

Alphabet-Based Multisensory Data Fusion and Classification Using Factor Graphs

Aliaksei Makarau, Gintautas Palubinskas, and Peter Reinartz

Abstract—The way of multisensory data integration is a crucial step of any data fusion method. Different physical types of sensors (optic, thermal, acoustic, or radar) with different resolutions, and different types of GIS digital data (elevation, vector map) require a proper method for data integration. Incommensurability of the data may not allow to use conventional statistical methods for fusion and processing of the data. A correct and established way of multisensory data integration is required to deal with such incommensurable data as the employment of an inappropriate methodology may lead to errors in the fusion process. To perform a proper multisensory data fusion several strategies were developed (Bayesian, linear (log linear) opinion pool, neural networks, fuzzy logic approaches). Employment of these approaches is motivated by weighted consensus theory, which lead to fusion processes that are correctly performed for the variety of data properties.

As an alternative to several methods, factor graphs are proposed as a new approach for multisensory data fusion. Feature extraction (data fission) is performed separately on different sources of data to make an exhausting description of the fused multisensory data. Extracted features are represented on a finite predefined domain (alphabet). Factor graph is employed for the represented multisensory data fusion. Factorization properties of factor graphs allow to obtain an improvement in accuracy of multisensory data fusion and classification (identification of specific classes) for multispectral high resolution WorldView-2, TerraSAR-X SpotLight, and elevation model data. Application and numerical assessment of the proposed method demonstrates an improved accuracy comparing it to well known data and image fusion methods.

Index Terms—Classification, factor graphs, fusion, graphical models, multisensor data, TerraSAR-X, WorldView-2.

I. INTRODUCTION

THE practical use of spaceborne very high resolution multispectral data (e.g., IKONOS, WorldView-2, or GeoEye-1) is still growing but the information gathered from the multispectral data is less in comparison to full spectral imaging. An improvement can be achieved by employment of fusion approaches with data from other sensors or sources since this may increase the quality of scene classification. Fusion of Synthetic Aperture Radar (SAR) and optical data is employed for several topics in remote sensing data interpretation, e.g., for landcover classification [1]–[3], change detection [4], object detection [5].

Manuscript received October 30, 2011; revised February 03, 2012, April 26, 2012, and August 31, 2012; accepted September 11, 2012. This work was supported by the German Aerospace Center and German Academic Exchange Service (DLR-DAAD) postdoctoral fellowship under Award A/09/95629.

The authors are with the German Aerospace Center (DLR), 82234 Wessling, Germany (corresponding author e-mail: aliaksei.makarau@dlr.de).

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

Digital Object Identifier 10.1109/JSTARS.2012.2219507

Incommensurability of different sources of data (e.g., optical, SAR, and DEM) requires a proper design of the fusion process. J. Benediktsson *et al.* [6], [7] first investigated statistical versus neural network approaches for multisensory data fusion and classification. Linear opinion pool and logarithmic opinion pool optimized by a multilayer neural network are proposed for the combination of multisensory data (multispectral, elevation, slope, aspect, and SAR). Several approaches for multisensory data fusion following consensus theory and employing different techniques such as Bayesian or neural networks were developed. F. Pacifici *et al.* [3] developed the best fusion algorithm for the 2007 GRSS Data Fusion Contest. The algorithm is based on a neural network classification enhanced by preprocessing and postprocessing. Principal component analysis is 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, Vegetation) provided a high Kappa coefficient equal to 0.9393. M. Fauvel *et al.* [8] applied decision fusion for classification of urban areas. The fusion approach consists of two steps. In the first step, data are processed by each classifier separately and the algorithms provide for each pixel membership degrees for the considered classes. In the second step, a fuzzy decision rule is used to combine the results provided by algorithms according to the classifiers' capabilities. The method is tested and validated with two classifiers on IKONOS images from 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 %.

F. Rottensteiner *et al.* [9] presented a method for building detection from a combination LIDAR data and multispectral images. They showed its applicability in a test site of heterogeneous building shapes. The method is based on the application of Dempster-Shafer theory for data fusion. The authors note that achieved results are satisfactory but in some cases buildings and trees could not be accurately separated, either because of shadows or because the resolution of the LIDAR data is not sufficient.

Graphical models (a type of factor graphs) were primarily employed for natural image classification and annotation. Fei-Fei Li *et al.* [10] proposed a new probabilistic directed graphical model for jointly modeling the image, its class label and its annotations. The model treats the class label as a global

description of the image, and treats annotation terms as local descriptions of parts of the image. Approximate inference and estimation algorithms based on variational methods were employed.

M. Lienou *et al.* [11] employed latent Dirichlet allocation model for the annotation of satellite images. The annotation task combines a supervised classification of the image patches and the integration of the spatial information between the patches. The model represents each patch of a collection as a random mixture of latent topics, where each topic is characterized by a distribution over words. The capability of the model is used to assign probabilities to unseen images to classify the patches of the large image into the semantic concepts. The authors in paper [12] developed a solution for bridging the gap between the results of classification algorithms and high semantic terminology of cartographic data. The Latent Dirichlet Allocation model is employed to map heterogeneous pixels with similar intermediate-level semantic meaning into land cover classes of various mapping products. A big problem in generating cartographic information from a fully automatic classification map is solved and demonstrated for Landsat images.

M. Datcu *et al.* [13] developed information mining system to retrieve remote sensing imagery from a database. The system allows semantic interpretation of the imagery using Bayesian networks. Unsupervised clustering of the features leads to obtaining an abstract vocabulary of signal classes. The vocabulary of signal classes is linked to the user-defined semantic landcover types using Bayesian networks.

The concept of factor graph (FG) was first devised in 1997 [14] and since then the application of FGs for signal/image processing and recognition is gradually emerging. B. Frey *et al.* [15] performed a thorough and exhaustive work on comparison of learning and inference methods for probabilistic graphical models (Bayesian networks, Markov random fields, factor graphs). For complex models that accurately describe many problems, direct application of Bayes rule leads to an intractable number of computations. A graphical model (factor graph) identifies the modules in the system and can be used to derive algorithms that achieve exponential speedups. Factor graphs subsume properties of Bayesian networks and Markov random fields. Any Bayesian network or Markov random field can be easily converted to a FG, without loss of information. Further, there exist models that have independence relationships that cannot be expressed in a Bayesian network or a Markov random field, but that can be expressed in a FG. FGs are more explicit about the factorization of the distribution than Bayesian networks and Markov random fields. Another advantage of FGs is that, because they explicitly identify functions, they provide a useful graph for message-passing algorithms, such as belief propagation [15].

Application of factor graphs for sensor fusion is performed by J. Moura *et al.* [16] on synthetic data. In this study a fusion of data collected by several heterogeneous sensors is performed to obtain a common goal. Successful fusion results were obtained on simulated scenarios with high sensing resolution and small

sensor network, or low sensing resolution and large sensor network (e.g., 150 sensors and 200 targets). Application on sonar data is performed by K. Kampa *et al.* [17]. In this work a dynamic factor graph is employed for data fusion (3 real targets taken by 2 sensors) and segmentation (in the sense of probabilistic framework). The linear-time inference is achieved on a tree-structured network using the sum-product algorithm. R. Naphade *et al.* [18] employed factor graphs for semantic annotation of video sequences. A factor graph is used for mapping low-level features to high-level semantics. To reduce the complexity and computational cost of the factor graph (probability function is exponential in the number of variables), a factorization of the function is enforced (instead of one joint probability mass function, several mass functions of two argument variables were used).

Remotely sensed data have a high complexity and the processing or interpretation requires a definition of a model. The model should have properties like, for example, an established mathematical basis, tractable learning and inference, good generalization capabilities, calculation time, absence of curse of dimensionality. Explicit factorization properties of factor graphs allow to compose complex models from simpler modules and establish links among the simpler modules using the rules of probability theory. Approximate inference methods applied for factor graphs allow to receive fast and plausible decisions; plausible decisions by inference on non-full data are possible to perform. These properties give a good motivation to employ factor graph model for data and imagery interpretation especially in remote sensing applications. Nevertheless, factor graphs are not widely employed yet especially for remotely sensed data processing and interpretation.

This research is motivated by previous works of the authors on single-/multisensory data classification [19] of satellite data. The used method employs input data/feature representation (separately for each data source) using an unsupervised clustering. k -means clustering or entropy based k -means (allows automatic computation of a proper number of clusters [20]) can be used. The unsupervised clustering allows to combine objects and structures with similar (i.e., spectral or textural) properties, reduce the size of the data, and make a higher level of feature abstraction. Since the represented data have a finite number of states a proper fusion method following consensus theory should be used. Considering the properties, factor graphs were chosen to perform multisensory data integration.

This paper is organized as follows. In Section II a detailed introduction and description of the proposed method for multisensory data fusion is given. Section III gives a short introduction to factor graphs, presents description of relationships of variables and factors in the factor graph, and a proper model of the factor graph is selected. Information on configuration and inference methods is also given. Description of the employed multisensory data, fusion strategies, results of the experimental part and discussion are given in Section IV. Conclusion and prospective development of the fusion model are given at the end of the paper.

II. ALPHABET-BASED DATA FUSION

Different types of data (qualitative and quantitative) are represented using different scales, e.g. nominal scale (color, flavour, or specie, an unordered set of qualitative “values”, numbers are used as labels and express no mathematical properties); ordinal scale (the result of qualitative or quantitative data, e.g., student rank in a school, numbers indicate the relative position of items); interval scale (opinion scale, time difference, or date in calendar, numbers indicate difference between items, there is no absolute zero point); ratio scale (length in meters, duration in hours, numbers indicate difference and there is a fixed zero point); absolute scale (the number of books in a library, numbers are properties of the attribute).

Performing fusion and classification of features represented on different scales and obtained from different acquisition devices (e.g., the features possess different statistical properties and distributions) may be difficult using parametric statistical methods and may lead to errors. An assumption on a parametric distribution for the employed features may not always hold. To overcome these difficulties we perform feature transformation to make an intermediate representation of a feature (e.g., assume a categorical (multinomial) distribution [21] separately for each transformed feature). This transformation (representation on the alphabet, or a vocabulary of signal classes [13]) makes another level of feature abstraction, generalized by similarity in the feature space, simultaneously performing data reduction. k -means unsupervised clustering is the most popular and easy way to perform this transformation and used in works [11], [13], [19], [21], [22]. The number of clusters can be selected empirically or by using approaches to determine the optimal number. In work [11] the optimal number of clusters is selected by modeling the features as a Gaussian mixture and using the minimum description length criteria to accede to the optimal complexity of the model [23]. The unsupervised clustering method by entropy minimization [20] performs clustering of data with automatic definition of a proper number of clusters.

Factor graph is a discrete graphical model (each variable can take any value of a finite predefined domain) therefore the input features can only be discrete. A trade-off between the number of clusters and the accuracy of fusion and classification is investigated. A factor graph is employed to integrate the multisensory data represented in a finite domain. Employment of this type of graphical models is motivated by the fact that factor graphs were used effectively in a wide range of application areas such as Decoding of codes, Behavioral modeling, Probabilistic modeling, Fast Fourier transform [14].

Detailed explanation of the fusion framework is given in the following.

A. Framework

The fusion framework consists of three main steps:

- 1) **Information fission: feature extraction from input datasets.** The aim of this step is to extract information and to make a full description of the input data (to provide the quasi-full description) [19]. For each data source relevant

features are extracted. These features are expected to characterize different properties of structures and objects in each data source. After feature extraction a large amount of redundant information is obtained.

- 2) **Feature representation on an alphabet.** The aim of this step is to represent a feature on a finite predefined domain—alphabet. The number of states of the feature value is reduced, i.e., a kind of “quantization” is performed. The objects with similar properties are combined and the data size is decreased. This representation can be made using several methods, e.g., unsupervised clustering. k -means clustering is used. All features extracted from multisensory data are processed in this way and a vector of features is composed. This vector is used as input evidence for a configured (learned on training data) factor graph.
- 3) **Fusion and classification of coded data or features** is performed using the factor graph [14], [24]. Configurations (parameter sets) of the FG are calculated according to supervisory selected classes and training areas. Configured FG is used for inference on evidence data (i.e., clustered input features). The posterior probability (maximum principle) or calculated marginal distribution of a latent variable is employed for data classification.

The overall scheme is given in Fig. 1. Different feature types can be extracted to make an exhausting description of the data. For example, a multispectral image can be used for extraction of spectral information, DVI indexes and texture features (Haralick, Gabor, or Laws). For some data sources (e.g., DEM) feature extraction is not carried out and the data are directly represented on the alphabet. The size of the alphabet should be appropriately defined for data or features (multispectral, textural, or DEM). At the final step the evidence (a code composed from features represented on an alphabet) is given to factor graph (factor graph is configured for several classes) and a class label is selected according to the maximum likelihood probability.

III. FACTOR GRAPH DEFINITION

A. Feature Representation on Finite Alphabet

Input data and extracted features have varying value ranges, statistics, and physical nature. For example, WorldView-2 has 11-bit coded multispectral data, TerraSAR-X has 16-bit coded data, while other sources of data and extracted features may have other value ranges. Clustering of input data is performed to fit the nature and properties of the data fusion method or classifier (e.g. neural network, Bayesian network, Markov random field, or Factor graph).

The aim of such feature coding is to represent a feature using a predefined finite domain (alphabet). The finite domain refers to the unique values (or a list of values) the feature can have. Here we use a finite domain consisting of natural numbers. To represent the input data source (or a feature) on the finite domain, the data are proposed to be processed by unsupervised clustering [20]. A cluster’s number is assumed as the value from the defined domain. The number of clusters is set empirically or can be automatically computed by the clustering method (according

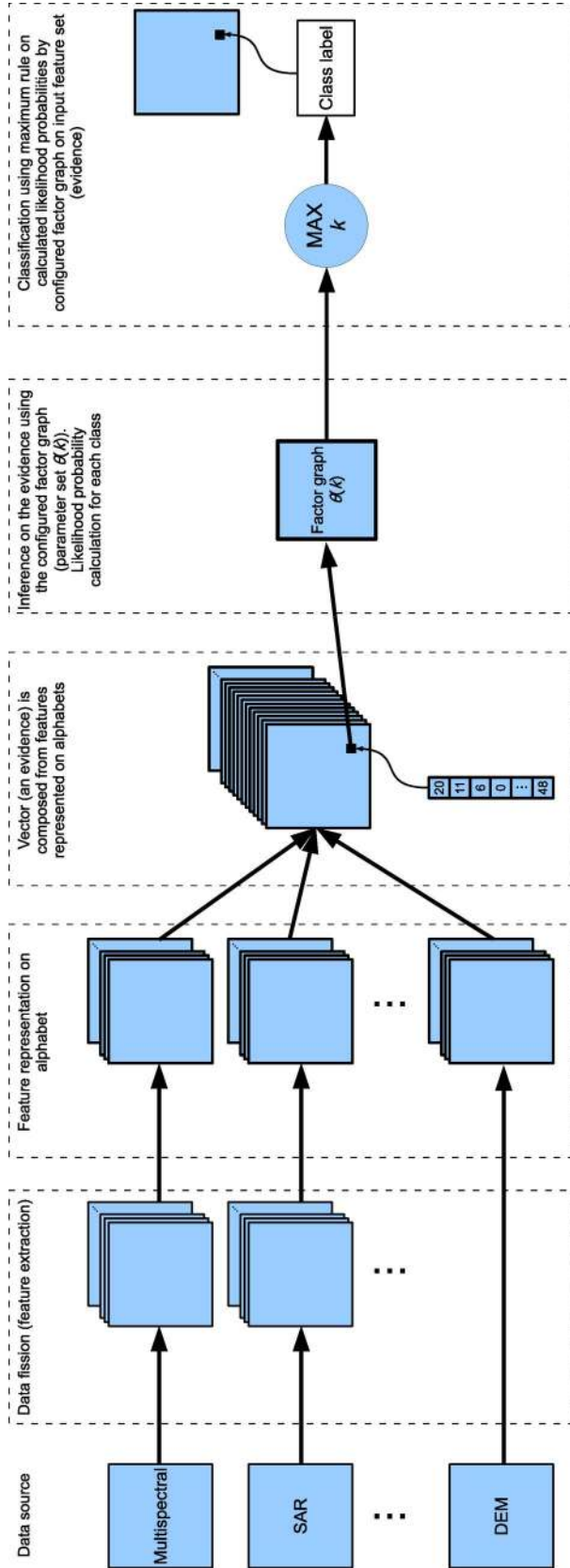


Fig. 1. Inference (classification) is performed using a configured factor graphon evidence (input features for the graph). After belief propagation on the graph the likelihood probability is calculated and a pixel is labeled by the class k that has the maximum probability (minimum energy) of the configured factor graph.

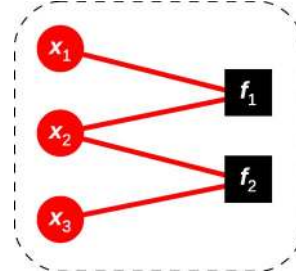


Fig. 2. An illustrative example of simple factor graph with three variables x_1, x_2, x_3 and two function nodes $f_1(x_1, x_2)$ and $f_2(x_2, x_3)$.

to the scene complexity) and equals to the size of the domain. Unsupervised clustering allows to group objects in the data source by the properties described by the sensor (intensity, DN value, or spectral properties) and to reduce the complexity of the data set.

After clustering, each feature is represented on a domain with a particular size (the number of clusters can be different and should be defined for each feature) and a vector is composed from the represented features. This vector (code vector, under some assumptions) is the input data (evidence) for a factor graph.

B. Factor Graph for Discrete Data

Factor graphs are more general graphical models than Bayesian networks or Markov random fields (according to the Hammersley-Clifford theorem [25], Bayesian networks and Markov random fields are particular cases of corresponding factor graphs). A FG possesses properties of Bayesian network and Markov random field and allows to describe relationships among parts of a modeled system using the rules of probability theory. Probability propagation in a Bayesian net is equivalent to the application of the sum-product algorithm to the corresponding factor graph [18]. FGs were found efficient for discrete-valued data recognition in various applications [16], [26], [27].

A factor graph is a bipartite graph containing two types of nodes: variable nodes ($x_i, i = 1 \dots n$) and function nodes (factors) ($f_j(x_1, x_2, \dots, x_n), j = 1 \dots m$), where a variable node x_i takes value on a finite domain (alphabet A_i) [14]. A variable node x_i is connected to a factor node f_j if and only if x_i is an argument of the f_j . A factor $f_j(x_1, x_2, \dots, x_n)$ is a function of the variables x_i with a configuration space O :

$$O = A_1 \times A_2 \times \dots \times A_n. \quad (1)$$

Fig. 2 presents an example of a factor graph with three variables x_1, x_2, x_3 and two function nodes f_1 and f_2 with factorization: $g(x_1, x_2, x_3) = f_1(x_1, x_2) * f_2(x_2, x_3)$.

The sum-product algorithm [14] works by computing messages at the nodes using a simple rule and then passing the messages between nodes according to a selected schedule [18]. A message from a function node to a variable node is the product

of all messages incoming to the function node with the function itself, summarized for the variable associated with the variable node. A message from a variable node to a function node is simply the product of all messages incoming to the variable node from other functions connected to it. Consider a message on the edge connecting function node f to variable node x . Let $\text{msg}_{x \rightarrow f}(x)$ denote the message sent from node x to node f in the operation of the sum-product algorithm. Also let $\text{msg}_{f \rightarrow x}(x)$ denote the message sent from node f to node x . Further, let $\text{neigh}(v)$ denote the set of neighbors of a given node v in a factor graph and let \downarrow indicate the summary operator [14], [18]. Consider a function $f(x_1, x_2, x_3)$ then a possible summary operator could be the summation operator in (2):

$$f(x_1, x_2, x_3) \downarrow x_1 = \sum_{x_2, x_3} f(x_1, x_2, x_3). \quad (2)$$

The message computations performed by the sum-product algorithm can be expressed as follows. Variable to local function:

$$\text{msg}_{x \rightarrow f}(x) = \prod_{h \in \text{neigh}(x) \setminus \{f\}} \text{msg}_{h \rightarrow x}(x), \quad (3)$$

Local function to variable:

$$\begin{aligned} & \text{msg}_{f \rightarrow x}(x) \\ &= \left(f(x_{\text{neigh}(f)}) \prod_{h \in \text{neigh}(f) \setminus \{x\}} \text{msg}_{h \rightarrow f}(h) \right) \downarrow x. \quad (4) \end{aligned}$$

In the case where a local function $f_j, j = 1 \dots m$ has arguments from subset $X_j, j = 1 \dots m$ (each f_j has a particular subset X_j of variables $\{x_1, x_2, \dots, x_n\}$), the product (global function, $g(x_1, x_2, \dots, x_n)$) of local functions f_j can be expressed as:

$$g(x_1, x_2, \dots, x_n) = \prod_j f_j(X_j). \quad (5)$$

C. Factor Graph Structure and Relationships Definition

The task of classification consists of determining the probability of a particular hypothesis given some observed evidence. Usually, this is solved by calculation of the marginal probability of a latent variable, or by calculating the posterior probability (likelihood on the configured factor graph given the evidence):

$$\arg \max_k P(c_k | E), \quad (6)$$

where the c_k is the class, E is the evidence (the evidence is a set of features: $E = \{x_1, x_2, \dots, x_n\}$).

Bayes rule allows to expand this rule to:

$$\arg \max_k P(c_k) P(E | c_k) / P(E), \quad (7)$$

where the $P(E)$ is the evidence E prior probability (fixed during inference), $P(c_k)$ is the k -th class prior probability (the probability can be flat over classes).

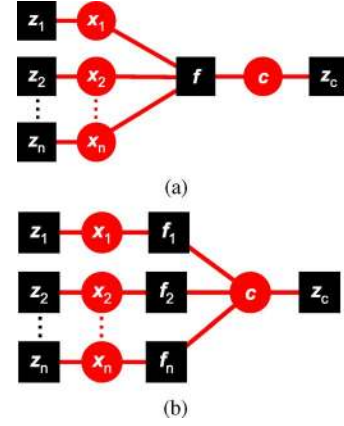


Fig. 3. Fusion modelling using factor graph: (a) the exact model (high complexity of learning and inference due to one configuration factor); (b) the independent model (low complexity).

Assuming that the factor graph is configured ($\theta(k)$ is the configuration for class k) the joint distribution of $P(E | c_k)$ is the following [28]:

$$\arg \max_k P(c_k) P(E, \theta(k) | c_k). \quad (8)$$

Conditioning on $\theta(k)$ we can write:

$$\arg \max_k P(c_k) P(\theta(k) | c_k) P(E | c_k, \theta(k)). \quad (9)$$

The following models can be defined for (7): the exact model (Fig. 3(a)) and the independent model (Fig. 3(b)). The exact model assumes that the class configuration is a single factor with dependent features. The independent model assumes that each input feature is independent.

1) *Exact Model*: Assume that we have n coded input features (denote as x_i) as an evidence E . The exact model factor graph g_{exact} (Fig. 3(a)) can be written as:

$$\begin{aligned} & g_{\text{exact}}(x_1, x_2, \dots, x_n) \\ &= z_c(c_k) f(x_1, x_2, \dots, x_n, c_k) z_1(x_1) z_2(x_2) \dots z_n(x_n) \quad (10) \end{aligned}$$

where the variables x_1, x_2, \dots, x_n are the input features (evidence E), c_k is the k -th class, z_c is the factor defining prior probability of class variable c_k , $f(x_1, x_2, \dots, x_n, c_k)$ is the factor of input features and class variable c_k , $z_1(x_1) z_2(x_2) \dots z_n(x_n)$ are the factors defining prior probabilities $(1)/(p(x_1)), (1)/(p(x_2)), \dots, (1)/(p(x_n))$ (normalizing factors).

Message updating for the exact model factor graph (Fig. 3(a)) is described in Fig. 4.

2) *Independent Model*: The independent model factor graph $g_{\text{independent}}$ (Fig. 3(b)) can be written as:

$$\begin{aligned} & g_{\text{independent}}(x_1, x_2, \dots, x_n) \\ &= z_c(c_k) f_1(x_1, c_k) f_2(x_2, c_k) \dots f_n(x_n, c_k) \\ & z_1(x_1) z_2(x_2) \dots z_n(x_n) \quad (11) \end{aligned}$$

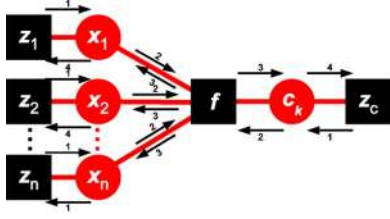


Fig. 4. Message update diagram for function node $f(x_1, x_2, \dots, x_n, c_k)$ (an equivalent to the joint mass function), the z_c node defining the prior probability of class variable c_k , and the normalization functions $z_1(x_1)z_2(x_2)\dots z_n(x_n)$ in the exact model factor graph (Fig. 3(a)) (also see Naphade *et al.* [18]).

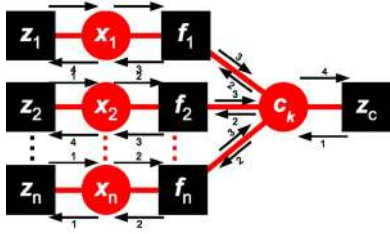


Fig. 5. Message update diagram for function nodes $f_1(x_1, c_k), f_2(x_2, c_k), \dots, f_n(x_n, c_k)$, the z_c node defining the prior probability of class variable c_k , and the normalization functions $z_1(x_1)z_2(x_2)\dots z_n(x_n)$ in the independent model factor graph (Fig. 3(b)).

where the factors f_1, f_2, \dots, f_n are the factors of the features x_1, x_2, \dots, x_n and share the class variable c_k ; c_k is the k -th class variable; z_c is the factor defining the prior probability of class variable c_k .

Message updating for the independent model graph (Fig. 3(b)) is described in Fig. 5. Detailed step-by-step message propagation map for the independent model is given in the Appendix.

Different topologies of the graph pose some difficulties and advantages on the graph use. A single configuration factor (f) in the exact model allows to perform exact inference but makes learning and inference problematic on this graph due to its high complexity, while the independent model is an equivalent to a tree-structured Bayesian network and has low complexity leading to easier configuration and inference.

Several modifications of the exact model can be made in order to allow usage of the arbitrary number of input features:

- 1) *Factorization of functions.* The factor f in (10) has exponential nature on the number of input variables (features), therefore the cost of computation increases quickly. Increasing the number of input features as well as the alphabet size, the factor graph will obtain a high and intractable state space and the configuration of this factor graph as well as inference may become difficult to perform. To overcome this disadvantage a factorization of mass functions can be done (e.g., [18]).
- 2) *Features as binary variables.* Additionally an input feature can be represented using several binary variables to reduce the state space of an input feature variable (then the number of input feature variables increases).

D. Model Learning: Graph Configuration

The structure of a factor graph (the exact or independent model) defines a dependency of class variable node c_k on input features x_1, x_2, \dots, x_n (evidence E). The probability $p(c_k|E)$ can be calculated using the configured (learned) factor graph. The use of training data allows to calculate a configuration (parameter set $\theta(k)$) for the factor graph (the $\theta(k)$ is calculated to maximize the $p(c_k|E)$, (6)–(9)).

An important issue is a proper way to configure (make supervised learning) the graph. The configuration of the FG consists of two steps. In the first step we maximize the total likelihood subject to all the factors (estimated factors are $z_1, \dots, z_n, f_1, \dots, f_n$, and z_c) using training data for all the classes (the number of samples for each class should be almost equal). In the second step, the factors z_1, \dots, z_n and z_c (prior probabilities over the features and class) are fixed and the factors f_1, \dots, f_n are estimated for each class k using only training data for the class k . Such a two step configuration procedure allows the comparison of factor graph (parametrized by $\theta(k)$) posterior probabilities for the classes using the maximum rule.

The gradient ascent method is employed for the calculation of the graph configuration $\theta(k)$. An inference on the configured factor graph given the evidence data allows to calculate a probability. A comparison of the calculated probabilities allows to make a classification of the input scene (maximum rule is used).

E. Inference

Usually the task of inference means one of two scenarios: 1) to compute a configuration of latent variables to maximize the posterior probability (in our task a latent variable is c_k^* and the task is to maximize the $p(c_k^*|E)$); 2) to compute the marginal distribution $p(c_k|E)$ for a single node, or marginal distributions over sets of nodes: $p(c_{(s)}|E)$ (see [29]).

The task of marginalization could be difficult to perform because of a high complexity of the model together with a high number of input features (the number of variables can be up to 300) as well as a quite high number of classes (23 classes in this work). In order to avoid such marginalization problems the first scenario (maximization of the posterior probability) is selected. Mean field inference [15] is employed.

IV. EXPERIMENTS, RESULTS, AND DISCUSSIONS

The main aim of this section is to illustrate fusion and classification advantages of the proposed method in comparison with other known methods. Fusion and classification accuracy of a method can be numerically assessed on multisensory data. The same multisensory data and fixed training and test samples are employed by all compared fusion methods and numerical measures are calculated using the test samples. Comparison of the numerical assessment results can reveal the method providing the best accuracy. Standard assessment measures such as overall accuracy, Cohen's Kappa, McNemar's test, and calculation of confusion matrices are used for the

TABLE I
PARAMETERS OF THE WORLDVIEW-2 AND TERRASAR-X DATA FOR MUNICH TEST SCENE

Parameter	WorldView-2 [33]	TerraSAR-X
Product	Ortorectified Standard Imagery	EEC
Sensor mode	Multispectral, PAN	Spotlight HS
Orbit	Descending	Descending
Acquisition date and time (UTC)	10 July 2010, 10:30:17	7 June 2008, 05:17:48
Look angle	5.2°, Left	49.2218°, Right
Ground pixel size, m	0.5 × 0.5	0.5 × 0.5
Polarization	-	Single, VV
Bits per pixel	11	16

comparison. Well known and employed methods such as maximum likelihood classification (based on statistics, not following consensus theory) and neural network (widely used for multisensory data fusion, following consensus theory) are selected for comparison. Single sensor (multispectral WorldView-2) and multisensory data (WorldView-2+Elevation or WorldView-2+Texture (SAR and Optical)+Elevation) were employed for comparison of fusion and classification accuracy. Experimental analysis is run on two multisensory data sets (WorldView-2, TerraSAR-X, and elevation data) acquired for Munich (23 classes) and London (14 classes) cities.

A. Munich Test Area

1) *Multisensory Data and Features*: For an experimental evaluation, a combination of very high resolution (VHR) satellite data from the optical spectrum (WorldView-2) and the microwave range (TerraSAR-X) together with a digital surface model (DSM) are used. Multispectral data allow to classify objects by spectral properties, but the single-angle acquisition images do not allow to reveal the elevation of an object, therefore some classes can be confused (e.g., asphalt material can belong to a road or can be a roof material). SAR data (here X-band is assumed) allows to assess the homogeneity, object surface structure, and other properties of the objects. This kind of information can be used for proper description and delimitation of objects with similar spectral properties. The DSM data allows to obtain information on object height in the scene, but employment of the DSM for scene classification can not allow to define a high number of classes. An adequate fusion method of the multisensory data allows to detect a higher number of landcover classes with increased accuracy. Urban area (Munich city) is especially selected for data fusion because it contains a variety of objects, structures and allows to define a high number of landcover classes.

Acquisition geometries of the employed WorldView-2 (5.2° look angle, almost nadir view) and TerraSAR-X (49.2218° look angle) data allow to use the multisensory data for fusion [30]. WorldView-2 (WV-2) multispectral data were pan-sharpened by the General Fusion Framework method [31]. The optical and SAR data were orthorectified (SRTM 30 m DEM) and distortions introduced by terrain are decreased. Ortorectified WorldView-2 and SpotLight Level-1B Product TerraSAR-X (TSX) data were used. A detailed description of the employed data sets is given in Table I. The registration of optical and radar data is made in ENVI using manual selection of control points.

In more complicated cases other registration methods should be employed, e.g., [32].

Highly detailed Digital Surface Model (DSM) of urban scene is generated using the Semiglobal Matching algorithm using in this case two Worldview-2 stereo pairs with small convergence angles (less than 20 degrees) [34].

Specific features should be extracted from the data to make an exhaustive description of landcover classes, structures, and objects. For example, a multispectral image can be used for extraction of spectral information, Difference Vegetation Index (DVI) indexes, well known texture features (Haralick features, Gabor, or Laws), while the TSX data is primarily suitable for extraction of texture features to describe specific properties of objects. For some data sources (e.g., DSM) feature extraction is not carried out and the data are directly represented on the domain. The size of the domain should be appropriately defined for each different feature (multispectral, textural, or the DSM).

The TSX image is employed for the characterization of surface and textural properties. Optical WV-2 data are also used for textural feature extraction and for providing spectral information on the objects of the scene. In our experiment Gabor features [35] were calculated on TSX data and on Red color channel (630–690 nm) from WV-2 data. A bank of Gabor wavelets consists of 18 filters (6 orientations ($0, \pi/6, \pi/3, \pi/2, (2)/(3)\pi, (5)/(6)\pi$), 3 different periods of filter's sine component ($\pi/3, (2)/(3)\pi, \pi$), and 1 sigma value ($\sigma = 4$)). A recursive implementation of Gabor filtering is employed [36].

A subscene (7115×4516 pixels) is used in the experiments. k -means clustering is employed for feature representation on the alphabet. In this experiment we try to set the size of the alphabet as small as possible to illustrate that the represented data allows to reach a higher accuracy of the fusion and classification comparing to the full range of values. We assume that multispectral data requires a larger alphabet size to preserve more spectral information, while the texture and the DSM data can be represented on the alphabet with a smaller size. The number of clusters is selected empirically and set to 10 for Gabor features (calculated on SAR and optical data), 10 for the DSM, and 50 for the WV-2 multispectral data.

2) *Fusion Strategies and Classification*: One of the main interests is to assess the influence of multisensory data employment on classification accuracy and to compare the fusion results with the results of single sensor data classification. Quantitative assessment allows to illustrate usefulness of multisensory

TABLE II

MUNICH TEST AREA. CLASS OF INTEREST, ACRONYM, NUMBER OF TRAINING AND TEST SAMPLES *-VALIDATED GROUND TRUTH CREATED BY DR. WIEKE HELDENS (AVAILABLE ONLY FOR A LIMITED NUMBER OF OBJECTS), †-VALIDATED GROUND TRUTH SELECTED FROM ATKIS VECTOR MAP, ‡-TRAIN AND TEST AREAS WERE SELECTED USING VISUAL INTERPRETATION OF WV-2 VISIBLE RANGE DATA TOGETHER WITH BING MAPS, GOOGLE EARTH, AND GOOGLE STREET VIEW. PER CLASS FUSION AND CLASSIFICATION ACCURACIES OF THE ML, NN, AND FG METHODS ON WV-2, WV-2+DSM, AND WV-2+TEXTURE+DSM

	Class	Acronym	Training samples	Test samples	WV-2			WV-2+DSM			WV-2+Texture+DSM		
					ML	NN	FG	ML	NN	FG	ML	NN	FG
1	Water†	WA	13231	14335	81.35	98.71	72.95	83.01	100	76.61	88.92	99.99	91.52
2	Forest/trees†	FO	10564	27730	99.45	99.60	79.07	96.72	87.99	82.33	98.71	86.91	96.83
3	Grass/Low vegetation†	GR	2654	32896	43.25	53.26	73.43	25.58	34.47	62.31	11.42	39.43	77.52
4	Bare soil‡	BS	7180	14046	84.15	67.53	0.15	57.36	62.32	1.96	61.40	64.62	92.99
5	Construction site‡	CS	15258	17693	70.01	84.45	63.42	74.29	82.33	95.99	88.11	86.47	77.11
6	Swimming pool†	SP	2667	3489	100.0	99.97	99.97	100.0	100	99.97	98.37	100.0	99.46
7	Asphalt road†	RO	2681	60033	58.23	51.38	29.02	59.65	56.67	31.62	40.33	69.38	71.37
8	Football field†	FF	21829	27263	91.82	98.06	39.32	99.45	94.97	43.28	94.66	93.69	70.34
9	Tennis field†	TF	4722	3300	99.27	100	98.18	90.76	99.94	98.18	93.85	99.55	98.88
10	Green house‡	GH	11920	12734	91.97	93.09	0	92.98	97.56	0.55	94.33	95.79	16.05
11	Rail road†	RR	5888	23279	68.18	79.44	50.10	61.96	82.61	71.01	43.15	89.65	86.71
12	Tram line†	TL	1009	1017	27.83	34.32	42.18	34.81	20.35	55.06	12.19	29.89	44.05
13	Cemetery†	CE	4195	8481	94.67	86.15	60.75	89.66	74.64	64.17	90.17	84.51	86.51
14	Parking/car‡	PA	3882	13628	61.43	37.61	0	54.40	29.49	3.00	59.30	52.83	93.33
15	Shadow‡	SH	3551	8473	61.93	93.21	98.48	64.67	99.24	98.48	50.40	98.19	84.86
16	Concrete‡	CO	46	137	2.92	80.29	72.99	2.92	82.48	78.10	0	78.10	100
17	Red roofing tiles*	R-RT	915	1050	98.95	99.71	94.19	99.24	100	99.24	0	99.81	57.62
18	Roofing concrete*	R-CO	887	1198	40.23	87.65	12.44	31.47	50.17	15.94	0	50.08	39.57
19	Vegetation roof*	R-VE	1605	1723	93.73	93.62	74.17	98.08	99.71	99.01	0	100.0	62.86
20	Dark roofing tiles*	R-DT	1164	2563	63.87	89.70	55.48	29.50	44.83	48.61	0	30.28	45.81
21	Zinc roof*	R-ZI	968	6418	75.77	89.98	26.94	71.42	65.36	32.36	0	65.13	28.22
22	Roofing copper*	R-CP	912	953	93.81	88.98	85.31	94.86	98.74	85.31	0	97.80	79.85
23	Grey roofing tiles*	R-GT	1354	2492	86.12	75.60	69.82	95.30	96.07	75.28	0	94.02	70.99

sory data for proper landcover class identification. Comparison of different fusion and classification methods on multisensory data is also of an interest. Availability of 8-band WV-2 multispectral data allows to increase the accuracy of fusion and classification comparing to well known IKONOS, Quickbird, or Geoeye sensors acquiring only 4-bands. The following combinations of multisensory and single-sensor data can be created:

- 1) WV-2 (8 features (spectral bands)),
- 2) WV-2+DSM (9 features),
- 3) WV-2+Texture (SAR and Optical)+DSM (45 features).

Single data sets like DSM or only texture were not selected for comparison since the data do not allow to obtain a comparable classification accuracy. Altogether, 23 classes were defined, the number of training and test samples is given in Table II. Selection of training and test regions is made manually on a color composite of WorldView-2 according to available ground truth data. The training and test samples are spatially uncorrelated. It should be noted that the validated ground truth is limited in size (e.g., vector data on class Nr. 12 and 16–23 is available only for a small number of objects and buildings). The ground truth for the area under investigation is proofed by the ATKIS vector map provided by Bavarian State Agency for Surveying and Geoinformation (Landesamt für Vermessung und Geoinformation). Vector data on the materials available in the scene is created and provided by W. Heldens [37].

3) *Results and Discussion:* Table III presents results (overall accuracy and Cohen's Kappa) on fusion and classification of single sensor and multisensory data. Results of two other methods: Maximum Likelihood (ML) (not following consensus theory) and Neural Network (NN) are also given for comparison. All the methods were run on the same feature set and the same training/test regions were employed. Neural

TABLE III

MUNICH TEST AREA. CLASSIFICATION ACCURACY USING DIFFERENT METHODS TOGETHER WITH THE PROPOSED APPROACH. ML-MAXIMUM LIKELIHOOD (NOT FOLLOWING CONSENSUS THEORY), NN-NEURAL NETWORK, FG-FACTOR GRAPH. OVA—OVERALL ACCURACY, KAPPA—COHEN'S KAPPA

Method	Employed features	OVA, %	Kappa
ML	WV-2 (8)	72.51	0.7010
ML	WV-2+DSM (9)	68.98	0.6620
ML	WV-2+Texture+DSM (45)	58.53	0.5502
NN	WV-2 (8)	75.00	0.7282
NN	WV-2+DSM (9)	71.18	0.6867
NN	WV-2+Texture+DSM (45)	76.33	0.7411
FG	WV-2 (8)	47.89	0.4536
FG	WV-2+DSM (9)	52.46	0.4909
FG	WV-2+Texture+DSM (45)	76.75	0.7438

network (multilayer perceptron) is chosen since it is shown to be an efficient and popular solution for multisensory data fusion and provides a high classification accuracy [3], [6]. The Neural Network is a 3 layer (2 hidden layers) feed-forward net trained with Kalman filter, implemented in IDL [38]. The input features are normalized to one. The number of neurons in the hidden layer is selected experimentally. Running the neural network on the same feature set (WV-2+Texture (SAR and Optical)+DSM) and employing different number of neurons (5, 10, 20, 60, and 80) in the hidden layers it is found that the three layer network with 20 neurons in the two hidden layers allows to obtain the best result on fusion and classification (Table IV). The same number of neurons is employed for the WV-2 and WV-2+DSM feature sets. For the ML classification ENVI software package is used.

The independent model of the factor graph is selected for fusion and classification since less calculation time is required

TABLE IV

MUNICH TEST AREA. CLASSIFICATION ACCURACY USING THE NN ON WV-2+TEXTURE (SAR AND OPTICAL)+DSM DATASET (ALL TOGETHER 45 FEATURES). THE NN IS A 2 LAYER NETWORK (1 HIDDEN LAYER) OR A 3 LAYER NETWORK (2 HIDDEN LAYERS) WITH A DIFFERENT NUMBER OF NEURONS IN THE HIDDEN LAYER/LAYERS (5, 10, 20, 60, AND 80) ARE EMPLOYED TO FIND THE MOST SUITABLE ARCHITECTURE OF THE NETWORK AND REACH THE HIGHEST FUSION AND CLASSIFICATION ACCURACY

Number of neurons	1 hidden layer		2 hidden layers	
	OVA, %	Kappa	OVA, %	Kappa
5	72.17	0.6966	73.67	0.7118
10	75.83	0.7357	75.70	0.7338
20	72.53	0.7005	76.33	0.7411
60	74.88	0.7251	74.15	0.7176
80	55.13	0.5158	75.21	0.7288

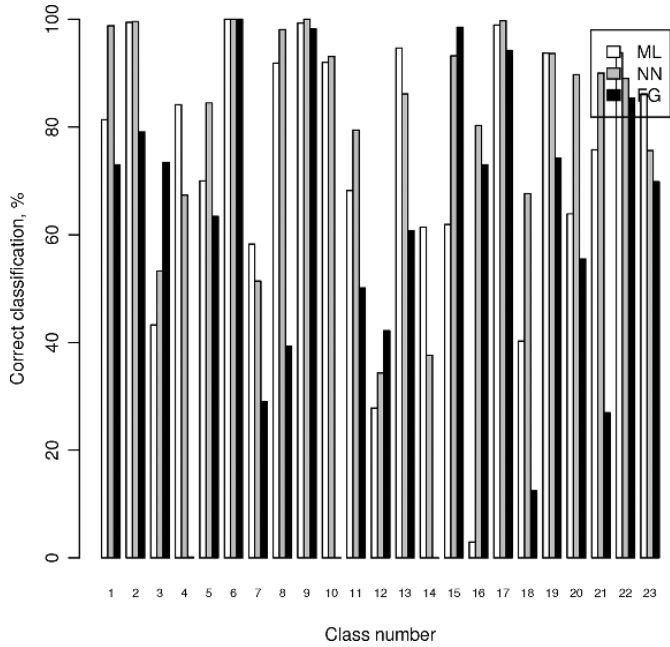


Fig. 6. Munich test area. Per-class accuracy for the WV-2 data classification using the ML, NN, and FG.

for the configuration and inference. Estimation of marginal distribution for class variable can be used when there is a relatively low number of input features and the number of classes is not high. Employment of a high number of input features and classes (e.g., 45 features and 23 classes) makes it problematic to estimate the posterior marginal of class variable (confusion among classes is high). To avoid the confusion among classes, total maximum likelihood probability of the factor graph is calculated to produce a probability map separately for each class, and maximal MAP selection allows to assign the class labels.

Figs. 6, 7, and 8 illustrate the fusion and classification accuracy using the defined combinations of multisensory data. Introduction of the textural features and the DSM allowed to increase the overall accuracy for the NN and the FG fusion methods from 75.00% and 47.89% up to 76.33% and 76.75%, respectively. Kappa values were increased from 0.7282 and 0.4536 up to 0.7411 and 0.7438 for the NN and FG, respectively. Confusion matrices for the fusion and classification of the WV-2,

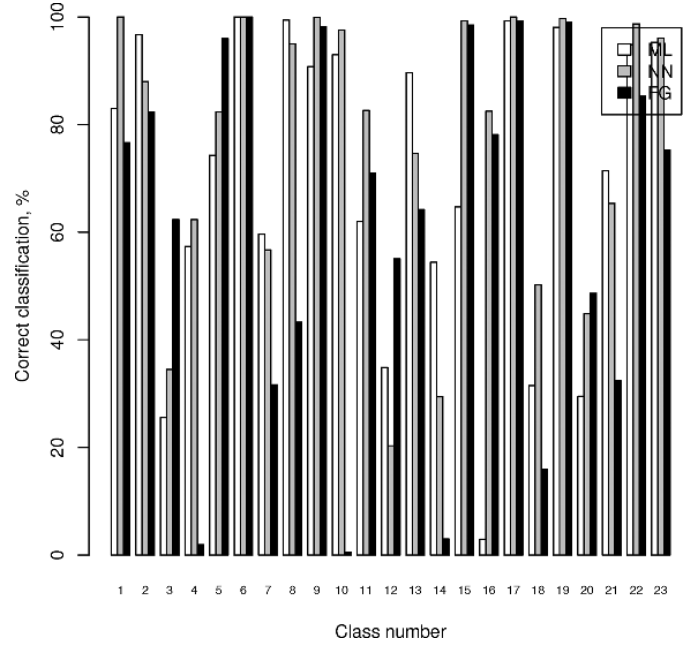


Fig. 7. Munich test area. Per-class accuracy for the WV-2+DSM data fusion and classification using the ML, NN, and FG.

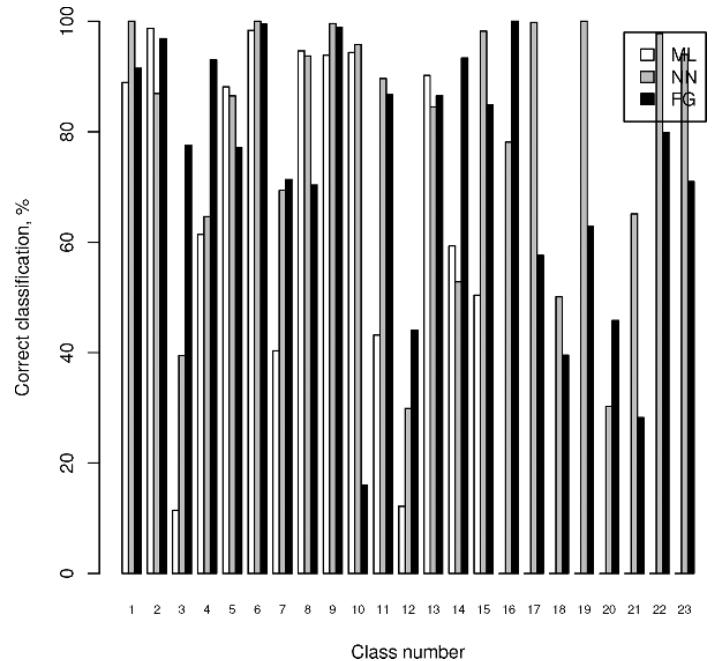


Fig. 8. Munich test area. Per-class accuracy for the WV-2+Texture(SAR and Optical)+DSM data fusion and classification using the ML, NN, and FG.

Texture, and DSM data (Tables V, VI, and VII) allow to compare per class confusion.

The introduction of the textural features extracted from the SAR data increase the accuracy of the ground class labeling but the accuracy of building classification is decreased (Table II). A strong backscatter from a building can reduce the accuracy of the building and nearby objects classification. Therefore the fusion with SAR for urban area classification should be used with caution and employment of multispectral and texture

TABLE V
MUNICH TEST AREA. CONFUSION MATRIX, THE ML CLASSIFICATION USING WV-2+TEXTURE (SAR AND OPTICAL)+DSM DATA.
OVERALL ACCURACY = 58.53, KAPPA = 0.5502

Class	Ground truth (Percent)																				User Acc.	Comission					
	WA	FO	GR	BS	CS	SP	RO	IF	TF	GH	RR	TL	CE	PA	SH	CO	R-RT	R-CO	R-VE	R-DT			R-ZI	R-CP	R-GT		
WA	88.92	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0		
FO	0	98.71	11.02	0	0	0	0	0	0	0	0	9.48	0.06	8.94	0	0	0	0	2.18	0	0	5.10	83.59	16.41			
GR	0	0	11.42	0	0	0	0	0	0	0	0	0.02	0	0	0	0	0	0	0	0	0	0	99.95	0.05			
BS	0	0	0	61.40	0.14	0	0.34	0	0	0	0	0.69	0	0	0	29.93	0	0	0.70	0	0	0	96.76	3.24			
CS	10.23	0.01	0	38.52	88.11	0.06	16.48	0.38	3.27	4.92	4.10	2.95	0	37.98	1.86	13.14	41.14	63.69	9.81	46.98	68.67	62.75	52.93	32.19	67.81		
SP	0	0	0	0	0	98.37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0			
RO	0	0	0	0	0	0	40.33	0	0	0.42	0.07	5.11	0	0	0	0	0	0	0	0	0	0	99.50	0.50			
IF	0.76	0.15	71.12	0	0	0	0.01	94.66	0	0	2.07	2.65	0.27	0	28.34	0	0	0	0	10.30	0	0	40.99	51.01			
TF	0	0	0	0	0	0	0	0	93.85	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0			
GH	0.07	0.06	0.04	0.09	2.67	0	35.34	0	0	94.33	49.82	68.53	0.06	2.66	9.49	6.57	36.48	0.17	0	30.94	19.43	4.51	3.13	24.14	75.86		
RR	0	0	0	0	0	0	4.72	0	0	0	43.15	7.87	0	0	0	0	0	0	0	2.38	0	0	77.16	22.84			
TL	0	0	0	0	0	0	1.46	0	0	0	0	12.19	0	0	0	0	0	0	0	0	0	0	12.39	87.61			
CE	0	0.95	6.40	0	0	0	0	4.96	0	0	0	0	90.17	0	0	0	0	0	0	0	0	0	67.30	32.70			
PA	0.02	0.12	0.01	0	9.08	1.58	1.32	0	2.88	0.34	0.79	0	59.30	0.97	50.36	9.90	35.89	0	0	0.31	0.21	0	69.63	30.37			
SH	0	0	0	0	0	0	0	0	0	0	0	0	0	50.40	0	0	0	0	0	0	0	0	100	0			
CO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12.48	0	0	0	0.56	0	0	0	100	0		
R-RT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0	100	0		
R-CO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	89.50	2.61	0.59	4.51	0.84	0	100	
R-VE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.60	0	0	0	0	100	0	
R-DT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10.44	0	0	0	100	0	
R-ZI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28.02	0	0	0	100	0	
R-CP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	37.60	0	100	0	
R-GT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	
Prod. Acc.	88.92	98.71	11.42	61.40	88.11	98.37	40.33	94.66	93.85	94.33	43.15	12.19	90.17	59.30	50.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Omission	11.08	1.29	88.58	38.60	11.89	1.63	59.67	5.34	6.15	5.67	56.85	87.81	9.83	40.70	49.60	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	

TABLE VI
MUNICH TEST AREA. CONFUSION MATRIX, THE NN FUSION AND CLASSIFICATION USING WV-2+TEXTURE (SAR AND OPTICAL)+DSM DATA.
OVERALL Accuracy = 76.33, Kappa = 0.7411

Class	Ground truth (Percent)																				User Acc.	Comission			
	WA	FO	GR	BS	CS	SP	RO	IF	TF	GH	RR	TL	CE	PA	SH	CO	R-RT	R-CO	R-VE	R-DT			R-ZI	R-CP	R-GT
WA	99.99	0	0	0	0	0	0	0	0	0	0	0	0	0	0.40	0	0	0	0	0	0	0	0.04	99.76	0.24
FO	0	86.91	0.46	0	0	0	0	0	0	0.24	0	2.72	1.12	0.01	0	0	0	0	0	0	0	0	0	97.60	2.40
GR	0	2.51	39.43	0	0	0	6.26	0	0	0	0	11.31	0	0	0	0	0	0	0	0	0	0	79.45	20.55	
BS	0	0	0	64.62	0.37	0	0.07	0	0	0	0	0	0.01	0	7.30	0	0	0	0	0	0	0	98.71	1.29	
CS	0	0	0	35.28	86.47	0	0.73	0	0	0.63	0.56	0	20.85	0	14.60	0	1.59	0	0	0.39	0	0	64.25	35.75	
SP	0	0	0	0	0	100.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0	0	
RO	0	0	0	0	0.14	0	69.38	0	0	1.26	0.58	39.72	0	1.63	0	0	0	0	0	0.70	0.09	0	97.73	2.27	
IF	0	0.10	46.30	0.01	0	0	0.03	93.69	0	0.21	0	1.46	0.12	0	0	0	0	0	0	0.27	0	0	62.17	37.83	
TF	0	0	0	0	0	0	0	99.55	0	0	0	0	0	0	0	0	0	0	1.67	0	0	0	98.47	1.53	
GH	0	0	0	0	0.29	0	7.29	0.01	0	95.79	4.37	4.42	0	18.21	0.06	0	0	0	0	6.36	1.85	1.05	0.16	59.58	40.42
RR	0	0	0	0	0	0	2.67	0	0	1.38	89.65	4.72	0	2.80	0.02	0	0	0	0.17	0	3.63	2.38	0	89.46	10.54
TL	0	0	0	0	0.19	0	8.39	0	0	0.11	3.62	29.89	0	0.40	0	0	0	0	0	0	0.02	0	0	4.84	95.16
CE	0	10.27	13.82	0	0	0	0.03	0	0	0.09	0	84.51	0.91	0.01	0	0	0	0	0	0	0	0	48.70	51.30	
PA	0	0	0	0	12.52	0	11.37	0	0	0.02	0.57	21.24	0	52.83	0.05	0	0.19	26.13	0	0	0.09	0	42.55	57.45	
SH	0.01	0.18	0	0	0.01	0	0	0	0	0	0	0	0	0.17	98.19	0	0	0	0	0	0	4.82	97.70	2.30	
CO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	78.10	0	0	0	0	0	0	0	100.0	0	0
R-RT	0	0	0	0.02	0.02	0	0	0.45	0	0.01	0	0	0.30	0	0	99.81	20.37	0	0	28.73	0	0	32.74	67.26	
R-CO	0	0	0	0	0	0	0	0	0	0	0	0	0.07	0	0	50.08	0	0	0	0.28	0	0	95.69	4.31	
R-VE	0	0.01	0	0	0	0	0.01	0	0	0.20	0.01	0	0	0.37	0	0	0	0	100.0	7.14	0	0	0.68	85.72	14.28
R-DT	0	0.01	0	0	0	0	0	0	0	0.09	0	0	0.09	0.07	0	0	0	0	0	30.28	0	0	94.75	5.25	
R-ZI	0	0	0	0	0	0	0.04	0	0	0.37	0	0	0.05	0	0	0	0	0	0	65.13	0	0	98.19	1.81	
R-CP	0	0	0	0	0	0	0	0	0	0.24	0	0	0.08	0	0	0	0	0	0	0.69	97.80	0.28	91.02	8.98	
R-GT	0	0	0	0	0	0	0.01	0	0	0	0	0	0	0	1.19	0	0	0	0	51.62	0	1.15	94.02	61.89	38.11
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Prod. Acc.	99.99	86.91	39.43	64.62	86.47	100.0	69.38	93.69	99.55	95.79	89.65	29.89	84.51	52.83	98.19	78.10	99.81	50.08	100.0	30.28	65.13	97.80	94.02	91.02	8.98
Omission	0.01	13.09	60.57	35.38	13.53	0	30.62	6.31	0.45	4.21	10.35	70.11	15.49	47.17	1.81	21.90	0.19	49.92	0	69.72	34.87	2.20	5.98	0	0

TABLE VII
MUNICH TEST AREA. CONFUSION MATRIX, THE FG FUSION AND CLASSIFICATION USING WV-2+TEXTURE (SAR AND OPTICAL)+DSM DATA.
OVERALL Accuracy = 76.75, Kappa = 0.7438

Class	Ground truth (Percent)																				User Acc.	Comission			
	WA	FO	GR	BS	CS	SP	RO	IF	TF	GH	RR	TL	CE	PA	SH	CO	R-RT	R-CO	R-VE	R-DT			R-ZI	R-CP	R-GT
WA	91.52	0.17	0	0	0	0	0.10	9.47	0	0	0	0	0	0	14.09	0	0	0	0	3.47	0	0.10	10.23	75.70	24.30
FO	0	96.83	7.03	0	0	0	0	0	0	0	0	5.41	0	0.50	0	0	0	0	0.23	0	0	0	0.36	90.47	9.53
GR	0	0.06	27.32	0	0	0	16.09	0	0	0	0	7.78	0	0	0	0	0	0	0	0	0	0	83.48	16.52	
BS	0	0	0	92.99	5.90	0	0.07	0.31	0	0.08	0	0	0.04	0	0	0	0	0	0	0	0	0	16.22	85.42	14.58
CS	0	0	0	7.01	77.11	0	0	0	0	8.39	0	0	0.71	0	0	30.38	2.32	0	0	4.46	0	0	82.77	17.23	
SP	0	0	0	0	0.09	99.46	0.02	0	0	0.69	0	0	0.04	0	0	0	0	0	0	2.18	0.63	0.20	92.66	7.34	
RO	0	0	0	0	3.18	0	71.37	0.66																	

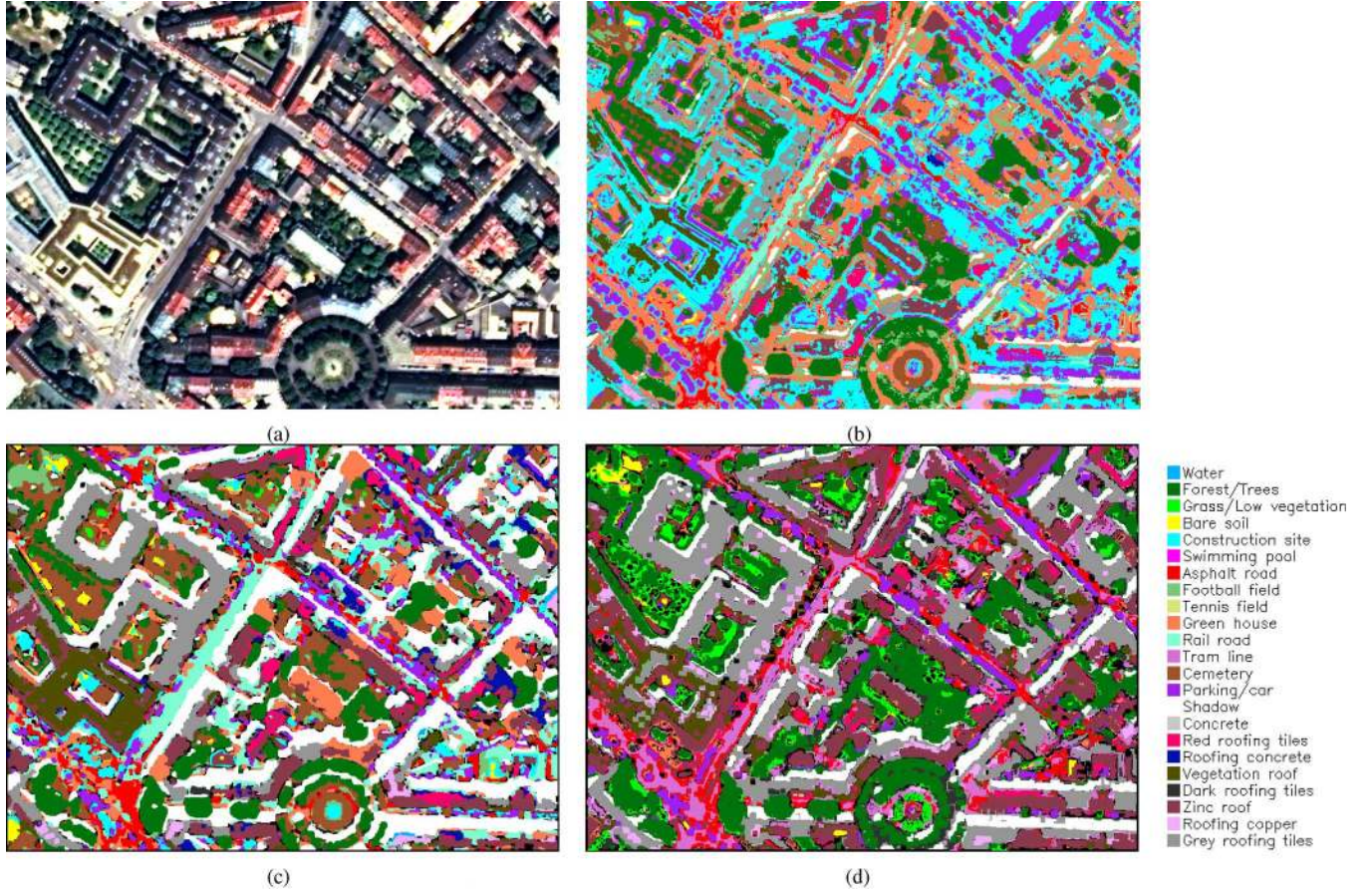


Fig. 9. Munich test area. A region of the classification map (WV-2+Texture (SAR and Optical)+DSM): (a) visible range multispectral image (bands 5,3,2), (b) fusion and classification by the ML; (c) fusion and classification by the NN; (d) fusion and classification by the FG.

information extracted from the multispectral data, as well as elevation data should be employed.

Overall, the ML provides a high confusion of buildings with construction sites in the whole scene on all features (Fig. 9(b)). Such low accuracies of the ML classification method may be achieved, since the ML classifier can not efficiently deal with different distributions of data (spectral and textural) and features, or the multisensory data is not classified in the way of consensus classification [7]. The ML fusion and classification of the WV-2+Texture (SAR and Optical)+DSM data illustrated zero accuracy of the selected material classes (classes 16–23).

Having a high classification accuracy on single sensor data (WV-2) the ML illustrates difficulties to employ multisensory data (see Table III), therefore with addition of the other features the fusion and classification accuracy decreased from $OVA = 72.51\%$ and $Kappa = 0.7010$ down to $OVA = 58.53\%$ and $Kappa = 0.5502$, respectively. Here, fusion and classification using the ML (employing stacked data) may not satisfy the assumptions on different statistical properties of multisensory data.

Fusion of multisensory data using the FG method ($OVA = 76.75\%$, $Kappa = 0.7438$) allowed to obtain an increase of the accuracy comparing to the fusion and classification results obtained by the neural network ($OVA = 76.33\%$, $Kappa = 0.7411$).

FG based fusion allows better labeling of roads, and comparing to the NN the FG allows also better labeling of tram lines and railroads. Also, the FG fusion illustrates less accurate detection of the class Green house comparing to the NN. Low accuracy for classification of single source data by the FG method (WV-2, 8 features) as well as fusion of WV-2+DSM data (9 features) can be caused since the size of the alphabet is low and might not be sufficient (50 for multispectral). Therefore a loss of information during clustering influences the accuracy comparing to the methods dealing with original 11-bit single source data. It should be noted that the specially selected classes (roofing materials of buildings) are difficult to identify using the employed data, while hyperspectral data allows better classification of materials [39]. Separability of these classes is better obtained using the fusion method based on neural network. Fig. 10 presents labels taken from FG classification map, Fig. 11 illustrates the full size classified image for Munich test area.

Given the contingency table (see example Table IX) McNemar's test [40] allows to compare the performances of two classification methods. McNemar's test has a chi-square distribution with 1 degree of freedom and the χ^2 is computed as follows:

$$\chi^2 = \frac{(|n_{01} - n_{10}| - 1)^2}{n_{01} + n_{10}}. \quad (12)$$

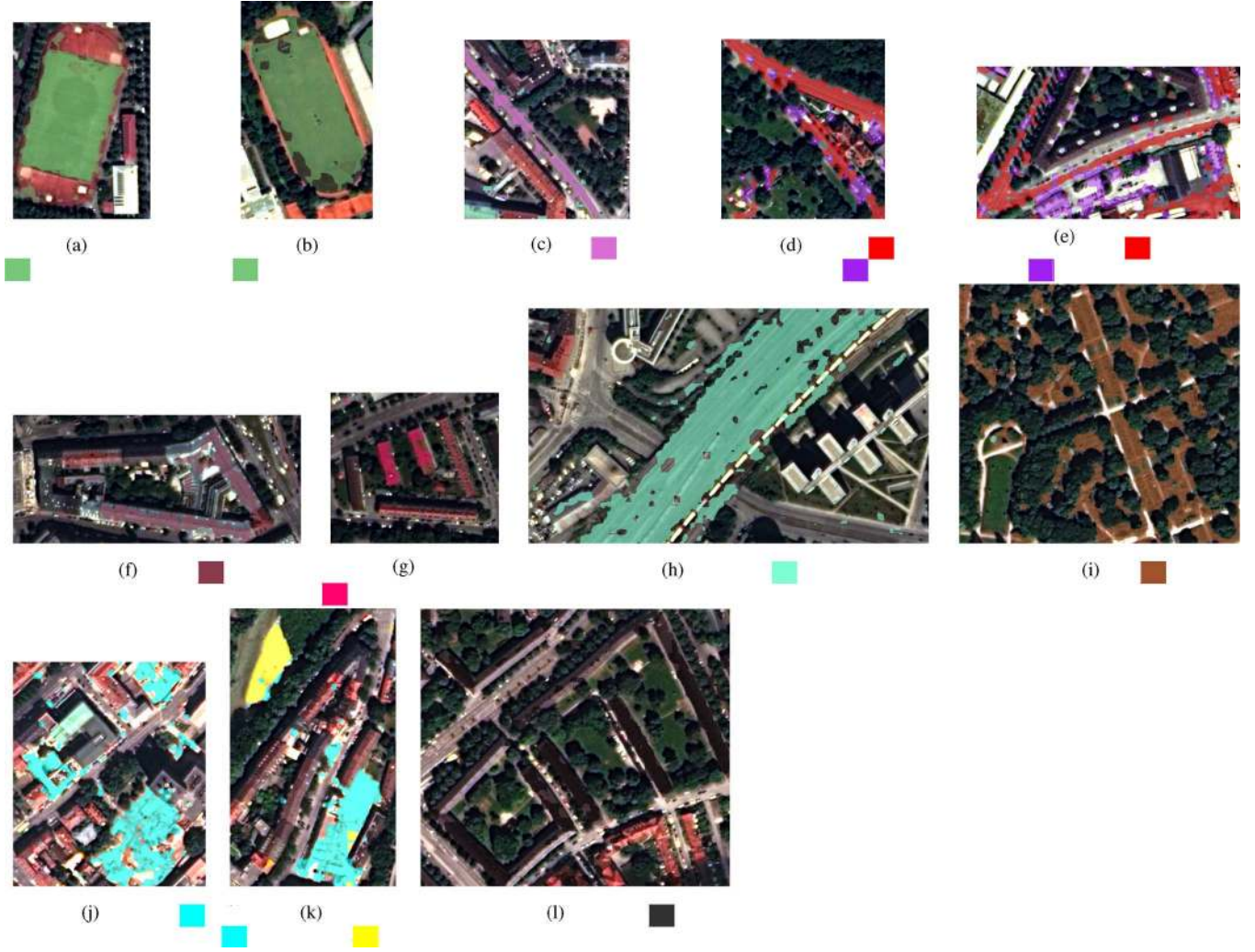


Fig. 10. Munich test area. Examples of class labels produced by WV-2+Texture+DSM data fusion and classification using factor graphs: (a), (b) Football field; (c) Tram line; (d), (e) Asphalt road and Parking/car classes; (f) Zink roof; (g) Red roofing tiles; (h) Rail road; (i) Cemetery; (j) Construction site; (k) Construction site vs. Bare soil classes; (l) Dark roofing tiles.

TABLE VIII

MUNICH TEST AREA. 2×2 CONTINGENCY TABLE FOR THE ML AND FG FUSION (WV-2+TEXTURE(SAR AND OPTICAL)+DSM). $\chi^2 = 4139.0665$. n_{00} IS THE NUMBER OF SAMPLES MISCLASSIFIED BY BOTH METHODS, n_{01} IS THE NUMBER OF SAMPLES MISCLASSIFIED BY METHOD I BUT NOT II, n_{10} IS THE NUMBER OF SAMPLES MISCLASSIFIED BY METHOD II BUT NOT I, n_{11} IS THE NUMBER OF CORRECTLY CLASSIFIED SAMPLES BY BOTH METHODS

$n_{00} = 28702$	$n_{01} = 57333$
$n_{10} = 37518$	$n_{11} = 161218$

TABLE IX

MUNICH TEST AREA. 2×2 CONTINGENCY TABLE FOR THE NN AND FG FUSION (WV-2+TEXTURE(SAR AND OPTICAL)+DSM). $\chi^2 = 16.4581$

$n_{00} = 24224$	$n_{01} = 43181$
$n_{10} = 41996$	$n_{11} = 175370$

A low p -value calculated from the χ^2 suggests that the null hypothesis should be rejected meaning that the two classifiers achieve different results. For the ML and FG (Table VIII), and the NN and FG (Table IX) fusion on the WV-2+Texture

TABLE X

FUSION AND CLASSIFICATION USING THE FG ON THE INPUT DATA REPRESENTED ON THE ALPHABET WITH VARYING SIZE (MUNICH AREA, WV-2+TEXTURE (SAR AND OPTICAL)+DSM FEATURES, 23 CLASSES)

Alphabet size	OVA, %	Kappa
20	74.12	0.7140
50	75.25	0.7266
100	72.89	0.7010

(SAR and Optical)+DSM features the χ^2 values are 4139.06 and 16.45, respectively. The p -values for the ML and FG, and the NN and FG are less than 0.05, meaning that the FG classification have different performances (better accuracy) on the same data.

The FG fusion and classification approach allows to achieve an acceptable accuracy even on a small size of the representation domain (in the paper experiment the domain sizes are 10 for Gabor features, 10 for DSM and 50 for multispectral data). Even a simple structure of the FG allows to increase the fusion



Fig. 11. Munich test area. Full size image employed for the FG fusion and classification. (a) visible range multispectral image (bands 5,3,2), (b) FG fusion and classification (WV-2+Texture (SAR and Optical)+DSM).

accuracy. The independent model of the factor graph is easy to configure and apply for real data.

4) *Alphabet Size Influence on the Fusion Accuracy*: A trade-off between the size of the alphabet (the number of clusters) and the accuracy of the FG fusion is of interest. An experiment is run to assess the influence of the alphabet size on the FG fusion accuracy. The FG structure, the training

and test samples, and the data (WV-2+Texture (SAR and Optical)+DSM features) are as in the Munich test area experiment. The multisensory data (all features) were represented using different size of the alphabet: 20, 50, and 100 (Table X).

Here, the increase of the alphabet size does not allow to rise the accuracy of the FG fusion and classification and it is possible to conclude that there is no trend on the accuracy improvement

TABLE XI

THE RANK OF THE FIRST 10 FEATURES SELECTED BY THE MRMR METHOD FROM THE FEATURE SET EMPLOYED FOR MUNICH AREA FUSION AND CLASSIFICATION USING FACTOR GRAPHS (10 CLUSTERS FOR GABOR FEATURES, 10 FOR THE DSM, AND 50 FOR THE WV-2 MULTISPECTRAL DATA REPRESENTATION)

Rank	Feature number	Feature type	Parameters
1	41	WV-2 spectral	Band 5 (Red, 630-690 nm)
2	45	DSM	
3	30	Gabor feature (Optical)	$\sigma = 4, \omega = 2/3\pi, \theta = 5/6\pi$
4	34	Gabor feature (Optical)	$\sigma = 4, \omega = \pi, \theta = \pi/2$
5	44	WV-2 spectral	Band 8 (NIR2, 860-1040 nm)
6	23	Gabor feature (Optical)	$\sigma = 4, \omega = 2/3\pi, \theta = 0$
7	27	Gabor feature (Optical)	$\sigma = 4, \omega = 2/3\pi, \theta = \pi/3$
8	26	Gabor feature (Optical)	$\sigma = 4, \omega = 2/3\pi, \theta = \pi/6$
9	31	Gabor feature (Optical)	$\sigma = 4, \omega = \pi, \theta = 0$
10	37	WV-2 spectral	Band 1 (Coastal Blue, 400-450 nm)

TABLE XII

PARAMETERS OF THE WORLDVIEW-2 AND TERRASAR-X DATA FOR LONDON TEST SCENE

Parameter	WorldView-2 [33]	TerraSAR-X
Product	Ortorectified Standard Imagery	EEC
Sensor mode	Multispectral, PAN	Spotlight HS
Orbit	Descending	Ascending
Acquisition date and time (UTC)	22 Oct 2011, 11:34:14	12 Jan 2009, 17:52:29
Look angle	6.3°, Left	48.23°, Right
Ground pixel size, m	0.5 × 0.5	0.5 × 0.5
Polarization	-	Single, VV
Bits per pixel	11	16

TABLE XIII

LONDON TEST AREA. CLASS OF INTEREST, ACRONYM, NUMBER OF TRAINING AND TEST SAMPLES (TRAIN AND TEST AREAS WERE SELECTED USING VISUAL INTERPRETATION OF WV-2 VISIBLE RANGE DATA TOGETHER WITH BING MAPS). PER CLASS FUSION AND CLASSIFICATION ACCURACIES OF THE ML, NN, AND FG METHODS ON WV-2, WV-2+DSM, AND WV-2+TEXTURE+DSM

	Class	Acronym	Training samples	Test samples	WV-2			WV-2+DSM			WV-2+Texture+DSM		
					ML	NN	FG	ML	NN	FG	ML	NN	FG
1	Water	WA	1678	50401	100	100	100	100	99.92	99.97	99.77	99.97	98.27
2	Forest/Trees	FO	1936	12709	97.26	94.40	89.97	96.77	92.93	93.0	98.31	97.76	93.01
3	Grass/Low vegetation	GR	229	7396	23.97	13.01	61.74	11.15	5.92	62.51	3.68	10.64	67.54
4	Bare soil	BS	630	2556	19.80	29.81	82.51	31.30	57.98	87.25	29.19	99.06	82.63
5	Football field	FF	4316	1640	94.82	89.33	28.29	96.40	87.87	60.49	89.57	95.98	69.70
6	Rail road	RR	699	15170	88.30	64.30	52.24	90.82	54.34	77.36	76.88	34.70	70.20
7	Parking/car	PA	596	4288	23.83	7.14	12.41	8.0	16.23	7.32	26.94	28.38	4.69
8	Asphalt road	RO	1174	88378	81.49	49.57	68.81	78.04	80.05	71.78	69.58	65.08	75.93
9	Shadow	SH	1461	41789	78.33	81.97	76.56	70.40	54.55	84.85	83.01	54.30	91.75
10	High-rise building	B-HR	1281	68674	26.37	26.68	38.76	90.29	82.93	55.41	88.18	83.96	83.53
11	Low-rise building	B-LR	8666	117373	20.73	50.36	28.20	32.79	58.99	10.72	26.21	56.17	33.27
12	Medium-rise building	B-MR	1549	39903	19.43	10.59	16.06	42.24	41.57	36.76	33.57	34.13	39.29
13	Tennis field	TF	112	451	0	0	60.53	0	0	67.63	0	0	48.12
14	Dock	DO	2599	31491	98.03	95.99	100	99.47	99.54	100	98.70	99.34	99.99

TABLE XIV

LONDON TEST AREA. CLASSIFICATION ACCURACY USING DIFFERENT METHODS TOGETHER WITH THE PROPOSED APPROACH (WV-2+TEXTURE (SAR AND OPTICAL)+DSM DATA). ML-MAXIMUM LIKELIHOOD (NOT FOLLOWING CONSENSUS THEORY), NN-NEURAL NETWORK, FG-FACTOR GRAPH. OVA—OVERALL ACCURACY, KAPPA—COHEN'S KAPPA

Method	Employed features	OVA, %	Kappa
ML	WV-2 (8)	55.33	0.4907
ML	WV-2+DSM (9)	67.83	0.6322
ML	WV-2+Texture+DSM (45)	64.30	0.5926
NN	WV-2 (8)	55.08	0.4738
NN	WV-2+DSM (9)	70.95	0.6632
NN	WV-2+Texture+DSM (45)	66.96	0.6157
FG	WV-2 (8)	55.60	0.4860
FG	WV-2+DSM (9)	57.67	0.5231
FG	WV-2+Texture+DSM (45)	68.37	0.6387

TABLE XV

LONDON TEST AREA. CLASSIFICATION ACCURACY USING THE NN ON WV-2+TEXTURE (SAR AND OPTICAL)+DSM DATASET (ALL TOGETHER 45 FEATURES). THE NN IS A 2 LAYER NETWORK (1 HIDDEN LAYER) OR A 3 LAYER NETWORK (2 HIDDEN LAYERS) WITH A DIFFERENT NUMBER OF NEURONS IN THE HIDDEN LAYER/LAYERS (5, 10, 20, 60, AND 80) ARE EMPLOYED TO FIND THE MOST SUITABLE ARCHITECTURE OF THE NETWORK AND REACH THE HIGHEST FUSION AND CLASSIFICATION ACCURACY

Number of neurons	1 hidden layer		2 hidden layers	
	OVA, %	Kappa	OVA, %	Kappa
5	60.88	0.5438	62.48	0.5635
10	64.65	0.5878	65.41	0.5964
20	64.63	0.5878	66.96	0.6157
60	61.06	0.5433	64.06	0.5810
80	64.27	0.5859	61.79	0.5545

with enlargement of the alphabet size. It should be noted that the increase of the alphabet size grows the time for the factor

graph configuration and further fusion. A moderately low and sufficient size of the alphabet allows to reach a competitive accuracy of the FG fusion and classification and to have a low calculation time.

TABLE XVI
LONDON TEST AREA. CONFUSION MATRIX, THE ML CLASSIFICATION USING WV-2+TEXTURE (SAR AND OPTICAL)+DSM DATA.
OVERALL ACCURACY = 64.30, KAPPA = 0.5926

Class	Ground truth (Percent)														User Acc.	Comission
	WA	FO	GR	BS	FF	RR	PA	RO	SH	B-HR	B-LR	B-MR	TF	DO		
WA	99.77	0	0	0	0	0	0	0.16	0	0	0.07	0.38	20.18	0	99.08	0.92
FO	0	98.31	41.52	0	0	0	1.26	0.18	0.52	0.05	2.01	0.11	0.22	0	67.78	32.22
GR	0	0	3.68	0	0	0	0	0	0	0	0	0	0	0	100	0
BS	0	0.01	1.54	29.19	0.73	0.02	0	0.94	0	0.02	1.03	0.02	12.20	0	25.00	75.00
FF	0	0.64	52.87	0	89.57	0.28	0	0.01	1.55	0	3.07	0.07	0	0	15.02	84.98
RR	0	0	0	0	0	76.88	1.00	19.81	0.01	0.01	4.11	0.75	9.31	0	33.91	66.09
PA	0	0.02	0	2.23	0	0.04	26.94	0.22	0.01	0.04	4.53	0.78	0	0	16.31	83.69
RO	0	0	0	0	0	8.10	1.17	69.58	2.56	2.51	13.64	25.02	53.88	0	66.98	33.02
SH	0	0	0	0	0	0	0	0.22	83.01	2.44	0.85	0.27	0	1.21	91.19	8.81
B-HR	0.23	0.98	0.37	33.06	5.73	1.78	21.27	2.38	9.64	88.18	20.15	17.40	0.89	0.09	60.73	39.27
B-LR	0	0.04	0.03	35.49	3.96	11.63	48.34	6.26	0.35	2.13	26.21	21.63	3.33	0	59.89	40.11
B-MR	0	0	0	0.04	0	1.27	0.02	0.24	0.06	4.61	24.33	33.57	0	0	29.41	70.59
TF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DO	0	0	0	0	0	0	0	0	2.30	0	0	0	0	0	98.70	3.00
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0		
Prod. Acc.	99.77	98.31	3.68	29.19	89.57	76.88	26.94	69.58	83.01	88.18	26.21	33.57	0	98.70		
Omission	0.23	1.69	96.32	70.81	10.43	23.12	73.06	30.42	16.99	11.82	73.79	66.43	100.00	1.30		

5) *Feature Rank*: Feature selection is run to analyze and find the features, which are more relevant for the FG fusion and classification. Peng *et al.* [41] developed a new and popular method for feature selection-the mRMR (minimum Redundancy Maximum Relevance Feature Selection). The features selected by the method have a good performance on various types of classifiers with different methodology such as support vector machines, Naive Bayes, Linear discriminant analysis. Table XI presents the rank of the first ten features selected for Munich scene. The first five features were taken for the FG fusion and classification experiment.

The FG fusion and classification on the selected five features resulted in the overall accuracy equal to 74.44% and Kappa = 0.7184. The experimental analysis illustrated that the feature selection does not result in the accuracy increase. Generally, a proper configuration (learning) of the factor graph makes it sensitive to the more informative features (for a particular class), while the less informative input features are chosen by the FG less significant for the fusion and classification. According to the feature rank, WorldView-2 multispectral (places 1,5,10), Textural (Optical) (places 3,4,6–9), and the DSM (2nd place) data bring the most significant contribution for the classification accuracy. The feature rank illustrates that the SAR textural features are less important for the classes identification in comparison to Gabor features calculated on optical data (places 3,4,6–9). The employment of the Gabor features improves identification of the classes characterized by spatial context properties. The DSM is placed on the 2nd place noticing an importance of elevation data. The importance of the DSM (classification of the objects with the help of the height information) also corresponds to the rankings obtained in the classification experiment on WorldView-2 and DSM data on urban area carried out by Longbotham *et al.* [42].

B. London Test Area

1) *Multisensory Data and Features*: A combination of WorldView-2 multispectral, TerraSAR-X (SpotLight Level-1B product), and a digital surface model is used. Acquisition geometry of the employed WorldView-2 data is 6.3° off-nadir

view angle. Acquisition geometry of the TerraSAR-X data is 48.23° off-nadir look angle. A detailed description of the employed data sets is given in Table XII. The registration of optical and radar data is made in ENVI using manual selection of control points.

WorldView-2 multispectral data were pan-sharpened by the General Fusion Framework method [31]. Digital Surface Model (DSM) is generated using the Semiglobal Matching algorithm [34]. A surface model generated from other sensor type (LIDAR, for example) can be also employed. The optical data were orthorectified. Gabor features were calculated on the TSX data and Red color channel (630–690 nm) from WorldView-2 data. A bank of Gabor wavelets consists of 18 filters (6 orientations $(0, \pi/6, \pi/3, \pi/2, (2)/(3)\pi, (5)/(6)\pi)$, 3 different periods of filter's sine component $(\pi/3, (2)/(3)\pi, \pi)$, and 1 sigma value $(\sigma = 4)$). A recursive implementation of Gabor filtering is run [36].

A subscene (3101 × 3041 pixels) is used for the fusion and classification experiments. k -means clustering is employed for feature representation on the alphabet. The number of clusters is selected empirically and set to 10 for Gabor features (calculated on SAR and optical data), 10 for the DSM, and 50 for the WV-2 multispectral data.

2) *Fusion Strategies and Classification*: The following combinations of multisensory and single-sensor data are created:

- 1) WV-2 (8 features (spectral bands)),
- 2) WV-2+DSM (9 features),
- 3) WV-2+Texture (SAR and Optical)+DSM (45 features).

Altogether, 14 classes were defined, the number of training and test samples is given in Table XIII. Selection of training and test regions is made manually on visible color composite of WorldView-2 multispectral data and Bing maps. The training and test samples are spatially uncorrelated.

3) *Results and Discussion*: All the methods were run on the same feature set and the same training/test regions were employed (Table XIV presents overall accuracy and Cohen's Kappa for the compared methods: the ML, NN, and FG). The Neural Network is a 3 layer (2 hidden layers) feed-forward net trained with Kalman filter, implemented in IDL [38]. The input

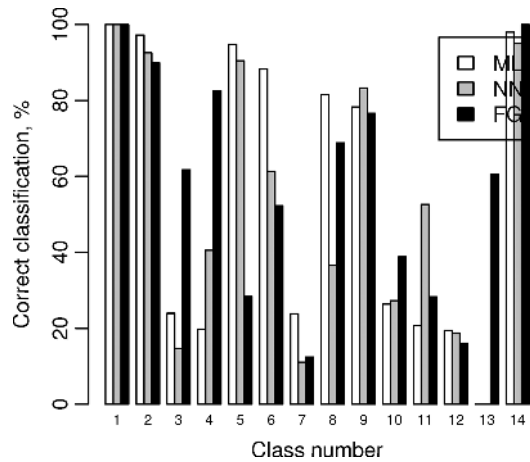


Fig. 12. London test area. Per-class accuracy for the WV-2 data classification using the ML, NN, and FG.

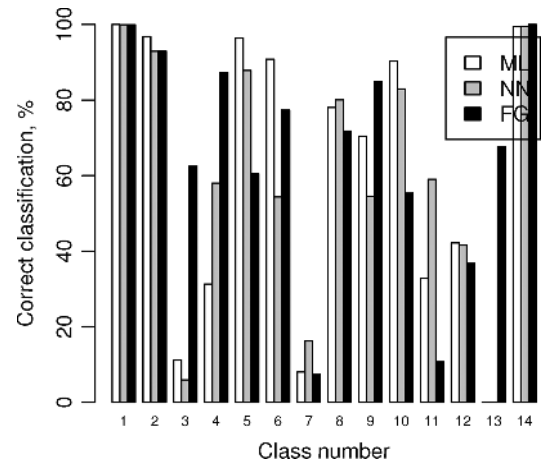


Fig. 13. London test area. Per-class accuracy for the WV-2+DSM data fusion and classification using the ML, NN, and FG.

features are normalized to one. The number of neurons in the hidden layer is selected experimentally. Running the neural network on the same feature set (WV-2+Texture (SAR and Optical)+DSM) and employing different number of neurons (5, 10, 20, 60, and 80) it is found that the network with two hidden layers and 20 neurons in a hidden layer allowed to obtain the best result on fusion and classification (Table XV). The same number of neurons is employed for the WV-2 feature set classification. For the ML classification ENVI software package is used. An independent model of the factor graph is selected for fusion and classification. Probability maps are calculated for each class and maximal MAP selection allows to assign the class labels.

Table XIII and Figs. 12, 13, and 14 illustrate the fusion and classification accuracy using single and multisensory data. Introduction of the textural features and the DSM allowed to increase the overall accuracy and Kappa values for the ML, NN, and FG fusion methods. Confusion matrices for the fusion and classification of the WV-2, Texture, and DSM data (Tables XVI, XVII, and XVIII) allow to compare per class confusion.

Fusion of multisensory data using the FG method (OVA = 68.37%, Kappa = 0.6387, example in Fig. 15(d)) allowed to obtain a better accuracy comparing to the fusion and classification results obtained by the neural network (OVA = 66.96%, Kappa = 0.6157), and by maximum likelihood classifier (OVA = 64.30%, Kappa = 0.5926). The introduction of the DSM and textural features increased the accuracy of the FG for classes Forest, Grass, Football field, Rail road, Asphalt road, Shadow, High-rise building, Medium-rise building, Low-rise building (Table XIII). In comparison to the NN (Table XIII), the employment of the WV-2+Texture (SAR and Optical)+DSM allowed the FG to reach better labeling of the classes: Grass/low vegetation, Rail road, Asphalt road, Shadow, Medium-rise building, Tennis field, Dock.

Overall, the ML provides a comparable accuracy of fusion and classification on all combinations of data (Table XIII). Introduction of the textural features and the DSM increased the accuracy for classes: Forest/Trees, Bare soil, Parking/car, Shadow, High-rise building, Low-rise building, Medium-rise

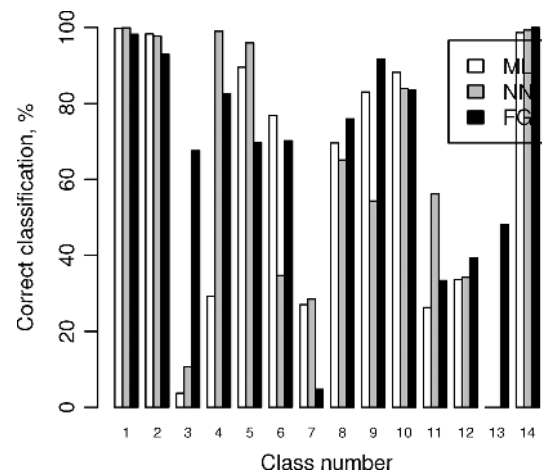


Fig. 14. London test area. Per-class accuracy for the WV-2+Texture(SAR and Optical)+DSM data fusion and classification using the ML, NN, and FG.

building comparing to the employment of single-sensor multispectral data. The reduce of the accuracy for the other classes may be caused by the statistical assumptions on the input data in the ML method. The ML fusion and classification of the WV-2 and WV-2+Texture (SAR and Optical)+DSM data illustrated zero accuracy for class Tennis field.

The NN provides a less accurate overall accuracy on the multisensory data in comparison to the FG. The NN allowed better labeling of Water, Forest, Bare soil, Football field, Parking, High-rise building, Low-rise building. Labeling of Tennis field is difficult to perform by the NN; Parking class is difficult to label by the FG.

Employment of single-sensor data (WV-2 multispectral data) for classification allowed to reach a high and comparable accuracy by all the methods (Table XIV; the ML resulted with a slightly higher Kappa value than the others). Usually maximum likelihood allows to reach a high accuracy for single-sensor optical multispectral data.

For the McNemar's test on the ML and FG (Table XIX), and the NN and FG (Table XX) fusion results on the WV-2+Texture (SAR and Optical)+DSM features, the χ^2 values are 3087.335 and 328.6979, respectively. The p -values for the ML and

TABLE XVII
LONDON TEST AREA. CONFUSION MATRIX, THE NN CLASSIFICATION USING WV-2+TEXTURE (SAR AND OPTICAL)+DSM DATA.
OVERALL ACCURACY = 66.96, KAPPA = 0.6157

Class	Ground truth (Percent)															User Acc.	Comission
	WA	FO	GR	BS	FF	RR	PA	RO	SH	B-HR	B-LR	B-MR	TF	DO			
WA	99.97	0	0	0.04	0.18	0.01	1.84	0.86	0	0.13	0.15	0.07	0	0	97.8	2.2	
FO	0	97.76	20.48	0	0	0	1.21	0.09	2.89	0	1.09	0.01	0	0	75.01	24.99	
GR	0	1.26	10.64	0	0	0	0.54	0.03	0	0.01	0.39	0	0	0	53.87	46.13	
BS	0	0.01	1.60	99.06	3.60	0	0.96	0.06	0	0.92	9.49	1.04	0	0	16.9	83.1	
FF	0	0.72	67.28	0	95.98	0	2.22	0.14	0.24	0.08	3.32	0.15	0	0	14.34	85.66	
RR	0	0.02	0	0	0	34.70	2.52	8.12	0	0.02	1.32	0.01	2.00	0	37.26	62.74	
PA	0	0	0	0	0	0.03	28.38	0.61	0.18	0.53	1.62	1.92	0	0	25.01	74.99	
RO	0	0.02	0	0.04	0	0.63	6.60	65.08	0	1.39	3.34	0.43	56.54	0	91	9	
SH	0	0.02	0	0	0	0	0.79	0.34	54.30	0	0.98	0.09	0	0.66	92.91	7.09	
B-HR	0	0	0	0.04	0	0.01	0.02	0	38.31	83.96	1.72	0.82	0	0	75.85	24.15	
B-LR	0.01	0.12	0	0.82	0.24	64.63	54.48	24.65	2.00	1.38	56.17	61.33	41.46	0	52.18	47.82	
B-MR	0	0.07	0	0	0	0	0	0	0.04	11.53	20.27	34.13	0	0	30.02	69.98	
TF	0	0.01	0	0	0	0	0.12	0.01	0	0.01	0.04	0.02	0	0	0	100	
DO	0.02	0	0	0	0	0	0.33	0.01	2.03	0.01	0.10	0	0	99.34	96.86	3.14	
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			
Prod. Acc.	99.97	97.76	10.64	99.06	95.98	34.7	28.38	65.08	54.3	83.96	56.17	34.13	0	99.34			
Omission	0.03	2.24	89.36	0.94	4.02	65.3	71.62	34.92	45.7	16.04	43.83	65.87	100	0.66			

TABLE XVIII
LONDON TEST AREA. CONFUSION MATRIX, THE FG CLASSIFICATION USING WV-2+TEXTURE (SAR AND OPTICAL)+DSM DATA.
OVERALL ACCURACY = 68.37, KAPPA = 0.6387

Class	Ground truth (Percent)															User Acc.	Comission
	WA	FO	GR	BS	FF	RR	PA	RO	SH	B-HR	B-LR	B-MR	TF	DO			
WA	98.27	0	0	0	0	0	0	0.14	0	0.05	0	0	0	0	99.68	0.32	
FO	0	93.01	28.38	0	0	0	0.54	0.06	4.69	0.24	3.16	0.27	0	0	59.29	40.71	
GR	0	2.24	67.54	0	0	0	0	0.12	0	0	0.49	0	9.53	0	83.22	16.78	
BS	0	0	0	82.63	0	0	8.12	0	0	0.53	27.33	0.99	0	0	5.98	94.02	
FF	0	1.45	0.12	5.32	69.70	0	3.96	0.36	0	0.27	2.01	0.31	0	0	24.73	75.27	
RR	0	0.42	0	0	0	70.20	0.09	0.20	0.44	0	4.67	2.20	0	0	61.10	38.90	
PA	0	0.24	0	2.11	11.77	0	4.69	0.78	0	0	0.54	0.04	0	0	11.05	88.95	
RO	0.11	0	3.15	0	1.46	10.61	9.07	75.93	0	0	4.33	0.04	41.24	0	89.84	10.16	
SH	0	0.09	0	0	0	0	0.26	0.36	91.75	0.68	1.38	0.45	0	0.01	93.62	6.38	
B-HR	0	0.15	0	0.16	0.06	0	0.09	0.02	0.11	83.53	5.12	22.05	0	0	79.37	20.63	
B-LR	1.61	2.37	0.81	5.44	16.95	19.18	65.62	19.51	1.08	9.04	33.27	34.37	1.11	0	46.50	53.50	
B-MR	0.01	0.02	0	4.34	0.06	0	7.56	0.23	0	5.66	17.62	39.29	0	0	38.33	61.67	
TF	0	0	0	0	0	0	0	2.30	0	0	0.07	0	48.12	0	9.33	90.67	
DO	0	0	0	0	0	0	0	0	1.92	0	0	0	0	99.99	97.51	2.49	
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			
Prod. Acc.	98.27	93.01	67.54	82.63	69.70	70.20	4.69	75.93	91.75	83.53	33.27	39.29	48.12	99.99			
Omission	1.73	6.99	32.46	17.37	30.30	29.80	95.31	24.07	8.25	16.47	66.73	60.71	51.88	0.01			

TABLE XIX
LONDON TEST AREA. 2 × 2 CONTINGENCY TABLE FOR THE ML AND FG FUSION (WV-2+TEXTURE(SAR AND OPTICAL)+DSM). $\chi^2 = 3087.335$

$n_{00} = 99949$	$n_{01} = 72201$
$n_{10} = 52573$	$n_{11} = 257496$

TABLE XX
LONDON TEST AREA. 2 × 2 CONTINGENCY TABLE FOR THE NN AND FG FUSION (WV-2+TEXTURE(SAR AND OPTICAL)+DSM). $\chi^2 = 328.6979$

$n_{00} = 85180$	$n_{01} = 74163$
$n_{10} = 67342$	$n_{11} = 255534$

FG, and the NN and FG McNemar's tests are less than 0.05, meaning that the FG classification have different performances (better accuracy) on the same data.

Fig. 16 illustrates the full size classified image for London test area. Fig. 17 illustrates artifacts which can appear during fusion in the areas with high-rise buildings. This is an illustrative example where the NN classifier is influenced by the DSM data and produces a wrong decision (a confusion of shadowing and High-rise building classes appears).

TABLE XXI
FUSION AND CLASSIFICATION USING THE FG ON THE INPUT DATA REPRESENTED ON THE ALPHABET WITH VARYING SIZE (LONDON AREA, WV-2+TEXTURE (SAR AND OPTICAL)+DSM FEATURES, 14 CLASSES)

Alphabet size	OVA, %	Kappa
20	68.34	0.6361
50	64.66	0.5964
100	61.62	0.5581

4) *Alphabet Size Influence on the Fusion Accuracy:* A trade-off between the size of the alphabet and the accuracy of the FG fusion is also performed for London test area. The FG structure, the training and test samples, and the data (WV-2+Texture (SAR and Optical)+DSM features) are as in the London test area experiment. The multisensory data (all features) were represented using different size of the alphabet: 20, 50, and 100 (Table XXI).

In this experiment we may conclude that there is no trend on the accuracy increase of the fusion and classification accuracy with the increase of the alphabet size (the same conclusions are made for Munich test area).

5) *Feature Rank:* Feature selection is run to find the most relevant and important features for the FG fusion and classification. The same feature selection method (mRMR, Peng *et al.*

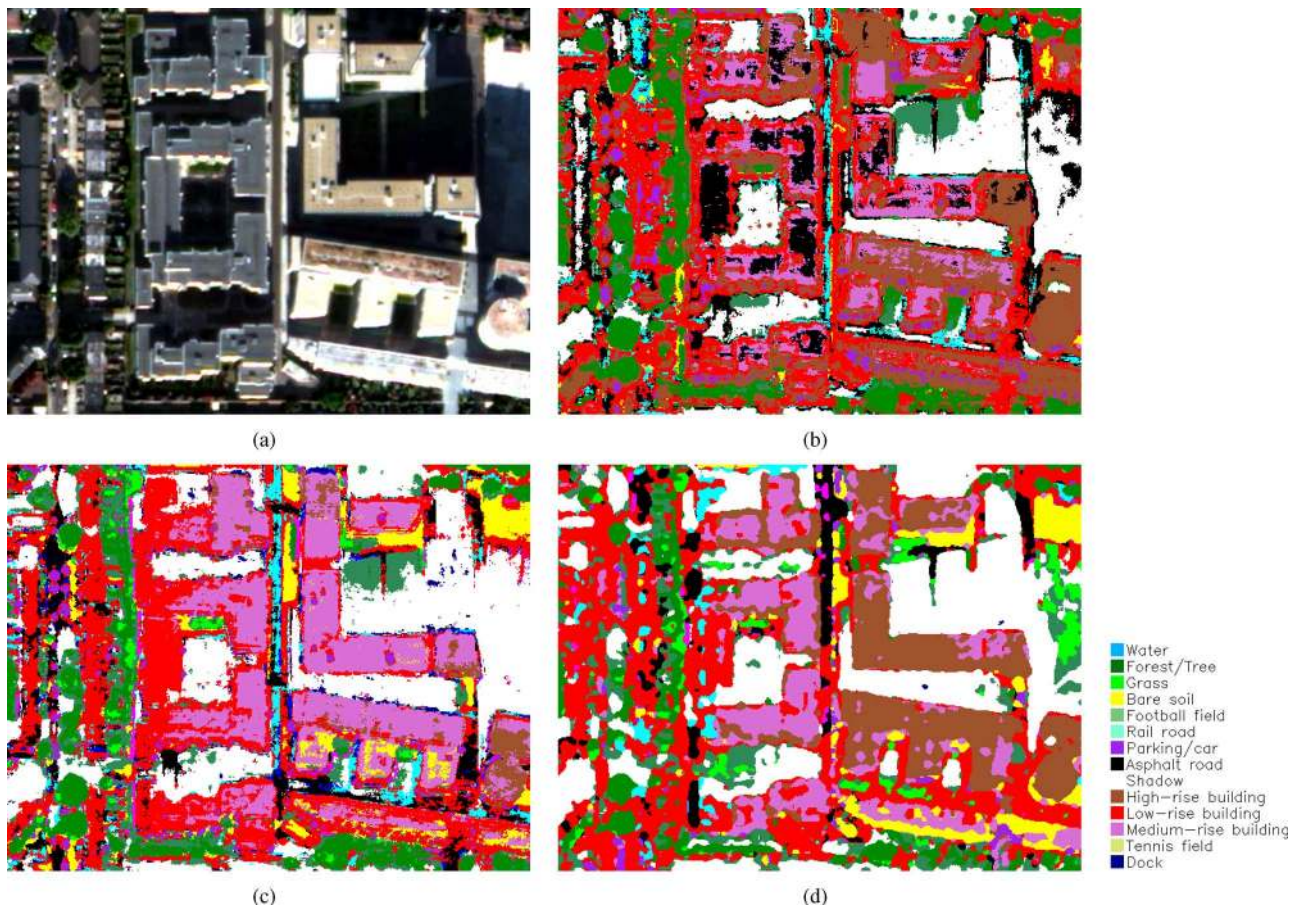


Fig. 15. London test area. A region of the classification (WV-2+Texture (SAR and Optical)+DSM data) map: (a) visible range multispectral image (bands 5,3,2), (b) fusion and classification by the ML; (c) fusion and classification by the NN; (d) fusion and classification by the FG.

TABLE XXII

THE RANK OF THE FIRST 10 FEATURES SELECTED BY THE MRMR METHOD FROM THE FEATURE SET EMPLOYED FOR LONDON AREA FUSION AND CLASSIFICATION USING FACTOR GRAPHS (10 CLUSTERS FOR GABOR FEATURES, 10 FOR THE DSM, AND 50 FOR THE WV-2 MULTISPECTRAL DATA REPRESENTATION)

Rank	Feature number	Feature type	Parameters
1	45	DSM	
2	25	Gabor feature (Optical)	$\sigma = 4, \omega = 2/3\pi, \theta = 0$
3	37	WV-2 spectral	Band 1 (Coastal Blue, 400-450 nm)
4	10	Gabor feature (SAR)	$\sigma = 4, \omega = 2/3\pi, \theta = \pi/2$
5	22	Gabor feature (Optical)	$\sigma = 4, \omega = \pi/3, \theta = \pi/2$
6	2	Gabor feature (SAR)	$\sigma = 4, \omega = \pi/3, \theta = \pi/6$
7	44	WV-2 spectral	Band 8 (NIR2, 860-1040 nm)
8	31	Gabor feature (Optical)	$\sigma = 4, \omega = \pi, \theta = 0$
9	18	Gabor feature (Optical)	$\sigma = 4, \omega = \pi/3, \theta = 0$
10	7	Gabor feature (SAR)	$\sigma = 4, \omega = 2/3\pi, \theta = 0$

[41]) is employed as in the experiment on Munich test area data. Table XXII presents the rank of the first ten features selected from the feature set calculated for London area experiment. The first five features were taken for the FG fusion and classification experiment.

The FG fusion and classification on the selected five features resulted in the overall accuracy equal to 64.74% and $Kappa = 0.5877$. As in the experiment for Munich area, the analysis illustrated that the feature selection does not result in the accuracy increase. The DSM is placed on the 1st place noticing a high importance of elevation data for the identification of the classes (corresponds to the rankings obtained in the classifica-

tion experiment on urban area carried out by Longbotham *et al.* [42]). Availability if the DSM makes possible identification and reduces the confusion of the following classes characterized by elevation like building type (High-rise, Low-rise, and Medium-rise building), Asphalt road, and Bare soil. The identification of buildings, for example, can be difficult without elevation data. In this experiment, three Gabor features calculated on SAR data were selected (places 4,6,10) among the selected WorldView-2 spectral bands (places 3,7) and Gabor features calculated on optical data (places 2,5,8,9). The feature rank for London area contains several features also selected in Munich area experiment: the DSM, WorldView-2 Band 1 (Coastal Blue,



Fig. 16. London test area. Full size image employed for the FG fusion and classification. (a) visible range multispectral image (bands 5,3,2), (b) FG fusion and classification (WV-2+Texture (SAR and Optical)+DSM).

400–450 nm), Band 8 (NIR2, 860–1040 nm), and Gabor features on optical data (parameter sets are: $\sigma = 4, \omega = 2/3\pi, \theta = 0$; $\sigma = 4, \omega = \pi, \theta = 0$) as in Munich experiment. Selected features can be dependent on landcover classes and can vary with the change of the classes.

V. CONCLUSION

Multisensory remote sensing data fusion allows to perform more precise decisions about the landcover present in the scene and allows to identify a higher number of specific classes. In most of the cases some specific landcover classes and objects as well as a high number of classes in the scene are not possible to identify using only single-sensor data.

Among several choices for multisensory data fusion, the selection of factor graphs for the fusion allows to perform a classification into an extended set of classes. It also provides a possibility for the fusion model development and opens possibilities to augment the model to define more complex systems.

Representation of multisensory data and extracted features using an alphabet (a predefined domain with finite states) allows

to deal with incommensurable features and data of different nature, statistical properties, and distributions. Such representation allows easier processing of data using factor graph. Separate processing of input features (spectral bands, textural features, or other multimodal features) and employment of the presented data fusion model is not influenced by the limitations of data dimensionality (i.e., there is no the curse of dimensionality). Proper selection of a factor graph structure allows the fusion model to be tractable for real use and application for remotely sensed data of arbitrary size. Approximate inference on independent models makes it possible to perform inference on input evidence of moderately high size. Application of factor graph on real multisensory data (WorldView-2, TerraSAR-X, and DSM) allows to obtain better results than the results reached by application of a multilayer perceptron.

The proposed fusion model is influenced and can be limited by the following properties. The quality of factor graph configuration depends on the learning method. An appropriately chosen learning method and its parameter set (iterations number, learning step, convergence guarantee, and other) can increase the generalization of the model and reduce the fusion error. The initialization of the learning method according to the knowledge of landcover classes, data type, or data acquisition parameters should lead to a faster convergence and an increase of the learning accuracy. Representation of input multisensory data on the alphabet reduces the data range and a chance on information loss can exist. A low size of the alphabet can lead to the fusion and classification error while a very high size can make the model intractable. To reduce the possibility of information loss the size of the alphabet should be chosen sufficient enough to preserve the necessary information from the input data. A relevant and sufficient method for the representation should be selected (knowledge on the data should be taken into account). The structure of a factor graph is one of the main aspects for an efficient solution. An inappropriate structure of the factor graph can increase the model complexity leading to intractability. The graph structure should be designed carefully to model the task and keep the tractability of the model. Inference method directly influences the decision making error in a factor graph. Recently developed approximate inference methods allow to benefit decision accuracy and inference time.

The fusion and classification model is managed to be augmented and developed in several directions. The first direction is on extension of the graph structure to preserve spatially uniform land-cover labeling and to include prior information on data to perform landcover extraction in unsupervised way. Another main direction is towards the employment of multisensory data and semantic information for modelling of evolution processes in a system. Higher levels of data interpretation are to be modeled by the employment of latent variables, approximate inference methods, and inference on non-full data. More thorough validation of the method is going to be performed on new available ground truth data.

APPENDIX

Detailed step-by-step message propagation for the independent model is as follows:

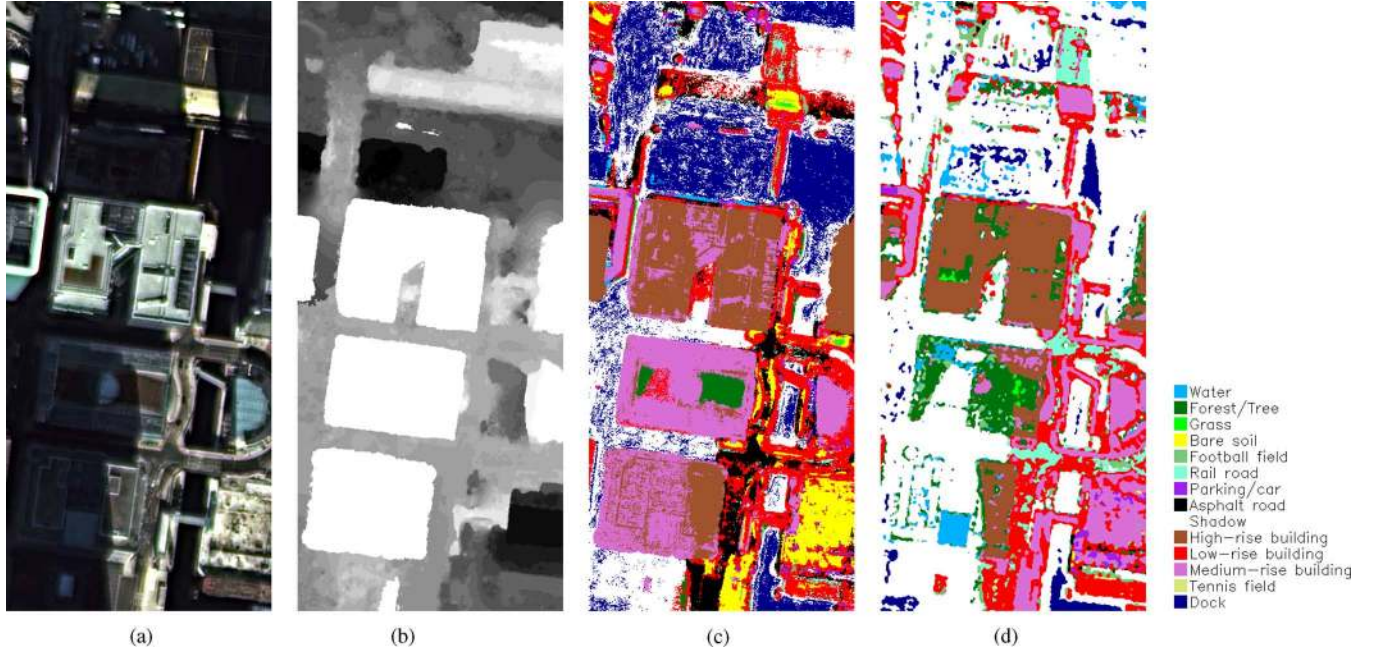


Fig. 17. London test area. An example of artifacts (shadowing) appearing in the area with high-rise buildings. Here, a wrong decision on a class is made by the neural network (shadowed building is classified as high-rise building, but not shadow; here a high influence of the DSM in the fusion and classification process). Also, the neural network classification demonstrate a confusion of shadow with class Dock. The same region is classified correctly by the FG fusion and classification method. (a)RGB, (b) DSM, (c) NN fusion and classification, (d) FG fusion and classification.

Step 1:

$$\begin{aligned} \text{msg}_{z_1 \rightarrow x_1}(x_1) &= \sum_{\sim\{x_1\}} z_1(x_1) = z_1(x_1), \\ &\vdots \\ \text{msg}_{z_n \rightarrow x_n}(x_n) &= \sum_{\sim\{x_n\}} z_n(x_n) = z_n(x_n); \\ \text{msg}_{z_c \rightarrow c_k}(c_k) &= \sum_{\sim\{c_k\}} z_c(c_k) = z_c(c_k). \end{aligned}$$

Step 2:

$$\begin{aligned} \text{msg}_{x_1 \rightarrow f_1}(x_1) &= \text{msg}_{z_1 \rightarrow x_1}(x_1), \\ &\vdots \\ \text{msg}_{x_n \rightarrow f_n}(x_n) &= \text{msg}_{z_n \rightarrow x_n}(x_n); \\ \text{msg}_{c_k \rightarrow f_1}(c_k) &= \text{msg}_{z_c \rightarrow c_k}(c_k), \\ &\vdots \\ \text{msg}_{c_k \rightarrow f_n}(c_k) &= \text{msg}_{z_c \rightarrow c_k}(c_k). \end{aligned}$$

Step 3:

$$\begin{aligned} \text{msg}_{f_1 \rightarrow c_k}(c_k) &= \sum_{\sim\{c_k\}} \text{msg}_{x_1 \rightarrow f_1}(x_1) f_1(x_1, c_k), \\ &\vdots \\ \text{msg}_{f_n \rightarrow c_k}(c_k) &= \sum_{\sim\{c_k\}} \text{msg}_{x_n \rightarrow f_n}(x_n) f_n(x_n, c_k); \\ \text{msg}_{f_1 \rightarrow x_1}(x_1) &= \sum_{\sim\{x_1\}} \text{msg}_{c_k \rightarrow f_1}(c_k) f_1(x_1, c_k), \\ &\vdots \\ \text{msg}_{f_n \rightarrow x_n}(x_n) &= \sum_{\sim\{x_n\}} \text{msg}_{c_k \rightarrow f_n}(c_k) f_n(x_n, c_k). \end{aligned}$$

Step 4:

$$\begin{aligned} \text{msg}_{c_k \rightarrow z_c}(c_k) &= \text{msg}_{f_1 \rightarrow c_k}(c_k) \dots \text{msg}_{f_n \rightarrow c_k}(c_k); \\ \text{msg}_{x_1 \rightarrow z_1}(x_1) &= \text{msg}_{f_1 \rightarrow x_1}(x_1), \\ &\vdots \\ \text{msg}_{x_n \rightarrow z_n}(x_n) &= \text{msg}_{f_n \rightarrow x_n}(x_n). \end{aligned}$$

ACKNOWLEDGMENT

The authors would like to thank European Space Imaging (EUSI) for the collection and provision of Digitalglobe WorldView-2 and IKONOS-2 data over Munich city. TerraSAR-X data were provided by DLR through the Science Projects MTH0505 and MTH0948.

REFERENCES

- [1] B. Waske and S. van der Linden, "Classifying multilevel imagery from SAR and optical sensors by decision fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1457–1466, 2008.
- [2] B. Waske and J. Benediktsson, "Fusion of support vector machines for classification of multisensor data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3858–3866, 2007.
- [3] F. Pacifici, F. D. Frate, W. Emery, P. Gamba, and J. Chanussot, "Urban mapping using coarse SAR and optical data: Outcome of the 2007 GRSS data fusion contest," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 331–335, 2008.
- [4] N. Longbotham, F. Pacifici, T. Glenn, A. Zare, M. Volpi, D. Tuia, E. Christophe, J. Michel, J. Inglada, J. Chanussot, and Q. Du, "Multimodal change detection, application to the detection of flooded areas: Outcome of the 2009–2010 data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. (JSTARS)*, vol. 5, no. 1, pp. 331–342, 2012.

- [5] F. Tupin and M. Roux, "Detection of building outlines based on the fusion of SAR and optical features," *ISPRS J. Photogramm. Remote Sens.*, vol. 58, no. 1–2, pp. 71–82, 2003.
- [6] J. Benediktsson, P. Swain, and O. Ersoy, "Neural network approaches versus statistical methods in classification of multisource remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 4, pp. 540–552, Jul. 1990.
- [7] J. Benediktsson, J. Sveinsson, and P. Swain, "Hybrid consensus theoretic classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 4, pp. 833–843, Jul. 1997.
- [8] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Decision fusion for the classification of urban remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2828–2838, 2006.
- [9] F. Rottensteiner, J. Trinder, S. Clode, K. Kubik, and B. Lovell, "Building detection by Dempster-Shafer fusion of LIDAR data and multispectral aerial imagery," in *Proc. 17th Int. Conf. Pattern Recognition*, Los Alamitos, CA, 2004, vol. 2, pp. 339–342.
- [10] C. Wang, D. Blei, and L. Fei-Fei, "Simultaneous image classification and annotation," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2009, pp. 1903–1910.
- [11] M. Lienou, H. Maitre, and M. Datcu, "Semantic annotation of satellite images using latent Dirichlet allocation," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 1, pp. 28–32, 2010.
- [12] D. Bratananu, I. Nedelcu, and M. Datcu, "Bridging the semantic gap for satellite image annotation and automatic mapping applications," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. (JSTARS)*, vol. 4, no. 1, pp. 193–204, Mar. 2011.
- [13] M. Datcu, H. Daschiel, A. Pelizzari, M. Quartulli, A. Galoppo, A. Coplacichioni, M. Pastori, K. Seidel, P. Marchetti, and S. D'Elia, "Information mining in remote sensing image archives: System concepts," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 12, pp. 2923–2936, 2003.
- [14] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sumproduct algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
- [15] B. Frey and N. Jovic, "A comparison of algorithms for inference and learning in probabilistic graphical models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 27, no. 9, pp. 1392–1416, Sept. 2005.
- [16] J. Moura, L. Jin, and M. Kleiner, "Intelligent sensor fusion: A graphical model approach," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Hong Kong, China, Apr. 2003, vol. 6, pp. VI–733–6, vol. 6.
- [17] K. Kampa, J. Principe, and K. Slatton, "Dynamic factor graphs: A novel framework for multiple features data fusion," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Dallas, TX, US, 2010, pp. 2106–2109.
- [18] M. R. Naphade, I. Kozintsev, and T. Huang, "Factor graph framework for semantic video indexing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 1, pp. 40–52, Jan. 2002.
- [19] G. Palubinskas and M. Datcu, "Information fusion approach for the data classification: An example for ERS-1/2 InSAR data," *Int. J. Remote Sens.*, vol. 29, no. 16, pp. 4689–4703, 2008.
- [20] G. Palubinskas, W. von der Linden, Ed. *et al.*, "An unsupervised clustering method by entropy minimization," in *Maximum Entropy and Bayesian Methods*. Boston, MA: Kluwer Academic, 1999, pp. 327–334.
- [21] S. Aksoy, K. Koperski, C. Tusk, G. Marchisio, and J. Tilton, "Learning Bayesian classifiers for scene classification with a visual grammar," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 581–589, 2005.
- [22] M. Schroder, H. Rehrauer, K. Seidel, and M. Datcu, "Interactive learning and probabilistic retrieval in remote sensing image archives," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 5, pp. 2288–2298, 2000.
- [23] I. O. Kyrgyzov, O. O. Kyrgyzov, H. Maître, and M. Campedel, "Kernel mdl to determine the number of clusters," in *Proc. 5th Int. Conf. Machine Learning and Data Mining in Pattern Recognition*, 2007, pp. 203–217.
- [24] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. Kschischang, "The factor graph approach to model-based signal processing," *Proc. IEEE*, vol. 95, no. 6, pp. 1295–1322, Jun. 2007.
- [25] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Statist. Soc. Series B (Method.)*, vol. 36, no. 2, pp. 192–236, 1974.
- [26] H. Shen, J. Coughlan, and V. Ivanchenko, "Figure-ground segmentation using factor graphs," *Image Vis. Comput.*, vol. 27, no. 7, pp. 854–863, 2009.
- [27] B. J. Frey, N. Mohammad, Q. D. Morris, W. Zhang, M. D. Robinson, S. Mnaimneh, R. Chang, Q. Pan, E. Sat, J. Rossant, B. G. Bruneau, J. E. Aubin, B. J. Blencowe, and T. R. Hughes, "Genome-wide analysis of mouse transcripts using exon microarrays and factor graphs," *Nature Genetics*, vol. 37, no. 9, pp. 991–996, 2005.
- [28] M. R. Boutell, J. Luo, and C. M. Brown, "Factor graphs for region-based whole-scene classification," in *Proc. 2006 Conf. Computer Vision and Pattern Recognition Workshop*, Washington, DC, 2006, pp. 104–111.
- [29] S. Roth, "High-Order Markov Random Fields for Low-Level Vision," Ph.D. dissertation, Brown Univ., Providence, RI, May 2007.
- [30] G. Palubinskas, P. Reinartz, and R. Bamler, "Image acquisition geometry analysis for the fusion of optical and radar remote sensing data," *Int. J. Image and Data Fusion*, vol. 1, no. 3, pp. 271–282, 2010.
- [31] G. Palubinskas and P. Reinartz, "Multi-resolution, multi-sensor image fusion: general fusion framework," in *Proc. Joint Urban Remote Sensing Event 2011 (JURSE 2011)*, Munich, Germany, 2011, pp. 313–316.
- [32] S. Suri and P. Reinartz, "Mutual-information-based registration of TerraSAR-X and IKONOS imagery in urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 2, pp. 939–949, 2010.
- [33] Digitalglobe, WorldView-2 Spacecraft Data Sheet.
- [34] P. d'Angelo and P. Reinartz, "Semiglobal matching results on the ISPRS stereo matching benchmark," in *ISPRS Hannover Workshop 2011: High-Resolution Earth Imaging for Geospatial Information*, Hannover, Germany, Jun. 2011.
- [35] J. Daugman, "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 7, pp. 1169–1179, 1988.
- [36] I. Young, L. van Vliet, and M. van Ginkel, "Recursive Gabor filtering," *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2798–2805, Nov. 2002.
- [37] W. Heldens, U. Heiden, M. Bachmann, T. Esch, A. Müller, and S. Dech, "Scaling issues in validation of abundance maps derived from HyMAP data of an urban area," in *6th EARSeL SIG IS Workshop on Imaging Spectroscopy*, Tel Aviv, Israel, Mar. 16–19, 2009.
- [38] M. Canty, *Image Analysis, Classification, and Change Detection in Remote Sensing: With Algorithms for ENVI/IDL*, 2nd ed. Boca Raton, FL: CRC Press, 2009.
- [39] W. Heldens, "Use of Airborne Hyperspectral Data and Height Information to Support Urban Micro Climate Characterisation," Ph.D. dissertation, Universität Würzburg, Würzburg, Germany, 2010.
- [40] T. G. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," *Neural Comput.*, vol. 10, pp. 1895–1923, 1998.
- [41] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [42] N. Longbotham, C. Chaapel, L. Bleiler, C. Padwick, W. Emery, and F. Pacifici, "Very high resolution multiangle urban classification analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1155–1170, Apr. 2012.



Aliaksei Makarau received the Diploma (Dipl.-Ing.) in computer science from Belarusian State University of Informatics and Radioelectronics in 2003, and the Ph.D. in technical sciences from the United Institute of Informatics Problems, Minsk, Belarus, in 2008. His dissertation was on fast methods for multispectral image fusion and processing.

He is at the Department of Photogrammetry and Image Analysis at the German Aerospace Centre (DLR), Remote Sensing Technology Institute (IMF). His research interests include atmospheric correction, multimodal data fusion, and pattern recognition.



Gintautas Palubinskas received the M.S. and Ph.D. degrees in mathematics from Vilnius University, Lithuania, in 1981 and the Institute of Mathematics and Informatics (IMI), Vilnius, in 1991. His doctoral dissertation was on spatial image recognition.

He was a Research Scientist at the IMI from 1981 to 1997. Since 1997, he has been a Research Scientist at Remote Sensing Technology Institute, German Aerospace Center (DLR), Oberpfaffenhofen, Germany. From 1993 to 1997, he was a Visiting Research Scientist at German Remote Sensing Data Center, DLR; the Department of Geography, Swansea University, Wales, U.K.; Institute of Navigation, Stuttgart University, Germany; and Max-Planck-Institute of Cognitive Neuroscience, Leipzig, Germany. His current interests are in image processing, classification and change detection, data fusion for optical and SAR remote sensing applications.



Peter Reinartz received the Diploma (Dipl.-Phys.) in theoretical physics from the University of Munich, Germany, in 1983, and the Ph.D. (Dr.-Ing) in civil engineering from the University of Hannover, Germany, in 1989. His dissertation was on statistical optimization of classification methods for multispectral image data.

He is head of the Department of Photogrammetry and Image Analysis at the German Aerospace Centre (DLR), Remote Sensing Technology Institute (IMF) and holds a professorship in Geoinformatics at the University of Osnabrueck. He has more than 20 years of experience in image processing and remote sensing and over 150 publications in these fields. His main interests are in direct georeferencing, stereo photogrammetry and data fusion, generation of digital elevation models and interpretation of VHR image data from sensors like WorldView, GeoEye a.o. He is also engaged in using remote sensing data for disaster management and using high frequency time series of airborne image data for real time application in case of disasters as well as for traffic monitoring.