

Fusion of high-resolution InSAR data and optical imagery for building detection using Conditional Random Fields

Jan Dirk Wegner, Uwe Soergel, Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover, Germany

Ronny Haensch, Computer Vision and Remote Sensing Department, Berlin Institute of Technology, Germany
Antje Thiele, Fraunhofer Institute of Optronics, System Technologies and Image Exploitation, Germany

Abstract

State-of-the-art satellite SAR sensors (e.g., TerraSAR-X, CosmoSkyMed) provide imagery of one meter resolution and airborne SAR sensors achieve even higher resolutions. In those data objects in urban areas become visible in high detail. However, SAR-typical effects like layover, shadowing, and interfering backscatter of multiple objects complicate interpretability. Thus, additional information about those objects may be obtained from optical imagery. In this work we combine features of high-resolution airborne interferometric SAR (InSAR) data with features of an orthophoto in order to detect buildings. A Conditional Random Field (CRF) is set up in order to integrate context-knowledge. We show that CRFs are a suitable method for integration of both context-knowledge and multi-sensor features for building extraction.

1 Introduction

Modern space borne and airborne SAR sensors are capable of mapping urban areas with high detail. They may acquire imagery at night or through cloud coverage due to their active sensor principle and the dielectric properties of the used microwave signal. Those properties make them a valuable tool for many applications, for example for rapid hazard response after natural disasters. However, interpretability in urban areas is complicated by occlusions and interfering backscatter of multiple objects. Therefore, automatic scene analysis may be facilitated by additional data like optical images.

Xiao et al. [1] and Hepner et al. [2] detect and reconstruct building blocks combining InSAR data with high-resolution multi-spectral images or hyperspectral images, respectively. Tupin and Roux [3] represent segments of an aerial photo in a region adjacency graph, which is then used within a Markov Random Field framework to regularize building heights determined by means of radargrammetry. Poullain et al. [4] combine high-resolution optical and SAR data with vector data with the goal of change detection using Dempster-Shafer evidential theory. Sportouche et al. [5] detect and three-dimensionally reconstruct large industrial buildings semi-automatically based on features of Quickbird imagery and TerraSAR-X data. We recently proposed a segment-based approach for building detection [6]. Segments of an orthophoto are classified into building and non-

building segments based on features of an orthophoto and InSAR double-bounce lines.

In order to exploit context-knowledge we use a Conditional Random Field (CRF) framework [7]. CRFs have already been successfully applied to various computer vision tasks [8,9], whereas they have only rarely been applied to remote sensing data [10,11], yet.

In the following we will first explain the building features in the orthophoto and the InSAR data. Then, CRFs are described. Finally, we present some experimental results and evaluate the overall contribution of the InSAR double-bounce lines to the detection performance.

2 Features

In order to detect buildings we have to first discover features that discriminate them from their environment. We use high-resolution InSAR data (**Figure 1(a,b)**) and an orthophoto as input (**Figure 1(c,d)**) and thus we have to first investigate the typical appearance of buildings in such data. In very high-resolution aerial imagery details of buildings like superstructures on the roof become visible. Additionally, facades may be partially visible due to the central perspective of the camera (and since we are not dealing with a true orthophoto). High-resolution SAR data provides complementary information. A flat-roofed building is usually characterized by layover, a double-bounce line where the building wall facing the SAR sensor meets

the ground, direct backscatter from the roof (depending on the width of the building in range direction), and a shadow area. Particularly the double-bounce line is a reliable building hint in urban areas [12].

We subdivide the images into patches and generate rather simple feature vectors for each patch since our current focus is on the overall suitability of CRFs for combining high-resolution optical and SAR data for building detection. Feature selection is accomplished empirically by testing various features and feature combinations.

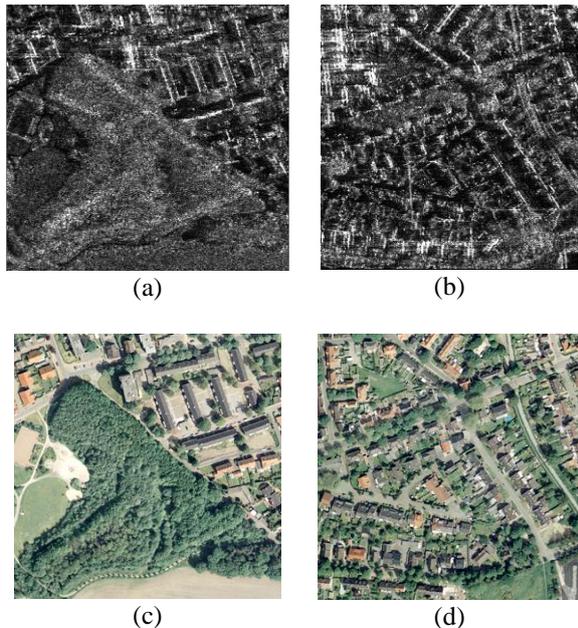


Figure 1 Test regions of the Dorsten scene: (a,b) orthorectified SAR magnitude images of the Aes-1 interferometric image pair (range from right to left), (c,d) corresponding regions of the orthophoto.

2.1 Orthophoto features

The most discriminative orthophoto features are found to be based on colour, intensity, and gradient. We take mean and standard deviation of red and green channel normalized by the length of the RGB vector as well as mean and standard deviation of the hue channel as features. In addition, variance and skewness of the gradient orientation histogram proved to be discriminative. We generate all those features in multiple scales [13] in order to mitigate shortcomings due to instable features of very small patches. Features of large patches integrate over bigger areas thus excluding, for example forests or agricultural areas, whereas small patches provide details. We tested various numbers of scales and scale combinations. Three different scales with sizes 10x10, 15x15, and 20x20 pixels were found to be appropriate in our case.

2.2 InSAR features

Buildings in InSAR data appear differently than in traditional optical data (cf. **Figure 1 (a,b)** and **(c,d)**). Although layover and shadowing effects carry valuable geometric and radiometric information about the object of interest, they often complicate automatic analysis of urban areas because of multiple backscatter interferences. Therefore, we utilize double-bounce lines as characteristic building hints. They are the most reliable feature in urban areas [12] because they show high coherence in InSAR data indicating high signal-to-noise-ratio. In addition, the distribution of the interferometric heights around those lines facilitates distinguishing between building lines and bright lines caused by other effects.

We segment double-bounce lines in the InSAR data applying the technique we developed in [6]. Those lines are then projected from slant range to the coordinate system of the orthophoto using the local mean interferometric height at each line position. Next, we smooth the intensity channel of the orthophoto with an edge preserving anisotropic diffusion filter, segment the smoothed image into homogenous regions using watershed segmentation, and overlay the double-bounce lines to the those segments. Intersecting segments are set to one (all others to zero) and a distance map is generated (**Figure 2(a,b)**).

3 Conditional Random Fields

CRFs have been widely used in several fields of applications, particularly in computer vision [9,13]. As discriminative models they directly estimate the posterior distribution $P(\mathbf{y}|\mathbf{x})$ of class labels \mathbf{y} given the data \mathbf{x} without wasting resources on modelling the joint probability, which is mostly not of interest. This paper applies a functional form of CRFs as proposed for example in [13] given by **Equation 1**:

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

where the association potential $A_i(\mathbf{x}, y_i)$ and the interaction potential $I_{ij}(\mathbf{x}, y_i, y_j)$ define the capabilities of the CRF. The partition function $Z(\mathbf{x})$ is merely a normalization constant (depending only on data \mathbf{x}). The image \mathbf{x} is assumed to consist of a set of label sites S (the image patches in our case), and the indices i and j denote an arbitrarily label site in S and a label site in the set N_i of all available adjacent label sites, respectively. Although the association potential $A_i(\mathbf{x}, y_i)$ (**Equation 2**) measures the probability of a certain class label at site i given the feature vector $\mathbf{h}_i(\mathbf{x})$ of this site, it does not have to be a probability by itself.

$$A_i(\mathbf{x}, y_i) = \exp(y_i \mathbf{w}^T \mathbf{h}_i(\mathbf{x})) \quad (2)$$

Note that $\mathbf{h}_i(\mathbf{x})$ depends on all available data \mathbf{x} and can therefore potentially use other parts of the data than those of the current label site. We use features calculated over three different scales to capture scale dependent characteristics, which are expected to be distinguishable between buildings and non-buildings. The weight vector \mathbf{w} contains the first set of free parameters, which have to be adjusted during training. As proposed in [13] a quadratic expansion is applied during calculation of $\mathbf{h}_i(\mathbf{x})$ generating a non-linear decision surface.

The interaction potential $I_{ij}(\mathbf{x}, y_i, y_j)$ (Equation 3) measures the cost of assigning different labels to two adjacent label sites.

$$I_{ij}(\mathbf{x}, y_i, y_j) = \exp(y_i y_j \mathbf{v}^T \boldsymbol{\mu}_{ij}(\mathbf{x})) \quad (3)$$

In contrary to standard MRFs it is data dependent by $\boldsymbol{\mu}_{ij}(\mathbf{x}) = \mathbf{g}_i(\mathbf{x}) - \mathbf{g}_j(\mathbf{x})$, where $\mathbf{g}_i(\mathbf{x})$ is the feature vector for label site i , which again could potentially use the whole data set \mathbf{x} . Note that different functionals are used to calculate feature vectors in $A_i(\mathbf{x}, y_i)$ and $I_{ij}(\mathbf{x}, y_i, y_j)$. Therefore, different features can be used in both potentials. Our current implementation defines the interaction potential only in terms of single scale features and no quadratic expansion is used. The weight vector \mathbf{v} of the interaction potential completes the set of system parameters, which have to be optimized during training.

Various training and inference methods were investigated. The Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) [14] method and Loopy Belief Propagation (LBP) [15] were found to lead to the most promising results.

4 Experiment

The InSAR data we use was acquired with the AeS-1 sensor of Intermap Technologies. The spatial resolution is about 0.38 m in range and 0.16 m in azimuth. The effective baseline between both X-band sensors was set to approximately 2.4 m. The optical image was taken with an analogue aerial camera Zeiss RMK and scanned leading to a pixel size of 0.31 m on the ground.

In order to assess the impact of the InSAR double-bounce lines on the overall building detection performance we compare CRF results based on merely the orthophoto features (Figure 2(c,d) and Table 1 left) with those exploiting the combination with InSAR double-bounce lines (Figure 2(e,f) and Table 1 right). For testing purposes we define four test regions and accomplish 4-fold cross validation. The true posi-

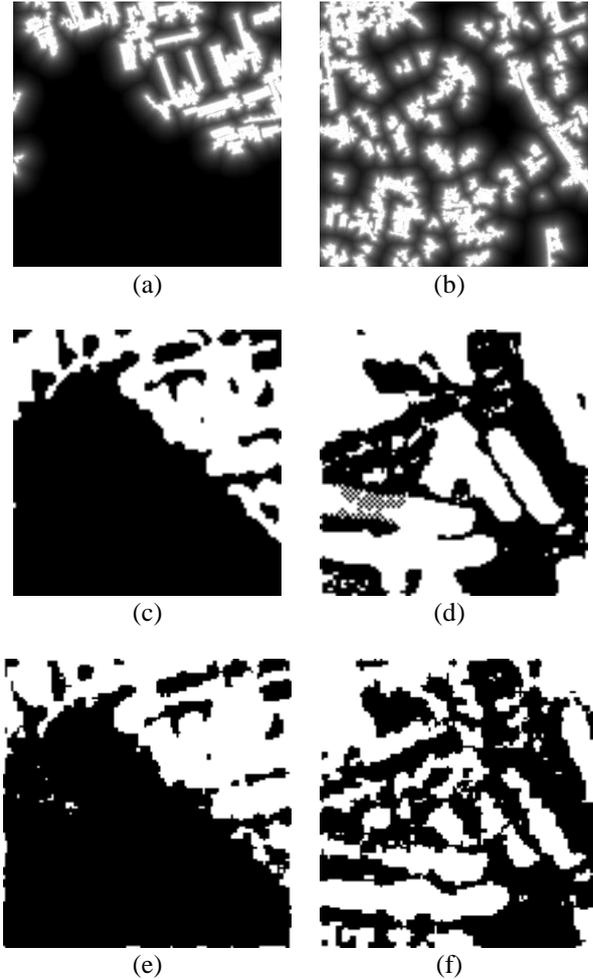


Figure 2 Test regions of the Dorsten scene as shown in Fig. 1: (a,b) distance maps of the segments that intersect with double-bounce line, (c,d) CRF building detection results using only orthophoto features, (e,f) CRF building detection results based on combined orthophoto and InSAR features.

tive rate (TPR) on a per-pixel level of both results is 85%. However, the InSAR double-bounce lines decrease the false positive rate (FPR) from 30% to 27%. Those relatively high FPRs are due to over-smoothing caused by the standard interaction potential. Small gaps between adjacent buildings are misclassified as buildings.

Table 1 CRF building detection results using only orthophoto features versus a combination of orthophoto and InSAR features: Mean and standard deviation of true positive rate (TPR) and false positive rate (FPR) evaluated on a per-pixel level.

| Orthophoto | | | | Orthophoto+InSAR | | | |
|------------|----------|-------|----------|------------------|----------|-------|----------|
| TPR | | FPR | | TPR | | FPR | |
| μ | σ | μ | σ | μ | σ | μ | σ |
| 85% | 6% | 30% | 13% | 85% | 7% | 27% | 8% |

5 Conclusions and future work

In conclusion, CRFs are suitable for building detection using multi-sensor data and the overall detection performance benefits from the use of complementary optical and InSAR features. However, the current standard interaction potential function needs improvements. It is appropriate for usual computer vision tasks where a single relatively large object has to be detected in an image (since it is more or less a smoothing term) but inconvenient for our task of detecting multiple densely distributed small objects. Our next efforts will thus go into the introduction of sophisticated discontinuity constraints. Furthermore, we will investigate the benefits of setting up the CRF on irregularly distributed segments generated by, for example Normalized Cuts or Mean Shift.

References

- [1] Xiao, R., Leshner, C., and Wilson, B. 1998. *Building Detection and Localization Using a Fusion of Interferometric Synthetic Aperture Radar and Multispectral image*. In: Proceedings of ARPA Image Understanding Workshop, pp. 583-588.
- [2] Hepner, G.F., Houshmand, B., Kulikov, I., and Bryant, N. 1998. *Investigation of the Integration of AVIRIS and IFSAR for Urban Analysis*. Photogrammetric Engineering and Remote Sensing, Vol. 64, No. 8, pp. 813-820.
- [3] Tupin, F., and Roux, M. 2005. *Markov Random Field on Region Adjacency Graph for the Fusion of SAR and Optical Data in Radargrammetric Applications*. IEEE Transactions on Geoscience and Remote Sensing, Vol. 42, No. 8, pp. 1920-1928.
- [4] Poulain, V., Inglada, J., Spigai, M., Tourneret, J.-Y., and Marthon, P. 2009. *Fusion of high resolution optical and SAR images with vector data bases for change detection*. In: Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS'09), 4 pages.
- [5] Sportouche, H., Tupin, F., and Denise, L. 2009. *Building Extraction and 3D Reconstruction in Urban Areas from High-Resolution Optical and SAR Imagery*. In: Proceedings of Joint Urban Remote Sensing Event (URBAN '09), 11 pages.
- [6] Wegner J.D., Thiele, A., and Soergel, U. 2009. *Fusion of optical and InSAR features for building recognition in urban areas*. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 38, Part 3/W4, 2009, pp. 169-174.
- [7] Lafferty, J., McCallum, A., and Pereira, F. 2001. *Conditional Random Fields: Probabilistic Models for segmenting and labeling sequence data*. In: Proceedings of International Conference on Machine Learning, 8 pages.
- [8] Kumar, S., and Hebert, M. 2003. *Discriminative Random Fields: A Discriminative Framework for Contextual Interaction in Classification*. In: Proceedings of IEEE International Conference on Computer Vision (ICCV '03), Vol. 2, pp. 1150-1157.
- [9] Korč, F., and Förstner, W. 2008. *Interpreting Terrestrial Images of Urban Scenes using Discriminative Random Fields*. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 37, part B3a, 2008, pp. 291-296.
- [10] Zhong, P., and Wang, R. 2007. *A Multiple Conditional Random Fields Ensemble Model for Urban Area Detection in Remote Sensing Optical Images*. IEEE Transactions on Geoscience and Remote Sensing, Vol. 45, No. 12, pp. 3978-3988.
- [11] He, W., Jäger, M., Reigber, A., and Hellwich, O. 2008. *Building Extraction from Polarimetric SAR Data using Mean Shift and Conditional Random Fields*. In: Proceedings of European Conference on Synthetic Aperture Radar, Vol. 3, pp. 439-443.
- [12] Thiele, A., Wegner, J., Soergel, U. 2010. *Building reconstruction from multi-aspect InSAR data*. In U. Soergel (Ed), Radar Remote Sensing of Urban Areas, Springer, 1st Edition, ISBN-13: 978-9048137500.
- [13] Kumar S., and Hebert, M. 2006. *Discriminative Random Fields*. International Journal of Computer Vision, Vol. 68, No. 2, pp. 179-201.
- [14] Liu, D.C., and Nocedal, J. 1989. *On the limited memory BFGS for large scale optimization*. Mathematical Programming, Vol. 45, No. 1-3, pp. 503-528.
- [15] Frey, B.J., and MacKay, D.J.C. 1998. *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, Vol. 10, MIT Press.