Multi-sensor remote sensing information fusion
for urban area classification and change detection

Gintautas Palubinskas*, Aliaksei Makarau, and Peter Reinartz

Remote Sensing Technology Institute, German Aerospace Center DLR, Oberpfaffenhofen, Germany

*Gintautas.Palubinskas@dlr.de, phone 49 8153 28-1490, fax 49 8153 28-1444, www.dlr.de/eoc

ABSTRACT

Information extraction from multi-sensor remote sensing imagery is an important and challenging task for many applications such as urban area mapping and change detection. A special acquisition (orthogonal) geometry is of great importance for optical and radar data fusion. This acquisition geometry allows to minimize displacement effects due inaccuracy of Digital Elevation Model (DEM) used for data ortho-rectification and existence of unknown 3D structures in a scene. Final data spatial alignment is performed by recently proposed co-registration method based on a Mutual Information measure. For a combination of features originating from different sources, which are quite often non-commensurable, we propose an information fusion framework called INFOFUSE consisting of three main processing steps: feature fission (feature extraction aiming at complete description of a scene), unsupervised clustering (complexity reduction and feature representation in a common dictionary) and supervised classification realized by Bayesian or Neural networks. An example of urban area classification is presented for the orthogonal acquisition of space borne very high resolution WorldView-2 and TerraSAR-X Spotlight imagery over Munich city, South Germany. Experimental results confirm our approach and show a great potential also for other applications such as change detection.

Keywords: multi-sensor, remote sensing, optical image, SAR, information fusion, classification, change detection

1. INTRODUCTION

Data fusion is a rapidly developing topic in various application areas during the last decades. Image fusion in remote sensing is one of them. However fusion of different sensor data such as optical and radar imagery is still a challenge. In this paper the term 'radar' is equivalent to Synthetic Aperture Radar (SAR). Different modalities of data can be obtained by different sensors for the same area, and more properties can be revealed on the area structure, contents, properties, etc. Several novel and competitive approaches on urban area and land cover classification using fusion were proposed, e.g. see results of fusion contest[6]. The algorithm is based on a neural network classification enhanced by pre-processing and post-processing. Principal component analysis was applied on SAR data. Altogether, 14 inputs to the Neural Network were given: 2 SAR images, 6 Landsat-5 spectral images, and 6 Landsat-7 spectral images. The classification into 5 classes (city center, residential area, sparse buildings, water and vegetation) provided Kappa coefficient equal to 0.93.

Fauvel[2] applied decision fusion for classification of urban area. The fusion approach consisted of two steps. In the first step, data were processed by each classifier separately and the algorithms provided for each pixel membership degrees for the considered classes. In the second step, a fuzzy decision rule was used to aggregate the results provided by algorithms according to the classifiers' capabilities. The method was tested and validated with two classifiers on IKONOS images for urban areas. The proposed method improves classification results when compared with separate use of different classifiers. The overall accuracy of classification for 6 classes (large buildings, houses, large roads, streets, open areas and shadows) is 75.7 %.

We approach the joint classification task using a more general view on the whole data fusion problem. Thus data acquisition planning and pre-processing become very important steps for a successful data fusion.

The paper is organized as follows. First, in Section 2 we illustrate by the example of optical and SAR data the difficulty of image fusion. Then, we introduce our data fusion concept which we follow further in solving data fusion tasks such as multi-sensor classification and change detection. The section 2 ends with a list of pre-processing tasks which are

necessary for a successful application of data fusion. In Section 3 the special (orthogonal) multi-sensor formation geometry for optical/SAR data acquisition is proposed aiming at easier data evaluation/interpretation in further processing steps. INFOFUSE classification framework and an example of the classification for WorldView-2 and TerraSAR-X data are presented in Sections 4 and 5 respectively.

2. PREPARATION FOR FUSION

A proper preparation of data is a very important prerequisite for a successful data fusion. Following sub-sections present our approach methodologically including the data fusion concept and pre-processing step list.

2.1 Fusion problem

For the fusion of data from sensors exhibiting different acquisition geometries such as optical and radar systems it is important to understand their influence on the fusion process and to optimize it if necessary. For example, in Figure 1 it is practically impossible to recognize Frauenkirche (famous church - tourist attraction – in Munich city center) in SAR image even with the help of optical image. Special data acquisition geometry can help enormously as can be seen later in Section 3.
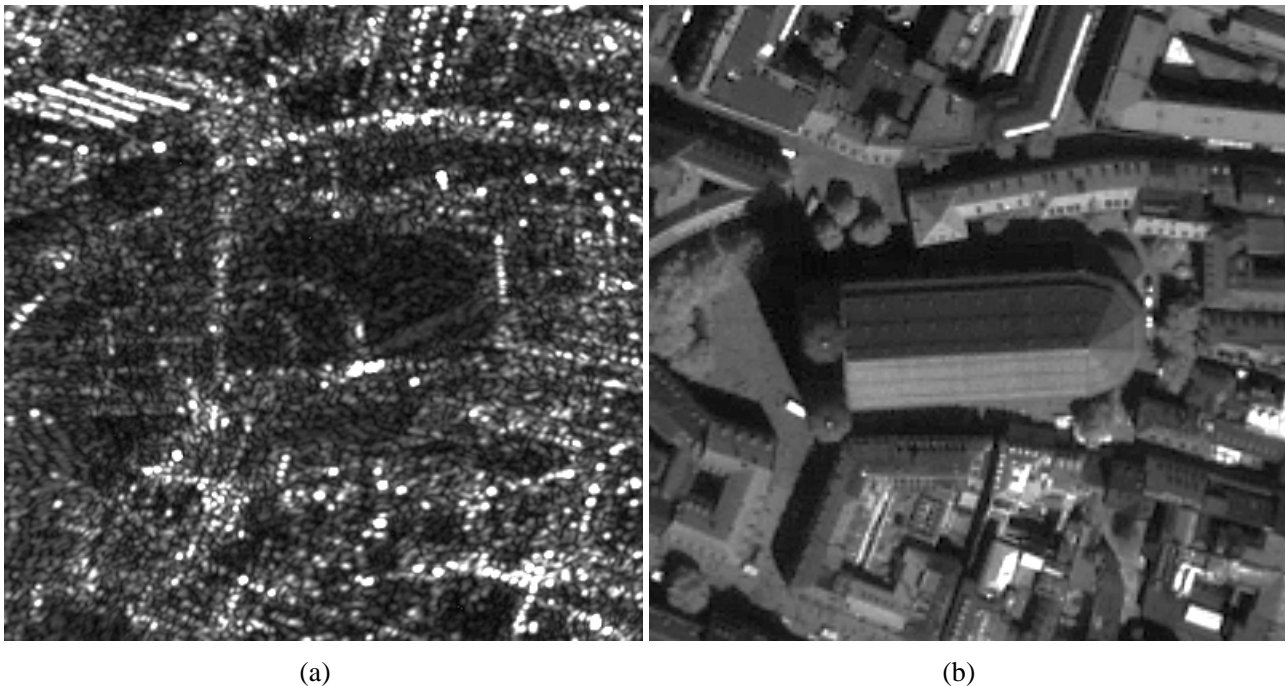

| (a) | (b) |

Figure 1. Part of Munich city center with Frauenkirche acquired by TerraSAR-X HS Spotlight mode (a) and IKONOS panchromatic mode (b) using the accidental satellite formation.

2.2 Fusion concept

After extensive literature review and our experience in this topic we propose the following fusion concept as shown in Figure 2.

The concept consists of three parts: sensors, methods and applications. Sensors are used to acquire data from various sources sometimes incommensurable. Special attention should be paid for sensors exhibiting different acquisition geometries as optical and radar sensors. Methods can be divided into two groups: pre-processing, e.g. ortho-rectification, radiometric normalization, co-registration and fusion methods itself. Fusion methods can be applied in different levels: pixel or signal (iconic), feature (symbolic) and knowledge (decision). Among them are Artificial Neural Networks, Bayesian Networks, Support Vector Machines just to mention few. Various tools e.g. simulation of optical/SAR images from Digital Surface Models (DSM) belong to methods group too. The third but not least group is applications. We have to note at this point that the fusion is always application dependent. Achieved results should be validated by quality assessment to prove usefulness of data fusion.

As we have seen already, according the proposed fusion concept already data pre-processing steps such as ortho-rectification and co-registration introduced in the following Section belong to data fusion itself.
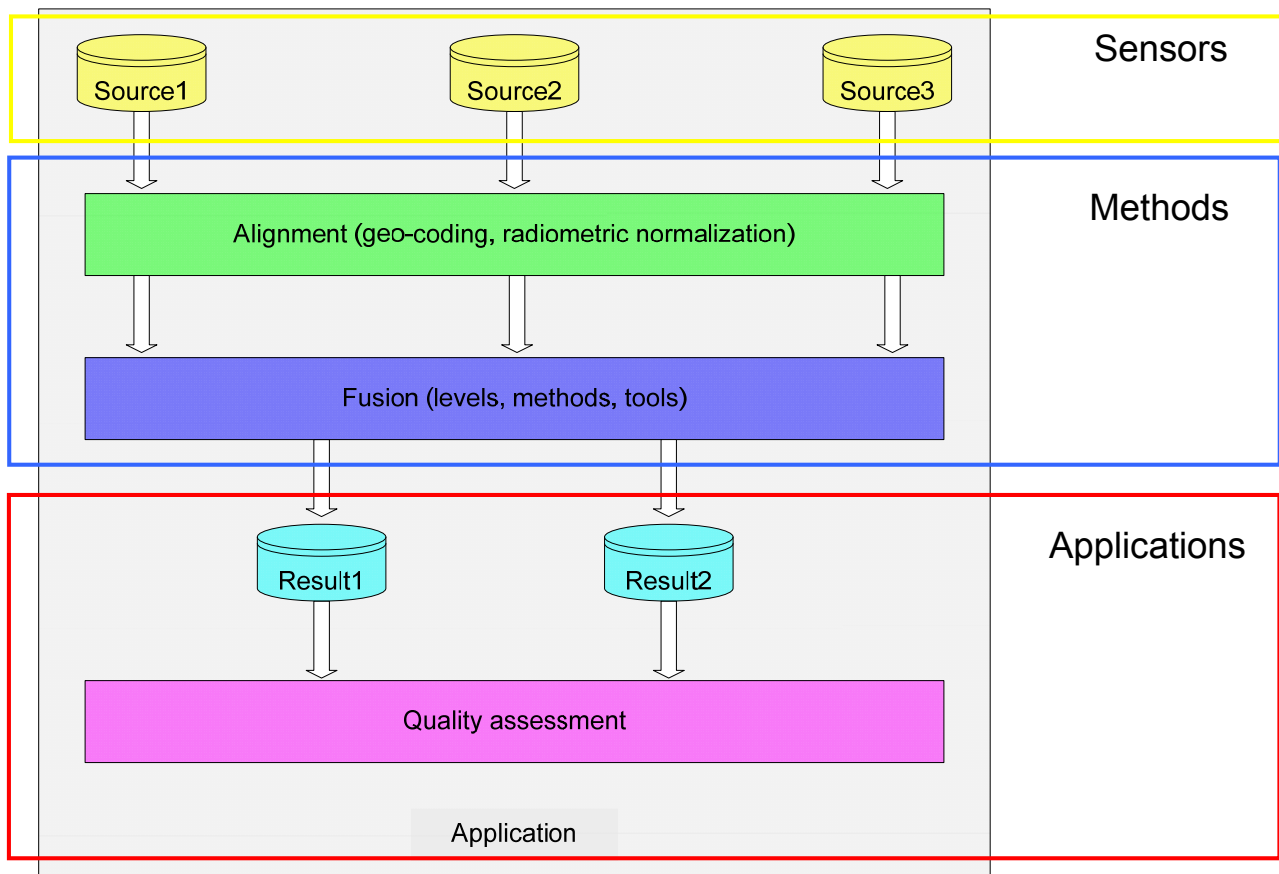

Figure 2. Proposed fusion concept.

2.3 Pre-processing

For the data fusion on the lowest level (pixel or image-based) data pre-processing such as ortho-rectification and co-registration is an important prerequisite for a success. The main tasks are summarized in Table 1.

Table 1. Main pre-processing tasks.

Task	Method	Reference
Orthogonal acquisition geometry	Minimize displacement effects[9]	(Palubinskas 2010)
Ortho-rectification	Optical imagery (using TS-X GCPs)[11] SAR imagery	(Reinartz 2011)
Co-registration	Mutual Information[12]	(Suri 2010)
Pan-sharpening (multispectral + panchromatic)	General Fusion Framework[10]	(Palubinskas 2011)
De-speckling (SAR)	Non-local means filter[1]	(Deledalle 2009)

3. ORTHOGONAL ACQUISITION GEOMETRY

In this Section we propose an optimal optical and radar sensor formation for an image acquisition compensating/minimizing ground displacement effects of different sensors[9]. A sum of look angles should give approximately 90° (Figure 3a). Flight directions should be as parallel as possible and perpendicular to look directions which are opposite for different sensors (Figure 3b). Same flight directions are not required in general e.g. airborne case. This sensor configuration allows e.g. a recovery of 3D object shadows during further data fusion, except a case when the Sun illumination direction is the same as for SAR look direction. Displayed left looking radar and right looking optical sensor formation can be preferable due to the Sun illumination direction which is from an optical sensor to the target on the Earth in order to see that side of a 3D object which is in shadow in the radar image and thus enable full reconstruction of a 3D object. Of course, the second possible sensor formation with a right looking radar and left looking optical sensor can be useful for data fusion too.
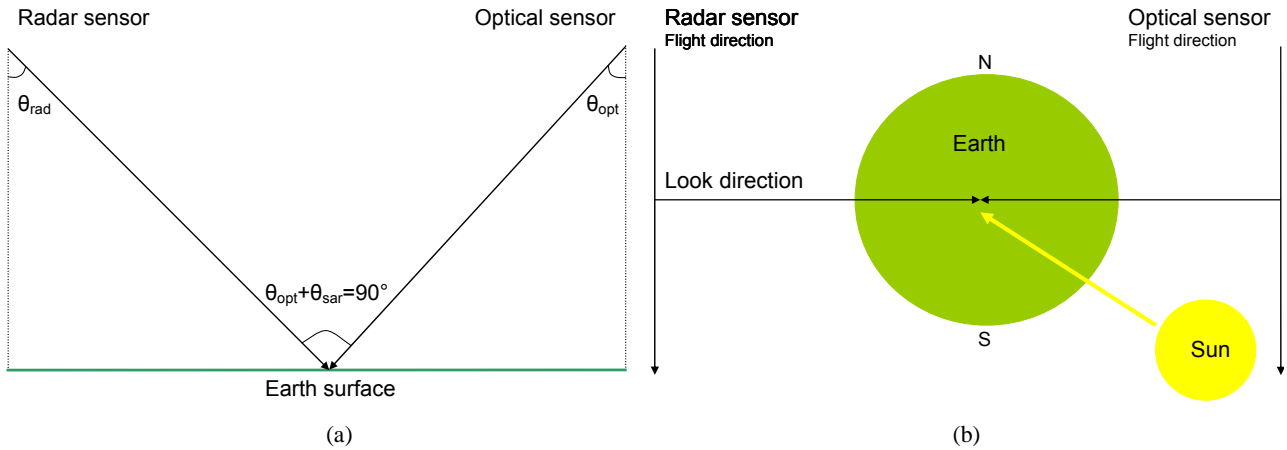

Figure 3. Proposed optical and radar sensor formation is illustrated. A sum of look angles should give 90° (a). Flight directions should be parallel, in same direction and perpendicular to look directions which are opposite for different sensors (b). Sun illumination direction is from an optical sensor to the target on the Earth.

Part of Munich center with Frauenkirche (tourist attraction) acquired by WorldView-1 (a) and TerraSAR-X (b) using the proposed satellite formation is shown in Figure 4. Ground objects like streets and plazas can be easily detected and found at the same geographical position in both images. Other structures: buildings (e.g. building block in the upper left corner of the image, church with two towers in the center of the image) and trees can be easily indentified in both images. Only the feet of the buildings, which are differently projected in the radar image due to foreshortening in radar are found at slightly different positions. So the roofs and tree crowns are well in place and can be overlaid correctly for any further processing.

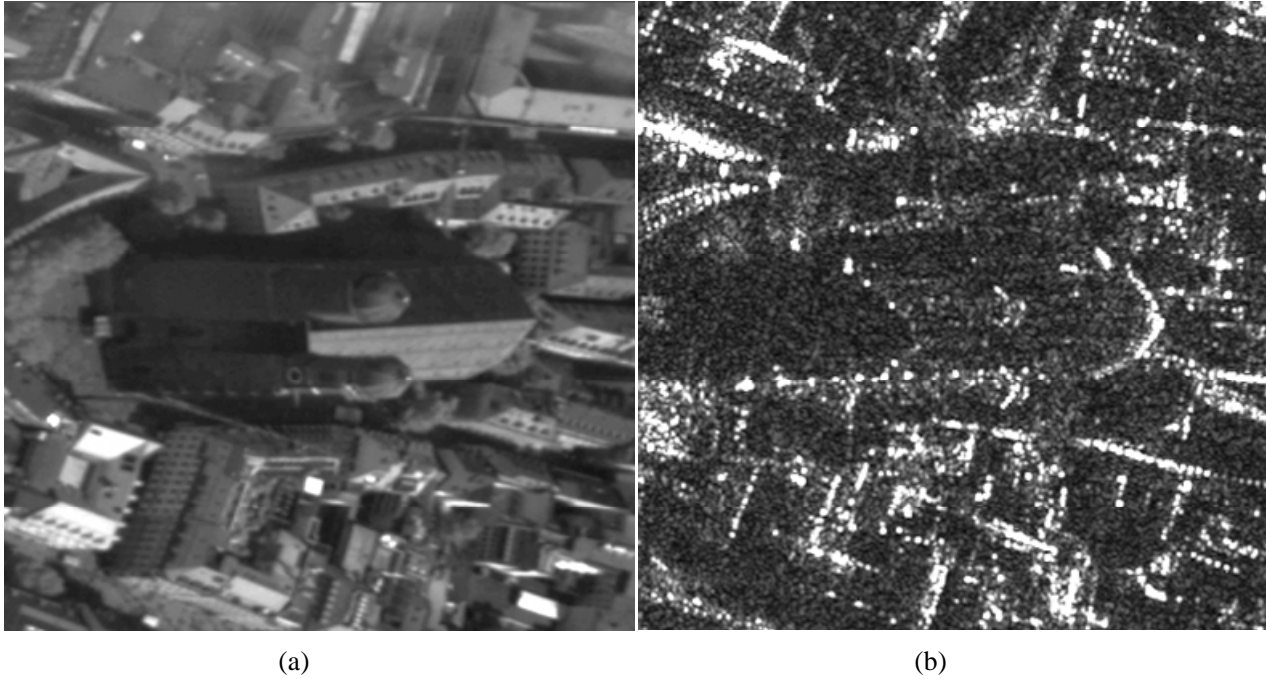(a)   (b)

Figure 4. Part of Munich center with Frauenkirche acquired by WorldView 1 panchromatic mode (a) and TerraSAR-X
    Spotlight mode (b) using the proposed orthogonal satellite formation.

4. INFOFUSE CLASSIFICATION

INFOFUSE classification approach is based on a combination of both unsupervised clustering and supervised classification, thus allowing the usage of different features and scales for data and easy inclusion of the prior information using Bayesian/Neural Networks. Data fusion framework[8] consists of three main stages as shown in Figure 5.
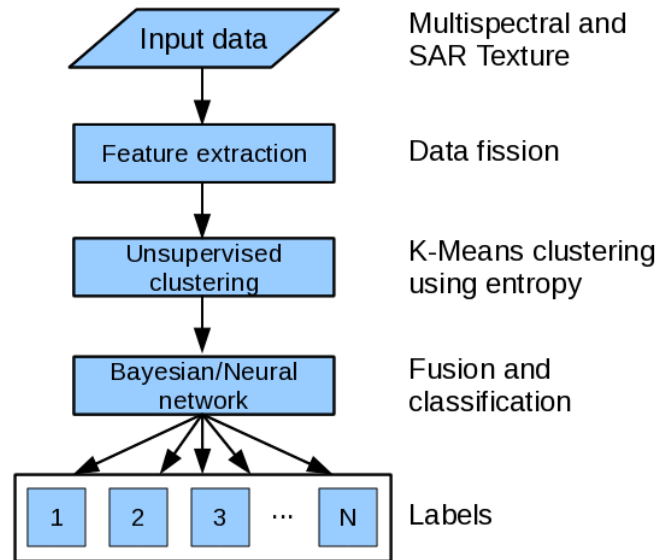
Figure 5. INFOFUSE framework for multi-sensor data fusion and classification.

4.1 Feature extraction from input datasets

Raw signal data are usually quite difficult to interpret or classify, so the pre-processing step—feature extraction—is often unavoidable. One tries to extract as much as possible various features from one source (information fission) or set of sources in order to produce a good classification. These features are expected to characterize different properties of structures and objects. After the feature extraction a large amount of redundant information is obtained.

4.2 Dimensionality reduction using unsupervised clustering

Since the aim of this step is to combine features with similar properties and to reduce the dimensionality of the calculated feature data, any unsupervised clustering method can be employed for this task. K-means clustering based on entropy[7] was applied on each extracted feature separately. The number of clusters for each feature can be different and defined individually according to the type of the feature. This step is performed to acquire the unique description of the data in terms of clusters and to reduce the dimensionality of the extracted features.

4.3 Fusion of the clustered features

A Bayesian network or a Neural network is employed to fuse the extracted features and to produce the inference (i.e. classification through fusion). Bayesian or Neural network allows to combine information from different sources of measurement, therefore the fusion of incommensurable features (numerical, logical, semantic, etc.) can be performed. Supervised training of the network allows to estimate the network state and a classification is possible to perform.

5. CLASSIFICATION WV-2 VNIR AND TS-X HS DATA (MUNICH CENTER)

In order to investigate and illustrate the effectiveness of the proposed approach we have chosen the area of Munich city as a test scene. Munich contains variety of urban building types and structures, such as old town, residential area, low- , medium-, and high-rise buildings, rail road, water regions, bare soil, etc. Two very high resolution datasets were chosen: WorldView-2 multispectral imagery and TerraSAR-X Spotlight HS mode data. WorldView-2 multispectral data were

obtained at the 12-th July 2010, 10:30:17 local time. Multispectral data contains 8 11-bit bands in 2m spatial resolution, the panchromatic data contain one 11-bit band in 0.5m spatial resolution. The spectral bands were pan-sharpened using an image fusion method based on high-frequency image data addition to low-resolution spectral image[10]. This method provides minimal distortion of spectral and spatial characteristics of multispectral imagery[4]. VNIR bands were especially used in our experiment since most of the very high resolution space borne multispectral sensors (e.g. IKONOS, GeoEye-1, Quickbird, etc.) acquire only VNIR data. TerraSAR-X data (Spotlight HS, EEC, VV polarization) were acquired at the 7-th June 2008, 06:17:48 local time. TerraSAR-X data were registered to the pan-sharpened multispectral image. Co-occurence texture features[3] for SAR data were calculated (among the texture features are the Mean, Variance, Homogeneity, Contrast, Dissimilarity, Entropy, Second Moment, and Correlation).

Table 2. List of classes/sub-classed used for classification.

Label	Classes/subclasses
Buildings	8 subclasses
Roads	2 subclasses
Water	1 class
Forest/Trees	1 class
Grass	1 class
Shadows	1 class

In this experiment we have selected 6 main classes for the urban scene: 1) Building; 2) Road; 3) Water; 4) Forest/Tree; 5) Grass; 6) Shadow (see Table 2). It should be noted that, for example, buildings have different material of the roofs, therefore highly varying spectral characteristics of the material (tiles, concrete, highly reflecting metal, etc.) make difficult to classify such inhomogeneous objects into one class of interest. In this experiment the building class contains the following types of building roofs: tiles roof, concrete roof, dark color roof, green color metal roof, blue color metal roof, glass roof, highly reflecting roof, grass on the roof. Roads class contains the following types of pavements: asphalt pavement and concrete pavement. The ground truth for the area under investigation was proofed by the ATKIS vector map provided by Bavarian State Agency for Surveying and Geoinformation (Landesamt für Vermessung und Geoinformation). The number of clusters in the unsupervised clustering usually has value between 40 and 100. Experimental search found that the value of 80 provides significant dimensionality reduction with high accuracy of the land cover classification. A Multilayer Perceptron (MLP) was employed for the data fusion and classification implemented in the IDL. A feed-forward neural network based was employed. The network contains two hidden layers with 16 neurons in each layer. Training of the MLP made 1000 training epochs.

INFOFUSE classification maps for Munich city center are presented in Figure 6: WV-2 RGB composite (a), joint classification of WV-2 VNIR bands and TS-X features (b).

(a)   (b)
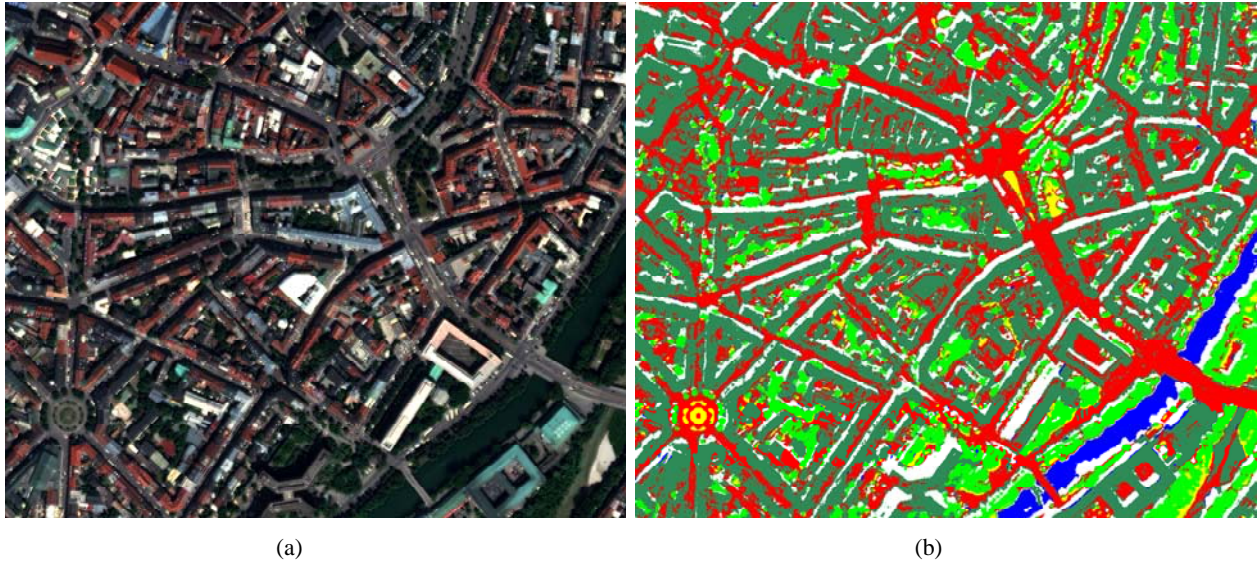
Figure 6. WV-2 RGB composite (a), INFOFUSE classification WV-2 VNIR+TS-X (b) for Munich city center.

Zoomed part of INFOFUSE classification for Munich city center is presented in Figure 7: WV-2 RGB composite (a), classification of WV-2 VNIR bands (b), classification of WV-2 VNIR bands and TS-X features (c). We see how the addition of SAR information helps to extract correctly the building seen in lower part of an image.

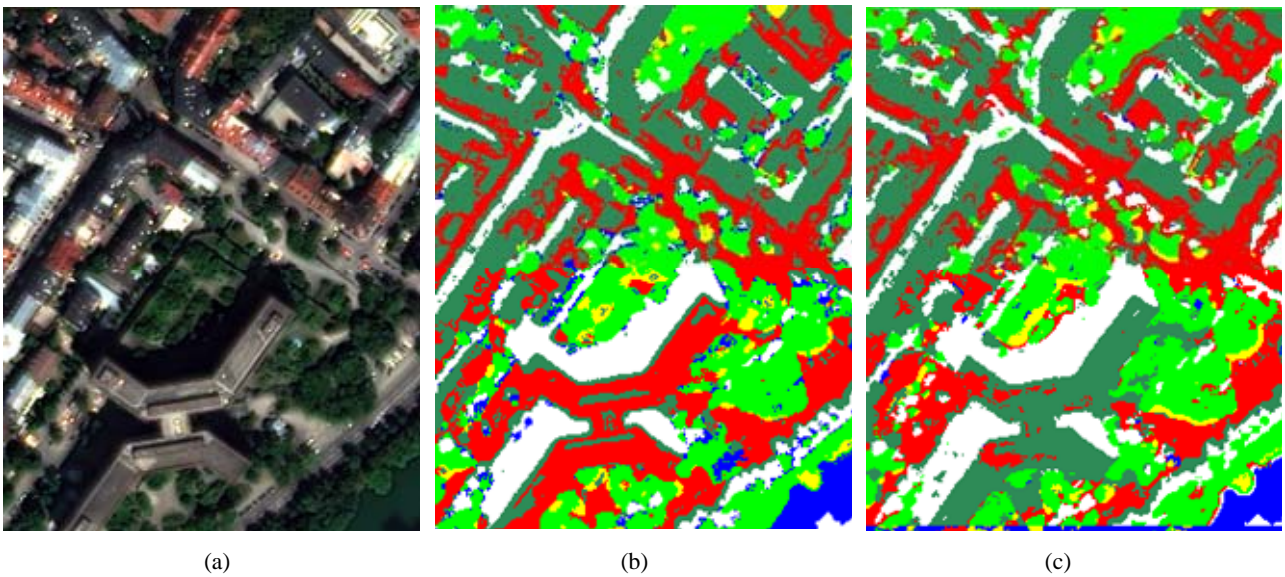(a)   (b)   (c)

Figure 7. WV-2 RGB composite (a), INFOFUSE classification WV-2 VNIR (b), INFOFUSE classification WV-2 VNIR+TS-X (c) for Munich city center zoom.

For the quantitative analysis of the quality of INFOFUSE classification confusion matrices are presented for WV-2 VNIR classification in Table 3 and for joint WV-2 VNIR and TS-X features classification in Table 4. When comparing the Tables we see that the addition of SAR information allows to decrease confusion between road and building classes significantly (compare numbers marked in red and green respectively).

Table 3. Confusion matrix of WV-2 VNIR classification.

Class	Water	Grass	Trees	Buildings	Road	Shadow
Water	**100.00**	0.00	0.00	0.00	0.00	0.00
Grass	0.00	**85.47**	0.00	0.00	0.00	0.00
Trees	0.00	14.53	**100.00**	0.00	0.00	0.00
Buildings	0.00	0.00	0.00	**60.05**	42.25	4.72
Road	0.00	0.00	0.00	39.95	**57.75**	0.00
Shadow	0.00	0.00	0.00	0.00	0.00	**95.28**
Total (%)	100.00	100.00	100.00	100.00	100.00	100.00

Table 4. Confusion matrix of WV-2 VNIR + TS-X features classification.

Class	Water	Grass	Trees	Buildings	Road	Shadow
Water	**100.00**	0.00	0.00	0.00	0.00	0.00
Grass	0.00	**98.20**	0.00	0.00	0.00	0.00
Trees	0.00	1.80	**98.90**	0.00	0.00	0.00
Buildings	0.00	0.00	0.00	**96.36**	3.91	0.47
Road	0.00	0.00	1.10	3.64	**96.06**	0.00
Shadow	0.00	0.00	0.00	0.00	0.00	**99.53**
Total (%)	100.00	100.00	100.00	100.00	100.00	100.00

6. CHANGE DETECTION

The classification using an INFOFUSE allows to obtain posterior probabilities of the classification map. Image area with a low probability is difficult to classify into any of the possible classes. This situation may be caused that these areas represent land cover classes not learned by the used classifier, or the used datasets provide different, insufficient, or changed information about the area. For example, the time gap between the acquisitions may be long enough and some changes in the area may happen (pavement change, construction of buildings, etc.) and different data types can represent different types of land cover for the same region. Such areas with low probabilities can not be classified with high certainty, therefore labeled as changed land cover.

CONCLUSIONS

This paper presents a multi-sensor data fusion method for urban area classification. The fusion model is based on information fission, dimensionality reduction, and information aggregation and employs relevant ways of multisource data combination. Utilized multi-sensor data (multispectral and SAR) allow to increase the number of classes and to boost the accuracy of the classification. The results of the classification are used for common land cover and urban classes. Multi-sensor data are processed separately and the fusion and classification method follows consensus rules of multisource data classification. The data classification is not influenced by the limitations of dimensionality and the

calculation complexity primarily depends on the step of dimensionality reduction. The shown method has also a high potential for the task of change detection, which is a matter for future research.

ACKNOWLEDGEMENTS

REFERENCES

[1] Deledalle, C.A., Denis, L. and Tupin, F., "Iterative Weighted Maximum Likelihood Denoising with Probabilistic Patch-Based Weights," IEEE Transactions on Image Processing, 18(12), 2661-2672 (2009).

[2] Fauvel, M., Chanussot, J. and Benediktsson, J. A., "Decision fusion for the classification of urban remote sensing images," IEEE Transactions on Geoscience and Remote Sensing, 44(10), 2828–2838 (2006).

[3] Haralick, R., Shanmugam, K. and Dinstein, I., "Textural features for image classification," IEEE Transactions on Systems, Man, and Cybernetics, 3(6), 610-621 (1973).

[4] Makarau, A., Palubinskas, G. and Reinartz, P., "Multiresolution Image Fusion: Phase Congruency for Spatial Consistency Assessment," Proc. ISPRS Technical Commision VII Symposium - 100 Years ISPRS, 5-7 July, 2010, Vienna, Austria, Wagner, W., Szekely, B. (eds.), IAPRS, XXXVIII(7B), 383-388 (2010).

[5] Makarau, A., Palubinskas, G. and Reinartz, P., "Multi-sensor data fusion for urban area classification," Proc. of Joint Urban Remote Sensing Event JURSE, 11-13 April, 2011, Munich, Germany, 1-4 (2011).

[6] Pacifici, F., Del Frate, F., Emery, W., Gamba, P. and Chanussot, J., "Urban mapping using coarse SAR and optical data: Outcome of the 2007 GRSS data fusion contest," IEEE Geoscience and Remote Sensing Letters, 5(3), 331–335 (2008)

[7] Palubinskas, G., "An unsupervised clustering method by entropy minimization," Maximum Entropy and Bayesian Methods, W. von der Linden et al., Eds., Kluwer Academic Publishers, Netherlands, 327-334 (1999).

[8] Palubinskas, G. and Datcu, M., "Information fusion approach for the data classification: an example for ERS-1/2 InSAR data", International Journal of Remote Sensing 29(16), 4689-4703 (2008).

[9] Palubinskas, G., Reinartz, P. and Bamler, R., "Image acquisition geometry analysis for the fusion of optical and radar remote sensing data," International Journal of Image and Data Fusion 1(3), 271-282 (2010).

[10] Palubinskas, G. and Reinartz, P., "Multi-resolution, multi-sensor image fusion: general fusion framework," Proc. of Joint Urban Remote Sensing Event JURSE, 11-13 April, 2011, Munich, Germany, 1-4 (2011).

[11] Reinartz, P., Müller, R., Schwind, P., Suri, S. and Bamler, R., "Orthorectification of VHR optical satellite data exploiting the geometric accuracy of TerraSAR-X data," ISPRS Journal of Photogrammetry and Remote Sensing, 66, 124-132, (2011).

[12] Suri, S. and Reinartz, P., "Mutual-Information-Based Registration of TerraSAR-X and Ikonos Imagery in Urban Areas," IEEE Transactions on Geoscience and Remote Sensing 48(2), 939-949 (2010).