Building detection in urban areas from combined optical and InSAR data exploiting context

Jan Dirk WEGNER[a,1], Jens R. ZIEHN[a] and Uwe SOERGEL [a]

[a] *IPI Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover, Hannover, Germany - {wegner, soergel}@ipi.uni-hannover.de*

Abstract. Up-to-date Synthetic Aperture Radar (SAR) sensors are able of acquiring imagery of sub-meter resolution. This very high resolution makes them an appropriate tool for the mapping of buildings in urban areas. Optical data may help to fill in gaps that are due to occlusions or signal mixtures caused by layover. We combine features of airborne interferometric SAR (InSAR) data and optical aerial images for classifying an urban scene into building and non-building sites. A Generalized Linear Model (GLM) within a Conditional Random Field (CRF) framework is, first, trained on features and, second, applied to a test site for inference. We then provide some concepts how the two different sensor geometries may be exploited for building height estimation once buildings have been detected.

Keywords SAR, Fusion, Classification, Building detection, Urban, Height estimation

Introduction

Single objects or even parts of them can be distinguished in data of very high-resolution SAR sensors. This can be seen in Fig. 1 where we compare a high resolution spotlight image of the TerraSAR-X satellite with an aerial optical image. The main building of the Leibniz Universität Hannover and various parts of it are captured by the SAR sensor. Layover and shadowing effects that are due to the slant range measuring principle occur. The optical data provides complementary information (e.g., color and texture) to the SAR data and may thus facilitate automatic object extraction [7-9]. We also see that in both images the main building is surrounded by various other objects like streets, trees, and other buildings. Those objects are the context of the object of interest. We may exploit this typical context in urban scenes in order to derive a more powerful and expressive building detection approach. But instead of introducing a model-based approach, which would potentially work well for a particular scene but fail for others, our aim is to learn context generically within a probabilistic framework. A method that meets the aforementioned requirements is Conditional Random Fields (CRF). CRFs were originally introduced by Lafferty et al. [10] for labeling one-

[1] Corresponding Author.

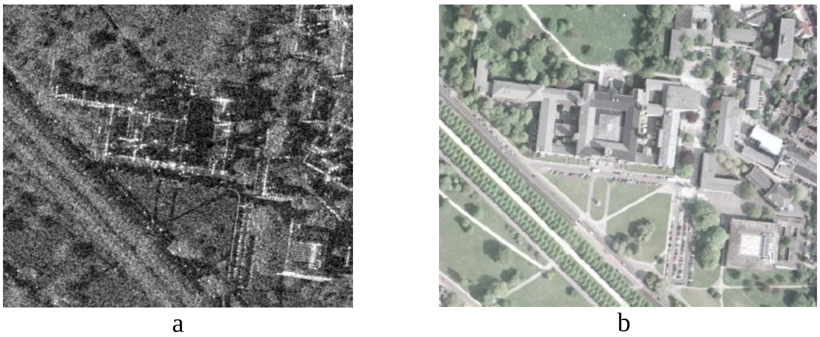

Figure 1. Comparison of SAR and optical image: the main building of the Leibniz Universität Hannover imaged by (a) the TerraSAR-X satellite (© DLR) (high resolution spotlight mode, resolution approx. 1m , range direction left to right) and by (b) an optical aerial sensor (© Geoinformation Stadt Hannover)

dimensional data and then adapted to imagery by Kumar and Hebert [11]. They are a probabilistic discriminative approach providing high modeling flexibility in terms of context integration. CRFs have already been applied to various computer vision tasks [11,16-17] and also for object detection in remote sensing data [12,15,19].

Once buildings have successfully been detected we may think of reconstructing them three-dimensionally [1-6,18]. Inherent effects like layover and shadowing contain height information. Additionally, the different sensor geometries of the SAR and the optical sensor may be used in order to combine features of both data sets to estimate building heights under the assumption of locally flat terrain.

The paper is organized as follows: First, Conditional Random Fields are introduced before we explain some inherent optical and SAR effects that are helpful for building height determination. We then apply our methods to some test data: features are extracted in SAR and optical data, buildings are detected with CRFs and building heights are estimated. Results are discussed and finally conclusions are drawn and ideas for future improvements are presented.

1. Conditional Random Fields

An overview of the general CRF framework we use is given in this section. Then, we introduce the basic formulae of our modified approach that better adapts to the task of building detection in urban areas.

CRFs are graphical models. They model relations between image sites through a network of nodes and edges. Since we are dealing with images, nodes may for example represent spatial entities like pixels, square image patches or irregular image segments. Edges link the nodes and carry information about how they should interact. This property enables us to introduce some prior knowledge we have by choosing a particular design of the edges. CRFs belong to the family of undirected graphical models (i.e., Random Fields) as opposed to directed graphical models like Bayesian networks. In contrast to Markov Random Fields (MRF), which are generative models because they model the joint probability $P(\boldsymbol{x},\boldsymbol{y})$, CRFs are discriminative models. They directly model the posterior probabilities $P(\boldsymbol{y}|\boldsymbol{x})$ of labels $\boldsymbol{y}$ given observations $\boldsymbol{x}$ through products of local marginal and conditional probabilities of adjacent nodes. Instead of providing only crisp decisions whether a pixel belongs to class building or non-building, we obtain probabilities for each node. This becomes convenient if we want to be flexible in terms of post-processing.

In the standard formulation, CRFs consist of two main terms (Eq. 1): the association potential $A_i(\boldsymbol{x},y_i)$ and the interaction potential $I_{ij}(\boldsymbol{x},y_i,y_j)$. The association potential evaluates how likely it is that a node i is labeled with label y_i given all data $\boldsymbol{x}$. We may use any discriminative classifier for the association potential. In the interaction potential we model context-knowledge. It describes how two label sites i and j interact considering all observations $\boldsymbol{x}$. In order to transform the node potentials to probabilities we have to divide the exponential of the sum of association potential and interaction potential through the partition function $Z(\boldsymbol{x})$. $Z(\boldsymbol{x})$ acts as a normalization factor and is a constant for a given data set.

$$P(\boldsymbol{y} \mid \boldsymbol{x}) = \frac{1}{Z(\boldsymbol{x})} \exp\left(\sum_{i \in S} A_i(\boldsymbol{x}, y_i) + \sum_{i \in S} \sum_{j \in N_i} I_{ij}(\boldsymbol{x}, y_i, y_j) \right) \qquad (1)$$

We use a Generalized Linear Model (GLM) for both the association potential $A_i(\boldsymbol{x},y_i)$ (Eq. 2) and the interaction potential $I_{ij}(\boldsymbol{x},y_i,y_j)$(Eq. 3). Vector $\boldsymbol{h}_i(\boldsymbol{x})$ contains all node features and vector $\boldsymbol{w}^T$ contains the weights of the features in $\boldsymbol{h}_i(\boldsymbol{x})$ that are tuned during the training process. Vector $\boldsymbol{v}^T$ contains the weights of the features, which are adjusted during the training process. y_i is the label of the site of interest and y_j the label it is compared to.

$$A_i(\boldsymbol{x}, y_i) = \exp\left(y_i \boldsymbol{w}^T \boldsymbol{h}_i(\boldsymbol{x}) \right) \qquad (2)$$

$$I_{ij}(\boldsymbol{x}, y_i, y_j) = \exp\left(y_i y_j \boldsymbol{v}^T \boldsymbol{\mu}_{ij}(\boldsymbol{x}) \right) \text{ where } \mu_{ij}(\boldsymbol{x}) = \left| h_i(\boldsymbol{x}) - h_j(\boldsymbol{x}) \right| \qquad (3)$$

Various designs of $I_{ij}(\boldsymbol{x},y_i,y_j)$ exist [11,17] that are designed for typical computer vision tasks. The usual consists of detecting a single rather large instance of an object in a relatively small image. We face a different challenge: Our images are large and many small instances of the same object occur with narrow gaps in-between. In order to avoid over-smoothing effects we introduce an explicit discontinuity constraint. High gradients are often found at building boundaries in the optical intensity image. They well isolate buildings from their environment. The CRF standard interaction potential (Eq. 3) does not fully incorporate this discriminating feature because it only compares features of adjacent nodes but nothing in between them. Therefore, we extend the edge feature vector $\boldsymbol{\mu}_{ij}(\boldsymbol{x})$ to $\boldsymbol{\mu}_{ij,mod}(\boldsymbol{x})$ by element-wise multiplication with a scalar weight $w_{disc,ij}$ [19]. This weight is a function of the mean gradient between two adjacent nodes i and j.

$$\mu_{ij,\mathrm{mod}}(\boldsymbol{x}) = \left| h_i(\boldsymbol{x}) - h_j(\boldsymbol{x}) \right| w_{disc,ij} \qquad (4)$$

$$w_{disc,ij} = \left(1 + \left(\frac{1}{1 + \exp\left(-\alpha \left(g_{ij} - \kappa \right) \right)} \right) \right) / 2 \qquad (5)$$

We introduce the mean gradient into a sigmoid function (Eq. 5) with the inflexion position at κ = 0.5 because an investigation of its histogram suggests that values above 0.5 separate objects of different classes. In order to still allow for the separation of two nodes in the absence of high gradients we shift the sigmoid function in y-direction.

In order to learn the parameters of our CRF, which are simply the feature weights within the vectors ***v*** and ***w***, we have to set up an objective function *O(**w**,**v**)*. In order to ensure a global optimum our objective function should either be concave (global maximum) or convex (global minimum). A common way to achieve a concave objective function is to use the log-likelihood of the posterior probabilities *P(**y**|**x**)* (Eq. 6).

$$O(\boldsymbol{w},\boldsymbol{v}) = \log P(\boldsymbol{y} \mid \boldsymbol{x}) \tag{6}$$

By substituting the CRF posterior probabilities of Eq. 1 into Eq. 6 we obtain the objective function (Eq. 7).

$$O(\boldsymbol{w},\boldsymbol{v}) = \left(\sum_{i \in S} A_i(\boldsymbol{x}, y_i) + \sum_{i \in S} \sum_{j \in N_i} I_{ij}(\boldsymbol{x}, y_i, y_j) \right) - \log Z(\boldsymbol{x}) \tag{7}$$

We use the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) [20] method as optimizer to train the association potential and the interaction potential simultaneously. For inference we use Loopy Belief Propagation (LBP) [21].

2. Building height estimation

Once we have detected buildings we may wish to get some knowledge about their heights. The previous two-dimensional detection provides horizontal building boundaries and an additional height information would allow for a three-dimensional reconstruction. Our aim is to investigate effects inherent in the data that contain height information.

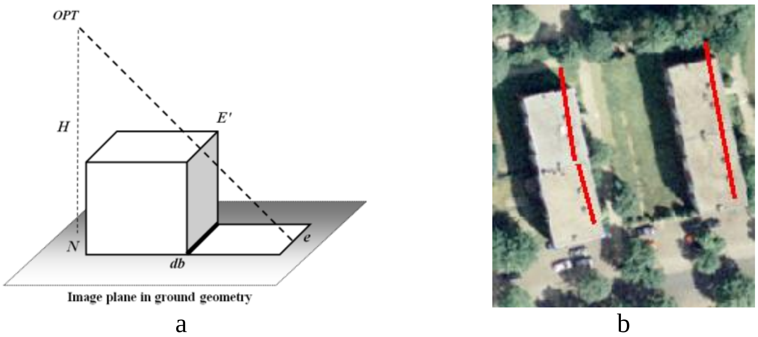

Figure 1. (a) Sketch of height estimation using SAR double-bounce line *db* (black) and overlapping building roof edge (*E'* is projected to *e*) of the optical image, (b) double-bounce lines (red) overlaid to a cut-out of the optical orthophoto.

An obvious height information source that first comes into one's mind is the InSAR heights. Currently, a rather simple InSAR height extraction procedure is applied for testing purposes. Initially, the previously extracted double-bounce lines of one of the two SAR amplitude images are extended to parallelograms towards the SAR sensor position in slant range geometry. We then turn to the interferometric heights in slant range geometry and overlay the parallelograms. Thus, the InSAR layover phase ramp is fully contained within the parallelogram of a particular building. We set the width of the parallelogram according to the SAR acquisition parameters and the expected maximum building height of the particular urban scene. Next, the maximum height within each parallelogram (i.e., the maximum height of each phase ramp) is assigned to the corresponding building. It should be noticed that we have not done any phase-unwrapping yet which is hard to conduct in urban environments with strong signal mixtures and abrupt height changes. In order to help circumvent this issue we make use of the optical data. Buildings in the optical image appear distorted due to the central perspective of the aerial camera. In the image this translates to a building's facade being visible and its roof being mapped with an offset of the building's footprint. A distortion of a particular building (given the acquisition parameters) is a function of its height *h* and its distance to the nadir point *N* of the camera (within the image). If we know those parameters, we simply have to measure the offset between a roof edge and the building footprint. An increasing offset indicates a higher building. Reconsidering our feature extracted from the SAR data, we see that double-bounce lines are located exactly at the boundary of the building footprint facing the SAR sensor. All double-bounce lines are located at ground height. If we overlay the extracted double-bounce lines (see details of extraction procedure in [13]) with the optical image, the building roof edge (*E'*) in the optical image falls over the double-bounce line *db* (*E'* is mapped to e). Then, we are able to calculate building height *h* from the flying altitude of the aerial camera *H* and the ratio of distances *Δdb* and *Δe*. *Δdb* describes the distance of the double-bounce line to the nadir *N*. *Δe* is the distance between the roof edge of a building to the nadir *N*.

$$h = H \cdot \left(1 - \frac{\Delta db}{\Delta e}\right) = H \cdot \left(1 - \sqrt{\frac{(db_x - N_x)^2 + (db_y - N_y)^2}{(e_x - N_x)^2 + (e_y - N_y)^2}}\right) \tag{8}$$

3. Results and discussion

In order to assess the current status of our research we do some experiments with test data. The proposed methods are tested using an aerial orthophoto of 0.31 meters ground sampling distance and airborne InSAR data acquired by the Intermap Technologies Aes-1 sensor of approximately the same resolution. The test scene shows a part of the city of Dorsten in Germany. First, we perform building detection based on combined optical and InSAR features. Second, we manually extract the footprints of some large flat-roofed buildings in the scene and apply our height determination methods.

3.1. Building detection

We have introduced the CRF classification framework in the first section. Vector ***h**$_i$**(x)*** in Eq. 2 contains the features that discriminate buildings from the rest of the data. We remind the reader that the context term (the interaction potential) is also based on those features as shown in Eq. 4. For our testing purpose we take rather simple features as input to ***h**$_i$**(x)***. Mean and variance of the red channel, the blue channel, the hue, features based on the gradient orientation histogram of the intensity image, and some

Haralick features are used. The most reliable InSAR feature is the already previously mentioned building double-bounce line [5,13], which is located where the near-range building walls meet the ground (see *db* in Fig. 1a). Using those simple features and the modified interaction potential (Eq. 4,5) we achieve a true positive rate (TPR) of 85% and a false positive rate (FPR) of 29% (see results of three test images in Fig. 2). A more detailed analysis of the results (with a slightly modified interaction potential) can be found in [15].

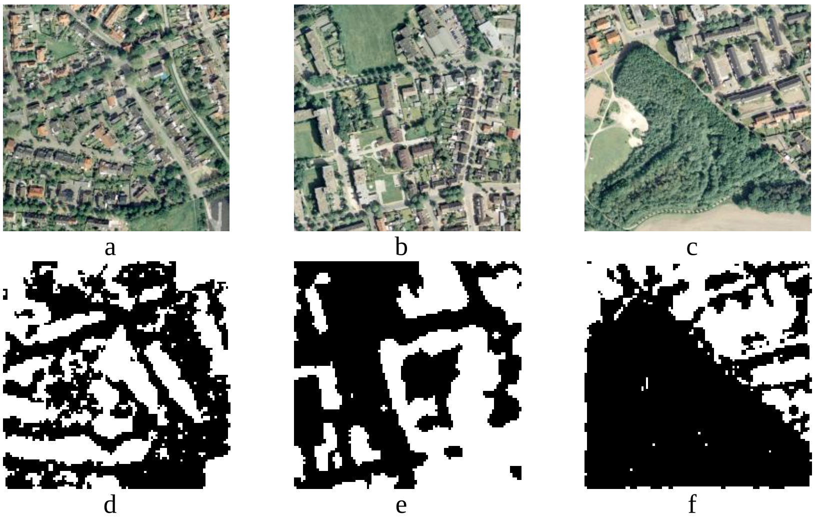

Figure 2. (a) Sketch of height estimation using SAR double-bounce line *db* (black) and overlapping building roof edge (*E'* is projected to *e*) of the optical image, (b) double-bounce lines (red) overlaid to a cut-out of the optical orthophoto.

3.2. Building height estimation

A small test region containing relatively high flat-roofed buildings (see test region overview in Fig. 3c) and estimate their heights in the two different ways described in section 3 (see results in Fig. 3a,b). The flying height *H* above ground was 3900 m. We compare the estimated building heights (Fig. 3a,b) to the LIDAR reference. Over-estimated building heights are shown in dark grey and under-estimated heights in light grey for each building separately. In Fig. 3a it can be observed that, in general, the InSAR height slightly under-estimates the building heights (mean error -2.8 meters). This is due to ambiguous phase to height conversions because we have not done any phase-unwrapping. We can see this at the high buildings on the right side of Fig. 3a where the actual building height has been under-estimated more than 50%. Thus, a next step will be to integrate the height estimated with the supplementary optical data in order to support phase-unwrapping. The combined use of the double-bounce line and the overlapping buildings in the optical image gives better results. It slightly over-estimates the building heights (mean error 1.6 meters, Fig. 3b).

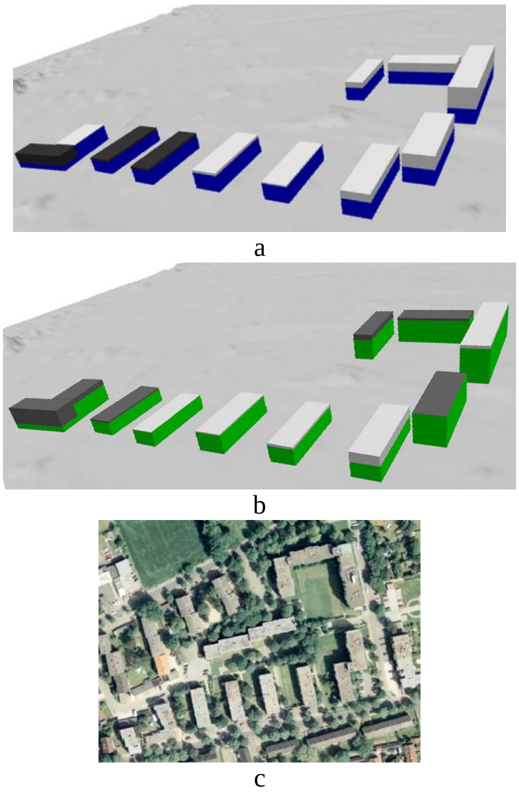

Figure 3. Estimated building heights assigned to building footprints: (a) differences of InSAR heights to LIDAR reference and (b) differences of heights calculated from overlapping optical data and double-bounce lines to LIDAR heights (light grey: height estimated too low, dark grey: height estimated too high), (c) aerial orthophoto of the test region

4. Conclusions and outlook

We have shown that a context-based approach using GLMs within a CRF framework provide good results for building detection in urban areas. Heights of the detected buildings can then be determined based on a combination of SAR double-bounce lines and distortions in the optical image that are caused by the central perspective of the aerial camera. One factor limiting the performance of the CRFs is the regular grid of image patches we use for classification. Those patches do not consider image information and thus building and non-building areas are often merged within one patch. This is the reason for less discriminative feature distributions. Therefore, we are currently testing CRFs set up on an initial segmentation that well preserves object boundaries. A graphical model on image segments would also allow us a more expressive integration of the image gradients as discontinuity constraint. In addition, we will have to test our classification approach on a second test site with data of different sensors in order to get a more generally valid performance evaluation.

Considering building height estimation we will integrate all possibilities into one joint least squares adjustment. We also need to test the presented ideas on a second data set.

References

[1] P. Gamba, B. Houshmand, and M. Saccani, “Detection and extraction of buildings from interferometric SAR data,” IEEE Trans. Geoscience and Remote Sensing, vol. 38, no. 1, part 2, 2000, pp. 611–617.

[2] U. Soergel, U. Thoennessen, and U. Stilla, “Reconstruction of Buildings from Interferometric SAR Data of built-up Areas”, Proc. of PIA, International Archives of Photogrammetry and Remote Sensing, vol. 34, part 3/W8, 2003, pp. 59-64.

[3] C. Tison, F. Tupin, and H. Maitre, “A Fusion Scheme for Joint Retrieval of Urban Height Map and Classification From High-Resolution Interferometric SAR Images”, IEEE Trans. Geoscience and Remote Sensing, vol. 45, no. 2, 2007, pp. 496-505.

[4] F. Xu and Y.-Q. Jin, “Automatic Reconstruction of Building Objects From Multiaspect Meter-Resolution SAR Images”, IEEE Trans. Geoscience and Remote Sensing, vol. 45, no.7, 2007, pp. 2336-2353.

[5] A. Thiele, J.D. Wegner and U. Soergel, “Building reconstruction from multi-aspect InSAR data”, In U. Soergel (Ed), *Radar Remote Sensing of Urban Areas*, Springer, 1st Edition, ISBN-13: 978-9048137500.

[6] R. Bolter, “Buildings from SAR: detection and reconstruction of buildings from multiple view high resolution interferometric SAR data”, PhD thesis, University of Graz, Austria, 2001.

[7] U. Soergel, E. Cadario, A. Thiele, and U. Thoennessen, “Feature Extraction and Visualization of Bridges over Water from high-resolution InSAR Data and one Orthophoto”, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 1, no.2, 2008, pp. 147-153.

[8] F. Tupin and M. Roux, “Detection of building outlines based on the fusion of SAR and optical features”, ISPRS Journal of Photogrammetry and Remote Sensing, vol. 58, 2003, pp. 71-82.

[9] F. Tupin and M. Roux, “Markov Random Field on Region Adjacency Graph for the Fusion of SAR and Optical Data in Radargrammetric Applications”, IEEE Trans. Geoscience and Remote Sensing, vol. 43, no. 8, 2005, pp. 1920-1928.

[10] J. Lafferty, A. McCallum, and F. Pereira, “Conditional Random Fields: Probabilistic Models for segmenting and labeling sequence data”, in *Proc. Int. Conf. on Machine Learning,* 2001.

[11] S. Kumar and M. Hebert, “Discriminative Random Fields: A Discriminative Framework for Contextual Interaction in Classification”, in *Proc. IEEE Int. Conf. on Computer Vision*, 2003, vol. 2, pp. 1150-1157.

[12] P. Zhong and R. Wang, “A Multiple Conditional Random Fields Ensemble Model for Urban Area Detection in Remote Sensing Optical Images”, IEEE Trans. Geoscience and Remote Sensing, vol. 45, no. 12, 2007, pp. 3978–3988.

[13] J.D. Wegner, A. Thiele, and U. Soergel, “Fusion of optical and InSAR features for building recognition in urban areas”, IntArchPhRS, vol. 38, part 3/W4, 2009, pp. 169-174.

[14] J.D. Wegner, S. Auer and U. Soergel, “Extraction and geometrical accuracy of double-bounce lines in high-resolution SAR images”, Photogrammetric Engineering & Remote Sensing, 2010, in press.

[15] J.D. Wegner, R. Hänsch, A. Thiele and U. Soergel, “Building detection from one Orthophoto and high-resolution InSAR Data using Conditional Random Fields”, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2010, in press.

[16] S.B. Wang, A. Quattoni, L.-P. Morency, D. Demirdjian, and T. Darrell, “Hidden Conditional Random Fields for Gesture Recognition”, in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2006, 7 p..

[17] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie, “Objects in Context”, in *Proc. IEEE Int. Conf. on Computer Vision*, 2007, 8 p..

[18] D. Brunner, G. Lemoine, and L. Bruzzone, “Extraction of building heights from VHR SAR imagery using an iterative simulation and match procedure”, in *Proc. IEEE Int. Geoscience and Remote Sensing Symposium*, 2008, 4 p..

[19] J.D. Wegner, J.R. Ziehn, and U. Soergel, “Building detection and height estimation from high-resolution InSAR and optical data”, in *Proc. IEEE Int. Geoscience and Remote Sensing Symposium*, 2010, 4 p..

[20] D.C. Liu and J. Nocedal, “On the limited memory BFGS for large scale optimization”, Mathematical Programming, 1989, vol. 45, no. 1-3, pp. 503-528.

[21] B.J. Frey and D.J.C. MacKay, “A revolution: Belief propagation in graphs with cycles”, in M.I. Jordan, M.J. Kearns, S.A. Solla (Eds), *Advances in Neural Information Processing Systems*, 1998, vol. 10, MIT Press.