Information Content of Very-High-Resolution SAR Images: Semantics, Geospatial Context, and Ontologies

Corneliu Octavian Dumitru, Shiyong Cui, Gottfried Schwarz, and Mihai Datcu, *Fellow, IEEE*

Abstract—Currently, the amount of collected Earth Observation (EO) data is increasing considerably with a rate of several Terabytes of data per day. As a consequence of this increasing data volume, new concepts for exploration and information retrieval are urgently needed. To this end, we propose to explore satellite image data via an image information mining (IIM) approach in which the main steps are feature extraction, classification, semantic annotation, and interactive query processing. This leads to a new process chain and a robust taxonomy for the retrieved categories capitalizing on human interaction and judgment. We concentrated on land cover categories that can be retrieved from high-resolution synthetic aperture radar (SAR) images of the spaceborne TerraSAR-X instrument, where we annotated different urban areas all over the world and defined a taxonomy element for each prevailing surface cover category. The annotation resulted from a test dataset comprising more than 100 scenes covering diverse areas of Africa, Asia, Europe, the Middle East, and North and South America. The scenes were grouped into several collections with similar source areas and each collection was processed separately in order to discern regional characteristics. In the first processing step, each scene was tiled into patches. Then the features were extracted from each patch by a Gabor filter bank and a support vector machine with relevance feedback classifying the feature sets into user-oriented land cover categories. Finally, the categories were semantically annotated using Google Earth for ground truthing. The annotation followed a multilevel approach that allowed the fusion of information being visible on different resolution levels. The novelty of this paper lies in the fact that a semantic annotation was performed with a large number of high-resolution radar images that allowed the definition of more than 850 surface cover categories. This opens the way toward an automated identification and classification of urban areas, infrastructure (e.g., airports), geographic objects (e.g., mountains), industrial installations, military compounds, vegetation, and agriculture. Applications that may result from this work can be a semantic catalog of urban images to be used in crisis situations or after a disaster. In addition, the proposed taxonomies can become a basis for building a semantic catalog of satellite images. Finally, we defined four powerful types of high-level queries. Querying on such high levels provides new opportunities for users to search an image database for specific parameters or semantic relationships.

I. INTRODUCTION

WITH THE increased availability of Earth Observation (EO) data, due to new satellite missions, their various sensors, and the interoperability of data archives, today's remote sensing community is faced with a dramatic increase in data volume, data content, and processing levels. This coincides with the era of "Big Data" [1], [2] that refers to increasing data *volume* (amount of data), *velocity* (speed of data to be transferred), and *variety* (range of data types and sources). For instance, the German Aerospace Center is hosting more than 1.5 million remote sensing data products occupying about 2 PB of storage [4].

In this study, we address two of three properties of "Big Data," namely the *variety* of data represented by the diversity of data formats and content and the *volume* of data represented by the number of available products. For solving typical satellite image analysis tasks within this framework, we propose a semiautomated methodology to explore the content of given synthetic aperture radar (SAR) satellite images and to extract the categories (i.e., the object classes) that are contained in them. Further, we define a specific taxonomy for each retrieved category based on the local geospatial context, and we propose some high-level querying techniques to be applied to remote sensing data archives.

In order to illustrate the diversity of categories that can exist in remote sensing images, Fig. 1 shows a set of subimages (called patches) tiled from product images of the German TerraSAR-X mission with a pixel spacing of about 1 m [3]. It is quite evident that the diversity of categories that can be retrieved from a high-resolution image is much higher than in the case of low- or medium-resolution images.

The main contributions of the present study are: 1) a semiautomated semantic annotation of image patches based on a large collection of high-resolution land cover SAR images; 2) a reliable classification of urban areas and infrastructure; 3) a general multilevel taxonomy to be used for the analysis of remote sensing images contained in databases; 4) fusion of the semantic annotation with information coming from external databases (e.g., GIS data); and 5) integration of high level querying with image data archives.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

2   IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING
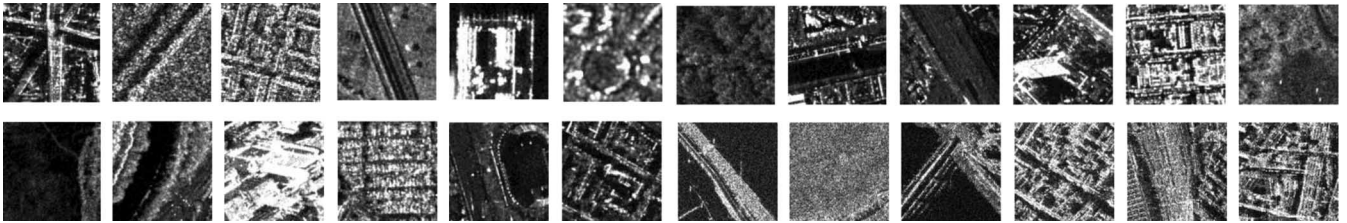
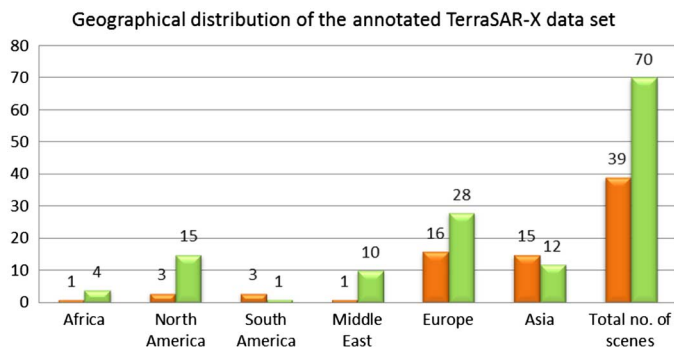Fig. 1. Diversity of patches tiled from high-resolution SAR images.

Fig. 2. Geographical distribution of the selected scenes acquired in 2012 (orange) and 2013 (green).

The organization of this paper is as follows. Section II describes the state-of-the-art of semantic annotation for multimedia and remote sensing images. Section III presents the typical characteristics of high-resolution SAR satellite images. Section IV describes our methodology used for semantic annotation. Section V outlines the elements of image content ontologies, whereas Section VI investigates image content taxonomies. Section VII explains several types of queries that can be applied to image data archives. Section VIII describes current developments and Section IX summarizes the conclusion of this paper.

II. SEMANTIC ANNOTATION FOR MULTIMEDIA AND REMOTE SENSING IMAGES

When we attach text labels to elements of images that explain their content or meaning, we call this step "labeling" or "semantic annotation." If this annotation can be automated, we have a tool for automated image content understanding and many researchers have tried to make some progress in this field. Traditionally, the conceptions and the performance of semantic annotations have been verified based on publicly available collections of typical image data sets. Most of them have been built by individual groups with the intention to solve specific problems. Thus, for the different domains of multimedia (e.g., press and television archives) and remote sensing applications, the available data sets contain a large variety of retrievable objects and categories. While most multimedia applications aim at the recognition of single objects in front of mostly irrelevant background (*a cow on a meadow*), typical remote sensing applications call for the identification of land cover or sea surface details covering the full image area (*icebergs on the Atlantic Ocean*). During the last years, we also saw a

growing interest in satellite images in order to support disaster and emergency relief. Within this context, rapid mapping services provide information support by delivering rapid mapping products emphasizing the extent and impact of an event. Rapid mapping activations cover a large range of applications, such as natural and environmental hazards (e.g., floods, hurricanes, fires, earthquakes, and oil spills) and humanitarian relief activities (e.g., mapping of refugee camps).

We start with a summary of multimedia databases because they represent a mature domain with more than 25 years of experience in close range digital photography; we have to mention image data collections such as ImageCLEF Wikipedia [5], PASCAL [6], ImageNet [7], and LabelMe [8]. A rather complete list of data sets for multimedia applications can be found in [8].

The ImageCLEF Wikipedia [5] collection consists of 237 434 images for which 137 users provided annotations. The collection covers different topics in several languages. The sources of the data sets are heterogeneous combining images with metadata, text, and additional links. During the annotation, users identified about 70 topics (e.g., *Oktoberfest beer tent, trains and locomotives, harbors, civil airplanes,* and *race cars*).

The ImageNet Large Scale Visual Recognition Challenge [7] is a project that had as its objective the classification of 1 229 413 images (a subset of the ImageNet data set) into about 1000 categories with spatial localization of the objects. This evaluation completed a part of the PASCAL project [6]. The source data having been combined here are images from ImageNet and additional information from WordNet (source of the labels, semantical hierarchy, English nouns, etc.). For these categories, different taxonomies are proposed (e.g., *mammal→placental→carnivore→canine→dog→working dog →husky*). The main goal in [6] is to recognize objects from a number of visual object categories in real world scenes. There are different tasks that are being discussed in [6]: classification/detection (20 classes), segmentation, person layout, action classification (11 classes), and ImageNet large scale visual recognition (these data sets contain 10 000 000 labeled images that represent 10 000 semantically annotated objects).

LabelMe [8] is another project that provides a data set of images with their associated annotations. The text data set used in order to label the objects is the same WordNet that was used in the PASCAL project. The LabelMe tool is available online and allows computer vision researchers to annotate large data sets. Between its implementation in 2005 and the end of 2010, the LabelMe tool produced 62 197 annotated images (from a
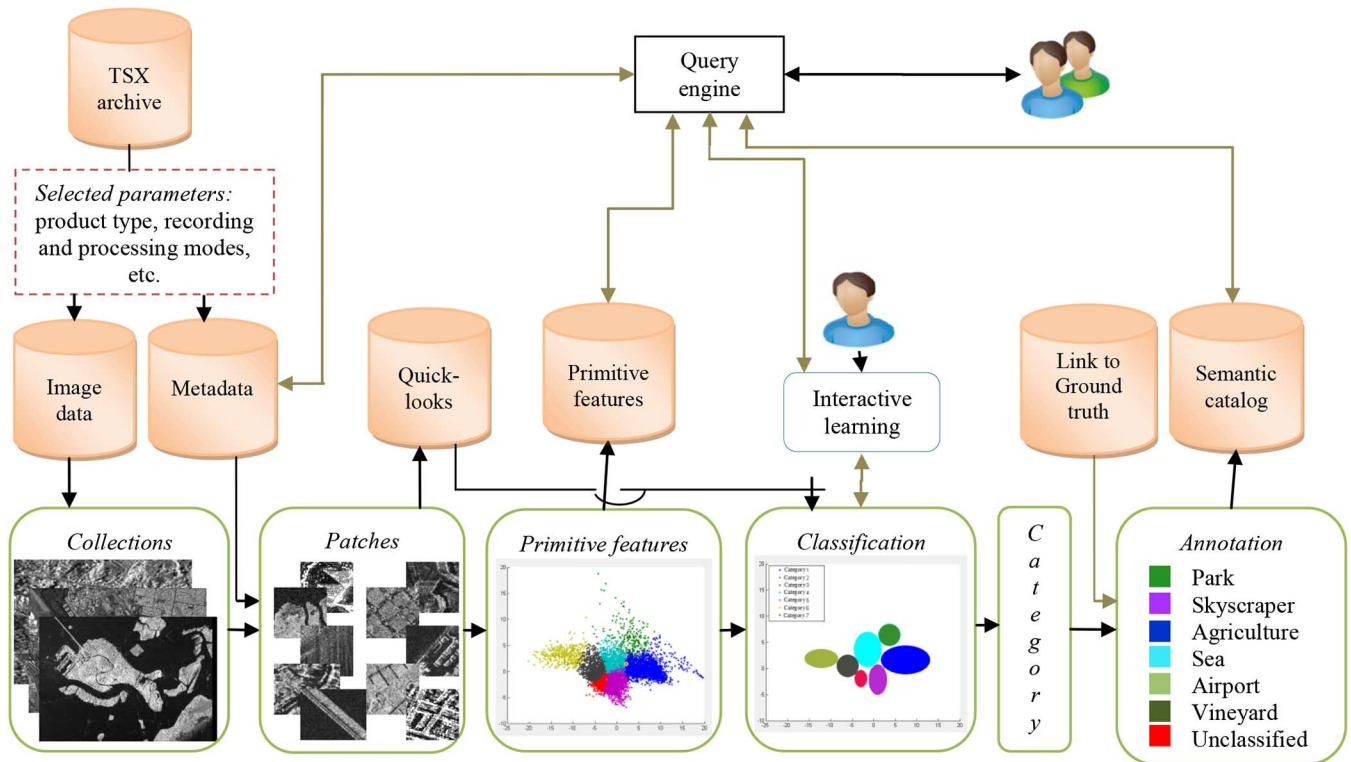
Fig. 3. Annotation methodology: download and store the selected TerraSAR-X products, group the images into collections, tile each image into patches, and generate a quick-look of each patch; extract the primitive features from each patch, group the extracted features into categories using interactive learning, and provide semantic annotation for each patch using Google Earth as visual support. Once this process has been completed, an operator can put various queries.
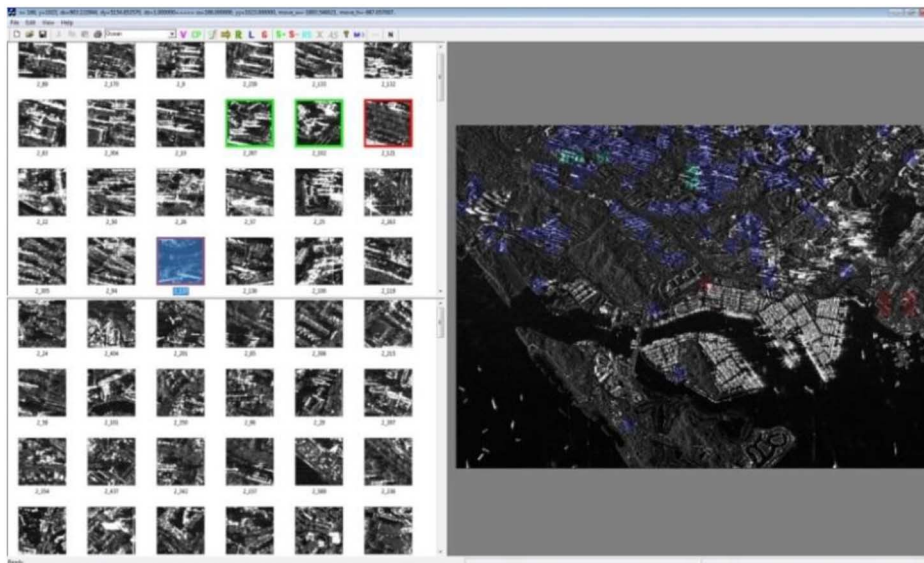

Fig. 4. Interactive interface to retrieve images belonging to categories that exist in a collection. The upper-left half shows relevant retrieved patches, whereas the lower-left half shows irrelevant retrieved patches. The large panel on the right shows the image that is being worked on and which can be zoomed. Users can see the distribution of the retrieved patches and all the training samples. They can also verify the selected training samples by checking their surroundings, as there is a link between the patches in the upper-left half and the right half.

data set of 187 240 images) that contain 658 992 annotated objects.

As for remote sensing, image classification and annotation can be divided into two steps: the first step defines the number of classes that can be retrieved using different features extracted from an image patch and the second step is linked to specific applications (e.g., rapid mapping after a disaster).

A typical database is the GeoIRIS image data collection. The GeoIRIS system is capable of automatically retrieving a large volume of selected multispectral data sets from existing
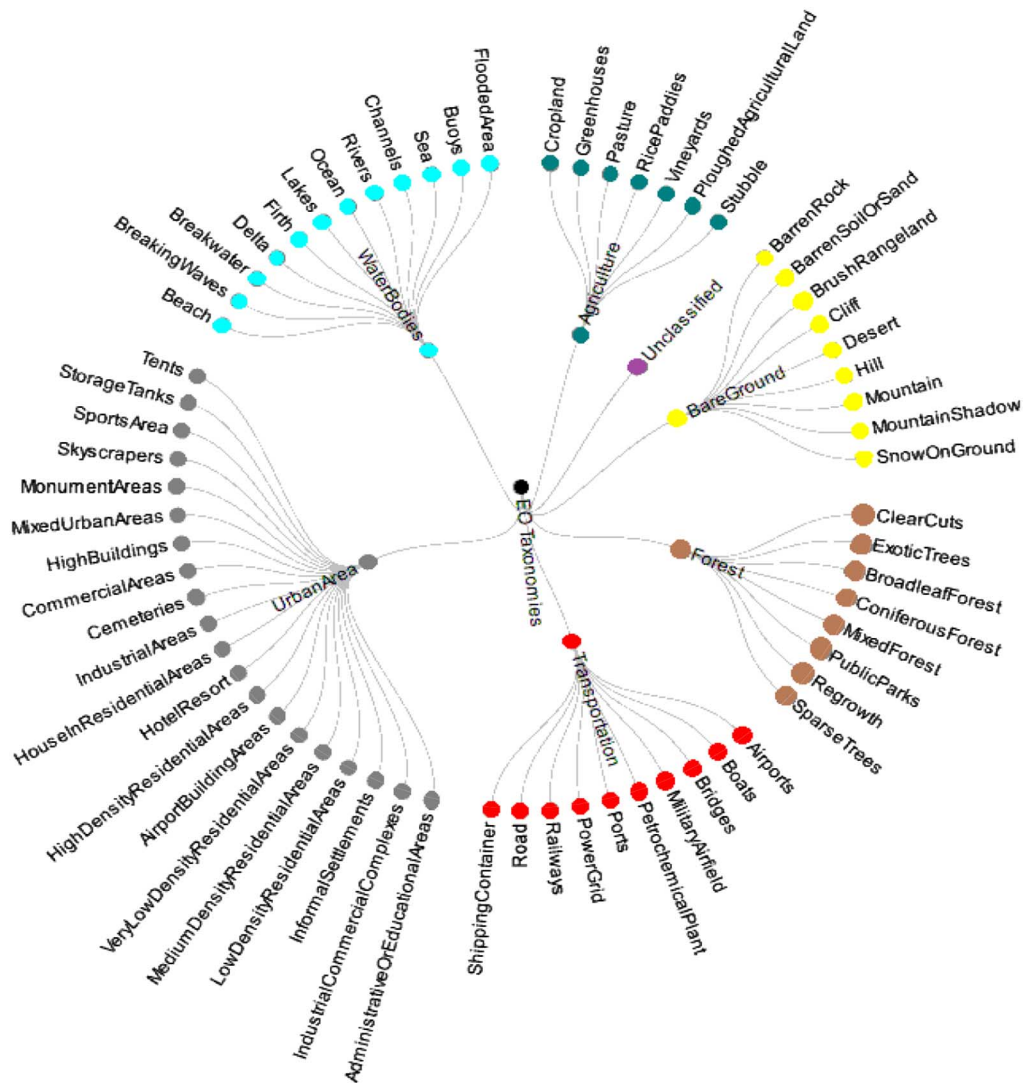
Fig. 5. Our proposed two-level hierarchical annotation scheme: Level 1 gives us a general information (e.g., *urban area*, *transportation*, *forest*, and *agriculture*), whereas level 2 provides additional details (e.g., *transportation* contains *airports*, *bridges, boats, ports,* and *railways*).

satellite image archives. Here, optical satellite images are tiled into patches of 250×250 pixels obtaining in the case of [9] 70 824 patches from which 531 208 objects were identified.

Another data set was described in [10] illustrating a patch contextual approach based on topographic and independent component analysis for very-high-resolution optical satellite images. This method is particularly useful for geometrical structures being contained in images of urban areas. The method was tested with QuickBird satellite images. Each test image was tiled into patches of 200×200 pixels thus generating 20 000 patches together with their associated features (gradients, topographic and independent components, as well as Gabor filter feature vectors) that were further used to retrieve 18 categories.

Still another method for feature extraction from image patches that captures the spatial context is presented in [11]. In this application, high-resolution SAR images yielded a set of 7000 patches with a size of 200×200 pixels that were grouped into 30 categories using a short time Fourier transform and a gray level co-occurrence matrix as feature extractors.

Finally, we generated a SAR image test data set when we tested the performance of selected feature extraction algorithms with high-resolution SAR images [12]. Each image was tiled into patches of different sizes generating 2170 patches that finally were grouped into 30 categories.

In the case of remote sensing, one can obtain a variable number of categories. In general, one obtains less than 10 categories for low-resolution images and more than 30 categories for high-resolution images, depending on the properties of the categories selected (i.e., manually extracted) by a user, while the number of categories that were identified in rapid mapping applications after a disaster is typically limited to 2 to 7 [13], [14]. In summary, the number of identified and annotated remote sensing categories is very limited compared with multimedia.

III. SELECTED TEST DATA SET

In this section, we prepare the analysis of the information content and the annotation of SAR images by selecting typical TerraSAR-X products and the generation of a dedicated test
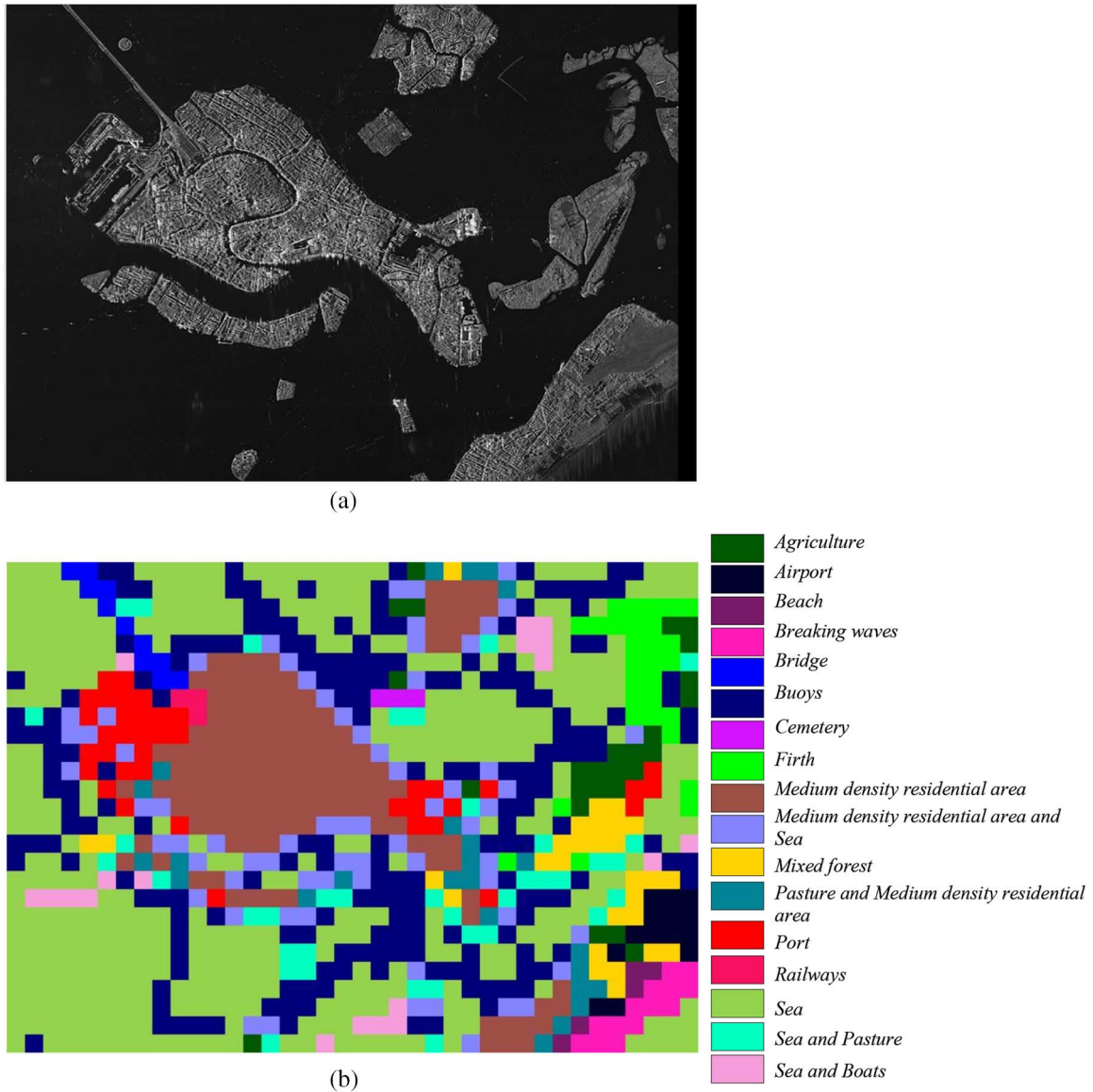
(a)

(b)

Fig. 6. (a) Original TerraSAR-X Venice image. (b) Categories for the Venice image using our proposed annotation scheme.

data set. TerraSAR-X is a German radar satellite launched in June 2007. It operates in X-band and is a side-looking SAR instrument that offers a wide selection of operating modes and product generation options [15], [16].

We selected high-resolution spotlight mode images because they provide the highest resolution data of the target areas. As for the product generation options, we took multilook ground range-detected (MGD) data because they are not affected by geometrical interpolation effects over mountainous terrain and thus are most suited for feature extraction. This was also the reason for choosing radiometrically enhanced products that are optimized with respect to radiometry (i.e., reduced speckle).

Finally, we took horizontally polarized (HH) images, as this option is most frequently used. The images have a pixel spacing of 1.25 m and a resolution about 2.9 m. The average size of the images is 4200 rows × 6400 columns.

Our data set contains scenes that cover different urban areas all over the world (see Fig. 2): 5 scenes from Africa, 27 scenes from Asia, 44 scenes from Europe, 11 scenes from the Middle East, and 22 scenes from North/South America.

IV. IMAGE ANNOTATION METHODOLOGY

In this section, we propose a semiautomatic procedure in order to annotate TerraSAR-X during a learning phase [17]. Fig. 3 presents the flowchart of the proposed semantic annotation methodology.

The selected TerraSAR-X products are downloaded from their archives and stored into our database. Once the images are available, they are grouped into geographically similar collections and processed as illustrated in Fig. 3.

This annotation methodology is semiautomated, i.e., the first three functions of the methodology (patch tiling,
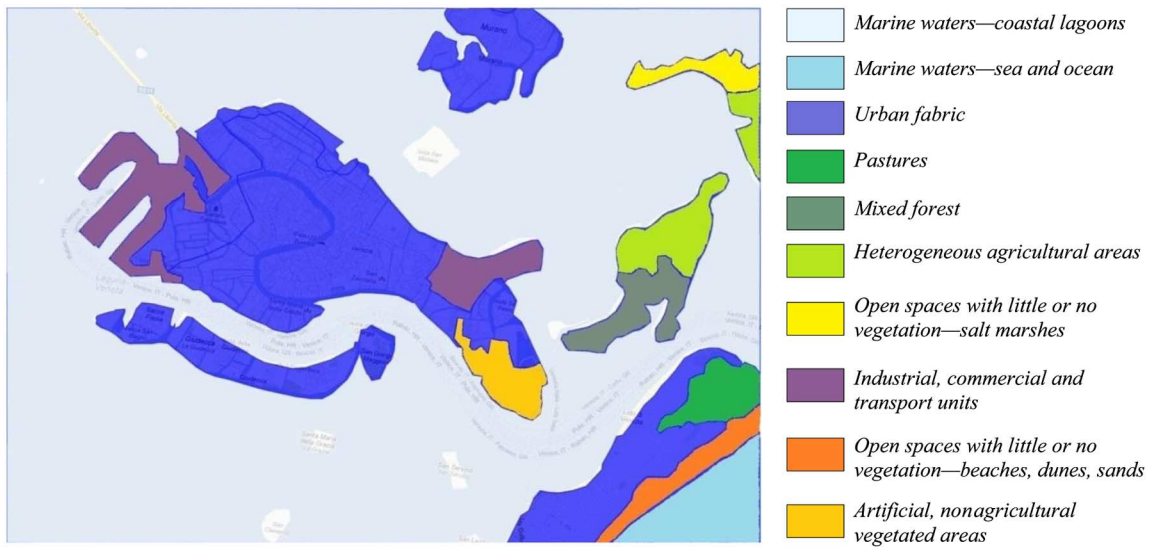
Fig. 7. Categories for the Venice image using the CORINE Land Cover nomenclature.
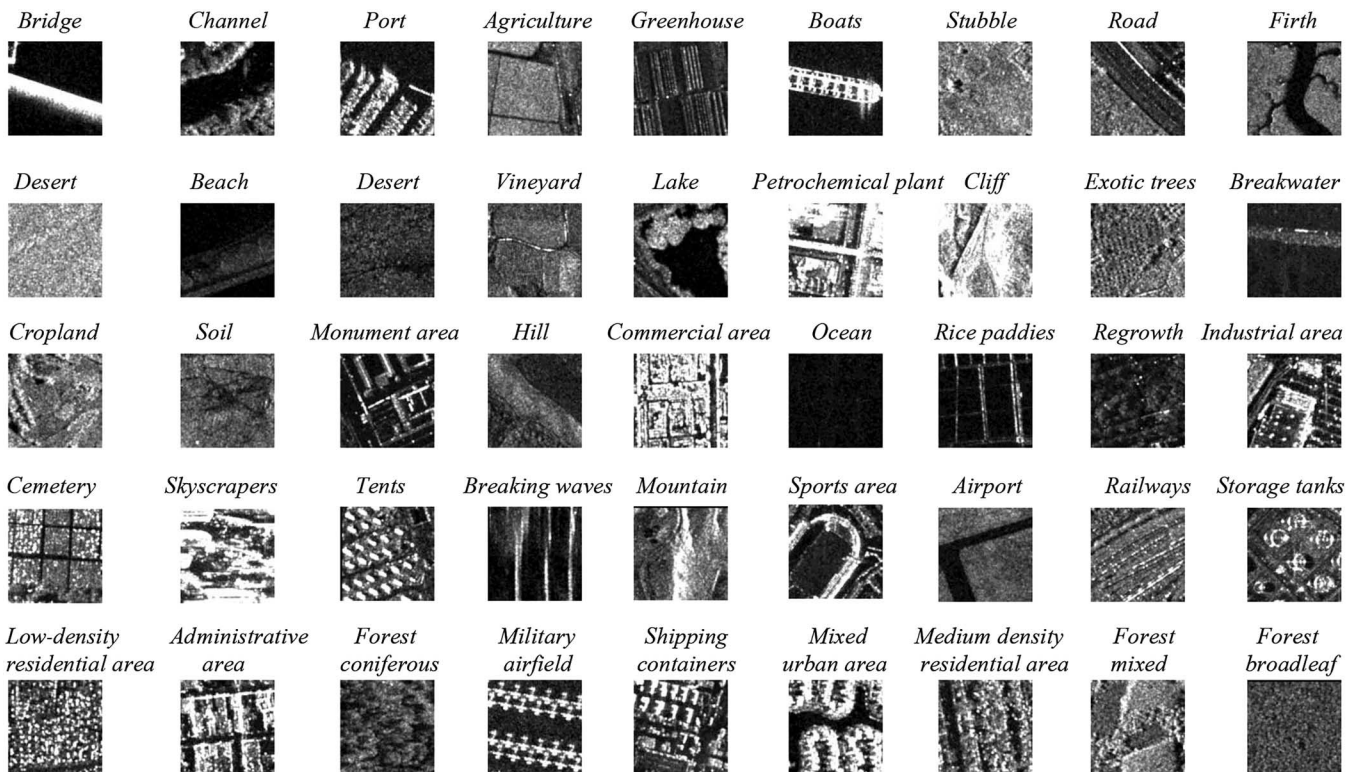

Fig. 8. Typical categories retrieved from Bagdad (IQ), Bandar Imam Khomeini (IR), Binhai (CN), Genoa (IT), Nazca Lines (PE), Singapore (SG), Trento (IT), and Tucson (USA).

quick-look generation, and feature extraction) are fully automated while the classification and annotation functions require manual operator interaction. They require interaction because classification includes an operator to rank the patches via human–machine interaction (i.e., active learning) and annotation calls for the selection of the proper semantic labels for each category. Later, in order to exploit the image content, users can put different interactive queries for comfortable image content mining.

A. Selection of Collections

Because a scene-wise annotation without a chance for inter-comparisons may become difficult, we propose to group the scenes into collections of typical and similar cases and then to annotate them. The scenes will be grouped together based on the following three criteria.

1) First, we select a single scene for annotation to get an idea about how many categories can be retrieved for a given city/country/continent.
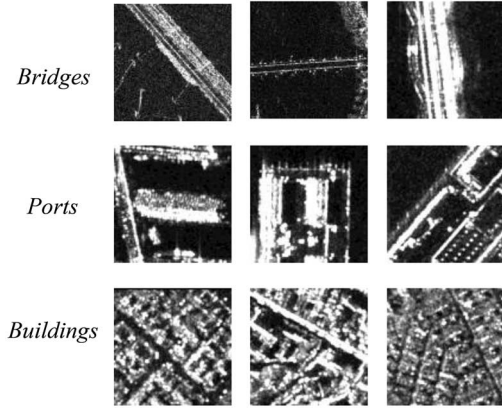
Fig. 9. Examples of categories that cannot be grouped together: *Bridges* in Venice (left column), Toulouse (center column), and Torun (right column); *Ports* in Havana (left column), Oran (center column), and Venice (right column); and *Buildings* in Dubai (left column), Venice (center column), and Vasteras (right column).
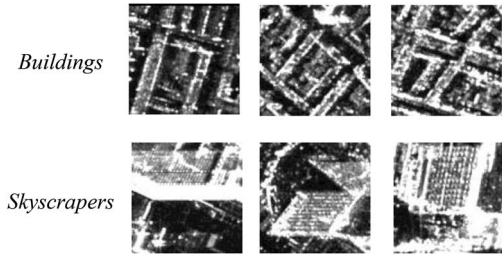

Fig. 10. Examples of categories that can be grouped together: *Buildings* in Munich (DE) (left column), Basel (SZ) (center column), and Kiel (DE) (right column) and *Skyscrapers* in Tokyo (JP) (left column), Dubai (center column), and Los Angeles (USA) (right column).

2) Then, we group together two scenes from different geographical areas (e.g., countries) to see whether the same urban categories can be found in both scenes.
3) Finally, we perform the grouping of more scenes with similar geographical location and/or architectural characteristics of the target areas to help us to annotate large areas.

The data set presented in Section III was split into about 45 collections following all three criteria. The annotation results are shown in Section VI.

B. Semantic Annotation

To annotate the data, the following steps are applied to each collection.
1) Tile the selected product images into patches of 160×160 pixels and discard empty edge patches of an image [17]. No further despeckling is applied to the SAR data.
2) Generate unscaled JPEG quick-look images of each patch and of the full image. Store all quick-looks into the database. This allows a quick interactive visualization of the images.
3) Extract a feature vector from each image patch. We use Gabor texture filters for feature extraction. Gabor texture features describe the statistics of subbands generated

by a set of Gabor filter banks. Generally, a Gabor function is a kind of wavelet transform, which has shown very good performance in texture discrimination. A two-dimensional Gabor function $g(x, y)$ is defined as follows:

$$g(x, y) = \left(\frac{1}{2\pi \sigma_x \sigma_y} \right) \exp \left[-\frac{1}{2} \left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) + 2\pi j W_x \right]. \tag{1}$$

For a detailed nomenclature, see [18].

Gabor filter banks based on the general Gabor function can be obtained by appropriate dilations and rotations. Thus, the number of scales and orientations are two important parameters that have to be carefully selected when extracting feature vectors from image patches. A set of filter responses is generated each time when applying a Gabor filter bank. The statistics of the subbands, i.e., mean and variance, are then used as feature vectors. We compared different feature extraction algorithms and found that Gabor filters show the highest accuracy in the comparison of similar image patches [12], [19].

4) Select a classifier in order to group the extracted features into categories. For this task, a support vector machine (SVM) tool with relevance feedback was built. The objective of an SVM is to separate two classes with a maximum margin. A two class linear SVM can then be formulated as a quadratic optimization problem [17]. In our case, we applied a chi-square kernel function.

In our implementation, users can search for patches of interest in a large repository having a full image on the right side of the display as a support (see Fig. 4). Our tool automatically suggests patches with rankings which are grouped into categories of relevance. This new visual support tool allows enhancing the quality of search results by giving positive and negative examples directly for a full image. Each patch is assigned to a single category based on its dominant content.

5) Annotate semantically each category using Google Earth for visual ground truthing based on the proposed hierarchical annotation scheme (see Fig. 5). The remaining unannotated patches are grouped in a category called "*unclassified*." The number of patches belonging to this category represents at most 10% of the total number of patches annotated in each collection, but for some collections this percentage is lower than 5%.

After the annotation of our test data, we can use the SVM parameters and the set of semantic categories for a fully automated semantic annotation of newly arriving images.

C. Querying

A query engine is an important part of an image mining system. It is an interactive component which allows normal and expert users to exploit the content of images. In our system, we implemented the following types of queries [20].
1) Query by metadata: One can exploit all metadata belonging to an image product.
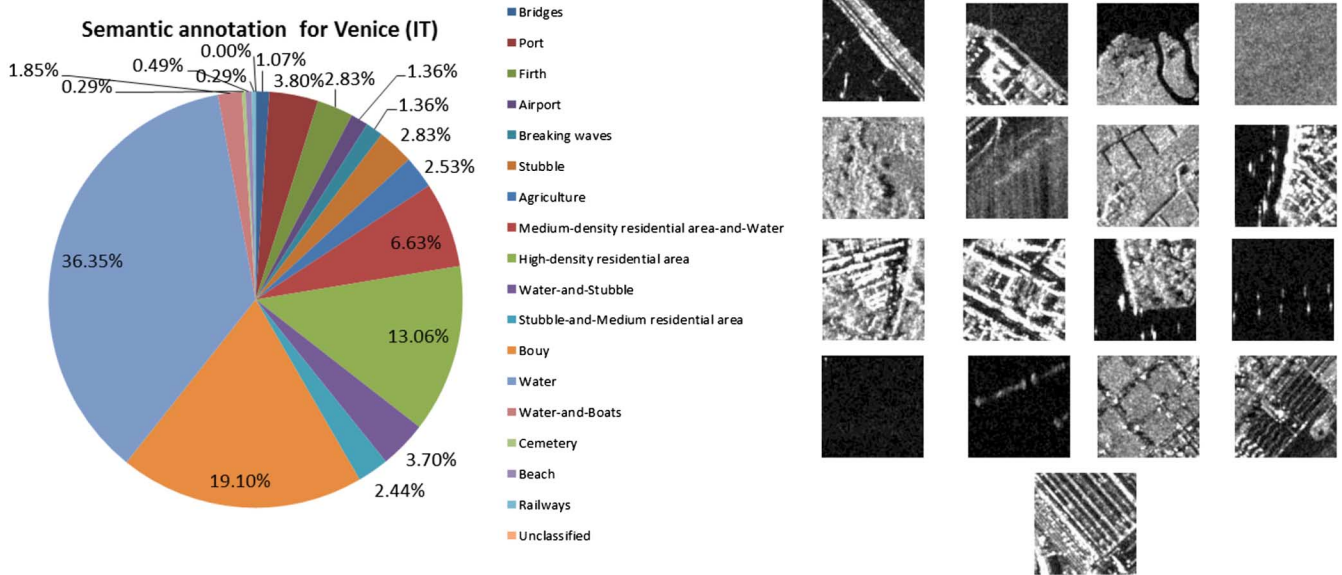
Fig. 11. Percentage of patches per semantic category for Venice (on the left), and a typical patch per category (on the right).
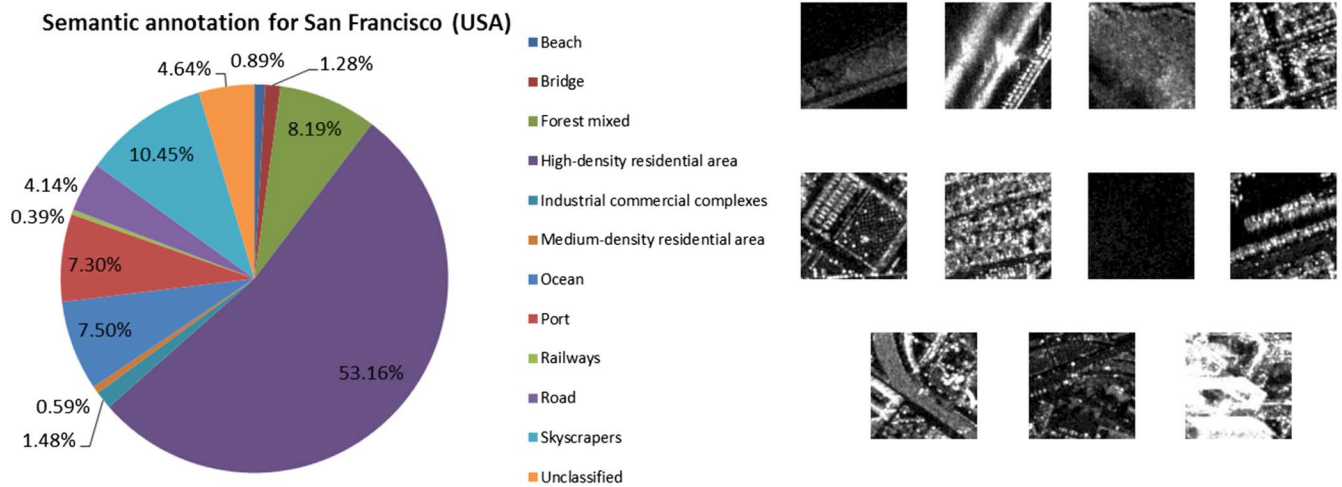

Fig. 12. Percentage of patches per semantic category for San Francisco (USA) (on the left), and a typical patch per category (on the right).

2) Query by extracted features: Expert users can statistically analyze extracted features from each patch.

3) Query by examples: All users can find images (patches) with similar content based on their feature vectors contained in the database.

4) Query by semantics: Expert users can use a semantic label in order to perform a query.

Several detailed examples of queries by semantics are given in Section VII [21].

V. ELEMENTS OF IMAGE CONTENT ONTOLOGIES

Geospatial ontologies appeared first in the fields of cartography and geographic information systems that have to cope with a lot of interrelated information. For satellite image analysis, ontologies are a rather new topic, and we try to use them for SAR images that have been studied in less detail until now.

Early research in the field of ontologies and semantics with respect to the representation and use of ontologies for
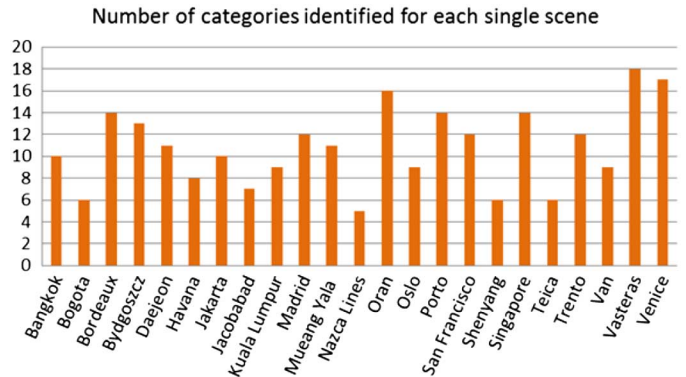

Fig. 13. Number of retrieved categories for each scene when annotated separately.

geographic features has been summarized by Smith and Mark [22]. Some years later, the *SemQuery* approach was published [23] supporting visual queries for heterogeneous features of
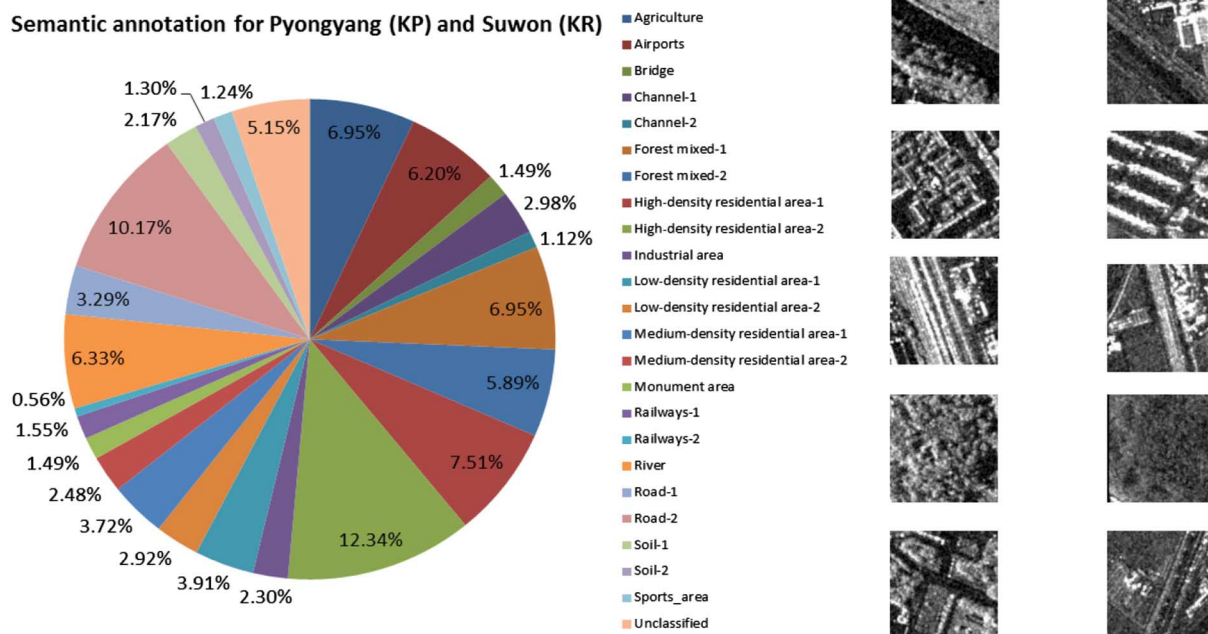
Fig. 14. Percentage of patches for Pyongyang and Suwon. On the right from top to bottom appear the categories: *river, high-density residential area, railways, forest*, and *road* [the left column corresponds to Pyongyang (KP) and the right column corresponds to Suwon (KR)].
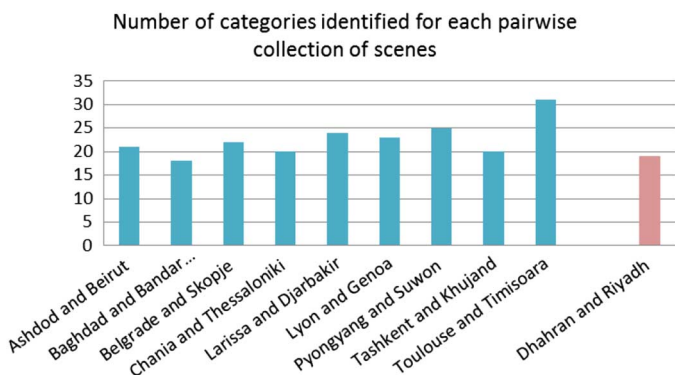

Fig. 15. Number of retrieved categories when two urban areas from different source regions are grouped together in a collection.

images. It represents a semantics-based clustering approach for the classification of database images. The proposed clustering approach can split an image into different clusters based on its features (e.g., texture, color, or object shape). SemQuery successfully combines semantics-based and template-based clustering. In addition, a query strategy was implemented to support visual queries for heterogeneous features. Note that, since the features are generated with different methods, this may require different similarity measures.

Recently, ontologies and semantic issues began to be studied in more detail and several publications have been presented during the International Geoscience and Remote Sensing Symposium (IGARSS 2012) held in Munich, Germany. Some representative papers will be described below.

A hierarchical ontology for land use data (covering Taiwan) has been published by Kuo and Hong [24]. Here the ontology represents an important rule set for GIS applications describing the hierarchical semantics relationships between different land use types; it includes the definition of a land use vocabulary that helps users to infer land use types by reasoning rules. In order to provide different granularities, the proposed method has three hierarchical levels: the first level subdivides the land use into nine basic categories (*agriculture, forest, hydrology, traffic, building architecture, public use, leisure, mineral/salt*, and *other land use categories*); the second level has 41 categories and the third level contains 103 categories that are derived from the midlevel to provide more detailed differences.

Web service technologies have shown great promise for rapid feature discovery from large volumes of remote sensing images. A corresponding ontology approach for the detection of complex features is described in [25]. A complex feature is spatially composed of elementary features and the spatial relationships among elementary features can be used further to find complex features. An application case is geospatial image understanding and feature extraction. While conventional approaches can extract elementary geospatial features from images such as *buildings, fences, bridges*, and *railways*, geoprocessing services within a Web service environment can be used to implement "ontologies" that are facilitating the expression and reuse of problem-solving knowledge for feature discovery. An alternative approach to develop a semantic feature catalog service that combines the Open Geospatial Consortium Catalogue Service with distributed Web Feature Services was published in [26]. It includes a feature type semantic catalog, while feature ontologies are connected through relations or quantities.

Still another approach is to reduce overfitting when ranking high-resolution satellite image by domain semantics [27]. A high-resolution image database containing 877 geospatial
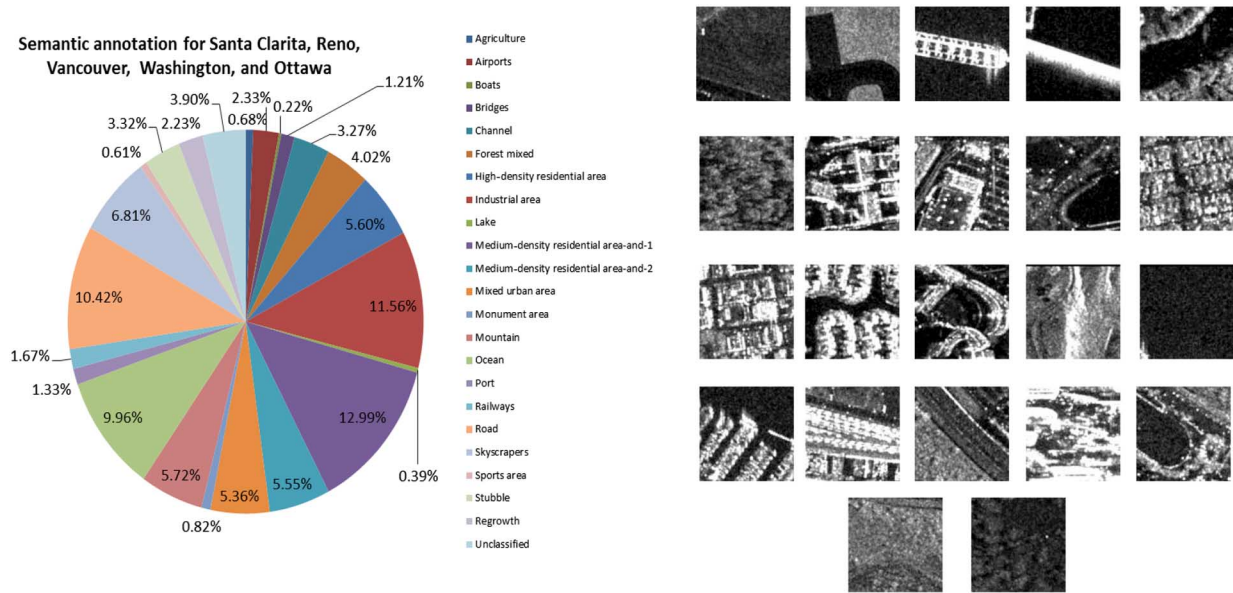
Fig. 16. Number of patches per semantic category for North America part II grouping five cities from Canada and the USA. On the right side typical patches of each category are shown.
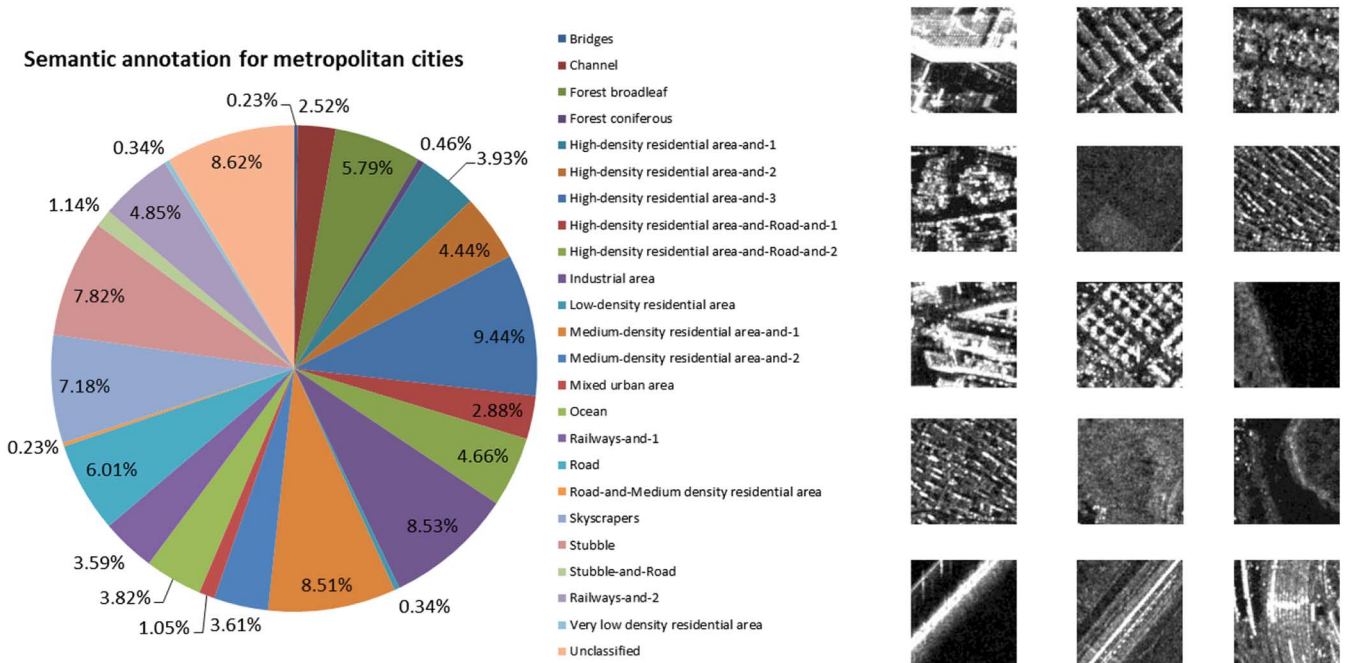

Fig. 17. Number of patches per semantic category for the metropolitan cities of Dubai, Los Angeles, Shenyang (CN), Tokyo, and Washington (USA). On the right side typical patches of each category are shown.

patches covering the Missouri area was generated after tiling the original multispectral images. Each patch results in a feature vector with 227 elements that includes histogram information, cooccurrence texture, linear structures, and aggregate object features. Then, one of the following categories was used to semantically annotate the patches: *construction, isolated road, industrial, grassland, cropland, commercial, residential,* and *forest*. From the total number of tiled patches, 154 patches were mapped into multiple semantics while the majority has only one semantic annotation.

If the semantic annotation of a patch can only be derived by support from neighboring patches, one needs an approach as described in [28] where groups of similar patches are compared by exploiting their feature vector compressibility. The method was tested with patches of 64×64 pixels and 128×128 pixels. For the first patch size, the categories *sea*, *seashore*, *farming area*, *car parking*, *golf course*, and *green area* could be identified, while for the second patch size, the categories *sea, seashore, farming area, seashore resort, car parking, golf course*, and *green area* could be retrieved. The difference
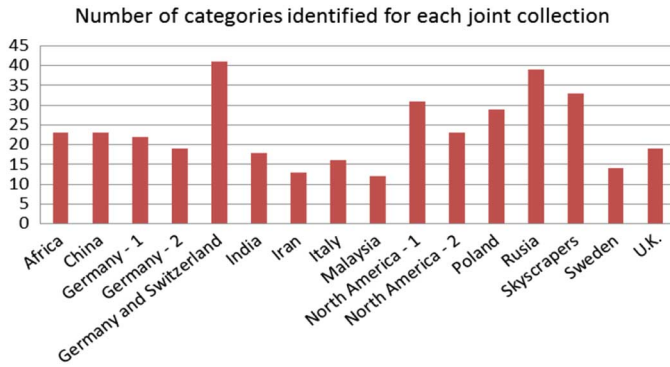
Fig. 18. Number of retrieved categories for joint collections: Grouping based on the geographical source regions or based on the architectural characteristics of the urban areas.

between the two cases is that for the latter one the categories *farming* and *seashore resort* are merged together.

VI. Image Content Taxonomies

In this section, we present the categories that can be retrieved from our TerraSAR-X data set. We explain the general annotation of patches and continue by defining ontologies/taxonomies for our data set using the methodology presented in this paper. The reason for this is that more and more researchers use ontologies to model the terminologies of different domains and to find a common representation of them.

There are many classification schemes that are proposed in literature but these are not satisfying our needs. The results of [29] show that the land cover categories retrieved for South America are not compatible with the categories identified in North America. Another important issue that is discussed in [29] is that for the nonurban areas the annotation needs to take into account the seasonal land cover of the investigated area, while our goal is to create a large annotated TerraSAR-X data set that covers many urban regions over the world and to find a criterion to help users to group the scenes. Once the categories are identified, the next step is to semantically annotate them using the hierarchical annotation scheme shown in Fig. 5 [30], [31]. Our solution is a two-level annotation scheme: Level 1 creates general information about the content of a patch, whereas level 2 details the general information of level 1.

This scheme provides semantic content annotation and is embedded within a hierarchical multilevel approach allowing us to integrate detailed annotations within higher level annotations. The multilevel approach seems to be a mandatory step for the annotation of high-resolution SAR images where pixel-based information from a local neighborhood is not sufficient for context recognition. Note, however, that a multilevel approach may lead to ambiguous higher level annotations.

In this vein, we selected a scene from our data set (Venice, Italy, Fig. 6a) and we compared the annotation categories from our methodology with the current CORINE Land Cover (CLC) nomenclature [32] and generated a map [33] with both

categories projected on Google Maps [34]. The results are shown in Figs. 6b and 7, where on the left side, we have the map with categories projected on an image of Venice, and on the right side we have the color legend of each category. For each case, we took the most detailed level of annotation and we reached a number of 17 categories for our annotation method and 10 categories in the case of CLC. We observed that our proposed categories are more detailed than the CLC nomenclature. For instance, *bridge, buoy,* and *sea* are included in *marine waters—coastal lagoons*, etc. In order to evaluate the accuracy of our proposed semantic annotation, the precision/recall metric was computed for each category and the overall average of this metric was 88.85% for precision and 73.37% for recall considering the Venice image.

When we consider the content of our entire data set, we have to know more about the diversity of semantic categories that can be retrieved from TerraSAR-X images and the obtained results. This will be described in the Section IV-A and IV-B together with some additional remarks.

A. Annotation Tests

The total number of patches obtained after tiling all available scenes is about 110 000. Then a feature vector and a semantic annotation are available for each patch [30]. Fig. 8 presents some typical categories extracted from images taken over Bagdad (IQ), Bandar Imam Khomeini (IR), Binhai (CN), Genoa (IT), Nazca Lines (PE), Singapore (SG), Trento (IT), and Tucson (USA).

For each category, we display the semantic annotation and a typical quick-look image. In contrast, Figs. 9 and 10 depict a set of patches taken from different regions all over the world [Belgaum (IN), Dubai (AE), Havana (CU), Lodz (PL), Oran (DZ), Toulouse (FR), Torun (PL), Vasteras (SE), and Venice (IT)] together with their associated semantic meanings [33]. What can be seen in Fig. 9 is that these patches contain objects with identical semantics. From a semantics point of view, these patches should be grouped into the same category but, unfortunately, this does not work as the geographical and architectural characteristics of the patches differ too much. In Fig. 10, we identified patches with similar architectural characteristics that can be grouped together. In the first example, we selected *Buildings* from German speaking countries (Munich and Kiel, Germany, and Basel, Switzerland) where the architecture of the buildings is similar. In the second example, we collected a few cosmopolitan cities (Tokyo, Dubai, and Los Angeles) that have *Skyscrapers* in common.

B. Annotation Results

We ask now under which criteria scenes shall be grouped together and how many categories exist in each scene or grouped scenes. In order to answer these questions the full test data set containing hundreds of scenes was split into collections in order to find the number of identifiable categories and to annotate them. We focused on urban areas because they are less
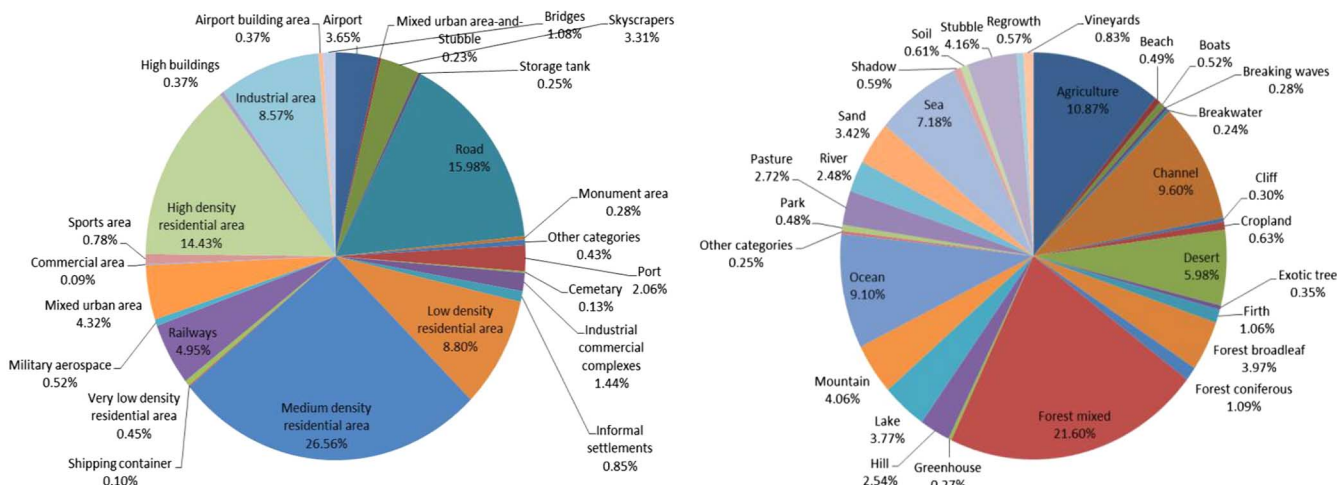
Thisarticlehasbeenacceptedforinclusioninafutureissueofthisjournal.Contentisfinalaspresented,withtheexceptionofpagination.

12   IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING

Fig. 19. Results of a query: *"How many urban and nonurban categories can be identified in our data acquired in 2013?"* On the left side are shown the urban categories, and on the right side the nonurban categories.

dependent on seasonal effects. In future, we will also consider nonurban areas.

1) Grouping of Scenes: In order to get an idea about how many categories can be retrieved for each *continent/country/city* we selected the following cities in different continents:

Africa: Oran, Algeria.

Asia: Bangkok, Thailand / Daejeon, South Korea / Jacobabad, Pakistan / Jakarta, Indonesia / Kuala Lumpur, Malaysia / Mueang Yala, Thailand / Shenyang, China / Singapore, Singapore / Van, Turkey.

Europe: Bordeaux, France / Bydgoszcz, Poland / Madrid, Spain / Oslo, Norway / Porto, Portugal / Teica, Romania / Trento, Italy / Vasteras, Sweden / Venice, Italy.

South and North America: Bogota, Colombia / Havana, Cuba / Nazca Lines, Peru / San Francisco, USA.

a) Results: Figs. 11 and 12 show the patch class distributions for Venice, Italy, and San Francisco, USA, and one typical patch for each annotated category. Similar results are obtained for other scenes and summarized in Fig. 13.

b) Remarks: For this type of grouping, scenes that cover Asia, Europe, and North America contain more than 10 categories per scene (except for Kuala Lumpur, Malaysia / Oslo, Norway and Van, Turkey with 9 categories). Bogota, Columbia / Jacobabad, Pakistan / Nazca Lines, Peru / Shenyang, China and Teica, Romania contain the least number of categories (4–7 categories per scene); for Nazca Lines and Teica 90% of these categories are nonurban categories.

Fig. 13 shows the number of categories gained by a separate annotation of each scene. The annotation used our hierarchical semantic annotation method.

2) Pairwise Grouping of Scenes From Different Source Regions: After we obtained an idea about how many categories can be identified for individual urban areas, we tried to see whether different urban scenes can be grouped together. For this investigation, pairs of scenes with different geographical source regions (e.g., countries) were grouped together. The scenes cover the following areas.
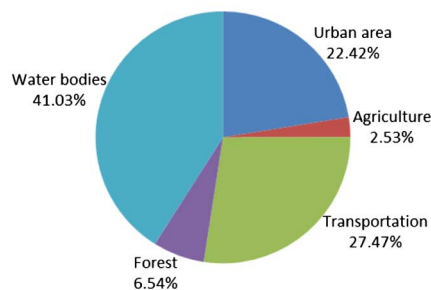

Fig. 20. Results of a query: *"How much level 1 urban area, agriculture, transportation, forest, and water bodies can be identified for the given scene?"*.

Asia: Pyongyang, North Korea and Suwon, South Korea / Tashkent, Uzbekistan and Khujand, Tajikistan.

Europe: Belgrade, Serbia and Skopje, Macedonia / Chania, Greece and Thessaloniki, Greece / Larissa, Greece and Djarbakir, Turkey / Lyon, France and Genoa, Italy / Toulouse, France and Timisoara, Romania.

Middle East: Ashdod, Israel and Beirut, Lebanon / Baghdad, Iraq and Bandar Imam Khomeini, Iran.

a) Results: We show as an example the pair of scenes of Pyongyang and Suwon (see Fig. 14). The results of this type of grouping are presented in Fig. 15. One can see the distribution of patches per category, and a typical patch of each category.

b) Remarks: The results of pairwise grouping show that the number of retrieved categories lie between 20 categories per collection (Chania and Thessaloniki) and 25 categories per collection (Pyongyang and Suwon) with one exception (Baghdad, Iraq, and Bandar Imam Khomeini, Iran) where 18 categories could be found.

Fig. 15 details the number of identified categories. The collection scenes stem from different countries (except for Dhahran and Riyadh, Saudi Arabia marked in pink in Fig. 15).
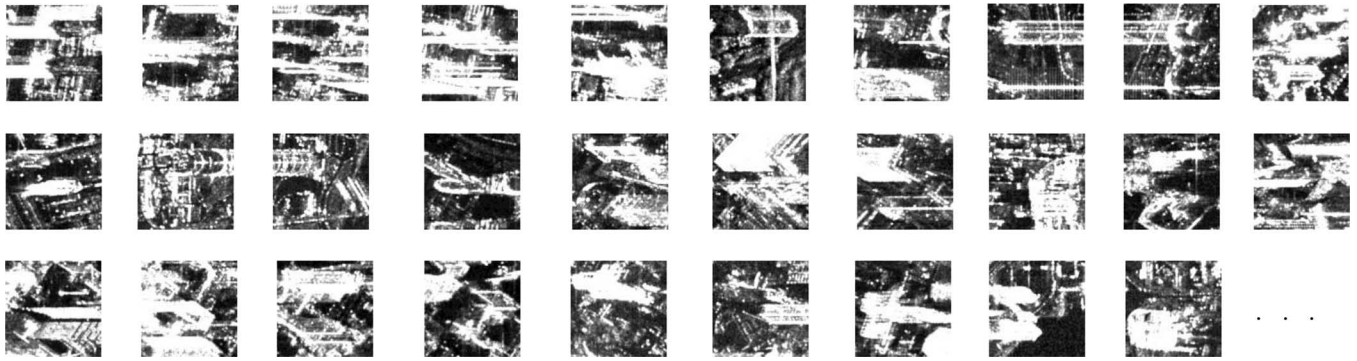
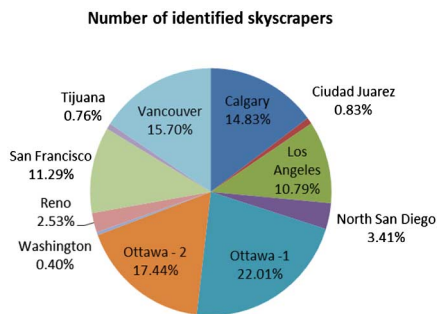Fig. 21. Results of a query: *"Find skyscrapers in our database."*.

Fig. 22. Percentage of *skyscrapers* retrieved for several given North American cities (Canada, USA).

If different source regions are combined, the identified categories are completely separated. Even for Dharan and Riyad the retrieved categories are separated which leads us to the conclusion that not only the geographical source region is important but also the architectural characteristics of each city.

3) Joint Grouping of Multiple Scenes With Similar Characteristics: Now we perform the grouping based on similar source regions and/or architectural characteristics of the target areas.

In the first step, the collections are based on the source regions of the scenes. In the second step, the scenes are grouped based on the architectural characteristics of the cities.

Step 1) Geographical grouping:
   Africa: Port Elizabeth, Bulawayo, Abuja, and Lomé.
   China: Binhai, Jinan, and Dalian.
   India: Belgaum, Pune, and Vadodara.
   Iran: Bandar-e'Abbas and Mahabad.
   North America part I: Calgary, Ciudad Juarez, North San Diego, Ottawa, Poway, South San Diego, Sun Lakes, Tijuana, and Tucson (taken in 2012).
   North America part II: Santa Clarita, Reno, Vancouver, Washington DC, and Ottawa (taken in 2013).
   Poland: Czestochowa, Lodz, and Torun.
Step 2) Architectural grouping:
   German speaking countries/cities: Bonn, Berlin, Cologne, Kiel, Munich, Oldenburg, and Basel.
   Germany part I: Bremen and Berlin (taken in 2012).
   Germany part II: Karlsruhe, Lindau, Mannheim, and Stuttgart (taken in 2013).

Italy: Naples and Taranto.
Malaysia: Alor Setar and Seremban.
Russia: Krutorozhino, center of Moscow, northern part of Moscow, southern part of Moscow, Perm, Rostov on Don, and Tula.
Sweden: Stockholm and Uppsala.
United Kingdom: London, Plymouth, and Portsmouth
Skyscrapers in metropolitan cities: Dubai, Los Angeles, Shenyang, Tokyo, and Washington.

a) Results: For visual representation, we show the results for different collections and their grouping. Figs. 16 and 17 show the distribution of patches in North America and of metropolitan cities.

b) Remarks: The grouping of multiple scenes shows that the number of retrieved categories ranges from 12 (when two scenes from Asia are annotated together) to about 30 (when nine scenes from North America are grouped together). We also need to distinguish between groupings based on the geographical source regions of the scenes and groupings based on the architectural characteristics of urban areas. In both cases, there are some categories that are common to all grouped scenes (e.g., *industrial areas, roads, railways*), others coincide only partially (e.g., *skyscrapers, medium-density residential areas,* and *low-density residential areas*), and some categories are rather unique (e.g., *mixed urban areas*).

Fig. 18 shows the number of identified categories for each collection when the grouping was made based on the geographical or the architectural characteristics of the scenes.

VII. Querying Examples

From the various query types that have been implemented, we select three example queries referring to annotation semantics. The first example shall find how many urban and nonurban categories are retrieved and annotated from our data set acquired in 2013 (70 scenes). Fig. 19 shows the semantic meaning of each urban or nonurban category and the percentage (%) of each category.

A second example is to find for a selected scene (e.g., Venice, Italy) all the level 1 categories that could be retrieved and annotated. The result of such a query (see Fig. 20) can give us information about the land cover percentages of *urban*

TABLE I
IMPROVED CLASSIFICATION ACCURACIES BY ROTATION INVARIANT FEATURE VECTORS

	Conventional Gabor filtering		Rotation invariant Gabor filtering	
	Precision (%)	Recall (%)	Precision (%)	Recall (%)
Forest broadleaf	82.71	67.48	68.00	78.60
Channel	25.75	93.75	31.72	93.75
Railway	44.75	47.22	27.38	63.89
Agriculture	35.85	92.68	19.51	97.56
High density residential area	69.05	69.88	69.12	90.36
Medium density residential area	20.21	46.34	16.03	62.19
Industrial area	22.71	25.74	40.00	30.26
Sports area	33.34	50.00	24.93	55.56

areas, transportation, water bodies, forest, agriculture, or *bare ground*.

A third example is to find all patches within our database that contain *skyscrapers* as a semantic label. As a result of this query, several thousand patches were retrieved. Fig. 21 shows a few examples of these patches. The results of the third query can also be displayed as percentages of *skyscrapers* retrieved for several given North American cities (see Fig. 22).

VIII. CURRENT DEVELOPMENTS

The most important current development is the validation of an enhanced rotation invariant feature extractor. It allows us to compare linear features independently of their actual orientation [37]. This can be accomplished rather simply by rearranging the elements of the extracted feature vectors, where we sort the Gabor filter bank results by magnitude.

Table I shows the gain in classification obtained by this new technique when applied to a typical mixed content scene.

Note that the improved recall results are mainly evident for urban categories. This is also due to an enlarged set of candidate elements becoming available after rotation realignment of the feature vectors.

IX. CONCLUSION

In this paper, we discussed the semantic categories that can be retrieved from TerraSAR-X data and we generated a semantic catalog for satellite images [21]. To our knowledge, this is the first time that such a large set of SAR images has been annotated covering specific signatures from all over the world. In addition, we defined a set of ontologies for high-resolution SAR images based on CORINE Land Cover and Urban Atlas ontologies.

The main contributions of this paper are as follows.
1) The results rely on a large database of high-resolution SAR data where scenes can be grouped geographically or architecturally.
2) The classification has a significant impact on the identification of urban characteristics; in future, this study will be extended to landscape scenes.
3) The number of categories retrieved from each scene can help us to better understand the content of SAR data [36]. In total, we identified 850 categories from which about 75 are independent.
4) The generated taxonomies generated have been used to design a semantic catalog [21].
5) The semantic annotation can be improved by a systematic definition of scene collections with increasing heterogeneity and additional information coming from cartographic or other sources such as [35].

The annotated data can be considered as a reference data set and can be used for different tasks such as feature extraction, machine learning, query building, and validation since we created more than 100 000 image patches, generated more than 40 collections from which we identified more than 850 categories.

The impact of increasing the number of scenes (*volume* of data) is that the diversity (*variety* of data) of the identified categories will become higher. As a consequence, more categories can be identified. Our experiments may result in a number of new applications in the fields of urban cartography, forest and water level monitoring, crisis and disaster interpretation, and satellite ground segment architectures. All components of the proposed approach and the hierarchical annotation scheme are now being integrated with a satellite ground segment.

As a continuation of our work, we are going to 1) to annotate another 200 scenes with new target categories (e.g., *landscapes, industrial sites, military facilities, mines, vegetation, agriculture, polar ice*, and *ocean water*) [21]); 2) extend our

hierarchical annotation scheme to three levels; and 3) design and implement an advanced query builder capable of handling the new three level annotation scheme.

REFERENCES

[1] Big Data Space Conference, Frascati, Italy, 2013 [Online]. Available: http://www.congrexprojects.com/2013-events/13c10/announcement

[2] D. Laney. (2001). *3D Data Management Controlling Data Volume Velocity and Variety* [Online]. Available: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

[3] M. Wolfmueller *et al.*, "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.

[4] K. Molch, M. Wegner, and E. Mikusch, "DLR long term data preservation experience and approach," in *Proc. 2nd Long Term Data Preserv. Workshop*, Frascati, Italy, May 2011, pp. 1–12.

[5] A. Popescu, T. Tsikrika, and J. Kludas. (2010). *Overview of the Wikipedia Retrieval Task at ImageCLEF 2010* [Online]. Available: http://www.imageclef.org/node/118/revisions/1691/view

[6] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, 2010.

[7] IMAGENET. (2011). *Large Scale Visual Recognition Challenge 2011 (ILSVRC)* [Online]. Available: http://www.image-net.org/challenges/LSVRC/2011/index

[8] A. Torralba, B. C. Russell, and J. Yuen, "LabelMe: Online image annotation and applications," *Proc. IEEE*, vol. 98, no. 8, pp. 1467–1484, Aug. 2010.

[9] C.-R. Shyu *et al.*, "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.

[10] P. Birjandi, "Modelling, extraction and description of intrinsic cues of high resolution satellite images: independent component analysis based approaches," Ph.D. dissertation, 2011 [Online]. Available: https://tel.archives-ouvertes.fr/pastel-00677956/file/Thesis_Presentation_final.pdf

[11] A. Popescu, 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.

[12] C. O. Dumitru and M. Datcu, "Information content of very high resolution SAR images: Study of feature extraction and imaging parameters," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 8, pp. 4591–4610, Aug. 2013.

[13] S. Voigt *et al.*, "Satellite image analysis for disaster and crisis management support," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1520–1528, Jun. 2007.

[14] J.-B. Henry, K. Fellah, S. Clandillon, B. Allenbach, and P. De Fraipont, "Earth observation and case-based systems for flood risk management," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Toronto, ON, Canada, Jun. 2002, vol. 3, pp. 1496–1498.

[15] TerraSAR-X Basic Products Specification Document, Issue: 1.6, TX-GS-DD-3302, 2009 [Online]. Available: http://sss.terrasar-x.dlr.de/

[16] EOWEB portal, 2014 [Online]. Available: https://centaurus.caf.dlr.de:8443/eoweb-ng/template/default/welcome/entryPage.vm

[17] S. Cui, C. O. Dumitru, and M. Datcu, "Semantic annotation in earth observation based on active learning," *Int. J. Image Data Fusion*, vol. 5, no. 2, pp. 152–174, Apr. 2014.

[18] 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.

[19] C. O. Dumitru, J. Singh, and M. Datcu, "Selection of relevant features and TerraSAR-X products for classification of high resolution SAR images," in *Proc. 9th Eur. Conf. Synth. Aperture Radar (EUSAR)*, Nuremberg, Germany, Apr. 2012, pp. 243–246.

[20] D. Espinoza-Molina and M. Datcu, "Earth-observation image retrieval based on content, semantics and metadata," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 11, pp. 5145–5159, Nov. 2013.

[21] (2014). *EOLib—Earth Observation Image Librarian Project* [Online]. Available: http://wiki.services.eoportal.org/tiki-index.php?page=EOLib

[22] B. Smith and D. M. Mark, "Ontology and geographic kinds," in *Proc. Int. Symp. Spat. Data Handling*, Vancouver, BC, Canada, Jul. 1998, pp. 308–320.

[23] 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./Oct. 2002.

[24] C.-L. Kuo and J.-H. Hong, "Hierarchical ontology development and semantics retrieval for land use data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Munich, Germany, Jul. 2012, pp. 6051–6054.

[25] L. Di, P. Yue, and Z. Sun, "Ontology-supported complex feature discovery in a web service environment," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Munich, Germany, Jul. 2012, pp. 2887–2890.

[26] P. Yue, L. Di, Y. Wei, and W. Han, "Semantic feature catalogue service," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Munich, Germany, Jul. 2012, pp. 2914–2917.

[27] A. Barb and C.-R. Shyu, "Associative semantic ranking of satellite images using path finder network scaling ensemble methods," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Munich, Germany, Jul. 2012, pp. 5289–5292.

[28] M. Nakajima, T. Watanabe, and H. Koga, "Compression-based semantic-sensitive image segmentation: PRDC-SSIS," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Munich, Germany, Jul. 2012, pp. 4303–4306.

[29] S. Durbha, R. King, and N. Younan, "Ontology matching approach for geospatial domain interoperability," in *Proc. IIM*, Ispra, Italy, Apr. 2011, pp. 75–80 [Online]. Available: http://rssportal.esa.int/tiki-index.php?page=ESA-EUSC-JRC-2011

[30] J. Anderson, E. Hardy, J. Roach, and R. Witmer. (1976). *A Land Use and Land Cover Classification System for Use With Remote Sensor Data* [Online]. Available: http://landcover.usgs.gov/pdf/anderson.pdf

[31] P. Blanchart and M. Datcu, "A semi-supervised algorithm for auto-annotation and unknown structures discovery in satellite image databases," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 3, no. 4, pp. 698–717, Dec. 2010.

[32] CORINE Land Cover [Online]. Available: http://sia.eionet.europa.eu/CLC2000/docs. CORINE Land Cover—Categories [Online]. Available: http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2000-clc2000-100-m-version-9-2007/corine-land-cover-2000-classes-and-rgb-color-codes/clc2000legend.xls

[33] TELEIOS—Virtual Observatory Infrastructure for Earth Observation Data, 2013 [Online]. Available: http://www.earthobservatory.eu

[34] Imagery ©2013 Google, Venice, Italy.

[35] (2013). *LUCAS—A Multi-Purpose Land Use Survey* [Online]. Available: http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/LUCAS_%E2%80%94_a_multi-purpose_land_use_survey

[36] C. O. Dumitru and M. Datcu, "Study and assessment of selected primitive features behaviour for SAR image description," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Munich, Germany, Jul. 2012, pp. 3596–3599.

[37] S. Newsam, L. Wang, S. Bhagavathy, and B. Manjunath, "Using texture to analyze and manage large collections of remote sensed image and video data," *Appl. Opt.*, vol. 43, pp. 210–217, 2004.

Corneliu Octavian Dumitru received the B.S. and M.S. degrees in applied electronics from the Faculty of Electronics, Telecommunications and Information Technology, Politehnica University Bucharest, Bucharest, Romania, in 2001 and 2002, respectively, the Ph.D. degree in engineering from Politehnica University Bucharest in 2006, and the Ph.D. degree in telecommunications from Pierre and Marie Curie University, Paris, France, in 2010.

At the Politehnica University, he had a teaching activity as a Lecturer, delivering lectures and seminars and supervising laboratory works in the fields of information and estimation theory, communication theory, and signal processing. From 2005 to 2006 and in 2008, he was a Coordinator for two national grants awarded by the Romanian Ministry of Education and Research. Since 2010, he has been a Scientist with the Remote Sensing Technology Institute (IMF), Earth Observation Center (EOC), German Aerospace Center (DLR), Oberpfaffenhofen, Germany. Recently, he is involved in several projects in the frame of the European Space Agency and FP7 Programmes for information extraction, taxonomies, and data mining using remote sensing imagery. His research interests include stochastic process information, model-based sequence recognition and understanding, basics of man–machine communication, information managing, and retrieval in extended databases.

Shiyong Cui received the M.S. degree in photogrammetry and remote sensing from Chinese Academy of Surveying and Mapping, Beijing, China, in 2009. Currently, he is pursuing the PhD. degree in electronic engineering at the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, Germany.

His main research interests include statistical image modeling, computer vision, and machine learning.

Gottfried Schwarz received the Diploma degree in electrical engineering from the Technical University of Munich, Munich, Germany.

For many years, he has been involved in a number of national and international space projects with the German Aerospace Center (DLR), Munich, Germany; among them were deep space missions and earth observation missions. In particular, he has been engaged in the design of deep space instruments from initial engineering studies to detailed design work, modeling of instrument performance, instrument assembly and testing, real-time experiment control, instrument check-out and calibration, data verification and validation as well as data processing and scientific data analysis. Besides instrument-related aspects, he has also many years of experience in the processing and analysis of various instrument data within ground segments, in particular of optical and SAR remote sensing data, in the interpretation of geophysical data with emphasis on retrieval algorithms with forward modeling, inversion techniques, and data mining. Special experience in signal processing resulted from engagement in image data compression and feature analysis together with performance analysis of image classification.

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. In 1999, he received the title "Habilitation à diriger des recherches" in computer science from Université 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), Oberpfaffenhofen, Germany. Currently, he is a Senior Scientist and Image Analysis Research Group Leader with the Remote Sensing Technology Institute of DLR, Oberpfaffenhofen, Germany. Since 2011, he has also been leading the Immersive Visual Information Mining Research Laboratory, the Munich Aerospace Faculty, Munich, Germany, and is the Director of the Research Center for Spatial Information, UPB. He has held Visiting Professor appointments with the University of Oviedo, Oviedo, Spain, the University Louis Pasteur and the International Space University in Strasbourg, France, the University of Siegen, Siegen, Germany, the University of Camerino, 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 led the Competence Centre on Information Extraction and Image Understanding for Earth Observation, ParisTech, Telecom Paris, Paris, France, a collaboration of DLR with the French Space Agency (CNES). He has been a Professor holder of the DLR–CNES Chair with ParisTech, Telecom Paris. He is the author of more than 200 scientific publications, among them are about 50 journal papers and a book on number theory. His research interests include information and complexity theory, stochastic processes, Bayesian inference, and image information mining (IIM).

Prof. Datcu is a member of the European Image Information Mining Coordination Group and of the Data Archiving and Distribution Technical Committee of the IEEE Geoscience and Remote Sensing Society.