FACTOR GRAPH MODELS FOR MULTISENSORY DATA FUSION: FROM LOW-LEVEL FEATURES TO HIGH LEVEL INTERPRETATION

Aliaksei Makarau, Gintautas Palubinskas, and Peter Reinartz

Photogrammetry and Image Analysis
Remote Sensing Technology Institute
German Aerospace Center DLR
Oberpfaffenhofen, D-82234 Wessling, Germany

ABSTRACT
A solution of difficult tasks in remotely sensed data information extraction can be reached by the development of more complex models. The most important step is in the selection of a relevant and universal methodology for data interpretation, classification, fusion, object detection, etc. Probabilistic graphical models [1] become a more and more popular way for image data annotation and classification [2, 3]. Factor graphs possess important properties such as probabilistic nature, explicit factorization properties, approximate inference, plausible inference of non-full data, easy augmenting, etc., and become relevant for the use in data interpretation systems.

In this paper we present several applications of factor graphs for single/multisensory data fusion, classification, and an extension of the graph structure to extract landcover from unseen data. The application of factor graphs allow to obtain an improvement in data fusion/classification accuracy.

Index Terms— Fusion, classification, factor graphs, WorldView-2

1. INTRODUCTION

Factor graph is a more general graphical model than Bayesian network or Markov random field. A factor graph (FG) possesses properties of Bayesian network and Markov random field and allows to describe more complex relationships among parts of a modeled system. A factor graph is a bipartite graph containing two types of nodes: variable nodes \( x_i, i = 1..n \) and factor nodes \( f_j(x_1, x_2, \ldots, x_n), j = 1..m \), where a variable node \( x_i \) takes value on a finite domain [4]. Figure 1 presents an example of a factor graph with three variables \( x_1, x_2, x_3 \) and two factors \( f_1(x_1, x_2) \) and \( f_2(x_2, x_3) \).

Explicit factorization properties of factor graphs allow to develop complex models to perform a desired interpretation of single/multisensory input data. Configuration (learning) of the model on training data and further inference by approximate inference methods allow to reach plausible decisions.

2. MULTISENSORY DATA FUSION

The fusion framework consists of three main steps: 1. Information fission: feature extraction from input data (to provide the quasi-full description of the scene [5]). 2. Feature representation on the alphabet (to represent a feature on a finite predefined domain). This transformation makes another level of feature abstraction, generalized by similarity in the feature space, simultaneously performing data reduction. This representation is made using unsupervised clustering (e.g. \( k \)-means). 3. Fusion and classification of the represented features is performed using a factor graph [4, 6]. Configuration (learning) of the FG is calculated according to supervised selected training samples. Configured FG is used for fusion of input data (inference on evidence). 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. The joint probability mass function for the fusion and classification can be defined as:

\[
p(x|c) = \prod_{k=1}^{K} \frac{1}{p(c_k)} \prod_{i=1}^{N} \frac{1}{p(x_i)} p(x_i|c_k),
\]  

Fig. 1. An example of a simple factor graph with three variables \( x_1, x_2, x_3 \) and two factors \( f_1(x_1, x_2) \) and \( f_2(x_2, x_3) \)

We would like to thank European Space Imaging (EUSI) for the collection and provision of WorldView-2 data. We would like to thank DigitalGlobe for providing the imagery used in this study, and the IEEE GRSS Data Fusion Technical Committee for organizing the Data Fusion Contest.
where $\mathbf{x}$ is the input feature vector ($x_i$ is the $i$-th input feature variable); $c_k$ is the $k$-th class variable; $\frac{1}{p(x_i)}$ and $\frac{1}{p(c_k)}$ are normalizing functions. The factorization of the function (factor graph in Figure 2) is as follows:

$$g(\mathbf{x}, c) = \prod_{k=1}^{K} \prod_{n=1}^{N} z_i(x_i) f_i(x_i, c_k) z_c(c_k), \quad (2)$$

where $\mathbf{x}$ is the input evidence (input feature vector); $f_i(x_i, c_k)$ is the $i$-th factor of $i$-th input feature and class variable; $z_i(x_i) = \frac{1}{p(x_i)}$ and $z_c(c_k) = \frac{1}{p(c_k)}$ are normalizing factors.

Fig. 2. Factor graph model (plate notation) for multisensory data fusion (boxes are representing replicates)

A combination of WorldView-2 multispectral data and a digital surface model is used (London city) for an experiment. Acquisition geometry of the employed WorldView-2 data is $6.3^\circ$ off-nadir view angle. WorldView-2 multispectral data are pan-sharpened by the General Fusion Framework method [7]. The optical data are orthorectified. Gabor features are calculated on Red color channel (630-690 nm) from WorldView-2 data. A bank of Gabor wavelets consists of 18 filters (6 orientations ($0, \pi/6, \pi/3, \pi/2, \frac{2\pi}{3}, \frac{5\pi}{6}, \pi$), 3 different periods of filter’s sine component ($\pi/3, \frac{2\pi}{3}, \pi$), and 1 sigma value ($\sigma = 4$)). A subscale (2880 × 2815 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 optical data), 10 for the DSM, and 50 for the WV-2 multispectral data. Altogether, 14 classes are defined: 1-Water, 2-Forest/Trees, 3-Grass/Low vegetation, 4-Bare soil, 5-Football field, 6-Rail road, 7-Parking/car, 8-Asphalt road, 9-Shadow, 10-High-rise building, 11-Low-rise building, 12-Medium-rise building, 13-Tennis field, 14-Dock. Selection of training and test regions is made manually on a color composite of WorldView-2 and Bing maps. The training and test samples are spatially uncorrelated. Configuration is performed using gradient ascent method by expectation-maximization (EM).

Table 1 presents results (OVerall Accuracy and Cohen’s Kappa) for fusion and classification of multisensory and single sensor data. Results of two other methods: Maximum Likelihood (ML) (not following consensus theory) and Neural Network (NN) are also given for comparison (the ML and NN were run in ENVI). Neural network (multilayer perceptron) is chosen since it is shown to be an efficient solution for multisensory data fusion and provides high accuracy of classification [8, 9]. Neural Network employs 2 hidden layers, 20 neurons in each layer. The ML and NN use original data without presentation on the alphabet. Learning time of the FG is faster than NN, the FG inference is slower. A classification map produced by the FG fusion using WV-2+Texture+DSM is presented in Fig. 3. The FG allows better accuracy of the fusion and classification of the multisensory data.

### Table 1. Classification accuracy using different methods together with the FG approach

<table>
<thead>
<tr>
<th>Method</th>
<th>Employed features</th>
<th>OA %</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML</td>
<td>WV-2+Texture+DSM (27)</td>
<td>78.46</td>
<td>0.7159</td>
</tr>
<tr>
<td>ML</td>
<td>WV-2 (8)</td>
<td>70.55</td>
<td>0.6208</td>
</tr>
<tr>
<td>NN</td>
<td>WV-2+Texture+DSM (27)</td>
<td>66.10</td>
<td>0.6693</td>
</tr>
<tr>
<td>FG</td>
<td>WV-2+Texture+DSM</td>
<td>70.05</td>
<td>0.6632</td>
</tr>
</tbody>
</table>

### 3. LANDCOVER EXTRACTION FROM NOT ANNOTATED DATA

Classification of a new image normally requires selection of training samples (being time-consuming). To automate this step a defined and configured model on a database of previously annotated images can be used. A knowledge on the distribution of features in a landcover class together with a distribution of classes in an image should be employed. The joint probability mass function for the landcover extraction model can be defined as:

$$p(\mathbf{x} | \mathbf{d}) = \prod_{m=1}^{M} \frac{1}{p(d_m)} \prod_{i=1}^{N} \frac{1}{p(x_i)} p(x_i | c_k) \prod_{k=1}^{K} \frac{1}{p(c_k)} p(c_k | d_m), \quad (3)$$

where $\mathbf{x}$ is the input feature vector ($x_i$ is the $i$-th feature variable); $c_k$ is the $k$-th class variable in the image $d_m$; $d_m$ is the $m$-th image in image database; $\frac{1}{p(d_m)}$, $\frac{1}{p(x_i)}$, and $\frac{1}{p(c_k)}$ are normalizing functions. $p(x_i | c_k)$ is the probability of the $x_i$ feature contributing to the $c_k$ class, $p(c_k | d_m)$ is the probability of the $c_k$ class in the $d_m$ image. The factorization of the function (factor graph in Figure 4) is as follows:

$$g(\mathbf{x}, \mathbf{c}, \mathbf{d}) = \prod_{m=1}^{M} z_d(d_m) \prod_{i=1}^{N} z_i(x_i) f_i(x_i, c_k) \prod_{k=1}^{K} f_d_m(c_k, d_m), \quad (4)$$

where $\mathbf{x}$ is the input evidence (input feature vector); $f_i(x_i, c_k)$ is the $i$-th factor of the $i$-th input feature and class variable.
c_k (the c_k in the m-th image); f_{dm}(c_k, d_m) is the factor of the k-th class (c_k) in the m-th image; \( z_d(d_m) = \frac{1}{p(d_m)} \) and \( z_i(x_i) = \frac{1}{p(x_i)} \) are normalizing factors; c_k, d_m are latent variables, observed during inference. The term \( \prod_{i=1}^{N} p(x_i | c_k) \) (factors \( \prod_{i=1}^{N} z_i(x_i) f_i(x_i, c_k) \)) is specific for a separate image in the database while the term \( \prod_{k=1}^{K} p(c_k | d_m) \) (factor \( \prod_{k=1}^{K} f_{dm}(c_k, d_m) \)) is expected to be common for all the images. Configuration is performed using EM. Factors \( \prod_{i=1}^{N} z_i(x_i) f_i(x_i, c_k) \) are configured for an annotated image in the database separately, factors \( \prod_{k=1}^{K} f_{dm}(c_k, d_m) \) are configured using all the images. Inference is performed using Mean Field method [10], which could be interpreted as a way of minimizing Kullback-Leibler divergence between the measured empirical distribution and the distribution given an unseen image (factors \( z_i(x_i)f_i(x_i, c_k) \) are updated; the \( f_{dm}(c_k, d_m) \) are fixed).

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

Experimental analysis of the model and preliminary results were obtained for WorldView-2 multispectral images. The images in the database (see Table 2) were manually annotated and configurations of the factor graph are learned for the following landcover classes: 1-Water, 2-Forest/Trees, 3-Grass, 4-Bare soil, 5-Road (asphalt), 6-Man-made structures/Building. The classes are expected to be characterized by spectral properties and the images were atmospherically corrected (ATCOR vers. 3, msrura mode) and represented on the alphabet (k-means clustering, 200 clusters). The annotated images 1-4 are employed for the configuration of the graph, the image acquired for Rio-de-Janeiro area is not employed for configuration/training and used for test. Figure 5 illustrates the result of the classification of a new image. The landcover classes are properly labeled (OVA: Water-84%, Forest-85%, Low vegetation-99%, Bare soil-38%, Road-66%, Man-made structure-30%) with several expected misclassifications due to the use of only multispectral data: bitumen roofs of buildings are labeled as asphalt road, concrete pavement is confused with man-made object, shadows and runway (partly) are labeled as unclassified (a decision is difficult to make since low probabilities are reached for all the classes).

Table 2. Acquisition parameters for the images in database

<table>
<thead>
<tr>
<th>No.</th>
<th>Location</th>
<th>Acq. date and time</th>
<th>Off-nadir angle</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>London, UK</td>
<td>22.10.2011, 11:34:15</td>
<td>6.3°</td>
</tr>
<tr>
<td>2</td>
<td>Munich, Germany</td>
<td>12.07.2010, 10:30:17</td>
<td>5.2°</td>
</tr>
<tr>
<td>3</td>
<td>Rome, Italy</td>
<td>10.12.2009, 10:30:20</td>
<td>26.1°</td>
</tr>
<tr>
<td>Test image</td>
<td>Rio de Janeiro, Brazil</td>
<td>19.01.2010, 13:10:46</td>
<td>7.5°</td>
</tr>
</tbody>
</table>

4. CONCLUSION

Explicit factorization properties of factor graphs allow easy creation of different models for interpretation of remote sens-
Fig. 5. Example of landcover class extraction from not annotated data (Rio de Janeiro): (a) visible range multispectral image (bands 5,3,2), (b) FG landcover class extraction

Definition of data abstraction levels (e.g. signal/feature, class, decision, etc.) in a model is easy to make using factor graphs where the interpretation of information is performed in a required way. This methodology allows successful solution of several important topics: multisensory data fusion/classification, landcover extraction from unseen data. The probabilistic nature of factor graphs allows easier interpretation of the learned graph configuration. Augmenting properties exhibit an easy way to extend graph structure and to include additional information on the processed data, features, decisions, etc. Employment of an approximate inference (Mean field) allows to obtain plausible decisions with a low calculation time. Future work will be on the employment of semantic input in landcover extraction model.

5. REFERENCES


