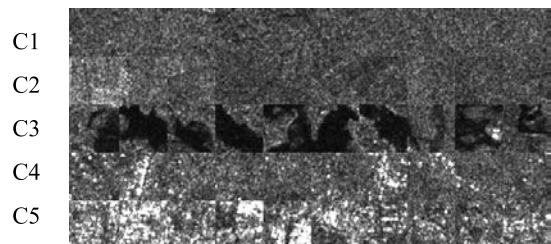


FIGURE 15. Samples form the training set used for the analysis of the ENVISAT image. The selected classes are: forest (C1), agriculture (C2), water (C3), sparse urban (C4), dense urban (C5).

A. 1st USE CASE SCENARIO: ENVISAT IMAGE ANALYSIS

For the first use case scenario, the analyzed data was provided by the Envisat sensor. This is the case of a 30 m spatial resolution synthetic aperture radar image. A preview of the data is illustrated in FIGURE 13. The size of the patch is 60×60 pixels. The rate distortion analysis recommends that the image classification to have at least 5 semantic categories of elements, but no more than 15 (FIGURE 14). Consequently, the initial setting ranges inside the optimum interval.

FIGURE 15 provides an insight on the sensor's record over the areas containing forest, agriculture, water, sparse urban and dense urban.

TABLE 2. Precision - recall for the ENVISAT image analysis.

SVM						
		C1	C2	C3	C4	C5
G	P	82.14%	46.43%	88.57%	60.71%	68.57%
	R	63.54%	69.89%	68.89%	62.04%	74.42%
WLD	P	57.14%	39.29%	48.57%	53.57%	76.43%
	R	43.24%	43.31%	42.50%	74.26%	78.10%
H	P	97.14%	28.57%	90.00%	48.57%	72.14%
	R	51.13%	60.61%	75.90%	69.39%	86.32%
GH	P	95.00%	28.57%	92.86%	57.86%	78.57%
	R	56.84%	68.97%	67.71%	74.31%	82.71%
WLDH	P	68.57%	40.71%	87.14%	67.86%	85.00%
	R	49.74%	57.58%	83.56%	71.43%	90.15%
k-NN						
G	P	98.57%	30.00%	72.86%	53.57%	67.86%
	R	50.92%	55.26%	96.23%	66.96%	80.51%
WLD	P	65.00%	39.29%	58.57%	12.86%	71.43%
	R	44.17%	41.35%	30.60%	85.71%	73.53%
H	P	94.29%	28.57%	84.29%	55.71%	65.00%
	R	50.57%	45.98%	96.72%	67.24%	86.67%
GH	P	97.14%	28.57%	80.00%	59.29%	66.43%
	R	50.94%	51.95%	100.00%	68.03%	86.11%
WLDH	P	93.57%	27.14%	80.00%	49.29%	80.71%
	R	48.16%	69.09%	81.16%	70.41%	83.09%

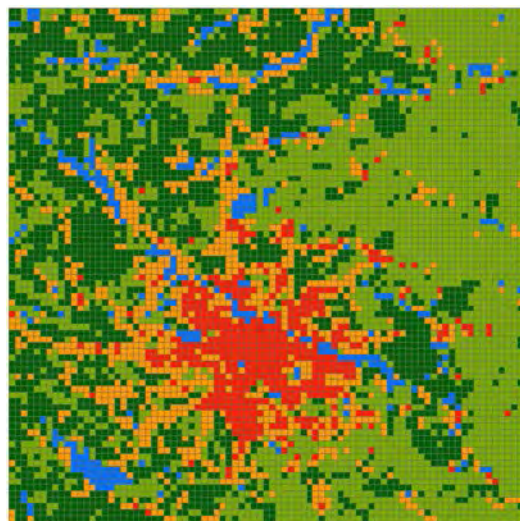


FIGURE 17. ENVISAT image classification using GH descriptor and SVM classifier. Legend: C1-dark green, C2-light green, C3-blue, C4-orange, C5-red.

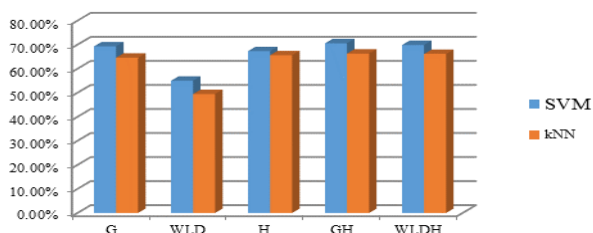


FIGURE 16. k-NN vs SVM classification of the ENVISAT image.

Further, the data was modeled using all the combinations feature extraction – feature classification that can be achieved using the algorithms included in the prototype software. The list includes 2 supervised classifiers and 5 feature extraction algorithms, excluding the methods dedicated to the multispectral analysis (Spectral Indexes and Bag of Spectral Indexes). The precision and recall computed against the reference dataset is presented in TABLE 2. We can observe from the table that all the patches including water were fully retrieved. However, SVM learning based on Gabor_Histogram descriptor has the highest average recognition of all the 5 classes (FIGURE 16). We applied this procedure to the entire scene and the result is illustrated in FIGURE 17.

B. 2nd USE CASE SCENARIO: SENTINEL 1 IMAGE ANALYSIS

In the second use case scenario, the focus is also on a synthetic aperture radar image (FIGURE 18). The acquisition sensor, Sentinel 1, encourages the discernment of details up to 10 m spatial resolution. The size of the patch is 60x60 pixels. Following the rate distortion analysis it appears that the optimum number of classes ranges between 5 and 15 (FIGURE 19). The initial setting ranges inside the optimum interval.

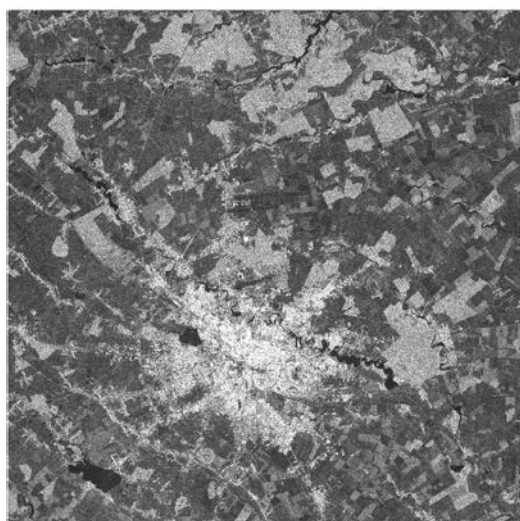


FIGURE 18. Sentinel 1 image, Bucharest area and surroundings.

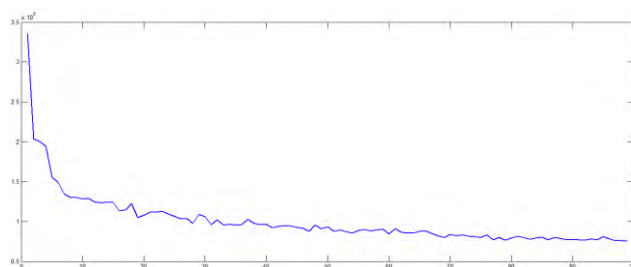


FIGURE 19. The rate distortion function for the Sentinel 1 image content classification. The optimum class number lies between 5 and 15.

FIGURE 20 offers a preview of the training set employed for this experiment. As in the previous case, only 5 of the available feature extraction algorithms are applicable

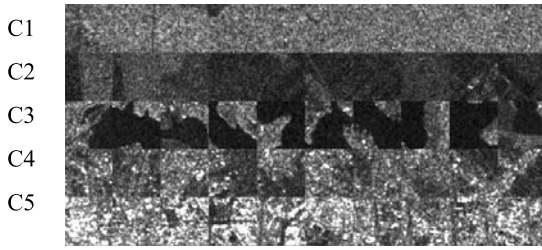


FIGURE 20. Samples from the training set used for the analysis of the Sentinel 1 image. The selected classes are: forest (C1), agriculture (C2), water (C3), sparse urban (C4), dense urban (C5).

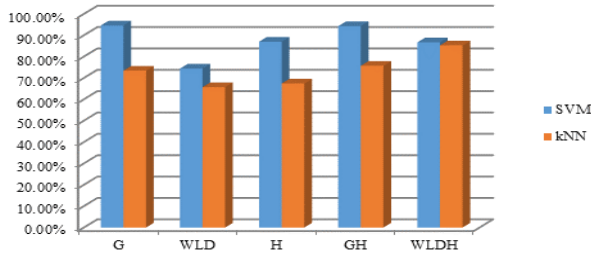


FIGURE 21. k-NN vs SVM classification of the Sentinel 1 image.

TABLE 3. Precision - recall for the sentinel 1 image analysis.

SVM						
		C1	C2	C3	C4	C5
G	P	97.14%	97.86%	94.29%	96.43%	87.14%
	R	99.27%	97.16%	95.65%	85.99%	96.83%
WLD	P	78.57%	59.29%	64.29%	86.43%	83.57%
	R	65.87%	65.87%	63.38%	85.21%	94.35%
H	P	95.00%	75.00%	91.43%	85.00%	88.57%
	R	91.72%	92.11%	62.14%	86.23%	95.38%
GH	P	95.71%	95.00%	95.71%	97.14%	87.86%
	R	96.40%	95.00%	97.10%	87.18%	97.62%
WLDH	P	96.43%	92.86%	62.86%	87.86%	93.57%
	R	94.41%	81.76%	78.57%	91.79%	94.93%
k-NN						
G	P	97.86%	96.43%	27.14%	64.29%	81.43%
	R	83.03%	69.59%	52.78%	76.92%	96.61%
WLD	P	85.00%	42.86%	38.57%	75.00%	87.14%
	R	54.59%	68.97%	52.94%	82.68%	82.99%
H	P	91.43%	82.86%	11.43%	72.86%	78.57%
	R	96.24%	65.54%	13.79%	68.00%	98.21%
GH	P	95.00%	96.43%	14.29%	90.00%	82.86%
	R	95.68%	67.84%	62.50%	80.25%	97.48%
WLDH	P	96.43%	92.86%	55.71%	91.43%	90.00%
	R	94.41%	79.75%	84.78%	87.07%	96.18%

for SAR imagery. The feature extraction – feature classification combination that provides the highest average semantic recognition is Gabor descriptor – SVM classifier (FIGURE 21). TABLE 3 presents a more detailed information regarding the precision and recall of the other combinations. The Sentinel 1 image classification using Gabor descriptor and SVM classifier is depicted in FIGURE 22.

C. 3rd USE CASE SCENARIO: LANDSAT 8 IMAGE ANALYSIS

The third experiment will address the analysis of a 30m spatial resolution optical image (FIGURE 23). Landsat 8 is

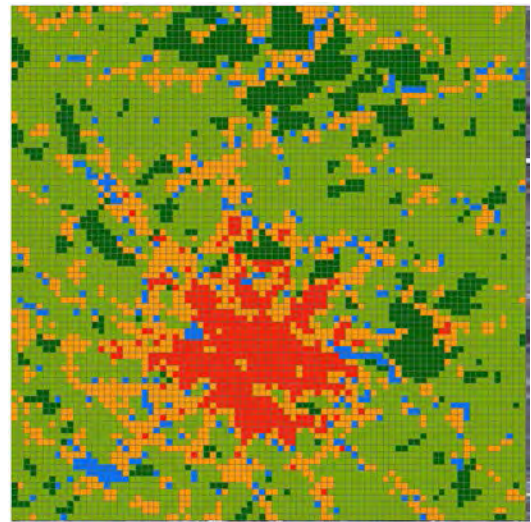


FIGURE 22. Sentinel 1 image classification using Gabor descriptor and SVM classifier. Legend: C1-dark green, C2-light green, C3-blue, C4-orange, C5-red.



FIGURE 23. Landsat 8 image, Bucharest area and surroundings.

the acquisition sensor. The size of the patch is 20x20 pixels. Samples from the training set are shown in FIGURE 25. As expected, the multispectral images enable the discernment of more semantic classes. The interval suggested through the rate distortion analysis lies from 5 to 20 (FIGURE 24). The information extraction process is a little more performant when dealing with multispectral images. There are few situations when one or two classes are retrieved 100% (TABLE 4). However, the highest average recognition of all the 5 classes is obtained when the BSI descriptors are grouped using the SVM learning method (FIGURE 26). The final Landsat 8 image classification result is presented in FIGURE 27.

D. 4th USE CASE SCENARIO: SENTINEL 2 IMAGE ANALYSIS

For the last use case scenario we considered a Sentinel 2 multispectral image (FIGURE 28). All the spectral bands

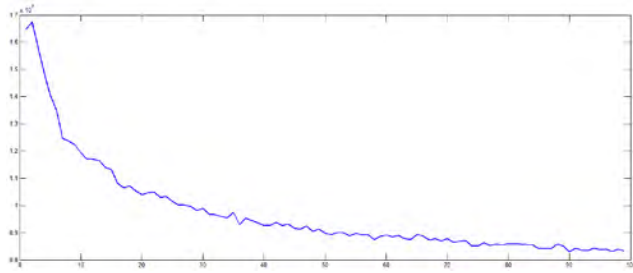


FIGURE 24. The rate distortion function for the Landsat 8 image content classification. The optimum class number lies between 5 and 20.

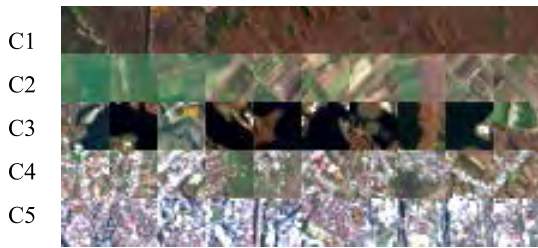


FIGURE 25. Samples from the training set used for the analysis of the Landsat 8 image. The selected classes are: forest (C1), agriculture (C2), water (C3), sparse urban (C4), dense urban (C5).

TABLE 4. Precision - recall for the landsat 8 image analysis.

		SVM				
		C1	C2	C3	C4	C5
G	P	97.86%	95.00%	100.00%	85.00%	97.86%
	R	95.14%	87.50%	98.59%	94.44%	100.00%
WLD	P	92.86%	79.29%	80.00%	92.86%	80.71%
	R	81.25%	84.09%	88.89%	81.76%	97.41%
H	P	27.14%	14.29%	95.71%	40.00%	7.86%
	R	27.54%	20.41%	98.53%	31.28%	7.48%
GH	P	97.14%	88.57%	98.57%	76.43%	82.14%
	R	91.89%	81.58%	97.18%	78.68%	93.50%
WLDH	P	72.14%	42.86%	92.86%	75.71%	85.00%
	R	50.50%	72.29%	100.00%	70.20%	90.84%
SI	P	97.14%	92.14%	97.14%	65.71%	97.14%
	R	92.52%	74.57%	95.77%	90.20%	99.27%
BSI	P	100.00%	90.71%	100.00%	90.71%	99.29%
	R	93.33%	92.03%	100.00%	95.49%	100.00%
		k-NN				
G	P	99.29%	70.71%	71.43%	75.71%	94.29%
	R	78.53%	78.57%	100.00%	79.70%	91.67%
WLD	P	95.71%	23.57%	60.00%	89.29%	76.43%
	R	52.96%	58.93%	93.33%	76.22%	95.54%
H	P	20.00%	16.43%	80.00%	69.29%	17.86%
	R	30.43%	21.90%	100.00%	30.89%	39.68%
GH	P	98.57%	90.00%	90.00%	58.57%	77.86%
	R	86.25%	69.23%	100.00%	78.85%	90.08%
WLDH	P	79.29%	27.14%	80.00%	76.43%	67.14%
	R	46.64%	74.51%	87.50%	60.11%	94.95%
SI	P	100.00%	27.14%	81.43%	97.86%	92.86%
	R	87.50%	100.00%	100.00%	57.32%	95.59%
BSI	P	99.29%	73.57%	100.00%	86.43%	98.57%
	R	80.35%	97.17%	81.40%	95.28%	100.00%

were resampled to the best spatial resolution that is 10 m. Consequently, the patch size is 60×60 pixels. The rate distortion function is pointing a 5 to 20 number of classes (FIGURE 29). Samples from the training set are shown in

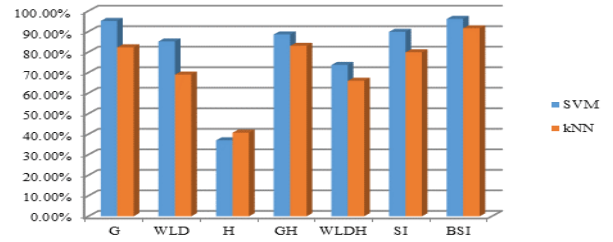


FIGURE 26. k-NN vs SVM classification of the Landsat 8 image.

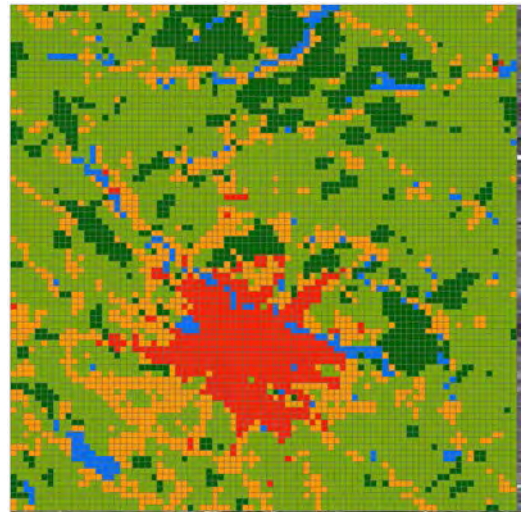


FIGURE 27. Landsat 8 image classification using BSI descriptor and SVM classifier. Legend: C1-dark green, C2-light green, C3-blue, C4-orange, C5-red.



FIGURE 28. Sentinel 2 image, Bucharest area and surroundings.

FIGURE 30. TABLE 5 explains how well the data processing adjust to the human perception. The highest average retrieval is provided when the content is described in terms of Gabor descriptor and the semantic associations are performed by the SVM classifier (FIGURE 31). FIGURE 32 illustrates the best achievable classification result for the Sentinel 2 image.

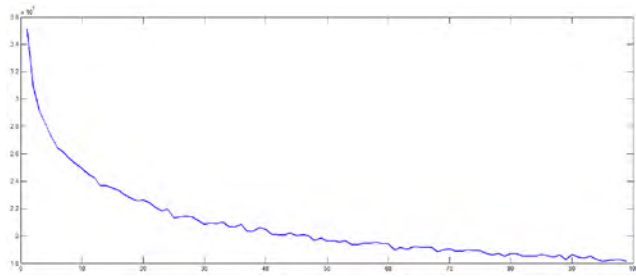


FIGURE 29. The rate distortion function for the Sentinel 2 image content classification. The optimum class number lies between 5 and 20.



FIGURE 30. Samples form the training set used for the analysis of the Sentinel 2 image. The selected classes are: forest (C1), agriculture (C2), water (C3), sparse urban (C4), dense urban (C5).

TABLE 5. Precision - recall for the sentinel 2 image analysis.

SVM						
		C1	C2	C3	C4	C5
G	P	93.57%	94.29%	98.57%	96.43%	96.43%
	R	94.93%	95.65%	89.61%	96.43%	98.54%
WLD	P	87.14%	92.86%	100.00%	99.29%	89.29%
	R	94.57%	100.00%	90.91%	82.25%	100.00%
H	P	82.86%	75.00%	100.00%	90.00%	95.71%
	R	98.31%	88.24%	100.00%	66.67%	100.00%
GH	P	85.71%	79.29%	100.00%	87.14%	92.86%
	R	96.77%	84.73%	100.00%	69.71%	100.00%
WLDH	P	90.71%	90.00%	100.00%	95.00%	92.14%
	R	95.49%	91.97%	100.00%	82.61%	100.00%
SI	P	85.00%	55.71%	100.00%	95.00%	96.43%
	R	98.35%	88.64%	78.65%	67.51%	100.00%
BSI	P	81.43%	89.29%	100.00%	98.57%	96.43%
	R	98.28%	82.24%	95.89%	89.61%	100.00%
k-NN						
G	P	100.00%	73.57%	47.14%	99.29%	85.71%
	R	83.83%	99.04%	100.00%	75.96%	83.92%
WLD	P	97.14%	27.14%	91.43%	96.43%	93.57%
	R	56.20%	100.00%	100.00%	90.00%	96.32%
H	P	85.00%	33.57%	92.86%	100.00%	95.00%
	R	95.20%	100.00%	91.55%	57.14%	93.66%
GH	P	88.57%	41.43%	92.86%	100.00%	95.00%
	R	94.66%	100.00%	100.00%	59.07%	95.68%
WLDH	P	89.29%	59.29%	98.57%	100.00%	97.14%
	R	93.28%	100.00%	94.52%	68.97%	99.27%
SI	P	100.00%	34.29%	84.29%	100.00%	90.71%
	R	98.59%	100.00%	100.00%	57.61%	92.03%
BSI	P	100.00%	93.57%	100.00%	90.00%	93.57%
	R	97.22%	92.91%	89.74%	92.65%	100.00%

E. DISCUSSIONS

There is a certain similarity between the 4 classification maps. There are visual elements expressing well known land structures whose symbolic representations are correlated with the user knowledge in all the experiments. *Morii*,

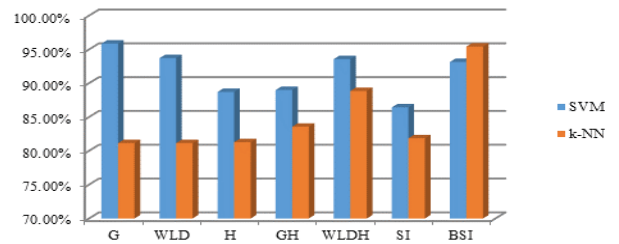


FIGURE 31. k-NN vs SVM classification of the Sentinel 2 image.

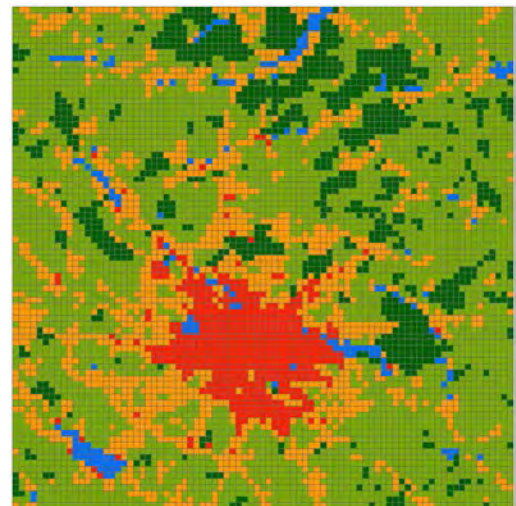


FIGURE 32. Sentinel 2 image classification using Gabor descriptor and SVM classifier. Legend: C1-dark green, C2-light green, C3-blue, C4-orange, C5-red.

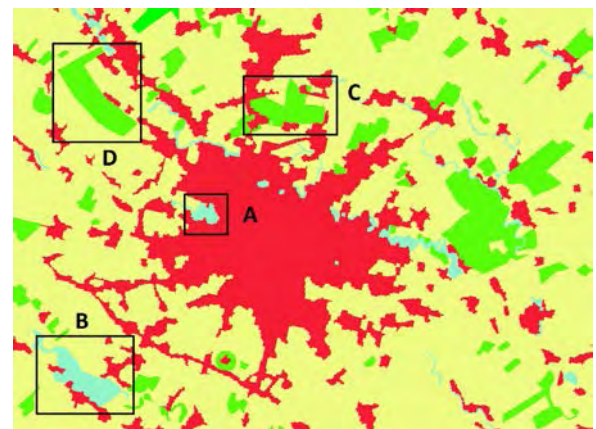


FIGURE 33. Corine Land Cover reference map. A-Morii Lake; B-Mihailesti Lake; C-Baneasa Forest; D-Raioasa Forest.

Lake, Mihailesti Lake, Baneasa Forest or Raioasa Forest (FIGURE 33) are reference elements for the analyzed area.

TABLE 6 introduces an analytical comparison between the average achievable precision and recall for the described experiments. A rate of the obtained semantic coverage proves the capability of the framework to learn visual perception from the human interaction.

TABLE 6. Analytical comparison between the scene classification using four different EO images.

Sensor	Mean Precision	Mean Recall	Mean Coverage				
			C1	C2	C3	C4	C5
Envisat	70.57%	70.11%	31.92%	39.77%	4.80%	15.98%	7.52%
Sentinel 1	94.57%	94.98%	12.57%	55.55%	4.43%	20.09%	7.36%
Landsat 8	96.14%	96.17%	14.76%	56.47%	3.34%	17.87%	7.56%
Sentinel 2	95.86%	95.03%	14.40%	52.97%	2.64%	22.73%	7.26%

Besides the flexibility and versatility, the approach has two important drawbacks. Firstly, the user must put more effort to create a relevant test dataset. This has to include a significant number of samples, larger than the usual training set. Secondly, the results obtained using the created reference dataset are roughly approximating the performance on data modelling. It is likely that the values or precision recall to decrease when the analysis is applied on a larger data set. However, they will indicate the most relevant descriptor-classifier combination for the content of the analyzed image.

IX. CONCLUSION

The paper advances the main aspects and shortcomings arising in the field of EO data processing. Intensively studied over the years, scientific progress is still required as the continuous development of remote sensing technologies yields new challenges with every new satellite mission. The heterogeneity of acquired data is particularly confusing due to the different perception and information the sensors are providing as compared to the human awareness. The literature calls this issue “sensory gap” and its main consequence consists in the fact that it prevents the definition of a general procedure for information mining from EO data collection. In this paper, the authors are summarizing all the related issues and their general impact on image understanding. Further, they present an integrated framework for image information mining and knowledge discovery from EO database to overcome them. The proposed procedure is targeting the analysis of EO data in general, offering flexibility and interoperability to the user, allowing him to design and develop unique application scenarios, according to his needs. The architecture is carefully described, with emphasis on the role and the functionality of the five component modules. A review of state of the art techniques for the main processes was meant to support the selection of appropriate algorithms for each function. The proposed approach adds human vision to the algorithmically computation to connect visual elements and similarities obtained by data processing. A dialogue is envisaged to support the knowledge transfer from the user to the computer that results in a hierarchical representation of the data content. The communication is performed by means of a graphical interface and wrapped around a mechanism to define “signal-symbol-signal” analogies that enables semantic annotations. After presenting all the details of the conceptual framework

for EO image information extraction, the paper continues with the engineering approach and a series of experimental results. There were considered 4 EO images centered on the area of Bucharest city, Romania, acquired by the following sensors: Envisat, Landsat 8, Sentinel 1 and Sentinel 2. The proposed approach has proven to be successful for each of the presented use case scenarios.

REFERENCES

- [1] M. Quartulli and I. G. Olaizola, “A review of EO image information mining,” *ISPRS J. Photogram. Remote Sens.*, vol. 75, pp. 11–28, Jan. 2013.
- [2] M. Schroder, H. Rehrauer, K. Seidel, and M. Datcu, “Interactive learning and probabilistic retrieval in remote sensing image archives,” *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 5, pp. 2288–2298, Sep. 2000.
- [3] M. Datcu et al., “EliA, “Information mining in remote sensing image archives: System concepts,” *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 12, pp. 2923–2936, Dec. 2003.
- [4] G. Sheikholeslami, W. Chang, and A. Zhang, “SemQuery: Semantic clustering and querying on heterogeneous features for visual data,” *IEEE Trans. Knowl. Data Eng.*, vol. 14, no. 5, pp. 988–1002, Sep. 2002.
- [5] S. Aksoy, K. Koperski, C. Tusk, G. Marchisio, and J. C. Tilton, “Learning Bayesian classifiers for scene classification with a visual grammar,” *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 581–589, Mar. 2005.
- [6] C.-R. Shyu, M. Klaric, G. J. Scott, A. S. Barb, C. H. Davis, and K. Palaniappan, “GeoIRIS: Geospatial information retrieval and indexing system—Content mining, semantics modeling, and complex queries,” *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 839–852, Apr. 2007.
- [7] A. Fahad et al., “A survey of clustering algorithms for big data: Taxonomy and empirical analysis,” *IEEE Trans. Emerg. Topics Comput.*, vol. 2, no. 3, pp. 267–279, Sep. 2014.
- [8] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, “Color and texture descriptors,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 703–715, Jun. 2001.
- [9] K. van de Sande, T. Gevers, and C. Snoek, “Evaluating color descriptors for object and scene recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1582–1596, Sep. 2010.
- [10] D. Espinoza-Molina et al., “Very-high-resolution SAR images and linked open data analytics based on ontologies,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 4, pp. 1696–1708, Apr. 2015.
- [11] Y. Rui, T. S. Huang, and S.-F. Chang, “Image retrieval: Current techniques, promising directions, and open issues,” *J. Vis. Commun. Image Represent.*, vol. 10, no. 1, pp. 39–62, Mar. 1999.
- [12] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, “Content-based image retrieval at the end of the early years,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000.
- [13] R. Datta, D. Joshi, J. Li, and J. Wang, “Image retrieval: Ideas, influences, and trends of the new age,” *ACM Comput. Surv.*, vol. 40, no. 2, Apr. 2008, Art. no. 5.
- [14] D. Peijun, C. Yunhao, T. Hong, and F. Tao, “Study on content-based remote sensing image retrieval,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2005, pp. 707–710.
- [15] D. Faur, I. Gavut, and M. Datcu, “Salient remote sensing image segmentation based on rate-distortion measure,” *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 855–859, Oct. 2009.
- [16] S.-F. Chang, A. Eleftheriadis, and R. McClintock, “Next-generation content representation, creation, and searching for new-media applications in education,” *Proc. IEEE*, vol. 86, no. 5, pp. 884–904, May 1998.
- [17] N. Vasconcelos and A. Lippman, “A probabilistic architecture for content-based image retrieval,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2000, pp. 216–221.
- [18] F.-A. Georgescu, C. Vaduva, D. Răducanu, and M. Datcu, “Feature extraction for patch-based classification of multispectral earth observation images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 6, pp. 865–869, Jun. 2016.
- [19] F. A. Georgescu, M. Datcu, and D. Răducanu, “Patch-based multispectral image classification assessment for sentinel-2 data analysis,” in *Proc. Int. Conf. Big Data Space (BiDS)*, Santa Cruz de Tenerife, Spain, Mar. 2016, pp. 81–84.

- [20] R. Bahmanyar, and A. M. M. de Oca, "Evaluating the sensory gap for earth observation images using human perception and an LDA-based computational model," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 566–570.
- [21] A. A. Popescu, I. Gavatu, and M. Datcu, "Contextual descriptors for scene classes in very high resolution SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 1, pp. 80–84, Jan. 2012.
- [22] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.
- [23] Y. Q. Chen, M. S. Nixon, and D. W. Thomas, "Statistical geometrical features for texture classification," *Pattern Recognit.*, vol. 28, no. 4, pp. 537–552, Apr. 1995.
- [24] Z.-C. Huang, P. P. K. Chan, W. W. Y. Ng, and D. S. Yeung, "Content-based image retrieval using color moment and Gabor texture feature," in *Proc. IEEE Int. Conf. Mach. Learn.*, Jul. 2010, pp. 719–724.
- [25] G. Pass, R. Zabih, and J. Miller, "Comparing images using color coherence vectors," in *Proc. 4th ACM Int. Conf. Multimedia*, 1997, pp. 65–73.
- [26] J. Huang, S. R. Kumar, and M. Mitra, "Combining supervised learning with color correlograms for content-based image retrieval," in *Proc. 4th ACM Int. Conf. Multimedia*, 1997, pp. 325–334.
- [27] F. Alamdar and M. Keyvanpour, "A new color feature extraction method based on dynamic color distribution entropy of neighborhoods," *Int. J. Comput. Sci. Issues*, vol. 8, no. 5, pp. 42–48, Sep. 2011.
- [28] G. Marchisio, F. Pacifici, and C. Padwick, "On the relative predictive value of the new spectral bands in the WorldView-2 sensor," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2010, pp. 2723–2726.
- [29] S. Cui, G. Schwarz, and M. Datcu, "Remote sensing image classification: No features, no clustering," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 11, pp. 5158–5170, Nov. 2015.
- [30] N. Kurinjivendhan and K. Thangadurai, "Modified K-means algorithm and genetic approach for cluster optimization," in *Proc. Int. Conf. Data Mining Adv. Comput. (SAPIENCE)*, Mar. 2016, pp. 53–56.
- [31] M. Nixon and A. Aguado, *Feature Extraction and Image Processing*. San Diego, CA, USA: Academic, 2002.
- [32] Y. Zhou, H. Chen, and Q. Zhu, "The research of classification algorithm based on fuzzy clustering and neural network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, vol. 4, Jun. 2002, pp. 2525–2527.
- [33] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2006.
- [34] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, vol. 1, 1967, pp. 281–297.
- [35] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [36] A. Mathur and G. M. Foody, "Multiclass and binary SVM classification: Implications for training and classification users," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 241–245, Apr. 2008.
- [37] P. Hart, "The condensed nearest neighbor rule," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 3, pp. 515–516, May 1968.
- [38] P. Mitra, C. A. Murthy, and S. K. Pal, "Unsupervised feature selection using feature similarity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 301–312, Mar. 2002.
- [39] Y. Liu, W. Chen, P. Arendt, and H.-Z. Huang, "Toward a better understanding of model validation metrics," *J. Mech. Design*, vol. 133, no. 7, pp. 071005-1–071005-13, Jul. 2011.
- [40] D. M. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, 2011.
- [41] W. Aoudi and A. M. Barbar, "Support vector machines: A distance-based approach to multi-class classification," in *Proc. IEEE Int. Multidisciplinary Conf. Eng. Technol. (IMCET)*, Nov. 2016, pp. 75–80.
- [42] N. M. Ali, S. W. Jun, M. S. Karis, M. M. Ghazaly, and M. S. M. Aras, "Object classification and recognition using Bag-of-Words (BoW) model," in *Proc. IEEE 12th Int. Colloq. Signal Process. Appl. (CSPA)*, Mar. 2016, pp. 216–220.



CORINA VADUVA received the B.S. and Ph.D. degrees in electronics and telecommunications from the University Politehnica of Bucharest (UPB), Romania, in 2007 and 2011, respectively. In 2010, she performed a six-month internship at the German Aerospace Center (DLR), Oberpfaffenhofen, Germany, as a Visiting Ph.D. Student with the Remote Sensing Technology Institute (IMF). In 2012, she returned at DLR as a Visiting Scientist for two months. She has been a Research Engineer with the Research Centre for Spatial Information, UPB, since 2007. Her research fields include earth observation image processing and data mining based on information theory, information retrieval, data fusion, change detection, temporal analysis in image time series, and semantic annotation for very high-resolution image understanding and interpretation. In 2010, she received a competition Prize in the Digital Globe 8 bands Research Challenge. She was the Chair at the ESA Conference on Big Data from Space BiDS in 2016 and 2017.



FLORIN A. GEORGESCU received the B.S. degree in civil engineering (topogeodesy and automatization of topogeodesy systems) from the Military Technical Academy, Bucharest, in 2011, the M.S. degree in electrical engineering (advanced techniques for image processing) in 2013, and the Ph.D. degree in civil engineering from the Military Technical Academy and the University Politehnica of Bucharest (UPB), under a study agreement, in 2016. His areas of research include geospatial content-based information retrieval, data mining, computer vision, and pattern recognition. In 2013, he joined the Research Centre for Spatial Information, UPB, as a Research Assistant, where he conducted research on these topics within the frameworks of several national and international projects. In 2015, he held a five-month internship at the German Aerospace Center (DLR), Oberpfaffenhofen, Germany, as a Visiting Ph.D. Student with the Remote Sensing Technology Institute.



MIHAI DATCU (SM'04-F'13) received the M.S. and Ph.D. degrees in electronics and telecommunications from the University Politehnica of Bucharest (UPB), Bucharest, Romania, in 1978 and 1986, respectively. Since 1981, he has been a Professor in electronics and telecommunications with UPB. Since 1993, he has been a Scientist with the German Aerospace Center (DLR), Munich, Germany. From 1991 to 1992, he was a Visiting Professor with the Department of Mathematics, University of Oviedo, Oviedo, Spain. From 1992 to 2002, he was a Longer Invited Professor with the Swiss Federal Institute of Technology, Zürich, Switzerland. In 1994, he was a Guest Scientist with the Swiss Center for Scientific Computing, Manno, Switzerland. From 2000 to 2002, he was with Universitouis Pasteur, Strasbourg, France, and International Space University, Strasbourg. In 2003, he was a Visiting Professor with the University of Siegen, Siegen, Germany. He is currently a Senior Scientist and the Image Analysis Research Group Leader with the Remote Sensing Technology Institute (IMF), DLR, a Coordinator of the CNES-DLR-ENST Competence Centre on Information Extraction and Image Understanding for Earth Observation, and a Professor with the Paris Institute of Technology/GET Telecom Paris. His research interests include Bayesian inference, information and complexity theory, stochastic processes, model-based scene understanding and image information mining for applications in information retrieval and understanding of high-resolution SAR, and optical observations. He is a member of the European Image Information Mining Coordination Group. In 1999, he received the title *Habilitation à diriger des recherches* from Universitouis Pasteur.

• • •