

Discrete Graphical Models for Alphabet-based Multisensory Data Fusion and Classification

Aliaksei Makarau, Gintautas Palubinskas and Peter Reinartz
 German Aerospace Center (DLR)
 Remote Sensing Technology Institute (IMF)
 82234 Oberpfaffenhofen, Germany
 {aliaksei.makarau, gintautas.palubinskas, peter.reinartz}@dlr.de

Abstract—The way of multisensory data integration is a crucial step of any data fusion method. Different physical types of sensors (optic, thermal, acoustic, radar, etc.), different resolution, and different types of GIS digital data (elevation, vector maps, etc.) 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. Correct and established way of multisensory data integration is required to deal with such incommensurable data, while employment of an inappropriate methodology may lead to errors in the fusion.

To perform a proper multisensory data fusion several methods were developed (weighted Bayesian, linear (log linear) opinion pool [1], [2], neural networks [1]–[3], fuzzy logic approaches [4], etc.). Employment of these approaches is motivated by weighted consensus theory, leading the fusion of incommensurable data to be performed in a correct way.

In this paper data fusion is proposed to perform using a finite predefined domain – alphabet. Feature extraction (data fission) is performed separately on different sources of data. Extracted features are processed to be represented on the predefined domain (alphabet). Alternative method such as factor graph (discrete graphical model) is employed for data and feature aggregation. The nature of factor graphs in application on data coded on a finite domain allows us to obtain an improvement in accuracy of real data fusion and classification for multispectral high resolution WorldView-2, TerraSAR-X SpotLight, and elevation model.

Index Terms—Multisensor data, fusion, classification, graphical models, factor graphs, WorldView-2, TerraSAR-X

I. INTRODUCTION

The practical use of spaceborne very high resolution multispectral data is still growing (IKONOS, Quickbird, GeoEye-1, etc.) but for classification purposes the number of bands is limited in comparison to full spectral imaging. These limitations may lead to the confusion of materials such as roofs, pavements, roads, etc. and therefore may provide wrong interpretation and classification of objects. Employment of hyperspectral data is another solution, but their low spatial resolution (comparing to multispectral data) restrict their usage for many applications.

An improvement can be achieved by employment of fusion approaches of multisensory data since this may increase the quality of scene classification. Integration of Synthetic Aperture Radar (SAR) and optical data is widely performed for automatic classification, interpretation, and change detection.

Incommensurability of different sources of data (e.g. optical, SAR, and DEM) requires a proper design of fusion process. Benediktsson et al. [1], [2] investigated statistical versus neural network approaches for multisensory data fusion and classification. Linear opinion pool, logarithmic opinion pool, and logarithmic opinion pool optimized by multilayer neural network are proposed for combination of multisensory data (multispectral, elevation, slope, aspect, and SAR) [1], [2]. Several approaches for multisensory data fusion following consensus theory and employing different techniques such as Bayesian networks, neural networks, etc. were developed.

II. ALPHABET-BASED DATA FUSION

Instead of continuous representation of data a discrete representation of the data on a finite domain (alphabet) is proposed for use. Discrete representation is motivated by a fact that fusion of multisensory data with different nature and statistics could be difficult using conventional statistical methods. To overcome this difficulty, a kind of “discretization” of continuous data is employed resulting in data with several possible states (i.e. multinomial distribution, see example [5]). Fusion framework consists of three main steps:

- 1) **Information fission: feature extraction from input datasets.** The aim of this step is to extract as much as possible information from input data [6]. 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 stage is to represent a feature on a finite predefined domain – alphabet. Here, objects with similar properties are grouped and data or feature dimensionality is reduced. Unsupervised clustering is used (k -means, entropy based k -means [7]).
- 3) **Fusion and classification of coded data or features** is performed using factor graphs (FG) [8], [9]. Configurations (parameter sets) of the FGs are calculated according to supervisory selected classes and training areas. Configured FGs are used for inference on evidence data (i.e. clustered input features). Maximal likelihood probability or calculated marginal probability of a latent variable is employed for data classification.

III. GRAPHICAL MODEL DEFINITION

A. Factor graphs for discrete data

Factor graphs are more general graphical models than Bayesian networks or Markov random fields (Bayesian networks and Markov random fields are particular cases of corresponding factor graphs). An FG possess properties of Bayesian network and Markov random field and allows to describe more complex relationships among parts of a modeled system.

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

$$\arg \max_k P(c_k|E), \quad (1)$$

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 [10]:

$$P(c_k)P(E|c_k)/P(E), \quad (2)$$

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).

Assuming that the factor graph is configured (θ_k is the configuration for class k) the joint distribution of $P(E|c_k)$ is the following [10]:

$$P(c_k)P(E, \theta_k|c_k), \quad (3)$$

where the θ_k is the parameter for the k -th class estimation (i.e. different θ_k are used for different classes).

Independent structure graphical model can be defined for (2), illustrated in Fig. 1. The independent model was selected since it has a low complexity leading to easier configuration and inference comparing to the exact model.

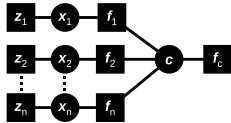


Fig. 1. Fusion modelling using factor graph: independent model (low complexity)

The independent model factor graph $g_{independent}$ ($g_{independent}$ is a factorization function) (Fig. 1) can be described as:

$$g_{independent}(x_1, x_2, \dots, x_n) = f_c(c_k) f_1(x_1, c_k) f_2(x_2, c_k) \dots f_n(x_n, c_k) z_1(x_1) z_2(x_2) \dots z_n(x_n) \quad (4)$$

where the functors f_1, f_2, \dots, f_n are the functors of the features x_1, x_2, \dots, x_n and share the class variable c_k ($P(E|c_k)$ in (2)); $z_1(x_1)z_2(x_2) \dots z_n(x_n)$ are the functors defining prior probabilities $\frac{1}{p(x_1)}, \frac{1}{p(x_2)}, \dots, \frac{1}{p(x_n)}$ (normalizing functors, $P(E)$ in (2)); c_k is the k -th class variable; f_c is the functor defining prior probability for class variable c ($P(c_k)$ in (2)).

B. Model learning: graph configuration

The structure of a factor graph defines a dependency of class variable node on input features. Use of training data allows to calculate a configuration (parameter set θ_k) for a factor graph for each class k , or the θ is calculated for one factor graph to obtain a marginal distribution $P(c_k|E)$ for a single class variable node k after inference.

A configured factor graph with the configuration θ_k is expected to have maximum likelihood probability (low energy state) on the evidence which most likely (similar) to the employed training data. Expectation maximization (EM) method is employed for learning the graph configuration. Belief propagation method is employed for inference (a configuration of latent variables is used to maximize the posterior probability).

IV. EXPERIMENTS, RESULTS, AND DISCUSSION

A. Multisensory data and features

Employment of SAR data together with optical data requires proper registration as well as a suitable acquisition model [11]. Acquired WorldView-2 (WV-2) (10-th July 2010, 10:30:17, Look angle 5.2° Left) and TerraSAR-X (TSX) (7-th June 2008, 05:17:48, Look angle 49.2218°, Right) data possess acquisition model requirements. The data were acquired for Munich city. WorldView-2 multispectral data were pan-sharpened by the General Fusion Framework method [12]. The optical and SAR data were orthorectified (SRTM 30m DEM) and distortions introduced by terrain are decreased. Orthorectified WorldView-2 and SpotLight Level-1B Product TerraSAR-X data were used. Registration of optical and radar data was made semi-automatically using an automatic registration method [13].

Highly detailed Digital Surface Model of urban scene is generated using the Semiglobal Matching algorithm using Worldview-2 stereo pairs or triplets with small convergence angles (less then 20 degrees) [14].

TSX image is employed for characterization of objects surface structure and textural properties (e.g. grass land versus football field, bare soil versus construction site, etc.). Multispectral data is also used for textural feature extraction and for providing spectral information on the objects of a scene. In our experiment Gabor features [15] were calculated on TSX data and on Red color channel from WV-2 data. A bank of gabor wavelets consists of 24 filters (6 orientations ($0, \pi/6, \pi/3, \pi/2, \frac{2}{3}\pi, \frac{5}{6}\pi$), 4 different periods of filter's sine component ($\pi/4, \pi/2, \frac{3}{4}\pi, \pi$), and 1 sigma value ($\sigma = 4$)), recursive implementation of Gabor filtering is employed [16].

A subscene (7115×4516) was used in the experiments. The number of clusters for feature representation on the alphabet

was set to 10 (was selected for experimental purposes to illustrate that even the low size of the alphabet allows to obtain a comparable fusion and classification accuracy). The alphabet size is used for representation of all features.

B. Fusion strategies and classification

One of the main interests is to compare the influence of data fusion for classification accuracy, and to compare fusion with single sensor classification results. The following combinations of multisensory and single-sensor data can be created:

- 1) WV-2 (single sensor, 8 features (spectral bands)),
- 2) WV-2 + DSM (9 features),
- 3) WV-2 + Texture (SAR and Optical) + DSM (57 features).

Altogether, 23 classes were defined: 1-Water, 2-Forest/Trees, 3-Grass/Low vegetation, 4-Bare soil, 5-Construction site, 6-Swimming pool, 7-Asphalt road, 8-Football field, 9-Tennis field, 10-Green house, 11-Rail road, 12-Tram line, 13-Cemetery, 14-Parking/car, 15-Shadow, 16-Concrete, 17-Red roofing tiles, 18-Roofing concrete, 19-Vegetation roof, 20-Dark roofing tiles, 21-Zinc roof, 22-Roofing copper, 23-Grey roofing tiles. Selection of training and testing regions was made manually according to available ground truth data. It should be noted that the validated ground truth is limited in size (e.g. vector data on classes 12, 16-23 is available only for a small number of objects and buildings). The ground truth for the area under investigation was proofed by the ATKIS vector map provided by Bavarian State Agency for Surveying and Geoinformation (Landesamt für Vermessung und Geoinformation). Vector data on the materials available in the scene was created and provided by Dr. Wieke Heldens [17].

C. Results and discussion

Table I presents results (overall accuracy and Cohen's Kappa) for fusion and classification of multisensory for single sensor data. Results of two other methods: Maximum Likelihood (ML) (not following consensus theory) and Neural Network (NN) are also given for comparison. Neural network (multilayer perceptron) was chosen since it was shown to be an efficient solution for multisensory data fusion and provides high accuracy of classification [1], [3]. Neural Network employs 2 hidden layers, 60 neurons in each layer. The ML was run in the ENVI software. The ML and NN use original data without representation on the alphabet.

Subscene of the classification maps produced by the ML, NN, and the FG fusion using WV-2+Texture+DSM are presented in Fig. 2.

Feature representation on a finite domain allows to convert incommensurable features and data with different statistical properties and distributions into one type of distribution (e.g. multinomial distribution [5]). Fusion of multisensory data using FG method (OVA=76.77, Kappa=0.7449) allowed to obtain comparable accuracy to the fusion and classification

TABLE I
CLASSIFICATION ACCURACY USING DIFFERENT METHODS TOGETHER WITH THE PROPOSED APPROACH. ML - MAXIMUM LIKELIHOOD (NOT FOLLOWING CONSENSUS THEORY), NN - NEURAL NETWORK, FG - FACTOR GRAPH (THE SIZE OF THE ALPHABET IS 10). 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 (57)	58.53	0.5502
NN	WV-2 (8)	67.17	0.6448
NN	WV-2+DSM (9)	71.34	0.6881
NN	WV-2+Texture+DSM (57)	74.88	0.7251
FG	WV-2 (8)	47.89	0.4536
FG	WV-2+DSM (9)	52.46	0.4909
FG	WV-2+Texture+DSM (57)	76.77	0.7449

results obtained by the two layer neural network (OVA=74.88, Kappa=0.7251).

Low accuracies of the ML classification on WV-2+Texture+DSM data may be caused that the ML classifier can not efficiently deal with different distributions of data and features, or the multisensory data is not classified in the way of consensus classification [2]. The ML provides a high confusion of building with construction site in the whole scene.

A 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 (i.e. the of clusters) is low. Therefore, a loss of multispectral information after clustering influences fusion and classification accuracy comparing to the methods dealing with the original 11-bit single source data.

The FG fusion and classification approach allows to receive a comparable accuracy even on small size of the representation domain (in the paper experiment the domain size is 10). Independent model of the factor graph is easy to configure and apply for real data. It is expected that an increase of the domain size should allow to increase the accuracy of fusion and classification, or to work with higher number of classes or to define more difficult classes of interest (high inhomogeneity, spectral properties).

V. CONCLUSION

Among not many choices for multisensory data fusion, selection of factor graphs for the fusion allows to perform a classification into an extended set of classes and provides a possibility for fusion model development to define more complex systems.

Representation of multisensory data and extracted features using an alphabet (a domain with finite states) allows to deal with incommensurable features and data with different nature, statistical properties and distributions. Such representation (e.g. representation of data using multinomial distribution [5]) makes possible and easier processing of data using aggregation methods (i.g. factor graphs). The data classification is not influenced by the limitations of dimensionality (i.e. there is no the curse of dimensionality).

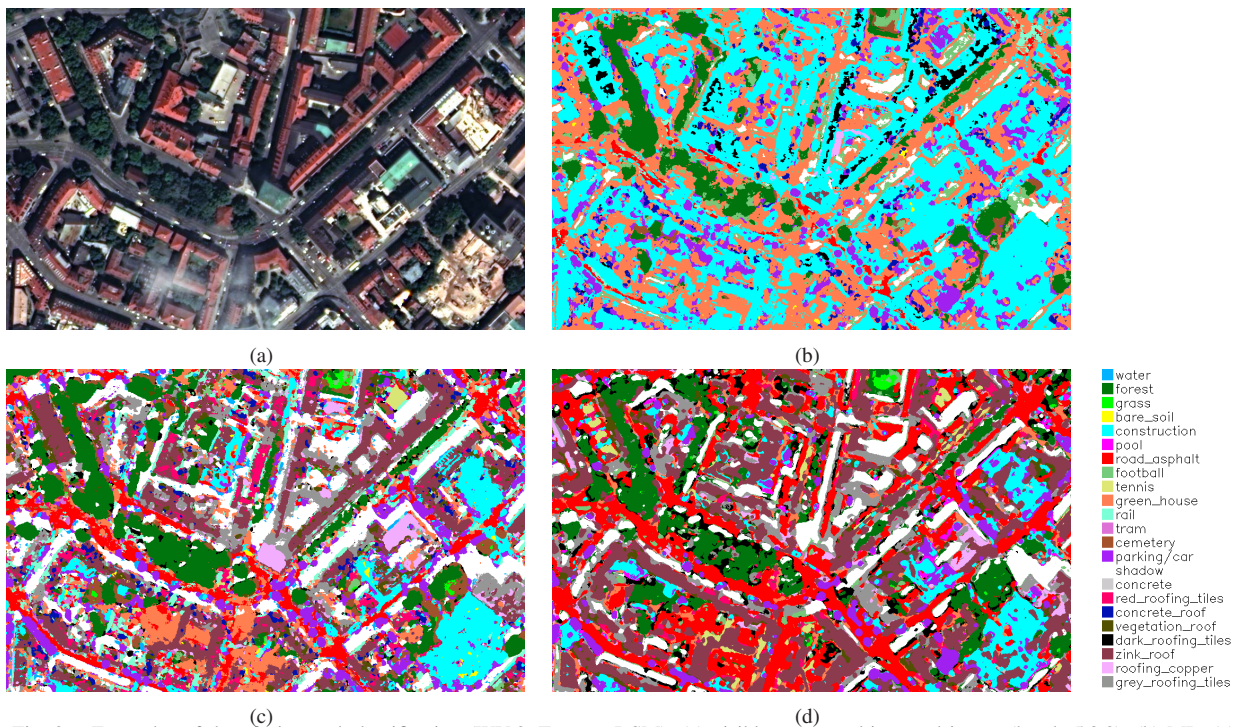


Fig. 2. Examples of data fusion and classification (WV-2+Texture+DSM): (a) visible range multispectral image (bands 5,3,2), (b) ML, (c) NN, (d) FG

Proper selection of factor graph structure allows to fusion model tractable for real use and application for remotely sensed data of arbitrary size. Approximate inference on independent model comparing to exact model makes possible to perform inference on input evidence of moderately high size (over 100 features).

ACKNOWLEDGMENT

This work was supported by the DLR-DAAD research grant (A/09/95629). We would like to thank European Space Imaging (EUSI) for provision of DigitalGlobe WorldView-2 data. TerraSAR-X data were provided by DLR through the Science Projects MTH0505 and MTH0948.

We would thank Dr. Pablo d' Angelo for generation of DEM data and valuable comments and recommendations.

REFERENCES

- [1] J. Benediktsson, P. Swain, and O. Ersoy, "Neural network approaches versus statistical methods in classification of multisource remote sensing data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, no. 4, pp. 540–552, Jul. 1990.
- [2] J. Benediktsson, J. Sveinsson, and P. Swain, "Hybrid consensus theoretic classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 833–843, Jul. 1997.
- [3] F. Pacifici, F. Del 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 Geoscience and Remote Sensing Letters*, vol. 5, no. 3, pp. 331–335, 2008.
- [4] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Decision fusion for the classification of urban remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 10, pp. 2828–2838, 2006.
- [5] S. Aksoy, K. Koperski, C. Tusk, G. Marchisio, and J. Tilton, "Learning bayesian classifiers for scene classification with a visual grammar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 581–589, 2005.
- [6] G. Palubinskas and M. Datcu, "Information fusion approach for the data classification: an example for ERS-1/2 InSAR data," *International Journal of Remote Sensing*, vol. 29, no. 16, pp. 4689–4703, 2008.
- [7] G. Palubinskas, "An unsupervised clustering method by entropy minimization," in *Maximum Entropy and Bayesian Methods*, W. von der Linden et al., Eds., Kluwer Academic Publisher, 1999, pp. 327–334.
- [8] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
- [9] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. Kschischang, "The factor graph approach to model-based signal processing," *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1295–1322, Jun. 2007.
- [10] M. R. Boutell, J. Luo, and C. M. Brown, "Factor graphs for region-based whole-scene classification," in *Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop*, ser. CVPRW '06, Washington, DC, USA: IEEE Computer Society, 2006, p. 104.
- [11] G. Palubinskas, P. Reinartz, and R. Bamler, "Image acquisition geometry analysis for the fusion of optical and radar remote sensing data," *International Journal of Image and Data Fusion*, vol. 1, no. 3, pp. 271–282, 2010.
- [12] G. Palubinskas and P. Reinartz, "Multi-resolution, multi-sensor image fusion: general fusion framework," in *Joint Urban Remote Sensing Event 2011 (JURSE 2011)*, Munich, Germany, 2011, pp. 313–316.
- [13] S. Suri and P. Reinartz, "Mutual-information-based registration of TerraSAR-X and IKONOS imagery in urban areas," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 2, pp. 939–949, 2010.
- [14] 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, June 2011.
- [15] J. Daugman, "Complete discrete 2-d gabor transforms by neural networks for image analysis and compression," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 7, pp. 1169–1179, 1988.
- [16] I. Young, L. van Vliet, and M. van Ginkel, "Recursive Gabor filtering," *IEEE Transactions on Signal Processing*, vol. 50, no. 11, pp. 2798–2805, Nov. 2002.
- [17] 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, March 16th-19th 2009.