Earth-Observation Image Retrieval Based on Content, Semantics, and Metadata

Daniela Espinoza-Molina, *Member, IEEE*, and Mihai Datcu, *Fellow, IEEE*

***Abstract*—Advances in the image retrieval (IR) field have contributed to the elaboration of tools for interactive exploration and extraction of the images from huge archives associating the content of the images with semantic meaning. This paper presents an Earth-observation (EO) IR system based on enriched metadata, semantic annotations, and image content called EO retrieval. EO retrieval generates an EO-data model by using automatic feature extraction, processing the EO product metadata, and defining semantics, which later is fully exploited for supporting complex queries. In order to demonstrate the functionality of the system, we have created a semantic catalog of TerraSAR-X as application scenario. The database is composed of 39 high-resolution TerraSAR-X scenes comprising about 50 000 image patches (160×160 pixels) with their feature descriptors, 100 of metadata entries for each scene, and about 330 semantic annotations. Many query examples combining semantics, metadata, and image content for full exploitation of the image database are presented.**

***Index Terms*—Content-based queries, database model, databases, Earth-observation (EO) images, image retrieval (IR), learning methods, numerical queries, raster information, semantic queries, synthetic aperture radar (SAR) images.**

I. Introduction

THE problem of image retrieval (IR) from huge image archives has been a subject of significant research in the recent past years. Basically, IR is defined by how the image archive is organized and how the search mechanism is used.

In the *framework of Earth observation (EO)*, early approaches introduced the concept of IR based on basic metadata such as geographical location, time of acquisition, and type of sensor [1]. Here, the search is based on either text or numbers, but it does not provide insight into the actual information content of the image (i.e., lakes, bridges, etc.). Later, in the early 1990s, the content-based IR (CBIR) was proposed mainly to overcome the problems of text-based IR. The query by visual example and query by subjective description as mechanisms of visual interaction for image database systems were presented in [2]. Later, Kato [3] formulated the term CBIR to describe his experiments in automatic retrieval of images from a database by color and shape features. The term has since then been widely used for describing the process of retrieving desired images from large archives on the basis of features (such as color, texture, shape, etc.) that can be automatically extracted from the images themselves. The first practical example of CBIR was query by image and video content (QBIC) [4], revolutionizing the multimedia community, where later many successful implementations have appeared [5], [6], etc. Along the years, CBIR systems have been improved by different ways, such as incorporating more primitive features, improving the search mechanism by using artificial intelligence algorithms, and adding relevance feedback (RF) methods. These progresses in multimedia inspired the EO community in retrieving raster satellite images using its content as example. Thus, among the few practical implementations of such systems, the Knowledge-driven content-based Image information Mining [7] and Geospatial Information Retrieval and Indexing [8] enable scalable processing and retrieval of a large volume of data by automatically preprocessing and indexing satellite images. However, it was quickly realized that strict visual similarity is, in most cases, weakly correlated with the measure of similarity adopted by humans for image comparison showing the so-called *semantic gap* [9]. This has motivated the use of semantic concepts for describing the image content and improving the queries in retrieval systems [10]. In *multimedia community*, the work in [11] used semantic classification methods based on the underlying assumption that semantically relevant images have similar visual characteristics or features, mainly focusing on binary classification of the images (textured and nontextured). The authors of [12] have demonstrated that the semantic representation has an intrinsic benefit for IR by introducing the concept of query by semantic example (semantic and content). Nowadays, there is a trend in tagging digital images (geotagged) with geographic information in order to improve the retrieval. As an example, the Tripod project [13] focused on using location (latitude and longitude coordinates) for improving the multimedia IR by generating both generic and specific semantic terms. The geotags and the underlying geocontext for an advanced visual search are exploited in [14]. These are clear examples of the IR tendency in combining several sources of information for improving the IR performance.

EO images are instrument measurements carrying information about physical parameters in addition to presenting the scene (Earth surface) as matrices of pixels, where each pixel has associated geographical location (latitude and longitude). However, with increase in resolution of optical and synthetic

Manuscript received September 28, 2012; revised January 11, 2013 and March 20, 2013; accepted April 15, 2013. Date of publication June 6, 2013; date of current version October 24, 2013.

D. Espinoza-Molina is with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234 Wessling, Germany (e-mail: daniela.espinozamolina@dlr.de).

M. Datcu is with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234 Wessling, Germany (e-mail: mihai.datcu@dlr.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2013.2262232

aperture radar (SAR) sensors, pixel sizes have become smaller than the ground objects, enabling automatic object extraction techniques. Moreover, EO images may be complemented with geographical information in vector format coming from geographical information systems (GISs), where the entities (objects) are well defined with the basic elements: points, lines, and polygons (areas). Every element in a vector model is described mathematically and bases on points that are defined by Cartesian coordinates. Vector data not only contain the geometry of a point, depending on the model, but also can include topology or neighborhood relations, e.g., areas next to a line or start and end points of a line, where the indexing of entities is easily done. In the case of EO images remains the problem of indexing pixels. EO images are delivered together with metadata. Usually, metadata comprise information about the data provenance (i.e., sensor type, mission information, acquisition parameters, quality process, etc.).

Concerning with the image archive organization, the traditional way is to store the images in a file-based system coded by name, dates, etc. In many IR systems such as Photobook [15], the data and the features are stored in set of files identified with their names. However, when the size of the data increases, the system performance decreases. The common problems in a file-based storage are the integrity of the data and low speed becoming a bottleneck in the retrieval process. It is clear that, when large image databases arise, the connection between IR and Database Management Systems (DBMSs) is inevitable [9]. The use of DBMS solves the integrity problems for data storage and allows dynamic updates. It also provides natural integration with features derived from other sources [9] and generates automatically indexes, which improves the query speed. In the *framework of multimedia*, one of the early contributions can be found in the CORE-System [16], whose architecture is centered around a general DBMS on top, where the database structure dominates knowledge management, feature calculation, and visualization tools because the modules for analysis, indexing, training, and retrieval are resolved in the database.

In the case of EO, compared to another database, an image database presents some special characteristics [17], e.g., the information has a spatial component, the implicit spatial information is critical for the interpretation of the image content, image characteristics could have multiple interpretations for the same visual patterns, and the values themselves may not be significant unless the context support them.

In this paper, we present an EO-retrieval system oriented toward remote sensing applications. Its architecture is founded on a database management system and on a multilayer concept connecting the EO-data sources (low level) with the end user (high level). EO retrieval attempts to discover knowledge from EO images, related geospatial data sources, and their associated metadata, mapping the extracted low-level data descriptors into semantic classes and symbolic representations and providing an interactive method for efficient semantic definition and querying the image archive. Combining these aspects, EO retrieval is able to respond to such questions as find all last year's vegetation images closer than 50 km to the Japan disaster area and find all images of Germany airports north of Munich. These queries are of huge importance for a broad community of users. The query itself needs to capture the nature and properties of the EO products, a complicated piece of information which currently remains hidden in the image archive. Hence, this paper contributes in combining enriched metadata, semantic annotations, and image content for advanced EO IR. We demonstrate the EO-retrieval functionality by creating an application scenario showing the ingestion of EO products and the semantic definition, with several examples of queries involving semantics, advanced metadata entries, and image content.

This paper is organized as follows. Section II presents a description of the EO products, the automatic extraction of the image content for image understanding, and our approach for semantic definition. Section III briefly describes the system architecture in terms of its components and functionality. Section IV presents the semantic catalog of TerraSAR-X data as study case. Finally, Section V concludes this paper and discusses further work.

II. Understanding of the High-Resolution EO Product

The volume of high-resolution EO products increases daily with the launch of new satellites that many petabytes are produced in just a few days. Moreover, the resolution of the satellite ranges from 1 to 10 m, allowing the development of new applications. For example, optical image library including images from platforms such as Quickbird, WorldView-1, and WorldView-2 may have more than 1.5 billion km^2 of data, 33% of which is less than one year old, with a daily collection rate of 1.5 million km^2 [18]. In the case of SAR products, Cosmo-Skymed [19], an Italian satellite, provides SAR images with 1 m of spatial resolution, and TerraSAR-X, the German satellite, offers products with spatial resolution ranging between 1 and 3 m, depending on the imaging mode [20]. TerraSAR-X archive currently holds about 1000 TB of satellite data and derived products and can be expanded to hold more than 2 PB with presently available storage media.

A TerraSAR-X EO product is mainly composed of the image data and its metadata, while the size of an EO image is, on average, 20 000 $\times$ 30 000 pixels with different pixel levels, i.e., 8, 16, and 32 b, and different types of data, i.e., float and unsigned integer; composed of one or multiple bands in GeoTIFF format,[1] the metadata is an Extensible Markup Language (XML) file containing related information in the forms of text and numbers. In the case of a TerraSAR-X image, apart from the image pixels, it contains a small set of internal annotation data defining the type of data, the dimension of the image, and some geographical coordinates (depending on the geometrical processing steps already done), and its XML metadata file comprises $\sim$250 entries grouped into categories such as product components, product information (i.e., pixel size, coordinates, format, etc.), processing, platform, calibration, and product quality [20]. Metadata constitute the provenance of the EO products because it summarizes the satellite parameters during

[1]GeoTIFF is an extension of the Tagged Image File Format (TIFF) standard which defines additional tags concerning map projection information.

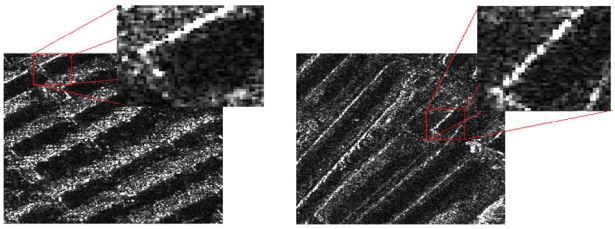

Fig. 1. In high-resolution images, the context shall be included in the analysis in order to avoid confusions between the scenes. (Left) Urban area and (right) railways are distinguished only when an analyzing window is large enough to capture the context of the scene.

the data acquisition along with the archive description, for example, product acquisition time, generation process, quality process, and calibration process. A complete description of TerraSAR-X products is presented in the Appendix as well as in [21].

A. Image Understating and Automatic Information Extraction

The availability of high-resolution optical and SAR data has assisted in recognizing more details within a scene like building extraction, road detections, growth of the cities, etc. For instance, using a TerraSAR-X image, buildings, roads, and vehicles are clearly distinguishable. The difference between vegetation and residential areas can also be observed. Ongoing research in automatic information extraction deals with the increased resolution of the image satellites, since the capacity of object recognition becomes difficult due to the very high resolution. In the case of feature extraction, the common approach is a pixel-based analysis, which extracts the primitive features using the pixels in the images. However, the object detection requires a new approach which considers the context of the image. An example is shown in Fig. 1; in here, at meter resolution, if context is not taken into count, very different scene classes may be confusing. The example shows urban area (left) and railways (right) which are comprehensible only when the analyzing window is sufficiently large to incorporate relevant context. Otherwise, both scene classes seem to be similar and can be confused. Therefore, we use a new approach for the analysis of the image content based on tiling the image into several patches. We establish the patch as a part of the image, which may contain one or several objects, depending on its size.

The advantages of patch-based approach have been demonstrated in several works, for instance, in [8], the selected patch size was 256 $\times$ 256 pixels in order to ensure that the extracted features capture the local characteristics within a patch rather the global features across the entire image. In [22], the TerraSAR-X high-resolution spotlight images were tiled into $\sim$7000 patches of 200 m $\times$ 200 m in order to characterize the large, as well as relatively small, structures available in an urban scene. The proposed method distinguished about 30 different object categories. In [23], the original image was tiled into patches of 16 $\times$ 16 pixels or 128 $\times$ 128 pixels. The result of the classification (city, forest, and sea) was better in the second case.

B. Semantic Definition of High-Resolution EO Images

The semantic definition or annotation of an image allows the user to describe its content with natural language. However, this is one of the most arduous and expensive tasks in IR that can be optimized following different methods such as semisupervised and autoannotation methods, machine learning methods, and neural networks.

The definition of semantic deals with some important aspects, such as the following: 1) The annotation can be allocated to either an entire image or regions considering that an image is composed of various regions with different meanings; 2) the annotation task can follow a machine-driven approach based on machine learning methods or human-driven approach, where the labels are assigned manually after visual inspection [24]; and 3) there is not a common and standard taxonomy or ontology defined for high-resolution EO images. An attempt was made in CORINE Land Cover (CLC), which introduced a land-cover/land-use classification scheme. Here, the mapping of the land cover and land use was performed on the basis of satellite remote sensing images on a scale of 1 : 100 000. The standard CLC nomenclature includes 44 land-cover classes. These are grouped in a three-level hierarchy [25]. However, in the case of high-resolution EO images, we found that EO images (both SAR and optical) contain many categories of objects, which are sometimes unexpected or difficult to be named, and their actual number is unknown. In order to solve these aspects, we established the following methodology: 1) to define an annotation scheme which supports about three levels of taxonomy and a hierarchical indexing structure; 2) to work with image regions (patches); and 3) to annotate the image patches using the semiautomatic machine learning methods and human expertise. The proposed annotation scheme is based on an adaptation of [26] and presented in Fig. 2. The annotation is context-dependent and follows the proposed scheme. Comparing the proposed scheme with CLC, we observe that our scheme includes more categories for urban areas as well as it is oriented to objects such as ports and building blocks as residential area. Some examples of semantic categories are presented in Fig. 3. Fig. 3(a) displays urban structures, and Fig. 3(b) shows scenes in natural areas such as vegetation areas and water bodies.

III. System Architecture

EO-retrieval system is best described by its architecture shown in Fig. 4. The architecture design and the implementation include characteristics such as the following: 1) The system is based on multilayer architecture, which is founded on a DBMS as central core; 2) the system uses a patch-based primitive feature extraction approach in order to analyze objects and to include the context information; 3) the system characterizes the image content in terms of the combination of several descriptors (basic primitive features, metadata, and semantic); and 4) the system implements semisupervised learning methods for semantic definition of the patch content. The intention of the EO-retrieval system is to implement a communication channel between the EO-data sources and the end user who receives the content of the data sources coded in an understandable format associated with semantics and ready for the exploitation.

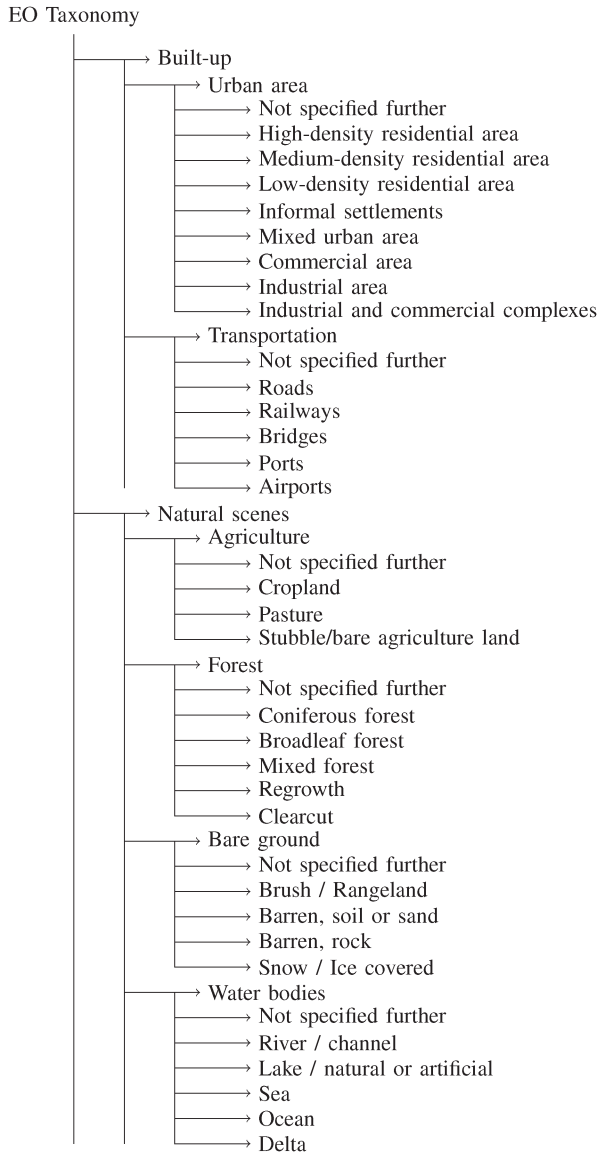

Fig. 2. Land-use/land-cover hierarchical scheme classification.

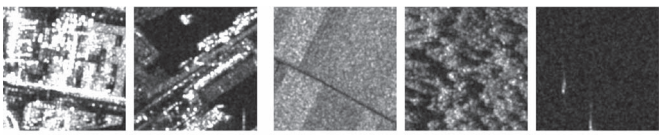

Fig. 3. Examples of built-up and natural scenes in TerraSAR-X. A patch example containing (left to right) industrial area, a port, some croplands, coniferous forest, and sea. (a) Built-up areas. (b) Natural scenes.

The three layers of the system are as follows. The *Presentation Layer* is in charge of showing the graphical user interfaces and supporting the user interactions. The *Logic Layer* coordinates the application logic, makes logical decisions, performs calculations, and supports the user operations; most of the system tasks are performed in this layer. The *Data Layer* stores and retrieves information from a database. It coordinates the storage as well as the provisions of the data. Observe that the system is oriented to three different users: 1) system operator who manipulates the EO sources; 2) EO expert/analyst who performs the semantic definition (annotation of the image content); and 3) end user who uses the main functionalities of the system. In the following, we start describing the architecture following the execution order (processing chain) for a better understanding.

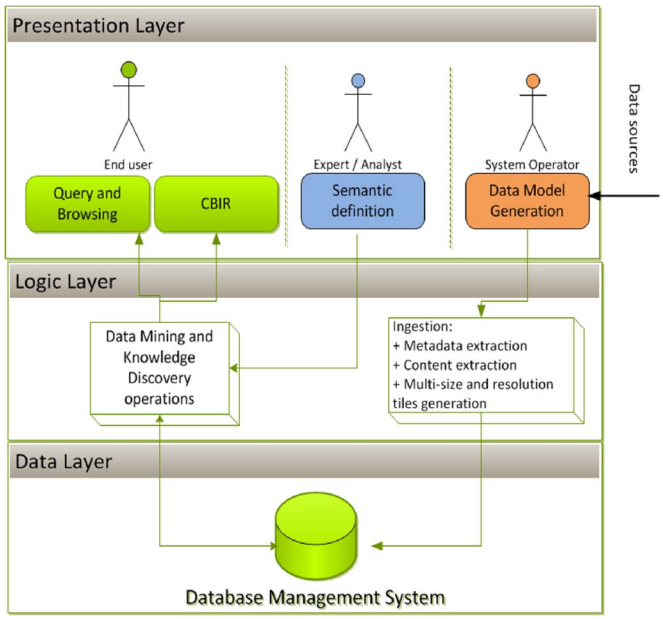

Fig. 4. Unified Modeling Language (UML) overview of a multitier architecture of the EO-retrieval system. The Presentation Layer provides the user interfaces. In the Logic Layer, the logical decisions, calculations are processed. The Data Layer stores and retrieves the information from the database.

A. Data Sources

The data sources are composed of the following: 1) EO images (either optical or SAR); 2) their associated metadata (i.e., acquisition time, incidence angles, etc., normally stored as XML file); and 3) some ancillary information in vector format coming from GIS sources that complements the information about the image, for instance, park boundaries or land uses represented as polygons, etc.

B. Data Model Generation

The data model generation (DMG) aims at creating a model of the EO image which will contain primitive features, associated metadata, and ancillary information. DMG focuses on the ingestion of new products and performs the metadata analysis, multiresolution patch generation, quick-look creation, and content analysis.

First, the metadata analysis processes the XML file, looking for the required entries, e.g., acquisition time, geographical coordinates, sensor specificities, number of columns and rows of the image, etc. Second, the EO image is divided into several patches with different sizes (i.e., 256×256, 128×128, 64×64 pixels, etc.), generating a multiresolution and multisize patch pyramid together with quick looks. Third, during the content analysis, the patch-based feature extraction methods are applied on these individual patches, converting their content into primitive feature vectors. Finally, all the generated information is stored in the database.

The feature extraction methods are implemented in order to deal with the specificities of EO images and to represent the relevant and correct information contained in the images known as descriptors or basic features. These features are the foundations of the CBIR systems and semantic definition modules. The most common approach for feature extraction considers the basic characteristic of an image as its relevant feature, for example, color, texture, and shape. However, [27] presents a pattern representation scheme based on compression techniques, deriving a compression ratio vector, which is considered as a descriptor of the data and is used for categorization and classification purposes. EO-retrieval system incorporates texture features as well as compression-based features.

1) Texture Features: EO-retrieval system uses Gabor filters (GFs) as feature extraction method. GFs can be considered as orientation and scale tunable edge and linear feature detectors, and the statistics of these microfeatures in a given region are often used to characterize the underlying texture information. The foundations of GF for texture description were proposed in [28], where the homogeneous texture descriptor was proposed, which was later used as one of the visual texture descriptors in the standard MPEG-7 [29]. We follow [28] and use the *mean* and *variance* of the magnitude of the Gabor coefficients to represent the image texture. Therefore, the feature vector is now constructed using mean and variance, as feature components. The dimension of the feature vector is equal to $2 \times \text{scale} \times \text{orientation}$, for instance, using four scales and six orientations results in a feature vector with dimension 48.

2) Compression-Based Features: The main idea is to extract directly dictionaries (sequence of repetitive patterns) from the image data by applying compression algorithms. A dictionary is extracted using the Lempel–Ziv–Welch (LZW) compression algorithm [30], which is a lossless dictionary-based compression method. This kind of compression algorithms scans a file for patterns (sequences) of data that occur more than once. These sequences are stored in a type of codification called dictionary, which later represents the image descriptor.

It is important to mention that, during the entire DMG, the data managed by the system are transformed, passing from an initial form of full EO products to much smaller value added products (quick looks, image descriptors, and metadata entries), named as *EO-data model*, as shown Fig. 5, which will be later available for immediate exploitation and analysis by the end user and EO expert. The main functionality of this module is performed in the *Logic Layer* (see Fig. 4).

C. Database Management System

The use of a DBMS enforces the referential integrity and consistency of the data (update, insert, and delete operations) as well as allows us to query the data using Standard Query Language (SQL) statements. For example, a typical SQL statement for searching products with resolution lower than three is

SELECT metadata-id, *resolution*, *startimeutc*, *stoptimeutc*
FROM metadata
WHERE groundrangeresolution < 3.

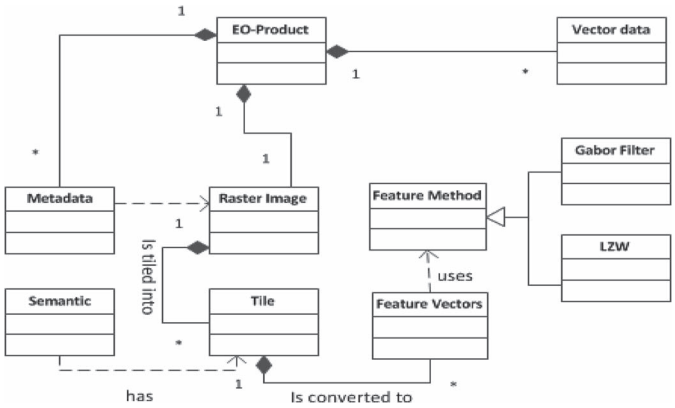

Fig. 5. UML representation of EO-data model. An EO product is composed of metadata, raster image, and vector data. The raster image is divided into patches, which are converted to feature vectors by applying feature extraction methods, and later will have associated semantic.

Our EO-retrieval system uses MonetDB as DBMS. MonetDB follows a column-store approach for high-performance data warehouses for business intelligence and eScience [31].

The generated *EO-data model*, depicted in Fig. 5, is stored into a relational database scheme, which acts as the core of the system. The database stores actionable information interacting with all the system components as well as it implements predefined functions based on SQL for supporting and augmenting system functionality, for example, computation of geodesic distance between two points given the location, returning the minimum bounded rectangle (MBR) of the given coordinates, and measuring similarity between primitive features.

D. Data Mining and Knowledge Discovery

Most of the modules in the presentation layer are supported by the operations performed in this component. For example, in the case of CBIR, the starting point is the *Query by Example*, where the image content is passed as query parameter activating later the data mining algorithms. Data mining is a repetitive process of sorting through a large amount of data and picking out the relevant information. It involves methods for dimensionality reduction, clustering and classifications, pattern recognition, active learning, and RF.

During the *semantic definition*, the expert participates actively in the active learning and RF, verifying/evaluating the retrieved results and stopping the learning process. Hence, the data are converted into information and later into knowledge (knowledge discovery). In EO-retrieval system, some data mining methods (clustering and similarity metrics) are implemented as SQL functions into the database, where the synergy between both modules causes a considerable improvement in the performance of the system.

E. Semantic Definition

The annotation module of EO-retrieval system is based on machine learning algorithms, the support vector machine (SVM) [32] together with RF. The SVM treats the problem of annotation as binary classification of the entire image database. The goal of SVM is to produce a model based on the training

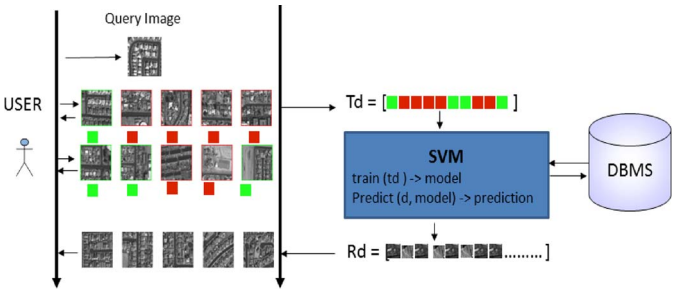

Fig. 6. Semantic definition based on SVM. The user gives positive and negative examples, which are passed as training data to the SVM. The SVM makes a predication based on these training data. The results are ranked and displayed on the screen.

data, which predicts the target values of the test data given only the attributes (features) of the training data. The RF asks the user whether certain results are relevant or not.

An example is illustrated in Fig. 6. The semantic definition is an interactive and iterative process. It starts randomly presenting the image patches on the screen. Then, the user judges the patch content according to his interest (i.e., houses). The user selects positive and negative examples, as shown Fig. 6 in green and red, respectively. The set of positive and negative examples is the *training data*, which is passed to the SVM (see Fig. 6). The SVM produces a *model* based on these training data. Later, using the generated model, the SVM makes a prediction of the rest of the data. The results are ranked according to the prediction and displayed on the screen (see the bottom part of Fig. 6), thus starting the second iteration. This process is repeated until the user is satisfied with the results or the tool remains in stable results (RF operation). When the training is finished, the new semantic label is stored into the database. It is worth to note that the training and prediction of SVM are part of the data mining and knowledge discovery component performed into the Logic Layer.

F. Query Module

The query module is a fundamental part of the EO-retrieval system. It is an interactive component that attempts to better exploit the generated *EO-data model* by providing several types of queries. This module relies on the following two types of queries.

1) Query by Attributes: Attribute queries ask for information from the database tables associated with entities. Attributes can be numeric values, text strings, or dates. Here, we mainly distinguish attributes coming from the metadata and predefined semantic labels.

1) *Query by metadata*: This query exploits the full metadata extracted from the XML file of the EO products. This kind of query allows one to look for specific products or patches using the metadata information, e.g., find TerraSAR-X products between a range of incidence angles or products belonging a given mission. Moreover, metadata can be used for creating more complex scenarios [33].
2) *Query by semantics*: The user can enter a simple semantic label in the form of a text or select one from the available labels in the semantic catalog to perform the query. The predefined semantic labels (semantic catalog) were previously obtained as results of the image annotation using the semantic definition module.

2) Query by Spatial Content: The answers to spatial queries are directly derived from the location of features on an image. Information about the proximity of one entity to another entity is not contained in an attribute table but is easily learned using a spatial query. For example, the end user could ask if one or more patches are located within a certain distance of other features or are contained by another feature. The literature [34] introduced five types of spatial queries: 1) point in polygon, which essentially asks the question: What do we have at this point (latitude/longitude)?; 2) region queries asking the question: What do we have in this region? (a region is defined as a rectangular box specified by the upper-left and lower-right geographic coordinates); 3) distance and buffer zone queries, questioning what do we have within some fixed distance of this point (latitude/longitude) (the buffer is created based on the fixed distance); 4) multimedia queries, which combine multiple georeferenced information in answering a query; and 5) path queries, which require network structures (sets of interconnected line segments); this query may ask what is the shortest route to a point. EO-retrieval system implements the first four types of queries. Most of the queries are solved by computing the geodesic distance and MBR.

G. CBIR

A CBIR module implements the concept of *query by example*. This type of query passes the image content as query parameter activating later the data mining algorithms for solving the query. The CBIR module of EO-retrieval system is based on the compression methods previously presented in Section III-B for feature extraction. The dictionaries (features) are extracted during the ingestion (DMG). In a following step, the similarities between the example patch and all the patches in the database are computed. The fast compression distance (FCD) [35] was adopted as similarity metric and is given by

$$D(x,y) = \frac{|D(x)| - \bigcap(D(x), D(y))}{|D(y)|} \tag{1}$$

where x and y are two patches (objects), $D(x)$ and $D(y)$ are the sizes of dictionaries of x and y, respectively, and the term $\bigcap(D(x), D(y)$ is the intersection between two dictionaries of x and y. The computation of similarities between two objects is performed on the basis of the size of the intersection set between the relative dictionaries [35].

The CBIR operation is simple; the end user selects a patch with the content of interest and asks the system to retrieve patches with similar content. The system computes the FCD between the query patch and the rest of the patches in the database and ranks the results according to the distance. Finally, the top 50 patches are displayed to the user. (See Fig. 17 as example.) An improvement of this version comparing with other traditional CBIRs is that the computation of the similarity metric inside the database improves the query speed.

IV. Application Scenario: The Semantic Catalog of TerraSAR-X Data

As application scenario, we chose to create a semantic catalog for TerraSAR-X products based on the extraction of the metadata and image content together with semantic definition. This scenario reflects all the system functionalities starting with the DMG (data ingestion), later the semiautomatic semantic definition, and finally performing different types of queries, for example, query based on location and query based on image content.

The semantic catalog was created using 39 high-resolution Multilook Ground range Detected (MGD) radiometrically enhanced (RE) TerraSAR-X products around the world. These products were grouped into 16 collections for analysis purposes.

A. Data Model Generation

In the DMG module, the operator configured the ingestion of the 39 products, setting input parameters such as the size of the patch, the feature extraction methods to be used, and the physical location of the EO products. Later, the EO products were ingested (metadata and primitive feature extraction, multiresolution patch generation, and quick-look creation), and their data model was stored into the database.

During the ingestion, each TerraSAR-X image was tiled in a patch size of 160×160 pixels, and in each patch, the Gabor method was applied for the texture feature extraction, with input parameter scale $= 4$ and orientation $= 6$, giving a feature vector of 48 descriptors, and LZW using the whole patch (see Section III-B) was applied for dictionary extraction. As result, the current database comprises about 50 000 patches with their feature vectors, dictionaries, and metadata.

B. Semantic Definition

The image annotation was done per collection, using the SVM with RF tool together with the ability of the EO expert. The semantic labels were assigned to the different patches, resulting in an enriched semantic catalog (about 330 labels) of the image database. The SVM used Gabor feature vectors with a length of 48 and seven to ten iterations on the average in the RF performed by the expert. For each collection, about 20% of the images patches were used as training data. One patch was assigned only to one semantic class based on the dominant content of the patch and context. After the semantic definition, the EO-data model (quick looks, image descriptors, enriched metadata, and patch information) was completed with the semantic labels.

C. Query by Attributes

Examples of queries that can be performed using metadata and semantics for exploring and exploiting the semantic catalog of TerraSAR-X products are presented as follows.

1) Query by Metadata: TerraSAR-X metadata file includes several information that can be used for various analyses, for example, image time series by querying the available images of a zone within a range of time and study of the incidence angle for determining available structures [33].

1) *Overview of the catalog*: We started with an overview of the created catalog in order to have an idea about the current products, the area of coverage, the different places, and the total of semantic labels found. Table I summarizes the organization of the semantic catalog. Observe that there are 16 collections, where some of them contain more than one scene. The resolution is a parameter contained in the metadata, and the area can be computed using extension of the image and resolution.

TABLE I
Collections and Products of the TerraSAR-X Semantic Catalog. The Catalog Comprises 39 High-Resolution MGD TerraSAR-X Scenes Taken Over Different Places and 330 Semantic Labels Covering ~1236.34 km^2. The Collections Are in Alphabetical Order

Collection	Place	Area(km^2)	Resolution	Labels
Algeria	Hai EL Moud-jahiddine	30.396	2.89	16
China	Shenyang	30.965	2.90	6
Colombia	Bogota	29.535	2.90	6
Cuba	Habana	38.050	3.36	8
Ex Soviet Union	Tajikistan	30.049	2.89	
	Uzbekistan	30.298	2.90	
	Total	60.347		20
Germany and Switzerland	Berlin	31.992	2.90	
	Cologne	30.859	2.88	
	Hamburg	30.328	2.88	
	Basel	30.895	2.90	
	Bonn	31.338	2.90	
	Munich 1	31.042	2.89	
	Munich 2	31.377	2.90	
	Total	217.82		42
Greece and Turkey	Larissa	30.172	2.88	
	Koyulhisar	31.590	2.90	
	Total	61.762		24
High buildings	Dubai	30.889	2.90	
	Las Vegas	32.206	2.91	
	Ansahn	32.665	2.92	
	Tokyo	32.680	2.92	
	Washington	32.965	2.92	
	Total	161.404		34
India	Belgaum	32.055	2.90	
	Pune	32.563	2.90	
	Sarsiya Talav	30.813	2.90	
	Total	95.431		25
Peru	Nazca	34.095	2.91	5
Poland	Czestochowa	31.654	2.90	
	Lodz	32.188	2.91	
	Torun	32.725	2.92	
	Total	96.567		30
Russia	Moscow 1	31.518	2.90	
	Moscow 2	31.101	2.89	
	Moscow 3	31.133	2.89	
	Perm	32.122	2.91	
	Rostov on Don	32.271	2.92	
	Tula	31.335	2.90	
	Rusia	30.657	2.90	
	Total	220.136		39
Sweden	Vasteras	31,675	2.89	18
Thailand	Bangkok	32.707	2.92	10
France and Romania	Toulouse	30.817	2.88	
	Timisoara	33.737	2.92	
	Total	64.553		31
Italy	Venice	30.888	2.89	17
	TOTAL	1236.34		330

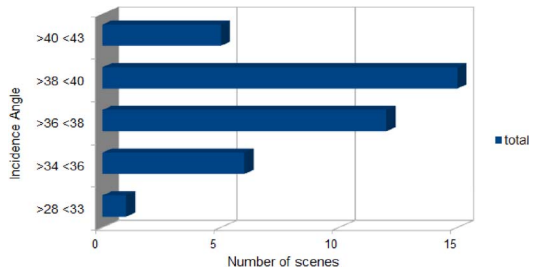

Fig. 7. Frequency distribution of the scenes in the catalog according to the incidence angle.

Table I shows that the coverage area of the catalog is relatively small $\sim$1236.34 km^2; however, the level of details that can be observed is huge (about 330 different semantic labels), for instance, the Venice scene has an area of 30 km^2, and about 17 semantic labels were identified.

2) *Statistical analysis*: Later, we were interested in a statistical analysis of the catalog, so we selected to study the frequency distribution of the 39 scenes in terms of the incidence angle. In the next example, we queried the catalog for finding the *frequency distribution* of the scenes according to the incidence angle. First, we found the minimum and maximum incidence angles and later counted the available products in a given range using SQL statements. Fig. 7 illustrates the histogram result, where most of the scenes were acquired between 38° and 40°.

   As an extension of this query, we can retrieve all the patches within the selected incidence angles.

3) *Acquisition ranges*: A common request is to find taken or available products in a specific time. This query can be combined with other attributes, for example, resolution.

More queries can be performed using several attributes of metadata like the mission, type of polarization, etc.

2) Query by Semantics: The importance of the performed annotation is reflected in the questions that can be answered by the queries, for example, how many classes or categories can be identified by a product or a patch, how is the content of the whole catalog divided, and what is the predominant semantic class contained in a catalog?

In the following, we presented some interesting queries for inspecting the semantic content of the image database.

1) *Overview of semantic catalog*: The semantic catalog is composed of about 330 different semantic labels that were assigned to the patches in the database. Fig. 8 depicts a part of the semantic catalog. We selected only 25 patches with their semantic labels in order to show the diversity of TerraSAR-X image content.

   The patches in the first row are labeled as urban area. In this case, the semantic classes were created according to the shape, level of brightness, and finding common characteristics between them. The second row shows examples of transportation, typical roads, railways, a bridge, and a port. Observe that the last two patches are difficult to identify unless the context supports the annotation. The third row depicts different types of agriculture, such as cropland and pasture; here, observe that they represent flat areas as main characteristic. The fourth row presents patches containing forest; in this class, the texture can be noted as main characteristic. In the last row, several water bodies are depicted. This catalog is a classical example of how the semantic helps in understanding the image content.

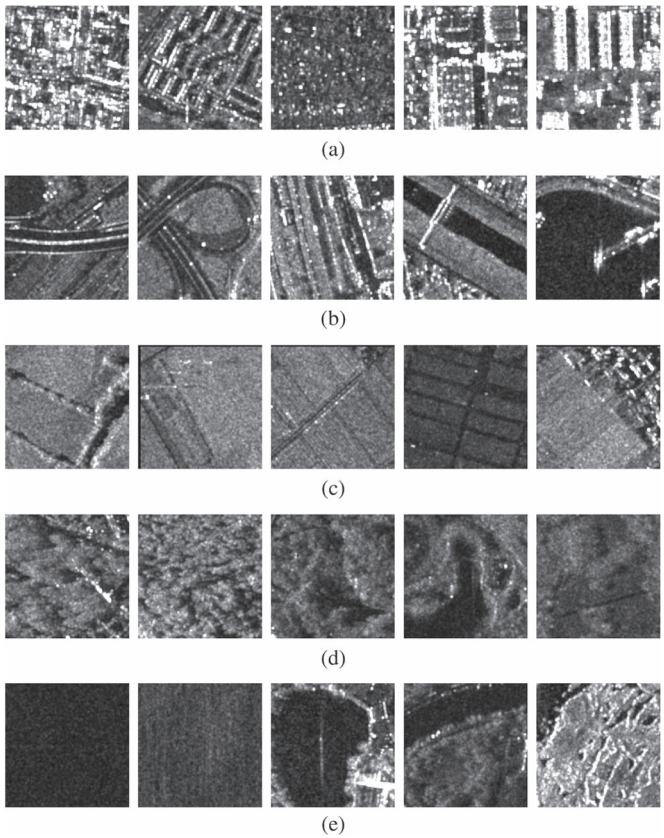

Fig. 8. Semantic catalog of TerraSAR-X images. The semantic classes are divided into two main groups: (First and second rows) Built-up area and (third to fifth rows) natural scenes. (From left to right) (a) Urban area showing examples of high-, medium-, and low-density urban areas as well as commercial and industrial areas. (b) Different kinds of transportation: Normal and curved roads, railways, bridge, and port. (c) Agriculture comprising examples of cropland, pasture, and mixed areas. (d) Different types of forest: Coniferous, broadleaf, mixed forest, etc. (e) Water bodies include ocean, sea, rivers, lakes, and deltas.

2) *Statistical analysis*: We analyzed the frequency distribution of the semantic catalog for having an example of land-cover application. We queried the whole archive counting the patches annotated with labels. Fig. 9 gives a complete view about the coverage of the semantic classes. For the sake of simplicity, only the labels with more than 1% were included in the figure. It can be noted that the dominant class is *water* with 16%, followed by *low-density urban area* with 15%, and *high-density urban area* with 13%.

   In the case of a deep study of the image content, we can select only one product for the query. For instance, Fig. 10 shows the frequency distribution of the semantic categories corresponding to TerraSAR-X scene over a city in Algeria. Here, we observe that 16 different semantic classes were found where *agriculture* contributes with only 3% of the total image coverage while around 35% of the image contains water, specifically *sea*.

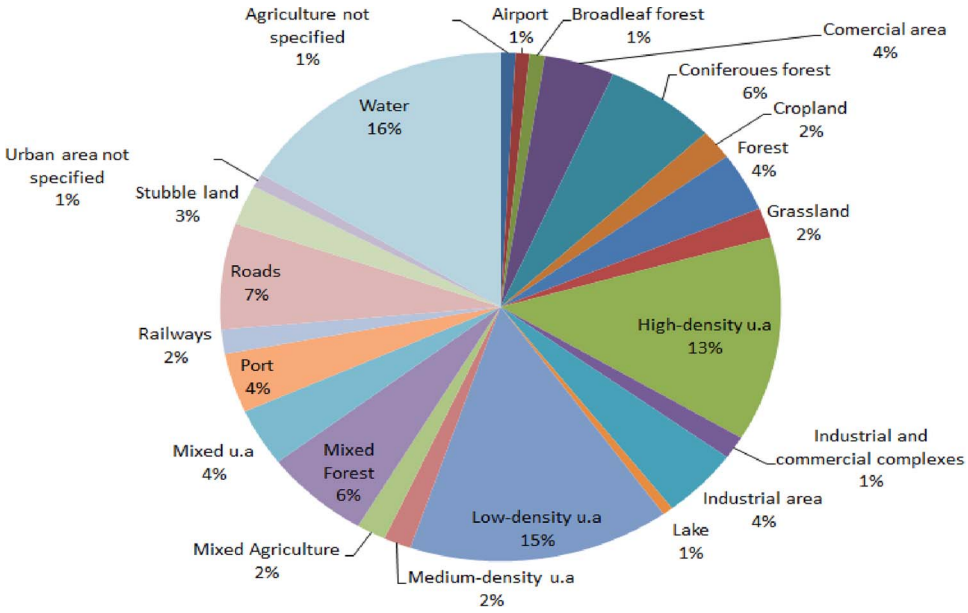

Fig. 9. Scenes categories and information content of the whole image archive. Water is the dominant class in the catalog, followed by urban area.

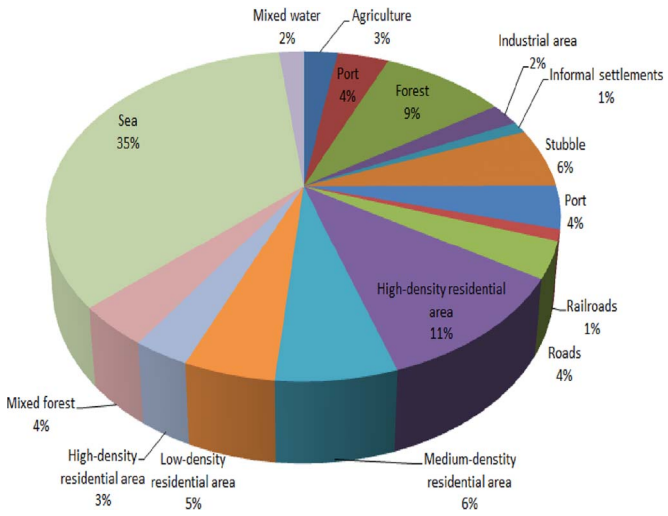

Fig. 10. Semantic content of the TerraSAR-X over a city in Algeria. Most of the image is covered by sea, followed by urban areas.

3) *Specific content*: This query is useful for finding particular information in the whole catalog. As an application example, we searched for all the scenes containing *bridges*. The result is summarized in Fig. 11, where the patches annotated as bridges from different collections are presented.

Here, it can be seen that the first two rows contain the Venice patches, while the third row shows the patches of Washington scene.

It is important to mention that, the same as metadata, more queries can be performed using semantic annotation, for example, *different levels of annotation*, where we can combine two semantic labels at different hierarchy levels, e.g., search for cities which have 20% or 30% of ports.

3) Query by Combining Semantics and Metadata: The most interesting part of the queries is that they are able to combine different sources of information.

We extended the statistical analysis of the incidence angle by plotting the frequency distribution of the combination of the semantic labels and number of scenes in intervals of angles. Fig. 12 shows the relation between the incidence angles and the semantic labels. We observe that most of the labels were found between 38° and 40°.

D. Query by Spatial Content

The queries based on location answer questions, for example, the proximity between two objects (*distances*), the objects within a given boundary (containment), or objects in a given direction. In the following, we present examples of distances and buffer zone queries answering the following question: What do we have within a fixed distance?

1) Proximity: The proximity of two objects is given by a distance. Computing distances between two geographic points represented by their latitude and longitude coordinates is a common task that can be easily used for different applications. In the next example, we calculated the geodesic distance [36] between Cologne with (50.94372, 6.9649677) coordinates and Dubai with (25.193075, 55.285995) coordinates; the geodesic distance approximates 4980.58 km.

2) Containment: A common query is to find all available products within a specific distance. This query is based on minimum bounding box algorithm and implemented using SQL statements.

3) Semantics and Location: Used in combination, both previous queries are a powerful way of exploring spatial patterns in the TerraSAR-X catalog. In the next example, we are interested in finding all the patches in Venice that contain forest within 2 km of St. Mark's Square. We assumed that St. Mark's

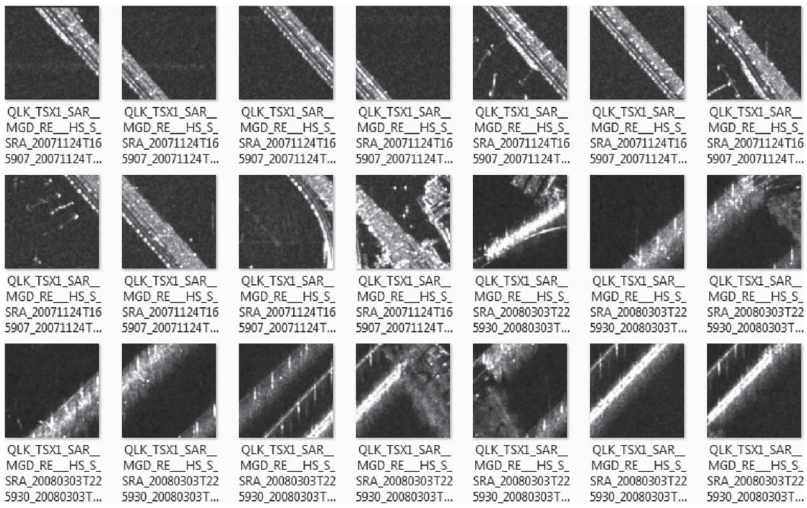

Fig. 11. Quick looks of patches containing *bridges* retrieved from different collections. (Top to bottom) First 11 patches were found in the Venice scene, and the last ten ones were found in the Washington scene.

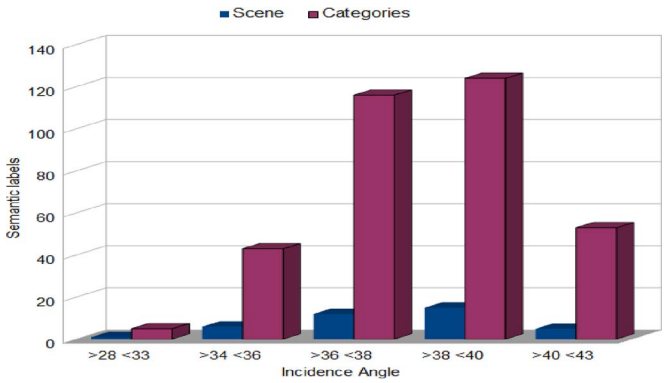

Fig. 12. Frequency distribution of the number of scenes and semantic categories versus incidence angle.

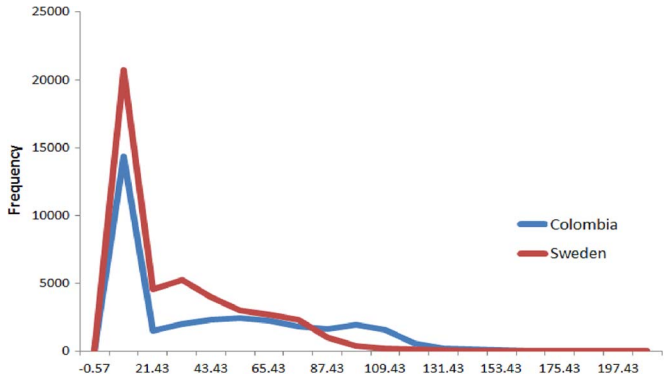

Fig. 14. Histogram of the feature vectors in the case of (in red) Sweden collection and (in blue) Colombia collection.

Fig. 13. Patches found within 2 km of St. Mark's Square that contain forest (image taken from Google Earth).

Square has (45.433919, 12.340568) coordinates. First, we selected all the patch annotated as forest within the Venice collection. Then, we computed the distance using the minimum bounding box algorithm. The query result was exported to a Keyhole Markup Language file and displayed in Google Earth, as shown Fig. 13.

E. Query Based on Basic Primitive Features

Generally, in most of the cases, the primitive features are used for the query based on example. However, we can also use these features for more tasks, such as the following: 1) statistical analysis of the feature space in terms of descriptive statistics and histograms; 2) analysis of the semantic based on the texture features; 3) classification of the image content or the archive content; and 4) CBIR. In the following, we present examples of the first three cases, which are based on texture features provided by the Gabor method (see Section III-B), while the last case is based on compression features and presented in the next section.

1) Statistical Analysis of the Feature Space: The histogram and descriptive statistics of two chosen collections, one containing high diversity (Sweden with 16 semantic labels) and the second low diversity (Colombia with 6 semantic labels) are presented. The diversity is given by the number of semantic labels. We retrieved the Gabor feature vectors of the patches corresponding to both collections and computed their histograms as is shown in Fig. 14. Here, in both cases, the major distribution is between -0.57 and 30.43.

TABLE II
SEMANTIC CATEGORIES AND MAIN STATISTICS OF SWEDEN AND COLOMBIA COLLECTIONS

Semantic	Mean	Variance	Min	Max
Sweden				
Airport	16.716	356.525	-0.299	128.655
Coniferous	22.772	677.503	-0.448	174.925
Forest not specified	22.938	645.809	-0.455	151.492
Mixed Forest	25.359	797.621	-0.549	196.455
Industrial area	23.810	726.273	-0.542	184.396
Commercial area	27.824	863.978	-0.359	135.374
Cropland	22.770	634.890	-0.457	180.104
Bare land	22.721	633.545	-0.459	155.291
Port	24.972	886.726	-0.345	212.287
Railway	25.161	738.191	-0.426	147.974
Roads	26.0814	802.500	-0.434	160.386
High-density urban area	28.396	902.474	-0.502	171.197
Medium-density urban area	24.483	742.574	-0.410	175.737
Low-density urban area	24.417	720.127	-0.452	171.864
Mixed urban area	25.727	744.242	-0.566	129.198
Industrial and commercial complexes	27.228	877.204	-0.449	164.261
Urban area not specified	20.243	492.578	-0.365	99.999
Water	19.941	606.378	-0.388	242.553
Colombia				
Commercial area	37.738	1544.487	-0.601	183.761
Cropland	35.0517	1303.826	-0.511	158.690
Forest	36.045	1418.460	-0.829	172.247
Roads	36.447	1414.175	-0.874	167.400
High-density urban area	37.012	1493.074	-0.781	165.229
Low-density urban area	35.458	1376.759	-1.216	235.225

2) Analysis of the Semantic Based on the Texture Features: We can combine the descriptive statistics of the texture features with the associated semantic class in order to analyze their relation. In the following, we study two cases: 1) descriptive statistics of the semantic classes based on the primitive features and 2) the behavior of one semantic class in the whole archive based on the texture features. However, the relation of all semantic classes in one collection together with the texture features can easily be done. *In the first case*, we computed some basic statistics, which are presented in Table II. Here, it can be observed that, in the Sweden collection, the semantic classes with low mean values are *coniferous forest* and *water*, showing low level of variation in their content, while high mean values stand for *high-density urban areas*, which show the high diversity of this class. Comparing *high-* and *low-density urban areas* between the Colombia and Sweden collections, we can conclude that the urban area in Colombia seems to be completely different to the urban area in Sweden since the descriptive statistics differ significatively.

In the second case, the query is oriented to analyze the behavior of one semantic class in the whole archive. For instance, we studied the behavior of *forest* as semantic class in two collections (Algeria, and France and Romania), trying to answer the following question: Is the forest class the same in all collections?

We plotted the feature vector of the patches annotated as forest in both collections (130 patches in the case of Algeria and 74 in the case of France and Romania), as shown Fig. 15. Here, it can be observed that, in the case of Algeria, the feature vectors have high values (in bright green), while in the case of France and Romania, the feature values are low (in dark green). Observing the figure, it can be deduced that the forest in Algeria has more variations (high variance values) while the forest in Romania seems to be more constant (low variance values).

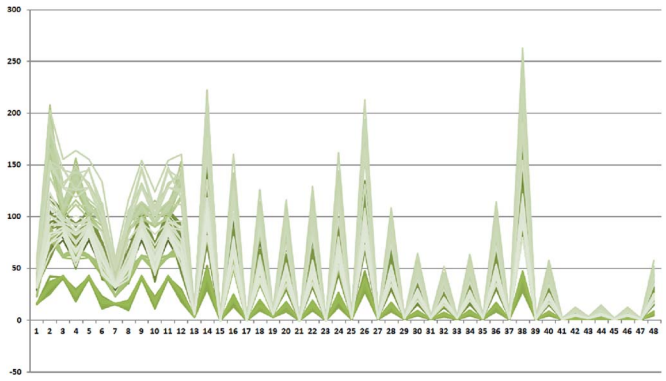

Fig. 15. Analysis of *forest* semantic class in two collections. Values in dark green correspond to France and Romania, and values in light green stand for the Algeria collection.

3) Classifications: Generally, the classification of the feature space is made by *clustering*, which is a common approach for grouping similar feature values together into clusters or generic classes (without semantic meaning). In the following example, a kind of clustering based on the Euclidean distance using the Gabor feature vectors is presented. The assumption is that, given a patch labeled with a semantic class, if we compute the Euclidean distance between its feature vector and all the feature vectors in the database using a threshold, short threshold values should retrieve similar patches (same semantic) while larger threshold values should retrieve different semantic classes. First, we randomly selected a patch annotated as *high-density urban area* belonging to Colombia scene and retrieved its feature vector. Second, we queried for texture parameters in intervals of ten and computed the Euclidean distance between the features.

The radar chart in Fig. 16 illustrates the results. Here, the distance from the center of the chart to the edge is presented. The center represents the shortest value (10), and the chart edge is the maximum value (100). The chart is also divided into six sections (series of data) according to the number of semantic classes, and the areas filled with different colors represent the retrieved semantic classes in a given threshold. Therefore, with threshold $= 10$, only the *high-density urban area* class is retrieved (in bright blue); with threshold $= 20$, *high-density urban area* and *medium-density urban area* classes are retrieved (in orange); and with threshold $= 100$, the six semantic classes are retrieved.

F. CBIR

This kind of query allows the user to find images with similar content as the one passed as example. The measurement of similarity is computed using a distance metric between the feature vectors. The CBIR component of EO retrieval is based on LZW compression method and the FCD distance. LZW is applied on the whole patch (160×160 pixels).

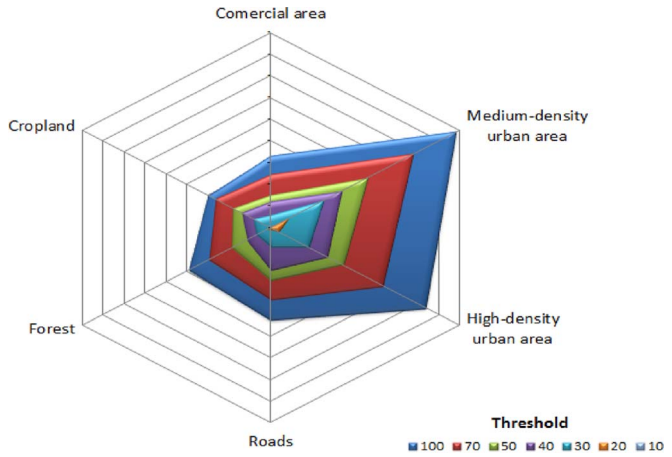

Fig. 16. Clustering of Colombia collection. The number of semantic classes increases according to the threshold. Short threshold (10–20) retrieves similar classes, which are in orange in the center of the chart, while larger threshold (100) retrieves all the classes, in dark blue.

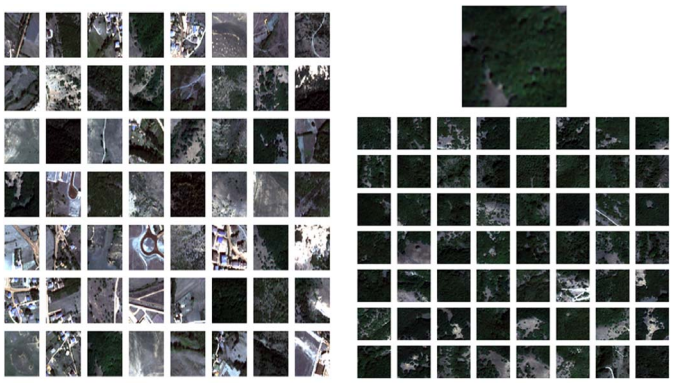

Fig. 17. Query by example based on compression methods. (a) Example of the image catalog and the different image content. (b) Query results of using forest as content example.

Fig. 17 presents examples using optical images in order to visually understand the concept of the system. Fig. 17(a) shows the system user interface where it is possible to select the content to be used as example by clicking on it. In Fig. 17(b), a patch containing forest is passed as example. It can be seen that the result is very good since all the retrieved patches contain forest.

We carried out some experiments using the semantic catalog of TerraSAR-X scenes. We searched for different patch contents and computed precision and recall as quantitative evaluation metrics. Recall measures how well the search engine is doing at finding all the relevant images for a query, and precision measures how well it is doing at rejecting nonrelevant images [37].

In the following experiments, using the Berlin scene with 1026 patches and 18 semantic labels, we randomly selected a patch containing a semantic label (forest, river medium-density urban area, and low-density urban area), and based on the first 20 retrieved patches, precision and recall were computed. The results are summarized in Fig. 18(a)–(d). The results of the

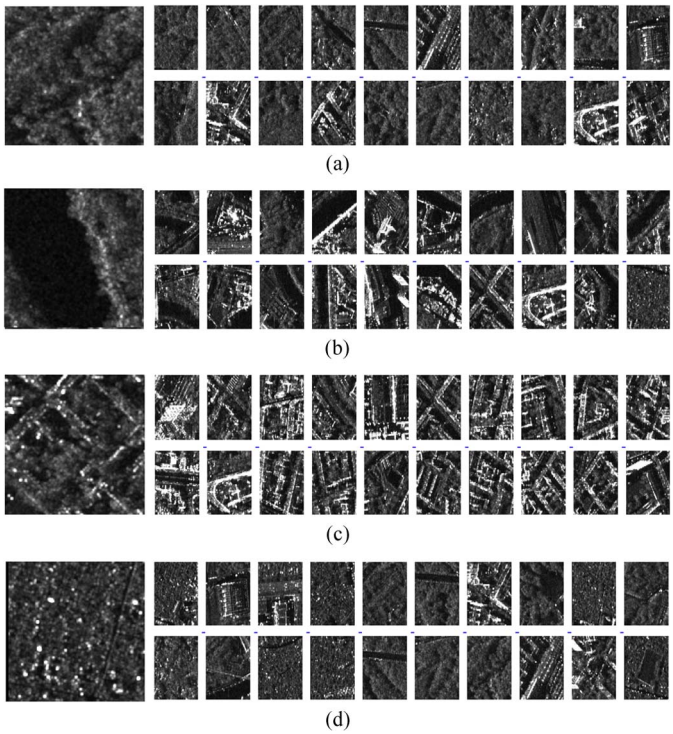

Fig. 18. CBIR results. (Left) Query image and (right) retrieved images presented as results. (a) Coniferous forest example. (b) Patch containing a river. (c) Example of medium-density urban area. (d) Low-density urban area patch. The precision and recall are computed at 20 retrieved images. (a) *Coniferous forest*: Precision $= 70\%$; Recall $= 16\%$. Total of 83 images. (b) *River*: Precision $= 70\%$; Recall $= 17\%$. Total of 80 images. (c) *Medium-density urban area*: Precision $= 35\%$; Recall $= 10\%$. Total of 66 images. (d) *Low-density urban area*: Precision $= 40\%$; Recall $= 36\%$. Total of 22 images.

system are acceptable in the case of *coniferous forest* and *river* [Fig. 18(a) and (b)], where precision of the retrieval is 70%. Both scenes are classified as natural scenes in our taxonomy. Unfortunately, in the case of urban area [see Fig. 18(c) and (d)], the precision is lower than 50%. In the case of *medium-density urban area*, the precision is 35%, showing that the system is not able to distinguish this class, and finally, in the case of *low-density urban area*, the precision is only 40%, demonstrating misclassifications. The recall shows poor performance, since we are using only the first 20 retrieved patches for its computation.

V. Conclusion

In this paper, an EO IR system has been introduced in its first version. The design concept of the system is based on a three-layer architecture, which is modular and user oriented. The system uses a patch-based feature extraction approach, the combination of several descriptors for characterizing the image, semantic definition based on semiautomatic machine learning methods, and, as a central core, a database management system.

In order to demonstrate the functionality of EO retrieval, we have created a semantic catalog using 39 high-resolution TerraSAR-X images, which were tiled into several patches and labeled with semantic annotations. Later, in order to exploit this catalog, we presented a series of advanced queries

using attributes like enriched metadata, semantic labels, spatial information derived directly from the metadata of the image (i.e., proximity, containment, etc.), and image content. A strong point of the system is that it can perform both attribute and spatial queries to get answers with far less effort or answer questions that would not be practical to answer using other traditional methods.

At this point of implementation, we have the following conclusions: From the implementation point of view, the use of database management system allows handling and exploring large volumes of data as well as the integration of several data resources in a proper and efficient way. From the research point of view, the combination of different descriptors like metadata, semantic, and image content enriches the image characterization which is reflected in the kind of queries than can be answered by the systems, becoming a powerful tool for IR.

The integration of ontologies for providing more advanced and user-friendly queries will be a future work. Increasing the size of the catalog is done by ingestion of more images, in order to reach a significative amount of data for testing the advantages of using a database scheme. The evaluation of the performance of the system in terms of speed of the retrieval and accuracy of the results is needed.

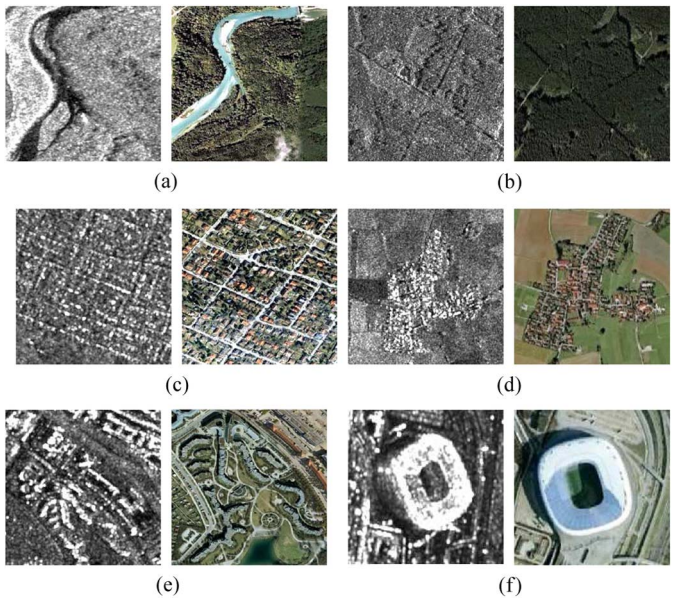

Fig. 19. Different types of natural and man-made scenes in TerraSAR-X image over Munich (side by side with optical Google-Earth reference images). (a) Isar river. (b) Woods and hedgerows. (c) Small houses with gardens. (d) Small town. (e) Curved high buildings. (f) Allianz Arena stadium.

APPENDIX

A. Basic Level 1b TerraSAR-X Products

TerraSAR-X can be operated in four *basic imaging modes* that have been designed to support a variety of applications, ranging from medium-resolution polarimetric imaging to high-resolution mapping. The standard imaging modes defined for the generation of basic product are as follows: Stripmap mode (SM), spotlight mode (SL), high-resolution spotlight mode (HS) in single or dual polarization, and ScanSAR mode (SC) in single polarization. The configuration of the imaging modes offers products with different resolutions in azimuth and range [20].

Considering the case of amplitude, TerraSAR-X level 1b products are generated and delivered as directories with subdirectories in compressed tar file, which comprises the image data in GeoTIFF format, a selection of quick-look and browse images provided as conventional TIFF images, a footprint map image in png format, a comprehensive list of image metadata including processing parameters in XML format, and a series of XML Schema Definition (XSD) files defining the XML layout [21].

1) TerraSAR-X Image: The GeoTIFF image contains, apart from the image pixels, a small set of internal annotation data defining the type of data, the dimension of the image, and some geographical coordinates (depending on the geometrical processing steps already done). TerraSAR-X images are characterized by their large size due to their large coverage with high resolution. Each TerraSAR-X image has to be ordered in a selectable data representation together with additional options. There are four alternative representations.

1) Single-look slant range complex, which is an image with complex values.
2) MGD, which is a detected multilook product with reduced speckle and approximately square resolution cells on ground.
3) Geocoded ellipsoid corrected corresponds to a multilook detected product. It is projected and resampled to the WGS84 reference ellipsoid assuming one average terrain height.
4) Enhanced ellipsoid corrected, which is a multilook detected product. It is projected and resampled to the WGS84 reference ellipsoid. The image distortions caused by varying terrain height are corrected using an external digital elevation model.

Each image can be either a spatially enhanced (SE) or RE product. The SE product is designed for the highest possible square ground resolution, while the RE product is optimized with respect to radiometry. In the latter, the range and azimuth resolutions are intentionally decreased to significantly reduce speckle [20]. Typical examples of TerraSAR-X image content are shown in Fig. 19. The subscenes were taken from a Munich TerraSAR-X scene, an RE Stripmap product in MGD with single polarization.

Fig. 19(a)–(e) displays scenes with low backscatterers, which appear in natural areas such as rivers, lakes, and medium brightness patterns which appear in vegetation areas. Locally extended brightness patterns are typical for urban structures and infrastructures (i.e., stadium).

2) TerraSAR-X Metadata: A typical product metadata is an XML file ASCII based and readable by common tools (e.g., a Web browser or simple text editors) and not a binary format; the indicated data types (strings, integers, doubles, etc.) for most of the annotations are the intrinsic default types. Formal definitions and their full XML nesting are given in the corresponding XSD files.

The TerraSAR-X XML file comprises ∼250 entries grouped into categories such as product components, product information, processing, scene information like geographic coordinates, platform, calibration, and product quality. The following list corresponds to the ordering of the entries within the overall metadata file and refers to a typical MGD image. Note that the actual parameters contained in the list are to be considered as typical examples; some parameters only appear when specific processing options have been selected. Omissions within the sequence of entries are marked by ⟨. . . . . .⟩.

⟨generalHeader⟩...
  ⟨itemName⟩Level 1B Product⟨/itemName⟩
  ⟨mission⟩TSX-1⟨/mission⟩
⟨/generalHeader⟩
⟨missionInfo⟩
  ⟨orbitDirection⟩DESCENDING⟨/orbitDirection⟩
  ⟨. . . . . .⟩
⟨/missionInfo⟩
⟨acquisitionInfo⟨...
  ⟨lookDirection⟩RIGHT⟨/lookDirection⟩
  ⟨polarizationMode⟩SINGLE⟨/polarizationMode⟩
        ⟨polLayer⟩VV⟨polLayer⟩
      ⟨...⟩
  ⟨/acquisitionInfo⟩
  ⟨productVariantInfo⟩
    ⟨productType⟩MGD-RE-HS-S⟨/productType⟩
    ⟨productVariant⟩MGD⟨/productVariant⟩
  ⟨/productVariantInfo⟩
  ⟨imageDataInfo⟩
    ⟨pixelvalueID⟩RADAR BRIGHTNESS⟨/pixelValueID⟩
    ⟨imageDataType⟩DETECTED⟨/imageDataType⟩
    ⟨. . . . . .⟩
    ⟨imageRaster⟩
      ⟨numberOfRows⟩8400⟨/numberOfRows⟩
      ⟨numberOfColumns⟩10 000⟨/numberOfColumns⟩
      ⟨rowSpacingunits = m⟩1.25000E + 00⟨/ . . .⟩
      ⟨columnSpacingunits = m⟩1.25000E + 00⟨/ . . .⟩
      ⟨groundRangeResolution⟩2.90357 . . . E + 00⟨/ . . .⟩
      ⟨azimuthLooks⟩2.62392E + 00⟨/azimuthLooks⟩
      ⟨rangeLooks⟩2.94849E + 00⟨/rangeLooks⟩
      ⟨/imageRaster⟩
⟨/imageDataInfo⟩
⟨sceneInfo⟩
    ⟨start⟩
      ⟨timeUTC⟩2010-12-26T04:35:39.158784z
      ⟨/timeUTC...⟩
    ⟨stop⟩
      ⟨timeUTC⟩2010-12-26T04:35.39.906253z
      ⟨/timeUTC...⟩
    ⟨sceneCenterCoord⟩
      ⟨. . . . . .⟩
    ⟨incidenceAngle⟩3.593E + 01⟨/incidenceAngle⟩
  ⟨/sceneCenterCoord⟩
⟨/sceneInfo⟩

REFERENCES

[1] M. Wolfmuller, D. Dietrich, E. Sireteanu, S. Kiemle, E. Mikusch, and M. Bottcher, "Data flow and workflow organization—The data management for the TerraSAR-X payload ground segment," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 44–50, Jan. 2009.

[2] T. Kato, "Cognitive view mechanism for multimedia database system," in *Proc. 22nd Annu. Can. Remote Sens. Symp.*, Victoria, BC, Canada, 1991, pp. 179–186.

[3] T. Kato, "Database architecture for content-based image retrieval," in *Proc. SPIE Conf. Image Storage Retrieval Syst.*, 1992, pp. 112–123.

[4] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by image and video content: The QBIC system," *IEEE Comput.*, vol. 28, no. 9, pp. 23–32, Sep. 1995.

[5] T. Gevers and A. W. M. Smeulders, "PicToSeek: Combining color and shape invariant features for image retrieval," *IEEE Trans. Image Process.*, vol. 9, no. 1, pp. 102–119, Jan. 2000.

[6] I. J. Cox, M. L. Miller, T. P. Minka, T. V. Papathomas, and P. N. Yianilos, "The Bayesian image retrieval system, PicHunter: Theory, implementation, and psychophysical experiments," *IEEE Trans. Image Process.*, vol. 9, no. 1, pp. 20–37, Jan. 2000.

[7] M. Datcu, H. Daschiel, A. Pelizzari, M. Quartulli, A. Galoppo, A. Colapicchioni, M. Pastori, K. Seidel, P. G. Marchetti, and S. D'Elia, "Information mining in remote sensing image archives: System concepts," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 12, pp. 2923–2936, Dec. 2003.

[8] 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.

[9] 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.

[10] R. Picard, "Digital libraries: Meeting place for high-level and low-level vision," in *Recent Developments in Computer Vision*, vol. 1035, *Lecture Notes in Computer Science*, S. Li, D. Mital, E. Teoh, and H. Wang, Eds. Berlin, Germany: Springer-Verlag, 1996, pp. 1–12.

[11] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: Semantics-sensitive integrated matching for picture libraries," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 9, pp. 947–963, Sep. 2001.

[12] N. Rasiwasia, P. J. Moreno, and N. Vasconcelos, "Bridging the gap: Query by semantic example," *IEEE Trans. Multimedia*, vol. 9, no. 5, pp. 923–938, Aug. 2007.

[13] Department of Information StudiesThe University of Sheffield, Tripod Automating Caption Creation, Jan. 2009. [Online]. Available: http://tripod.shef.ac.uk/

[14] X. Li, C. G. M. Snoek, M. Worring, and A. W. M. Smeulders, "Fusing concept detection and geo context for visual search," in *Proc. 2nd ACM ICMR*, 2012, pp. 4:1–4:8.

[15] A. Pentland, R. W. Picard, and S. Sclaroff, "Photobook: Content-based manipulation of image databases," *Int. J. Comput. Vision*, vol. 18, no. 3, pp. 233–254, Jun. 1996.

[16] J. K. Wu, A. Desai Narasimhalu, B. M. Mehtre, C. P. Lam, and Y. J. Gao, "CORE: A content-based retrieval engine for multimedia information systems," *Multimedia Syst.*, vol. 3, no. 1, pp. 25–41, Feb. 1995.

[17] W. Hsu, M. L. Lee, and J. Zhang, "Image mining: Trends and developments," *J. Intell. Inf. Syst.*, vol. 19, no. 1, pp. 7–23, Jul. 2002.

[18] N. Longbotham, F. Pacifici, and W. J. Emery, "Multi-temporal and multi-angular analysis of very high spatial resolution images," in *Proc. IEEE IGARSS*, 2012, pp. 1996–1999.

[19] F. Covello, F. Battazza, A. Coletta, G. Manoni, and G. Valentini, "Cosmo-skymed mission status: Three out of four satellites in orbit," in *Proc. IEEE IGARSS*, Jul. 2009, vol. 2, pp. II-773–II-776.

[20] DLR, TerraSAR-X Ground Segment Basic Product Specification Document, TX-GS-DD-3302, Dec. 2008. [Online]. Available: http://sss.terrasar-x.dlr.de/pdfs/TX-GS-DD-3302.pdf

[21] DLR, TerraSAR-X, Ground Segment, Level 1b Product Data Specification, TX-GS-DD-3307, Dec. 2007. [Online]. Available: http://sss.terrasar-x.dlr.de/pdfs/TX-GS-DD-3307.pdf

[22] P. Anca, I. Gavat, 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.

[23] P. Birjandi and M. Datcu, "Multiscale and dimensionality behavior of ica components for satellite image indexing," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 1, pp. 103–107, Jan. 2010.

[24] J. Steggink and C. G. M. Snoek, "Adding semantics to image-region annotations with the name-it-game," *Multimedia Syst.*, vol. 17, no. 5, pp. 367–378, Oct. 2011.
[25] EUR12585 Luxembourg, Office for Official Publications of the European Communities, Luxembourg Y. Heymann, C. Steenmans, G. Croissille, and M. Bossard, Corine Land Cover. Technical Guide 1994, EUR12585 Luxembourg, Office for Official Publications of the European Communities, Luxembourg.
[26] U.S. Geological Survey Professiona paper 671 J. R. Anderson, E. E. Hardy, and J. T. Roach, A Land Use and Land Cover Classification System for Use With Remote Sensor Data 1972, U.S. Geological Survey Professiona paper 671.
[27] T. Watanabe, K. Sugawara, and H. Sugihara, "A new pattern representation scheme using data compression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 579–590, May 2002.
[28] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 837–842, Aug. 1996.
[29] S.-F. Chang, T. Sikora, and A. Purl, "Overview of the MPEG-7 standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 688–695, Jun. 2001.
[30] T. A. Welch, "A technique for high-performance data compression," *Computer*, vol. 17, no. 6, pp. 8–19, Jun. 1984.
[31] CWI, Monetdb Copyright (c), Mar. 2008–2011. [Online]. Available: http://www.monetdb.org/Home
[32] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, Apr. 2011.
[33] C. Dumitru and M. Datcu, "Dependency of SAR image structure descriptors with incidence angle," in *Proc. 4th Int. Conf. Adv. SPACOMM*, Chamonix, France, Apr. 2012, pp. 92–97.
[34] R. R. Larson, "Geographic information retrieval and spatial browsing," in *Proc. GIS Libraries: Patrons, Maps Spatial Inf.*, Apr. 1996, pp. 81–124.
[35] D. Cerra and M. Datcu, "A fast compression-based similarity measure with applications to content-based image retrieval," *J. Vis. Commun. Image Represent.*, vol. 23, no. 2, pp. 293–302, Feb. 2012.
[36] K. A. Musiol, G. Muehling, H. Bronshtein, and I. N. Semendyayev, *Handbook of Mathematics*. New York, NY, USA: Springer-Verlag, 2004.
[37] C. D. Manning, P. Raghavan, and H. Schütze, *An Introduction to Information Retrieval*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

Daniela Espinoza-Molina (M'11) received the Engineering degree from Universidad de Cuenca, Cuenca, Ecuador, in 2002, the M.S. degree in geographical information systems from Universidad Pontificia de Salamanca, Madrid, Spain, in 2007, and the Ph.D. degree from Telecom Paris Tech, Paris, France, in 2011.

She is currently with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Wessling, Germany. Her main research interests include image processing and image information mining.

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), Romania, in 1978 and 1986, respectively. In 1999, he received the title "Habilitation á diriger des recherches" in computer science from University Louis Pasteur, Strasbourg, France.

Since 1981, he has been a Professor with the Faculty of Electronics, Telecommunications and Information Technology, UPB, working in signal/image processing and Electronic Speckle Interferometry. Since 1993, he has been a scientist with the German Aerospace Center (DLR), Wessling, Germany. He is developing algorithms for analyzing Very High Resolution Synthetic Aperture Radar (VHR SAR) and Interferometric SAR (InSAR) data. He is engaged in research related to information theoretical aspects and semantic representations in advanced communication systems. Currently, he is Senior Scientist and Image Analysis research group leader with the Remote Sensing Technology Institute of DLR, Wessling. Since 2011, he is also leading the Immersive Visual Information Mining research laboratory at the Munich Aerospace Faculty and is director of the Research Center for Spatial Information at UPB. He has held Visiting Professor appointments with the University of Oviedo, Spain, University Louis Pasteur and the International Space University, in Strasbourg, France, University of Siegen, Germany, University of Camerino, Italy, and the Swiss Center for Scientific Computing, Manno, Switzerland. From 1992 to 2002 he had a longer Invited Professor assignment with the Swiss Federal Institute of Technology, ETH Zurich. Since 2001, he has initiated and leaded the Competence Centre on Information Extraction and Image Understanding for Earth Observation, at ParisTech, Telecom Paris, a collaboration of DLR with the French Space Agency (CNES). He has been Professor holder of the DLR-CNES Chair at ParisTech, Telecom Paris. His interests are in information and complexity theory, stochastic processes, Bayesian inference, and Image Information Mining (IIM). He and his team have developed and are currently developing the operational IIM processor in the Payload Ground Segment systems for the German missions TerraSAR-X, TanDEM-X, and the ESA Sentinel 1 and 2. He is the author of more than 200 scientific publications, among them about 50 journal papers, and a book on number theory. He is a member of the European Image Information Mining Coordination Group (IIMCG) and of the Data Archiving and Distribution Technical Committee (DAD TC) of the IEEE Geoscience and Remote Sensing Society.