

Robust Object-Based Multipass InSAR Deformation Reconstruction

Jian Kang, *Student Member, IEEE*, Yuanyuan Wang, *Member, IEEE*, Marco Körner, *Member, IEEE*,
and Xiao Xiang Zhu, *Senior Member, IEEE*

Abstract—Deformation monitoring by multipass synthetic aperture radar (SAR) interferometry (InSAR) is, so far, the only imaging-based method to assess millimeter-level deformation over large areas from space. Past research mostly focused on the optimal retrieval of deformation parameters on the basis of a single pixel or a pixel cluster. Only until recently, the first demonstration of object-based urban infrastructure monitoring by fusing InSAR and the semantic classification labels derived from optical images was presented by Wang *et al.* Given such classification labels in the SAR image, we propose a general framework for object-based InSAR parameter retrieval, where the parameters of the whole object are jointly estimated by the inversion of a regularized tensor model instead of pixelwise. Our approach does not assume the stationarity of each sample in the object, which is usually assumed in other pixel cluster-based methods, such as SqueeSAR. In addition, to handle outliers in real data, a robust phase recovery step prior to parameter retrieval is also introduced. In typical settings, the proposed method outperforms the current pixelwise estimators, e.g., periodogram, by a factor of several tens in the accuracy of the linear deformation estimates. Last but not least, for a practical demonstration on bridge monitoring, we present a full workflow of long-term bridge monitoring using the proposed approach.

Index Terms—Bridge detection, joint deformation reconstruction, object-based, synthetic aperture radar (SAR), SAR interferometry (InSAR).

I. INTRODUCTION

A. Multipass Synthetic Aperture Radar Interferometry

MULTIPASS or multibaseline synthetic aperture radar (SAR) interferometry (InSAR) techniques, such as persistent scatterer interferometry (PSI) [5]–[12] and differential SAR tomography (D-TomoSAR) [13]–[18], are the most popular methods for long-term millimeter-level deformation mon-

Manuscript received October 20, 2016; revised January 5, 2017; accepted February 6, 2017. This work was supported in part by the China Scholarship Council, in part by the European Research Council (ERC) through the European Unions Horizon 2020 Research and Innovation Programme (Acronym: So2Sat) under Grant ERC-2016-StG-714087, and in part by the Helmholtz Association through the framework of the Young Investigators Group SiPEO under Grant VH-NG-1018. (*Corresponding author:* Xiao Xiang Zhu.)

J. Kang and Y. Wang are with Signal Processing in Earth Observation Group, Technical University of Munich, 80333 Munich, Germany (e-mail: jian.kang@tum.de).

M. Körner is with the Chair of Remote Sensing Technology, Technical University of Munich, 80333 Munich, Germany.

X. X. Zhu is with the Remote Sensing Technology Institute, German Aerospace Center (DLR), 82234 Wessling, Germany, and also with Signal Processing in Earth Observation, Technical University of Munich, 80333 Munich, Germany (e-mail: xiao.zhu@dlr.de).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2017.2684424

itoring over large areas. Through modeling the interferometric phase of the scatterers, we are able to reconstruct the 3-D position and the deformation history of individual scatterers.

The main focus of the past research on multipass InSAR was on the optimal retrieval of the phase history parameters of scatterers. On the one hand, single pixel-based methods, such as PSI [5]–[12] and D-TomoSAR [13]–[18], have been the workhorses for many applications on the monitoring of urban areas. Some advanced D-TomoSAR methods have been especially developed, for example, using compressive sensing [19]–[21] to achieve superresolution in the elevation reconstruction, and fusing SAR imaging geodesy [22] and TomoSAR inversion to obtain absolute *geodetic TomoSAR* [23] point clouds. On the other hand, pixel-cluster-based methods, such as SqueeSAR [4], [24]–[27], CAESAR [28], and TomoSAR based on distributed scatterers [29], [30], exploit statistical similarities between the neighboring pixels, in order to retrieve the phase history parameters from the covariance matrix. Statistical ergodicity of the selected pixel clusters is always assumed in these methods for the estimation of the covariance matrix. In a similar vein, nonlocal-InSAR [31]–[34] also selects similar pixels for a weighted averaging. Yet, the similarity is compared patchwise instead of pixelwise in the above-mentioned pixel-cluster-based methods. That is to say, the local morphology of the pixel has been taken into account. However, no explicit geometric information is utilized to guide the similarity comparison.

Although some of the above-mentioned techniques do exploit information from multiple neighboring pixels or patches, the natural semantic and geometric information in SAR images has not been explicitly employed. Such geometric information can be the object mask of a building façade where the height and deformation obey certain prior knowledge, such as continuity and smoothness. Introducing them into the retrieval of phase history parameters can be of great advantage, for example, to improve the accuracy of the estimates, or to reduce the number of images required to obtain a reliable estimation. Only until recently, a pioneer work [35] has demonstrated that by introducing building footprints from OpenStreetMap as a prior knowledge of pixels sharing similar heights, along with joint sparse reconstruction techniques, a highly accurate tomographic reconstruction can be achieved using six interferograms only, instead of the typically required 20–100.

In this paper, we seek to investigate the integration of semantic and geometric information into the retrieval of geophysical parameters, e.g., linear deformation rate, in a more

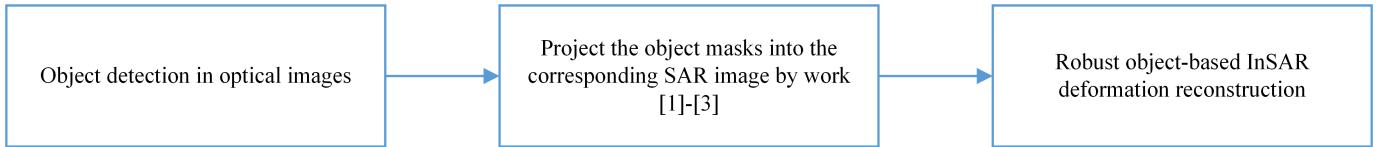


Fig. 1. Overall workflow.

TABLE I
MATHEMATIC NOTATION

$\mathcal{A}, \mathbf{A}, \mathbf{a}, a$	Tensor, Matrix, Vector, Scalar
$(\cdot)^H$	Conjugate transpose
$\ \mathbf{A}\ _*$	Matrix nuclear norm: the sum of its singular values, i.e. $\ \mathbf{A}\ _* := \sum_i \sigma_i$
$\mathbf{A}_{(n)}$	The mode-n unfolding of the tensor \mathcal{A}
$\text{Trank}(\mathcal{A})$	Tucker rank of tensor \mathcal{A}
$\ \mathcal{A}\ _F$	$\sqrt{\sum_{i_1}^{I_1} \sum_{i_2}^{I_2} \dots \sum_{i_N}^{I_N} a_{i_1, i_2, \dots, i_N} ^2}$, where a_{i_1, i_2, \dots, i_N} is the element of tensor \mathcal{A}
$\ \mathcal{A}\ _1$	Tensor L1 norm: $\ \text{vec}(\mathcal{A})\ _1$
$\text{vec}(\mathcal{A})$	The vectorization of \mathcal{A}

general framework, in multipass InSAR.

B. Object Detection in SAR Image

Due to the complex scattering mechanism and the layover effect in SAR images, the detection and the semantic classification of objects in SAR images present a greater challenge than in optical images. Still, several papers have been published in this regard. For example, [36]–[40] provided a classification of large urban areas in SAR images, but they did not achieve the classification for each object instance. In [41] and [42], the extraction of buildings and the estimation of their heights were well studied, with the assumptions that the buildings were following specific models or with flat roofs. In [43], the models of individual buildings were reconstructed that required high-quality TomoSAR point clouds as input, however. Only until recently, the SARptical technique [1]–[3] has provided a promising solution. Instead of using SAR images alone, it combines the semantic classification labels obtained from optical images with SAR images via a 3-D geometric coregistration.

C. Contributions of This Paper

The contributions of this paper are listed as follows.

- 1) Extended the multipass InSAR phase model of single pixel to an object-level by means of a tensor representation.
- 2) Proposed an object-based multipass InSAR parameter inversion technique that exploits the prior knowledge of the deformation pattern of a certain object.
- 3) Introduced a robust phase recovery step prior to the object-based parameter retrieval to deal with possible outliers in real data.

The overall workflow contains three main steps, which is shown in Fig. 1. In this paper, we choose bridges as the target objects. A bridge extraction method from optical images

is developed to achieve this task. Of course, the proposed approach can be used for other types of objects in urban areas, e.g., roads and roofs.

D. Structure of This Paper

The rest of this paper is organized as follows. Section II describes the tensor representation of the object-based InSAR phase and the robust object-based InSAR deformation reconstruction algorithm. Section III introduces the framework for the bridge extraction in optical images. Section IV provides a brief description of the SARptical algorithm, which projects the bridge masks from the optical images to the SAR images. In Section V, as the experiment part, we demonstrate the proposed algorithm on simulated and real SAR data. Section VI draws the conclusion of this paper.

E. Mathematical Notations

The mathematical notation in this paper is summarized in Table I.

II. OBJECT-BASED INSAR DEFORMATION RECONSTRUCTION

A. Single-Pixel Multipass InSAR Model

The interferometric phase history of a persistent scatter (PS) in a multipass InSAR stack is determined by the elevation (topography) of the scatter, as well as its deformation over the observed period. Correspondingly, a PS phase history $\mathbf{g}(s, p)$ can be modeled as follows:

$$\bar{\mathbf{g}}(s, p) = a \exp \left\{ -j \left(\frac{4\pi}{\lambda r} s \times \mathbf{b} + \frac{4\pi}{\lambda} p \times \boldsymbol{\tau} \right) \right\} \quad (1)$$

where a is the modeled amplitude of the PS, \mathbf{b} is the vector of the spatial baseline, $\boldsymbol{\tau}$ is the vector of the deformation model, e.g., $\boldsymbol{\tau} = \mathbf{t}$ for linear motion, and $\boldsymbol{\tau} = \sin(2\pi(\mathbf{t} - t_0))$

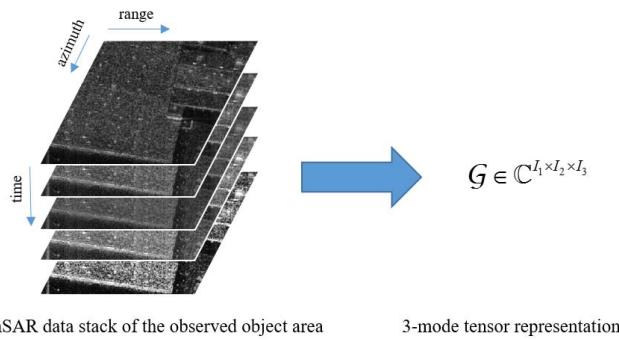


Fig. 2. Three-mode tensor representation of the InSAR data stack.

for the seasonal motion model with the temporal baseline t , s and p are the unknown elevation and deformation parameters to be estimated, respectively, λ is the wavelength of the radar transmitted signals, and r denotes the range between radar and the observed object.

The maximum-likelihood estimator of the parameters s and p is

$$\{\hat{s}, \hat{p}\} = \underset{s, p}{\operatorname{argmin}} \| \mathbf{g} - \bar{\mathbf{g}}(s, p) \|_2 \quad (2)$$

which is equivalent to the periodogram [44]–[46]

$$\{\hat{s}, \hat{p}\} = \underset{s, p}{\operatorname{argmax}} | \bar{\mathbf{g}}(s, p)^H \mathbf{g} | \quad (3)$$

where \mathbf{g} is the input-observed PS phase vector, and \hat{s} and \hat{p} represent the elevation and deformation parameter estimates, respectively. Since we mainly focus on the complex phase of PSs, the amplitude is ignored in the following content.

B. Tensor Extension of Multipass InSAR Model

Given an object area, e.g., roof, as shown in Fig. 2, its interferometric phase stack can be represented as a three-mode tensor $\mathcal{G} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$, where I_1 and I_2 represent the spatial dimension in range and azimuth, respectively, and I_3 denotes the number of SAR images. Similar to (1), its tensor extension (amplitude dropped) can be expressed as follows:

$$\bar{\mathcal{G}}(\mathbf{S}, \mathbf{P}) = \exp \left\{ -j \left(\frac{4\pi}{\lambda r} \mathbf{S} \otimes \mathbf{b} + \frac{4\pi}{\lambda} \mathbf{P} \otimes \boldsymbol{\tau} \right) \right\}. \quad (4)$$

Here, $\bar{\mathcal{G}}$ is the modeled complex phase tensor of the object, and \mathbf{S} and \mathbf{P} represent the matrices of elevation and deformation to be estimated, respectively. The symbol \otimes denotes the outer product, which plays a role in the dimension extension [47].

Given the observed complex phase tensor \mathcal{G} of a certain object, the goal is to jointly reconstruct all the parameters. In the proposed method, we exploit the prior knowledge of their spatial deformation patterns. To this end, the optimization problem of (2) is extended to the following expression:

$$\{\hat{\mathbf{S}}, \hat{\mathbf{P}}\} = \underset{\mathbf{S}, \mathbf{P}}{\operatorname{argmin}} \frac{1}{2} \| \mathcal{W} \odot (\mathcal{G} - \bar{\mathcal{G}}(\mathbf{S}, \mathbf{P})) \|_F^2 + \eta f(\mathbf{S}, \mathbf{P}) \quad (5)$$

where \mathcal{W} denotes a weighting tensor, and \odot denotes the elementwise product between two tensors. The first term is the weighted data fidelity term, which calculates the weighted Frobenius norm of the log likelihood between the observed

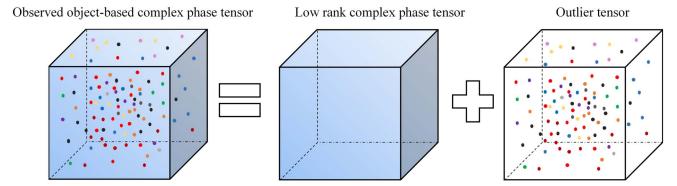


Fig. 3. Robust low-rank tensor decomposition for the observed object-based complex phase tensor.

tensor \mathcal{G} and the modeled tensor $\bar{\mathcal{G}}$. The weighting tensor is created by replicating the coherence matrix obtained from the pixelwise periodogram in the temporal dimension. Pixels of higher coherence are given higher weights. $f(\mathbf{S}, \mathbf{P})$ denotes the penalty term, which represents the spatial prior of \mathbf{S} and \mathbf{P} . The regularization parameter η controls the balance between these two terms. Therefore, (5) is also the maximum *a posteriori* (MAP) estimator of \mathbf{S} and \mathbf{P} .

One popular smoothness prior is the total variation (TV) norm, which is widely used in multiple image processing problems, e.g., image deblurring, denoising, and inpainting [48]–[50]. The optimization in (5) with a TV penalty can be written as

$$\begin{aligned} \{\hat{\mathbf{S}}, \hat{\mathbf{P}}\} = \underset{\mathbf{S}, \mathbf{P}}{\operatorname{argmin}} & \frac{1}{2} \| \mathcal{W} \odot (\mathcal{G} - \bar{\mathcal{G}}(\mathbf{S}, \mathbf{P})) \|_F^2 \\ & + \eta \sum_{i, j} | \mathbf{P}(i+1, j) - \mathbf{P}(i, j) | + | \mathbf{P}(i, j+1) - \mathbf{P}(i, j) | \end{aligned} \quad (6)$$

where i and j are the pixel row and column coordinates of the matrix, respectively.

Equation (6) is the minimization of a nonlinear and nonconvex function. It can be solved by the quasi-Newton method: Limited-memory-Broyden–Fletcher–Goldfarb–Shanno [51]–[53], which is developed for unconstrained nonlinear optimization. It has recently been applied as a comparison solver for nonconvex optimization in [54] and utilized as an efficient solver for patch-ordering regularization inverse problems [55].

C. Robust Object-Based InSAR Deformation Reconstruction

In real data, the observed data stack \mathcal{G} may contain outliers. The MAP we proposed is based on Gaussian noise that cannot robustly reconstruct the deformation matrix. Therefore, inspired by Wei [56], which provides a thorough analysis of exploiting low-rank information for multitemporal SAR data sets, the above-mentioned approach is extended to a robust version in case outliers exist. We introduce a robust object-based phase recovery step to the observed phase stack \mathcal{G} , prior to the reconstruction. In the object-based phase recovery step, as shown in Fig. 3, we first decompose the observed phase tensor \mathcal{G} into two parts: the low-rank tensor part \mathcal{X} and the sparse outlier tensor part \mathcal{E} , as the outlier-free complex phase stack \mathcal{X} can be considered as a low-rank tensor compared with the observed phase tensor \mathcal{G} . Estimating \mathcal{X} leads to the following optimization problem [57]:

$$\{\hat{\mathcal{X}}, \hat{\mathcal{E}}\} = \underset{\mathcal{X}, \mathcal{E}}{\operatorname{argmin}} \| \mathcal{X} \|_* + \gamma \| \mathcal{E} \|_1, \text{ s.t. } \mathcal{X} + \mathcal{E} = \mathcal{G} \quad (7)$$

where $\|\mathcal{X}\|_*$ denotes the tensor nuclear norm, $\|\mathcal{E}\|_1$ is the tensor L1 norm of sparse outliers, and γ is the regularization parameter. $\|\mathcal{X}\|_*$ can be calculated by the sum of the N nuclear norm $\sum_n \|\mathbf{X}_{(n)}\|_*$ of the mode- n unfoldings of \mathcal{X} , i.e., $\|\mathcal{X}\|_* = \sum_n \|\mathbf{X}_{(n)}\|_*$.

The convex optimization problem (7) can be solved by the general framework of the alternating direction method of multipliers (ADMM) [58]. The advantage of ADMM is that it tries to split one difficult optimization problem into several subproblems, where each of them has a closed-form solution. We first convert the constraint optimization problem in (7) to its augmented Lagrangian function as follows:

$$\begin{aligned} L_\mu(\mathcal{X}, \mathcal{E}, \mathcal{Y}) = & \|\mathcal{X}\|_* + \gamma \|\mathcal{E}\|_1 - \langle \mathcal{Y}, \mathcal{X} + \mathcal{E} - \mathcal{G} \rangle \\ & + \frac{1}{2\mu} \|\mathcal{X} + \mathcal{E} - \mathcal{G}\|_F^2 \end{aligned} \quad (8)$$

where \mathcal{Y} denotes the introduced dual variable and μ is the penalty parameter. Accordingly, the minimization of L_μ , with respect to each variable, can be solved by the following optimization subproblems.

1) \mathcal{X} Subproblem: The subproblem of L_μ , with respect to \mathcal{X} , can be rewritten as

$$\min_{\mathcal{X}} \|\mathcal{X}\|_* + \frac{1}{2\mu} \|\mathcal{X} + \mathcal{E} - \mathcal{G} - \mu\mathcal{Y}\|_F^2. \quad (9)$$

It can be solved by the singular value thresholding (SVT) operator [59], [60] of mode- n ($n = 1, 2, \dots, N$) unfolding of the tensor $\mathcal{E} - \mathcal{G} - \mu\mathcal{Y}$, where SVT operator is defined as $\mathcal{T}_\mu(\mathbf{A}) := \mathbf{U}\text{diag}(\max(\sigma_i - \mu, 0))\mathbf{V}$ with \mathbf{U} , \mathbf{V} , and σ_i obtained from singular value decomposition from the matrix \mathbf{A} .

2) \mathcal{E} Subproblem: The subproblem of L_μ , with respect to the outlier tensor \mathcal{E} , has the following form:

$$\min_{\mathcal{E}} \gamma \|\mathcal{E}\|_1 + \frac{1}{2\mu} \|\mathcal{X} + \mathcal{E} - \mathcal{G} - \mu\mathcal{Y}\|_F^2. \quad (10)$$

This L1-norm-induced subproblem can be efficiently solved by applying the soft-thresholding operator defined as $\mathcal{S}_\gamma(\mathcal{A}) := \text{sign}(\mathcal{A}) \odot \max(|\mathcal{A}| - \gamma, 0)$, where \odot denotes the elementwise product (Hadamard product) of two tensors, and $|\mathcal{A}| = \text{sign}(\mathcal{A}) \odot |\mathcal{A}|$.

3) Multiplier Updating: The multiplier \mathcal{Y} can be updated by the following:

$$\mathcal{Y} = \mathcal{Y} - \frac{1}{\mu} (\mathcal{X} + \mathcal{E} - \mathcal{G}). \quad (11)$$

The detailed ADMM pseudocode for solving (7) is summarized in Algorithm 1. By the predefined convergence condition, the optimal solution $\hat{\mathcal{X}}, \hat{\mathcal{E}}$ can be obtained. Then, the outlier-free tensor $\hat{\mathcal{X}}$ replaces \mathcal{G} in (5) for deformation parameter retrieval.

It is worth noting that [28] also exploits the low rankness of the data covariance matrix using principle component analysis in order to separate different scattering components within individual pixels. This requires the phase ergodicity among the spatial samples. In contrast, our approach exploits the low rankness in both the spatial dimension and temporal dimension. The phase ergodicity among the samples in the object is not strictly assumed.

Algorithm 1 Equation (7) Solved by ADMM

Input: $\mathcal{G}, \gamma, \mu, N$

- 1: Initialize $\mathcal{X} = \mathcal{E} = \mathcal{Y} = 0$
- 2: **for** $k = 0$ to maxIter **do**
- 3: Singular value thresholding (SVT) for mode- n unfolding of $\mathcal{G} + \mu\mathcal{Y} - \mathcal{E}$,
 mode- n folding as N tensors, and then average
 them by N :

$$\mathcal{X}^{(k+1)} \leftarrow \frac{1}{N} \sum_{n=1}^N \mathcal{T}_{n,\mu N}(\mathbf{G}_{(n)} + \mu \mathbf{Y}_{(n)}^{(k)} - \mathbf{E}_{(n)}^{(k)}),$$

 where $\mathcal{T}_{n,\mu N}(\cdot) := \text{fold}_n(\mathcal{T}_{\mu N}(\cdot))$
- 4: Elementwise L1 norm thresholding of tensor $\mathcal{G} + \mu\mathcal{Y} - \mathcal{X}^{(k+1)}$:

$$\mathcal{E}^{(k+1)} \leftarrow \mathcal{S}_{\mu\gamma}(\mathcal{G} + \mu\mathcal{Y}^{(k)} - \mathcal{X}^{(k+1)})$$
- 5:
$$\mathcal{Y}^{(k+1)} \leftarrow \mathcal{Y}^{(k)} - \frac{1}{\mu} (\mathcal{X}^{(k+1)} + \mathcal{E}^{(k+1)} - \mathcal{G})$$
- 6: **end for**

Output: $(\mathcal{X}, \mathcal{E})$

III. BRIDGE EXTRACTION IN OPTICAL IMAGES

For completeness, this section introduces the proposed bridge detection algorithm. However, it is not the focus of this paper. The proposed object-based inversion can also be applied to other objects. For the study of other objects in urban areas, such as roads, roofs, and façades, we refer the readers to the papers [61]–[63] and the survey [64].

The proposed bridge detection algorithm is based on the following prior knowledge of bridges.

1) Bridges are above rivers.

2) One bridge should completely cross and divide the river.

Accordingly, we develop the algorithm as shown in Fig. 4.

A. Texture Feature for River Classification

For the first step, considering the homogenous texture of river, we classify the river and the rest pixels using local binary pattern (LBP) [65] as the feature. LBP describes the local structure within a certain neighborhood. Compared with the intensity of the center pixel, the neighboring pixels are thresholded and multiplied by the binomial weights. In this paper, we utilize the rotation-invariant uniform LBP [66] to describe the texture and generate the LBP histogram feature for each patch created by a sliding window through the whole image. A support vector machine (SVM) is employed as the supervised classifier, which is trained with dozens of training patches.

B. Retrieval of River Segments by Active Contour

With the above-mentioned method, we can obtain a coarse classification of the river regions. It is not sufficient to precisely determine the discontinuous regions of the rivers (which are bridges). Therefore, given the initial river mask, we employ the Chan–Vese (CV) segmentation [67] to refine the river segments.

By using the level set formulation $C = \{(x, y) | \phi(x, y) = 0\}$ introduced in [68], where C represents the closed curve and $\phi(x, y) = 0$ is the zero-level set function. CV basically solves

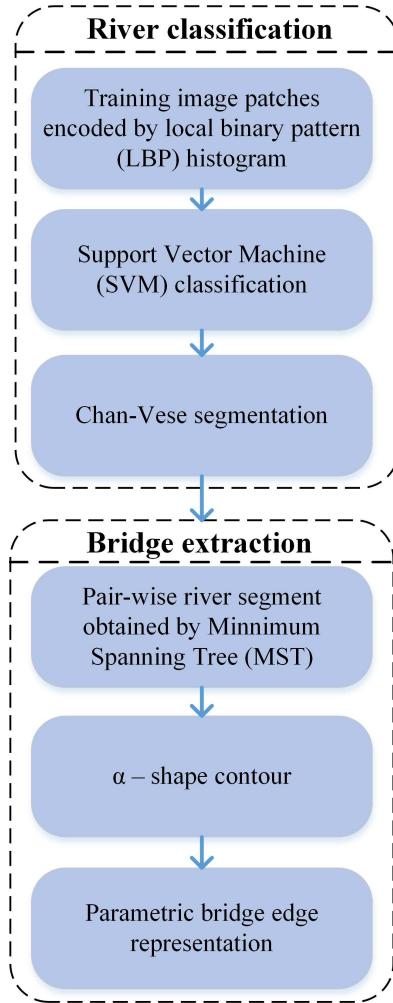


Fig. 4. Bridge extraction workflow.

the following minimization function [67]:

$$F(c_1, c_2, \phi) = \mu \int_{\Omega} \|\nabla H(\phi)\| + \int_{\Omega} (I(x, y) - c_1)^2 H(\phi) dx dy + \int_{\Omega} (I(x, y) - c_2)^2 (1 - H(\phi)) dx dy \quad (12)$$

where c_1 and c_2 are the two unknown constants, $I(x, y)$ is the image pixel intensity with the spatial coordinate (x, y) , Ω denotes the domain to be segmented, and $H(\cdot)$ is the Heaviside step function. By setting the partial derivatives with respect to the unknown constants c_1, c_2 , and ϕ to be zeros and updating them recursively, the minimization problem can be solved. More detailed introductions about CV segmentation and implementation can be seen in [69].

C. Bridge Extraction

For bridge extraction, we start from determining the centroid position of each river segment in the obtained river binary mask. Based on this, a graph of the centroids can be built. By exploiting the minimum spanning tree, each pair of river segments with bridges located in between can be determined.

Then, we use an α -shaped contour approximation [70] to obtain the contours of river segments as shown in Fig. 5 (top)



Fig. 5. (Top) α -shaped contours of the river segments example, and each color corresponds to each river segment. (Bottom) Proposed parametric bridge edge representation of the top bridge in (top). Given the α -shaped contours, each line segment (red line) on the bridge edges can be parametrically modeled by the normal direction \mathbf{d} (black line with arrow) and the coordinate \mathbf{c} of the middle point (black point).

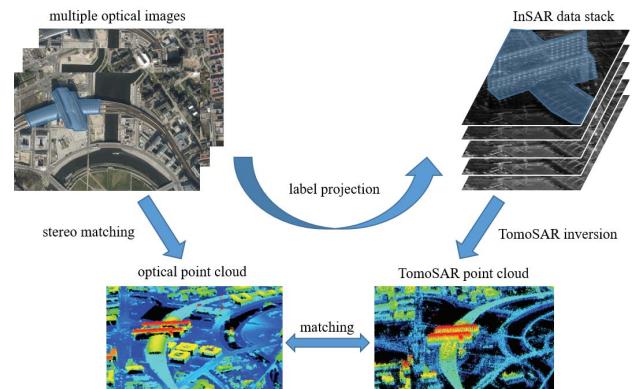


Fig. 6. SARptical processing steps as introduced in Section IV.

and each small piecewise line segment on the α -shaped contour can be represented in a parametric way. The parameters we choose, here, are the normal direction \mathbf{d} and the coordinate \mathbf{c} of the middle point of the line segment. As a result,

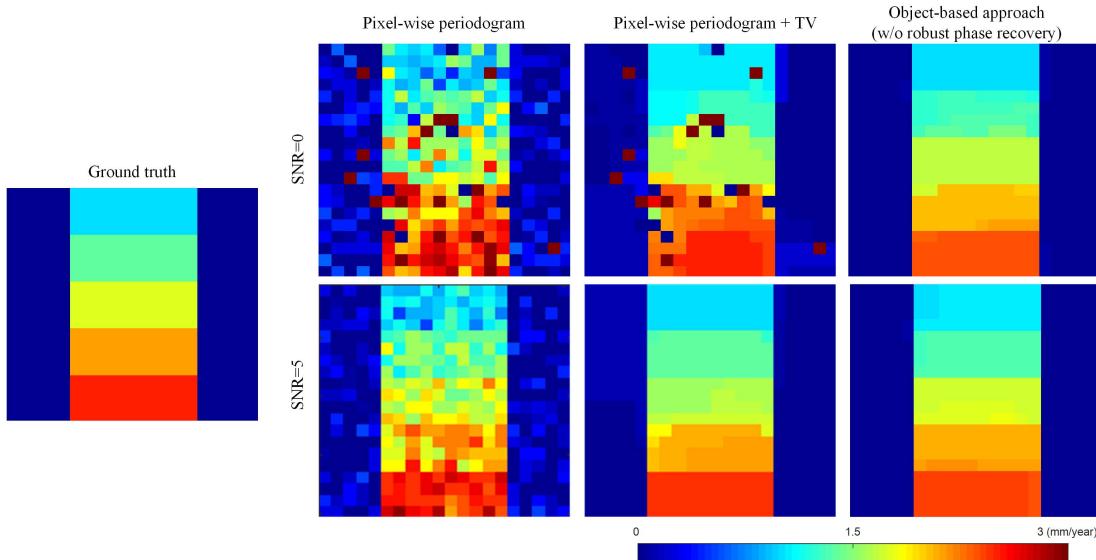


Fig. 7. Simulated ground truth linear deformation rate and the estimations obtained by comparing three methods in the case of two SNRs (0 and 5 dB). We simulate a multipass InSAR stack of 20 SAR images with the ground truth linear deformation rate shown in (left) ranging from 1 to 2.5 (mm/year). The spatial baseline is set to be comparable to that of TerraSAR-X and the temporal baseline is set to have regular spacing from 0 to 5 years. (Right) Deformation estimations by the three comparing methods, i.e., the pixelwise periodogram, the pixelwise periodogram + TV refers to a direct TV filtering on the pixelwise periodogram estimates. The deformation pattern cannot be exactly revealed by the pixelwise periodogram method. Although TV filtering postprocessing can improve the result, the performance is severely influenced by large bias existing in the periodogram result, especially in the case of low SNR (0 dB). As a comparison, our proposed approach can achieve much more reliable results than the others, especially in the case of low SNR. A quantitative comparison of the proposed approach with the other two is listed in Table II.

the bridge edges can be retrieved based on the following assumptions.

- 1) The pair of line segments is parallel or approximately parallel.
- 2) The pair of line segments is relatively close to each other.

A similarity function expressed in (13) is introduced to incorporate the two above-mentioned criteria

$$\begin{aligned} sim_{i,j} = \exp \left(- \left| \frac{\langle \mathbf{d}_1(i), \mathbf{d}_2(j) \rangle}{\|\mathbf{d}_1(i)\| \|\mathbf{d}_2(j)\|} \right| \right) \\ \times \exp \left(\frac{\|\mathbf{c}_1(i) - \mathbf{c}_2(j)\|_2^2}{\max(\|\mathbf{c}_1(i) - \mathbf{c}_2(j)\|_2^2)} \right) \quad (13) \end{aligned}$$

where $\mathbf{d}_1(i)$ denotes the normal direction of the i th line piece of river segment 1, $\mathbf{d}_2(j)$ is the normal direction of the j th line piece of river segment 2, and the first exponential term measures the normalized similarity of the normal directions of the two line pieces. Similarly, $\mathbf{c}_1(i)$ represents the coordinate vector of the middle point of the i th line piece of river segment 1, $\mathbf{c}_2(j)$ denotes the coordinate vector of the middle point of the j th line piece of river segment 2, $\|\mathbf{c}_1(i) - \mathbf{c}_2(j)\|_2^2$ is the Euclidean distance of the two points, and the second exponential term in (13) calculates the distance similarity normalized by the maximum distance of the two line pieces from the pairwise river segment. Fig. 5 (bottom) shows the parametric bridge edge representation proposed in this section. Finally, based on the similarity function defined in (13), the pair of line pieces under a certain similarity threshold th_{sim} can be selected as the bridge edges.

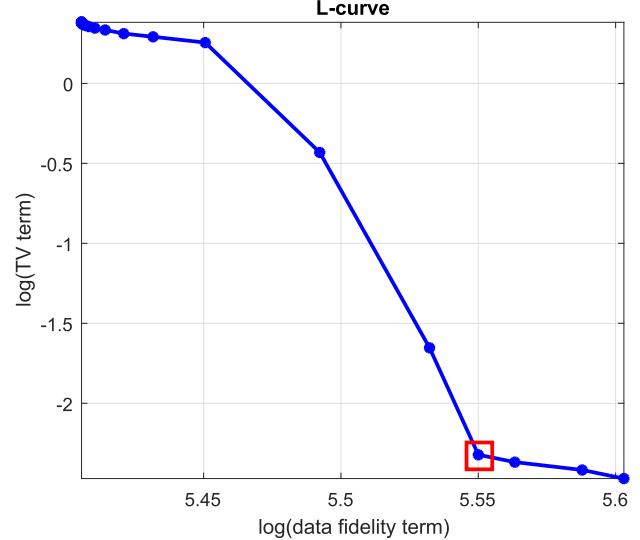


Fig. 8. L-curve plot (in log scale) of the regularization term with respect to the data fidelity term in (5) for a range of η . The optimal η can be automatically chosen by finding the maximum curvature of the curve. In this simulation (SNR = 0 dB), η is accordingly selected as 350 (red rectangle).

IV. SARPTICAL PROCESSING

The bridge masks derived from optical images are projected to SAR images using the SARptical method [1]–[3]. SARptical links the pixels between SAR images and the corresponding optical images. A schematic of SARptical is shown in Fig. 6. The basic idea is to coregister the 3-D models independently derived from two data sources, in order to establish a link between the 2-D SAR and optical images. Its key steps are briefly listed as follows.

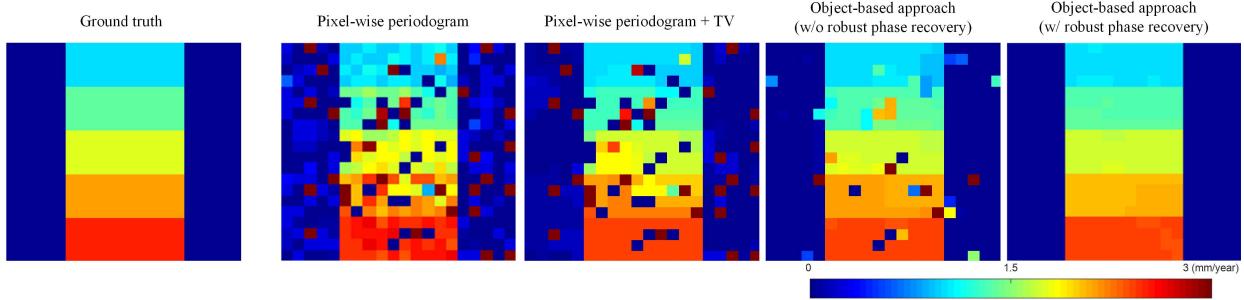


Fig. 9. Comparisons of the performances of the four approaches, i.e., the pixelwise periodogram, the pixelwise periodogram + TV, and the proposed approach without and with robust phase recovery, at 5-dB SNR and 20% outliers. (Left) Similar ground truth InSAR stack simulation as in Fig. 7. (Right) Linear deformation rate estimations by the four comparing approaches. Since the estimation performance of the pixelwise periodogram is severely influenced by SNR, the amount of outliers, and the number of the reconstructed SAR images, its result contains both the noise and the outliers. The TV filtering can greatly mitigate its noise, but the outliers still remain in the result, since it tends to preserve sparse outliers. Consistent with the above-mentioned analysis, the object-based approach without robust phase recovery does outperform the former two methods. However, some outliers still persist in its result. In contrast, the approach with the robust phase recovery can better remove the outliers and achieves a much more reliable result than the others.

TABLE II
NUMERICAL PERFORMANCE FOR THE RESULTS SHOWN IN FIG. 7

		pixel-wise periodogram	pixel-wise periodogram + TV	the proposed approach
Standard Deviation (SD) [mm/year]	SNR = 0dB	2.64	2.39	6.32×10^{-2}
	SNR = 5dB	2.31×10^{-1}	6.72×10^{-2}	3.94×10^{-2}

- 1) Retrieve 3-D point cloud from SAR image stacks, i.e., using PSI or TomoSAR, not necessarily object-based. In this paper, it has been done by Tomo-GENESIS—DLR’s tomographic SAR processing system [71], [72].
- 2) Retrieve 3-D point cloud from optical images using stereo matching.
- 3) Coregister the TomoSAR and optical point clouds, and correct the camera positions of the optical images with respect to the TomoSAR 3-D point cloud.
- 4) Estimate an average height of the target object label from the matched point cloud.
- 5) Project the object label to SAR image coordinate (range-azimuth).

V. EXPERIMENTAL RESULTS

A. Simulated Data

We simulate a multipass InSAR stack with the ground truth spatial linear deformation rate ranging from 1 to 2.5 (mm/year), as shown in Fig. 7 (left). We choose a spatial baseline comparable to that of TerraSAR-X and a temporal baseline with regular spacing from 0 to 5 years. The number of SAR images is 20. Uncorrelated complex circular Gaussian noise is added to the simulated stack with an SNR of 0 and 5 dB, respectively. Since outliers are not simulated in this experiment, the proposed approach without the robust phase recovery is employed.

The performance is compared with those obtained by the pixelwise periodogram and the pixelwise periodogram + TV filtering, which is shown in Fig. 7. Here, the pixelwise periodogram + TV refers to a direct TV filtering [73] on the pixelwise periodogram estimates instead of optimizing the periodogram and the TV term jointly, as done in the proposed algorithm. As shown from the results, without considering

the prior knowledge of the deformation rates along spatial directions, the pixelwise periodogram result is much noisier than the results of the other two. Especially, in the case of the lower SNR (0 dB), one cannot exactly interpret the whole deformation pattern. Large bias (showing as deep red/blue outliers) often appears in the estimates, as the periodogram is only asymptotically optimal. What is more interesting is the comparison between the proposed method and the pixelwise periodogram + TV. The results show that applying TV filtering afterward can indeed achieve a certain level of denoising. However, it cannot mitigate the large bias. The reason is that the TV filtering is not a robust image denoising method, since minimizing the L1 norm of the TV term tends to keep sparse outliers in the image. The detailed quantitative comparisons of the performances of the three above-mentioned algorithms can be seen in Table II. Consistent with the above-mentioned analysis, in the case of low SNR (0 dB), the performance of separate TV filtering is almost the same as that of the pixelwise periodogram, due to the dominating large bias in the result. Besides, we have also tested the efficiency of the proposed estimator at different object sizes (15×15 , 20×20 , and 25×25 pixels, not shown in the paper). No significant difference in the efficiency is found. Their standard deviations are all around 4×10^{-2} (mm/year) under the same noise level (SNR = 5 dB).

The parameter η in (5) can be determined by the “L-curve” method, which is a plot of the regularization term with respect to the data fidelity term for a range of regularization parameters [74]. Generally, the L-curve is made up of a flat part and a steep part. The optimal parameter is selected as the corner point of the L-curve, which provides the maximum curvature of the curve. Concretely, we utilize the above-mentioned simulation with SNR = 0 dB and plot its L-curve (in log scale) shown in Fig. 8. Accordingly, the

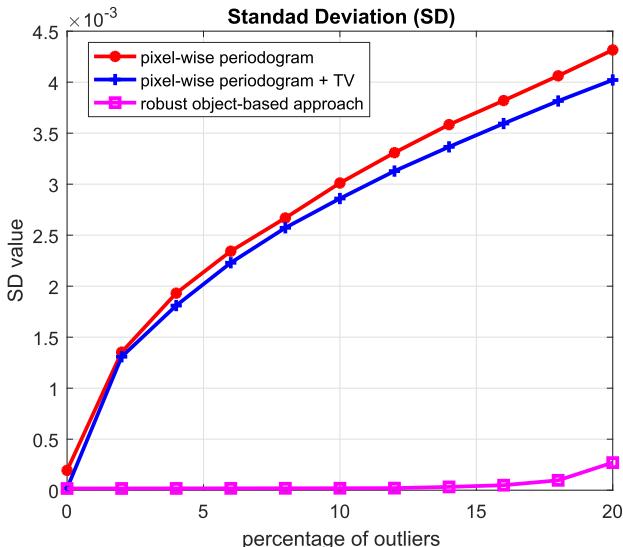


Fig. 10. Comparisons of the standard deviation of the linear deformation rate estimated by the pixelwise periodogram, the pixelwise periodogram + TV, and the robust object-based algorithm at different percentages of the outliers. The estimation performances of the former two methods are severely influenced by the existing outliers, while our proposed method can robustly reconstruct the result with small errors.

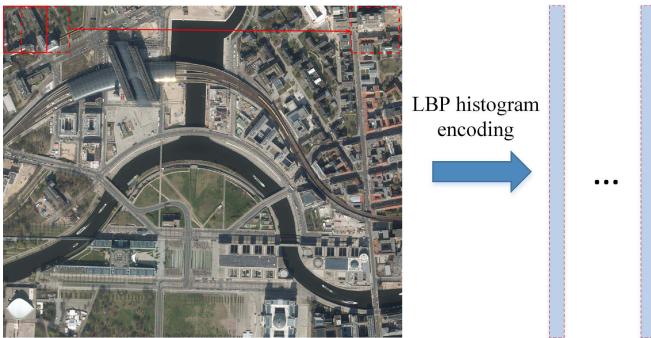


Fig. 11. Optical image of our study area with the size of 980×1180 and a pixel spacing of 0.2 m/pixel . We manually select dozens of training patches (size 50×50) encoded by LBP histograms and we train a binary classifier by linear SVM. The river classification on the whole test image is then achieved in a 50×50 sliding window manner.



Fig. 12. (Left) Coarse river segments result obtained by a trained binary linear SVM. (Right) Refined river segments by CV segmentation.

optimal η can be determined as 350. Nevertheless, the reasonable range of η is not very restricted (about tens to hundred). Experiments show that the standard deviations of the estimates are in the same order of magnitude for $\eta = 250$ and $\eta = 300$ at SNR = 5 dB.

Moreover, to investigate the proposed object-based deformation reconstruction with the robust phase recovery, we first add complex Gaussian noise to the ground truth phase tensor with

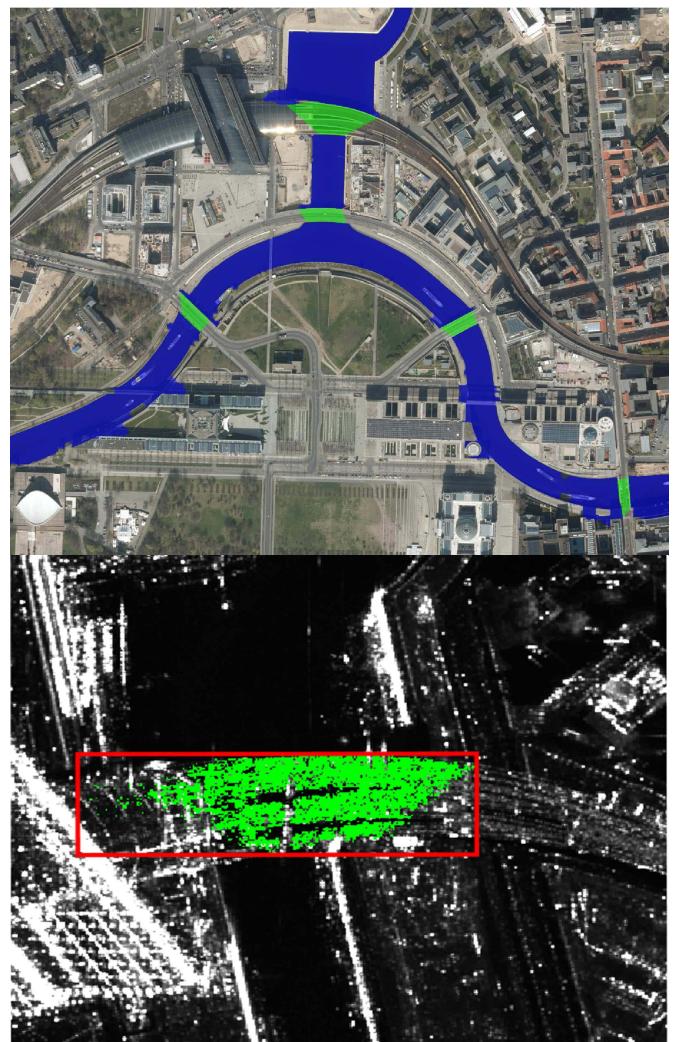


Fig. 13. (Top) Classifications of the rivers and bridges by the proposed approach covered with blue and green masks, respectively. As we can see from the result, some building shadows are also classified as rivers, since they share similar RGB values with those of rivers. Some bridges do show irregular shapes, especially the top one, since the bridge mask depends on the boundary of the river segments. Yet, this does not affect the bridge monitoring too much, since the bridge masks cover most parts of the bridges. (Bottom) Bridge mask (green) in the SAR image obtained by projecting the corresponding mask from the optical image (SARptical), which is the top bridge shown in (top), and the red rectangular area is used for the robust object-based deformation reconstruction.

SNR = 5 dB. Outliers are then simulated by replacing 20% randomly selected pixels with uniformly distributed phases, i.e., completely complex Gaussian noise without signals. In Fig. 9, we show the deformation estimates of the four methods, i.e., the pixelwise periodogram, the pixelwise periodogram + TV, and the proposed approach without and with the robust phase recovery. For the parameters of the proposed approaches, we set η and γ as 200 and 0.7, respectively. We can see from Fig. 9 that the estimations of the pixelwise periodogram contain both the noise and outliers, because its performance is severely influenced by the SNR, the amount of outliers, and the number of SAR images. Consistent with the above-mentioned analysis, the TV filtering keeps those sparse outliers in its result. The object-based approach without robust

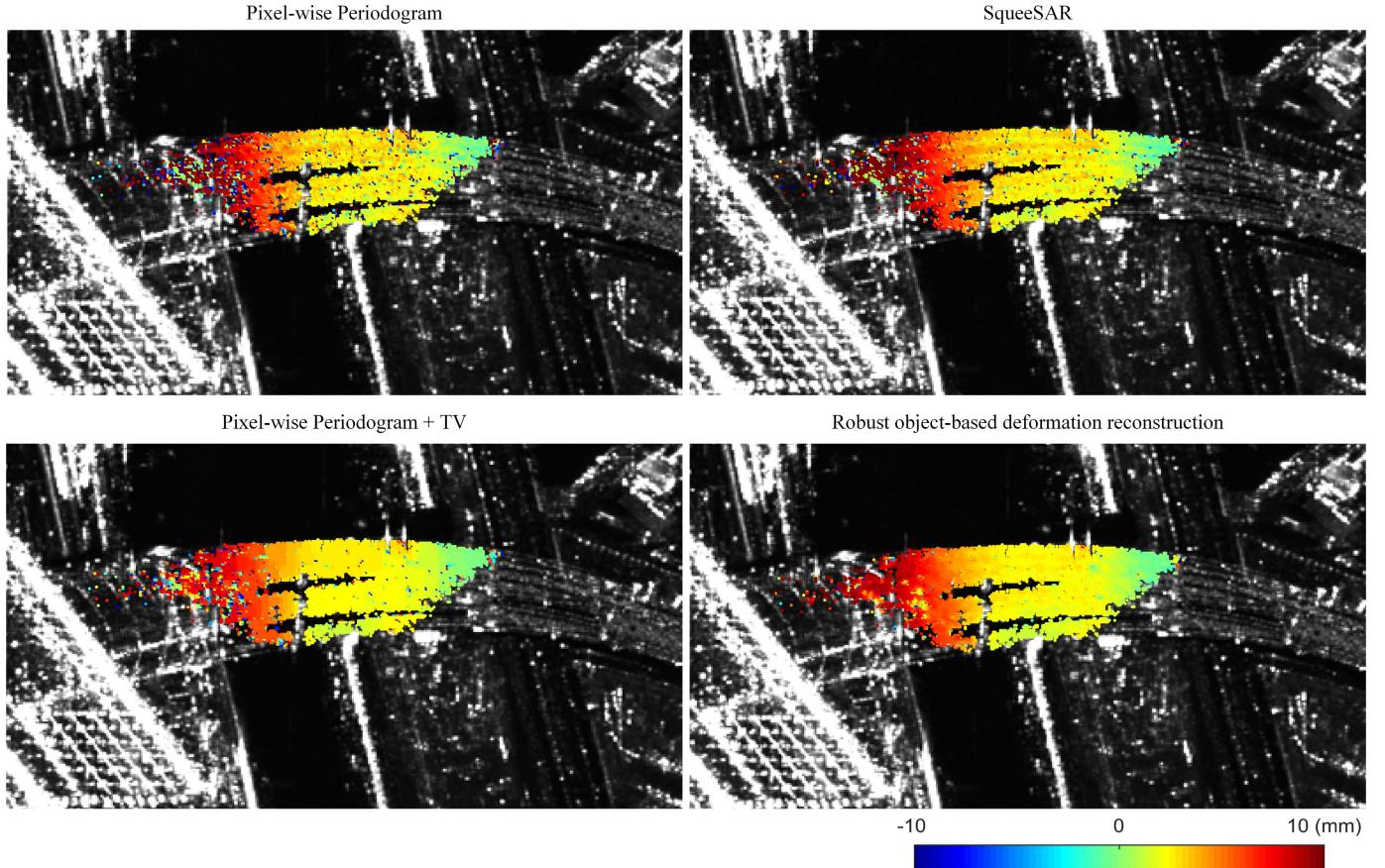


Fig. 14. Seasonal periodic motion amplitudes of the classified bridge area shown in Fig. 13 (bottom) reconstructed by the (top left) pixelwise periodogram, (top right) SqueeSAR, the (bottom left) pixelwise periodogram + TV, (bottom right) and the proposed robust object-based deformation reconstruction method. The pixelwise periodogram result shows the noisier estimates of the amplitude of seasonal motion than the other methods. SqueeSAR utilizes the similar statistic behaviors of multiple pixels, which indeed obtains the less noisy result by averaging those pixels, especially in the left part of the bridge. However, some prominent noise still persists, without exploiting the geometric or semantic information to jointly reconstruct the deformations. Both the pixelwise periodogram + TV and the proposed method incorporate geometric information. Consistent with the analysis in Section V-A, the separate TV filtering does filter out the majority of the noise, but it heavily depends on the results of the pixelwise periodogram, which can be corrupted by outliers. Thereby, some large outliers still exist in the left part of the bridge. Separate TV filtering also tends to oversmooth and underestimate the deformation pattern. As a comparison, our proposed approach can both mitigate the noise and the outliers, which outperforms the other methods.

phase recovery does outperform the former two methods. However, some outliers still persist in its result. In contrast, the approach with the robust phase recovery can better remove the outliers and achieve much more reliable result than the others. We also perform the numerical analysis through Monte Carlo experiments to compare the three methods at different percentages of outliers, as shown in Fig. 10. It is demonstrated that the accuracies of the reconstructions by the pixelwise periodogram and the pixelwise periodogram + TV are very sensitive to any percentage of outliers, while the proposed algorithm can robustly reconstruct the deformation rates with little influence from the outliers.

B. Real Data

1) Bridge Detection: In this test, we take the central area in Berlin, which contains the target bridges for the following object-based deformation reconstruction. As shown in Fig. 11, the size of our study area is 980×1180 with a pixel spacing of 0.2 m. We manually select dozens of training patches (size 50×50) encoded by LBP histograms. A binary classifier

(river versus background) is then trained by linear SVM. The river classification on the whole test image is then achieved in a 50×50 sliding window manner. The classified river segments are shown in Fig. 12 (left). For the refinement of the classified river segments, the iteration number of CV segmentation is set to be 150. This parameter influences the boundary of the segments, larger number may induce overflow of segmentation, otherwise, insufficient segmentation. The refined river segments are demonstrated in Fig. 12 (right). The corresponding river contours are obtained by α -shape with the probe radius $R_\alpha = 50$. The similarity threshold for bridge edge detection is 0.3. As shown in Fig. 13 (top), the rivers and the bridges are classified by the proposed approach and covered with blue and green masks, respectively. As we can see from the result, some building shadows are also classified as rivers, since they share similar RGB values with those of rivers. Some bridges do show irregular shapes, especially the top one, since the bridge mask depends on the boundary of the river segments. Yet, this does not affect the bridge monitoring too much, since the bridge masks cover most parts of the bridges.



Fig. 15. Estimated amplitudes of the seasonal motion of all the bridges in central Berlin by the proposed robust object-based deformation reconstruction. Except for the top one, the motions of all other bridges are not significant enough to draw any obvious conclusions. Interesting to note, however, is that the motion of the leftmost bridge tends to be increasing from the left side to the right side. Its corresponding orthorectified optical image, with a pixel spacing of 7 cm, is shown to the left (image is provided by the DLR Institute of Robotics and Mechatronics produced by semiglobal matching [75]). As shown by the two highlighted positions of the red ellipses, we can see that the bridge is separated from the roads. In this case, there may be one reason that the motion allowances of the bridge on the two sides are different, i.e., its right side is higher than the left.

2) *Object-Based Deformation Retrieval:* Given the bridge masks obtained in the optical image, the corresponding bridge areas in the SAR image can be determined by the SARptical work [1]–[3]. For example, the top bridge mask in Fig. 13 (top) projected in the SAR image is shown in Fig. 13 (bottom). The area extracted for the deformation reconstruction is chosen by the bounding box of the bridge mask, as indicated by the red rectangle. The dimension of this data set is 243×66 pixels with 109 SAR images in total. This area undergoes a seasonal periodic motion that is primarily caused by the thermal dilation of the steel railways on the bridge. The estimated amplitudes of this periodic motion using the proposed method and the other state-of-the-art methods are shown in Fig. 14. The result obtained by the pixelwise periodogram is at the top-left. In the following, it is the result from the pixelwise periodogram + TV. The SqueeSAR result is located at the top-right, where the searching window size for the neighborhood pixels is 15×15 pixels, which corresponds to at most 225 statistically homogeneous samples per pixel for this data set, and the result of the proposed approach is demonstrated at the bottom-right, where η is 1000 and γ is 0.45.

The pixelwise periodogram result shows the noisier estimates of the amplitude of seasonal motion than the other methods. SqueeSAR utilizes the similar statistic behaviors of multiple pixels, which indeed obtains the less noisy result by averaging those pixels, especially in the left part of the bridge. However, some prominent noise still persists, without exploiting the geometric or semantic information to jointly reconstruct the deformations. Both the pixelwise periodogram + TV and the proposed method incorporate geometric information. Consistent with the analysis in Section V-A, the separate TV filtering does filter out the majority of the noise, but it heavily depends on the results of the pixelwise periodogram, which

can be corrupted by outliers. Thereby, some large outliers still exist in the left part of the bridge. The separate TV filtering also tends to oversmooth and underestimate the deformation pattern. As a comparison, our proposed approach can both mitigate the noise and the outliers, which outperforms the other methods.

The estimated amplitudes of the seasonal motion of all the bridges in central Berlin are shown in Fig. 15. Except for the top one, the motions of all other bridges are not significant enough to draw any obvious conclusions. Interesting to note, however, is that the motion of the leftmost bridge tends to be increasing from the left side to the right side. Its corresponding orthorectified optical image, with a pixel spacing of 7 cm, is shown at the left (image is provided by the DLR Institute of Robotics and Mechatronics produced by semiglobal matching [75]). As shown by the two highlighted positions with red ellipses, we can see that the bridge is separated from the roads. In this case, there may be one reason that the motion allowances of the bridge on the two sides are different, i.e., its right side is higher than the left.

VI. CONCLUSION

In this paper, we proposed a general framework for object-based multipass InSAR parameter estimation, i.e., introducing a spatial regularization term based on given object labels. Moreover, considering the significant amount of outliers existing in real data, we develop a robust InSAR stack filtering approach by minimizing the rank of the InSAR stack tensor. To demonstrate the application of the proposed method for bridge monitoring, we also propose a bridge detection method in optical images.

The simulation demonstrates that the proposed approach has an outstanding performance. In the outlier-free case,

it outperforms the pixelwise approach by a factor of 40 at an SNR of 0 dB, in terms of the standard deviation of the estimates. The experiments show that the regularization parameter η does not sensitively influence the efficiency of the reconstruction result, especially at the high SNR (5 dB). Its optimal value can be automatically determined by the “L-curve” method within a given range of η . This range setting depends on the data set. We find it to be 100–350 in our simulated experiments. Besides, any postprocessing based on the pixel-wise method, such as TV filtering on the result, can still carry large bias to the final result. The joint optimization of the data fidelity and the regularization term is superior to the former approaches by simultaneously balancing the two terms. It achieves a better efficiency and detailed preservation, which in turn reduces the number of images required for a reliable estimation. Experiments also demonstrate that the efficiency of the joint estimation does not heavily depend on the object size. The accuracies of the estimates under different object sizes (15×15 , 20×20 , and 25×25) show similar orders of magnitude (around 4×10^{-2} mm/year at 5-dB SNR). That is to say, the object size can be kept small for large area processing, in order to increase the parallelization of the processing.

The proposed approach with robust phase recovery is also proven to be effective against outliers. It outperforms the nonrobust pixelwise approach by a factor of 20 in terms of the standard deviation of the estimates, at 5-dB SNR with an outliers percent of 20%. Also, it achieves more reliable results than the one without the robust phase recovery step. Therefore, in case outliers exist, this step is necessary for the object-based deformation reconstruction.

Furthermore, in the river segmentation part, we find that some building shadows are wrongly classified as rivers, due to the reason that the shadows share similar RGB values with rivers. It would be better to not only consider exploiting color but also some other features, such as texture, for the river classifications.

To summarize, the proposed robust object-based approach is a novel framework that combines geometric information and multipass InSAR methods. It is suited for areas with homogenous pixels, such as SqueeSAR, as well as for urban areas where the pixels are highly nonergodic. The proposed approach can be efficiently solved by smart optimization methods, which renders it suitable for operational processing. It also demonstrates the advantage of fusing optical and SAR images, especially for the monitoring of urban areas. For the future development of multipass InSAR algorithms, we shall see a growing relationship of multisensory data [76], e.g., fusion of multispectral and SAR images, and joint reconstruction using SAR images with different resolutions.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their valuable suggestions.

REFERENCES

- [1] Y. Wang, X. X. Zhu, B. Zeisl, and M. Pollefeyns, “Fusing meter-resolution 4-D InSAR point clouds and optical images for semantic urban infrastructure monitoring,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 14–26, Jan. 2017.
- [2] Y. Wang and X. X. Zhu, “InSAR forensics: Tracing InSAR scatterers in high resolution optical image,” in *Proc.FRINGE WORKSHOP*, 2015, pp. 1–8.
- [3] Y. Wang and X. X. Zhu, “Semantic fusion of SAR interferometry and optical image with application to urban infrastructure monitoring,” *ISPRS-Int. Arch. Photogram., Remote Sens. Spatial Inf. Sci.*, vol. XL-3/W3, pp. 153–160, 2015, doi: 10.5194/isprsarchives-XL-3-W3-153-2015.
- [4] A. Ferretti, A. Fumagalli, F. Novali, C. Prati, F. Rocca, and A. Rucci, “A new algorithm for processing interferometric data-stacks: SqueeSAR,” *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 9, pp. 3460–3470, Sep. 2011.
- [5] A. Ferretti, C. Prati, and F. Rocca, “Permanent scatterers in SAR interferometry,” *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 1, pp. 8–20, Jan. 2001.
- [6] N. Adam, B. Kampes, M. Eineder, J. Worawattanamateekul, and M. Kircher, “The development of a scientific permanent scatterer system,” in *Proc. ISPRS Workshop High Resolution Mapping Space*, Hanover, Germany, 2003, p. 6.
- [7] G. Fornaro, A. Pauciullo, and F. Serafino, “Deformation monitoring over large areas with multipass differential SAR interferometry: A new approach based on the use of spatial differences,” *Int. J. Remote Sens.*, vol. 30, no. 6, pp. 1455–1478, 2009.
- [8] J. J. Sousa, A. J. Hooper, R. F. Hanssen, L. C. Bastos, and A. M. Ruiz, “Persistent Scatterer InSAR: A comparison of methodologies based on a model of temporal deformation vs. spatial correlation selection criteria,” *Remote Sens. Environ.*, vol. 115, no. 10, pp. 2652–2663, 2011.
- [9] S. Gernhardt and R. Bamler, “Deformation monitoring of single buildings using meter-resolution SAR data in PSI,” *ISPRS J. Photogramm. Remote Sens.*, vol. 73, pp. 68–79, Sep. 2012.
- [10] M. K. Bert, *Radar Interferometry: Persistent Scatterer Technique*. Dordrecht, The Netherlands: Springer, 2006.
- [11] Y. Wang, X. X. Zhu, and R. Bamler, “An efficient tomographic inversion approach for urban mapping using meter resolution SAR image stacks,” *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 7, pp. 1250–1254, Jul. 2014.
- [12] M. Costantini, S. Falco, F. Malvarosa, F. Minati, F. Trillo, and F. Vecchioli, “Persistent scatterer pair interferometry: Approach and application to COSMO-SkyMed SAR data,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 7, pp. 2869–2879, Jul. 2014.
- [13] G. Fornaro, F. Serafino, and F. Soldovieri, “Three-dimensional focusing with multipass SAR data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 3, pp. 507–517, Mar. 2003.
- [14] F. Lombardini, “Differential tomography: A new framework for SAR interferometry,” *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 1, pp. 37–44, Jan. 2005.
- [15] X. X. Zhu and R. Bamler, “Very high resolution spaceborne SAR tomography in urban environment,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 12, pp. 4296–4308, Dec. 2010.
- [16] X. X. Zhu and R. Bamler, “Let’s do the time warp: Multicomponent nonlinear motion estimation in differential SAR tomography,” *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 735–739, Jul. 2011.
- [17] D. Reale, G. Fornaro, A. Pauciullo, X. Zhu, and R. Bamler, “Tomographic imaging and monitoring of buildings with very high resolution SAR data,” *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 661–665, Jul. 2011.
- [18] G. Fornaro, F. Lombardini, A. Pauciullo, D. Reale, and F. Viviani, “Tomographic processing of interferometric SAR data: Developments, applications, and future research perspectives,” *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 41–50, Jul. 2014.
- [19] X. X. Zhu and R. Bamler, “Tomographic SAR inversion by L_1 -norm regularization—The compressive sensing approach,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3839–3846, Oct. 2010.
- [20] A. Budillon, A. Evangelista, and G. Schirinzi, “Three-dimensional SAR focusing from multipass signals using compressive sampling,” *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 488–499, Jan. 2011.
- [21] X. X. Zhu and R. Bamler, “Superresolving SAR tomography for multidimensional imaging of urban areas: Compressive sensing-based TomoSAR inversion,” *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 51–58, Jul. 2014.
- [22] M. Eineder, C. Minet, P. Steigenberger, X. Cong, and T. Fritz, “Imaging geodesy—Toward centimeter-level ranging accuracy with TerraSAR-X,” *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 661–671, Feb. 2011.
- [23] X. X. Zhu, S. Montazeri, C. Gisinger, R. F. Hanssen, and R. Bamler, “Geodetic SAR tomography,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 18–35, Jan. 2016.

- [24] Y. Wang, X. X. Zhu, and R. Bamler, "Retrieval of phase history parameters from distributed scatterers in urban areas using very high resolution SAR data," *ISPRS J. Photogramm. Remote Sens.*, vol. 73, pp. 89–99, Sep. 2012.
- [25] K. Goel and N. Adam, "An advanced algorithm for deformation estimation in non-urban areas," *ISPRS J. Photogramm. Remote Sens.*, vol. 73, pp. 100–110, Sep. 2012.
- [26] S. Samiei-Esfahany, J. E. Martins, F. van Leijen, and R. F. Hanssen, "Phase estimation for distributed scatterers in InSAR stacks using integer least squares estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5671–5687, Oct. 2016.
- [27] N. Cao, H. Lee, and H. C. Jung, "A phase-decomposition-based PSInSAR processing method," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 2, pp. 1074–1090, Feb. 2016.
- [28] G. Fornaro, S. Verde, D. Reale, and A. Pauciullo, "CAESAR: An approach based on covariance matrix decomposition to improve multibaseline–multitemporal interferometric SAR processing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2050–2065, Apr. 2015.
- [29] M. Neumann, L. Ferro-Famil, and A. Reigber, "Estimation of forest structure, ground, and canopy layer characteristics from multibaseline polarimetric interferometric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1086–1104, Mar. 2010.
- [30] S. Tebaldini, "Single and multipolarimetric SAR tomography of forested areas: A parametric approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2375–2387, May 2010.
- [31] C.-A. Deledalle, L. Denis, and F. Tupin, "NL-InSAR: Nonlocal interferogram estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 4, pp. 1441–1452, Apr. 2011.
- [32] X. X. Zhu, R. Bamler, M. Lachaise, F. Adam, Y. Shi, and M. Eineder, "Improving TanDEM-X DEMs by non-local InSAR filtering," in *Proc. 10th Eur. Conf. Synth. Aperture Radar (EUSAR)*, 2014, pp. 1–4.
- [33] C.-A. Deledalle, L. Denis, G. Poggi, F. Tupin, and L. Verdoliva, "Exploiting patch similarity for SAR image processing: The nonlocal paradigm," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 69–78, Jul. 2014.
- [34] F. Sica, D. Reale, G. Poggi, L. Verdoliva, and G. Fornaro, "Nonlocal adaptive multilooking in SAR multipass differential interferometry," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 4, pp. 1727–1742, Apr. 2015.
- [35] X. X. Zhu, N. Ge, and M. Shahzad, "Joint sparsity in SAR tomography for urban mapping," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 8, pp. 1498–1509, Dec. 2015.
- [36] R. J. Dekker, "Texture analysis and classification of ERS SAR images for map updating of urban areas in The Netherlands," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1950–1958, Sep. 2003.
- [37] C. Tison, F. Tupin, and H. Maître, "A fusion scheme for joint retrieval of urban height map and classification from high-resolution interferometric SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 2, pp. 496–505, Feb. 2007.
- [38] C. Tison, J. M. Nicolas, F. Tupin, and H. Maître, "A new statistical model for Markovian classification of urban areas in high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2046–2057, Oct. 2004.
- [39] P. Gamba, B. Houshmand, and M. Saccani, "Detection and extraction of buildings from interferometric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 1, pp. 611–617, Jan. 2000.
- [40] A. Voisin, V. A. Krylov, G. Moser, S. B. Serpico, and J. Zerubia, "Classification of very high resolution SAR images of urban areas using copulas and texture in a hierarchical Markov random field model," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 1, pp. 96–100, Jan. 2013.
- [41] A. Thiele, E. Cadario, K. Schulz, U. Thonnessen, and U. Soergel, "Building recognition from multi-aspect high-resolution InSAR data in urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, pp. 3583–3593, Nov. 2007.
- [42] J. D. Wegner, J. R. Ziehn, and U. Soergel, "Combining high-resolution optical and InSAR features for height estimation of buildings with flat roofs," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5840–5854, Sep. 2014.
- [43] M. Shahzad and X. X. Zhu, "Automatic detection and reconstruction of 2-D/3-D building shapes from spaceborne TomoSAR point clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1292–1310, Mar. 2016.
- [44] A. De Maio, G. Fornaro, and A. Pauciullo, "Detection of single scatterers in multidimensional SAR imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2284–2297, Jul. 2009.
- [45] D. C. Rife and R. R. Boorstyn, "Single tone parameter estimation from discrete-time observations," *IEEE Trans. Inf. Theory*, vol. 20, no. 5, pp. 591–598, Sep. 1974.
- [46] Y. Wang and X. X. Zhu, "Robust estimators for multipass SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 2, pp. 968–980, Feb. 2016.
- [47] A. Cichocki *et al.*, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, Mar. 2015.
- [48] M. Zhu and T. Chan, "An efficient primal-dual hybrid gradient algorithm for total variation image restoration," *Dept. Math., UCLA CAM Rep.*, pp. 8–34, 2008.
- [49] M. Tao and J. Yang, "Alternating direction algorithms for total variation deconvolution in image reconstruction," *Dept. Math., Nanjing Univ., Nanjing, China, Tech. Rep. TR0918*, 2009.
- [50] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imag. Vis.*, vol. 40, no. 1, pp. 120–145, 2011.
- [51] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "L-BFGS-B: Fortran subroutines for large-scale bound constrained optimization," *Dept. Elect. Eng. Comput. Sci., Northwestern Univ., Evanston, IL, USA, Tech. Rep. NAM-11*, 1994.
- [52] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization," *ACM Trans. Math. Softw.*, vol. 23, no. 4, pp. 550–560, 1997.
- [53] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM J. Sci. Comput.*, vol. 16, no. 5, pp. 1190–1208, 1995.
- [54] P. Ochs, Y. Chen, T. Brox, and T. Pock, "iPiano: Inertial proximal algorithm for nonconvex optimization," *SIAM J. Imag. Sci.*, vol. 7, no. 2, pp. 1388–1419, 2014.
- [55] G. Vaksman, M. Zibulevsky, and M. Elad, "Patch ordering as a regularization for inverse problems in image processing," *SIAM J. Imag. Sci.*, vol. 9, no. 1, pp. 287–319, 2016.
- [56] L. Wei, "Quantification of low-rank information in multi-temporal SAR datasets," M.S. thesis, Signal Process. Earth Observ., Tech. Univ. Munich, Munich, Germany, 2015.
- [57] D. Goldfarb and Z. Qin, "Robust low-rank tensor recovery: Models and algorithms," *SIAM J. Matrix Anal. Appl.*, vol. 35, no. 1, pp. 225–253, 2014.
- [58] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
- [59] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [60] S. Gandy, B. Recht, and I. Yamada, "Tensor completion and low-n-rank tensor recovery via convex optimization," *Inv. Prob.*, vol. 27, no. 2, p. 025010, 2011.
- [61] G. Mátyus, S. Wang, S. Fidler, and R. Urtasun, "Enhancing road maps by parsing aerial images around the world," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1689–1697.
- [62] J. Liu and Y. Liu, "Local regularity-driven city-scale facade detection from aerial images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3778–3785.
- [63] D. Marmanis, F. Adam, M. Datcu, T. Esch, and U. Stilla, "Deep neural networks for above-ground detection in very high spatial resolution digital elevation models," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 2, no. 3, pp. 103–110, 2015.
- [64] G. Cheng and J. Han, "A survey on object detection in optical remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 117, pp. 11–28, Jul. 2016.
- [65] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, 1996.
- [66] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
- [67] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.
- [68] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton–Jacobi formulations," *J. Comput. Phys.*, vol. 79, no. 1, pp. 12–49, 1988.
- [69] P. Getreuer, "Chan–Vese segmentation," *Image Process. Line*, vol. 2, pp. 214–224, Aug. 2012.

- [70] H. Edelsbrunner, D. G. Kirkpatrick, and R. Seidel, "On the shape of a set of points in the plane," *IEEE Trans. Inf. Theory*, vol. 29, no. 4, pp. 551–559, Jul. 1983.
- [71] X. X. Zhu, Y. Wang, S. Gernhardt, and R. Bamler, "Tomo-GENESIS: DLR's tomographic SAR processing system," in *Proc. Joint Urban Remote Sens. Event*, Apr. 2013, pp. 159–162.
- [72] X. Zhu, *Very High Resolution Tomographic SAR Inversion for Urban Infrastructure Monitoring: A Sparse and Nonlinear Tour* (Verlag der Bayerischen Akademie der Wissenschaften), vol. 666. Munich, Germany, Deutsche Geodätische Kommission.
- [73] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 323–343, 2009.
- [74] P. C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," *SIAM J. Sci. Comput.*, vol. 14, no. 6, pp. 1487–1503, 1993.
- [75] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.
- [76] M. Schmitt and X. X. Zhu, "Data fusion and remote sensing: An ever-growing relationship," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 4, pp. 6–23, Dec. 2016.



Jian Kang (S'16) received the B.S. and M.E. degrees in electronic engineering from the Harbin Institute of Technology, Harbin, China, in 2013 and 2015, respectively. He is currently pursuing the Ph.D. degree with the Chair of Signal Processing in Earth Observation, Technical University of Munich, Munich, Germany.

His research interests include multidimensional data analysis, geophysical parameters estimation based on InSAR data, and machine learning in optical images.



Yuanyuan Wang (S'10–M'14) received the B.Eng. degree (Hons.) in electrical engineering from The Hong Kong Polytechnic University, Hong Kong, in 2008, and the M.Sc. and Dr.-Ing. degrees from the Technical University of Munich (TUM), Munich, Germany, in 2010 and 2015, respectively.

In 2014, he was a Guest Scientist with the Institute of Visual Computing, ETH Zürich, Zürich, Switzerland. He is currently with Signal Processing in Earth Observation, TUM. His research interests include optimal and robust parameters estimation in multibaseline InSAR techniques, multisensor fusion algorithms of SAR and optical data, nonlinear optimization with complex numbers, and the applications of these techniques in urban and volcanic areas.



Marco Körner (M'15) studied Computer Sciences with Psychology as minor subject and received his diploma (Dipl.-Inf.) and PhD (Dr. rer. nat.) degrees from Friedrich Schiller University of Jena, Germany, in 2009 and 2016, respectively. From 2009 to 2015, he was a member of the Computer Vision Group in Jena. Since 2015, he has been a senior researcher and deputy head at the chair of Remote Sensing Technology at Technical University of Munich (TUM), Germany. In 2012 and 2014, he was a visiting researcher at Instituto Politécnico Nacional (CIC-IPN) in Mexico City, Mexico, and the University of California in San Diego (UCSD), USA, respectively. His main research interests focus on machine learning in computer vision, particularly for application in automotive, remote sensing, and biomedical scenarios.



Xiao Xiang Zhu (S'10–M'12–SM'14) received the bachelor's degree in space engineering from the National University of Defense Technology, Changsha, China, in 2006, and the M.Sc., Dr.-Ing., and Habilitation degrees in signal processing from the Technical University of Munich (TUM), Munich, Germany, in 2008, 2011, and 2013, respectively.

Since 2011, she has been a Scientist with the Remote Sensing Technology Institute, German Aerospace Center (DLR), Cologne, Germany, where she is currently the Head of the Team Signal Analysis. Since 2013, she has also been a Helmholtz Young Investigator Group Leader and has been appointed as a TUM Junior Fellow. In 2015, she was appointed as the Professor for Signal Processing in Earth Observation with TUM. She was a Guest Scientist or a Visiting Professor with the Italian National Research Council (CNR-IREA), Naples, Italy, Fudan University, Shanghai, China, The University of Tokyo, Tokyo, Japan, and the University of California at Los Angeles, Los Angeles, CA, USA, in 2009, 2014, 2015, and 2016, respectively. Zhu is a member of the young academy at the Berlin-Brandenburg Academy of Sciences and Humanities and the German National Academy of Sciences Leopoldina and the Bavarian Academy of Sciences and Humanities, respectively. Her research interests include advanced InSAR techniques, such as high-dimensional tomographic SAR imaging and SqueezSAR, computer vision in remote sensing, including object reconstruction and multidimensional data visualization, big data analysis in remote sensing, and modern signal processing, including innovative algorithms, such as sparse reconstruction, nonlocal means filter, robust estimation, and deep learning, with applications in the field of remote sensing, such as multi/hyperspectral image analysis.

Dr. Zhu is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.