IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-0369-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/IGARSS47720.2021.9553691

GENERATION OF LARGE SCALE 3-D CITY MODELS USING INSAR AND OPTICAL DATA

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

[1] Chair of Remote Sensing Technology (LMF), Technical University of Munich, Munich, Germany
[2] Data Science in Earth Observation (SiPEO), Technical University of Munich, Munich, Germany
[3] Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Wessling, Germany

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

1. INTRODUCTION

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 [1][2][3][4]. 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 [4] or COSMO-Skymed [5]. Due to modern development in signal processing, the compressive sensing (CS) based methods like SL1MMER and others [6] [7] can achieve unprecedented accuracy for 3-D reconstruction and show the super-resolution (SR) power, which is not exhibited with conventional approaches.

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 [8], 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. There is also exsiting algorithm handling the micro-stack TomoSAR problem. In [9], 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.

Therefore, 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. 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. We use LiDAR data as reference to compare with our reconstructed TomoSAR point cloud.

2. 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.

2.1. The Multi-Master SAR Imaging Model

The TomoSAR imaging model can be formulated as :

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

Where the term $k_n = 4\pi b_n/\lambda r$ and the term g_n is the measurement of the nth 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. 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 $b = \lambda r/8s_0$, 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

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

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 Δk :

$$E\left\{\Gamma(k_n + \Delta k_n)\Gamma^*(k)\right\} = \int \sigma_0(s) \exp(-j\Delta k_n s) ds \tag{3}$$

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

$$\mathbf{g} = \mathbf{R}\mathbf{X} + \boldsymbol{\varepsilon} \tag{4}$$

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 $R_{nl} = \exp(-j\Delta k_n s_l)$.

2.2. Non-Local Procedure

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:

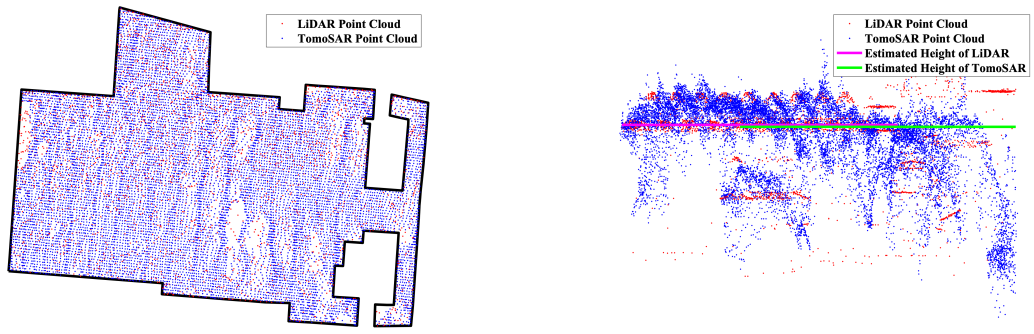
$$\hat{\boldsymbol{\Theta}}_c = \operatorname{argmax} \sum_s \mathbf{w}(i_s, j_s) \log p(\mathbf{g}_s|\boldsymbol{\Theta}) \tag{5}$$

where weights $\mathbf{w}(i_s, j_s)$ can be measured by calulating the patch-wise similarity [10] [11]. 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.

2.3. 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, which includes singular value decomposition (SVD) [2][3], compressive sensing (CS) [12][13].

In this work, we use similar strategy introduced in [14]. 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 [15] can speedup the processing of L_1 minimization.



(a) (b)

Fig. 1. 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).

2.4. 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 $\rho\left(\right)$ of the residuals:

$$\tilde{s} = \arg\min_{s} \sum_{i} \rho\left(\hat{s}_i - s\right), \tag{6}$$

where $\hat{s}_i$ represents the estimated elevation of the ith pixel. The alternative interpretation of Eq. (6) is a weighted averaing of the heights of the neighbouring pixels. The weighting function can be written in this formulation.

$$w\left(x\right) = \frac{-\partial\rho\left(x\right)}{x\partial x} \tag{7}$$

3. PRACTICAL DEMONSTRATION

3.1. 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).

3.2. 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. 1. Fig. 1 (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. 1 (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. There are still 34,054 buildings left after truncation. The standard deviation of height difference is about 1.96 m [16].

3.3. 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. 2 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.

Fig. 2. Visualization of fused 3-D urban model of Munich.

4. CONCLUSION

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

5. REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.*, vol. 58, pp. 8026-8037, May 2020.