

Diversity of Settlement Categories in Very High Resolution SAR Images

Corneliu Octavian Dumitru

Remote Sensing Technology Institute (IMF)
German Aerospace Center (DLR)
Oberpfaffenhofen – Wessling, Germany
corneliu.dumitru@dlr.de

Mihai Datcu

Remote Sensing Technology Institute (IMF)
German Aerospace Center (DLR)
Oberpfaffenhofen – Wessling, Germany
mihai.datcu@dlr.de

Abstract—In this paper we propose to identify the number of urban categories that can be retrieved from very high resolution TerraSAR-X data. For this task a semi-automated procedure was built in order to search in large Earth Observation database for similar sub-images (i.e. patches) and group them in the same semantic category. The dataset consists of 39 scenes over the world: 18 scenes in Asia, 15 scenes in Europe, 5 scenes in North and South America, and 1 scene in Africa. These scenes are grouped in three different types of collections (a total of 16 collections are generated) and the semantic categories in each collection are retrieved. A total of 320 categories are identified and those that contain man-made structures in the urban area are kept and the rest of categories (e.g., agriculture, water, etc.) are discarded. These categories represent between 50% and 65% from the total number of retrieved categories in each collection.

I. INTRODUCTION

Experiments reported in literature by different authors' show that is possible to detect a 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 bounds of categories which are defined by a user.

The GeoIRIS system [1] is capable to retrieve a large volume of multispectral dataset automatically after pre-processing and indexing of satellite images. The images were tiled in patches obtaining 70,824 patches where about 531,208 objects were identified.

For different urban scenes acquired in Las Vegas (USA), Venice (Italy), Gizah (Egypt), and Gauting (Germany) a large variety of objects and structures were identified in [2]. The very high resolution Spotlight TerraSAR-X images were tiled in patches and from about 7,000 patches 30 semantic categories (e.g., bridges, skyscrapers, industrial area, buildings, port, etc.) were identified.

Following the same idea of [2] in [3] the very high resolution Spotlight TerraSAR-X images from different areas over the world were used and 20 categories were identified during the evaluation from 2,000 patches. Some examples of these semantic categories are: urban coast line, aircraft stands, skyscrapers, industrial complex, urban, etc.

In this paper, we try to progress compared with the state-of-the-art by providing the number of categories that can be extracted from different areas and to find the most appropriate

semantic meaning for the retrieved categories. In order to understand the huge diversity of categories that exist, the scenes are grouped in collections based on different criteria which are explained further in the paper.

The paper structure is the following. Section 2 briefly describes the methodology used to semantically annotate the categories extracted from very high resolution TerraSAR-X data. The first part of Section 3 presents the configuration of TerraSAR-X product used in order to build the *dataset*, while the second part explains how to group these scenes in collections for a better retrieve of these categories in different geographical areas. Section 4 provides examples of urban categories retrieved in our *dataset*. The paper ends with conclusions in Section 5.

II. METHODOLOGY

The methodology that is used in order to semantically annotate the *dataset* is presented [4] further as a pseudo-code.

- 1) *Select the TerraSAR-X products for evaluation.*
Based on [6] the optim basic product (level1b) is Multi-look Ground range Detected (MGD), radiometrically enhanced (RE), high resolution Spotlight mode (HS).
- 2) *Group the scenes (TerraSAR-X products) in collections.*
In section III.B is presented how to group these scenes in collections.
- 3) *Tile the grouped scenes into small images (called patches).*
The size of these pathes depend by the resolution and the pixel specing of the product. For MGD-RE products the optimal patch size is 160 by 160 pixels [5].
- 4) *Generate the quick-looks of the tile patches, without rescalling the data, needed for visualisation.*
- 5) *Compute the primitive features for each patch.*
The quadrature mirror filters (QMFS) is used as primitive features and comparative results between four very well know features are presented in [6].
- 6) *Select a classifier in order to group the features of each patch in categories.*

To solve this, we built a tool based on support vector machine with relevance feedback (SVM-RF) [7]. In the last years the SVM is one of the most used classifier

having the capacity to perform very accurate classification with a very limited numbers of examples. The GUI of the tool allows to rank automatically suggested images (patches) which are expected to be grouped in the categories of relevance. Visual supported ranking allow enhancing the quality of search results by giving positive and negative examples.

7) *Annotate these categories and find an appropriate semantic meaning for each category.*

In order to annotate these categories, we proposed for each categorie an appropriate semantic meaning [8] and we considered only one level of annotation. We used as visual support the Google Earth as ground truth.

III. EARTH OBSERVATION TERRASAR-X DATASET

For the evaluation, we select from the huge diversity of modes, types, and geometric configurations that are available for EO TerraSAR-X [9], the following configuration [4]-[5]: (a) HS mode because is the most popular mode; (b) MGD because are not affected by the interpolation and are more suitable for image processing, feature extraction, and annotation than the geo-coded products; and (c) RE geometric configuration because is optimal for feature extraction and the relation between the SNR (Signal-to-Noise Ratio) and resolution is the best compare with spatially enhanced (SE).

The pixel spacing is equal to 1.25 pixels and the resolution about 2.9 meters. The TerraSAR-X products are with single polarization (HH), ascending orbit looking, and incidence angle between 32° and 42° .

A. Dataset

The *dataset* contains 39 scenes over the world and the locations are: Anshan (China), Bangkok (Thailand), Basel (Switzerland), Belgaum (India), Bogota (Columbia), Bonn (Germany), Berlin (Germany), Cologne (Germany), Czestochowa (Poland), Djarbakir (Turkey), Dubai (United Arab Emirates), Havana (Cuba), Khujand (Tajikistan), Kiel (Germany), Krutorozhino (Russia), Larissa (Greece), Lodz (Poland), Los Angeles (USA), Moscow – center part (Russia), Moscow – north part (Russia), Moscow – south part (Russia), Munich (Germany), Nazca Lines (Peru), Oldenburg (Germany), Oran (Algeria), Perm (Russia), Pune (India), Rostov on Don (Russia), Shenyang (China), Tashkent (Uzbekistan), Timisoara (Romania), Tokyo (Japan), Toulouse (France), Torun (Poland), Tula (Russia), Vadodara (India), Vasteras (Sweden), Venice (Italy), and Washington (USA).

The distribution of the *dataset* over the world is the following: most of the scenes are in Asia, 18 scenes, and in Europe, 15 scenes; in North and South America there are 5 scenes; and in Africa 1 scene.

B. Collections

Trying to semantically annotate the entire *dataset* as a single collection, we realized that this operation is very difficult or maybe impossible to be achieved with a good accuracy. For this reason, we started gradually the annotation first by generating collections with only one scene (type 1) and after that grouping two scenes with different geographical

location (type 2). The last type of grouping (type 3) is based on the geographical or architectural characteristics of the area covered by the respective scenes. For each type of grouping are presented the collections that can be generated using different locations that are available in our *dataset*.

Eight collections that contain only one scene (type 1) are generated and these collections are covering the next areas:

- Africa (Oran),
- Asia (Bangkok, Shenyang),
- Europe (Vasteras, Venice),
- North America (Havana) and South America (Bogota, Nazca Lines).

In the second type of grouping, two scenes having different locations are grouped together in three collections:

- one location is in Europe and the other one in Asia (Larissa and Djarbakir),
- both locations are in Europe (Toulouse and Timisoara),
- both locations are in central Asia (Khujand and Tashkent).

The last type of grouping is based on: geographical location of the scenes or some specific architectural characteristics. In the first case, the collections are grouped geographically:

- North Asia – Russia (Krutorozhino, center part of Moscow, north part of Moscow, south part of Moscow, Perm, Rostov on Don, and Tula),
- South Asia – India (Belgaum, Pune, and Vadodara),
- Europe – Poland (Czestochowa, Lodz, and Torun).

In the second case, the collections are grouped based on architectural characteristics of these scenes:

- Skyscrapers in North America (Los Angeles, Washington) and Asia (Anshan, Dubai, and Tokyo),
- Germanic space in Europe (Germany – Bonn, Berlin, Cologne, Kiel, Munich, Oldenburg, and Switzerland – Basel).

IV. URBAN CATEGORIES EXTRACTED FROM THE DATASET

This paper is one of the first papers dealing with a large set of high resolution TerraSAR-X images and providing for these a proper semantic annotation. For each category retrieved in all the collections we defined an appropriate semantic meaning. The methodology is the one presented in Section II and after the tiling procedure (considering all 39 scenes grouped in 16 collections) about 45,000 patches are obtained with size of 160×160 pixels. 320 categories are retrieved with SVM-RF and we find a semantic annotation for each category.

In Fig. 1 a synthesis of the number of scenes and the retrieved categories in each collection is presented. For example, for North Asia 7 locations (over Russia) are grouped together and 39 semantic categories are retrieved during the evaluation contrary with the South Asia where 3 scenes over India give us 18 categories. This led us to the idea that the location, architectural characteristics, and the number of scenes influence the number of retrieved categories and their semantics.

In the first part of this section, we presented for each type of grouping described in previous section, Section III B, the number of generated collections (each type of grouping separately) with their retrieved semantic categories.

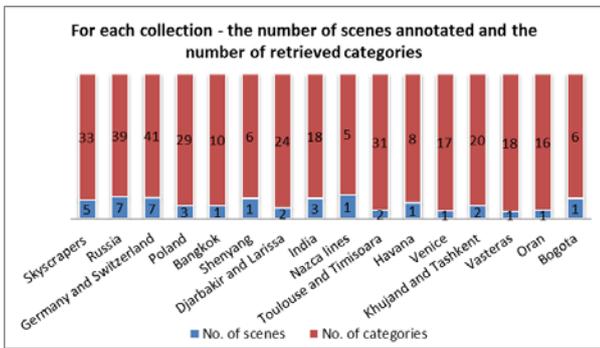


Figure 1. For each collection are presented the number of scenes and the annotated categories. For each collection the numbers of patches semantically annotated are: Skyscrapers – 5382 patches, Russia – 7187 patches, Germany and Switzerland – 7176 patches, Poland– 3169 patches, Bangkok – 1075 patches, Shenyang – 1014 patches, Djarkbakir and Larissa – 2015 patches, India – 3146 patches, Nazca Lines – 1150 patches, Toulouse and Timisoara – 2132 patches, Havana – 1008 patches, Venice – 1026 patches, Khujand and Tashkent – 1912 patches, Vasteras – 1053 patches, Oran – 962 patches, and Bogota – 900 patches.

In the next figures, Fig. 2 ÷ Fig. 4, for each type of grouping the number of scenes, the number of retrieved categories, and the number of categories that contain man-made structures (buildings, bridges, ports, streets, etc.) are shown. In order to find the man-made structures categories, we removed from the total number of retrieved categories obtained applying the methodology presented in Section II, the categories as agriculture, land, forest, water, frozen water, desert, mountains, etc.

In the next paragraphs, some particularities that we identified during the semantic annotation are detailed.

In the case of the first type of grouping (Fig. 2), the largest number of retrieved categories (16-18 categories) is obtained in Europe (Venice and Vasteras) and Africa (Oran). For North / South America (Havana and Nazca Lines) and Asia (Bangkok) this number of categories is limited to 10. From the total number of retrieved categories 57% are categories that contain man-made structures.

Note that, the Nazca Lines is the collection with the highest number of patches (775 patches) with the same semantic meaning (e.g., “desert”) but this scene has only one category that contains man-made structures (“roads”).

In the second case (Fig. 3), where two scenes with different location are grouped together; the number of retrieved categories varies between 20 and 30. From these categories 55% contain man-made structures.

In the last case (Fig. 4), the grouping is realized first based on geographical location of the scenes (e.g., India, Poland, and Russia) and second based on the architectural characteristics of the scenes (e.g., skyscrapers or cities in the Germanic space). The categories retrieved vary between 18 and 39 when is taking into account the geographical location and between 33 and 41 when considering the architectural information. The number of man-made structures categories represents 60% from the total number of retrieved categories. A higher number of categories that contain man-made structures are for those collections in which the grouping was based on architectural characteristics of the considered area (e.g., the scenes from the Germanic space). These categories represent 65% from the total number of retrieved categories.

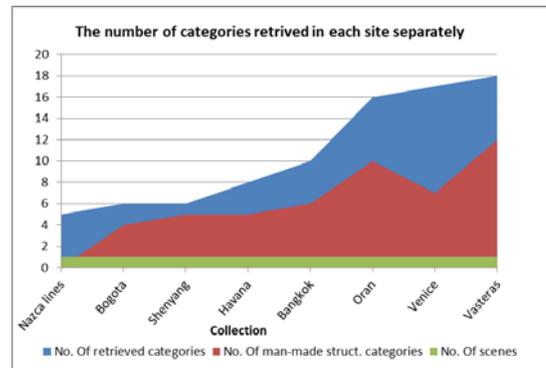


Figure 2. The scenes that are grouped in collections based on the first criterion with their categories retrieved in each collection.

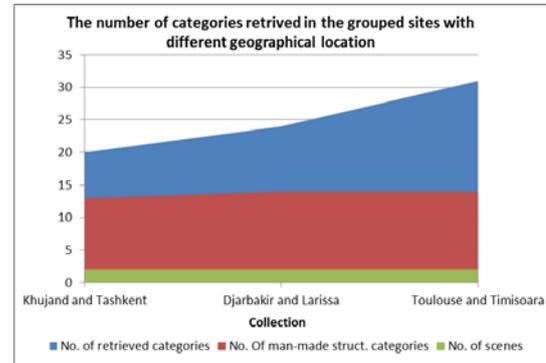


Figure 3. The scenes grouped in collections based on the second criterion (where the location of the scene was different) with their number of categories that are retrieved in each collection.

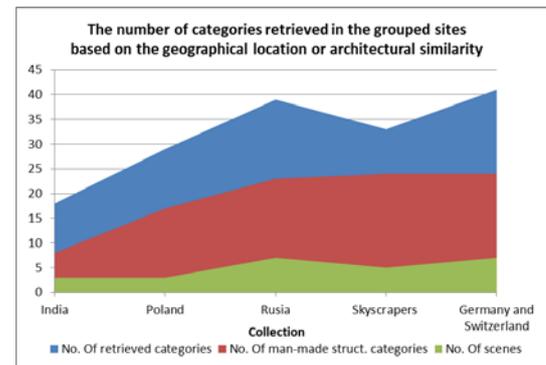


Figure 4. The scenes grouped in collections of type 3 (considering the geographical location or architectural similarity) with their categories retrieved in each collection.

In the second part, we selected from our *dataset* one scene for each region (Africa, Asia, Europe, and North/South America) and we show the retrieved categories having the dominant content in the patch ‘man-made structures’.

To illustrate the man-made categories retrieved, we selected one scene for each continent: Oran (Algeria), Bangkok (Thailand), Venice (Italy), Havana (Cuba), and Bogota (Columbia) and present these categories in the next figures (Fig. 5 ÷ Fig. 9). For each scene separately are presented: the semantic meaning, the number of patches in each category, and the quick-look of the center of each retrieved category.

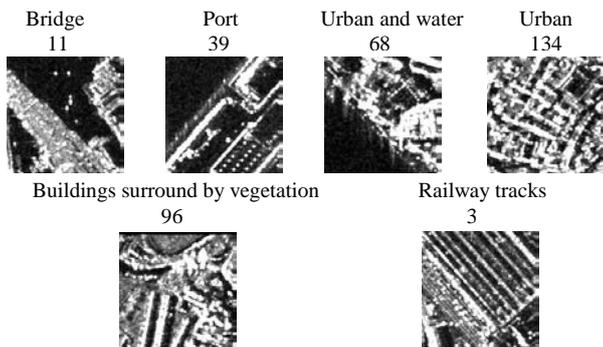


Figure 5. The man-made structures categories retrieved for Venice.

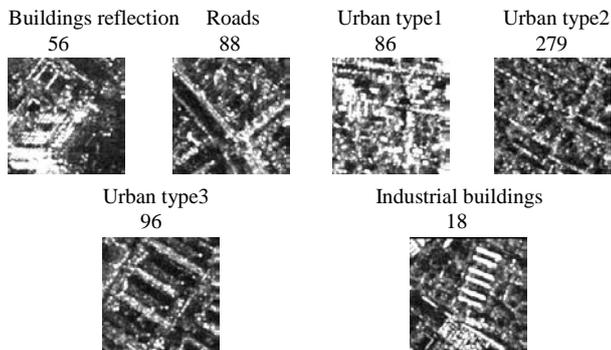


Figure 6. The man-made structures categories retrieved for Bangkok.

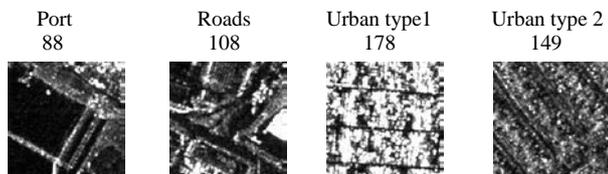


Figure 7. The man-made structures categories retrieved for Havana.

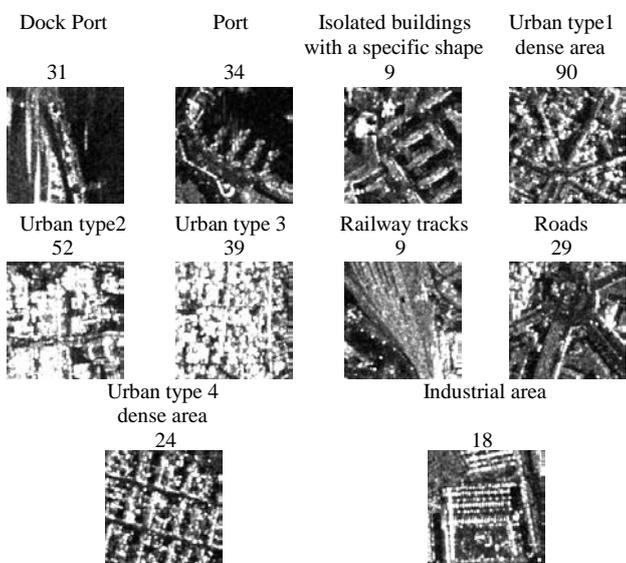


Figure 8. The man-made structures categories retrieved for Oran.

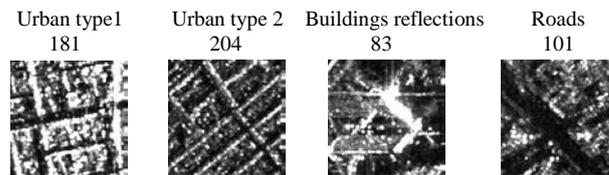


Figure 9. The man-made structures categories retrieved for Bogota.

V. CONCLUSION

Considering the first two types of grouping (type 1 and type 2), the highest numbers of retrieved categories are obtained for Europe (e.g., Venice, Vasteras, and Timisoara) followed by Africa (e.g., Oran). For the others the number of categories was lower this means the geographical location influence the number of categories, but it is not possible to group scenes with different location under the same semantic category. In the case of the third type of grouping (type 3) was possible to annotate patches retrieved from different scenes (between 5 and 7 scenes by collection) having the same geographical location or architectural characteristics with the same label.

Regarding the number of man-made structures categories these represents between 50% and 65% from all existing categories and depends also by the type of grouping.

As a final conclusion, we recommend to use the geographical or architectural grouping in order to semantically annotate large datasets.

As a future work, we try to extend the dataset in order to cover more areas around the world and to use this semantic annotated *dataset* for a query-builder in our new generation of data mining system [7].

REFERENCES

- [1] C.-R. Shyu, M. Klaric, G. Scott, A. Barb, C. Davis, and K. Palaniappan, "GeoIRIS: Geospatial Information Retrieval and Indexing System – Content Mining, Semantics Modeling, and Complex Queries", IEEE Transactions on Geoscience and Remote Sensing, 2007, pp. 839-852.
- [2] A. Popescu, I. Gavati, and M. Datcu, "Contextual Descriptors for Scene Classes in Very High Resolution SAR Images" IEEE Geoscience and Remote Sensing Letters, 2012, 80-84.
- [3] J. Singh and M. Datcu, "SAR Image Categorization with Log-Cumulants of the Fractional Fourier Transform Coefficients", IEEE Transaction on Geoscience and Remote Sensing, to be published.
- [4] C.O. Dumitru and M. Datcu, "Dependency of SAR Image Structure Descriptors with Incidence Angle", SPACOMM 2012, Chamonix / Mont Blanc, April 2012, pp. 92-97.
- [5] C.O. Dumitru and M. Datcu, "Study and Assessment of Selected Primitive Features Behaviour for SAR Image Description", Proc. of IGARSS 2012, Munich, July 2012, pp. 3596-3599.
- [6] C.O. Dumitru, J. Singh, and M. Datcu, "Selection of relevant features and TerraSAR-X products for classification of high resolution SAR images", Proc. of EUSAR, Nuremberg, April 2012, pp. 243-246.
- [7] Virtual Observatory Infrastructure for Earth Observation Data - TELEIOS project, Available: <http://www.earthobservatory.eu/knowledge-discovery-from-EO-data>
- [8] J. Anderson, E. Hardy, J. Roach, and R. Witmer, "A Land Use and Land Cover Classification System for Use with Remote Sensor Data", 1976, Available: <http://landcover.usgs.gov/pdf/anderson.pdf>.
- [9] TerraSAR-X: "Basic Products Specification Document", Issue: 1.6 (TX-GS-DD-3302)

