# 3D Buildings Reconstruction with SAR Tomography Guided by Partial Footprints Information

Clément Rambour, Loïc Denis, Florence Tupin

# 3D Buildings Reconstruction with SAR Tomography Guided by Partial Footprints Information

Clément Rambour[a,b], Loïc Denis[c], and Florence Tupin[d]
[a]CNAM, Paris, France
[b]ONERA, The French Aerospace Laboratory, 91761 Palaiseau, France
[c]Univ Lyon, UJM-Saint-Etienne, CNRS, Institut d Optique Graduate School, Laboratoire Hubert Curien UMR 5516, F-42023, SAINT- ETIENNE, France
[d]LTCI, Télécom Paris, Institut Polytechnique de Paris, France

## Abstract

SAR tomography is a powerful 3D reconstruction method. Nevertheless, in urban areas, results are usually rather sparse and some post-processing is needed to obtain 3D buildings. In this paper, external information about building footprints is introduced in the 3D reconstruction process. This is done by extending the recent method REDRESS based on a graph-cut method (Rambour *et al.*, *Computer Vision and Image Understanding* 2019). The graph construction is modified to take into account spatially variable weights depending on the footprints location. Experimental results show a clear improvement of the retrieved shapes.

## 1 Introduction

Thanks to the phase of the complex backscattered signal, SAR imaging has great capacities to recover information on the elevation and the movement of the illuminated structures. Nevertheless, in urban areas, this information is often sparse, restricted to specific local configurations like dihedral or trihedral shapes that back-scattering a strong echo. To "densify" the measurements, a possible solution is to combine SAR images with heterogeneous data like optical images or GIS information. One can cite the use of optical data for elevation recovery using single SAR image [1], interferometric [2] or radargrammetric data [3], or information provided by a GIS [4]. The shape information provided by the external source helps to regularize the SAR measurements.

In this paper, we investigate how the building footprints can improve the 3D information recovery in a tomographic framework. To this aim, we propose to extend our recently proposed approach [5], to take into account this information. This is done by locally modifying the weights in the graph in order to favor cuts in agreement with the footprints.

## 2 Urban SAR tomography

A SAR tomographic stack is built from a set of $N$ co-registered images of size $N_x \times N_r$. In the absence of scatterer motion and after proper phase calibration, the complex value of a pixel in the image $n$ at the azimuth-range position $(x, r)$ is given by:

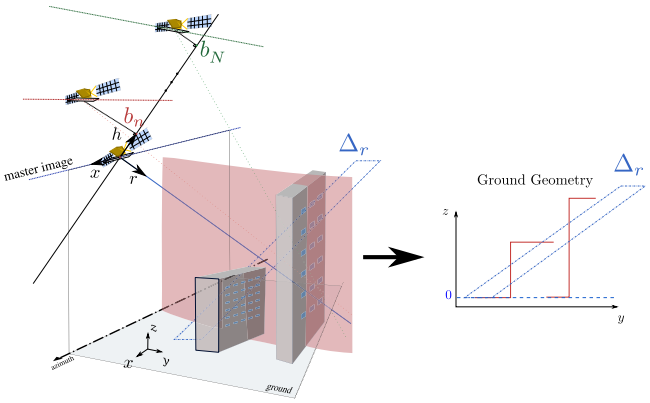


**Figure 1** Principle of SAR tomography. Using different sensor positions, it is possible to unmix the backscattered signals in $\Delta_r$ and recover the different contributions inside the resolution cell.

$$v_n(x,r) = \iint_{(y,z)\in\Delta_r} u(x,y,z)e^{-j\xi_n z}\mathrm{d}y\mathrm{d}z\,. \quad (1)$$

As depicted in Fig.1, the 3-D reflectivity function $u(x, y, z)$ is summed along the radar resolution cell extent $\Delta_r$ to model the value $v_n(x, r)$ at the pixel with range $r$ and azimuth $x$. The complex exponential contains the interferometric information that relates the $n^{th}$ antenna position to the height $z$ of the scatterers. Under some classical approximations [6], the interferometric phase can be considered directly proportional to $z$. The parameter $\xi_n = \frac{4\pi b_{\perp_n}}{\lambda r \sin\theta}$ corresponds to the angular frequency associated to the antenna $n$, with $b_{\perp_n}$ the orthogonal baseline, $\theta$ the incidence angle and $\lambda$ the wavelength.

Estimators such as MUSIC [7] can be used to recover

$u$ from an estimate of the complex covariance matrix computed at the pixel $(x, r)$. The Compressive Sensing (CS) approach directly inverts the scattering vector $\boldsymbol{v} = [v_1, \ldots, v_N]^{\mathrm{t}}$. It has been shown to be more adapted to dense urban landscapes [8, 9]. It is based on a regularized inversion with a sparsity-promoting $\ell_1$ term:

$$\hat{\boldsymbol{u}} = \arg\min_{\boldsymbol{u}} \quad \|\boldsymbol{P}\boldsymbol{u} - \boldsymbol{v}\|_2^2 + \mu\|\boldsymbol{u}\|_1 \,. \tag{2}$$

The vector $\boldsymbol{u} \in \mathbb{C}^{(N_x \cdot N_r \cdot N_z)}$ contains the estimated 3D distribution of reflectivities. The parameter $\mu$ controls the sparsity of the estimation. The matrix $\boldsymbol{P} \in \mathbb{C}^{(N \cdot N_x \cdot N_r) \times (N_x \cdot N_r \cdot N_z)}$ performs the projection of the voxels in $\boldsymbol{u}$ onto the SAR image stack and accounts for the phase shift induced in each image by the optical path length. Here, the matrix $\boldsymbol{P}$ is defined like in the REDRESS method [5] as a mapping from voxels uniformly sampled in the 3D ground coordinates system $(x, y, z)$ to pixels in the SAR coordinates $(x, r)$.

# 3 Urban surface segmentation

Raw reconstructions of dense urban areas obtained with SAR tomography are generally difficult to interpret for non-experts. Extracting meaningful information from these estimations is still a challenging task and generally requires good priors on the observed scene. For urban areas, a recent approach consists of modeling the spatial distribution of the scatterers as a surface corresponding to the ground, the walls and rooftops [5, 10]. These methods have proved to work well for urban areas and show promising result in dense urban areas. Here we propose to extend the REDRESS algorithm [5] to take into account additional information on the 3-D scene.

The original version of REDRESS alternatively performs an estimation of the reflectivity of the scene and a segmentation of the urban surface. To perform the segmentation, a cost function composed of a data fidelity term $P_d$ and a regularization term $P_s$ is defined on all possible segmentations. First, to enforce a good fidelity to the tomographic estimation, surface $\mathscr{S}$ located near high intensity voxels are favored. Since the radar wave does not penetrate inside the buildings, along each ray that emerges from the radar, only one intersection with the visible surface is possible. To account for these considerations, we define a cost function $D$ that is applied to all incident rays:

$$D(\rho) = \int_{\rho_{\min}}^{\rho} [C_f(s) - C_b(s)]^+ \mathrm{d}s \tag{3}$$

$$+ \int_{\rho_{\min}}^{\rho} [C_b(s) - C_f(s)]^+ \mathrm{d}s \,, \tag{4}$$

where $\rho$ is the position along the ray and $[.]^+$ stands for the so-called relu function (i.e., the positive part). The functions $C_f$ and $C_b$ are the the forward and backward cumulative sum of the reflectivity defined as:

$$C_f(\rho) = \int_{\rho_{\min}}^{\rho} |u(\rho)| \mathrm{d}\rho \qquad C_b(\rho) = \int_{\rho}^{\rho_{\max}} |u(\rho)| \mathrm{d}\rho \tag{5}$$

The function $D$ has a unique minimum which is met where the reflectivity is equal to its median along the ray. The global data penalty term $P_d$ is then defined for a surface $\mathscr{S}$ as the sum of the cost function $D$ over the set $\mathscr{R}$ of all incident rays in the scene:

$$P_d(\mathscr{S}) = \int_{\mathscr{R}} D_\tau(\rho_{\tau \to \mathscr{S}}) \mathrm{d}\tau \,, \tag{6}$$

where $\tau$ is a ray, $D_\tau$ is the cost function $D$ defined for that ray, and $\rho_{\tau \to \mathscr{S}}$ is the distance from the sensor to the surface $\mathscr{S}$ along ray $\tau$.

To take into account the typical geometrical patterns observed in urban areas, the surface should also favor smooth vertical and horizontal regions. This kind of behavior is introduced by penalizing the $\ell_1$ norm of its variations (i.e., the area of the surface):

$$P_s(\mathscr{S}) = \iiint \|\nabla \mathscr{S}(x, y, z)\| \, \mathrm{d}x\mathrm{d}y\mathrm{d}z \tag{7}$$

with $\nabla$ the gradient of the surface. Finally, the estimated urban surface should be the solution of the following optimization problem:

$$\hat{\mathscr{S}} = \arg\min_{\mathscr{S}} \quad P_d(\mathscr{S}) + \beta P_s(\mathscr{S}) \tag{8}$$

The parameter $\beta$ sets the importance of the data fidelity term with respect to the spatial regularization. The optimization problem (8) is solved using a graph-cut approach. We detail the structure of the graph in the next part.

# 4 Adaptive spatial regularization using the buildings footprint

The correct segmentation of the surface may be hard to achieve in configurations where large parts of walls do not backscatter any signal. In such cases, the spatial regularization may favor surfaces that skip isolated building parts (which is favored by the regularization term). The result may then show only the bottom part of these buildings. The buildings footprint can be used to reduce the regularization parameter $\beta$ where discontinuities of the surface are expected.

The structure of the graph is illustrated in Fig. 2. It is constructed based on the discretized volume of reflectivities and has a node for each voxel in $\hat{\boldsymbol{u}}$, plus two terminal nodes: the source and the sink. All the nodes are connected to their horizontal neighbors with edges of constant capacity $\beta$ except at the border of the footprints where $\beta$ is replaced by a small value $\epsilon \ll \beta$. In our experiments, we have set $\epsilon = 10^{-2}$. Finally, infinite-capacity edges connect each node to its next vertical neighbor in order to forbid the surface to intersect twice a vertical line (such a surface is deemed unlikely in urban areas). The graph is built in such a way that the cuts, i.e., the partition of the nodes into two distinct subsets, one containing the source, the other containing the sink, have associated costs that match the value of the cost function for the corresponding 3D surface. An efficient min-cut/max-flow algorithm can identify the minimal cut, i.e., the optimal surface according to the cost function that we designed.

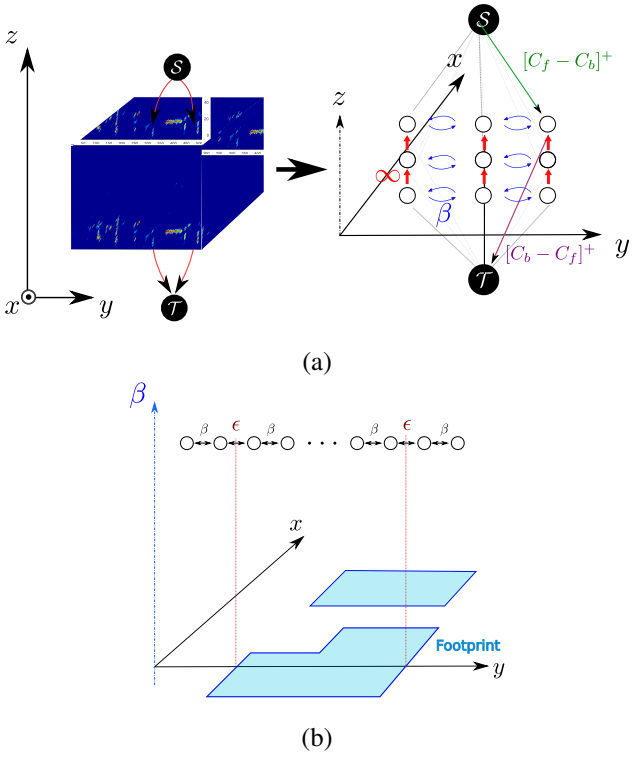


**Figure 2** (a) The graph is built from the reconstructed volume with one node for each voxel of the volume. Nodes are connected to the source, the sink and their immediate spatial neighbors. The value of a cut in the graph corresponds to the sum of the terms $[C_f - C_b]^+$ for the nodes above the cut, the terms $[C_b - C_f]^+$ for the nodes below the cut, and $\beta$ times the number of horizontal edges cut. The optimal surface is defined by the minimal cut in the graph. Infinite capacity edges connect each node to its upper neighbor to ensure that the surface does not intersect twice any given vertical line. (b) The value of the spatial regularization parameter $\beta$ is spatially adapted: it is reduced in the vicinity of the buildings footprint.

# 5 Results

The validation of the proposed approach is performed using a dataset of 20 TerraSAR-X images of Paris[1] and a map of the buildings footprint provided by the french National Institute of Geographic and Forest Information (IGN). The average image of the stack and the footprints are presented in Fig. 3. This dataset contains four buildings with a height above 75 meters and large footprints. Two of them backscatter a high-intensity signal and are clearly visible in the SAR images. However, only parts of the walls of the building in the left of the images are visible and the one in the middle is almost invisible in the SAR image. The result obtained with a spatially variant parameter $\beta$ is shown in Fig. 4. It is clear that the footprints provide an important additional information for the reconstruction of the scene. In the final version of the paper, a comparison and discussion on different strategies for the choice of $\beta$ will be provided.

[1]These data have been obtained in the framework of the DLR project LAN-176.

# 6 Conclusion

In this paper, we presented an extension to the urban surface and tomographic inversion algorithm REDRESS [5]. This extension includes exogenous information such as building footprints within the surface segmentation method. This kind of information is particularly useful when large parts of a building are not visible in the tomographic inversion. Improvement in the reconstruction were illustrated on a set of 20 TerraSAR-X images of Paris.

# 7 Literature


[1] H. Sportouche, F. Tupin, and L. Denise, "Extraction and 3D Reconstruction of Isolated Buildings in Urban Scenes from High-Resolution Optical and SAR Spaceborne Images," *IEEE TGRS*, Oct. 2011.

[2] L. Denis, F. Tupin, J. Darbon, and M. Sigelle, "Joint Regularization of Phase and Amplitude of InSAR Data: Application to 3D reconstruction," *IEEE TGRS*, vol. 47, pp. 3774–3785, Nov. 2009.

[3] F. Tupin and M. Roux, "Markov Random Field on Region Adjacency Graphs for the fusion of SAR and optical data in radargrammetric applications," *IEEE TGRS*, vol. 43, pp. 1920–1928, Aug. 2005.

[4] Y. Sun, Y. Hua, L. Mou, and X. X. Zhu, "Large-scale building height estimation from single VHR SAR image using fully convolutional network and GIS building footprints," in *Joint Urban Remote Sensing Event, JURSE, Vannes*, 2019.

[5] C. Rambour, L. Denis, F. Tupin, H. Oriot, Y. Huang, and L. Ferro-Famil, "Urban surface reconstruction in sar tomography by graph-cuts," *Computer Vision and Image Understanding*, vol. 188, p. 102791, 2019.

[6] G. Fornaro, F. Serafino, and F. Soldovieri, "Three-dimensional focusing with multipass SAR data," *IEEE TGRS*, vol. 41, pp. 507–517, March 2003.

[7] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, pp. 276–280, Mar 1986.

[8] X. X. Zhu and R. Bamler, "Tomographic SAR Inversion by $L_1$ -Norm Regularization ; The Compressive Sensing Approach," *IEEE TGRS*, vol. 48, pp. 3839–3846, Oct 2010.

[9] A. Budillon, A. Evangelista, and G. Schirinzi, "Three-Dimensional SAR Focusing From Multipass Signals Using Compressive Sampling," *IEEE TGRS*, vol. 49, pp. 488–499, Jan 2011.

[10] A. Ley, O. D'Hondt, and O. Hellwich, "Regularization and completion of tomosar point clouds in a projected height map domain," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, pp. 2104–2114, June 2018.

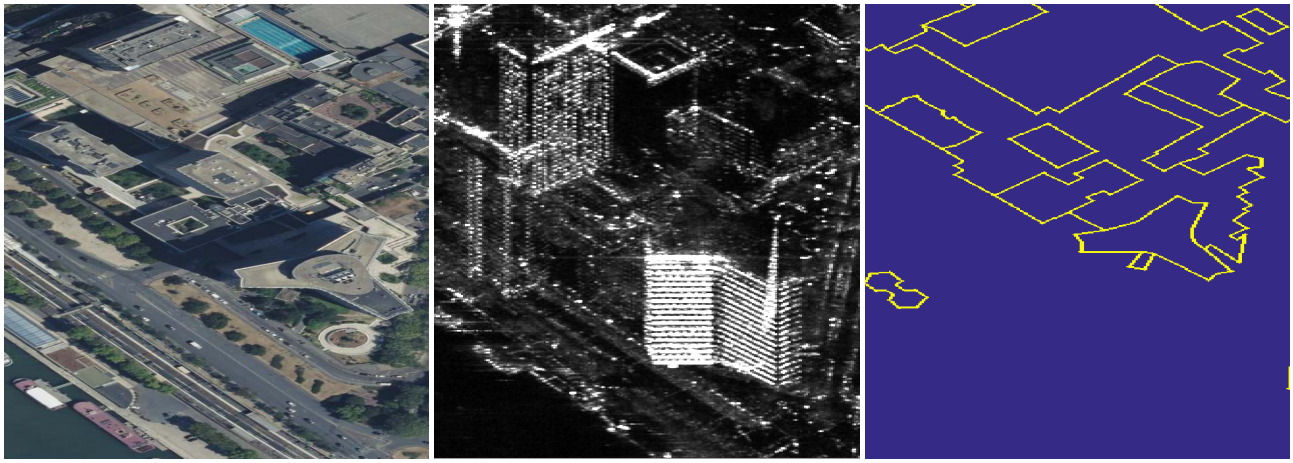

**Figure 3** From left to right: the optical image of the scene, the average of the SAR images stack and the available buildings footprint.

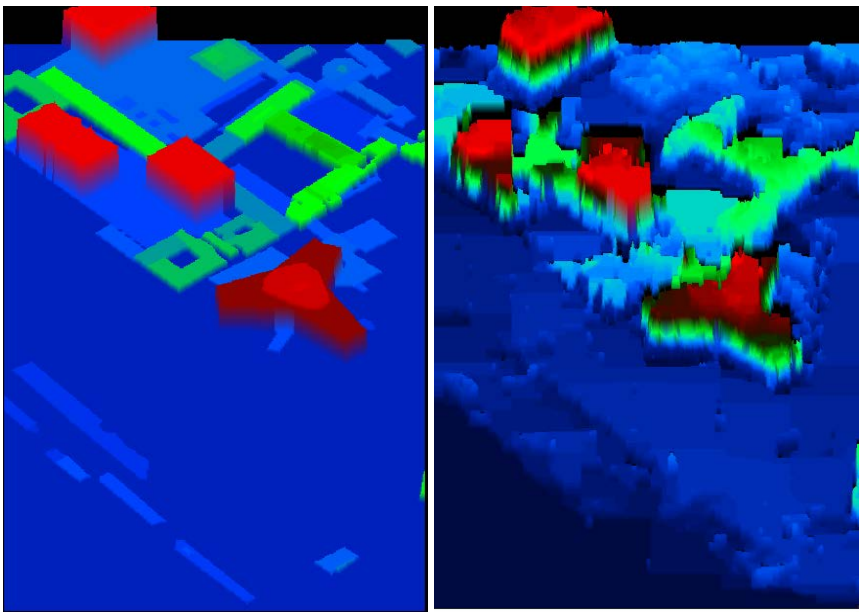

**Figure 4** From left to right: the urban surface ground truth, the surface obtained with a spatially-variant regularization parameter $\beta$ based on the footprints knowledge.