SEMANTIC INTERPRETATION OF INSAR ESTIMATES USING OPTICAL IMAGES WITH APPLICATION TO URBAN INFRASTRUCTURE MONITORING

Yuanyuan Wang [a], Xiao Xiang Zhu [a, b, *]

[a] Helmholtz Young Investigators Group "SiPEO", Technische Universität München, Arcisstraße 21, 80333 Munich, Germany.
wang@bv.tum.de
[b] Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, 82234 Weßling, Germany.
xiao.zhu@dlr.de

Commission III, WG III/4

KEY WORDS: optical InSAR fusion, semantic classification, InSAR, SAR, railway monitoring, bridge monitoring

ABSTRACT:

Synthetic aperture radar interferometry (InSAR) has been an established method for long term large area monitoring. Since the launch of meter-resolution spaceborne SAR sensors, the InSAR community has shown that even individual buildings can be monitored in high level of detail. However, the current deformation analysis still remains at a primitive stage of pixel-wise motion parameter inversion and manual identification of the regions of interest. We are aiming at developing an automatic urban infrastructure monitoring approach by combining InSAR and the semantics derived from optical images, so that the deformation analysis can be done systematically in the semantic/object level. This paper explains how we transfer the semantic meaning derived from optical image to the InSAR point clouds, and hence different semantic classes in the InSAR point cloud can be automatically extracted and monitored. Examples on bridges and railway monitoring are demonstrated.

1. INTRODUCTION

1.1 Deformation Monitoring by SAR Interferometry

Long term deformation monitoring over large area is so far only achievable through differential SAR interferometry (InSAR) techniques such as persistent scatterer interferometry (PSI) (Adam et al., 2003; Ferretti, Prati & Rocca, 2001; Gernhardt & Bamler, 2012; Kampes, 2006) and differential SAR tomography (TomoSAR) (Fornaro et al., 2015; Fornaro, Reale & Serafino, 2009; Lombardini, 2005; Zhu & Bamler, 2010a; Zhu & Bamler, 2010b). Through modelling the interferometric phase of the scatterers in the SAR image, we are able to reconstruct their 3-D positions and the deformation histories.

The focus of development in differential InSAR techniques has always been on the estimation of the phase history parameters (elevation, motion parameters, etc.) under different scattering models including single deterministic scattering (persistent scatterer), distributed scattering (distributed scatterer), and layover of multiple scatterings (TomoSAR).

In regard of large area deformation monitoring, PSI is the workhorse among the InSAR methods. Distributed scatterer (DS)-based methods such as SqueeSAR and its alternatives (Ferretti et al., 2011; Jiang et al., 2015; Wang, Zhu & Bamler, 2012) have emerged in the last few years to complement the drawback of few PS in nonurban area, while TomoSAR has become the most competent method for urban area monitoring because of its capability of separating multiple scatterers in a resolution cell. With meter-resolution SAR data, it is also demonstrated that even individual building could be monitored in very high level of detail from space.

In summary, great progress has been made in inversion problems of the coherent signals from SAR data stacks. Millimetre-precision in the linear deformation rate can be achieved (Ferretti, Prati & Rocca, 2001; Bamler et al., 2009).

1.2 Motivation

The current D-InSAR methods are able to produce excellent deformation estimates. However, they are based on pixel-wise parameters inversion and manual identification of the region of interest. It lacks a systematic way to monitor the region of interest, for example, the railway or the road network in a city. Therefore, the next generation InSAR techniques in urban area should be aimed towards exploiting and understanding the regularities and semantics of the manmade world that is imaged.

With such vision in mind, we aim to bridge the InSAR and optical field by complementing InSAR's high precision deformation measurement with the high interpretability of optical images.

1.3 Methodology

This work is aimed towards a future generation of InSAR techniques that are contextually aware of the semantics in a SAR image. It enables the object-level deformation reconstruction and analysis from SAR images. The proposed approach brings the first such analysis via a semantic classification in the InSAR point cloud.

The general framework of the proposed approach is shown in Figure 1. The semantic classification of the InSAR point cloud is achieved by co-registering the InSAR point cloud and an optical image to a common reference 3-D model, so that the semantic classification in the optical image can be transferred to the InSAR point cloud. The general procedures are as follows.

[*] Corresponding author

This contribution has been peer-reviewed.
Editors: U. Stilla, F. Rottensteiner, and S. Hinz
doi:10.5194/isprsarchives-XL-3-W3-153-2015

a. Retrieve the 3-D positions of the scatterers from SAR image stacks. Since urban area is of our main interest, tomographic SAR inversion should be employed in order to resolve a substantial amount of layovered scatterers.
b. Absolute geo-reference the 3-D InSAR point cloud, due to the relative position of the InSAR point cloud w.r.t. a reference point. This step is achieved by co-registering the InSAR point cloud with a reference 3-D model.
c. Texturing the reference 3-D model with high resolution optical images, so that each SAR scatterer can be traced in the optical image.
d. Classify the optical image pixels based on its semantic meaning, e.g. geometry, material, and so on.
e. Perform further analysis on object-level in the InSAR point cloud based on their semantic class.

Since the fusion of the InSAR point cloud and the optical image is done in pure 3-D, which required strict 3-D reconstruction from both SAR images and the optical images, the work described in this paper is also different from many early research on SAR and optical image fusion, such as (Gamba & Houshmand, 2002; Wegner, Ziehn & Soergel, 2014; Wegner, Thiele & Soergel, 2009).

To summarize, the proposed method requires in addition only a stereo pair of optical images of the same area. The InSAR point cloud is co-registered with the 3-D point cloud derived from the optical image pair, and is therefore, co-registered to the classification derived from the optical images. Additionally, the user could also use reference model from different sources, e.g., LiDAR.

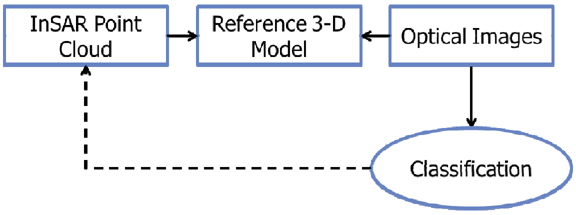


Figure 1. Flowchart of the proposed method. The semantic classification of the InSAR point cloud is achieved by co-registering the InSAR point cloud and the optical image to a reference model.

2. SAR TOMOGRAPHIC INVERSION

TomoSAR aims at separating multiple scatterers possibly layovered in the same pixel, and retrieving their third coordinates *elevation* in the SAR native coordinate system. Displacement of the scatterers can also be modeled and estimated, using stack of images acquired at different times. The technique is commonly known as differential SAR tomography (D-TomoSAR)

The D-TomoSAR processing was done by DLR's D-TomoSAR software Tomo-GENESIS. For an input data stack, Tomo-GENESIS retrieves the following information:

- the number of scatterers inside each pixel,
- the scattering amplitude and phase of each scatterer,
- and their 3D positions and motion parameters, e.g. linear deformation rate and amplitude of seasonal motion.

The scatterers' 3D positions in SAR coordinates are converted into a local Cartesian coordinate system, such as Universal Transverse Mercator (UTM), so that the results from multiple data stacks with different viewing angles can be combined. For our test area Berlin, two TerraSAR-X high resolution image stacks – one ascending orbit, the other descending orbit – are processed. These two point clouds are fused to a single one, following a feature-based matching algorithm which estimates and matches common building edges in the two point clouds (Wang & Zhu, 2015). Figure 2 is the fused point cloud which provides a complete monitoring over the whole city of Berlin.

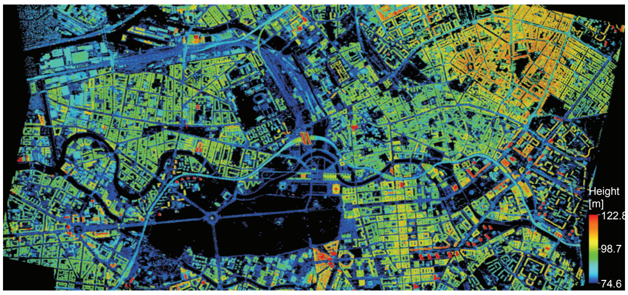


Figure 2. The fused TomoSAR point cloud of Berlin, which combines the result from an ascending stack and a descending stack. The height is color-coded.

3. GEOREFERENCE TOMOSAR POINT CLOUDS

Due to the relative position of the TomoSAR point cloud w.r.t. a reference point selected during the processing, it must be co-registered to a reference 3-D model in order to align with an optical image.

Our reference model is the 3-D point cloud derived from a pair of optical images by means of stereo matching. As a by-product, the optical images are also geo-localized in the optical point cloud. Other precise reference 3-D model such as LiDAR point cloud can also be used, with the additional effort of aligning it with the optical image.

3.1 Co-registration workflow

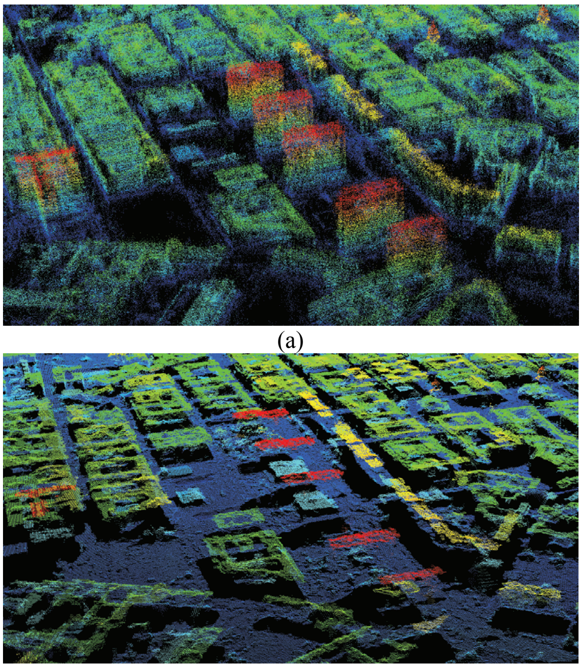
Since both the TomoSAR and optical point clouds are geo-coded into local coordinate systems, the co-registration problem is the estimation of translation between two rigid point clouds, subject to a certain tolerance on rotation and scaling. However, the optical point cloud is nadir-looking, in contrast to the side-looking geometry of SAR. In another word, façade point barely appears in optical point cloud while it is prominent in TomoSAR point cloud. This difference is exemplified in Figure 3, where the left and the right subfigures correspond to the TomoSAR and optical point clouds of the same area. The same conclusion should also apply to nadir-looking LiDAR point cloud. These unique modalities have driven our algorithm to be developed in the following way:

1 Edge extraction
  a. The optical point cloud is rasterized into a 2D height image.
  b. The point density of TomoSAR point cloud is estimated on the rasterized 2D grid.
  c. The edges in the optical height image and the TomoSAR point density image are detected.
2 Initial alignment
  a. Horizontally by cross-correlating the two edge images.

   b. Vertically by cross-correlating the height histogram of the two point clouds.

3 Refined solution

   a. The façade points in both point clouds are removed.
   b. The final solution is obtained using iterative closest point (ICP) applied on the two reduced point clouds.



(a)

(b)

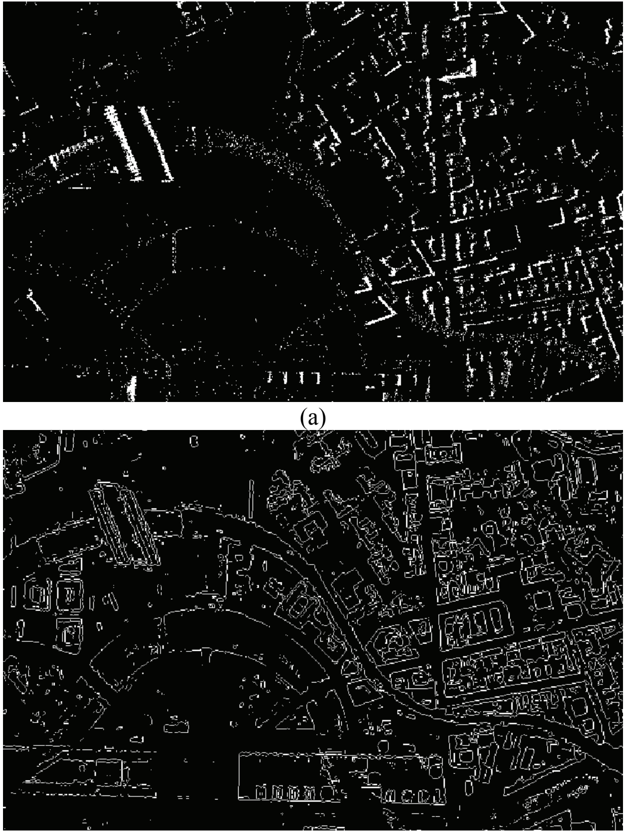
Figure 3. (a) TomoSAR point cloud of high-rise buildings, and (b) the optical point cloud of the same area. Building façades are almost invisible in the optical point cloud, while it is prominent in the TomoSAR point cloud.

3.2 2-D Edge Extraction

The 2-D edge images of the TomoSAR and optical point cloud are extracted from their rasterized height image, and point density image, respectively. Here we use 2×2 m for our dataset. For the optical point cloud, the mean height in each grid cell is computed, while for the TomoSAR point cloud, the number of points inside the grid cell is counted. The edges can be extracted from these two images using an edge detector, such as Sobel filter (Sobel, 1968). The thresholds in the edge detector are decided adaptively, so that the numbers of edge pixels in the two edge images are on the same order. Figure 4 is a close up view of the two edge images near downtown Berlin.

3.3 Initial Alignment

The initial alignment provides an initial solution to the iterative closest point (ICP) algorithm which is known to suffer from finding possibly a local minimum. The initial alignment consists of independently finding the horizontal and the vertical shifts. The horizontal shift is found by cross-correlating the edge images of the two point clouds. In most of the cases, a unique peak can be found, due to the complex, hence pseudorandom, structures of a city. Please see Figure 5 for the 2D correlation of two edge images, where a single prominent peak is found. The vertical shift is found by cross-correlating the height histogram of the two point clouds, which is shown in Figure 6. We also set the bin spacing of the height histograms to be 2m in our experiment. The accuracy of the shift estimates are of course limited by the discretization in the three directions. However, this is sufficient for the final estimation.



(a)

(b)

Figure 4. (a) A crop of the edge image of the TomoSAR point cloud in downtown Berlin, and (b) the edge image of the reference optical point cloud roughly at the same area.

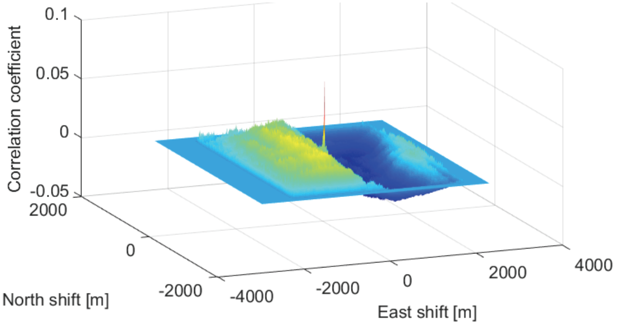
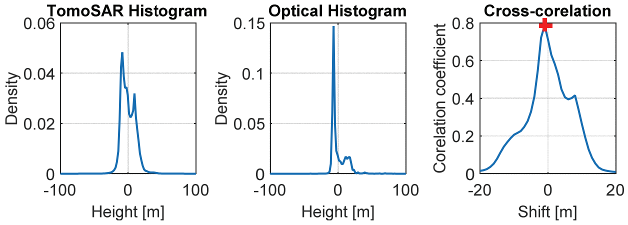


Figure 5. 2D cross-correlation of the edge images of TomoSAR and optical point clouds. Due to the pseudorandom nature of the urban infrastructure, a single prominent peak can always be found.



(a) (b) (c)

Figure 6. (a) The height histogram of TomoSAR point cloud, (b) the height histogram of LiDAR point cloud, and (c) the correlation of (a) and (b), where the red cross marks the peak position which is at -2 m.

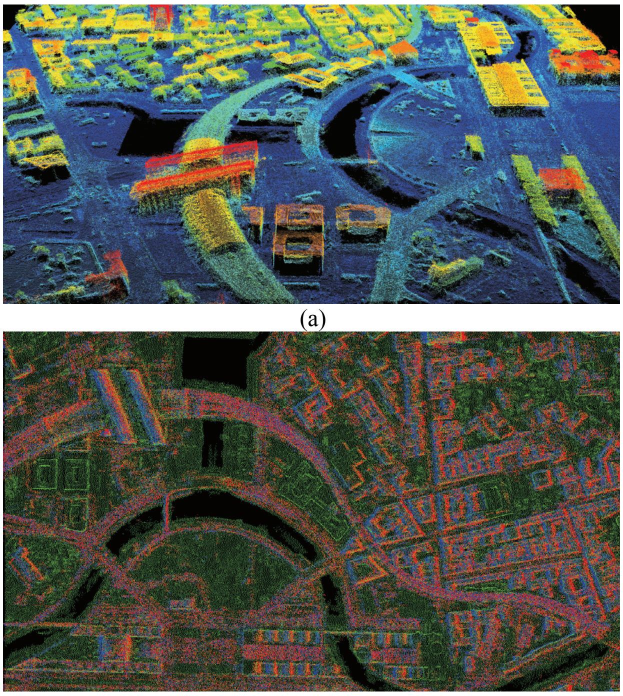
3.4 Final Solution

The final solution is obtained using a normal ICP algorithm based on the initial solution calculated from the previous step. ICP solves the following equation

$$\left\{\hat{\mathbf{R}},\hat{\mathbf{t}}\right\} = \min \sum_i \left\|\mathbf{x}_i - \mathbf{R}\mathbf{p}_i - \mathbf{t}\right\|_2^2 \tag{1}$$

where **R** and **t** are the rotation matrix and the translation vector, and $\mathbf{x}_i$ and $\mathbf{p}_i$ are the corresponding point pair. Given correct point pairs of the two point clouds, (1) can be easily solved. Assuming the closest points being the corresponding point pair, ICP iteratively improves the co-registration results. However, it suffers from finding local minimum. Therefore, giving a good initial estimate is the key to the success of ICP.

In our implementation, the façade points in the TomoSAR point clouds are removed to prevent ICP from finding a wrong solution. Figure 7(a) demonstrates the co-registered point cloud combining the optical images-derived one and two TomoSAR point clouds from ascending and descending viewing angles with color representing the height. Successful co-registration can be confirmed by seeing the correct location of the façade points in Figure 7(b) which shows the top view of fused point cloud with different colors representing different point clouds.


(a)

(b)

Figure 7. (a) the fused point cloud combining the optical images-derived one and two TomoSAR point clouds from ascending and descending viewing angle with color representing the height, and (b) the top view of co-registered point cloud where red, blue, green representing the points from descending TomoSAR, optical, and ascending TomoSAR, respectively.

4. SEMANTIC CLASSIFICATION IN OPTICAL IMAGE

The semantic classification is done in a sliding patch manner. Each patch is described using a dictionary, to be specific, the occurrence of the atoms in the dictionary. Such model is known as the Bag of Words (BoW) (Csurka et al., 2004). The final patch classification is achieved using support vector machine (SVM). The detailed workflow is as follows.

4.1 BoW Model

BoW originates from text classification, where a text is modeled as the occurrence of the words in a dictionary, disregarding the grammar as well as the order. This is also recently employed in computer vision, especially in image classification. Analogous to text, the BoW descriptor **w** of an image (in our case an image patch) **Y** is modeled as the occurrence of the "visual" words in a predefined dictionary **D**, i.e.:

$$\mathbf{w} = h_{\mathbf{D}}\left(\psi\left(\mathbf{Y}\right)\right) \tag{2}$$

where $h(\bullet)$ is the histogram operator, and $\psi(\bullet)$ is the transformation function from the image space to the feature space. Hence the visual words refer to the representative features in the image, whose ensemble constructs the dictionary.

4.2 Dictionary Learning

Define the dictionary matrix as $\mathbf{D} \in \mathbb{R}^{N \times k}$, where N is the dimension of the word, i.e. feature vector/atom, and k is the number of words. The k feature vector should include representative features appearing in the whole image, so that each patch can be well described. Therefore, the dictionary is usually overcomplete.

We adopt a dictionary learning approach commonly used in the computer vision community:

1. Sample the whole image sufficiently dense, and computing the feature at each sampled location. To this end, a large number of feature vectors are collected.

2. Reduce the number of feature vectors by quantization in the feature space. Here, we perform an unsupervised clustering, e.g. k-means. The cluster centroids are extracted as the final dictionary. Figure 8 exemplify the quantization in a 2-D feature space. The colored crosses are the features extracted from the whole image. The cluster centroids are the final words in the dictionary.

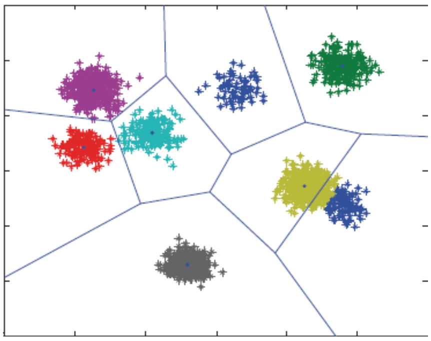


Figure 8. Demonstration of dictionary learning in two dimensional feature space. The colored crosses are the features collected from all the patches in the image. A k-means clustering is performed to get k cluster centers, i.e. the dictionary atoms. Image modified from (Cui, 2014).

4.3 Patch Description

Each patch is described using Equation (2). Similar to the feature extraction in the dictionary learning step, we calculate the dense local features of each patch, i.e. the feature is computed in a sliding window through the patch. This is described in Figure 9 where the red window traverses the patch, and computes one local feature vector at each position.

The descriptor of a patch is the occurrence (histogram) of the collected features in the dictionary. This is calculated by assigning the features to their nearest neighbours in the dictionary words. To this end, the patch descriptor is a vector $\mathbf{v} \in \mathbb{R}^k$.

Several commonly used features have been tested, which includes the most popular scale-invariant feature transform (SIFT) suggested by many literatures. In our problem, the simple vectorized RGB pixel values in a 3×3 sliding window turned out to be the most appropriate feature, and hence are selected for further classification.

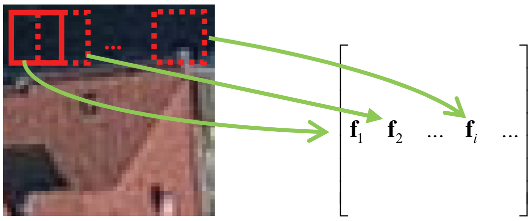


Figure 9. Demonstration of dense local feature computed in an image patch. The feature is computed in the red sliding window, and the local features are denoted as $\mathbf{f}_i$ in the

4.4 Classification

The classification is done using a linear SVM (Cortes & Vapnik, 1995) implemented in an open source library VLFeat (Andrea Vedaldi, 2010). The SVM classifier finds a hyperplane which separates two classes of training samples with maximal margin. Giving the patch descriptor **v**, its SVM classification is:

$$f(\mathbf{v}) = sign\left(\mathbf{w}^T\mathbf{v} + b\right) \tag{3}$$

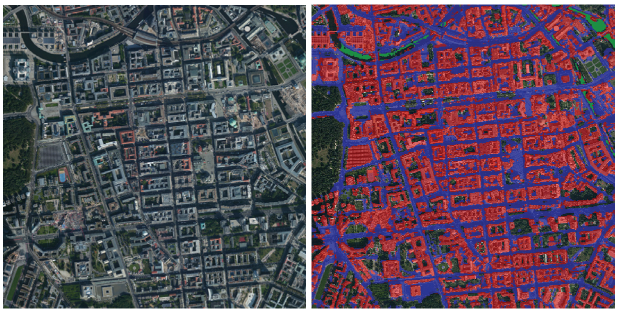
where $\mathbf{w} \in \mathbb{R}^k$ and b are the parameters of the hyperplane, and $sign(\bullet)$ is the sign operator which outputs ±1.

For an *m*-class (*m*>2) problem, we follow the one-against-rest approach. Different SVM is trained for each class. The final classification of a patch **v** is assigned to the one with the largest SVM score, i.e.:

$$f(\mathbf{v}) = \max\left(\mathbf{W}^T\mathbf{v} + \mathbf{b}\right) \tag{4}$$

where $\mathbf{W} \in \mathbb{R}^{k\times m}$ and $\mathbf{b} \in \mathbb{R}^m$ are the concatenated parameters of *m* hyperplanes.

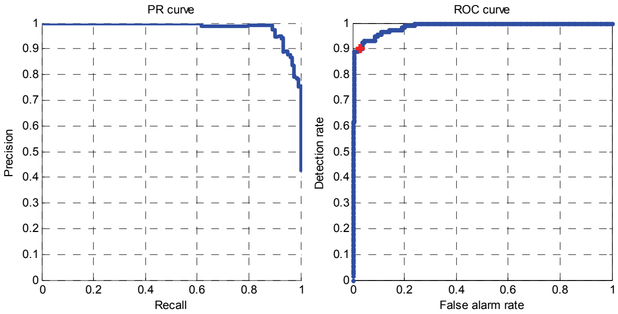
We classify every 4×4 pixel in our test image (5000×5000) taking into account the 50×50 pixel patch around it. We manually selected 570 50×50 pixel patches as training samples. Four classes are preliminarily defined: building, roads/rail, river, and vegetation. Each of them has 240, 159, 39, and 132 training patches, respectively. The feature in our experiment is simply the vectorized RGB pixel values in a 3×3 sliding window, which results in a feature space of 27 dimensions. Figure 10 shows the classification result of a region in the entire image, where the left image is the optical image, and in the right image, classified building, road, river, and vegetation depicted in red, blue, green, and blank respectively. Despite the extremely simple feature we used, the four classes are very well distinguished.



(a) (b)

Figure 10. (a) the test optical image, and (b) the classification of building, road, river, and vegetation, where they are colored in red, blue, green, and blank.

Since we are particularly interested in building, its classification performance is evaluated by classifying half of training samples using the SVM trained with the other half of the samples. The average precision of the current algorithm is 98%. The full precision and recall curve is plotted in Figure 11(a). The equivalent receiver operating characteristic curve is also shown in Figure 11(b), for the readers who are more familiar with it. The red cross marks our decision threshold which gives a detection rate of 90%, and false alarm rate of 3%.



(a) (b)

Figure 11. (a) precision and recall curve of the building classification with an average precision is 98%, and (b) the ROC curve of the classification. The red cross marks our decision point which gives a detection rate of 90%, and false alarm rate of 3%.

5. SEMANTIC-LEVEL ANALYSIS

5.1 Automatic Railway Monitoring

We applied the semantic classification scheme on an orthorectified optical image centered at the Berlin central station. For this analysis, we particularly classified the railway and river class. Figure 12 shows the classification map where the railway class and river class are labelled in green and red, respectively. The classification performance is consistent with the evaluation shown in Figure 11. Some false alarm appeared as small clusters, but they can be removed by post-processing.

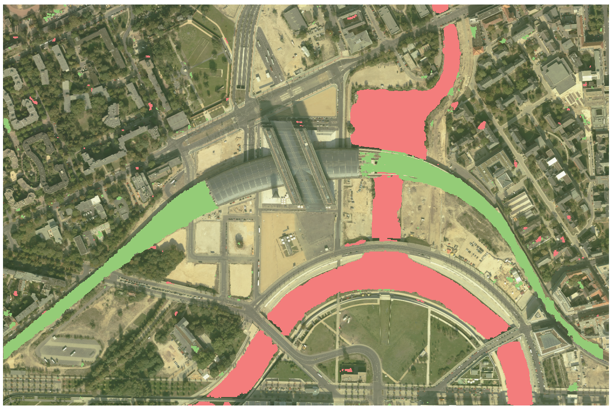


Figure 12. River (red) and railway (green) classified using the BoW method. The classification performance is consistent as the evaluation in Figure 11 shows.

5.1.1 Railway classification refinement using smooth spline

Based on the classification, the corresponding points in the TomoSAR point cloud can be extracted. Assuming the railway is smooth and continuous, a smooth spline function was fitted to the x and y (east and north) coordinates of the railway points to connect separated segments, i.e.:

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \left\{ \lambda \left\| \mathbf{y} - \mathbf{s} \right\|_2^2 + (1-\lambda) \left\| \Delta \mathbf{s} \right\|_2^2 \right\} \quad (5)$$

where $\mathbf{y}$ is the y (north) coordinates of the railway points, $\mathbf{s}$ is the spline function (quadratic or cubic) w.r.t. the x (east) coordinates of the railway points, Δ is the Laplace operator, and $\lambda \in [0,1]$ the smoothing parameter. The regularization on the L2 norm of the second order derivative grant the smoothness of the spline function. The smooth spline is centered in the railway, and the width of the railway is adaptively estimated at each position. Therefore, we are able to interpolate the gap of the railway due to miss classification (here due to the presence of the Berlin central station). Figure 13(a) shows the extracted continuous railway points overlaid on the optical image. The color shows the amplitude of seasonal motion caused by thermal dilation.

5.1.2 Total variation denoising

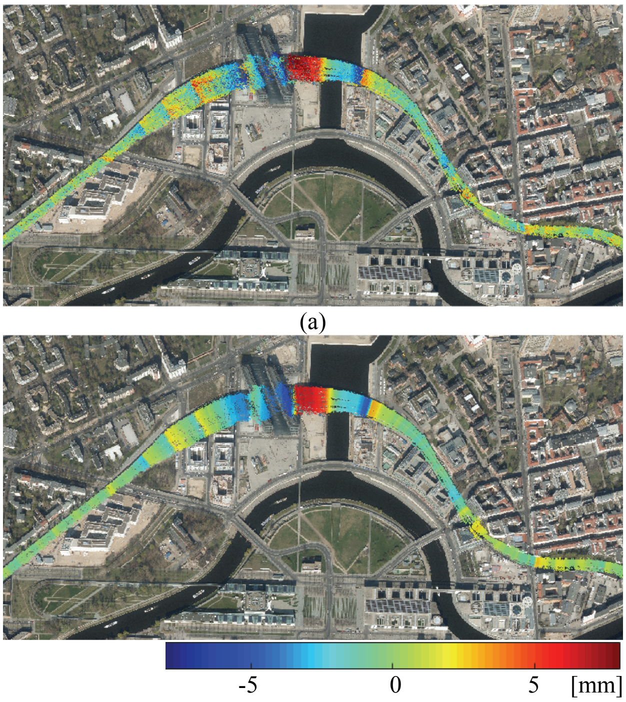
Due to the high dynamics of the motion in high resolution SAR data, we introduce an additional step of denoising in order to retrieve some higher level information such as the joints of the railway sections.

Because the thermal dilation of the steel railway beam is mostly proportional to its length (Kerr, 1978), it is plausible to assume that the scatterers on the same cross-section of the railway undergo identical deformation. Therefore, the scatterers' deformation along the railway can be transformed into one dimension, i.e. the railway distance. The original deformation estimates in the radar's line-of-sight direction must also be projected to the railway direction.

In order to preserve the edge and the piecewise linear structure of the railway deformation parameters which can be observed in Figure 13(a), we employ a minimization of total variation of the second order derivative of the deformation function along the railway:

$$\hat{\mathbf{g}} = \arg\min_{\mathbf{g}} \left\{ \frac{1}{2} \left\| \mathbf{g} - \mathbf{v} \right\|_2^2 + \lambda \left\| \Delta \mathbf{g} \right\|_1 \right\} \quad (6)$$

where $\mathbf{v}$ is the deformation estimates along the railway direction, $\mathbf{g}$ is its denoised version. As shown in literatures of total generalized variation (Bredies, Kunisch & Pock, 2010; Knoll et al., 2011), the L1 norm of second order derivative is convex and lower semi-continuous, one can solve it using convex optimization solvers. Figure 14 (a) shows the original and denoised deformation estimates as a function of the railway direction. The edges due to different railway segments are clearly preserved. Figure 13(b) shows the denoised deformation estimates re-projected back into the world coordinate, in order to have a visual comparison with Figure 13(a).



(a)

(b)

Figure 13. (a) Railway points extracted from the TomoSAR point cloud. The color shows the amplitude of seasonal motion due to the thermal expansion of the steel, and (b) denoised deformation estimates using minimization of total variation.

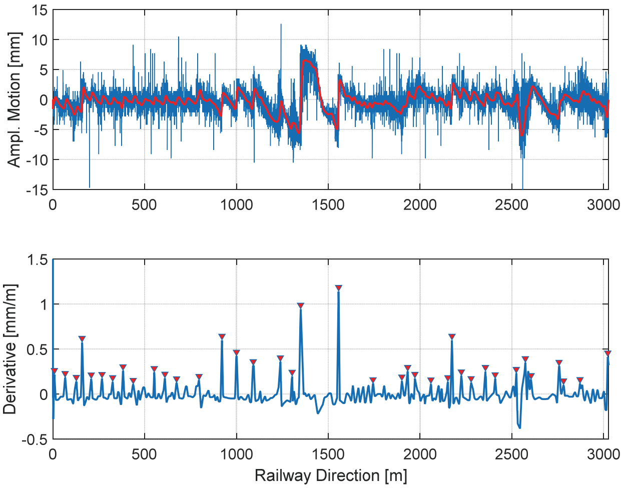


Figure 14. (a) Original estimates of the amplitude of seasonal deformation in blue, and the denoised deformation estimates in red estimated by regularizing on the total variation of its second order derivative, and (b) peaks detected in the derivative of the denoised deformation function along the railway direction. They are the locations of the railway segment joints.

5.1.3 Railway joints detection

By detecting the peaks in the derivative of the deformation function along the railway direction, the joints of railways can be detected. Constraint is put on the minimum distance between two peaks representing the minimum length of a railway segment. The detected peaks in the deformation's derivative can be seen in (b). The positions of the peaks on the railway are shown as the green dots in Figure 15(a). Each green dot represents the midpoint of the joint cross-section. In Figure 15(c), we provide the close up view of the two joints in the optical image with 20cm pixel size. It can be clearly observed that the railway joint shown up as dark lines in the optical image.

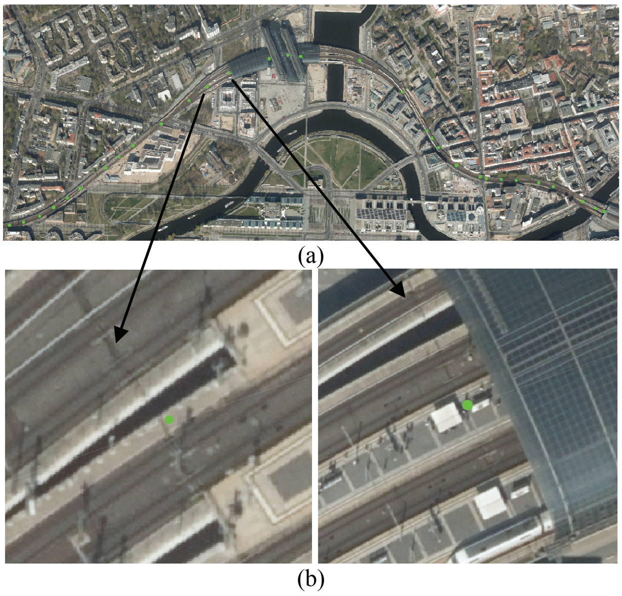


Figure 15. (a) the midpoint of the detected railway joint cross-section marked in green, and (b) close up view of the railway joint. The background optical image has a ground spacing of 20cm.

5.2 Automatic Bridge Monitoring

By analysing the discontinuity of the river segmentation and assuming the discontinuities are caused by bridges, the bridges' positions can be detected automatically. The workflow of identifying all the bridge is fairly simple. Starting with the initial river classification, small clusters are removed by setting threshold on the area of connected region. Followed by morphological closing operation, the gaps caused by bridges are closed as shown in Figure 16(b). Lastly, by finding the convex hull of each connected region in the difference image of Figure 16(b) and (a), the bridge mask can be identified. As we can see, all the bridges in Figure 12 are identified, except a very narrow one that was originally false classified as river.

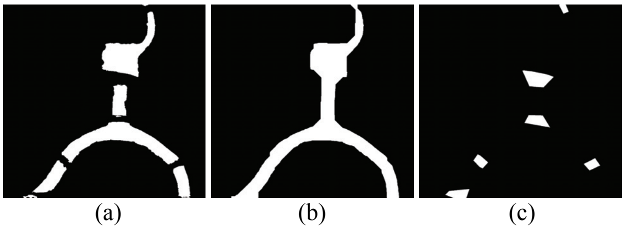


Figure 16. (a) initial river mask with false detected small cluster removed, (b) connected river mask by applying morphological closing, and (c) the final bridge mask by finding the convex hull of each region in the difference between (b) and (a).

The corresponding bridge points are extracted from the TomoSAR point cloud, and projected to the optical image. The projected bridge points are shown in Figure 17 where the color represents the amplitude of seasonal deformation. The upper most bridge belongs to a segment of the railway which is known to have thermal expansion. The middle bridge undergoes a 5mm seasonal motion in its west end and 2mm at the east end. This suggests a more rigid connection of the bridge with the foundation at its east end. The two lower bridges are stable according to the motion estimates.

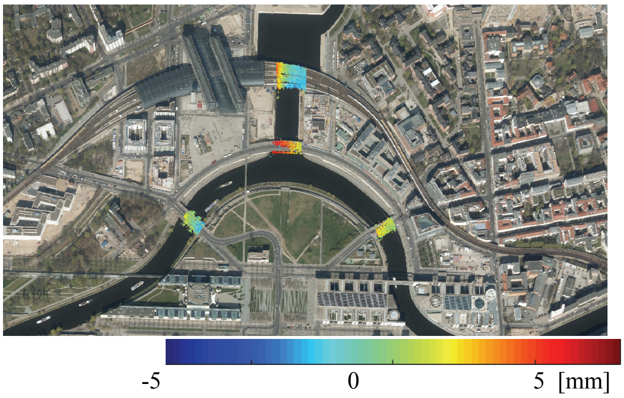


Figure 17. Overlay of the amplitude of seasonal motion of brides extracted from the TomoSAR point cloud on the optical image. The bridges are automatically detected from the classification map shown in Figure 12 using discontinuity analysis.

CONCLUSION

This paper presents the first systematic semantic analysis of very high resolution InSAR point cloud in urban area. Through co-registering optical image and InSAR point cloud to a common reference 3-D model, we are able to relate the semantic meaning extracted from the optical image to the

InSAR point cloud. The complementary information provided by the two data types enables an object-level InSAR deformation and 3-D analysis.

In the future, we will include more semantic classes, such as high-rise buildings, residential area, or even specific landmarks, and so on. To reduce the human interaction, we aim at a completely unsupervised semantic classification.

REFERENCES

Adam, N., Kampes, B., Eineder, M., Worawattanamateekul, J. & Kircher, M. 2003. The development of a scientific permanent scatterer system. In *ISPRS Workshop High Resolution Mapping from Space, Hannover, Germany*, pp. 6.

Andrea Vedaldi, B.F. 2010. VLFeat: an open and portable library of computer vision algorithms. In *Proceedings of the 18th International Conference on Multimedea 2010*, pp. 1469–1472. , Firenze, Italy.

Bamler, R., Eineder, M., Adam, N., Zhu, X. & Gernhardt, S. 2009. Interferometric Potential of High Resolution Spaceborne SAR. *Photogrammetrie - Fernerkundung - Geoinformation* 2009(5), 407–419.

Bredies, K., Kunisch, K. & Pock, T. 2010. Total Generalized Variation. *SIAM Journal on Imaging Sciences* 3(3), 492–526.

Cortes, C. & Vapnik, V. 1995. Support-vector networks. *Machine Learning* 20(3), 273–297.

Csurka, G., Dance, C., Fan, L., Willamowski, J. & Bray, C. 2004. Visual categorization with bags of keypoints. In *Workshop on statistical learning in computer vision, ECCV*, pp. 1–2.

Cui, S. 2014. Spatial and temporal SAR image information mining. Ph.D. thesis. Universität Siegen.

Ferretti, A., Prati, C. & Rocca, F. 2001. Permanent scatterers in SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing* 39(1), 8–20.

Ferretti, A., Fumagalli, A., Novali, F., Prati, C., Rocca, F. & Rucci, A. 2011. A New Algorithm for Processing Interferometric Data-Stacks: SqueeSAR. *IEEE Transactions on Geoscience and Remote Sensing* 49(9), 3460–3470.

Fornaro, G., Reale, D. & Serafino, F. 2009. Four-Dimensional SAR Imaging for Height Estimation and Monitoring of Single and Double Scatterers. *IEEE Transactions on Geoscience and Remote Sensing* 47(1), 224–237.

Fornaro, G., Verde, S., Reale, D. & Pauciullo, A. 2015. CAESAR: An Approach Based on Covariance Matrix Decomposition to Improve Multibaseline-Multitemporal Interferometric SAR Processing. *IEEE Transactions on Geoscience and Remote Sensing* 53(4), 2050–2065.

Gamba, P. & Houshmand, B. 2002. Joint analysis of SAR, LIDAR and aerial imagery for simultaneous extraction of land cover, DTM and 3D shape of buildings. *International Journal of Remote Sensing* 23(20), 4439–4450.

Gernhardt, S. & Bamler, R. 2012. Deformation monitoring of single buildings using meter-resolution SAR data in PSI. *ISPRS Journal of Photogrammetry and Remote Sensing* 73, 68–79.

Jiang, M., Ding, X., Hanssen, R.F., Malhotra, R. & Chang, L. 2015. Fast Statistically Homogeneous Pixel Selection for Covariance Matrix Estimation for Multitemporal InSAR. *IEEE Transactions on Geoscience and Remote Sensing* 53(3), 1213–1224.

Kampes, B.M. 2006. Radar *Interferometry - Persistent Scatterer Technique*, Dordrecht, The Netherlands: Springer, 211p.

Kerr, A.D. 1978. Analysis of thermal track buckling in the lateral plane. *Acta Mechanica* 30(1-2), 17–50.

Knoll, F., Bredies, K., Pock, T. & Stollberger, R. 2011. Second order total generalized variation (TGV) for MRI. *Magnetic resonance in medicine* 65(2), 480–491.

Lombardini, F. 2005. Differential tomography: a new framework for SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing* 43(1), 37–44.

Sobel, I. 1968. An Isotropic 3x3 Image Gradient Operator. *Presentation at Stanford A.I. Project 1968.*

Wang, Y. & Zhu, X. 2015. Automatic Feature-based Geometric Fusion of Multi-view TomoSAR Point Clouds in Urban Area. *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing* 8(3), 953 – 965.

Wang, Y., Zhu, X. & Bamler, R. 2012. Retrieval of Phase History Parameters from Distributed Scatterers in Urban Areas Using Very High Resolution SAR Data. *ISPRS Journal of Photogrammetry and Remote Sensing* 73, 89–99.

Wegner, J.D., Thiele, A. & 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* 38(Part 3), W4.

Wegner, J.D., Ziehn, J.R. & Soergel, U. 2014. Combining High-Resolution Optical and InSAR Features for Height Estimation of Buildings With Flat Roofs. *IEEE Transactions on Geoscience and Remote Sensing* 52(9), 5840–5854.

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

Zhu, X., & Bamler, R. 2010b. Very High Resolution Spaceborne SAR Tomography in Urban Environment. *IEEE Transactions on Geoscience and Remote Sensing* 48 (12), 4296–4308.