URBAN BUILDING DETECTION FROM OPTICAL AND INSAR FEATURES EXPLOITING CONTEXT

J. D. Wegner [a,*] , A. O. Ok [b], A. Thiele [c], F. Rottensteiner [a], U. Soergel [a]

[a] Institute of Photogrammetry and GeoInformation (IPI), Leibniz Universität Hannover, Hannover, Germany – (wegner, soergel)@ipi.uni-hannover.de

[b] Dept. of Geodetic and Geographic Information Technologies, Middle East Technical University, Ankara, Turkey – oozgun@metu.edu.tr

[c] Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB), Ettlingen, Germany – antje.thiele@iosb.fraunhofer.de

Commission III, WG III/4

KEY WORDS: Conditional Random Fields, Remote Sensing, Fusion, InSAR Data, Optical Stereo Data, Urban Area

ABSTRACT:

We investigate the potential of combined features of aerial images and high-resolution interferometric SAR (InSAR) data for building detection in urban areas. It is shown that completeness and correctness may be increased if we integrate both InSAR double-bounce lines and 3D lines of stereo data in addition to building hints of a single optical orthophoto. In order to exploit context information, which is crucial for object detection in urban areas, we use a Conditional Random Field approach. It proves to be a valuable method for context-based building detection with multi-sensor features.

1. INTRODUCTION

Building detection in urban areas based on merely a single aerial photo is often hard to conduct (Mueller and Zaum, 2005). Features of additional data sources may be introduced to improve detection completeness and correctness. In addition to features derived from an orthophoto we use building hints of high-resolution InSAR data and an optical stereo image pair.

Several works have already dealt with the integration of features derived from high-resolution optical and SAR (or InSAR) data with the goal of building detection. Xiao et al. (1998) detect and reconstruct building blocks combining high-resolution optical and InSAR data. They classify both data sets separately within a multi-layer neural network followed by morphological operations. Finally, rectangles are fit to building hypothesis and heights are derived. Hepner et al. (1998) jointly use hyper-spectral imagery and InSAR data acquired by airborne sensors to detect and three-dimensionally reconstruct large buildings in urban areas. Tupin and Roux (2003) propose an approach to extract footprints of large flat-roofed industrial buildings based on line features. In (Tupin and Roux, 2005) the same authors represent homogeneous regions of an aerial photo with a region adjacency graph. This graph is then used within a Markov Random Field framework to regularize building heights determined by means of radargrammetry. A discontinuity constraint based on the image gradient along segment boundaries is introduced into the prior term in order to preserve sudden height jumps. Poulain et al. (2009) combine high-resolution optical and SAR data with vector data in order to detect changes. Since no learning step is conducted all classification is performed based on prior knowledge. They generate features from previously extracted primitives and set up a score for each building site using Dempster-Shafer evidential theory. Sportouche et al. (2009) detect and three-dimensionally reconstruct large industrial buildings semi-automatically. They combine features of high-resolution optical satellite imagery (Quickbird) with high-resolution SAR data (TerraSAR-X). Building hypothesis of the optical data are validated or rejected based on a classification of the SAR image making use of roof textures, bright lines, and shadows. Building heights are derived simultaneously exploiting the different optical and SAR sensor geometries. We recently proposed a segment-based approach for building detection (Wegner et al., 2009). Segments of an orthophoto are classified in combination with InSAR double-bounce lines.

In this paper, we use a Conditional Random Field (CRF) framework, which is a probabilistic contextual classification framework originally introduced by Lafferty et al. (2001) for labelling 1D sequential data and later on extended to images by Kumar and Hebert (2003). CRFs have already been successfully applied to various computer vision tasks (e.g., Rabinovich et al., 2007; Korč and Förstner, 2008). Nonetheless, CRFs have only rarely been applied to remote sensing data (Zhong and Wang, 2007). Furthermore, to the authors knowledge only one publication exploits CRFs for the analysis of SAR data (He et al., 2008).

Our focus is on the suitability of CRFs for combining multi-sensor remote sensing data using context with the aim of single building detection. Although much more sophisticated features could potentially be derived from stereo and InSAR data we use rather simple ones in order to transparently assess the entire framework. More sophisticated features may then be introduced in future work.

We now first give an overview of the entire processing chain. Then, features we utilize are explained, the basic theory of CRFs is described, and finally building detection results with different feature sets as input are compared.

* Corresponding author

2. PROCESSING CHAIN

In this section we provide an overview of the proposed processing chain (Fig. 1). It can roughly be subdivided into five steps: 1) line extraction, 2) projection of all lines to a reference coordinate system, 3) extraction of features, 4) training of the CRF parameters using ground truth, and 5) classification into building and non-building sites. The output is a label image showing building and non-building sites.
First, 3D lines are computed from the optical stereo images (section 3.2) and double-bounce lines are segmented in the InSAR data (section 3.3). Both line sets are then projected from the sensors' coordinate systems to the reference coordinate system of the orthophoto. Thereafter, a feature vector is computed for each site. In our case, an image site corresponds to a square image patch as traditionally used for both computer vision (e.g., Kumar and Hebert, 2003) and remote sensing applications of CRFs (e.g., Zhong and Wang, 2007). In addition, we adapt the idea of Kumar and Hebert (2006) and compute those features in three different scales. Then, the parameters of the CRF are trained on a subset of the data using ground truth. Subsequently, inference is conducted and the test data are classified into building sites and non-building sites (see CRF details in section 4).

3. FEATURES

Usually, high-resolution multi-spectral orthophotos are widely available and thus we take an orthophoto as the basic source of features for building detection. In order to assess the impact of height data on the building detection results of the CRF framework we also investigate optical stereo imagery. In very high-resolution aerial imagery characteristic objects of urban areas, particularly buildings, become visible in great detail (Fig. 2(a)). High-resolution SAR data provides complementary information. Double-bounce lines occurring at the position where the building wall meets the ground are characteristic features (Thiele et al., 2010). Fig. 3(a) compares the sensor geometries and the projected lines in ground geometry. Disregarding all projection artefacts, the double-bounce line of a flat-roofed building (with vertical walls) is located at the same position as the stereo line representing the roof edge (neglecting overhang). Note that the roof segment of the building in the orthophoto we use falls over double-bounce line and stereo line since we are not dealing with a true orthophoto (cf. Fig. 3(b,c)).

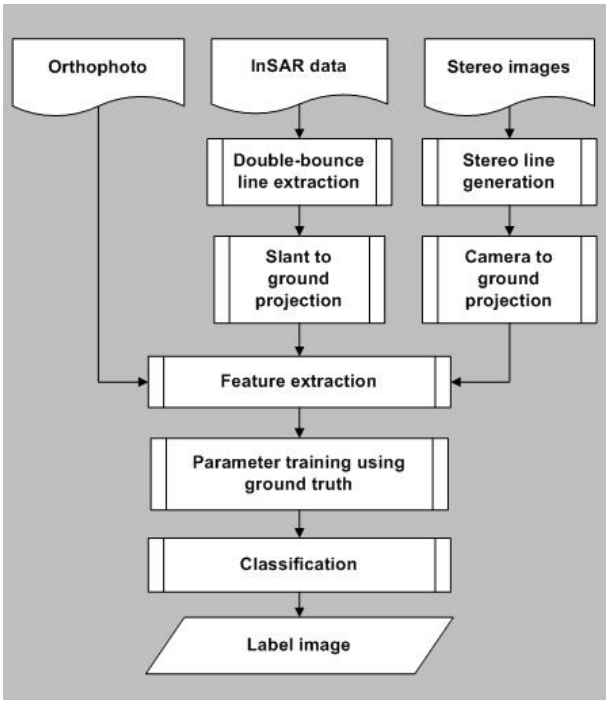


Figure 1. Flowchart of the processing chain for building detection

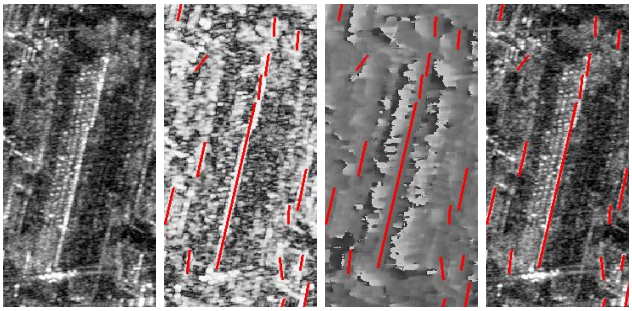
The focus of this research is neither on particularly sophisticated features nor on sophisticated feature selection techniques but on the overall suitability assessment of CRFs for building detection with multi-sensor data. Therefore, rather simple features are selected and feature selection is accomplished empirically.

3.1 Orthophoto features

We test various combinations of features (colour, intensity, and gradient) of the orthophoto within the CRF framework and choose those that provide the best results. The most suitable features are found based on colour, intensity, and gradient. As colour features we take mean and standard deviation of red and green channel normalized by the length of the RGB vector. Mean and standard deviation of the hue channel are found to be discriminative, too. Furthermore, variance and skewness of the gradient orientation histogram of a patch proved to be good features. The images are subdivided into square image patches and features are calculated within each patch. Of course, the choice of patch size is a trade-off. A small patch size is desirable in order to detect buildings in detail. However, too small patches lead to instable features resulting in less reliable estimates of the probability density distributions. We apply a multi-scale approach to mitigate those shortcomings (Kumar and Hebert, 2006). Each feature is calculated for different patch sizes and all scales are integrated into the same feature vector. We follow this approach and test various numbers of scales and scale combinations. Three different scales (10x10, 15x15, and 20x20 pixels) are found to provide good results. Features of large patches integrate over bigger areas thus excluding, for example forests or agricultural areas whereas the small patches provide details.

3.2 Stereo lines

We extract 3D lines from a pair of aerial images using the pair-wise line matching approach proposed by Ok et al. (2010). At this point we only briefly summarize the algorithm and refer the reader to the reference for further details. The entire algorithm consists of four main steps: pre-processing, straight line extraction, stereo matching of line pairs, and post-processing.
Pre-processing contains smoothing with a multi-level non-linear colour diffusion filter and colour boosting in order to exaggerate colour differences in each image. Next, straight lines are extracted in each of the stereo images. A colour Canny edge detector is applied to the pre-processed images. Thereafter, straight edge segments are extracted from the edge images using principal component analysis followed by random sampling consensus. Subsequently, a new pair-wise stereo line matching technique is applied to establish the line to line correspondences between the stereo images. The pair matches are assigned after a weighted matching similarity score, which is computed over a total of eight measures.



a b c d

Figure 2. (a) flat-roofed building signature in magnitude data of InSAR pair (range from left to right), extracted double-bounce lines overlaid to (b), the coherence image, (c) the interferometric heights, (d) the magnitude image.

A post-processing step is accomplished in order to reduce the number of mismatches, which occur due to multiple matches of individual lines. Finally, the stereo line segments are reconstructed exploiting the intersection of the stereo image rays. The stereo lines overlaid to a small part of the orthophoto are shown in Fig. 3(b) and a sketch showing the mapping geometries is given in Fig. 3(a). It can be seen that most of the stereo lines are located along the boundaries of the roofs particularly in case of flat roofs (see building A). In case of gable roofs some parts of the roof ridges are also extracted (see building group B).

In order to derive meaningful and consistent features we normalize the heights of the stereo lines. First, the local ground height is determined for each training and test image (of 310 m x 310 m size) assuming locally flat terrain. This assumption can readily be made because the test area is relatively flat. Second, the individual ground height of each image was subtracted from the heights of the stereo lines. Then, based on the assumption that the minimum building height is three meters, all stereo lines below this threshold are discarded. Then, we simply check if an image patch intersects with a line. In case it does the patch value is set to one and all other patches are set to zero (Fig. 4(c)). We compute this feature in all three scales.

3.3 InSAR features

Buildings in InSAR data appear differently compared to optical data due to the active illumination, the different wavelength, the side-looking viewing geometry, and the distance measurement. Furthermore, relevant building features occur in both magnitude and in phase data. An example is given in Fig. 2. It shows a typical magnitude signature of a flat-roofed building in (a) dominated by layover, double-bounce scattering, and shadow. A more in detail explanation considering different building types and illumination directions is provided in Thiele et al. (2010). Focusing on the coherence (b) and interferometric height data (c), especially the double-bounce line shows characteristic distributions. The high coherence value indicates high signal-to-noise-ratio in the InSAR data of this region. Furthermore, the interferometric height distribution at this line enables to discriminate between building lines and bright lines due to other effects. This double-bounce line is part of the building footprint, which is shown in Fig. 3(a). All these attributes make the double-bounce lines the most reliable building feature in urban areas and thus we extract features based on them.

First, those double-bounce lines are extracted as proposed by Wegner et al. (2009) based on the magnitude image, the coherence, and the InSAR heights in slant range. Those lines (given in (b), (c), and (d)) are projected from slant to ground projection using the local mean interferometric height at the line position. A schematic comparison of the extracted building hints of orthophoto, stereo images, and InSAR data is given in Fig. 3(a). In Fig. 3(c) the double-bounce lines of a flat-roofed (A) and a gable-roofed (B) building are superimposed to a small part of the orthophoto.

Again, double-bounce lines may not be introduced in vector format directly since we deal with image patches. Thus, we apply a segmentation to the orthophoto and overlay segments and double-bounce lines. All intersecting segments are set to one, all others to zero. Finally, a distance map is generated and minimum and maximum values within each patch are computed. This feature is only generated for the highest resolution (i.e., the smallest patch size) (Fig. 4(d)).

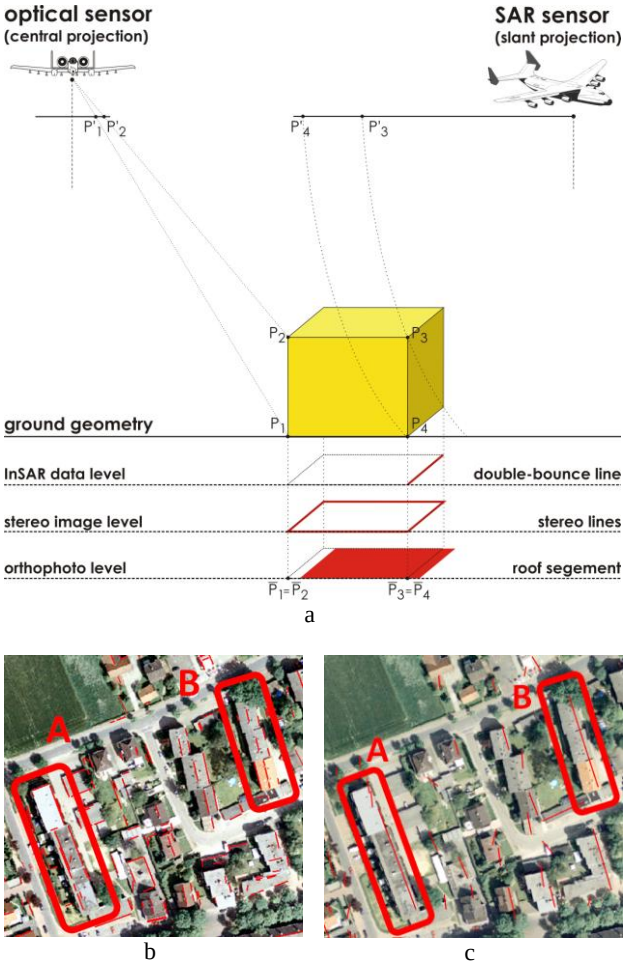


Figure 3. (a) Geometries of orthophoto, optical stereo images, and InSAR, (b) Buildings in orthophoto with flat roofs (A) and gable roofs (B) overlaid with 3D stereo lines, (c) same region as (b) overlaid with InSAR double-bounce lines

4. CONDITIONAL RANDOM FIELDS

High-resolution optical and InSAR data provide detailed information of urban area objects (see Fig. 2(a) and 2(e)). Single trees, gardens, and streets are mapped. Those objects, their typical spatial distribution and interrelations with buildings can be exploited in order to improve classification through context integration.

Conditional Random Fields, similar to Markov Random Fields (MRF), provide the possibility to integrate this context know-

ledge into a probabilistic classification framework. They belong to the family of graphical models and thus facilitate the use of well investigated learning and inference techniques. We use CRFs instead of MRFs because they allow integrating observations $\boldsymbol{x}$ and comparisons of labels $\boldsymbol{y}$ globally across the entire image as well as the use of observations within the prior term. Furthermore, the conditional independence assumption between features can be relaxed. Those properties make them a very flexible technique for context-based classification.

CRFs are discriminative models and thus model the posterior probabilities $P(\boldsymbol{y}|\boldsymbol{x})$ of labels $\boldsymbol{y}$ conditioned on observations $\boldsymbol{x}$ directly (Eq. 1) (unlike MRFs, which model the joint probability $P(\boldsymbol{x},\boldsymbol{y})$). We deal with a simple binary classification task and thus we only have two different labels $\boldsymbol{y}$, building and non-building. The set of all observations is denoted as observation vector $\boldsymbol{x}$, the label of the patch i that is currently investigated is denoted y_i, and its adjacent label it is compared to is denoted y_j. The set of all patches i to be labeled is S and the set of all patches j in the neighborhood of patch i is N_i (which naturally is a subset of S). $Z(\boldsymbol{x})$ is called the partition function (Eq. 2). It is a normalization constant (for a given data set) and transforms the sum of potentials to probabilities $P(\boldsymbol{y}|\boldsymbol{x})$.

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

where

$$Z(\boldsymbol{x}) = \sum_{\boldsymbol{y}}\exp\left(\sum_{i\in S} A_i(\boldsymbol{x},y_i)+\sum_{i\in S}\sum_{j\in N_i} I_{ij}(\boldsymbol{x},y_i,y_j)\right) \quad (2)$$

CRFs basically consist of two main terms (Lafferty et al., 2001), the association potential $A_i(\boldsymbol{x},y_i)$ and the interaction potential $I_{ij}(\boldsymbol{x},y_i,y_j)$. We use a standard approach for both similar to the one proposed by Kumar and Hebert (2006) in order to evaluate its performance for building detection. We use a generalized linear model for $A_i(\boldsymbol{x},y_i)$ (Eq. 3). However, various other classifiers, for example Maximum Likelihood or Logistic Regression could equally be used. The association potential $A_i(\boldsymbol{x},y_i)$ determines the most likely label y_i of a single patch (i.e., node) i considering all observations $\boldsymbol{x}$.

$$A_i(\boldsymbol{x},y_i) = \exp\left(y_i\boldsymbol{w}^T\boldsymbol{h}_i(\boldsymbol{x})\right) \quad (3)$$

Thus, all observations of the entire data set could potentially be used to label a single patch. In order to limit complexity we do not use all feature vectors but only a single feature vector $\boldsymbol{h}_i(\boldsymbol{x})$ for each patch i containing the features of three different scales described in Section 3. Vector $\boldsymbol{w}^T$ contains weights of features in $\boldsymbol{h}_i(\boldsymbol{x})$ that are adjusted during training. In order to generate a more accurate non-linear decision surface a quadratic expansion of $\boldsymbol{h}_i(\boldsymbol{x})$ is done (p.191, Kumar and Hebert, 2006). Thereafter, $\boldsymbol{h}_i(\boldsymbol{x})$ contains all features as described in section 3, their squares, and their pair-wise products.

$$I_{ij}(\boldsymbol{x},y_i,y_j) = \exp\left(y_i y_j\boldsymbol{v}^T\boldsymbol{\mu}_{\boldsymbol{ij}}(\boldsymbol{x})\right) \quad (4)$$

The interaction potential $I_{ij}(\boldsymbol{x},y_i,y_j)$ (Eq. 4) basically is a smoothing term comparing adjacent labels y_i and y_j that are either suppressed or supported by features $\boldsymbol{\mu}_{ij}(\boldsymbol{x})$. Those edge features $\boldsymbol{\mu}_{ij}(\boldsymbol{x})$ again could possibly be based on all observations globally. We simply define $\boldsymbol{\mu}_{ij}(\boldsymbol{x})$ as the difference $\boldsymbol{\mu}_{ij}(\boldsymbol{x}) = \boldsymbol{h}_i(\boldsymbol{x}) - \boldsymbol{h}_j(\boldsymbol{x})$ of the expanded single patch feature vector of the current node $\boldsymbol{h}_i(\boldsymbol{x})$ and its neighboring nodes $\boldsymbol{h}_j(\boldsymbol{x})$ within a 4-connectivity neighborhood.

We tested various training and inference methods and found the Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) (Liu and Nocedal, 1989) method and Loopy Belief Propagation (LBP) (Frey and MacKay, 1998) to deliver the best results for training and inference, respectively.

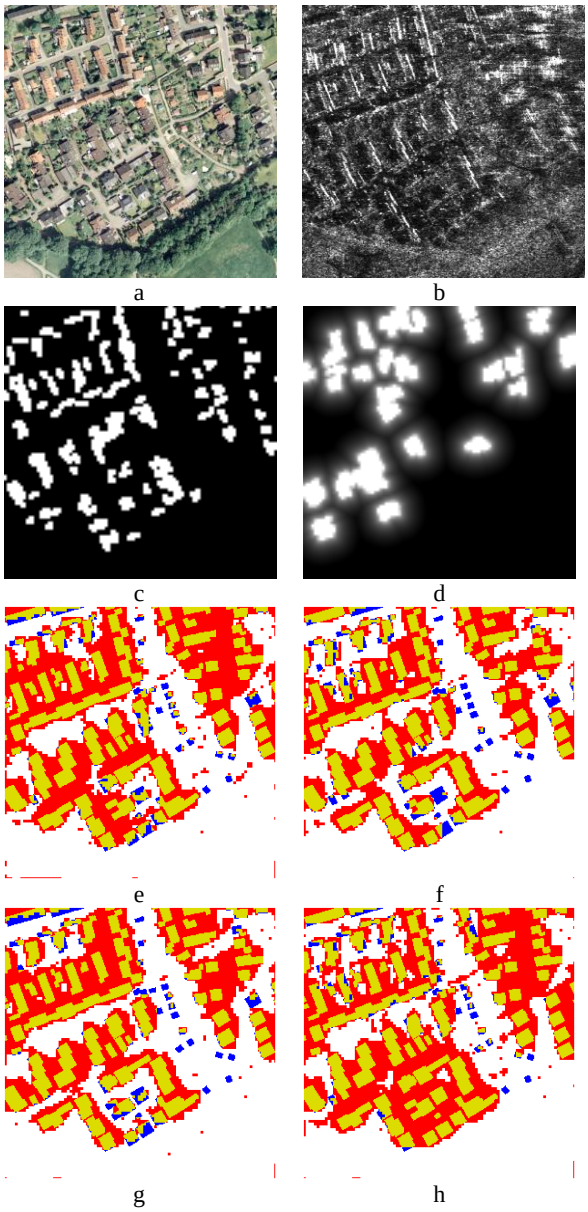


Figure 4. CRF classification results of one out of four test scenes (a) Orthophoto, (b) SAR amplitude image, (c) stereo line patches, (d) double-bounce line segments; true positive (orange), false positive (red), true negative (white), and false negative (blue) building detection results based on features of (e) the orthophoto, (f) orthophoto, 3D stereo lines, and InSAR, (g) orthophoto and 3D stereo lines, (h) orthophoto and InSAR

5. RESULTS

Our test data set consists of one orthophoto, an optical aerial image stereo pair (© Geobasisdaten: Land NRW, Bonn, 2111/2009), and one mono-aspect InSAR image pair of the city Dorsten, Germany. The orthophoto was acquired with the analogue aerial camera Zeiss RMK and scanned whereas the two stereo images were taken with the digital aerial camera Z/I Imaging DMC. The single-pass X-band InSAR data (wave length λ = 3.14 cm) were acquired by the AeS sensor of Intermap Technologies (Schwaebisch and Moreira, 1999). Spatial data resolution of the original single-look data is 38.5 cm in range and 18 cm in azimuth with a baseline of 2.4 m. Since the different test data were not acquired exactly at the same time we selected smaller blocks of 1000 x 1000 pixels size without significant changes between acquisitions.
In order to assess the quality of our results, they are compared to reference data, and the completeness and the correctness are determined on a per-pixel level. These numbers give a balanced estimate of the area that is classified correctly. We also determine the completeness of the results on a per-building level, using the method based on the area overlap as described in (Rutzinger et al., 2009). In this context, a building is considered to be a true positive if 70% of its area is covered by a building in the reference. The correctness of the results is not determined on a per-building label, because in our results most of the buildings are merged into a few large building segments, which makes a meaningful interpretation of the correctness impossible.

5.1 Orthophoto versus multi-sensor feature combination

We first compare CRF building detection results achieved with merely the orthophoto (Fig. 4(e)) to those based on all available features described in section three (Fig. 4(f)). Thus, we may empirically assess the improvements due to InSAR double-bounce lines and 3D stereo lines. Table 1 gives the average μ and the standard deviation σ of both completeness and correctness of this first test on pixel-level. The completeness on a per-building-level is shown in Table 2.

Orthophoto				Orthophoto+Stereo+InSAR			
Completeness		Correctness		Completeness		Correctness	
μ	σ	μ	σ	μ	σ	μ	σ
85%	5%	71%	7%	88%	5%	76%	7%

Table 1. Completeness and correctness on a per-pixel level of the CRF building extraction results using only orthophoto features vs. the combination of orthophoto, stereo line, and InSAR features. μ and σ are the mean and standard deviation of the results from four test scenes.

Orthophoto		Orthophoto+Stereo+InSAR	
μ	σ	μ	σ
85%	10%	81%	13%

Table 2. Completeness on a per-building level of the CRF building extraction results using only orthophoto features vs. the combination of orthophoto, stereo line, and InSAR features. μ and σ are the mean and standard deviation of the results from four test scenes.

On pixel-level, we achieve 85% correctly classified building pixels using the features generated from the orthophoto. However, the correctness (71%) is very low because small gaps between buildings are misclassified. This effect occurs in all four tests (see red areas in Fig. 4(e)-(h)) because of the simple standard interaction potential, which is basically a smoothing term. A combination of orthophoto features with stereo and InSAR helps increasing both completeness (88%) and Correctness (76%). Nonetheless, the strong smoothing effect caused by the smoothing effect of the interaction potential is still present.

5.2 Stereo lines versus InSAR lines

Secondly, we evaluate the impact of stereo lines and InSAR double-bounce lines separately on the overall CRF building detection performance. Results based on orthophoto features and stereo lines (Fig. 4(g)) are compared to those combining orthophoto features with InSAR double-bounce lines (Fig. 4h). Tables 3 and 4 summarize the evaluation on pixel-level and on object-level, respectively.

Orthophoto+Stereo				Orthophoto+InSAR			
Completeness		Correctness		Completeness		Correctness	
μ	σ	μ	σ	μ	σ	μ	σ
87%	6%	74%	7%	88%	6%	70%	10%

Table 3. Completeness and correctness on a per-pixel level of the CRF building extraction results using orthophoto features plus stereo lines vs. the combination of orthophoto and InSAR features. μ and □ are the mean and standard deviation of the results from four test sites.

Orthophoto+Stereo TPR		Orthophoto+InSAR TPR	
μ	σ	μ	σ
79%	9%	81%	9%

Table 4. Completeness on a per-building level of the CRF building extraction results using orthophoto features plus stereo lines vs. the combination of orthophoto and InSAR features. μ and σ are the mean and standard deviation of the results from four test sites.

The combination of the orthophoto features with the stereo lines increases the pixel-based correctness (74%) compared to the combination with InSAR double-bounce lines (70%) whereas the completeness is on the same level (87% vs. 88%).
Comparing the completeness on a per-building level given in Tables 2 and 4, the best is achieved using only orthophoto features because of over-smoothing. This is due to the reasons that in all other cases very small buildings are missed if neither InSAR lines nor stereo lines occur. They are strong features and thus gain high weights during CRF training. Nonetheless, we have seen in the pixel-based error analysis that those additional features increase the correctness significantly.

6. CONCLUSION AND OUTLOOK

In this work, first building detection results from combined features of an orthophoto, optical stereo images, and InSAR data using Conditional Random Fields was presented. CRFs proved to be a suitable technique for context-based classification. The introduction of very simple features derived from stereo lines and InSAR double-bounce lines helped increasing completeness and correctness on per-pixel level

although only slightly. This is due to, first, the very simple features derived from stereo and InSAR data and, second, the standard approach of the interaction potential, which basically is a smoothing term. Context is only modelled implicitly by either supporting or suppressing the label comparison $y_i y_j$ with the observations. This method works well if large single objects occur in an image as for instance shown by Kumar and Hebert (2006) and Korč and Förstner (2008). Our task of building detection in urban areas shows a different characteristic. Many relatively small objects are distributed over a large part of the scene with sometimes very small gaps between them. Therefore, our next step will be the introduction of an explicit discontinuity constraint similar to the one proposed by Tupin and Roux (2005). High gradients, double-bounce lines, and stereo lines at roof edges located between two patches could possibly be a hint for discontinuities.

Nonetheless, those discontinuity constraints and the context of the scene may only be exploited to their full extent if we also replace the regular patch grid by an irregular segmentation. We are currently working on setting up the CRF graph on irregularly distributed segments obtained with Normalized Cuts. In the long term we will also have to fine tune the features we use in order to optimize building detection results.

REFERENCES

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.

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.

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.

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.

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.

Kumar S., and Hebert, M. 2006. Discriminative Random Fields. International Journal of Computer Vision, Vol. 68, No. 2, pp. 179-201.

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.

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.

Ok, A. O., Wegner, J.D., Heipke, C., Rottensteiner, F., Soergel, U., and Toprak, V. 2010. A new straight line reconstruction methodology from mutli-spectral stereo aerial images. submitted to Photogrammetric computer vision and image analysis conference (PCV'10), accepted.

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.

Rabinovich, A., Vedaldi, A., Galleguillos, C., Wiewiora, E., and Belongie, S. 2007. Objects in Context. In: *Proceedings of IEEE International Conference on Computer Vision (ICCV'07)*, 2007, 8 p..

Rutzinger, M., Rottensteiner, F., Pfeifer, N., 2009. A comparison of evaluation techniques for building extraction from airborne laser scanning. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Vol. 2, No. 1, pp. 11-20.

Schwaebisch, M., and Moreira, J. 1999. The High Resolution Airborne Interferometric SAR AeS-1. In: *Proceedings of the Fourth International Airborne Remote Sensing Conference and Exhibition*, pp. 540-547.

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.

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.

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

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.

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, pp. 169-174.

Xiao, R., Lesher, 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.

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.