

DATA FUSION FOR BUILDING RECONSTRUCTION FROM MULTI-ASPECT INSAR DATA

O. Maksymiuk, U. Stilla

Photogrammetry & Remote Sensing, Technische Universitaet Muenchen (TUM), 80333 Muenchen,
oliver.maksymiuk@tum.de

KEY WORDS: InSAR, Multi-Aspect, MASAR, 3D Reconstruction, Building Reconstruction, Urban Areas, Probabilistic Graphical Model, Data Fusion, Radargrammetry, Compressive Sensing, Sparsity

ABSTRACT:

In this paper we present a novel approach for 3D reconstruction of point clouds based on single baseline Multi-Aspect InSAR (MASAR) data. The point clouds represent an intermediate result to achieve a comprehensive building reconstruction framework. The exact determination of scatterers based on SAR data is a non-trivial task since the optimal solution requires the knowledge of the number of scatterers within one range cell. In recent years many methods were proposed addressing this problem but most of them require multiple observations making them inapplicable to our task. We use a Probabilistic Graphical Model (PGM) to combine all aspects into a common framework exploiting all contradictions and redundancy in the data. The model is used iteratively together with local optimizations adjusting the hypothesis of the scatterer within one range cell to the corresponding observation. This makes it possible to find a global solution.

1 INTRODUCTION

In recent years much effort was done in the analysis of urban areas by remote sensing techniques. Buildings are an integral part of every human settlement and therefore building reconstruction is one of the most important parts of a comprehensive urban analysis system. A detailed knowledge of buildings substantially supports the planning of emergency response or urban management in general. Nowadays most of the data used for this task is acquired by laser scanners due to their high accuracy. In the recent developments these scanners are not directed in nadir direction but with a significant elevation. This setup allows to capture also the facades of the buildings allowing for a full 3D reconstruction. But as a drawback some areas residing behind the buildings are shadowed. To overcome this limitation it is necessary to use acquisitions from multiple, orthogonal aspects with a subsequent coregistration (Hebel and Stilla, 2007). In SAR imagery the elevation is always given due to the sidelooking geometry. Therefore it is strictly necessary to use MASAR data for a full description of the scene. It has been shown that decimeter resolution SAR imagery has great potentials to contribute to urban scene analysis (Stilla et al., 2005), (Brenner and Roessing, 2008). There are already some approaches working on MASAR data (Bolter, 2001), (Thiele et al., 2007a), (Schmitt and Stilla, 2011). In this work however we want to investigate a general framework using all context available to obtain a posterior estimation of the scatter locations. We use PGMs as a declarative representation to encode our MASAR domain.

2 PROBABILISTIC FRAMEWORK FOR MASAR DATA

The general problem of the estimation of scatterer along the elevation profile is thoroughly investigated (Zhu and Bamler, 2010), (Schmitt and Stilla, 2013). Since the number of available measurements is generally limited the estimation problem is ill-posed introducing the need to add further constraints on the solution. In the context compressive based SAR tomography the solution is achieved by using the reconstruction method LASSO (Tibshirani, 1996):

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 + \lambda \|\mathbf{x}\|_1. \quad (1)$$

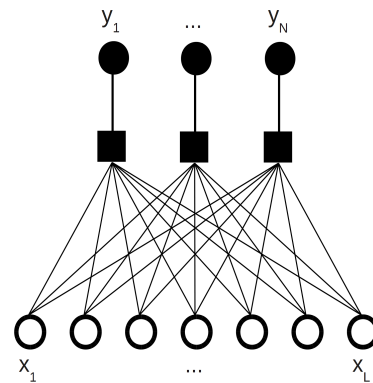


Figure 1: Factor graph that correspond to Eq. 3. Empty circles depict variables x_i , $i \in [L]$ and squares are representatives of factors. Filled circles represent the observations y_j , $j \in [N]$. They are often directly included in the factors. The factorized structure of Eq. 4 is obviously represented by this bipartite graph. Nonzero entries in \mathbf{A} correspond to edges in the graph.

where \mathbf{y} is a N -dimensional measurement vector, \mathbf{A} an $N \times L$ transformation matrix and \mathbf{x} the L -dimensional discrete reflectivity profile. The Lagrange multiplier λ depends on the number of nonzero entries in \mathbf{x} which represent the scatterers along the elevation profile. The determination of the correct value of λ is subject to active research. In some cases λ is set to the noise threshold but there are also other methods like GCV or L-curve (Batu and Cetin, 2011). Besides the methods found in the context of TomoSAR there are also graphical model approaches to such reconstruction problems (Montanari, 2012). These approaches find solutions by maximizing a joint probability distribution on (\mathbf{x}, \mathbf{y}) which are generally of the form

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y}|\mathbf{x}) p(\mathbf{x}). \quad (2)$$

The conditional distribution $p(\mathbf{y}|\mathbf{x})$ is the likelihood term, a data driven ... which includes the noise process/model. The prior distribution $p(\mathbf{x})$ on the other hand encodes some information and model knowledge e.g. the sparsity property within the compressive sensing. In the case of Gaussian distribution with zero mean

and variance σ^{-1} , the solution is the MAP estimate

$$\hat{\mathbf{y}} = \arg \max_y p(\mathbf{x}|\mathbf{y}) \quad (3)$$

$$p(\mathbf{x}|\mathbf{y}) = \frac{1}{Z(\mathbf{y})} \prod_{i=1}^N \exp\left(-\frac{\sigma}{2}(y_i - A_{i,\cdot}\mathbf{x})^2\right) p(\mathbf{x}) \quad (4)$$

where $A_{i,\cdot}$ is the i -th row of \mathbf{A} and $Z(\mathbf{y})$ ensures that $p(\mathbf{x}|\mathbf{y})$ is a proper density. Despite the explicit expression of the posterior the computation of expectations or marginals of it is in general very complicate due to the fact, that any pair of variables from the set $\{x_i\}, i \in [L]$ is conditionally dependent given any $y_k, k \in [N]$. The MAP solution factorizes nicely and can conveniently be described by a factor graph, which in this case is a bipartite graph, exploiting the structure of the formula clearly. This is depicted in Fig. 1.

A drawback of graphical models is that efficient inference algorithms exists only for discrete or Gaussian variables. Since in the presented estimation the variables are not within one of these categories some further developments were necessary. A new messaging framework called *Approximate Message Passing* (AMP) was introduced by (Donoho et al., 2009). As the name suggests the algorithm uses approximations in the sense, that for the large limit $N, L \rightarrow \infty$, the sum of many distributions is well approximated by a Gaussian. This procedure was expanded to complex valued variables by (Maleki et al., 2013 (in Press)) and is known as *Complex Approximate Message Passing* (CAMP).

2.1 Aspect Fusion and Graph Topology

The aim of the paper is not to investigate in new optimization procedures but to show how the MASAR domain can be modeled effectively in terms of graphical models. In order to introduce our concept consider a general factor graph $G = (V, \Phi, E)$ with V a set of variable nodes, F a set of factor nodes and the edges E connecting variables with factors. A *conditional random field* (CRF) encodes a conditional distribution as follows:

$$p(\mathbf{x}|\mathbf{y}) = \frac{1}{Z(\mathbf{y})} \tilde{p}(\mathbf{x}, \mathbf{y}) \quad (5)$$

$$\tilde{p}(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^m \Phi_i(\mathbf{D}_i) \quad (6)$$

where $Z(\mathbf{y}) = \sum_{\mathbf{x}} \tilde{p}(\mathbf{x}, \mathbf{y})$ is so called the partition function. Factors Φ and cliques \mathbf{D} are explained in more detail in the next section. At this point it is worth to emphasize that the CRF representation avoids the encoding of the distribution of \mathbf{y} . This is a advantage because in most cases it might be even impossible to derive a proper and descriptive distribution for the measurements due to the very complex interrelationships between them.

In our domain a variable $v \in V$ represents a scatterer in the scene. A typical scene is depicted in in Fig. 2. One can see that there are three scatterer in every elevation profile and that they share a common point, which is at the intersection of the elevation profiles. Obviously the measurements made in both acquisitions are dependent on each other. Graphical models enable to take this easily into account. Dependencies in graphical models are expressed by the factors Φ . Any subset of variables, which are constrained to each other, are attached to a single factor and this subset is then called a *clique*. As an example a factor is introduced for every measurement connecting all variables of one elevation profile to the measurement. Thus these factors represent the *likelihood*. Although we aim to estimate the correct scatterer location and amplitudes the underlying structure of our problem

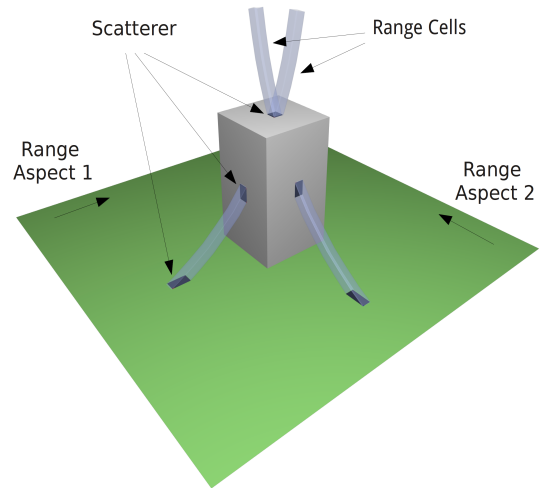


Figure 2: Three dimensional sketch of the intersection of multiple elevation profiles from different aspects. Obviously both measurements share a common scatterer. This way the aspects are not independent from each other. The corresponding structure in terms of the graph topology is shown in Fig. 3.

statement is completely different from the standard TomoSAR. In the MASAR domain we have obviously more evidence to estimation of the scatterer locations than just a single measurement yet we do not have a stack of observations originating from the very same scatterer. The implication is that we cannot built up a dense system matrix \mathbf{A} complying with the *Restricted Isometry Property* (RIP). Instead we would have a larger matrix which is sparse. The increase in size stems from the fact that the domain of the system matrix \mathbf{A} has increased to all scatterer in the scene. This prevents us from using powerful and well known spectral methods. For the same reasons we are also not able to apply the CAMP method. The sparse nature of the problem violates the large limit assumption that messages can be approximated with Gaussian densities within an tolerable error bound.

In general PGMs are an appropriate framework for sparse problems with complex interrelations between random variables. Their ability to encode conditional independencies very effective makes them a powerful tool. A typical graph topology of the MASAR domain is shown in Fig. 3. In our model we generate a set of variables for every aspect separately. This means, that a common scatterer as indicated in Fig. 2 is represented by multiple variables and not by just one. The reason for this is an increase in flexibility of the model if e.g. a point is not visible in every aspect. Of course the variables exhibiting are small pairwise spatial distance should be clustered to share a common behaviour. This is explained later in the next section.

2.2 Features

As stated in the previous section the dependencies between variables of subsets of V are expressed in terms of factors. One way to represent the factor is the *log-linear* model

$$\Phi_i(\mathbf{D}) = \exp(-\beta_i f_i(\mathbf{D})) \quad (7)$$

where $f_i(\mathbf{D})$ is a feature over a subset \mathbf{D} of variables and β_i a weighting factor controlling the influence. In a slightly different setup f_i can be interpreted as *energy function*. This section presents the different feature used to compose the joint probability of the MASAR domain.

It was already mentioned that graphical models are restricted to

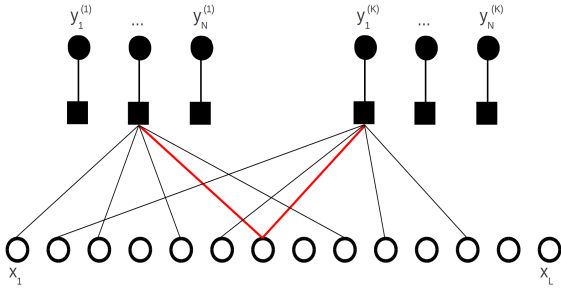


Figure 3: Sparse graph topology for the MASAR domain. The red lines indicate the edges of the scatterer (variable) which connects two observations from different aspects (denoted by the superscript). Note that this graph would correspond to a sparse system matrix \mathbf{A} .

discrete or Gaussian variables for practical use. The preceding section explained why it is not possible in this problem statement to introduce approximations to overcome these limitations. The general way of setting up a graphical model with discrete variables is to define a set of states θ for every $v \in V$. Assume that every variable $v \in V$ has a set of attached states $\theta = \{\theta_0, \dots, \theta_n\}$ where every state $\theta_i = (a_i, \rho_i)$ is a tuple comprised by the parameters a as the amplitude of the scatterer and ρ as its interferometric phase or position. Since some scatterer might be nonexistent there is a *null state* θ_0 attached to every variable. We define M scatterers per elevation profile thus we can resolve up to M scatterers per range cell.

2.2.1 Measurement likelihood At first we introduce the data term of the model which is the likelihood. Let $N_i^{(k)}$ be the set of scatterer which belong to an observation i in aspect k . Taking a look at Fig. 2 these are just the points along the elevation profiles. The feature is given by

$$f_d(\mathbf{D} = N_i^{(k)} | y_i^{(k)}) = \left\| \sum_{j \in N_i^{(k)}} \hat{a}^{(j)} \exp(j\hat{\rho}^{(j)}) - y_i^{(k)} \right\|_2^2 \quad (8)$$

where $\theta^{(j)} = (\hat{a}^{(j)}, \hat{\rho}^{(j)})$ is the current state of variable j . With this formulation we use the underlying assumption of a Gaussian distribution. This factor is introduced for every measurement in all aspects.

2.2.2 Sparsity The main objective on the quality of the solution is its sparsity and we prefer solutions with as few scatterers as necessary. In mathematical notation this would imply to minimize the L_0 -norm which is infeasible. Instead, similar to standard CS, we use the L_1 -norm as an approximation (Tibshirani, 1996). For every variable $v \in V$ we insert a new factor according to

$$f_s(\mathbf{D} = v) = |\hat{a}|. \quad (9)$$

into the graph. Concerning just a single variable this factor has a precedence for scatterers with small amplitudes which globally tends to produce solutions with only a few scatterers.

2.2.3 Spatial smoothness In most cases we prefer solutions which are in a certain sense simple. By this we mean solutions where points compose finite planes. Let $B_k(v_i)$ be the set of the k nearest variables of v_i with respect to the euclidean or Mahalanobis distance. A new factor is generated for every element of $B_k(v_i)$ and v_i . This factor tries to minimize the L_1 -norm of the horizontal and vertical distance, defined by $d_h(v_i, v_j)$ and

$d_v(v_i, v_j)$ respectively. In our setup we used this simple model but this can easily be enhanced by first estimating a common plane with e.g. PCA and then using a corresponding coordinate system.

$$f_p(\mathbf{D} = \{v_i, v_j\}) = g(\alpha_v d_v(v_i, v_j) + \alpha_h d_h(v_i, v_j)) \quad (10)$$

Not that the position of the scatterer can be inferred from the phase ρ and the acquisition geometry such that we simply write $d(v_i, v_j)$ to indicate the distance between two variables. It is important to allow discontinuities in the spatial domain as they occur frequently in urban areas e.g. on the transition from roof to facade. Therefore we use a bounded or truncated function g which limits the maximum penalty value e.g. the truncated quadratic potential function $g(x) \equiv g_\gamma(x) = \min\{\gamma, x^2\}$. Furthermore it is important to normalize $f_p(\mathbf{D})$ by the cardinality of $B_k(v_i)$.

2.3 Local Optimizations

As stated above the variables in a PGM are often discrete for practical reasons. This is a clear limitation considering the 3d reconstruction task. One solution would be to apply a very fine sampling of continuous variables but unfortunately the complexity of the PGM is exponential in the cardinality of its variables. To overcome these limitation we propose an iterative approach where we introduce new states for the variables and thus mimic the continuous nature. What we want to achieve is to adjust the variables in the way that they can reproduce the measurement.

Assume that $M_{i,j}^{(k)}$ is the set of variables which contribute to the measurement in aspect k at azimuth position i and range j . Let $I_{i,j}^{(k)}$ be an index set addressing all variables of $M_{i,j}^{(k)}$ which predicted state is distinct from the null state θ_0 . Then it is possible to reconstruct the observation by

$$\tilde{y}_{i,j}^{(a)} = \sum_{p \in I_{i,j}^{(a)}} \hat{a}^{(p)} \exp(j\hat{\rho}^{(p)}). \quad (11)$$

In cases the states θ_i are not properly defined, this reconstruction differs from the true observation. To allow the model to approximate the true observation in the next iteration we generate a set of improvements for the current states. Since the number of variables as well as their states are known we can use an estimation scheme without unknowns that just aims at adjusting the inconsistencies among the observations. In this case we use a strict Gauss-Helmert model. Our equation which should be fulfilled is

$$f(\theta) = \|y - \mathbf{a}^T \boldsymbol{\rho}\|_2^2 = 0 \quad (12)$$

constrained to $\mathbf{a} \geq \mathbf{0}$. After this computation we generate for every variable a new set of states composed of the null state, the current state and the newly generated state as a new hypothesis. Additionally we append also a new state generated from the positions of the local neighbourhood. This state serves as an opponent to the state obtained from the adjustment since it is more based on the prior of the variables than on the measurements.

It is the task of the inference algorithm to choose the most proper state as a new prediction. Using the iterative scheme we can generate a continuous solution space from discrete variables where the solution evolves over the time. Note that we use local optimizations on a global state in order to generate a global solution. The workflow of our approach is depicted in Fig. 4. Furthermore we can control the adjustment by providing weights for all observations we want to adjust. This way we can control the amount of improvement for every observed variable. The weights can be inferred from the local neighbourhood. A low $f_p(v_i, v_j)$ for every $v_j \in B_k(v_i)$ according to Eq. (10) is an indicator for a high reliability. Of course, it should not be modified more than other

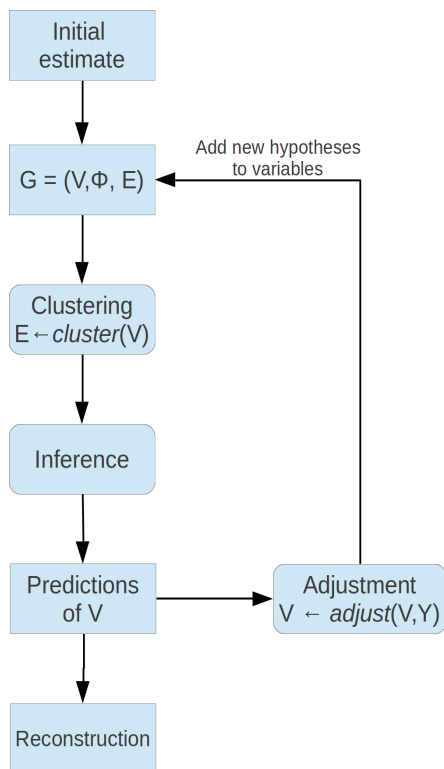


Figure 4: Workflow of the presented approach. The computation is divided into global and local parts. Whereas the inference computes the MAP based on the global parametrization of the graph G , the adjustment performs only a local operation based on the predicted states on the variables V and the corresponding observations Y . The iterations continues until a stop criterion is reached but in our experiments the number of iterations was fixed and chosen to be large enough.

points with a smaller support on their neighbourhood.

3 SIMULATION RESULTS

The presented method is evaluated on simulated single-pass InSAR data according to (Thiele et al., 2007b). The scene with an extend of $80 \text{ m} \times 80 \text{ m}$ contains two simple cubes as representative primitives for a building. The lower building block has a height of 12 m and the the other one 20 m. The sensor is at a height of 700 m and the slant range to the scene center is 1237 m. The baseline is $B = 5.5 \text{ cm}$ with an inclination of $\alpha = 50^\circ$. The SAR system operates in the Ka-Band with a center frequency of 35 GHz. The initial scatter position are placed randomly along their elevation profiles. This ensures that the state is not initialized nearby a local maximum in the search space. The resolution is approximately 5 m in range and azimuth. As inference method we use the *Tree Reweighted Belief Propagation* (TRWBP) algorithm which belongs to the family of message passing inference methods (Wainwright et al., 2003).

All scatterers were simulated with an amplitude uniformly distributed between 0.8 and 1.0. The maximum number of scatterers per range cell were set to $M = 4$. The reconstruction result for a SNR = 20 dB is shown in Fig. 5.

4 DISCUSSION

In the reconstruction result one can clearly recognize the building. Although the heights are not reconstructed exactly one can clearly see the transition from the building to the ground and even from the higher to the lower part of the building. The result seems

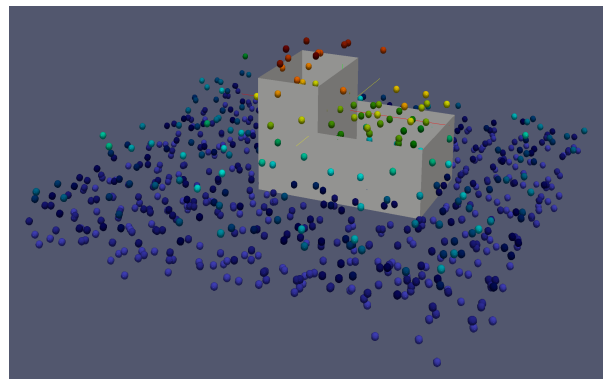


Figure 5: Reconstruction result obtained after 10 iterations with SNR = 20dB. The color encodes the height of the scatterers. The walls are visualized by surfaces although this might hide some scatterers which are located slightly behind the wall. The scene was illuminated from all four directions such that a wall is only visible in one aspect. Note that the resolution of $5 \text{ m} \times 5 \text{ m}$ is quite low. Note also that the points on the facades belong to just one aspect. This implies that here the point density is four times lower than in areas visible in all aspects.

to be promising for further research and investigation. Note especially the scatterers on the facades of the building. This indicates that the approach is indeed able to resolve multiple scatterer in a range cell. Of course, a numerical evaluation of the accuracy should be subject to further analysis in the future. The fact that the occasionally distributed scatterers over the ground are not aligned to the common height might be caused by the discontinuity preserving behaviour of the model.

5 CONCLUSIONS

This paper presented a novel approach for layover separation and scatter estimation based on the fusion of multi aspect InSAR data. The method has proven to work well on simulated data. As next step the approach needs to be evaluated on real data. This would be done in the future on data containing an urban scene in Munich. Although the method seems to be robust there are many parameters which have to be tuned carefully. Until now there is no comprehensive understanding on the effects of the parameters. This is also subject to further research and numerical evaluation. There are also many open issues concerning the speed of convergence and the initialization of the initial states. Furthermore the presented approach is just an intermediate step towards a comprehensive building reconstruction system which includes high level models. The next step would be to include geometrical primitives e.g. planes or cubes as simple primitives buildings are build from. These primitives have to be included into the reconstruction process which could be done via the EM-algorithm.

REFERENCES

- Batu, O. and Cetin, M., 2011. Parameter selection in sparsity-driven SAR imaging. *EEE Trans. Aerosp. Electron. Syst.* 47(4), pp. 3040–3050.
- Bolter, R., 2001. Buildings from SAR: Detection and reconstruction of buildings from multiple view high resolution interferometric SAR data. PhD thesis, University of Graz.

- Brenner, A. and Roessing, L., 2008. Radar Imaging of Urban Areas by Means of Very High-Resolution SAR and Interferometric SAR. *IEEE Trans. Geosci. Remote Sens.* 46(10), pp. 2971–2982.

Donoho, D. L., Maleki, A. and Montanari, A., 2009. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences* 106(45), pp. 18914–18919.

Hebel, M. and Stilla, U., 2007. Automatic registration of laser point clouds of urban areas. In: *International Archives of Photogrammetry and Remote Sensing*, Vol. 36number W49A, pp. 13–18.

Maleki, A., Anitori, L., Yang, Z. and Baraniuk, R., 2013 (in Press). Asymptotic analysis of complex LASSO via complex approximate message passing (CAMP). *IEEE Transactions on Information Theory*.

Montanari, A., 2012. *Compressed Sensing*. Cambridge University Press, chapter Graphical Models Concepts in Compressed Sensing, pp. 394–438.

Schmitt, M. and Stilla, U., 2011. Fusion of airborne multi-aspect InSAR data by simultaneous backward geocoding. In: *Joint Urban Remote Sensing Event (JURSE)*, pp. 53 –56.

Schmitt, M. and Stilla, U., 2013. Compressive Sensing Based Layover Separation in Airborne Single-Pass Multi-Baseline InSAR Data. *IEEE Geoscience and Remote Sensing Letters* 10(2), pp. 313–317.

Stilla, U., Soergel, U., Thoenessen, U. and Brenner, A. R., 2005. Potential and limits for reconstruction of buildings from high resolution SAR data of urban areas. In: *Proc. of International Union of Radio Science (URSI)*.

Thiele, A., Cadario, E., Schulz, K., Thoenessen, U. and Soergel, U., 2007a. Building Recognition From Multi-Aspect High-Resolution InSAR Data in Urban Areas. *IEEE Trans. Geosci. Remote Sens.* 45(11), pp. 3583–3593.

Thiele, A., Cadario, E., Schulz, K., U., T. and Soergel, U., 2007b. InSAR Phase Profiles at Building Locations. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 36, pp. 202–208.

Tibshirani, R., 1996. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society* 58(1), pp. 267–288.

Wainwright, M. J., Jaakkola, T. S. and Willsky, A. S., 2003. Tree-reweighted belief propagation algorithms and approximate ML estimation by pseudo-moment matching. In: *9th Workshop on Artificial Intelligence and Statistics*.

Zhu, X. X. and Bamler, R., 2010. Tomographic SAR Inversion by L1-Norm Regularization: The Compressive Sensing Approach. *IEEE Transactions on Geoscience and Remote Sensing* 48(10), pp. 3839–3846.