SAR Tomography at the Limit: Building Height Reconstruction Using Only 3 – 5 TanDEM-X Bistatic Interferograms

Yilei Shi, Member, IEEE, Richard Bamler, Fellow, IEEE, Yuanyuan Wang, Member, IEEE, and Xiao Xiang Zhu, Senior Member, IEEE

Abstract—This is the preprint version, to read the final version please go to IEEE Transactions on Geoscience and Remote Sensing on IEEE Xplore. Multi-baseline interferometric synthetic aperture radar (InSAR) techniques are effective approaches for retrieving the 3-D information of urban areas. In order to obtain a plausible reconstruction, it is necessary to use more than twenty interferograms. Hence, these methods are commonly not appropriate for large-scale 3-D urban mapping using TanDEM-X data where only a few acquisitions are available in average for each city. This work proposes a new SAR tomographic processing framework to work with those extremely small stacks, which integrates the non-local filtering into SAR tomography inversion. The applicability of the algorithm is demonstrated using a TanDEM-X multi-baseline stack with 5 bistatic interferograms over the whole city of Munich, Germany. Systematic comparison of our result with TanDEM-X raw digital elevation models (DEM) and airborne LiDAR data shows that the relative height accuracy of two third buildings is within two meters, which outperforms the TanDEM-X raw DEM. The promising performance of the proposed algorithm paved the first step towards high quality large-scale 3-D urban mapping.

Index Terms—SAR interferometric (InSAR), SAR tomography (TomoSAR), TanDEM-X, digital elevation models (DEM), 3-D urban mapping

I. INTRODUCTION

A. The TanDEM-X Mission

This work is supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement no. ERC-2016-StG-714087, acronym: S02Sat, www.s02sat.eu), the Helmholtz Association under the framework of the Young Investigators Group “SiPEO” (VH-NG-1018, www.sipeo.bgu.tum.de), Munich Aerospace e.V. Fakultät für Luft- und Raumfahrt, and the Bavaria California Technology Group "SiPEO" (VH-NG-1018, www.sipeo.bgu.tum.de), Munich Aerospace Center (DLR), 82234 Weßling, Germany and Signal Processing in Urban Mapping (TomoSAR), TanDEM-X, digital elevation models (DEM), 3-D urban mapping.

B. SAR Tomography Techniques

TANDEM-X satellite is a German civil and commercial high-resolution synthetic aperture radar (SAR) satellite which has almost identical configuration as its ‘sister’ TerraSAR-X satellite. Together with TerraSAR-X, they are aiming to provide a global high-resolution digital elevation model (DEM) [1]. Both satellites use a spiral orbit constellation to fly in tight formation in order to acquire the image pair simultaneously, which significantly reduces the temporal decorrelation error and the atmospheric interference. Since its launch in 2010, TanDEM-X has been continuously providing high quality bistatic interferograms that are nearly free from deformation, atmospheric and temporal decorrelation.

C. The proposed framework

As mentioned above, the accuracy of 3-D reconstruction replies on the product of SNR and the number of measurements $N$. Since the motivation of our work is the large-scale...
urban mapping, the data we adopted is TanDEM-X stripmap co-registered single look slant range complex (CoSSC), whose resolution is about 3.3 m in azimuth direction and 1.8 m in range direction. The typical number of available interferograms for most areas is 3 to 5 [16]. In [17], the pixels with similar height are grouped for the joint sparsity estimation, which leads to an accurate inversion of TomoSAR using only six interferograms. Although the unprecedented result is obtained, the accurate geometric information is usually not available for most areas. Therefore, the feasible way to keep the required precision of the estimates is to increase the SNR.

Recent works [18] [19] [20] show that SNR can be dramatically increased by applying non-local filters to the TomoSAR processing for different sensors, such as airborne E-SAR, COSMO-Skymed and TerraSAR-X. In [18], different non-local filters have been adopted to improve the estimation of the covariance matrix for distributed scatterers, which leads to a better height estimation for simulated data and airborne SAR data. Ferraioli et al. [19] introduced the non-local filter and the total variation regularizer to improve the multi-baseline phase unwrapping process. In [20], 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, we extend the concept of non-local compressive sensing TomoSAR in [20] [21] [22] and propose a new framework of spaceborne multi-baseline SAR tomography with TanDEM-X bistatic micro-stacks, i.e. 3 to 5 interferograms. The framework includes non-local filtering, spectral estimation, model selection and robust height estimation. Since the different spectral estimators have different estimation accuracy and computational cost, we compared the estimation accuracy of different estimators with micro-stacks.

The demonstration of different TomoSAR inversion methods for a large-scale area has been shown in [23] [24] [11]. Only a few works on the validation of single buildings were reported in [8] [10] [25]. Therefore, the validation of the specified quality of the TomoSAR result at a larger scale, would be of considerable interests for the scientific and commercial users. We choose Munich city as a test site because of a high quality LiDAR reference available to us, and we propose a complete workflow to compare the TomoSAR point cloud [26] generated by the proposed framework, TanDEM-X DEM product, and LiDAR data.

D. Contribution of this work

The major contributions of this work are summarized as follows:

- We make possible a new application of bistatic SAR data for global building height reconstruction.
- We have pointed out that for pixel-wise multi-master TomoSAR the well-known system equation is no longer valid in the multi-scatterer case.
- We have developed a framework for tomographic stacks with only 3 - 5 interferograms. A systematic investigation on the estimation accuracy and super-resolution power for the micro-stacks have been carried out, which was never done before.

- We use 5 TanDEM-X bistatic data to demonstrate the proposed framework. A systematic validation for large-scale TomoSAR reconstruction have been carried out and a method for comparing with other reference data is established. The results are quantitatively compared with LiDAR reference for more than 34,000 buildings.

II. NON-LOCAL TOMOSAR FOR MULTI-MASTER INSAR

In this section, we introduce the non-local TomoSAR framework for multi-master multi-baseline InSAR configuration. Fig. 1 illustrates multi-master multi-baseline SAR imaging. The framework consists of several steps: (1) non-local filtering; (2) spectral estimation; (3) model selection; (4) robust height estimation. Fig. 2 shows the flowchart of the non-local TomoSAR framework.

A. The multi-master TomoSAR imaging model

For a fixed azimuth-range position, $\gamma(s)$ represents the reflectivity profile along elevation $s$. The measurement $g_n$, i.e. the complex value of the corresponding azimuth-range pixel, in the $n^{th}$ SAR image is then a sample of $\Gamma(k)$ – the Fourier transform of $\gamma(s)$, where the elevation wavenumber $k$ is a scaled version of the sensors position $b_n$ projected on the cross-range-azimuth axis $b|s$:

$$g_n = \Gamma(k_n) = \int \gamma(s) \exp(-jk_n s) ds$$

with

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

Note that $b_n$ are no 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 $b_0$ and its phase is subtracted from all other acquisitions: $g_n g_0^∗ / |g_0|$. This operation renders the phase spatially smooth and is a prerequisite for spatial phase unwrapping, averaging and graph (network) processing. It does not reduce information, since the phase of any (master) acquisition is random. Note that the choice of the point $b = 0$ only defines the x-r-s coordinate system. This point need not necessarily be the master track position $b_0$. However, $b_0 = 0$ is a mathematically convenient choice and is assumed in all conventional TomoSAR system model like in \cite{27}. All the equations in \cite{27}, however, are actually independent of this particular choice. Only in the special case $b_0 = 0$ the $b_n$ are identical to baselines.

Here we are dealing with stacks of bistatic acquisitions, i.e. with the multi-master case. From each of these acquisitions we get a master $g_{n,m} = \Gamma(k_n)$ taken at $b_{n,\text{master}} = b_n$ and a slave $g_{n,s} = \Gamma(k_n + \Delta k_n)$ image taken at $b_{n,\text{slave}} = b_n + \Delta b_n$, where $\Delta b_n$ is the bistatic baseline (which takes the effective positions of the transmit-receive phase center into account).

If we used a standard, i.e. single-master, TomoSAR inversion algorithm, we would confuse $\Delta b_n$ and $b_n$. In the case of a single scatterer in $\gamma(s)$, this misinterpretation would do no harm, because 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). This is no longer true for two or more scatterers. The example of two symmetric and equally strong scatterers makes this clear:

$$\gamma(s) = \delta(s + s_0) + \delta(s - s_0)$$

$$\Gamma(k) = 2 \cos(s_0 k) = 2 \cos(2 \pi b_{s0} k b)$$

Hence, acquisitions with the same baseline $\Delta b$ are different depending on where the two sensors were located along $b$.

Stack Non-Local Filtering

$\mathcal{N}(g) = f(\theta)$

Combination of Spectral Estimators

$X = (R^* C \Delta^\dagger + R \Delta) \mathcal{N}(g)$

or

$X = \arg \min \|RX - \mathcal{N}(g)\|_2^2 + \lambda\|X\|_1$

Model Selection

$\hat{R} = \arg \min \{ -2 \ln p(\gamma|K) + 2C(K) \}$

Robust Height Estimation

$\hat{\delta} = \arg \min \sum \rho(\delta_i - s)$

Hence, acquisitions with the same baseline $\Delta b$ are different depending on where the two sensors were located along $b$.

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 $\angle \Gamma(k_n + \Delta k_n) = \Gamma^*(k_n)$, i.e. we must normalize the phase by the respective master in every acquisition, in order to become unaffected by deformation and atmospheric delay. Spectral estimation based conventional TomoSAR inversion algorithms, however, require complex spectral samples at several wavenumbers, phase-normalized to a single master phase. In a current parallel work by one of the authors \cite{28} it is shown that pixel-wise TomoSAR using multi-master acquisitions is a non-convex hard to solve problem.

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

$$\sigma_0(s) = E \{ |\gamma(s)|^2 \}$$

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 $\Delta k$:

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

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 $\Delta b_n$ counts. In other words we can use standard TomoSAR inversion algorithms in this case.
In this paper we use nonlocal filtering to improve SNR for micro-stacks. These filters perform ensemble averages with number of looks in the order of tens to hundreds. Hence, we tend to the assumption that we work with reasonably good estimates of \( E \{ \Gamma (k_n + \Delta k_n) \} \) and can use the bistatic interferograms for TomoSAR reconstruction.

By introducing a noise \( \varepsilon \), the matrix notation of TomoSAR model can be formulated as:

\[
g = RX + \varepsilon \tag{6}
\]

where \( g = [g_1, g_2, \ldots, g_n]^T \) is vector notation of the complex-valued measurement with dimension \( N \times 1 \), and \( X \sim \sigma_0(s_l) = E\{ |\gamma(s_l)|^2 \} \) is the expectation value of reflectivity profile along elevation uniformly sampled at \( s_l (l = 1, 2, \ldots, L) \). \( R \) 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 we have only limited number of acquisitions for large-scale area, the SNR need to be dramatically increased in order to obtain the required accuracy. As shown in [20], non-local procedure is an efficient way to increase the SNR of interferograms without notable resolution distortion. The idea of patch-wise non-local means considers all the pixels \( s \) in the search window, when the patch with the central pixel \( s \) is similar to the patch with central pixel \( c \), the value of \( s \) is selected for calculating the value of pixel \( c \). The value of pixel \( c \) is estimated by using a weighted maximum likelihood estimator (WMLE).

\[
\hat{\Theta}_C = \arg \max \sum_s w(i_s, j_s) \log p(g_s | \Theta) \tag{7}
\]

where weights \( w(i_s, j_s) \) can be calculated by using patch-wise similarity measurement [20]. Assuming that we have two expressions \( g = (I_1, I_2, \phi) \) and \( \Theta = (\psi, \mu, \sigma^2) \), where \( g \) denotes the complex-valued measurement. \( I_1 \) and \( I_2 \) are the intensity of two SAR images. \( \phi \) is the interferometric phase. \( \Theta \) is the true value of the parameters, where \( \psi \) is the noise-free interferometric phase, \( \mu \) is the coherence magnitude, and \( \sigma^2 \) is the variance. The likelihood function \( p(g_s | \Theta) = p(I_1, I_2, \phi_s | \psi, \mu, \sigma^2) \) is adopted from [29] with following formulaiton:

\[
p(I_1, I_2, \phi | \psi, \mu, \sigma^2) = \frac{1}{16\pi^2\sigma^4(1 - \mu^2)} \times \exp \left[ \frac{-I_1 + I_2 - 2\sqrt{I_1 I_2} \mu \cos(\phi - \psi)}{2\sigma^2(1 - \mu^2)} \right] \tag{8}
\]

\( \mathcal{N}(. \) denotes the non-local estimator, where \( \mathcal{N}(g) = f(\hat{\Theta}) \). \( \hat{\Theta} = (\hat{\psi}, \hat{\mu}, \hat{\sigma^2}) \) represents the parameters being estimated, where \( \hat{\psi} \) is the estimated interferometric phase, \( \hat{\mu} \) stands for the coherence magnitude, and \( \hat{\sigma^2} \) stands for the variance. \( f(\Theta) \) is the maximum likelihood estimator and the estimated parameters can be formulated as

\[
\hat{\psi} = -\arg \left( \sum_s w_s g_{1,s} g_{2,s}^* \right) \tag{9}
\]

\[
\hat{\mu} = \frac{2}{\sum_s w_s |g_{1,s}| |g_{2,s}|} \tag{10}
\]

\[
\hat{\sigma^2} = \frac{\sum_s w_s (|g_{1,s}|^2 + |g_{2,s}|^2)}{4\sum_s w_s} \tag{11}
\]

The patch size and search window size are set to be \( 7 \times 7 \) and \( 21 \times 21 \) according to experimental study, which is also reported by other works [30] [31]. Each pixel represents 2.17 m in azimuth and 1.36 m in range.

**C. Spectral Estimation**

After the non-local procedure, spectral estimation is applied. The most relevant spectral estimation algorithms, including singular value decomposition (SVD) [5] [7], compressive sensing (CS) are introduced in the following.

- **SVD:**

\[
\hat{X} = (R^H C_{xx}^{-1} R + C_{xx}^{-1})^{-1} R^H C_{xx}^{-1} N(g) \tag{12}
\]

- **CS:**

\[
\hat{X} = \arg \min_X \{ ||RX - N(g)||^2_2 + \lambda ||X||_1 \} \tag{13}
\]

where \( C_{xx} \) is the noise covariance matrix, which is defined as:

\[
C_{xx} = (g - RX) (g - RX)^H \tag{14}
\]

Under the assumption that the model errors are circular Gaussian distributed with zero mean, the noise covariance matrix is formulated as \( C_{xx} = |\sigma_c|^2 I \) and \( |\sigma_c|^2 \) is the noise power level. \( C_{XX} \) is the covariance matrix of the prior, if it is assumed to be white, i.e. \( C_{XX} = I \).

The choice of different combinations of spectral estimators depends on the required accuracy, the computational time and others. We follow the procedure proposed in [23]. It consists of three parts: (1) an efficient low-order spectral estimation; (2) the discrimination of the number of scatterers; (3) an accurate high-order spectral estimation. The elevation profile is first estimated by an efficient low-order spectral estimator in order to discriminate the number of scatterers in one resolution cell. Then, CS-based approach is adopted for the pixel which has multiple scatterers. This method decreases the amount of pixels that need the \( L_1 \) minimization, which leads to reduce the computational cost. Furthermore, the rest of pixels can be efficiently solved by randomized blockwise proximal gradient method [24].

**D. Model Selection**

The abovementioned spectral estimators retrieve a nonparametric reflectivity profile. Since our data is in urban area, we assume only a few dominant scatterers exist along the reflectivity profile. Therefore, the number of scatterers \( K \) is estimated by a model order selection algorithm as well as their
elevation in one azimuth-range pixel \( \hat{s} \). The estimator can be expressed as follows:

\[
\hat{K} = \arg \min_{K} \left\{ -2 \ln \rho \left( \theta \right) + 2C(K) \right\}
\]  

where \( C(K) \) is a model complexity penalty term which avoids more complicated model overfitting the observed data. The classical penalized likelihood criteria are the Bayesian information criterion (BIC), the Akaike information criterion (AIC), and the minimum description length (MDL) principle [32].

As mentioned in [7], the criteria of model order selection has to be chosen according to the experiments for the particular situation, because it is difficult to remove the bias of the selection.

E. Robust Height Estimation

To tackle the possible remaining outliers in the height estimates, the final height will be fused from the result of multiple neighbouring pixels as a post-processing. Instead of minimizing the sum of squared residuals in averaging, M-estimator minimizes the sum of a customized function \( \rho(\cdot) \) of the residuals:

\[
\hat{s} = \arg \min_{s} \sum_{i} \rho \left( \hat{s}_i - s \right),
\]

where \( \hat{s}_i \) is the elevation estimates of the \( i \)th neighbouring pixel. It is shown that the close-formed solution of Eq. (16) is simply a weighted averaging of the heights of the neighbouring pixels [33]. The weighting function can be expressed as follows, if the derivative of \( \rho(x) \) exists.

\[
w(x) = \frac{\partial \rho(x)}{x \partial x}
\]  

The robust estimated height can be written as follows:

\[
\hat{h} = \frac{\sum_{i} w(x_i) \cdot \hat{h}}{\sum_{i} w(x_i)}
\]  

where \( \hat{h} = \hat{s} \cdot \sin \theta \), and \( \theta \) is incident angle. The choice of the weighting function depends on the distribution of the heights. Without prior knowledge of the distribution, promising robust weighting functions are Tukey’s biweight or t-distributed weighting [33]. For instance, the formulation of Tukey’s biweight loss function can be written as:

\[
\rho(x) = \begin{cases} 
-\frac{(c_2-x^2)^2}{c_4} + \frac{x^2}{c_6} & \text{if } |x| < c_r \\
0 & \text{elsewhere}
\end{cases}
\]  

and the weighting function can be formulated as:

\[
w(x) = \begin{cases} 
1 - \frac{x^2}{c_r^2} - \frac{2x^2}{c_r^4} & \text{if } |x| < c_r \\
0 & \text{elsewhere}
\end{cases}
\]

III. Estimation Accuracy of TomoSAR with Small Stacks

This section will discuss the theoretical 3-D reconstruction accuracy of a micro-stack with 3-5 interferograms. The estimation accuracy of TomoSAR has been systematically investigated. It is exhaustively shown in [15] that the elevation estimation accuracy and SR power depend asymptotically on the multiplication \( N \cdot \text{SNR} \). In this section, we investigate the estimation accuracy of TomoSAR with the extremely small number of interferograms, which is 3 to 5.

A. The Lower Bound for Micro-stacks

In the case of pixel-wise TomoSAR inversion, i.e. without spatial averaging, each of our \( N \) bistatic pairs contain three
pieces of information, as mentioned before. If we want to reconstruct elevation profiles containing \( M \) discrete scatterers we need to infer \( 3M \) parameters, i.e. elevation, magnitude and phase for each scatterer. Hence an absolute lower bound of the micro-stack size is \( N \geq M \).

Distributed scatterers, on the other hand, are characterized by only two parameters each: elevation and backscatter coefficient. Likewise each interferogram provides only two parameters, magnitude and phase (difference). Since our goal is 3-D reconstruction based on bistatic data, we disregard motion-induced phase here. Hence, also in this case the absolute lower limit is \( N \geq M \). This limit is only a necessary condition, however not sufficient from the robustness point of view, because of ambiguities in the inversion cost functions.

For 3-D urban mapping the single and double scattering cases are the dominant ones. We investigate the cases \( N = 3 - 5 \) in this paper, because these are close to the mentioned limits and are relevant for TanDEM-X.

### B. CRLB

It is demonstrated in [7] that the Cramer-Rao lower bound (CRLB) of the elevation estimates for single scatterer can be expressed as:

\[
\sigma_s = \frac{\lambda r}{4\pi \cdot \sigma_b \cdot \sqrt{2 \cdot \text{SNR} \cdot N}} \tag{21}
\]

where \( \sigma_b \) is the standard deviation of the baseline distribution. \( N \) is the number of interferograms, and SNR is the signal-to-noise ratio.

For the double scatterers’ case, the CRLB can be written as:

\[
\sigma_{s_q} = c_0 \cdot \sigma_{s_q,0} \tag{22}
\]

where \( \sigma_{s_q,0} \) represents the CRLB on the elevation estimation of the \( q \)th scatterer without the interference with the others. \( c_0 \) is the correction factor of the interference for the scatterers, which are closely located [15]. It is nearly free from \( N \) and SNR, which can be written as:

\[
c_0 = \max \left\{ \sqrt{\frac{40\kappa^{-2}(1 - \kappa/3)}{9 - 6(3 - 2\kappa)\cos(2\Delta\varphi) + (3 - 2\kappa)^2 + 1}}, 1 \right\} \tag{23}
\]

where \( \Delta\varphi \) is the phase difference of the two scatterers, \( \kappa \) is the normalized distance between two scatterers (defined in next section). Since \( \Delta\varphi \) is a random variable, the approximated formulation of \( c_0 \) can be calculated by integrating the variances over \( \Delta\varphi \).

\[
c_0 = \max \left\{ 2.57(\kappa^{-1.5} - 0.11)^2 + 0.62, 1 \right\} \tag{24}
\]

A note on the baseline distribution is worth mentioning: all the Eqs. (21)-(24) are satisfied with large stacks. In the micro-stacks with only 3 - 5 acquisitions, the baseline distribution may be unfavorable, even if the baseline spread \( \sigma_b \) is acceptable. For example, if two baselines were very similar, the information content would be reduced. It is therefore desirable to have the baselines possibly statistically uniformly distributed.

### C. Monte Carlo Simulations

In this section, we compare different spectral estimators using simulated data. Two cases were carried out. The first case considers only a single scatterer in the interest of exploring the effect of \( N \) and SNR on the estimation accuracy for micro-stacks and the performance of different estimators. The
Fig. 5. Visual comparison of NL-TomoSAR point clouds and TanDEM-X DEM over Munich, Germany. Color code: 565 m (blue) 596 m (red), scene size: 15 km \(\times\) 9 km, north = top. The violet bounding box indicates the region of interest (ROI) over the area of European bureau of patent and the white bounding box indicates the ROI near Munich central station. (a) Point Clouds generated by NL-TomoSAR with five interferograms. (b) TanDEM-X DEM.
second case considers double scatterers to investigate the estimation accuracy and the super-resolution power for different estimators. The inherent (Rayleigh) elevation resolution $\rho_s$ is inversely proportional to the maximal elevation aperture $\Delta b$.

$$\rho_s = \frac{\lambda_r}{2\Delta b}$$

(25)

The normalized distance is defined as

$$\kappa = \frac{s}{\rho_s}$$

(26)

For the first test case, only one scatterer is placed at $s = 0$, and the SNR is in the range between 0 and 30 dB. For each $N \cdot \text{SNR}$ value, 100 different baseline distributions were generated. We carried out a Monte Carlo simulation for each baseline distribution with 10,000 realizations. Afterwards, the CRLB was evaluated by averaging the value of 100 different baselines. Fig. 3(a) shows a performance comparison between SVD, and CS on simulated data with five acquisitions for a single scatterer. X-axis presents $N \cdot \text{SNR}$ in dB. Y-axis is the normalized CRLB $\sigma_s/\rho_s$. As one can see, both approaches have similar estimation accuracy. They are asymptotically towards the CRLB and collapse it when $N \cdot \text{SNR}$ is large. More interestingly is when $N \cdot \text{SNR}$ is fixed, leaving $N$ as a variable. Fig. 3(b) presents the estimation accuracy of SVD with $N = 3, 4, 5$. It shows that the accuracy when $N = 3$ is the smallest. This indicates that SNR carries more weight than $N$ on the estimation accuracy when $N$ is very small.

With the 3-D reconstruction accuracy of single scatterer clearly analyzed, we switch to the double scatterers case. In the simulation, the elevation of one scatterer is fixed at 0. The normalized elevation of the other scatterer is increased from 0.1 to 1.5, in order to mimic the layover of a ground layer and a facade layer. The number of acquisitions is set to $N = 3 - 5$, same as the first simulation. SNR is set to be 10 dB, since the SNR of TanDEM-X bistatic data is usually higher than this value in urban area.

The Monte Carlo simulation result is shown in Fig. 4. The x-axis represents the true normalized elevation distance $\kappa$ of the simulated facade and ground layers. Y-axis is the estimated normalized elevation distance $\hat{\kappa}$ of the simulated facade and ground layers. The two solid lines in the Fig. 4 (a) (b) (c) represent the true position of the building facade and the ground, respectively. The dashed lines imply the true position plus and minus the CRLB. The blue bar and dot imply the standard deviation and the mean of the estimated elevation of the facade scatterers, whereas the red ones represent those of the ground scatterers. The green dot indicates that the detection rate of double scatterers is below 5% and denotes the estimated result of the single scatterer. Fig. 4 (b) (c) show the estimated results by SVD and CS, respectively.

As one can see in Fig. 4(b) (c), the result of SVD has larger bias and slightly bigger standard deviation than CS. Note that, comparing to SVD, CS can give the better result, not only the accuracy of the estimation but also the super-resolution power. As one can see that the SVD has scarcely no super-resolution power, which can only distinguish double scatterers tilie one Rayleigh resolution $\rho_s$. In contrast, CS can achieve until 0.6 $\rho_s$.

IV. PRACTICAL DEMONSTRATION

A. Data Description

We make use of a stack of five co-registered TanDEM-X bistatic interferograms to evaluate the proposed algorithm. The dataset is over Munich, Germany, whose slant range resolution is 1.8 m and the azimuth resolution is 3.3 m. The images were acquired from July 2016 to April 2017. The most pertinent parameters of a TanDEM-X bistatic stripe map acquisition of Munich are listed in Table I and Table II. All the preprocessing steps, like deramping, are standard that are known from bistatic forest tomography. For interested readers, please refer to [2].

### Table I

| Name                      | Symbol | Value |
|---------------------------|--------|-------|
| Distance from the scene center | $r$     | 698 km|
| Wavelength                | $\lambda$ | 3.1 cm|
| Incidence angle at scene center | $\theta$ | 50.4° |
| Maximal elevation aperture | $\Delta b$ | 187.18 m|
| Number of interferograms  | $N$    | 5     |

### Table II

| No. | Date       | Baseline [m] | Height Ambiguity [m/cycle] |
|-----|------------|--------------|----------------------------|
| 1   | 2016-07-25 | 184.40       | 50.30                       |
| 2   | 2016-09-07 | 171.92       | 54.01                       |
| 3   | 2017-02-19 | 32.30        | 286.03                      |
| 4   | 2017-04-26 | -2.78        | -8710.99                    |
| 5   | 2017-07-01 | 9.30         | 1073.03                     |

B. Visual Comparison with TanDEM-X raw DEM

In this work, the TanDEM-X raw DEM is adopted for visual comparison with TomoSAR point clouds of the test area, which is formed by two TanDEM-X bistatic acquisitions using the Integrated TanDEM-X Processor (ITP).

A top view of the reconstructed point cloud of TomoSAR is shown in Fig. 5(a). The black regions in the figure is where the pixels are not coherent. The corresponding area of TanDEM-X raw DEM is presented in Fig. 5(b) as a comparison. It is clear that the result of TomoSAR point cloud preserves more detailed building structures. The road layer is also better represented in the TomoSAR result as well. In Fig. 5(b), the flat ground surface are well reconstructed. But when it comes to complex or high-rise buildings, their accuracy is compromised. For instance, the building of European bureau is compromised. For instance, the building of European bureau is compromised. For instance, the building of European bureau is compromised.
processing, the TanDEM-X raw DEM merges several buildings together and exhibits lower accuracy on the height of the buildings.

As another example, Fig. 7 shows the visual comparison over the area around Munich central station. It is clear that NL-TomoSAR result can show more detailed structures, such as the bridge, the central station, and roads.

V. QUANTITATIVE VALIDATION

In this section, we have quantitatively compared the TomoSAR point clouds with TanDEM-X raw DEM, as well as a much more precise LiDAR reference. The LiDAR dataset of Munich is provided by Bavarian State Office for Survey and Geoinformation with ten centimeter accuracy [35].

Since the TomoSAR point cloud is with respect to a reference point that was chosen during the TomoSAR processing, its location is not with respect to a geo-coordinate system. We coregistrated the point cloud of TomoSAR with the DEM and the LiDAR point cloud. In addition, in order to compare point clouds with DEM, we rasterize the two point clouds. These preparing steps are briefly explained in this section.

A. Geocoding

Since the result of TomoSAR inversion is a 3-D point cloud in the range-azimuth coordinate, the first step is to transform the result to Universal Transverse Mercator (UTM) coordinate with the Range-Doppler approach [36].

B. Coregistration of different point clouds

Consequently, when the TomoSAR point cloud is transformed to a UTM coordinate, its position may differ from the ground truth since the height of the reference point is unknown. Hence, the alignment of different point clouds is necessary. The most popular 3-D point cloud registration algorithm is iterative closest point (ICP) approach [37].

The performance of ICP depends on the initial alignment. Hence, a coarse alignment is adopted before applying ICP, which includes three steps: (1) the edge image is extracted by an edge detector, such as Sobel algorithm [38]; (2) the horizontal coregistration of two edge images is using cross-correlation of two edge images; (3) the vertical coregistration is using cross-correlation of the two height histograms. After the coarse alignment, ICP can be applied for the fine alignment [39].

C. Object-based raster data generation

The direct comparison of TomoSAR and LiDAR points is not feasible, as the central position of two corresponding points (one TomoSAR and one LiDAR point) and the footprint of the points are differing. Consequently, for comparing both data, an object-based raster need to be generated by using geographic information system GIS data.

D. Comparison of individual structure

In order to evaluate the estimation accuracy, nine test sites with high average SNR have been chosen for individual quantitative comparison. Fig. 8 shows the optical images of nine test sites for quantitative comparison of NL-TomoSAR point clouds and TanDEM-X DEM. They are (1) Munich central station, (2) European bureau of patent, (3) Technical University of Munich, (4) A railway signal light stand near Hirschgarten, (5) A train repair garage near Hirschgarten, (6) A residential building between two bridges (Hackerbrücke and Donnersbergebrücke), (7) Munich University of Applied Sciences, (8) A residential building near Lowenbrau beer company and (9) Karstadt (shopping mall). The summary of the results is shown in Tab. III.
From Tab. III we can see that the height differences between TomoSAR result and LiDAR data are within one meter and the height differences between TanDEM-X DEM product and LiDAR data vary from 2.5 m to 8.5 m. Similar performance is shown in standard deviation, for NL-TomoSAR is up to 1.4 m and for TanDEM-X DEM is up to 8.4 m.

E. Average accuracy

In order to have an assessment of the overall accuracy in a city scale, we compared all the 36,499 buildings in the area with the LiDAR point cloud. 38.7% buildings are within 1 m accuracy. 62.8% are within 2 m accuracy. A detailed distribution of accuracy is listed in Tab. IV. However, the two datasets (TanDEM-X CoSSC and LiDAR) were acquired at different time. It is almost certain that changes happened during the period. Therefore, in order to obtain a more realistic assessment, we truncated the distribution of height difference at ±15 m. The truncated histogram can be seen in Fig. 9. 34,054 buildings remains after the truncation. Their overall standard deviation is 1.96 m.

VI. CONCLUSION

A new SAR tomographic inversion framework tailored for very limited number of measurements is proposed in this paper. A systematic investigation of the estimation accuracy of TomoSAR with a micro-stacks is carried out using simulated data. Our experiments show that SVD and CS-based methods have almost identical performance on the estimation of single scatterer and the SNR plays a more important role than \( N \) for the estimation accuracy, when \( N \) is small. For the estimation
of double scatterers, CS-based approach outperforms the other spectral estimators. Experiments using TanDEM-X bistatic data shows the relative height accuracy of 2 m can be achieved in large scale. Thus it demonstrates the proposed framework being a promising solution 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] F. Lombardini, “Differential tomography: A new framework for SAR interferometry,” IEEE Trans. Geosci. Remote Sensing, vol. 43, no. 1, pp. 37-47, Jan. 2005.

[5] 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.

[6] G. Fornaro, D. Reale, and F. Serafino, “Four-dimensional SAR imaging for height estimation and monitoring of single and double scatterers,” IEEE Trans. Geosci. Remote Sensing, vol. 47, no. 1, pp. 224-237, Jan. 2009.

[7] 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.

[8] 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.

[9] X. X. Zhu, Y. Wang, S. Montazeri, N. Ge, “A Review of Ten-Year Advances of Multi-Baseline SAR Interferometry Using TerraSAR-X Data,” Remote Sensing, vol. 10, no. 9, pp. 1374, Aug. 2018.

[10] 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.

[11] X. X. Zhu, R. Bamler, and R. Bamler, “Tomo-GENESIS: DLR’s tomographic SAR processing system,” Proc. Joint Urban Remote Sensing Event, 2013 pp. 159-162.

[12] G. Fornaro, A. Pauciullo, D. Reale, and S. Verde, “Multilook SAR
