Generation of Large-Scale High Quality 3-D Urban Models

*Note: Sub-titles are not captured in Xplore and should not be used

Yilei Shi [1], Richard Bamler [1,3], Yuanyuan Wang [2] and Xiao Xiang Zhu [2,3]

[1] *Chair of Remote Sensing Technology, Technical University of Munich (TUM-LMF)*
[2] *Signal Processing in Earth Observation, Technical University of Munich (TUM-SIPEO)*
[3] *Institute of Remote Sensing Technology, German Aerospace Center (DLR-IMF)*
[1,2] Munich and [3] Weissling, Germany
yilei.shi@tum.de

***Abstract*—Interferometric synthetic aperture radar (InSAR) techniques are powerful tool for reconstructing the 3-D position of scatterers, especially for the urban areas. Since the estimation accuracy depends on the inverse of number of interferograms and signal-to-noise ratio (SNR), it is necessary to use as many as possible interferograms in order to achieve more accurate result. However, the number of interferograms of TanDEM-X data is generally limited for most areas. Therefore, in order to maintain the estimation accuracy, one feasible way is to increase the SNR. In this work, we proposed a novel framework, which integrates the non-local procedure into SAR tomography inversion and combines the robust estimation. A large-scale demonstration has been carried out with five TanDEM-X bistatic data, which covers the entire city of Munich, Germany. Quantitative evaluation of the reconstructed result with the LiDAR reference exhibits the standard deviation of the height difference is within two meters, which implies the proposed framework has great potential for high quality large-scale 3-D urban modeling.**

***Index Terms*—3-D urban models, InSAR, TomoSAR, TanDEM-X, global mapping**

I. Introduction

TanDEM-X satellite is a synthetic aperture radar (SAR) satellite, which is launched in 2010. It is still the civil and commercial SAR satellite with highest resolution. Together with its twin satellite TerraSAR-X, they flied together and formed into a helix orbit. The twin satellite constellation is aiming to generate global high-resolution digital elevation models (DEM). The two satellite can receive the echo almost same time, which enables the bistatic interferograms to avoid atomspheric and temporal decorrelation.

SAR tomography (TomoSAR) is advanced InSAR technique that is able to retrieve the 3-D position of scatterers. Many algorithms have been developed in last two decades [2] [3] [4] [5] [6]. The results indicate that TomoSAR is a promissing tool for urban reconstruction and monitering, when using different data from different sensors, such as TerraSAR-X [7] [8] or COSMO-Skymed [10]. Due to modern development in signal processing, the compressive sensing (CS) based methods like SL1MMER and others [11] [12] can achieve unprecedented accuracy for 3-D reconstruction and show the super-resolution (SR) power, which is not exhibited with conventional approaches. However, SR power is very important for distinguishing very close scatterers, which happens often in urban area, since the layover effect is dominant.

The previous mentioned InSAR algorithms, such as PSI, Capon, MUSIC, SVD-wiener or CS-based methods, ususlly needs very large InSAR stacks, e.g. empirically more than twenty interferograms, in order to achieve a reliable result. However, there is only limited number of images avaliable for global-scale. As we know from previous study [13], the estimation accuracy of TomoSAR is symptotically related to the product of signal-to-noise ratio (SNR) and number of images. Therefore, it is not suitable for our task that only micro-stacks are available for most areas.

The main goal of this work is the large-scale urban mapping. Therefore, we can only use TanDEM-X stripmap data, which has global coverage. The resolution of this data is dramatically reduced compared to spotlight data, which is around 3 m in azimuth direction and 1.5 m in range direction. As mentioned before, the accuracy of 3-D reconstruction stronly depends on the mutiplication of SNR and the number of acquisitions N. The typical number of available images for global-scale is about 3 to 5 [14]. There is also exsiting algorithm handling the micro-stack TomoSAR problem. In [15], authors use GIS information to group the pixels with similar height for the joint sparsity estimation, which enables use to obtain an accurate estimation with only six interferograms. Although the promising result has been obtained, the required GIS information is not avaiable for anywhere. Hence, a feasible way is to increase the SNR in order to maintain the estimation accuracy.

A successful trends to increase the SNR is non-local filtering. Recently, [16] [17] [18] indicate that SNR can be significantly improved by applying non-local filters into the tomographic inversion for different data, for instance, airborne E-SAR, spaceborne COSMO-Skymed and TerraSAR-X. In

978-1-7281-8942-0/20/$31.00 ©2020 European Union

[17], the non-local filter and total variation regularizer has been introduced to improve the multi-baseline phase unwrapping process. In [16], authors use different non-local filters to accurately estimate the covariance matrix of distributed scatterers, which leads to an improved height estimation for both simulated and real data. In [18], it is shown that we can achieve a reasonable reconstruction using only seven interferograms and better super-resolution properties when the number of interferograms is relative low.

In this work, a novel framework has been proposed, which introduces non-local fitering into tomographic processing for multi-master multi-baseline SAR tomography configuration with micro-stacks of TanDEM-X bistatic interferograms. The application of robust estimation for reconstructed building height can dramatically improve the accuracy of the result. The hybrid strategy for selection of spectral estimators can speed up the whole processing and maintain the estimation accuracy and the super-resolution power.

In order to evaluate the performance of the proposed framework, a micro-stack with five TanDEM-X bistatic interferograms which covers a large-scale area has been used. Until now, there are only few works about the large-scale demonstration with TomoSAR [22] [23] [8]. And the validation of TomoSAR result is only performed for some individuel buildings [6] [7] [9]. Hence, the validation of TomoSAR result at a large-scale is very important for the community. In this work, we use LiDAR data as reference to compare with our reconstructed TomoSAR point cloud.

II. Non-Local TomoSAR for Multi-Master InSAR

First, the non-local TomoSAR pipeline has been introduced in this section. The whole pipeline have four parts, which are non-local fitering, spectral estimation, model selection and robust height estimation. As one can see in Fig. 1, the multi-master multi-baseline SAR imaging configuration is illustrated.

A. The Multi-Master SAR Imaging Model

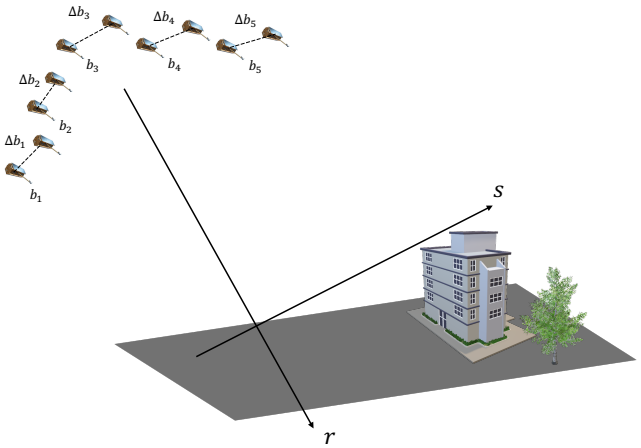


Fig. 1. Illustration of multi-master multi-baseline SAR imaging.

The TomoSAR imaging model can be formulated as :

$$g_n = \Gamma(k_n) = \int \gamma(s) \exp(-jk_n s) ds \tag{1}$$

with

$$k_n = -\frac{4\pi b_n}{\lambda r} \tag{2}$$

Where the term g_n is the measurement of the n^{th} SAR acquisition. $\gamma(s)$ is the reflectivity profile along the elevation direction s and $\Gamma(k)$ is the Fourier transform of $\gamma(s)$, where wavenumber k is a scaled version of the sensor's position b_n projected on the cross-range-azimuth axis $b||s$. Note that b_n for multi-master configuration are no longer baselines, but the positions of the sensor w.r.t. some origin. In case of monostatic multi-temporal data stacks, a single master g_0 is chosen with $b_0 = 0$ and interferograms to all other acquisitions are formed : $g_n g_0^*$.

Here we are using stacks of *bistatic* interferograms, i.e. the multi-master configuration. For each acquisitions we have a master $g_{n,m} = \Gamma(k_n)$ taken at $b_{master} = b_n$ and a slave $g_{n,s} = \Gamma(k_n + \Delta k_n)$ image taken at $b_{slave} = b_n + \Delta b_n$, where Δb_n is the bistatic baseline (which takes the *effective* positions of the transmit-receive phase center into account). The consequence is that the sinlge-master sar imaging model can not handle this configuration, since it will confuse Δb_n and b_n. When only single scatterer exists in $\gamma(s)$, this would be no problem, since the Fourier transform of a single point has a constant magnitude and a linear phase. In order to determine the slope of the phase ramp we can take any two samples and divide their phase difference by the difference in wavenumbers (= baseline). But for two or more scatterers, this will not work. Here we show an example of two symmetric and equally strong scatterers to make it clear:

$$\begin{gathered} \gamma(s) = \delta(s + s_0) + \delta(s - s_0) \\ \updownarrow \\ \Gamma(k) = 2\cos(s_0 k) = 2\cos(2\pi \frac{2s_0}{\lambda r} b) \end{gathered} \tag{3}$$

Hence, interferograms with the same baseline Δb are different depending on where the two sensors were located along b. If by chance one of the sensors is at a zero of $\Gamma(k)$, e.g. at $b = \lambda r / 8 s_0$, the interferogram would be zero. Obviously, every bistatic acquisition provides three pieces of information: the two magnitudes $|\Gamma(k_n)|$ and $|\Gamma(k_n + \Delta k_n)|$ as well as the phase difference $\angle \Gamma(k_n + \Delta k_n)\Gamma^*(k_n)$ which have to be accounted for by the inversion algorithm.

This is true for pixel-wise tomographic inversion or for point scatterers. The situation becomes different, though, once we talk about averages of pixels, i.e. estimates of expectation values. Let us assume Gaussian distributed scattering with a backscatter coefficient along elevation of

$$\sigma_0 = E\left\{|\gamma(s)|^2\right\} \tag{4}$$

Assuming further that $\gamma(s)$ is white, its power spectral density is stationary and is the autocorrelation function of $\Gamma(k)$, i.e.

the Fourier transform of $\sigma_0(s)$ as a function of the baseline wavenumber Δk :

$$E\left\{\Gamma(k_n+\Delta k_n)\Gamma^*(k)\right\}=\int \sigma_0(s)\exp(-j\Delta k_n s)ds \quad (5)$$

Instead of sampling the Fourier spectrum we sample its autocorrelation function by the bistatic data stack. Since this relationship is *independent* of $k \propto b$ because of stationarity, it makes no difference, where the two acquisitions have been taken, only their baseline Δb_n counts. In other words we can use standard TomoSAR inversion algorithms in this case.

In this work, a non-local concept has been introduced in order to improve the SNR for micro-stacks case. The non-filter can combine different patches into a weighted maximum likelihood estimation (WMLE). Therefore, we can assume that the expectation $E\left\{\Gamma(k_n+\Delta k_n)\Gamma^*(k)\right\}$ can be good estimated. In the presence of noise $\boldsymbol{\varepsilon}$, the discretized TomoSAR model can be expressed as

$$\mathbf{g}=\mathbf{R}\mathbf{X}+\boldsymbol{\varepsilon} \quad (6)$$

where $\mathbf{g}=[g_1,g_2,...,g_n]^{\mathrm{T}}$ is vector notation of the complex-valued measurement with dimension $N\times 1$, and $\mathbf{X}\sim\sigma_0(s_l)=E\{|\gamma(s_l)|^2\}$ is the expectation value of reflectivity profile along elevation uniformly sampled at $s_l(l=1,2,...,L)$. $\mathbf{R}$ is a sensing matrix with the dimension $N\times L$, where $R_{nl}=\exp(-j\Delta k_n s_l)$.

B. Non-Local Procedure

Since the number of available images is limited for our task, we need to improve the SNR to maintain the accuracy. A successful trends is to apply non-local concept into tomographic inversion. The non-local procedure can dramatically improve the SNR of images without notable resolution distortion. The patch-wise non-local concept is to combine different patches into a weighted maximum likelihood estimation. The value of pixel s in the search window is used to calculate the value of pixel c and the patch with central pixel s is used to measure the similarity of the patch with the central pixel c. The formulation of this process can be written as:

$$\hat{\boldsymbol{\Theta}}_c=\operatorname{argmax}\sum_s \mathbf{w}(i_s,j_s)\log p(\mathbf{g}_s|\boldsymbol{\Theta}) \quad (7)$$

where weights $\mathbf{w}(i_s,j_s)$ can be measured by calulating the patch-wise similarity [18] [19]. Assuming that we have two expressions $\mathbf{g}=(I_1,I_2,\phi)$ and $\boldsymbol{\Theta}=(\psi,\mu,\sigma^2)$, where $\mathbf{g}$ denotes the complex-valued measurement. I_1 and I_2 are the instensity of two SAR images. ϕ is the interferometric phase.

C. Spectral Estimation

Spectral Estimation is main step for SAR tomography. Since each estimator has different performance and efficiency, we need to consider the trade-off between the accuracy and computational efficiency. Hence, we have proposed a hybrid algorithm for spectral estimation. The formulation of singular value decomposition (SVD) [4] [5], compressive sensing (CS) are introduced in the following.

- SVD:

$$\hat{\mathbf{X}}=\left(\mathbf{R}^{\mathrm{H}}\mathbf{C}_{\varepsilon\varepsilon}^{-1}\mathbf{R}+\mathbf{C}_{XX}^{-1}\right)^{-1}\mathbf{R}^{\mathrm{H}}\mathbf{C}_{\varepsilon\varepsilon}^{-1}\mathcal{N}(\mathbf{g}) \quad (8)$$

- CS:

$$\hat{\mathbf{X}}=\arg\min_{\mathbf{X}}\{\|\mathbf{R}\mathbf{X}-\mathcal{N}(\mathbf{g})\|_2^2+\lambda\|\mathbf{X}\|_1\} \quad (9)$$

In this work, we use similar strategy introduced in [22]. The reflectivity profile is first estimated by a faster estimator, i.e. SVD-wiener, in order to discriminate the number of scatterers in one resolution cell. After that, more accurate CS-based approach is applied in order to distinguish multiple scatteres within one resolution cell. The hybrid algorithm can significantly reduce the number of pixels that requires the L_1 regularization, which is equvalent to reduce the computational cost. Moreover, our previous development [23] can speedup the processing of L_1 minimization.

D. Robust Height Estimation

In order to remove the outliers in the height estimates, the final result will be estimated by an *M-estimator* with multiple neighbouring pixels. Instead of minimizing the sum of squared residuals in averaging, M-estimator minimizes the sum of a customized function $\rho(\centerdot)$ of the residuals:

$$\tilde{s}=\arg\min_s\sum_i \rho(\hat{s}_i-s), \quad (10)$$

where $\hat{s}_i$ represents the estimated elevation of the ith pixel. The alternative interpretation of Eq. (10) is a weighted averaing of the heights of the neighbouring pixels. The weighting function can be written in this formulation.

$$w(x)=\frac{-\partial\rho(x)}{x\partial x} \quad (11)$$

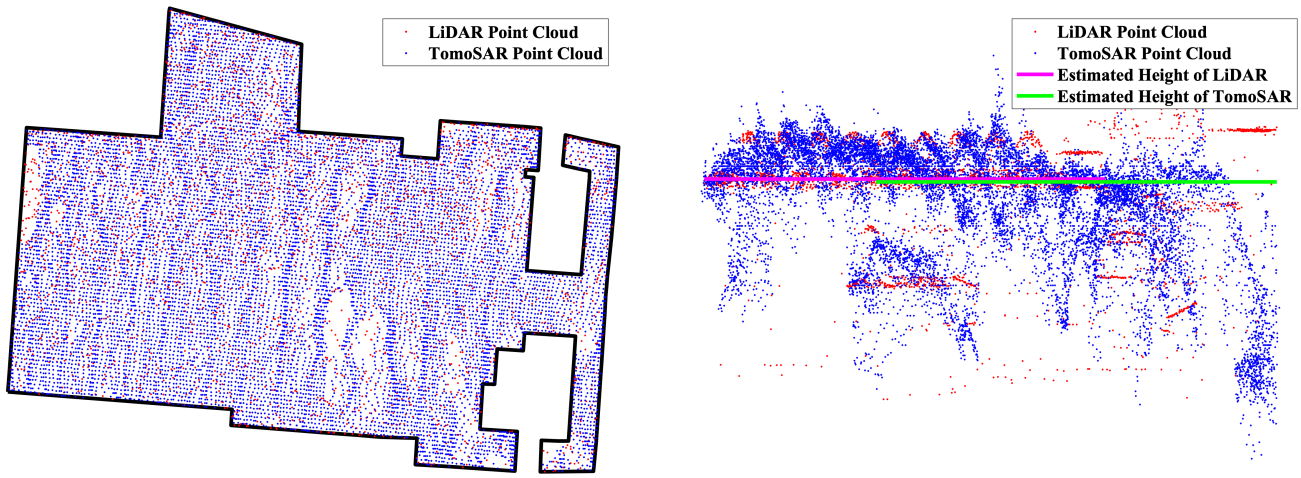
III. Practical Demonstration

A. Data Description

The data used in this work is a stack of TanDEM-X bistatic interferograms, which includes five acquisitions between July 2016 and April 2017. The data covers the city of Munich in Germany. It is preprocessed by integrated TanDEM-X processor (ITP) from German Aerospace Center (DLR). The main parameters of the acquisitions are summarized in Table I.

TABLE I
PARAMETERS OF TANDEM-X STRIPE MAP ACQUISITION OF MUNICH

Name	Symbol	Value
Distance of Range	r	698 km
Wavelength	λ	3.1 cm
Incidence angle at scene center	θ	50.4°
Maximal elevation aperture	Δb	187.18 m
Number of interferograms	N	5



(a) (b)

Fig. 2. Example of Robust Height Estimation of LiDAR and TomoSAR point clouds. (a) Top view of two point clouds, i.e., LiDAR (red) and TomoSAR (blue). Note that since LiDAR is too dense, only 1% is visuallized. (b) robust height estimation of two point clouds, LiDAR point cloud (red dots), TomoSAR point cloud (blue dots), estimated building height of LiDAR data (magenta solid line), estimated building height of TomoSAR data (green solid line).

B. Quantitative Validation

In order to systematically investigate the performance of proposed method, the precise LiDAR data is used as reference to compare with the result of reconstructed TomoSAR point clouds. The resolution of LiDAR data is ten centimeter, which is produced by Bavarian State Office for Survey and Geoinformation. As different data sources have different coordinates and quality, we apply the following steps on the data. (1) Geocoding of TomoSAR point cloud; (2) Co-registration of different point clouds; (3) Object-based raster data generation; (4) Robust height estimation. Here we show an example of these pre-processing steps for the structure "munich central station" in Fig. 2. Fig. 2 (a) shows the top view of two point clouds, i.e., LiDAR (red) and TomoSAR (blue). Note that since LiDAR is too dense, only 1% is visuallized. And Fig. 2 (b) shows the robust height estimation of two point clouds, LiDAR point cloud (red dots), TomoSAR point cloud (blue dots), estimated building height of LiDAR data (magenta solid line), estimated building height of TomoSAR data (green solid line). The height difference of the central station of Munich is about 0.25 m.

A quantitative assessment for the large-scale area has been carried out in this study. More than 36,000 buildings in the city has been compared with the LiDAR data. The result shows there are 38.7% buildings within 1 m accuracy and 62.8% buildings within 2 m accuracy. Since the TanDEM-X data and LiDAR data were obtained at different time, it is likely that old buildings were demolished or new buildings are constructed. Hence, the histogram of height difference is truncated with a threshold of 15 m in order to avoid outliers. Fig. 3 shows the truncated distribution of height differences. There are still 34,054 buildings left after truncation. The standard deviation of height difference is about 1.96 m [24].

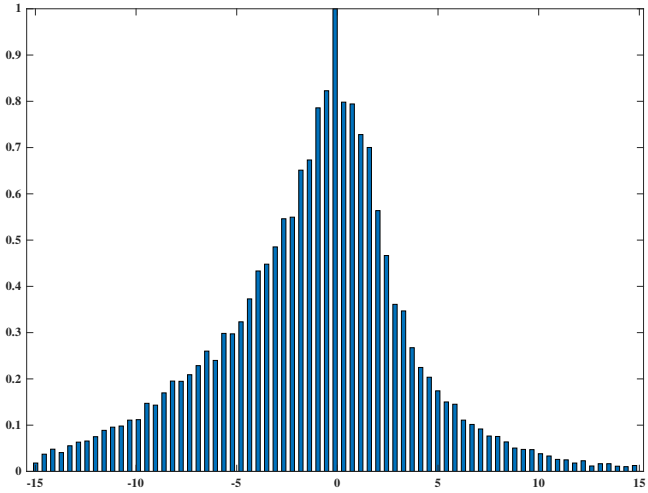


Fig. 3. Histogram of height differences of structures.

C. Fusion with Building Footprint

Finally, LOD1 polyhedral models are generated. The 3-D urban models are reconstructed by extruding OpenStreetMap (OSM) with the building height estimated by the proposed multi-master non-local TomoSAR approach. Fig. 4 shows the fused 3-D urban model of Munich. Color indicates the height of the buildings and 3-D models are overlayed on the Google Map images.

IV. Conclusion

In this work, we have proposed a novel TomoSAR approach with micro-stacks of interferograms to generate 3-D urban model. The introduction of Non-Local filter into TomoSAR

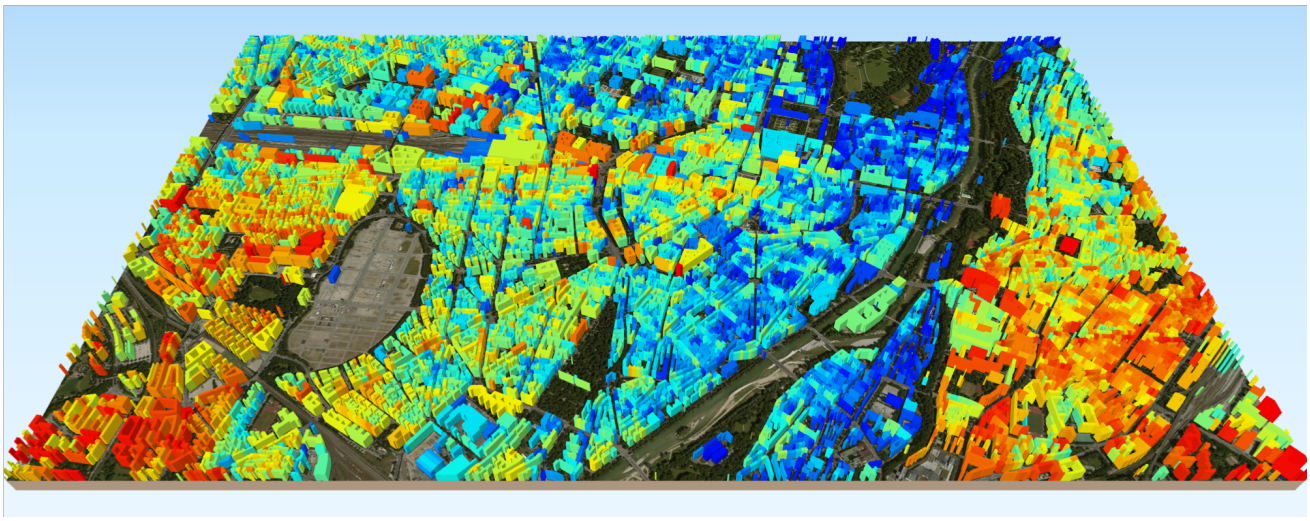


Fig. 4. Visualization of fused 3-D urban model of Munich.

inversion can significantly improve the SNR of the interferograms, which can maintain the estiamtion accuracy for the limited number of interferograms. A large-scale experiment with TanDEM-X bistatic stacks have been carried out to evaluate the performance of proposed framework. The result shows an unprecedented accuracy of height estimation can be achieved for a large-scale area. Therefore, it indicates the proposed approach can be a effective tool for high quality large-scale 3-D urban mapping.

REFERENCES

[1] G. Krieger, A. Moreira, H. Fiedler, I. Hajnsek, M. Werner, M. Younis, and M. Zink, "TanDEM-X: A satellite formation for high-resolution SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 3317-3341, Oct. 2007.

[2] A. Reigber, and A. Moreira, "First demonstration of airborne SAR tomography using multibaseline L-band data," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, no. 5, pp. 2142-2152, Sep. 2000.

[3] F. Gini, F. Lombardini, and M. Montanari, "Layover solution in multibaseline SAR interferometry," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 4, pp. 1344-1356, 2002.

[4] G. Fornaro, F. Lombardini, and F. Serafino, "Three-dimensional multipass SAR focusing: experiments with long-term spaceborne data," *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 4, pp. 702-714, Apr. 2005.

[5] 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.

[6] N. Ge, F. Rodriguez Gonzalez, Y. Wang, Y. Shi, and X. X. Zhu, "Spaceborne Staring Spotlight SAR Tomography–A First Demonstration With TerraSAR-X," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 10, pp. 702-714, Oct. 2018.

[7] X. X. Zhu and R. Bamler, "Demonstration of super-resolution for tomographic SAR imaging in urban environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 3150-3157, Aug. 2012.

[8] X. X. Zhu and Y. Wang and S. Gernhardt and R. Bamler, "TomoGENESIS: DLR's tomographic SAR processing system," *Proc. Joint Urban Remote Sensing Event, 2013* pp. 159-162.

[9] A. Budillon, A. C. Johnsy, and G. Schirinzi, "Extension of a Fast GLRT Algorithm to 5D SAR Tomography of Urban Areas," *Remote Sensing*, vol. 9, no. 8, pp, 844, 2017.

[10] G. Fornaro, A. Pauciullo, D. Reale, and S. Verde. "Multilook SAR Tomography for 3-D Reconstruction and Monitoring of Single Structures Applied to COSMO-SKYMED Data," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 7, no. 7, pp. 2776-2785, Jul. 2014.

[11] 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.

[12] 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.

[13] X. X. Zhu and R. Bamler, "Super-resolution power and robustness of compressive sensing for spectral estimation with application to spaceborne tomographic SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 1, pp. 247-258, Jan. 2012.

[14] P. Rizzoli, M. Martone, C. Gonzalez, C. Wecklich, D. Borla Tridon, B. Bräutigam, M. Bachmann, D. Schulze, T. Fritz, M. Huber, B. Wessel, G. Krieger, M. Zink, A. Moreira, "Generation and performance assessment of the global TanDEM-X digitalelevation model," *ISPRS J. Photogram. Remote Sens.*, vol. 132, pp. 119-139, 2017.

[15] X. X. Zhu, N. Ge, 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.

[16] O. D'Hondt, C. López-Martínez, S. Guillaso, and O. Hellwich, "Nonlocal Filtering Applied to 3-D Reconstruction of Tomographic SAR Data," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 272-285, Jan. 2018.

[17] G. Ferraioli, C.A. Deledalle, L. Denis, and F. Tupin, "PARISAR: Patch-Based Estimation and Regularized Inversion for Multibaseline SAR Interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1626-1636, Mar. 2018.

[18] Y. Shi, X. X. Zhu and R. Bamler, "Non-Local Compressive Sensing Based SAR Tomography," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 3015-3024, May 2019.

[19] X. X. Zhu, G. Baier, M. Lachaise, Y. Shi, F. Adam, R. Bamler, "Potential and limits of non-local means InSAR filtering for TandDEM-X high-resolution dem generation," *Remote sensing of environment*, vol. 218, pp. 148-161, 2018.

[20] Y. Shi, X. X. Zhu and R. Bamler, "SAR Tomography Using Non-Local Sparse Reconstruction," *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia, 2018, pp. 6091-6094.

[21] Y. Shi, Y. Wang, X. X. Zhu and R. Bamler, "Non-Local SAR Tomography for Large-Scale Urban Mapping," *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, Yokohama, Japan, 2019, pp. 5197-5200.

[22] 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.

[23] Y. Shi, X.X. Zhu, W. Yin, and R. Bamler, "A fast and accurate basis pursuit denoising algorithm with application to super-resolving tomographic SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 6148-6158, Oct. 2018.

[24] Y. Shi, R. Bamler, Y. Wang and X. X. Zhu, "SAR Tomography at the Limit: Building Height Reconstruction Using Only 3-5 TanDEM-X Bistatic Interferograms," *IEEE Trans. Geosci. Remote Sens.*, in press. 2020.