A Study of Multi-Sensor Satellite Image Indexing

Corneliu Octavian Dumitru, Shiyong Cui, Mihai Datcu
Remote Sensing Technology Institute (IMF)
German Aerospace Center (DLR)
Oberpfaffenhofen – Wessling, Germany
Email: {corneliu.dumitru, shiyong.cui, mihai.datcu}@dlr.de

Abstract—**In the context of earth observation, different sensors have been used to acquire satellite images and it becomes a research topic about how to analyse and use multi-sensor images. In this paper, we carry out a study of multi-sensor satellite image indexing. The goal is to study which kind of satellite image provides more information for classification. To this end, we prepared four datasets covering four typical cities. Each dataset consists of three kinds of images: multispectral and panchromatic images from WoldView-2, Synthetic Aperture Radar (SAR) images from TerraSAR-X satellite. Image indexing is performed at patch level with the same feature extraction method. The indexing is carried out using an active learning system we developed before. A series of independent and joint indexing by combining the features have been performed. Through this study, we found that the indexing accuracy on SAR images is the worst. By contrast, the joint indexing by concatenating the features computed from each kind of image could provide best accuracy. Thus, we conclude that combing information from multi-sensor images could achieve better results than using each kind of image independently.**

I. Introduction

In the context of earth observation, there are many satellites that have been launched and they provide a large variety of satellite images, such as Synthetic Aperture Radar (SAR) image, Hyperspectral and multispectral optical images, and panchromatic images with higher resolution. These images play an important role in different applications and research fields. However, most research subjects focus only on a particular kind of images. There are only few researches investigating combination of different kinds of image acquired by different sensors for indexing. A typical research topic about multi-sensor images is registration, such as [1]. However, there are no many works to analyse multi-sensor images. In [2], a combination of Landsat-7 enhanced thematic mapper panchromatic and SPOT data is used for urban land use change detection. Since these images are acquired by different sensors, they provide different information about of the covered area. It is worth to study their combination in order to achieve better indexing accuracy. Thus, in this paper, it motivates us to carry out a study of multi-sensor satellite image indexing.

II. Methodology

A. *Our methodology in seven steps*

- Select one of four dataset and process its three kinds of images: multi-spectral image, panchromatic image, TerraSAR-X image, and a joint of all three images.
- Tile each kind of images into patches with a size of 100 by 100 pixels.
- Generate the quick-looks of the patches and the full image, which are needed by our GUI tool [3].
- Extract bag-of-words (BoW) features from each image patch with a dimensionality of 200. In the case of joint sensor evaluations, we use the concatenation of single feature vectors. The details can be found from [4], [6].
- Iteratively select training samples and learn a Support Vector Machine classifier with relevance feedback (RF) in order to group the patches into categories.
- Annotate semantically each retrieved category using our hierarchical annotation scheme [5].
- Compute the precision/recall (P/R) based on the reference dataset [3].

B. *GUI Tool*

In order to easily work with the multi-sensor data a tool was developed for Earth Observation (EO) data annotation. There are mainly two modules, which are feature extraction and active learning based on SVM for annotation. Different kinds of tasks can be realized through this tool, depending on the ingested data set. The multi-temporal / multi-sensor images can be annotated individually or jointly. Individual annotation is the same as single image annotation, where a set of images are ingested and they are annotated in the same feature space. Joint annotation is the working mode in which the images are annotated jointly in the concatenated feature space.

Using this tool, the users can create a project by importing a set of images. The number of images that can be imported depends on the available memory. Categories can be discovered through active learning and the annotation can be saved and exported. These are further used to compute precision/recall measures. The overall architecture is shown in Fig. 1 and a detailed description of this tool can be found in [8].

III. Dataset

We selected four pairs of multi-sensor images covering the following cities: Bucharest (Romania), Munich (Germany), Venice (Italy), and Washington DC (USA). In figure 2 is presented the multi-sensor images covering the city of Munich.

The TerraSAR-X product-images are in the format of enhanced ellipsoid corrected (EEC) and radiometrically enhanced (RE) with HH polarization for Bucharest and Washington and VV polarization for Munich and Venice.

978-1-4799-6652-3/15/$31.00~ © 2015 IEEE

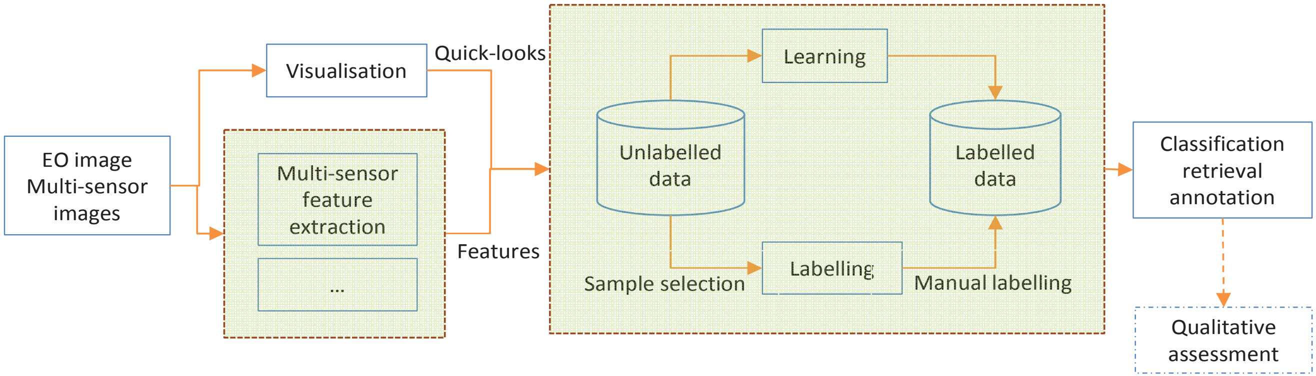

Figure 1. The overall architecture of the system.

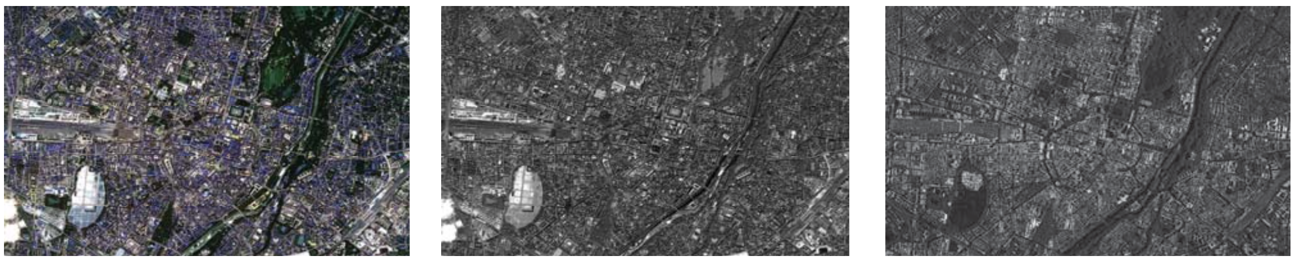

Figure 2. A multi-senor data: multi-spectral image (left side), panchromatic image (center), and TerraSAR-X image (right side) for the city of Munich.

The ground range resolution is 2.9 m with WGS-84 map projection. The size of these images is: for Bucharest 6800×9600 pixels,for Munich 8000×9200 pixels, for Venice 8000×8000 pixels, and for Washington 6800×8000 pixels.

The WoldView-2 products include both the panchromatic (0.46 m) and multi-spectral images (1.87 m). The map projection of WordView-2 is WGS-84 like the one for TerraSAR-X. The size of these images is: for Bucharest 47,399×37,463 pixels (panchromatic image) and 11,850×9366 pixels (multi-spectral image), for Munich 52,764×34,812 pixels (panchromatic image) and 13,191×8703 pixels (multi-spectral image), for Venice 47,113×43,452 pixels (panchromatic image) and 11,778×10,863 pixels (multi-spectral image), and for Washington 39,532×33,786 pixels (panchromatic image) and 9883×8447 pixels (multi-spectral image).

A difficulty arises when trying to co-align these images because the data has different pixel spacings. To solve this problem, we resample the panchromatic image in order to co-align with TerraSAR-X image. The disadvantage of this process is a loss of details for panchromatic images.

IV. Experimental Results and Discussion

To evaluate the performances of indexing using the methodology proposed in Section II, each sensor image is tiled in patches of 100×100 pixels and the BoW features [7] from each patch are computed. The total number of patches is approximatively 25,000 patches (Bucharest, Munich, Venice, and Washington DC). All experiments conducted hereinafter are based on an active learning system for interactive multi-sensor image indexing [7].

Separately, for each sensor image we tried to discover categories among all patches corresponding to each city. For each category we give 20% of the patches for training and we try to index similar patches with about 7 to 10 iterations. The evaluation stops when the indexed patches do not change anymore. The procedure is repeated 2-3 times for the same category, city and sensor image giving initially the same sequence of positive and negative examples (patches). All these identified categories are semantically annotated using our hierarchical annotation scheme with two levels [5] or three levels [8].

We start the indexing first for multi-spectral sensor data, second for panchromatic data, third for TerraSAR-X data, and finally for the combination of WoldView-2 (multi-spectral and panchromatic) and TerraSAR-X data.

For quantitative assessment, we compare the indexing results with the reference dataset and compute precision/recall for each category, city and sensor. The precision/recall is shown in Tables 1-4 for each sensor image and joint sensor images separately. For each retrieved category is given the semantic meaning, the number of patches in each category and the precision/recall.

Analysing each table separately, we can observe that the overall average of precision/recall obtained for joint sensor images is higher than the precision/recall of individual sensor images. The best average value of precision/recall is marked with green colour, while the lower value of precision/recall is marked with bright red colour. Taking into account each category separately, we can see from the results presented in tables that the highest precision/recall is obtained by using as input one sensor image or another sensor image.

TABLE I. PRECISION / RECALL (%) RESULTS FOR MUNICH (DE).

No.	Label	No. patches	Multi-spectral		Panchromatic		TerraSAR-X		Joint images	
			Precision	*Recall*	*Precision*	*Recall*	*Precision*	*Recall*	*Precision*	*Recall*
1	**Amusement parks**	156	85.06	83.87	59.60	75.64	48.80	51.92	87.06	94.87
2	**Bridges**	35	76.00	54.29	60.87	40.00	53.33	22.86	53.39	42.86
3	**Broadleaf forest / Sparse Trees**	3620	97.67	34.72	95.34	46.38	91.08	20.03	96.69	32.26
4	**Clouds**	233	94.53	96.57	97.54	51.07	0.00	0.00	98.51	48.33
5	**Grassland**	256	70.21	38.67	85.45	36.72	43.59	29.92	93.55	33.98
6	**Industrial commercial areas**	503	62.84	22.86	48.96	32.80	36.22	31.59	46.04	38.97
7	**Mixed urban areas**	4637	71.95	57.79	69.64	61.70	61.63	63.10	58.15	79.53
8	**Railway tracks**	349	80.69	74.21	51.89	74.79	34.49	74.50	88.62	82.52
9	**River**	118	61.34	61.86	77.27	57.63	45.40	66.95	67.63	79.66
10	**Roads**	816	24.33	25.98	17.21	19.68	53.64	72.30	35.18	27.03
			72.46	**55.08**	**66.38**	**49.64**	**52.02**	**48.13**	**72.48**	**56.00**

TABLE II. PRECISION / RECALL (%) RESULTS FOR VENICE (IT).

No.	Label	No. patches	Multi-spectral		Panchromatic		TerraSAR-X		Joint images	
			Precision	*Recall*	*Precision*	*Recall*	*Precision*	*Recall*	*Precision*	*Recall*
1	**Boats**	76	68.89	81.58	84.00	55.26	33.33	6.58	87.95	90.53
2	**Bridges**	11	100.00	100.00	100.00	72.72	100.00	100.00	100.00	100.00
3	**Broadleaf forest**	15	77.78	46.67	72.72	53.33	40.00	80.00	100.00	80.00
4	**Buoys**	48	0.00	0.00	0.00	0.00	79.17	90.48	0.00	0.00
5	**Cemeteries**	3	100.00	75.00	100.00	75.00	75.00	75.00	100.00	100.00
6	**Channel and Medium density area**	29	83.34	68.97	85.00	58.62	82.76	82.76	100.00	82.76
7	**Firth**	17	100.00	94.12	100.00	52.94	100.00	82.35	100.00	94.12
8	**Harbour infrastructure**	31	90.90	64.52	95.45	67.74	66.67	77.42	100.00	87.10
9	**Industrial buildings**	24	52.38	45.83	47.06	33.33	39.28	45.83	85.00	70.83
10	**Medium density residential area**	96	87.50	72.92	86.36	79.17	91.76	81.25	95.88	96.88
11	**Railway tracks**	4	100.00	100.00	100.00	100.00	100.00	75.00	100.00	100.00
12	**Sea**	210	97.11	77.78	94.26	54.76	77.42	80.00	99.49	92.38
13	**Sea and Medium density area**	43	79.17	88.37	62.50	46.51	60.00	48.84	100.00	100.00
			86.42	**76.31**	**85.61**	**62.45**	**78.78**	**77.13**	**97.36**	**91.22**

TABLE III. PRECISION / RECALL (%) RESULTS FOR BUCHAREST (RO).

No.	Label	No. patches	Multi-spectral		Panchromatic		TerraSAR-X		Joint images	
			Precision	*Recall*	*Precision*	*Recall*	*Precision*	*Recall*	*Precision*	*Recall*
1	**Administrative and Monument areas**	646	50.29	36.47	53.75	28.17	44.49	42.30	94.78	73.21
2	**Bridges**	24	42.42	58.33	44.38	45.83	33.45	37.50	80.95	70.83
3	**Broadleaf forest**	1061	82.96	41.67	79.47	61.55	56.57	52.87	95.39	76.06
4	**Cemeteries**	72	44.45	36.67	36.11	38.10	41.10	36.57	91.67	30.56
5	**Grassland**	201	41.94	71.14	45.28	81.08	40.29	77.62	78.00	84.03
6	**High density residential areas**	617	46.45	58.99	44.78	54.94	43.64	39.66	96.98	57.37
7	**Medium density residential areas**	3120	73.97	57.12	69.40	78.36	51.51	42.05	94.75	89.58
8	**Mixed urban areas**	374	56.00	39.21	52.94	32.46	53.24	38.72	80.21	40.11
9	**Parking areas**	143	60.61	43.97	68.75	35.39	50.00	37.00	52.76	46.85
10	**River**	120	69.37	64.17	64.14	67.50	59.08	47.50	80.00	80.33
11	**Roads**	949	56.37	45.39	59.90	44.24	47.84	42.33	98.60	22.34
12	**Sports grounds**	21	100.00	80.95	89.26	65.19	52.31	58.10	85.45	79.00
			60.40	**52.84**	**59.01**	**52.73**	**47.79**	**46.02**	**85.80**	**62.52**

TABLE IV. PRECISION / RECALL (%) RESULTS FOR WASHINGTON DC (USA).

No.	Label	No. patches	Multi-spectral		Panchromatic		TerraSAR-X		Joint images	
			Precision	*Recall*	*Precision*	*Recall*	*Precision*	*Recall*	*Precision*	*Recall*
1	**Boats**	5	100.00	100.00	100.00	100.00	100.00	60.00	100.00	87.00
2	**Bridges**	12	92.31	100.00	100.00	100.00	90.00	75.00	100.00	91.67
3	**High density residential area**	182	79.12	79.12	84.21	70.33	77.14	74.18	98.01	81.32
4	**Medium density residential area**	159	82.68	93.08	82.56	89.31	89.68	87.42	99.32	92.45
5	**Mixed forest**	107	56.07	96.77	92.06	54.21	84.88	68.22	100.00	98.13
6	**Railway tracks**	26	72.00	69.23	81.82	69.23	41.18	26.92	65.71	88.46
7	**River**	35	97.06	94.29	96.97	91.43	96.97	91.43	91.67	94.29
8	**Streets**	48	51.52	35.42	53.85	29.17	80.00	16.67	89.47	35.42
			78.85	**83.49**	**86.43**	**75.46**	**82.48**	**62.48**	**93.02**	**83.59**

In Figure 3 and 4 are shown typical examples of categories that we identified applying the methodology. We selected 3 categories (*bridges*, *channel and medium density residential area and harbour infrastructure*) out of 13 categories for the city of Venice, Italy. The same number of categories was selected for exemplification from the city of Washington DC, USA.

These categories are: *boats*, *bridges* and *medium density* residential *area*. The example patches tiled from multi-spectral image, from panchromatic image and from TerraSAR-X image are displayed from the left to right in figure 3 and figure 4. These two examples are "happy" cases in which tiled patches coming from different sensor images have the same semantics.

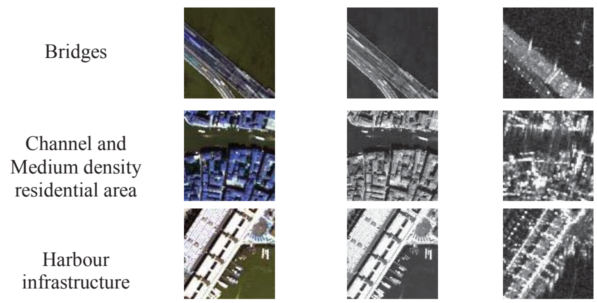

Figure 3. Three annotated categories (of the city of Venice, Italy) are selected for visualisation. The patches from the left side are tiled from multi-spectral image, the patches from the center are tiled from panchromatic image and the last patches from the right side are tiled from the TerraSAR-X image.

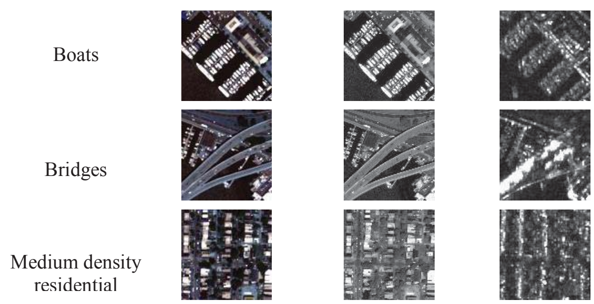

Figure 4. Three annotated categories (of the city of Washington, USA) are selected for visualisation. The patches from the left side are tiled from multi-spectral image, the patches from the center are tiled from panchromatic image and the last patches from the right side are tiled from the TerraSAR-X image.

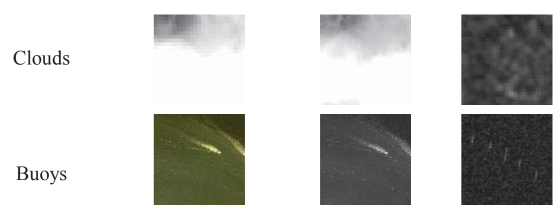

Figure 5. Two annotated categories (of the city of Munich, Germany first row and city of Venice, Italy second row) are shown for visualisation. In first case (first row) for the first two patches the semantic annotation is clouds and for the last one is broadleaf forest. In the secand case (second row) for the first two patches the semantic annotation is boats and for the last one buoys. The patches from the left side are tiled from multi-spectral image, the patches from the center are tiled from panchromatic image and the last patches from the right side are tiled from the TerraSAR-X image.

Unfortunately, there are also "unhappy" cases in which for the same patch tiled from different sensor image (in case of WorldView-2) we have a different semantic in case of TerraSAR-X. This is the case of category *clouds* from the city of Munich, Germany and the category *buoys* from the city of Venice, Italy. Such examples are presented in Figure 5.

For the first exception, namely clouds, this category is not present in TerraSAR-X data this is one of the advantages of the SAR sensors. For second exception, the performance of buoys category that was retrieved only for TerraSAR-X data has impact in lowering the performances of boats category.

V. CONCLUSION

In this paper, we carried out a study of multi-sensor satellite image indexing. To this end, we prepared four datasets consisting of multispectral and panchromatic images from WoldView-2, and Synthetic Aperture Radar (SAR) images from TerraSAR-X satellite. The indexing is carried out based on an active learning system. A series of independent and joint indexing by combining the features have been performed. Through this study, we found that the indexing accuracy on SAR images is the worst. By contrast, the joint indexing by concatenating the features computed from each kind of image could provide best accuracy. To conclude, the joint sensor data enable us to discriminate more accuracy the retrieved categories (over all investigated cities).

ACKNOWLEDGMENT

The TerraSAR-X image data being used in this study were provided by the TerraSAR-X Science Service System (Proposal MTH 1118), while the World-View–2 image data were provided by the European Space Imaging (EUSI).

REFERENCES

[1] S. Xing, Q. Xu, D. Ma, "Multi-sensor Satellite Image Sub-pixel Registration" Conference on Image and Graphics, 2007. ICIG 2007. Fourth International, 22-24 Aug. 2007, pp.713-718.

[2] J. S. Deng, K. Wang, J. Li, and Y. H. Deng, „Urban land use change detection using multisensor satellite images", Pedosphere, Vol. 19, No. 1, 2009, pp. 96-103.

[3] C.O. Dumitru and M. Datcu, "Information Content of Very High Resolution SAR Images: Study of Feature Extraction and Imaging Parameters", IEEE Trans. on Geoscience and Remote Sensing, Vol. 51, No. 8, August 2013, pp. 4591-4610.

[4] D.G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints., International Journal of Computer Vision, Vol. 60, No. 2, November 2004, pp. 91–110.

[5] C.O. Dumitru, S. Cui, G. Schwarz and M. Datcu, Information Content of Very High Resolution SAR Images: Semantics, Geospatial Context, and Ontologies, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (in press).

[6] S. Cui, Spatial and Temporal SAR Image Information Mining, PhD thesis, August 2014.

[7] S. Cui, C.O. Dumitru, and M. Datcu, Semantic Annotation in Earth Observation Based on Active Learning, International Journal of Image and Data Fusion, Vol. 5, No. 2, April 2014, pp. 152-174.

[8] C.O. Dumitru, G. Schwarz and M. Datcu, Land Cover Semantic Annotation Derived from High Resolution SAR Images, International Journal of Remote Sensing (under the second review).