BUILDING DETECTION AND HEIGHT ESTIMATION FROM HIGH-RESOLUTION INSAR AND OPTICAL DATA

Jan Dirk Wegner and Uwe Soergel

Institute of Photogrammetry and GeoInformation (IPI), Leibniz University Hannover, Germany

1. INTRODUCTION

State-of-the-art satellite SAR sensors acquire data of one meter geometric ground resolution, airborne sensors achieve even higher resolutions. Nonetheless, layover and occlusion hamper interpretability of such data particularly in urban scenes. In order to overcome this drawback, SAR data may be in analyzed in combination with optical data in urban areas [7][8][9]. We use additional information from aerial photos to detect buildings. Features are extracted from both data sets and introduced to a common feature vector followed by a classification into building sites and non-building sites. We use a Conditional Random Field (CRF) [10] framework for classification, which provides high modeling flexibility by both relaxing the conditional independence assumption and incorporating global context information (as compared to Markov Random Fields (MRF)). Furthermore we show that the different sensor geometries of the SAR and the optical sensor may be used to estimate building heights (once buildings are detected) under the assumption of locally flat terrain. Such height estimate may be helpful to resolve phase unwrapping problems and model urban scenes three-dimensionally. The proposed methods are tested using an aerial orthophoto and airborne InSAR data acquired by the Intermap Aes-1 sensor [15] of the city of Dorsten, Germany.

2. FEATURES

We take very simple features for the building detection task because our focus is on the overall suitability of CRFs as a method for building detection. Mean and variance of the red channel, the blue channel, and the hue are found to be descriptive color features. Additional features are calculated based on the gradient orientation histogram of the intensity image as already used for building detection in terrestrial images by Kumar and Hebert [11]. We use mean, variance, the maximum value, and the sinusoidal difference between the two highest values (i.e., detection of right angles) as features. The analysis of the InSAR data is focused on the detection of stable building features and their recognition. The most reliable feature is the building corner [5]. Furthermore, the InSAR phase distribution along the corner lines facilitates the separation of these lines from other bright lines in the scene (see details in [12]).

3. CONDITIONAL RANDOM FIELDS

Like all graphical models, CRFs (see Eq. 1) have the advantage of assigning probabilities to the final labeling instead of only providing final decisions (like for instance Support Vector Machines). In contrast to MRFs, CRFs are discriminative models and therefore model only the posterior distribution $P(\mathbf{y}|\mathbf{x})$ of the labels $\mathbf{y}$ given data $\mathbf{x}$.

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

The association potential $A_i(\mathbf{x},y_i)$ measures how likely a label site i is labeled with y_i given data $\mathbf{x}$, while the interaction potential $I_{ij}(\mathbf{x},y_i,y_j)$ describes how two label sites i and j interact. Both the association potential and the interaction potential are defined over all data, the entire orthophoto and all InSAR data in our case. Hence, we may introduce both local and global context knowledge, which is a major advantage concerning automatic analysis of high-resolution remote sensing data of urban areas. In order to obtain a posterior probability $P(\mathbf{y}|\mathbf{x})$ of labels $\mathbf{y}$ conditioned on data $\mathbf{x}$ the exponential of the sum of association potential and interaction potential is normalized by division through the partition function $Z(\mathbf{x})$, which is a constant for a given data set.

Our modeling of the association potential $A_i(\mathbf{x},y_i)$ and the interaction potential $I_{ij}(\mathbf{x},y_i,y_j)$ is closely related to the approach proposed by Kumar and Hebert in [11]. For our binary classification task of distinguishing building sites from non-building sites, labels y_i may either become 1 or -1, respectively. The association potential $A_i(\mathbf{x},y_i)$ measures how likely it is that a site i takes label y_i given all data $\mathbf{x}$ (see Eq. 2). We use logistic regression to distinguish building and non-building sites in the association potential.

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

We deal with a binary classification task and hence we use the sigmoid function $\sigma(t) = 1/(1+e\text{-}t)$ as activation function σ. Vector $\mathbf{h_i(x)}$ contains all node features whereas vector $\mathbf{w^T}$ contains the weights of the features in $\mathbf{h_i(x)}$ that are tuned during the training process.

The interaction potential $I_{ij}(\mathbf{x},y_i,y_j)$ determines how two sites i and j should interact regarding all data $\mathbf{x}$ (see Eq. 3). In our case, feature vector $\boldsymbol{\mu}_{\mathbf{ij}}(\mathbf{x})$ is simply calculated by subtracting the single scale feature vector from site j from such of the site i of interest $\boldsymbol{\mu}_{\mathbf{ij}}(\mathbf{x}) = \mathbf{h_i(x)}\text{-}\mathbf{h_j(x)}$. Vector $\mathbf{v^T}$ contains the weights of the features, which are adjusted during the training process. y_i is the label of the site of interest and y_j the label it is compared to. Unlike clique potentials in MRFs, label y_j does not necessarily have to be a label of a site j in the local neighborhood of y_i.

$$I_{ij}\left(\boldsymbol{x}, y_i, y_j\right) = y_i y_j \boldsymbol{v}^T \boldsymbol{\mu}_{ij}\left(x\right) \tag{3}$$

Only single scale features are used for the interaction potential and no quadratic expansion of the feature vector is currently done. The comparison of labels y_i and y_j follows the Ising model $\beta y_i y_j$. With $\beta = 1$, the product $y_i y_j$ becomes -1 if labels y_i and y_j do not belong to the same class whereas their product is 1 in case both labels are equal. We use the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) method as optimizer to train the association potential and the interaction potential simultaneously. For inference we use Loopy Belief Propagation (LBF).

4. HEIGHT ESTIMATION

Various possibilities of estimating building heights based on SAR [4][9] and InSAR [1][2][3][6] data exist. However, to our knowledge, only very few attempts have been made to exploit the different sensor geometries of optical and SAR sensors for building height estimation. As shown in Fig. 1, building top P is imaged to point PS in the SAR image whereas it is mapped to point PO in the optical image. Basic concepts to exploit this effect for building height estimation are developed in [16] and further extended in [14]. Due to the central projection of the optical aerial sensor, buildings are projected away from the sensor's nadir point (see Fig. 1b). The corner line extracted from the InSAR data, however, is located exactly at the location where the building walls meet the ground (corresponding to point P' in Fig. 1a). Hence, the roof of the building falls over the corner line (Fig. 1c). A rough height estimate can simply be determined with $h = d / \tan \theta_2$ where d corresponds to the distance between P' and PO measured in the image. The further away from the nadir of the optical sensor and the higher the building, the more accurate may the building height be estimated. In our paper we will present first height estimation results for flat roofed buildings and compare them to ground truth derived from LIDAR data.

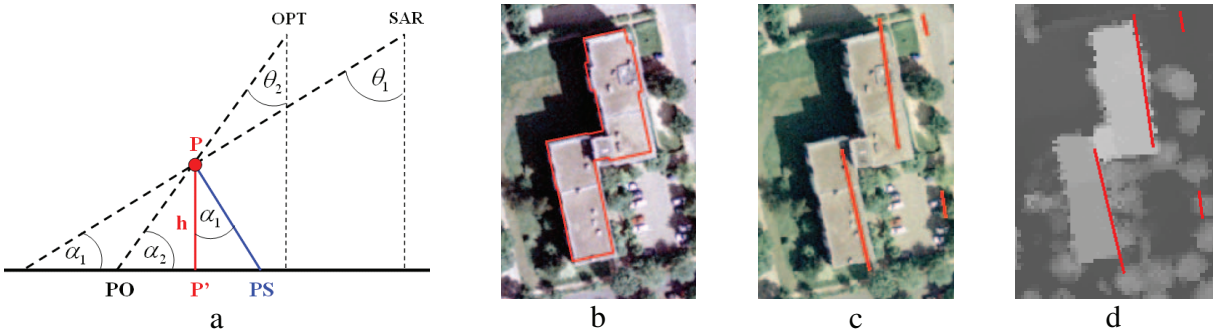

Figure 1. Comparison of SAR and optical viewing geometry under the assumption of locally flat terrain (a); optical data (b) overlaid with cadastral building footprint; optical data (e) and LIDAR data (d) overlaid with detected building corner

[1] P. Gamba, B. Houshmand, and M. Saccani, “Detection and extraction of buildings from interferometric SAR data,” IEEE Trans. Geoscience and Remote Sensing, vol. 38, no. 1, part 2, 2000, pp. 611–617.
[2] U. Soergel, U. Thoennessen, and U. Stilla, “Reconstruction of Buildings from Interferometric SAR Data of built-up Areas”, Proc. of PIA, International Archives of Photogrammetry and Remote Sensing, vol. 34, part 3/W8, 2003, pp. 59-64.
[3] C. Tison, F. Tupin, and H. Maitre, “A Fusion Scheme for Joint Retrieval of Urban Height Map and Classification From High-Resolution Interferometric SAR Images”, IEEE Trans. Geoscience and Remote Sensing, vol. 45, no. 2, 2007, pp. 496-505.
[4] F. Xu and Y.-Q. Jin, “Automatic Reconstruction of Building Objects From Multiaspect Meter-Resolution SAR Images”, IEEE Trans. Geoscience and Remote Sensing, vol. 45, no.7, 2007, pp. 2336-2353.
[5] A. Thiele, E. Cadario, K. Schulz, U. Thoennessen, and U. Soergel, “Building Recognition From Multi-Aspect High-resolution InSAR Data in Urban Areas”, IEEE Trans. on Geoscience and Remote Sensing, vol. 45, no. 11, 2007, pp. 3583-3593.
[6] R. Bolter, “Buildings from SAR: detection and reconstruction of buildings from multiple view high resolution interferometric SAR data”, PhD thesis, University of Graz, Austria, 2001.
[7] U. Soergel, E. Cadario, A. Thiele, and U. Thoennessen, “Feature Extraction and Visualization of Bridges over Water from high-resolution InSAR Data and one Orthophoto”, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 1, no.2, 2008, pp. 147-153.
[8] F. Tupin and M. Roux, “Detection of building outlines based on the fusion of SAR and optical features”, ISPRS Journal of Photogrammetry and Remote Sensing, vol. 58, 2003, pp. 71-82.
[9] F. Tupin and M. Roux, “Markov Random Field on Region Adjacency Graph for the Fusion of SAR and Optical Data in Radargrammetric Applications”, IEEE Trans. Geoscience and Remote Sensing, vol. 43, no. 8, 2005, pp. 1920-1928.
[10] J. Lafferty, A. McCallum, and F. Pereira, “Conditional Random Fields: Probabilistic Models for segmenting and labeling sequence data”, in *Proc. Int. Conf. on Machine Learning,* 2001.
[11] S. Kumar and M. Hebert, “Discriminative Random Fields: A Discriminative Framework for Contextual Interaction in Classification”, in *Proc. IEEE Int. Conf. on Computer Vision*, 2003, vol. 2, pp. 1150-1157.
[12] J.D. Wegner, A. Thiele, and U. Soergel, “Fusion of optical and InSAR features for building recognition in urban areas”, IntArchPhRS, vol. 38, part 3/W4, 2009, pp. 169-174.
[13] J.D. Wegner and U. Soergel, “Bridge height estimation from combined high-resolution optical and SAR imagery”, IntArchPhRS, vol. 37, part 7/3, 2008, pp. 1071-1076.
[14] M. Schwaebisch and J. Moreira, “The High Resolution Airborne Interferometric SAR AeS-1", in *Proc. Fourth Int. Airborne Remote Sensing Conf. and Exhibition*, 1999, pp. 540-547.
[15] J. Inglada and A. Giros, “On the possibility of Automatic Multisensor Image Registration”, IEEE Trans. Geoscience and Remote Sensing, 42(10), pp. 2104-2120.