Recognition of building features in high-resolution SAR and optical imagery

Jan Dirk Wegner, Alexander Schunert & Uwe Sörgel

Institute of Photogrammetry and GeoInformation
Leibniz Universität Hannover
email: {wegner, schunert, soergel}@ipi.uni-hannover.de

Abstract

In data of modern high-resolution SAR sensors like TerraSAR-X (TSX) man-made objects become visible in high detail. However, layover, occlusions, and interfering backscatter of different objects within the same resolution cell complicate automatic analysis particularly in urban areas. Two possibilities to facilitate interpretability are the use of additional data sources and time series of SAR images. We will present the current status of two research projects concerning those possibilities. First, we show building detection results achieved with a combination of SAR and optical aerial images. Features are extracted and analysed in a Conditional Random Field (CRF) framework exploiting context-knowledge. Second, the persistent scatterer (PS) interferometric SAR (InSAR) technique is applied to discover building deformations in a TSX time series. Those PS are grouped and interpreted using a 3D city-model.

Keywords: Building detection, Conditional Random Fields, Persistent Scatterer, Data fusion.

1 Introduction

In this paper we deal with the idea of introducing some kind of prior knowledge (or context information) in order to automatically interpret high-resolution SAR data. Furhermore, we introduce additional data to facilitate the recognition and analysis of characteristic patterns. First, Conditional Random Fields (CRF), introduced by Lafferty et al. (2001) and adapted to imagery by Kumar and Hebert (2003), are applied to combine high-resolution airborne InSAR data with an orthophoto for building detection. Second, persistent scatterers (PS) are estimated from time series of TSX images of the city of Berlin. In order to more robustly estimate deformations, the PS are grouped and assigned to particular buildings with the aid of a 3D city model. In the next section an overview of the CRF approach is given followed by the PS analysis supported by a 3D city model.

2 Conditional Random Fields

The basic idea is to extract characteristic features for buildings in both optical and InSAR data, to insert both feature sets into a single feature vector, and to finally classify the data based on this feature vector using a CRF. CRFs are graphical models and thus provide probabilities of the final labeling instead of just decisions. Those probabilities are very useful for post-processing or decision making. Moreover, they are undirected graphical models (i.e., random fields) as opposed to, for example Bayesian nets, which are directed graphical models. Their main advantage is that they do not suffer from the label bias problem, which states that labels with fewer successors in the tree are preferred Lafferty et al. (2001). Furthermore, unlike Markov Random Fields (MRF) CRF are globally conditioned on all observations **x** making them highly flexible for context modelling. They are discriminative models and thus model only the posterior distribution $P(\mathbf{y}|\mathbf{x})$ of labels **y** given data **x** as opposed to MRF that model the joint distribution of data and labels. CRFs as introduced by Lafferty et al. (2001) are defined as follows (**x** contains all observations and **y** all labels):

> *Let $G = (N, E)$ be a graph such that $\boldsymbol{y} = (\boldsymbol{y}_v)_{v \in V}$, so that* **y** *is indexed by the vertices of G. Then $(\boldsymbol{x}, \boldsymbol{y})$ is a conditional random field in case, when conditioned on* **x***, the random variables $\boldsymbol{y}_v$ obey the Markov property with respect to the graph: $p(\boldsymbol{y}_v|\boldsymbol{x}, \boldsymbol{y}_w, w \neq v) = p(\boldsymbol{y}_v|\boldsymbol{x}, \boldsymbol{y}_w, w \sim v)$, where $w \sim v$ means that w and v are neighbors in G.*

The most common CRF approach is based on sufficient statistics of exponential functions:

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

The association potential $\mathbf{A}_i(\mathbf{x}, y_i)$ measures how likely it is that a site i takes label y_i given all data **x** (see Eq. 2). Data **x** in our case are the orthophoto and the InSAR data. We use a generalized linear model to distinguish building and non-building sites in the association potential.

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

Vector $\mathbf{h}_i(\mathbf{x})$ contains all node features. We take mean of red channel, green channel, hue (Fig. 1(b)) and saturation as orthophoto features. In addition, features based on the gradient orientation histogram are used. As InSAR features we extract bright double-bounce lines, overlay them with a segmentation of the orthophoto, and calculate distance maps of such segments (Wegner et al., 2009) (Fig. 1(d)). Vector $\mathbf{w}^T$ contains the weights of the features in $\mathbf{h}_i(\mathbf{x})$ that are tuned during the training process.

The interaction potential $\mathbf{I}_{ij}(\mathbf{x}, y_i, y_j)$ determines how two sites i and j should interact regarding all data **x** (see Eq. 3). In our case, feature vector $\mu_{ij}(\mathbf{x})$ is simply calculated by subtracting the single scale feature vector from site j from such of the site i of interest $\mu_{ij}(\mathbf{x}) = \mathbf{h}_i(\mathbf{x}) - \mathbf{h}_j(\mathbf{x})$. However, in general $\mu_{ij}(\mathbf{x})$ could also be chosen based on other features than such already used for the association potential and other methods of comparing the features are possible, too. 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.

$$I_{ij}(\mathbf{x}, y_i, y_j) = \exp\left(y_i y_j \mathbf{v}^T \mu_{ij}(\mathbf{x}) \right) \tag{3}$$

In order to obtain a posterior probability $P(\mathbf{y}|\mathbf{x})$ of labels **y** conditioned on data **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.

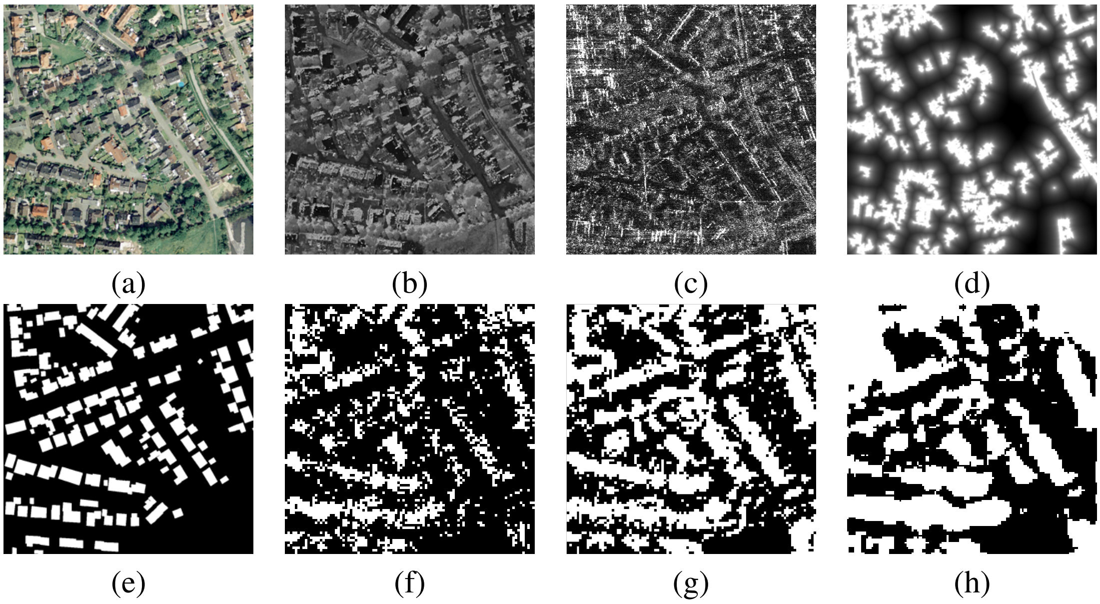

Figure 1: (a) One test region of the orthophoto, (b) corresponding hue image, (c) Aes-1 SAR image of the same region (©Intermap), (d) distance map of segments that overlap with SAR double-bounce lines, (e) building ground truth, and building detection results with Maximum Likelihood (f), Support Vector Machines (g), and Conditional Random Fields (h)

CRF version	DTR μ	DTR σ	FPR μ	FPR σ
Maximum Likelihood	61	6	13	6
Support Vector Machine	85	5	24	7
Conditional Random Field	85	7	27	8

Table 1: Mean μ and standard deviation σ of detection rate and false positive rate of first ML, SVM, and CRF experiments in percent

Table 1 shows preliminary results using the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) (Liu and Nocedal, 1989) approach for training and Loopy Belief Propagation (LBP) (Frey and MacKay, 1998) for inference. In order to evaluate the performance of the current CRF set-up we compare building detection results to those of a Maximum Likelihood classifier (ML), and Support Vector Machines (SVM) (see corresponding label images in Fig. 1 (f,g,h)). The CRF performs better than ML and delivers results on the same level as the SVM. In order to further improve the CRF performance compared to the SVM context has to be modelled in a more sophisticated way in the interaction potential.

3 Persistent Scatter InSAR

Persistent Scatterer InSAR (PSI) is an extension to the classical InSAR approach mitigating the effects of atmospheric disturbances and temporal decorrelation. This is achieved by the use of a stack of SAR images and the restriction of the processing to a set of radar targets referred to as Persistent Scatterers, which exhibit a coherent backscattering behaviour over time. The outcome of a PSI analysis is an estimate of the deformation and the height for each scatterer with respect to some reference point. The PSI algorithms of the first generation, which were mainly applied to stacks of ERS and ENVISAT images essentially applied signal processing techniques to discriminate between deformation and height signal on the one hand and nuisance terms like atmosphere on the other hand (Ferretti et al., 2000). Thereby only little knowledge about the relationship of the PS to each other was used, which was mainly due to the quite low spatial resolution of the ERS and ENVISAT data. With the advent of high resolution SAR sensors like TerraSAR-X, which provide a spatial resolution of up to one meter, this has fundamentally changed. While it is even hard to assign PS to single buildings in the ERS case, there are usually plenty of PS found for every building in the TerraSAR-X case, which can be exploited for the PSI analysis (Bamler et al., 2009). The main idea is to group the PS into reasonable clusters and to work with these clusters instead of working with the single PS. One way to cluster the PS is the use of building features like building outlines, which can be extracted from 3D city models. After the projection of these building features into the radar geometry a matching with the PS set is conceivable, which results in an assignment of PS to building features. Finally the relationship of the PS to each other can be inferred offering the opportunity to introduce constraints into the PSI analysis. An intermediate result for this procedure applied to a stack of TSX high resolution spotlight images of Berlin is shown in Figure 2. On the left side the 3D city model is depicted, which is taken from googleearth. The two buildings of interest are framed with a dashed and a solid rectangle respectively. On the right hand side the appropriate part of the SAR image is depicted together with the radar coded building features, which are illustrated in dashed and solid style respectively corresponding to the target buildings. The features used here are the approximate building outline and rows of windows.

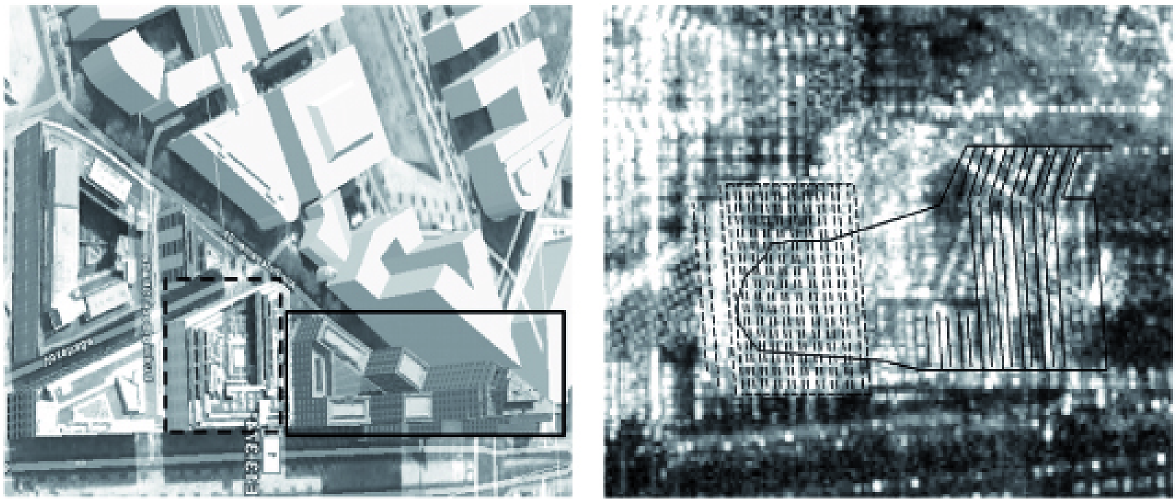

Figure 2: Left: 3D city model of a site in Berlin with two buildings of interest framed by the two rectangles. Right: TSX image of this site overlayed with the radar coded building features.

References

Frey, B.J. and MacKay, D.J.C.: 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, 1998, vol. 10, MIT Press.

Kumar, S. and Hebert, M.: 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.

Lafferty, J., McCallum, A. and Pereira, F.: Conditional Random Fields: Probabilistic Models for segmenting and labeling sequence data. in Proc. Int. Conf. on Machine Learning, 2001, 8 p..

Liu, D.C. and Nocedal, J.: On the limited memory BFGS for large scale optimization. Mathematical Programming, 1989, vol. 45, no. 1-3, pp. 503-528.

Wegner, J.D., Thiele, A. and Soergel, U.: 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.

Bamler R., Eineder, M., Adam, N., Zhu, X., Gernhardt, S.: Interferometric Potential of High Resolution Spaceborne SAR. In: Photogrammetrie - Fernerkundung - Geoinformation, no. 5, 2009, pp. 407-419

Ferretti, A., Prati, C., Rocca, F.: Permanent scatterers in SAR interferometry. In: IEEE Trans. Geosci. Remote Sens., vol. 39, no. 1, 2000, pp. 8-20.