Discriminative Learning of Conditional Random Fields Applied to the Classification of Urban Settlement Types

TESSIO NOVACK[1] & UWE STILLA[1]

Abstract: Dieser Artikel zeigt die ersten Experiments zur Klassifikation von Urban Structure Types (USTs) unter Verwendung von hochaufgelösten, satellitebasierten InSAR Daten und Conditional Random Fields (CRFs). Es wurde eine Prozedur entwicklet um die Parameter des CRFs diskriminativ zu lernen und zu testen. Die Klassifikation dieses Modells sowie zwei weiterer Verfahren werden mittels einer handsegmentierten Karte verglichen. Obwohl das Lernenverfahren konvergiert hat, sind die Genauigkeiten der Klassificationergebnisse nicht zufriedenstellend. Methoden zur Verbesserung der Genauigkeit werden aufgezeigt und sollen in Zukunft bewertet werden.

1 Introduction and Goals

Information about the spatial distribution of different types of settlements found in cities is important for several urban planning actions as well as for modelling the spatial behaviour of different urban phenomena (e.g. energy balance and urban climate, population estimation, traffic behaviour etc.). In Germany, these types of urban settlements are often categorized under the concept of Urban Structure Types (USTs) (PAULEIT AND DUHME, 2000; HEIDEN ET AL., 2012). USTs are categorized considering aspects like the geometry, density and spatial configuration of buildings, their social usages (e.g. residential, industrial etc.), as well as their environmental properties (e.g. presence and type of vegetation and water bodies). According to HEIDEN ET AL. (2012), mostly cost-and-time-intensive approaches, such as field surveys or manual interpretation of aerial photographs, have been used to create or update USTs inventories. Until now, remote sensing research explicitly focused on the automatic classification of USTs have exclusively used multispectral (WURM ET AL., 2009; BANZHAF AND HÖFER, 2008) or hyperspectral (HELDENS, 2010; HEIDEN ET AL., 2012) data. Motivated by that, we aim to assess the feasibility of mapping USTs based on the classification of high-resolution spaceborne InSAR data. Assuming that contextual information is important for predicting the most probable type of urban settlement, we propose the use of undirected Probabilistic Graphical Models (PGM), more specifically of Conditional Random Fields (CRFs), to attain this goal.

CRFs enable the contextual relations between the classes of neighboring image segments and their attributes to be statistically modeled. Besides eliminating the necessity of manually defining hundreds of parameters and the subjectivity that it involves, automatic learning of the parameters of PGMs give rise to more accurate models (KOLLER AND FRIEDMAN, 2009). From the methodological point of view, the goals of this work are hence twofold: (1) to evaluate the performance of automatically learned CRFs regarding the classification of USTs using InSAR data and (2) to compare the performance of the CRFs model with other standard classification approaches, namely Nearest-Neighbors and Maximum-Likelihood.

1) Photogrammetrie & Fernerkundung, Technische Universität München, Arcisstrasse 21, 80333 München; E-Mail: tessio.novack@tum.de

2 USTs Classes and Utilized Data

Given that this USTs official map considers fifty classes, most of which are correlated to each other and cannot be distinguished based solely on remote sensing data, we grouped similar classes into more general ones. We ended up in this way with eight final classes. These are: Green Areas (*Grünflächen*), Railroad (*Gleisanlagen*), Detached and Semi-detached Houses (*Hausbebauung*), Perimeter Block Development (*Blockrandbebauung*), Block Development (*Blockbebauung*), High-rise Buildings (*Geschoßbaukomplexe*).
To test the performance of the learned CRFs model on the classification of these USTs, we utilized an interferometric pair of images from the city of Munich (Germany) obtained by the TerraSAR-X satellite operating at Spotlight mode. The images were acquired on May 2011 with a ground sampling distance of approximately 1.1 meters in azimuth and 5.8 meters in slant range. A Digital Elevation Model created through interferometry and also the coherence image were used afterwards for the calculation of image attributes.
In order to test our classifications, we used the official USTs map (*Flächennutzung – Strukturkartierung*) from the city of Munich for the year of 2011 (MÜNCHEN, 2014).

3 Methods

In this section we describe the methodological steps performed in this work. These comprise the partition of the scene into image objects, the selection of samples from each USTs class, the structuring and learning of the CRFs model, as well as its application on the interferometric TerraSAR-X data from Munich.

3.1 Image Segmentation and Sample Selection

Since individual pixels are not expressive analysis units for the classification of USTs, we decided to approach this problem by a region-based approach. In order to create image-regions from our InSAR scene, we applied a grid-based segmentation approach in which the scene was divided into grid-cells (**Figure 1**, a). Following, we collected 38 grid-cells as samples from each of the considered USTs classes (**Figure 1**, b). Considering that each of these samples have exactly four neighbours, the total number of samples collected was of 152. These samples were exported with the image attributes *Mean Coherence, Mean Intensity, Standard Deviation of DEM pixels* and *Relative Area of High Intensity Objects*. For the calculation of this last feature, we applied a threshold over the intensity image in order to select the pixels with intensity higher than this threshold. The relative area of these groups of pixels was then calculated for each of the grid-cells. These features were intuitively chosen assuming that: (1) *Mean Coherence* is a good feature for distinguishing Green Areas and Detached and Semi-detached Houses from other USTs where vegetation is less present; (2) Perimeter Block Development and Block Development are USTs with higher *Mean Intensity* and *Relative Area of High Intensity Objects* and yet they have among them different values for these features (i.e. Block Development has higher values than Perimeter Block Development) and (3) that *Standard Deviation of DEM pixels* is a good feature for detecting High-rise Buildings. The neighbouring segments of each

sample containing these image attributes and their reference USTs classes was also exported and associated to its corresponding samples.

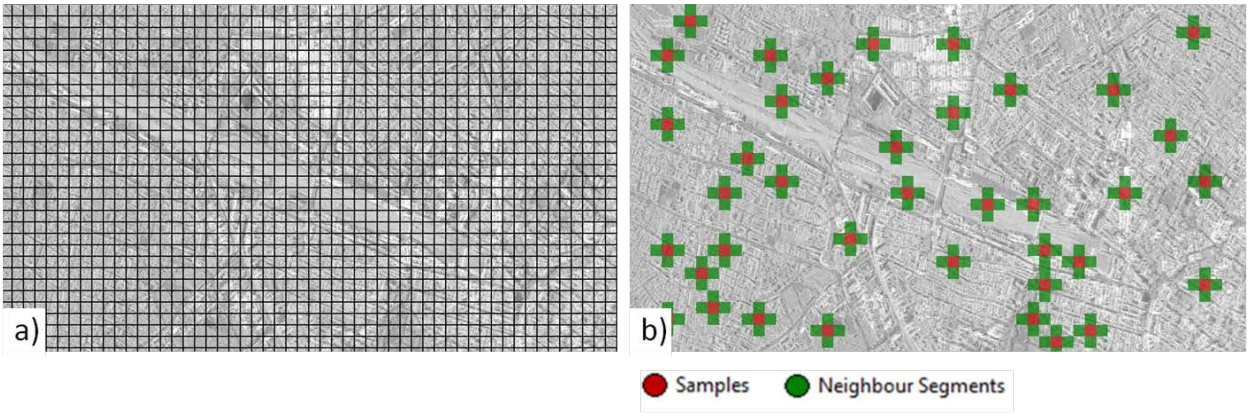

Figure 1 – Grid-based segmentation (a) and the samples used for the training of the CRFs model.

3.2 Reference Map Creation

In order to evaluate the USTs classifications, we manually created a reference map. Firstly, the study area (**Figure 2**, a) was segmented using the algorithm from BAATZ AND SHÄPE (2000) and the segments were manually classified based on the interpretation of optical data and on the official USTs map from Munich produced in 2011 (**Figure 2**, b). Then, the study area was grid-segmented and the grid-cells were automatically classified based on the criterion of the reference map's class inside the grid with larger relative area (**Figure 2**, c). Lastly, the grid-cells sampled for the learning of the CRFs model were removed from the grid-based reference map so that the final reference didn't contain any bias (**Figure 2**, d).

3.3 CRFs Model's Structure Definition and Other Pre-Learning Steps

Before learning the parameters of the CRFs model, we defined its structure, i.e. the factors involving the observed variables (the image attributes) and the unobserved variables (the unknown classes of the image segments). The structure of our CRFs model is comprised of third-order factors (i.e. factors containing three variables), where, according to the definition of CRFs, at least one of these variables must be an unobserved one. **Figure 3** shows the structure of the model, as well as the image attributes they involve. This structure was defined manually.
After choosing the observed variables and defining the structure of the model, we discretized all the observed attributes into four bins with the intention of reducing the number of parameters to be learned and therefore the number of necessary samples for learning them.

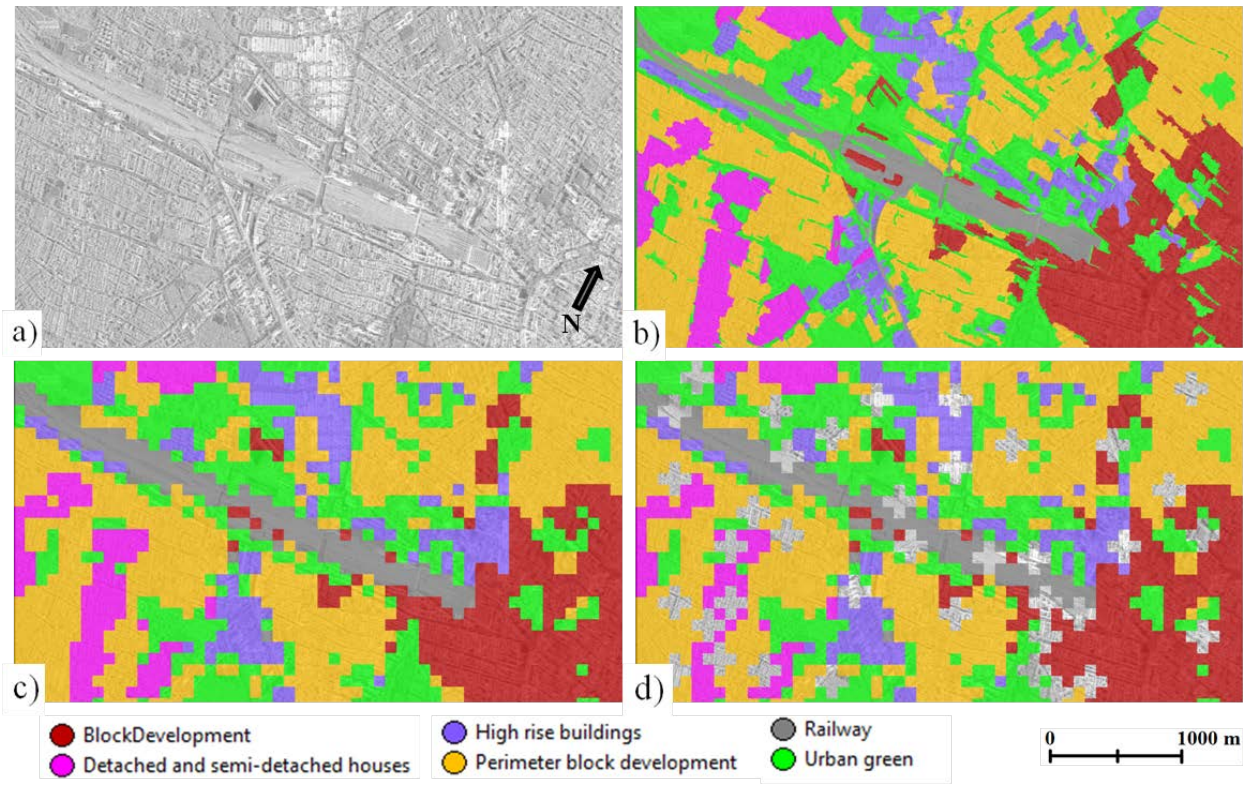

Figure 2 – The study area (a), the manually created reference map (b), the grid-based reference map (c) and the final reference map (d).

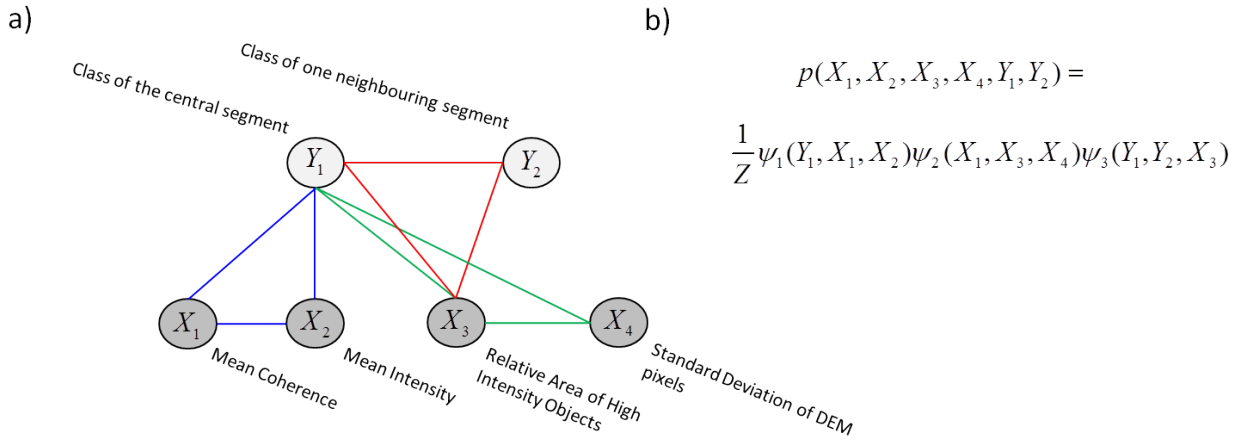

Figure 3 – Structure of the CRFs model (a) and its corresponding factorization (b). The different factors are also shown on (a) with different colours.

3.4 Learning of the CRFs Model Parameters

The joint distribution of the variables of an undirected PGM is modelled by the structure (i.e. the factorization) and the parameters of the model. The parameters *k* of an undirected PGM are positive real numbers associated to each possible combination of assignments of the variables *x*

from each of the factors of the model. Here, we note this association between an assignment combination and its parameter as $f_k(x)$. A simple approach for learning the parameters of undirected PGMs is the maximization of the Maximum Log-Likelihood function. The Maximum Log-Likelihood objective function has the form:

$$l(M,\theta:D) = \sum_{f_k \in F} \theta_k (\sum_m f_k(\xi[m])) - M \log Z(\theta) \quad (1)$$

where M is the number of samples, m is a single sample from the training data set and Z is the partition function, which makes it a valid probabilistic distribution. This function is a convex one, which means it has no local optima. However, it also has no closed-form solution. Hence, any of its global optima has to be found through an iterative optimization process (like Gradient Descent, for example). The gradients of this objective function are the differences between the empirical probabilities of the assignments (calculated directly from the sample data set) and their estimated probabilities given a parameter set θ. Formally, we want that for each assignment $f_k(x)$ these differences (i.e. the gradients) equal to zero:

$$(\sum_m f_k(\xi[m])) - ME_{x \sim P_\theta}[f_k(x)] = 0 \quad (2)$$

The first term of this expression is the empirical probability of an assignment and the second term is its marginal probability. In order to estimate the second term, inference has to be performed over the model considering the current parameter set θ. Initially it is usually a random parameter set. Many approximate and exact inference algorithms are proposed on the literature (FREY AND JOJIC, 2005, JÖRG ET AL., 2013), whereas on real world problems exact inference is hardly tractable.

Being a specific case of undirected PGMs, the parameters of CRFs are more properly learned in the discriminative way. In this setting, our training set consists of pairs $D = \{(y[m], x[m])\}_{m=1}^{M}$ specifying assignments to Y (unobserved variables) and X (observed variables). Here we want to optimize the likelihood of each $y[m]$ given the corresponding $x[m]$. This is a sum of convex log-likelihood functions, i.e. one for each data sample, with a region of global optima, i.e. redundant optimal parameterizations. The gradients in the discriminative training case have the form:

$$\sum_{m=1}^{M} (f_i(y[m], x[m]) - \mathrm{E}_\theta[f_i \mid x[m]]) \quad (3)$$

Whereas in the generative training regiment each gradient step required only a single execution of inference, training a model in the discriminative way is more cumbersome because we have to execute inference for every data sample conditioned on $x[m]$. On the other hand, the inference is executed on a considerably simpler model, since conditioning the model on the evidence X can only reduce the computational costs.

The inference and optimization algorithms we utilized where respectively the Sum-Product Loopy Belief Propagation (SP-LBP) (MURPHY ET AL. 1999; KOLLER AND FRIEDMAN, 2009) and the Gradient Descent (GD) algorithms. **Figure 4** shows the result of such a procedure for learning the parameters of a CRFs model later applied for the classification of USTs.

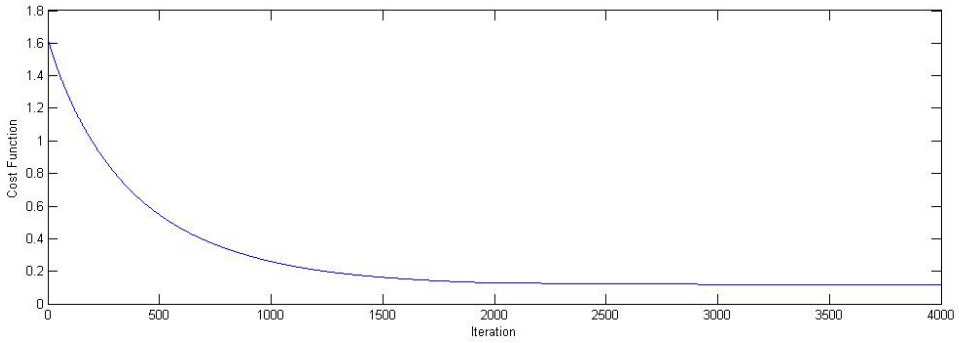

Figure 4 – Optimization result of the parameter learning procedure using SP-LBP and GD. The cost function is the sum of the absolute gradients of all data samples.

4 Results

Once the structure of the CRFs model was defined and all of its parameters were automatically learned, we could expand the model's structure to the whole scene and reduce all factors to the evidence, i.e. the values of the image attributes considered as the observed variables. Following, inference was performed in order to obtain the most probable classification (also known as the most probable assignment, or MAP) of the scene. The inference method used for that was the Max-Sum Loopy Belief Propagation (KOLLER AND FRIEDMAN, 2009). We also performed classifications with the Nearest-Neighbours and Maximum-Likelihood classifiers trained with the same sample set used for training the CRFs model. These classifications are shown on **Figure 5**. **Table 1** shows the overall accuracy and the Kappa index of accuracy (KIA) of these three classifications. It is noticeable from **Table 1** that the accuracy indexes from all classifications are very low. On the other hand, **Figure 5** shows that there is to some extent a certain visual correlation between the reference map and the classifications, especially regarding the one obtained with the CRFs model. Although arguably a qualitative analysis of the classifications concludes that the one obtained with the CRFs model is the best, this method achieved the worst KIA and an overall accuracy lower than the one obtained with the Nearest-Neighbour method.

	Overall Accuracy	Kappa Index of Accuracy
Nearest-Neighbours	0.397	0.215
Maximum Likelihood	0.374	0.191
Conditional Random Fields	0.376	0.167

Table 1 – Overall accuracy and Kappa index of the three performed classifications.

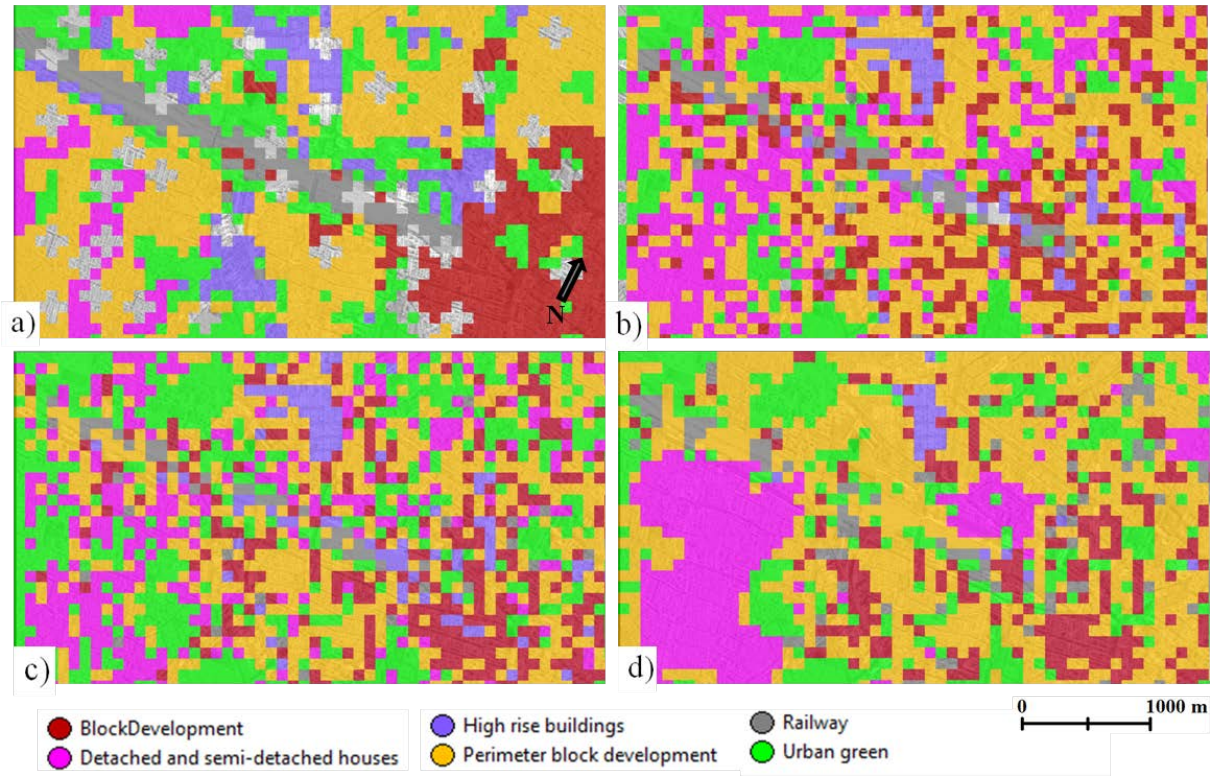

Figure 5 – Classifications performed with the Nearest-Neighbours (b) and Maximum-Likelihood (c) classifiers and the classification obtained with our learned CRFs model (d). The reference map used for the quantitative evaluation of these classifications is shown in (a).

The reason for these not promising results is most probably the low number of samples collected for training the classifiers, which contrast to the high geometrical and spectral complexity of the USTs classes. Also, the grid-based segmentation is not the best regiment for such a classification task, given that the grid-cells are not meaningful analysis units. On future work we plan on creating such meaningful analysis units through a segmentation procedure applied with adequate parameters. Also, an extensive exploratory analysis of image attributes has to be undertaken so that the ultimate CRFs model contains more expressive image attributes. We expect that these measures will increase the accuracy of future classification tests and express the advantages of using undirected PGMs for classifying USTs based on InSAR data.

5 Conclusions

This paper shows the first experiments on the classification of USTs using high-resolution space-borne InSAR data and CRFs. A procedure for learning in a discriminative way the parameters of a manually defined CRFs model was implemented and tested. Since classifications performed with two other standard classifiers also did not obtain good accuracy indexes, we assume that the negative results obtained with our learned CRFs model are not because of the approach itself or the CRFs model specifically. Actually our results show that, given our sample set and the considered image attributes, none of the approaches can be properly applied for the problem at

hand. We assume that the complexity of USTs demands special care in the choice of image attributes and in the generation of image objects through the segmentation process. Equally important though is the selection of a large and representative set of samples. It is possible that if more expressive image attributes related to meaningful image segments are considered and if the sample set is large and representative enough, CRFs model learned in the procedure shown here will deliver more accurate results regarding the classification of USTs. This also applies for the other two standard classifiers.

Finally, we suggest trying the option of defining the CRFs parameters based on statistical models. Also other classification algorithms like Random Forest or Support Vector Machine can be tried in other to, based on performance comparison, evaluate the adequacy of using learned CRFs for classifiying USTs on InSAR data.

6 References

BAATZ, M. AND A. SCHÄPE. 2000: Multiresolution Segmentation – An Optimization Approach for High Quality Multi-scale Image Segmentation. Journal of Photogrammetry and Remote Sensing, **58**, P.12–23.

FREY, B. J., JOJIC, N. 2005: A Comparison of Algorithms for Inference and Learning in Probabilistic Graphical Models. IEEE Transactions on Pattern Analysis and Machine Intelligence, **27**, 1-25.

HEIDEN, U., HELDENS, W., ROESSNER, S., SEGLL, K., ESCH, T. & MUELLER, A., 2012: Urban structure type characterization using hyperspectral remote sensing and height information. Landscape and Urban Planning, **105**, P.361-375.

HELDENS, W., 2010: Use of airborne hyperspectral data and height information to support urban micro climate characterisation. PhD Dissertation.

JÖRG, H. K., ANDRES, B., HAMPRECHT, F. A., SCHNÖRR, C., NOWOZIN, S., BATRA, D., KIM, S., KAUSLER, B. X., LELLMANN, J., KOMODAKIS, N., ROTHER, C. 2013: A Comparative Study of Modern Inference Techniques for Discrete Energy Minimization Problems. CVPR, 2013

KOLLER, D. & FRIEDMAN, N., 2009: Probabilistic Graphical Models – Principles and Techniques. The MIT Press: Cambridge: USA.

MÜNCHEN, 2014: Landeshauptstadt München – Referat für Gesundheit und Umwelt. http://maps.muenchen.de/rgu/strukturtypen

MURPHY, K., WEISS, Y. & JORDAN. M., 1999: Loopy belief propagation for approximate inference: an empirical study. In UAI, 1999.

PAULEIT, S. AND DUHME, F., 2012: Assessing the environmental performance of land cover types for urban planning. Landscape and Urban Planning, **52**, P.1-20.

WURM, M., TAUBENBÖCK, H., ROTH, A. AND DECH, S., 2009: Urban structuring using multisensoral remote sensing data – By the examples of german cities – Cologne and Dresden. In: Urban Remote Sensing Joint Event, Shanghai, China.