VERY HIGH RESOLUTION SAR TOMOGRAPHY VIA COMPRESSIVE SENSING

Xiao Xiang Zhu(1), Richard Bamler(1,2)

(1)Technische Universität München (TUM), Remote Sensing Technology, Arcisstraße 21, 80333 München, Germany
Email: xiao.zhu@dlr.de
(2)German Aerospace Center (DLR), Remote Sensing Technology Institute (IMF), Oberpfaffenhofen, 82234 Wessling, Germany
Email: richard.bamler@dlr.de

ABSTRACT

By using multi-pass SAR acquisitions, SAR tomography (TomoSAR) extends the synthetic aperture principle into the elevation direction for 3-D imaging. Since the orbits of modern space-borne SAR systems, like TerraSAR-X, are tightly controlled, the elevation resolution (depending on the elevation aperture size) is at least an order of magnitude lower than in range and azimuth. Hence, super-resolution algorithms are desired. The high anisotropic 3-D resolution element renders the signals sparse in elevation. This property suggests using compressive sensing (CS) methods. The paper presents the theory of 4-D (i.e. space-time) CS TomoSAR and compares it with classical tomographic methods. Super-resolution properties and point localization accuracies are demonstrated using simulations and real data. A CS reconstruction of a building complex from TerraSAR-X spotlight data is presented. In addition, the model based time warp method for differential tomographic non-linear motion monitoring is proposed and validated by reconstructing seasonal motion (caused by thermal expansion) of a building complex.

1. INTRODUCTION

SAR tomography (TomoSAR) extends the synthetic aperture principle of SAR into the elevation direction s for 3-D imaging. It uses acquisitions from slightly different viewing angles to reconstruct for every azimuth-range (x-r) pixel the reflectivity function along the elevation s. Differential TomoSAR, also referred to as 4-D focusing, obtains a 4-D scatterer map by estimating the elevation and the deformation velocity of multiple scatterers inside an azimuth-range cell.

Currently, very high spatial resolution (VHR) SAR satellites (e.g. TerraSAR-X and COSMO-Skymed) can provide data of up to 1m resolution. These are particularly helpful for tomographic imaging of buildings and infrastructure. We work with TerraSAR-X high resolution spotlight data. These data stacks of urban areas have some particular properties [1]: A very detailed view of individual buildings is possible. But also nonlinear motions of different building parts must be expected which introduce additional phase errors, if not modeled. Due to the tight orbit tube of TerraSAR-X [2], the inherent elevation resolution is about 50 times worse than in azimuth or range. This extreme anisotropy calls for super-resolution algorithms. We concentrate on single-look super-resolution methods to exploit the potential of VHR data. All methods that require multi-look estimates of covariance matrices (e.g. CAPON or MUSIC) would reduce the azimuth-range resolution and are not able to resolve structural building elements in the important meter scale.

Compressive sensing (CS) [3], [4] is a new and attractive technique for TomoSAR. It aims at minimizing the number of measurements to be taken from signals while still retaining the information necessary to approximate them well. It provides a good compromise between classical parametric and non-parametric spectral analysis methods. Compared to parametric spectral analysis, CS is more robust to phase noise, has lower computational effort, and does not require model selection to provide the prior knowledge about the number of scatterers in a resolution cell. Compared to non-parametric spectral estimation CS does not suffer from sidelobe interference and overcomes the limitation of elevation resolution caused by the limited extent of the elevation aperture, i.e. CS has super-resolution properties. In this paper the CS approach to TomoSAR is outlined. Its extension to differential (4-D) TomoSAR is introduced. Numerical simulations for realistic acquisition and noise scenarios will be presented to evaluate the potential and limits of the technique. The first CS TomoSAR results with TerraSAR-X spotlight data over urban areas will be presented.

2. TOMOSAR VIA COMPRESSIVE SENSING

2.1 System and Noise Model

In VHR X-band data, the signal contributions we expect are the following (see Figure 1) [5]: Strong returns (red part) from metallic structures or specular and dihedral or trihedral reflections result in points in points that would also be used in Persistent Scatter Interferometry. They are the dominating signal contributions. With VHR SAR data the density of these points can be very high. Weak diffuse scattering from – mostly horizontal or vertical – rough surfaces (roads, building walls). These objects have an elevation extent of $\rho / \tan(\theta - \alpha)$, where $\rho$ is the (slant) range resolution, $\theta$ the local incidence angle, and $\alpha$ the slope of the surface relative to horizontal. Except from large surfaces accidentally oriented along elevation, these responses are of much smaller extent than the elevation
resolution $\rho$, and, hence, they can be treated as discrete scatterers in the elevation direction (delta-functions). Returns from volumetric scatterers, e.g. from vegetation result in a continuous signal background in elevation. These ensembles of scatterers, however, often decorrelate in time, and their response is treated as noise.

The noise sources are the following: Gaussian noise which is caused by thermal noise and temporal decorrelation as mentioned above. Calibration errors in amplitude. The radiometric stability of TerraSAR-X, i.e. the amplitude variations within one stack is 0.14 dB and is therefore negligible compared to our typical SNR. Phase errors caused by atmospheric delay and unmodeled motion. They require robust and phase-error-tolerant estimation methods.

These considerations suggest that the elevation signal to be reconstructed is sparse in the object domain, i.e. it can be described by a few point-like contributions of unknown positions and unknown amplitudes and phases. Sparsity is the central concept of and a prerequisite for CS.

![Figure 1. Possible signal contributions in a single SAR image azimuth-range pixel. $\rho_r$ and $\rho_e$ : range and elevation resolution, respectively (size of resolution cells not to scale)](image)

### 2.2 TomoSAR Imaging Model

For a single SAR acquisition, the focused complex-valued measurement $g_n(x_o, \xi)$ of an azimuth-range pixel $(x_o, \xi)$ for the $n^{th}$ acquisition at aperture position $b_n$ and at time $t_n$ is the integral (tomographic projection) of the reflected signal along the elevation direction [6] (the deformation term is ignored here for simplicity) [7]:

$$g_n = \int_{x_o} \gamma(s) \exp(-j \xi s) ds, \quad n = 1,...,N \quad (1)$$

where $\gamma(s)$ represents the reflectivity function along elevation $s$. $\xi = -2b_n/\lambda r$ is the spatial (elevation) frequency. The continuous space system model of equation (1) can be approximated by discretizing the continuous reflectivity function along $s$ (ignoring an inconsequential constant):

$$g = R\gamma \quad (2)$$

where $g$ is the measurements vector with the $N$ elements, $R$ is an $N \times L$ mapping matrix with $R_{nl} = \exp(-j2\pi \xi s_l)$ and $\gamma$ is the discrete reflectivity vector with $L$ elements $\gamma_l = \gamma(s_l)$. $s_l (l = 1,...,L)$ are the discrete elevation positions. Equation (1) is an irregularly sampled discrete Fourier transform of the elevation profile $\gamma(s)$. The objective of TomoSAR is to retrieve the reflectivity profile for each azimuth-range pixel.

### 2.3 TomoSAR via Compressive Sensing

Compressive sensing (CS) is a favourable approach for sparse signal reconstruction. As described in the preceding section and in [5] [8], for VHR space-borne X-band TomoSAR the signal $\gamma$ to be reconstructed has typically 1-3 point-like contributions of unknown positions, amplitudes and phases, i.e. $\gamma$ is sparse in the identity orthogonal basis $I$. In order to measure the sparse signal $\gamma$ efficiently, the sensing matrix $R$ should spread out the information of highly localized sparse signals in the entire projection space and thus makes them insensitive to “undersampling”. Otherwise the reconstruction of non-zeros coefficients will be biased towards certain positions. This properties is the so called incoherence between the sensing matrix and orthogonal basis. According to equations (1) and (2) the sensing matrix $R$ is a randomly distributed Fourier sampling matrix which has the best incoherence property with $I$. The aim is to find the solution of $\gamma$ with least number of scatterers, i.e. minimal $L_0$ norm, to satisfy measurements with noise:

$$\min_{\gamma} \|R\gamma\|_0 \quad \text{s.t. } g = R\gamma \quad (3)$$

For $N = O\left(K \log(L / K)\right)$, which can be very easily fulfilled due to the small number $K$ of scatterers, it can be shown that $L_1$ norm minimization leads to the same result as $L_0$ norm minimization. Hence, $\gamma$ can be exactly recovered in the absence of noise by $L_1$ minimization:

$$\min_{\gamma} \|R\gamma\|_1 \quad \text{s.t. } g = R\gamma \quad (4)$$

In case there is no prior knowledge about $K$ and in the presence of noise, the solution can be approximated by:

$$\hat{\gamma} = \arg\min_{\gamma} \left\{ \|h - R\gamma\|_2^2 + \lambda_c \|\gamma\|_1 \right\} \quad (5)$$

$\lambda_c$ is a factor adjusted according to the noise level. The choice of $\lambda_c$ is described detailed in [9]. Equation (5) can be solved by Basis Pursuit (BP) methods. Instead of detecting $K$ most significant coefficients, it minimizes the residual by employing an $L_1$ norm regularization. By providing the over-completeness of $\gamma$ (i.e. several close spectral lines for one scatterer), it can provide more robust solutions.
2.4 Extension to differential TomoSAR

The extension to the 4-D (space-time) case is straightforward [10], [5]. Taking the motion term into account, the system model (1) can be extended to:

\[
g_n = \int \gamma(s) \delta(v-V(s)) \exp(-j2\pi(\frac{\Delta s}{\lambda} + \Delta v)) ds dv (6)
\]

where \(\Delta v\) is the range of possible velocities, \(V(s)\) is the deformation LOS velocity profile along elevation and \(\eta_s = 2r_s / \lambda\) may in analogy be called a “velocity frequency”. Equation (6) is a 2-D Fourier transform of \(\gamma(s) \delta(v-V(s))\) which is a delta-line in the elevation-velocity (s-v) plane along \(v = V(s)\). Its projection onto the elevation axis \(\gamma(s) \delta(v-V(s))\) is the reflectivity profile \(\gamma(s)\). If we accept \(\gamma(s) \delta(v-V(s))\) as the object to be reconstructed, the discretized system model of equation (2) is easily adopted and simply becomes a 2-D Fourier transform [10] [7]. Its inversion provides retrieval of the elevation and deformation information even of multiple scatterers inside an azimuth-range resolution cell and thus obtains a 4-D map of scatterers. It is required for reliable 3-D and 4-D city mapping from repeat-pass acquisitions.

2.5 Experiments

2.5.1 The Data Set

In this paper, we work with TerraSAR-X spotlight data with a resolution of 0.6m in range and 1m in azimuth. Our test site is Las Vegas, Nevada, USA. The orbit of TerraSAR-X is controlled in a predefined tube of 500m diameter. A data stack of 25 scenes is used for our experiment. The elevation aperture positions are shown in Figure 2 with an elevation aperture size \(\Delta b\) of about 269.5m.

\[
\text{Figure 2. Elevation aperture positions [m]}
\]

For non-parametric linear spectral analysis, the expected \(\rho_s\) is approximately 40.5m for. The exact value for \(\rho_s\) depends on additional factors (e.g. spectral weighting, the real elevation aperture positions). With our aperture sampling, the 3dB resolution is \(\rho_s = 33 \text{ m}\) or about 16m in height \(z\) (look angle =31.8°). This, however, does not mean the elevation localization accuracy of individual scatterer is poor. The Cramér-Rao lower bound (CRLB) [5] on elevation estimates suggests a 1.1m (i.e. 1/30 of \(\rho_s\)) accuracy can be achieved with this stack under 10dB SNR.

2.5.2 Simulated Data

In this section, the CS approach is compared to conventional non-parametric and parametric methods using simulated data. The data is simulated using the elevation sampling in Figure 2. The decorrelation effect is introduced by adding Gaussian noise with certain SNR. Phase noise due to unmodeled deformation and atmospheric effects are simulated by adding a uniformly distributed phase.

\[
\text{Figure 3. Reconstructed reflectivity profiles along elevation s: SVD-Wiener vs. CS with SNR=10dB. Red: SVD-Wiener; Blue: CS; Green: ±3CRLB. left: s=-20, 25m; right: s=0, 20m.}
\]

Figure 3 shows the reconstructed reflectivity profile along elevation \(s\) using a singular value decomposition method with Wiener-type regularization (SVD-Wiener) [5] and CS. Red lines represent the reconstruction using SVD-Wiener. Blue lines show the same result using CS. We start with two scatterers with elevation of -25m and 20m (SNR=10dB), both methods can distinguish the scatterers well (left). CS reconstructs spectral lines instead of a sinc-like point response function (PRF). However, once they move closely into one elevation resolution cell with elevation of 0m and 20m, SVD-Wiener is not able to distinguish them. CS detects very clearly two spectral lines with an accuracy within ±3 times CRLB (right). With higher SNR, CS is even able to distinguish closer scatterers.

In [8], the situation that two scatterers inside one resolution cell (one from the building façade and another from the ground) is simulated as an example to evaluate the performance of the spectral estimation methods. CS has been compared to with maxima detection (MD) [11] and nonlinear least-squares (NLS) among them MD simply uses the maxima of the non-parametric SVD-Wiener reconstruction as estimates and NLS is the theoretically best solution under Gaussian noise. Compared to MD, beside the super-resolution property, CS shows no sidelobe interference problem. Compared to NLS, CS shows comparable performance with lower computational effort and does not require the number of scatterers as a prior.
Figure 4 shows the elevation estimation accuracy of a single scatterer in phase noise free case using NLS and CS compared to the CRLB under different SNR levels. NLS as the MLE estimator with Gaussian noise shows an accuracy consistent to CRLB. And the estimation accuracy of CS is almost identical to NLS. Figure 5 shows the plot by adding a phase noise uniformly distributed in \([-\phi_0, \phi_0]\) under SNR=20dB. It shows that the estimation accuracy highly depends on the phase noise and CS has more robust performance against non-Gaussian phase noise. Taking all those aspects into account, CS provides the best of the two worlds of non-parametric and parametric spectral estimation methods and, hence, is proven very attractive for TomoSAR.

![Figure 4. Single scatterer elevation estimation accuracy of NLS and CS compared to the CRLB as a function of SNR](image)

![Figure 5. Same plot as in Figure 4 with different phase noise levels (SNR=20dB).](image)

2.5.3 Real Data

A. CS TomoSAR

We use the Las Vegas convention center as a test site. It has a height of about 20m, the critical distinguishable distance between two scatterers by using SVD-Wiener for our elevation aperture size. The presence of two scatterers within azimuth-range pixels is expected in layover areas and has been validated in [5]. Thus, we are able to compare the performance of CS at the layover areas to the SVD-Wiener. The left image in Figure 6 shows its TS-X intensity map. We choose a reference pixel (green point). The bright blue line shows the position of the analysis slice and the area marked by the red block is a layover area. There is a triangular shaped plaza on the ground made of the same material as the building. Thereby, multiple scatterers are expected. The middle lower image shows the estimated reflectivity with SVD-Wiener in the x-s plane. Multiple scatterers with marginally distinguishable distance appear (one from the building: blue line, the other from the ground structure: yellow line). Although it demonstrates the stability of SVD-Wiener, the resolution limit blurs the reflectivity profile. In contrast, the middle upper image shows the same result estimated by CS. Compared to SVD, besides the layover area can be separated, the elevation positions can be easily located.

B. Differential CS TomoSAR

The CS approach is implemented to differential TomoSAR. Figure 7 shows the 4-D reconstruction for the same pixel shown in Figure 6. Again two scatterers with slightly different velocities have been detected in the s-v plane by SVD-Wiener (left). But the scatterer on the ground appears much brighter and wider. It is very likely having two non-separable scatterers. The middle image in Figure 7 shows the result using CS. Two very close scatterers have been detected. The right image shows the projections to the elevation, i.e. the reflectivity reconstruction by differential TomoSAR. Compared to Figure 6, differential TomoSAR allows to locate the scatterers more precisely by separating coupled linear deformation, in particular, differential TomoSAR via CS provides super resolution up to 2m in height (i.e. about 4m in elevation).

Figure 8 shows the final surface model with deformation correction over-layered by the reflectivity map, i.e. 3-D SAR image. This visualizes for the first time in detail how the convention center would look like from the position of TerraSAR-X if our eyes could see X-band radiation. This may lead to a better understanding of the nature of scattering. For instance, an overview about the multiple bounce can be acquired by looking at the very bright structure in Figure 8. Also the very bright individual scatterers who behave as corner reflectors can be precisely located. This may help in finding natural corner reflectors. For each scatterer, there is also a LOS linear deformation velocity estimates. Most of them show split peaks in the velocity direction which suggests the appearance of non-linear motion. Since the deformation is presumably caused by thermal dilation, it rather follows a periodic seasonal model. This will be further discussed in the next section.
Figure 6. Left: TerraSAR-X intensity map of Las Vegas convention center; Right: estimated reflectivity with SVD-Wiener (top) & CS (bottom) shown in the azimuth-elevation plane (horizontal: azimuth; vertical: elevation, converted to height [m]).

Figure 7. 4-D reconstruction example for a pixel located at the layover area by using SVD-Wiener (left) and CS (middle). Right plot shows the projection on the elevation.

Figure 8. The world at X-band: Tomographic surface reconstruction overlayed by 3-D reflectivity map

Figure 9. Projecting the temporal baselines $t_n$ to artificial temporal baselines $\tau_n$.
Figure 10. 4-D reconstruction example with seasonal motion by using SVD-Wiener: Linear motion model (left) and time warp method with seasonal motion model (right). s=10, 70m; amplitude of seasonal motion=8, 2mm; SNR=10dB.

Figure 11. Left: reconstructed digital surface model (DSM) from differential ToMoSAR, [unit: m] and estimated amplitude of seasonal motion using time warp method [unit: mm]

3. LET’S DO THE TIME WARP

With the approximately 1 year time spread of our data set, non-linear (e.g. thermally induced) movements of different building parts must be expected [12]. In this section, the time warp method will be introduced for differential tomographic non-linear motion estimation. Taking the seasonal motion as an example, it suggests assuming the following model:

$$g_{n} = \int_{\Delta \tau} r(s) \exp \left( -j2\pi \left( \xi_{n} s + \frac{2}{\lambda} \sin(2\pi (t_{n} - \tau_{0})) a(s) \right) \right) ds$$  \hspace{1cm} (7)

where $a(s)$ represents the amplitude of seasonal motion along $s$ and $\tau_{0}$ is the initial temporal offset of the master image which can be estimated from the temperature history of the test area.

By projecting the temporal baseline $\tau_{n}$ to the artificial temporal baselines $\tau_{n} = \sin \left( 2\pi \left( t_{n} - \tau_{0} \right) \right)$, with introducing the new motion frequency $\eta_{n} = 2\tau_{n} / \lambda$, the system model (7) can be rewritten as:

$$g_{n} = \int_{\Delta \tau} r(s) \exp \left( -j2\pi \left( \xi_{n} s + \eta_{n} a(s) \right) \right) ds$$  \hspace{1cm} (8)

With this formulation, the motion term can be focused in motion parameter space which is amplitude of seasonal motion in this case. Figure 9 shows the time warp procedure where the $x$-axis represents the real temporal baseline distribution of the data stack and $y$-axis shows the distribution after projection (here the initial offset is 2 months). After time warping, we are able to estimate the elevations and the amplitudes of seasonal motion in the elevation-amplitude plane.

In Figure 10, the situation that two scatterers with elevations 10m and 70m and amplitudes of seasonal motion 8mm and 2mm, respectively, inside one azimuth-range cell has been simulated under 10dB SNR. The peak in velocity direction gets spread out along $v$ which leads to
The new class high resolution spotlight TerraSAR-X data is very attractive for 3-D and 4-D tomographic mapping in urban environment. A stack of TerraSAR-X high resolution spotlight data over the city of Las Vegas is used. Compared to the medium resolution SAR systems available so far, the information content and level of detail has increased dramatically. A full tomographic high resolution reconstruction of a building complex is presented. Recognizing the sparsity of the signal in elevation, Compressive Sensing, as a new and favorable technique for sparse signal reconstruction, has been implemented to TomoSAR and differential TomoSAR. It provides a very elegant compromise between conventional non-parametric and parametric tomographic methods. For instance, it shows good robustness w.r.t. unmodeled non-Gaussian phase noise. Compared to non-parametric methods, it provides super-resolution properties without sacrificing the azimuth or range resolution; it does not suffer from the sidelobe interference effect. Compared to parametric methods, like NLS, in the single scatterer case and under Gaussian noise CS approaches the accuracy of NLS with lower computational effort. In addition, CS does not need model selection, i.e. it “automatically” chooses the number of scatterers that can be resolved. A new model-based time warp method has been proposed for differential tomographic nonlinear motion monitoring. By forming an artificial temporal baseline, it provides the possibility to focusing the desired parameter, e.g. amplitude of seasonal motion, to the coefficient space. Time warp method has been validated by reconstructing seasonal motion caused by thermal expansion of a building complex.

REFERENCE


