

# Earth Observation Using SAR and Social Media Images

Yuanyuan Wang

Signal Processing in Earth Observation  
Technical University of Munich (TU Munich)  
wang@bv.tum.de

Xiao Xiang Zhu

Signal Processing in Earth Observation, TU Munich  
and German Aerospace Center  
<http://www.sipeo.bgu.tum.de>

## Abstract

*Earth Observation (EO) is mostly carried out through centralized optical and synthetic aperture radar (SAR) missions. Despite the controlled quality of their products, such observation is restricted by the characteristics of the sensor platform, e.g. the revisit time. Over the last decade, the rapid development of social media has accumulated vast amount of online images. Despite their uncontrolled quality, the sheer volume may contain useful information that can complement the EO missions, especially the SAR missions.*

*This paper presents a preliminary work of fusing social media and SAR images. They have distinct imaging geometries, which are nearly impossible to even coregister without a precise 3-D model. We describe a general approach to coregister them without using external 3-D model. We demonstrate that, one can obtain a new kind of 3-D city model that includes the optical texture for better scene understanding and the precise deformation retrieved from SAR interferometry.*

## 1. Introduction

Earth Observation (EO) is usually carried out through centralized missions, delivering products with controlled quality. Among them, optical and synthetic aperture radar (SAR) are the most popular ones among the EO missions, for example the SAR sensors TerraSAR-X, TanDEM-X, and the optical sensors Landsat, WorldView series. They provide geodetically highly accurate large scale information from space, which serve as an irreplaceable data resource for urban monitoring. Despite the controlled quality of the delivered images, such observation is restricted by the characteristics of the sensor platform, e.g. the revisit time of the satellites, and the spatial resolution of the sensors.

On the other hand, social media images, mostly terrestrial, have accumulated over the last decades to a nearly equal amount as the volume of EO missions' data. Estimates suggest hundreds of petabytes of them are available online in total. These images are complementary

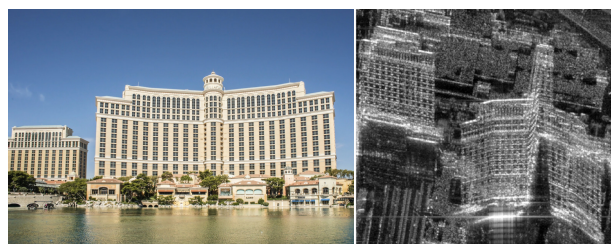


Figure 1. Left: a social media image of Bellagio hotel, Las Vegas, and right: the TerraSAR-X image of the same area. The white arrow in the SAR image indicates its range direction. Because of the unique range-azimuth imaging geometry, SAR image has a layover effect appeared in the images as “collapsed buildings”. And hence, the rich details in front of (towards the sensor) the building is buried in the strong reflection of the façade pixels.

to the remote sensing ones in the data characteristics. They are not restricted by regular temporal sampling. But their quality is uncontrolled.

The rapid increase of social media images enables a new type of remote sensing by fusing it with the EO data. This is especially beneficial for EO based on SAR images, as SAR image has a unique range-azimuth imaging geometry that is distinct from the usual perspective projection that human is familiar with. For example, Figure 1 shows a comparison of a social media image and a SAR image of Bellagio Hotel in Las Vegas. As one can see that the SAR image “suffers” from the *layover* effect which appears in the image as “collapsed buildings”. As a result, the rich details in front of the building are buried underneath the strong reflections of the façade pixels. Whereas, the social media image shows clear structure and texture of the façade. However, SAR images acquired in a repeat pass manner can provide valuable information of the subtle deformation of buildings to the millimeter level through SAR interferometric (InSAR) techniques. Therefore, fusing these two types of images can greatly improve the interpretability of the SAR images. It can lead to a new type of 3-D city model that includes the optical texture for scene understanding and the precise deformation parameters retrieved from SAR images.

This paper is a pioneer study of urban monitoring using SAR and social media images. The joint use of them

requires a coregistration between the two type of images which is nearly impossible in urban area without a precise 3-D model. Based on the framework presented in [1], we tackle this problem by performing the coregistration in 3-D space using 3-D models estimated from the optical and SAR images respectively. This requires the 3-D reconstruction techniques from multipass InSAR and multi-view stereo matching, respectively.

## 2. Related work

Only a handful of literatures address the problem of remote sensing and social media images fusion, e.g. joint classification of aerial and social media images [2], and flood damage mapping using governmental and crowdsource data [3], [4].

The fusion of SAR and social media images is nonexistent. Even the general coregistration of SAR and optical image in dense urban areas still poses a great challenge [5], because SAR image has an inevitable oblique viewing geometry and a unique cylindrical projection. The most relevant research in this context is joint analysis of SAR and optical images in urban areas, such as SAR simulation of dense urban areas [6], [7], height estimation of buildings with flat roof using SAR and optical images [8], and analysis of persistent scatterer using oblique images [9], just to name a few. Please refer to [10] for a detailed review of data fusion in remote sensing. Only until recently, [1] described a preliminary framework of pixel-level coregistration of SAR and aerial images in dense urban areas. Different from [1], this paper exploits the potential of social media images which on one hand have a similar oblique view as the SAR image, but on the other hand have uncontrolled quality and worse geo-localization.

Nevertheless, extensive research in the respective field of deformation monitoring by multipass InSAR, and 3-D reconstruction from social media images have been conducted in the past.

### 2.1. Multipass SAR interferometry

Multipass InSAR is so far the only imaging-based method to assess long-term millimeter-level deformation over large areas. Through modeling the interferometric phase history of a scatterer's time series, it is able to reconstruct the 3-D position as well as the deformation parameters of the scatterer. Among various multipass InSAR techniques, persistent scatterer interferometry (PSI) [11]–[14], SqueeSAR [15]–[18], SAR tomography (TomoSAR) and its differential form (D-TomoSAR) [19]–[26] are the workhorses for the deformation monitoring over large areas.

Among them, TomoSAR and D-TomoSAR have their unique ability to separate layovered scatterers within a

pixel, which renders them the most competent InSAR methods for monitoring urban areas. In particular, several advanced D-TomoSAR methods have been developed, e.g. using compressive sensing to achieve superresolution in separating closely spaced scatterers [23], and fusing SAR imaging geodesy [27] and TomoSAR inversion to obtain absolute Geodetic TomoSAR point clouds [28]. Although able to deliver precise deformation estimates, the abovementioned multipass InSAR techniques usually require a fairly large SAR image stack (>20 images) to achieve a reliable estimation.

### 2.2. 3-D reconstruction from optical images

3-D reconstruction from optical images is a classical topic in photogrammetry that is usually carried out by stereo matching [29]. But many algorithms were developed over the last decade to exploit the available unstructured image data. These algorithms are referred to as structure from motion (SfM) [30]–[34]. Through matching the corresponding points among the images, it allows the recovery of both the camera parameters and the scene structure. Many studies have shown that processing hundreds of thousand social media images is now even possible [32], [33]. Despite it is suitable to work with large quantity of unordered image, the availability of the social media images is only guaranteed for certain geographical areas.

## 3. Methodology

The general framework of the proposed method is shown in Figure 2. At the core of the method is the independent 3-D reconstructions from the two datasets, respectively, which are used to coregister the images. Therefore, the algorithm requires a stack of tens of SAR images, and tens to hundreds of social media images, for the 3-D reconstruction.

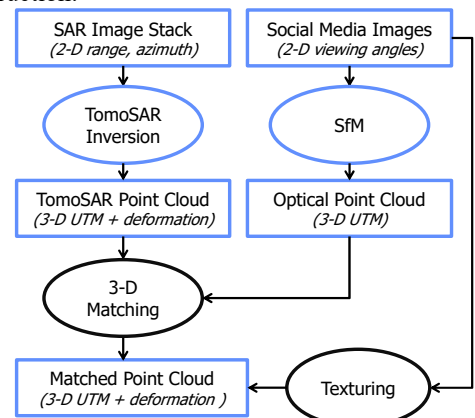


Figure 2. The flowchart of the SAR and social media image fusion algorithm. It matches the 3-D TomoSAR and optical point clouds, and then transfers the texture from optical images to the

TomoSAR point cloud. This paper focuses on the 3-D matching and texturing which are shown in black ellipses. The coordinate system of each dataset in the flowchart is indicated by the italic text in the bracket. The TomoSAR point clouds contains the attribute of deformation parameters in addition to the 3-D position.

A brief summary of the essential steps in the method is listed as follows.

- 3-D reconstructions: reconstructs 3-D point clouds from the SAR image stack and the social media images using D-TomoSAR and SfM, respectively.

D-TomoSAR retrieves the 3-D position, as well as the deformation parameters, e.g. linear deformation rate and the amplitude of periodic motion. Therefore, the result is a 3-D point cloud plus the attributes of deformation parameters.

SfM works on unordered optical images, which is well suited for social media images. But the reconstructed 3-D point cloud can have arbitrary scale and coordinate origin. Ground control points are required to geo-localize the point cloud.

- 3-D matching: coregister the TomoSAR point cloud and the optical point cloud.

Both point clouds are obtained from oblique looking images, which result in rich façade points. Therefore, the coregistration can be done using the conventional iterative closest point (ICP) algorithm.

However, one must take into account the different modalities of the two point clouds. The accuracy of TomoSAR point cloud is very anisotropic. It has an extremely good accuracy up to centimeter in range-azimuth [27], [35], but much poorer in elevation. A typical elevation accuracy for a stack of 50 TerraSAR-X images is about of 1~10m, which is about 100~1000 times worse than the other two dimensions. Transforming to Universal Transverse Mercator (UTM) coordinate, the error is mostly translated to the *up* and *east* direction. In addition, the accuracy is also not consistent over the whole point cloud, as the signal-to-noise ratio (SNR) of each pixel are very dynamic.

On the contrary, the accuracy of the optical point cloud is much less anisotropic, because of the various viewing direction of the images. The accuracy of the optical point cloud derived from social media images will depend on the number of images, as well as their quality.

### 3.1. Differential SAR tomography

D-TomoSAR is the most competent InSAR method for urban monitoring, since it has the ability to resolve the 3-D position and deformation parameters of multiple layovered scatterers in a pixel. Therefore, it outperforms other single scatterer-based approaches such as PSI in the monitoring of urban areas.

Figure 3 illustrates the layover phenomenon by the SAR imaging geometry at a fixed azimuth position. The satellite on the left refers to the SAR sensor hundreds of kilometers above the earth's surface. The gray diamonds near the satellite represent its repeated positions at a different times. Since SAR measures range (distance), the reflection at the same distance to the sensor, e.g. the reflection  $\gamma$  from the objects within the dashed lines, is integrated into a single pixel in the SAR image. On the contrary, an optical image having the same imaging geometry should be taken at a 90° angle from the SAR range direction, i.e. elevation.

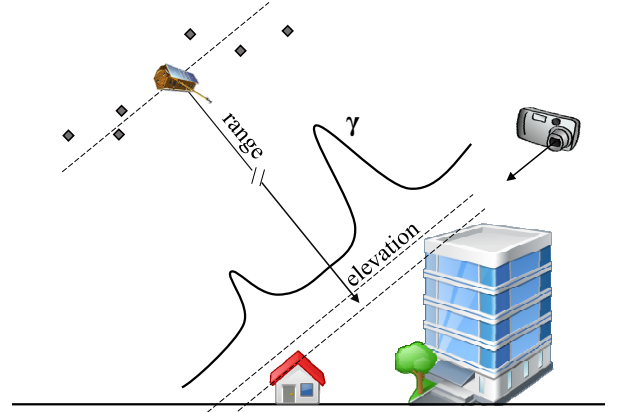


Figure 3. SAR imaging geometry at a fixed azimuth position. Reflection within one range resolution cell, e.g. the reflection  $\gamma$  coming from the objects within the two dashed lines, is integrated into a single pixel in the final SAR image. TomoSAR retrieves the whole reflectivity profile  $\gamma$  using multiple images acquired at slightly different antenna positions which is indicated by the diamonds near the satellite. On the contrary, an optical image having the same imaging geometry should be taken at a 90° angle from the SAR range direction, i.e. elevation.

This imaging model can be described as a linear model as follows

$$\mathbf{g} = \mathbf{F}\gamma \quad (1)$$

where  $\mathbf{g}$  is the observed pixel time series,  $\mathbf{F}$  is the Fourier sensing matrix depending on the baselines between the satellites. The number of observations in  $\mathbf{g}$  (typically 20~50) is usually much smaller than the dimension of  $\gamma$ . This is especially true when considering multiple deformation parameters. For example, when considering two deformation parameters in addition to the elevation, the dimension of  $\gamma$  in our processing is about 100,000, depending on the discretization level (e.g. 100×30×30) in each parameter space. This renders equation (1) an extremely underdetermined system. We solve it using Tikhonov regularization expressed in equation (2)

$$\hat{\gamma} = \arg \min_{\gamma} \frac{1}{2} \|\mathbf{g} - \mathbf{F}\gamma\|_2^2 + \lambda \|\gamma\|_2^2 \quad (2)$$

where  $\lambda$  is the regularization parameter.

### 3.2. Structure from motion

SfM is a well-studied topic. Therefore, it is not the focus of this paper. To briefly recap, SfM estimates the camera parameters of each image, i.e. the 3-D position, orientation, focal length, and sometimes the distortion parameters, given the matched corresponding points between multiple images. The SfM system model can be formulated as follows [33].

$$\tilde{\mathbf{x}}_{ij} = f_j \Pi(\mathbf{R}_j(\mathbf{x}_i - \mathbf{t}_j)) \quad (3)$$

where  $\tilde{\mathbf{x}}_{ij}$  is a set of corresponding image points, which is used to denote the  $i$ th points in image  $j$ ,  $\mathbf{x}_i$  is their real 3-D coordinates,  $f$ ,  $\mathbf{R}$ ,  $\mathbf{t}$  are the camera focal length in number of pixels, the camera orientation matrix and camera position vector, respectively, and  $\Pi$  is the projection function:  $\Pi(x, y, z) = (x/z, y/z)$ . The estimation is achieved through the minimization of the *reprojection* error which is the different between the predicted pixel position and the observed one:

$$\{\hat{\mathbf{x}}_i, \hat{\mathbf{R}}_j, \hat{\mathbf{t}}_j, \hat{f}_j\} = \arg \min_{\mathbf{x}_i, \mathbf{R}_j, \mathbf{t}_j, f_j} \sum_{i,j} \left\| \tilde{\mathbf{x}}_{ij} - f_j \Pi(\mathbf{R}_j(\mathbf{x}_i - \mathbf{t}_j)) \right\|_2^2 \quad (4)$$

Equation (4) is usually overdetermined, as many corresponding points can be found across images of a same object.

### 3.3. 3-D point clouds coregistration

Without any ground control points, the point cloud derived from social media images is always in an arbitrary scale and coordinate w.r.t. the real object. In this preliminary work, the ground control points were manually selected. The optical point cloud was scaled and oriented according to the ground control points.

The TomoSAR point cloud is also relative to a reference point (with unknown height) selected during processing. Such relative measurement is always performed for multipass InSAR technique to mitigate some common errors in the image, such as atmosphere delay. Hence, the TomoSAR point cloud is shifted from its true position.

Considering the anisotropic and nonergodic noise in the TomoSAR point cloud, an anisotropic ICP algorithm [36], [37] using robustly estimated covariance matrix [1] was employed for the coregistration.

Anisotropic ICP solves the following minimization at each iteration.

$$\{\hat{\mathbf{R}}, \hat{\mathbf{t}}\} = \arg \min_{\mathbf{R}, \mathbf{t}} \sum_i \left\| \mathbf{x}_i - \mathbf{R}\mathbf{p}_i - \mathbf{t} \right\|_{\mathbf{C}_{\varepsilon_i}}^2 \quad (5)$$

where  $\mathbf{R}$  and  $\mathbf{t}$  are the rotation matrix (included scaling) and the translation vector, and  $\mathbf{x}_i$ ,  $\mathbf{p}_i$  are the respective corresponding points in the reference (optical) and target

point clouds (TomoSAR).  $\mathbf{C}_{\varepsilon_i}$  is the covariance matrix of the residual  $\varepsilon_i = \mathbf{x}_i - \mathbf{R}\mathbf{p}_i - \mathbf{t}$ . ICP sometimes suffers the problem of finding local minima. A good initialization is usually required.

### 3.4. Texture mapping

Once the coregistered 3-D point cloud is obtained, one can texture the RGB color and the deformation parameters to the 3-D point cloud, or to the 2-D SAR or optical images.

Texturing basically solves the occlusion problem given a view point. It usually requires 3-D surface. As the accuracy of the point cloud may not be sufficient enough for a reliable surface reconstruction, we employed the hidden point removal algorithm [38], [39] which is a point-based rendering technique. However, we also demonstrate that simple surface reconstruction is possible for some popular buildings with simple geometry.

Since multiple social media images can be visible to a point/face, we also consider the appearance consistency between adjacent point/face. A method similar to [40] is employed. The images are weighted by their distance to each target point/face. A majority voting among the neighboring points is performed to select consistent images for texturing.

## 4. Experiments

In the experiment, we applied our algorithm to a few test sites in Las Vegas, and Berlin. Two stacks of TerraSAR-X high resolution spotlight images, with 25 and 102 images respectively, were used to reconstruction the TomoSAR point clouds of the two cities. 286 and 1,766 social media images (creative common) of Bellagio hotel in Las Vegas, and Berlin Cathedral were downloaded from social media websites (Flickr and Google), respectively.

### 4.1. Reconstruction from SAR images

The SAR interferogram formation and stacking were performed using the integrated wide area processor (*IWAP*) [41], [42] of German Aerospace Center (DLR). The D-TomoSAR reconstruction was performed using the DLR's large area tomographic inversion system: *Tomo-GENESIS* [43] which is based on [22]–[24], [44]–[47].

In the D-TomoSAR reconstruction, we considered both linear deformation rate and amplitude of seasonal motion as the deformation parameters. Figure 4 shows the reconstructed topographic heights of the two cities, and the amplitude of seasonal motion of Berlin. The height and the deformation are all color coded. The other motion parameters are not shown here due to space limit. The two point clouds contain around 17 million and 13 million points, respectively.



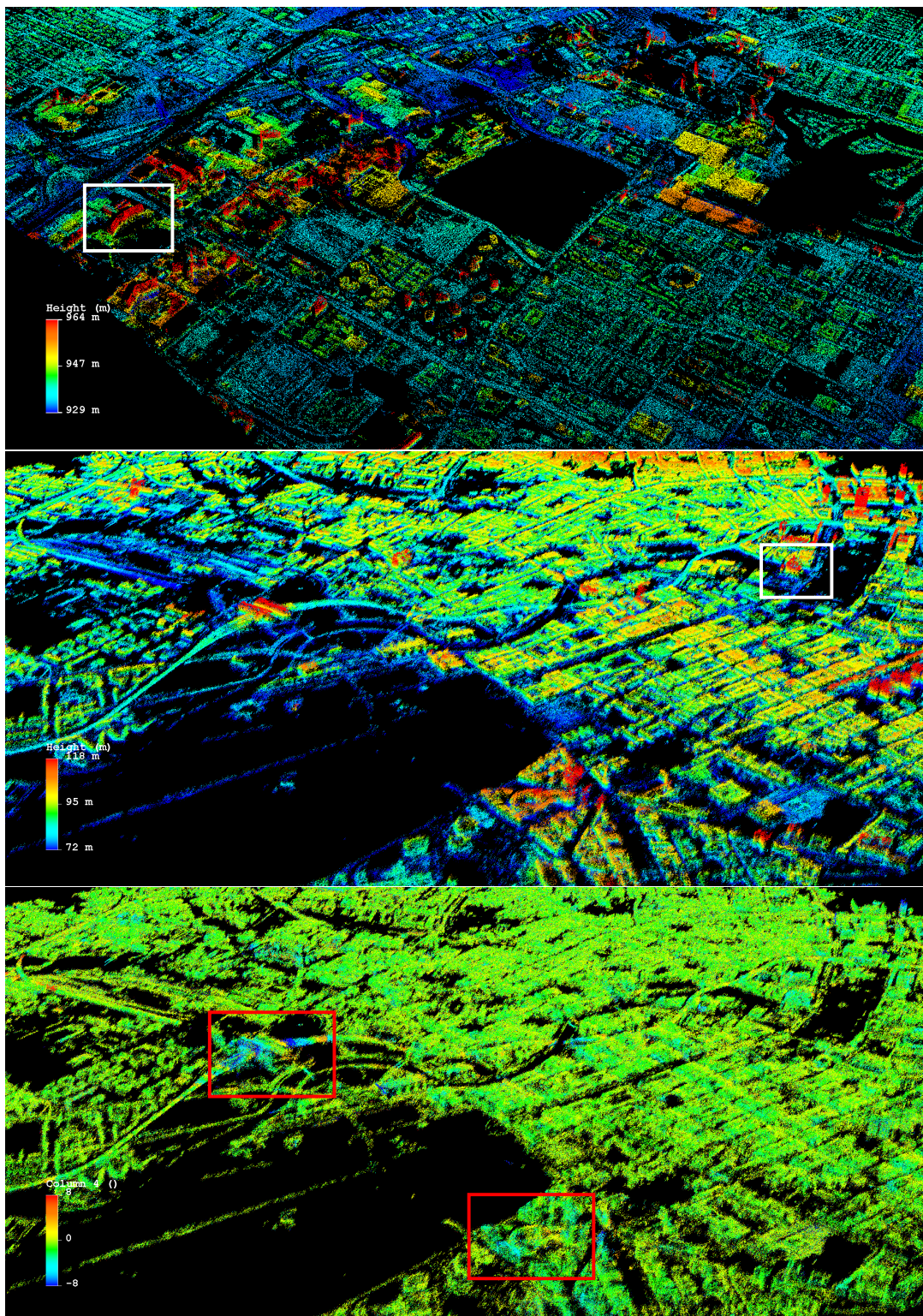


Figure 4. D-TomoSAR reconstruction of Las Vegas, and Berlin. Upper and middle: the topography of Las Vegas and Berlin in meters, respectively, and lower: the amplitude of seasonal motion of Berlin in millimeters. The white rectangles in the figures mark the Bellagio Hotel in Las Vegas, and the Berlin Cathedral which are the test sites for SfM from social media images. The red rectangles mark the Berlin central station and Berlin Potsdamer Platz where high seasonal motion is present.



D-TomoSAR is able to retrieve moderately accurate 3-D topography, but rather precise deformation up to a millimeter-level. As one can see, Berlin in general undergoes no deformation, except a few buildings, including the Berlin central station, and the high-rise buildings in Potsdamer Platz (marked in red rectangles) undergoing significant periodic motion. This is mostly the natural thermal dilation of the structure caused by seasonal temperature variation. The black area in the figures is the region where no coherent signal was acquired over the time span of the images. These areas are usually vegetation, and river whose surface is rapidly varying.

#### 4.2. Reconstruction from social media images

We selected Bellagio Hotel in Las Vegas, and the Berlin Cathedral as the test sites. 286 and 1,766 social media images (creative common) were download from Flickr and Google. The SfM and dense reconstruction were carried out by VisualSFM [34], [48]. Figure 5(a) shows the reconstructed 3-D models of the two buildings, where 230 thousand and 1.7 million points were reconstructed, respectively. Since Bellagio hotel has a relatively simple geometry, a Poisson surface was fitted. The Berlin Cathedral remains in point cloud form because of its rather complicated geometry. For reference, an optical image for each building was shown in Figure 5(b).

Compared to the TomoSAR point cloud, as shown in Figure 6, SfM reconstructs more accurate and dense 3-D point cloud for famous landmarks where social media images are abundant. A rough estimate of the SfM point cloud at flat surface shows an accuracy of about 50cm. This improvement can be used to enhance the TomoSAR point cloud.

#### 4.3. Coregistration and texturing

The optical point clouds are coregistered to their corresponding TomoSAR point clouds in UTM coordinate. As a result, the missing link between the 2-D spaces of the optical images and the SAR images is established, since each point in the optical or the TomoSAR point cloud has unique coordinate in their original image spaces. Hence, one can texture the TomoSAR point cloud or the 2-D SAR images with optical information, or texture the 3-D model with the deformation parameters from TomoSAR.

As an example, Figure 7 shows the 3-D model (identical to Figure 5(a) left) of the Bellagio hotel textured by the amplitude of seasonal deformation derived from TomoSAR. The deformation is color coded, with red meaning moving towards the sensor, and blue meaning moving away from the sensor. The textured model shows that the building undergoes significant seasonal motion. The deformation tends to be severer at higher floors. This is reasonable, as higher floors are less restricted to

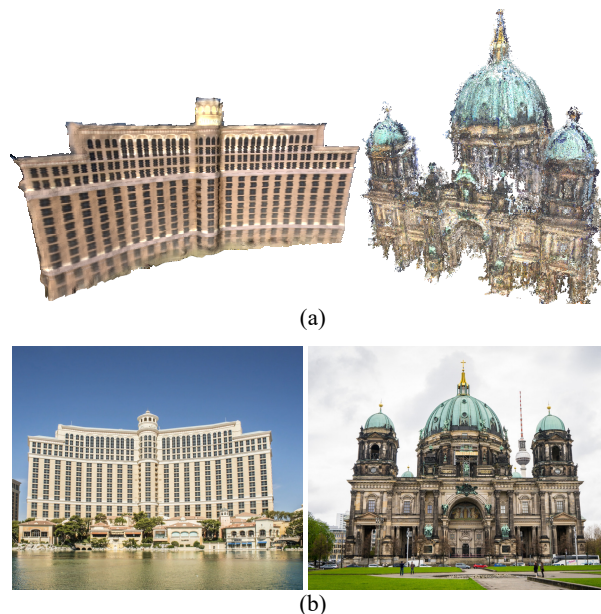


Figure 5. (a) SfM plus dense reconstruction of Bellagio Hotel and Berlin Cathedral, and (b) corresponding social media images as a visual reference.

displacement than the lower floors which are bounded to the foundation. But Figure 7 also shows that the right side of Bellagio hotel undergoes much larger seasonal motion, as usually one would expect similar thermal dilation at two symmetric parts of a building. This indicates either a strong temperature gradient across the building, or the internal structure is not equally rigid at both sides. Such information cannot be obtained from the optical appearance alone. This gives the user an intuitive way to understand the healthiness of the structure, even precise floors where undergo significant deformation.

A more innovative application of this research is to project the social media images to the SAR geometry for

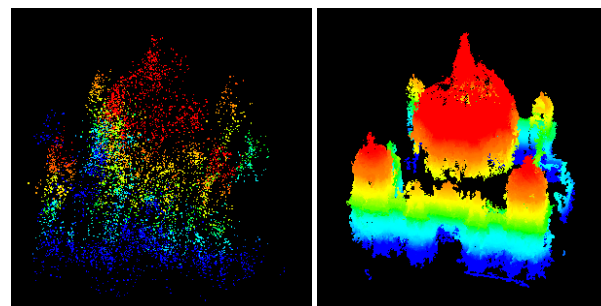


Figure 6. A comparison of the Berlin Cathedral 3-D reconstruction from: (left) TerraSAR-X images acquired 700km above the earth's surface, and (right) social media images taken at close distance. The color represents height. The social media produce denser and more accurate reconstruction because of the large amount of images. However, TomoSAR has the unique ability to measure millimeter-level deformation of urban infrastructures.

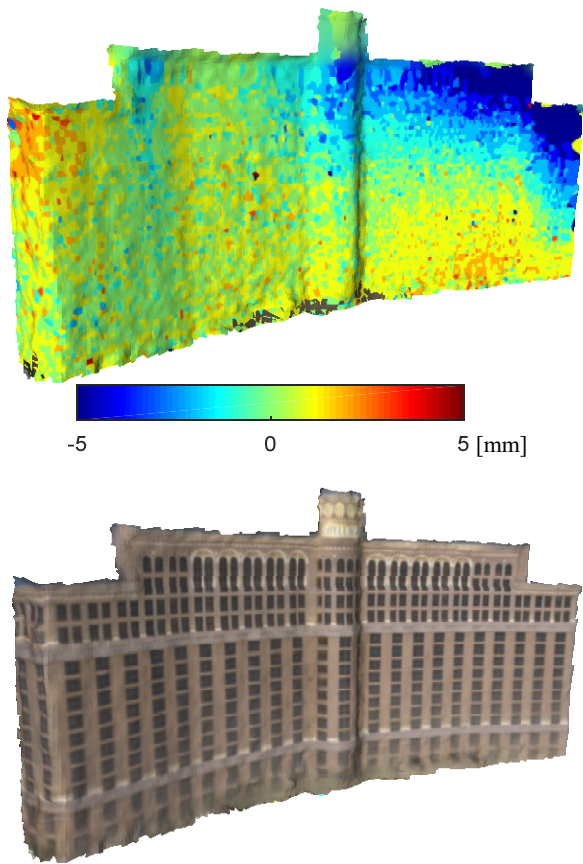


Figure 7. Upper: the 3-D model of Bellagio hotel, textured by the amplitude of seasonal deformation derived from TomoSAR, and lower: the 3-D model identical to that in Figure 5(a). The upper model shows that the building undergoes significant seasonal motion which generally increases with the height. This is reasonable, as the higher floors are more freely to displace than the lower ones. It also shows that the right side of the building undergoes much larger seasonal motion, whereas one usually would expect similar magnitude at two symmetric parts of a building. This indicates either a strong temperature gradient across the building, or the internal structure is not equally rigid at both sides. Such information cannot be obtained from the optical appearance alone.

better SAR image understanding. SAR image are difficult to interpret even for trained experts, as the formation of SAR image is a fairly complicated process that is difficult to interpret without a good knowledge of the 3-D geometry of the imaged object. Figure 8 shows the point cloud of Berlin Cathedral with colors from 1,766 social media images projected to the SAR range-azimuth geometry. The SAR amplitude image is also shown for comparison. With the help of optical information, the users are able to relate the bright scatterers in the SAR to objects seeing in visible spectrum.

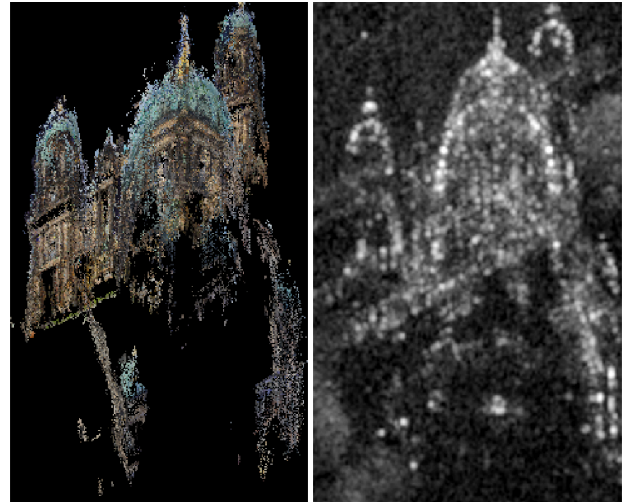


Figure 8. Left: the 3-D point cloud of Berlin Cathedral with colors from 1,766 social media images projected to the SAR geometry, and right: the SAR amplitude image as a reference. The SAR image is difficult to interpret even for trained experts. But social media images can help the understanding of SAR images.

As one can see that, the projected optical point cloud fits well to the original SAR image, thanks to the rather accurate SfM reconstruction. The comparison also shows that complicated structures do not always give a strong reflection in the SAR images, e.g. the left tower in Figure 8.

## 5. Discussion and Conclusion

This paper is the first demonstration of the fusion of multipass SAR interferometry and social media images. We show that by combining the InSAR unique ability of measuring millimeter-level deformation and the rich texture detail of large amount of close-range social media images, one is able to produce detailed 3-D model with precise deformation figure that can reveal unusual deformation behavior on specific parts of the building. Experiments show that the oblique view of terrestrial social media image is an advantage over the nadir-looking remote sensing images in the fusion with SAR images which have a similar oblique imaging geometry.

In the pipeline of the algorithm, D-TomoSAR is the most time consuming part. It also requires a fairly large stack of well-coregistered SAR images, typically more than twenty. Such data requirements may not be fulfilled for some cities. In the same vein, although SfM works on unordered optical images, which is well suited for social media images, the reconstruction quality depends on the availability of the images online. Therefore, the proposed is mostly suitable for famous cities in developed countries.

The point cloud coregistration usually gives rather accurate result, since only three-axis rotation, translation and a scaling need to be estimated from millions of points. However, the point cloud from social media images is in an

arbitrary scale and coordinate. In this work, ground control points were manually selected to correctly scale and orient the optical point cloud.

For a global scale application, future algorithm development needs to be focused on computationally efficient tomographic inversion, and very large scale SfM algorithm. Efficient algorithms on geo-localizing the SfM point clouds are also required.

## 6. Acknowledgement

The authors thank the NVIDIA Cooperation for providing the GPU which was used in this research. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V for providing computing time at the GCS Supercomputer SuperMUC at the Leibniz Supercomputing Centre. The authors also thank the reviewers for giving pertinent comments.

## References

- [1] Y. Wang, X. X. Zhu, B. Zeisl, and M. Pollefeys, "Fusing Meter-Resolution 4-D InSAR Point Clouds and Optical Images for Semantic Urban Infrastructure Monitoring," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 14–26, Jan. 2017.
- [2] S. Bhattacharya and A. M. Cheriyyadat, "Integrated Analysis of Ground Level and Aerial Image Data."
- [3] E. Schnebele and G. Cervone, "Improving remote sensing flood assessment using volunteered geographical data," *Nat. Hazards Earth Syst. Sci.*, vol. 13, no. 3, pp. 669–677, Mar. 2013.
- [4] E. Schnebele, G. Cervone, and N. Waters, "Road assessment after flood events using non-authoritative data," *Nat. Hazards Earth Syst. Sci.*, vol. 14, no. 4, pp. 1007–1015, Apr. 2014.
- [5] M. Schmitt and X. Zhu, "On the challenges in stereogrammetric fusion of SAR and optical imagery for urban areas," in *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2016.
- [6] S. Auer, S. Hinz, and R. Bamler, "Ray-Tracing Simulation Techniques for Understanding High-Resolution SAR Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1445–1456, Mar. 2010.
- [7] J. Tao, S. Auer, G. Palubinskas, P. Reinartz, and R. Bamler, "Automatic SAR Simulation Technique for Object Identification in Complex Urban Scenarios," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 7, no. 3, pp. 994–1003, Mar. 2014.
- [8] J. D. Wegner, J. R. Ziehn, and U. Soergel, "Combining High-Resolution Optical and InSAR Features for Height Estimation of Buildings With Flat Roofs," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5840–5854, Sep. 2014.
- [9] A. Schunert and U. Soergel, "Assignment of Persistent Scatterers to Buildings," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3116–3127, Jun. 2016.
- [10] M. Schmitt and X. X. Zhu, "Data Fusion and Remote Sensing: An ever-growing relationship," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 4, pp. 6–23, Dec. 2016.
- [11] A. Ferretti, C. Prati, and F. Rocca, "Permanent scatterers in SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 1, pp. 8–20, Jan. 2001.
- [12] B. M. Kampes, *Radar Interferometry - Persistent Scatterer Technique*, vol. 12. Dordrecht, The Netherlands: Springer, 2006.
- [13] N. Adam, B. Kampes, M. Eineder, J. Worawattanamateekul, and M. Kircher, "The development of a scientific permanent scatterer system," in *ISPRS Workshop High Resolution Mapping from Space, Hannover, Germany, 2003*, vol. 2003, p. 6.
- [14] S. Gernhardt and R. Bamler, "Deformation monitoring of single buildings using meter-resolution SAR data in PSI," *ISPRS J. Photogramm. Remote Sens.*, vol. 73, pp. 68–79, Sep. 2012.
- [15] A. Ferretti, A. Fumagalli, F. Novali, C. Prati, F. Rocca, and A. Rucci, "A New Algorithm for Processing Interferometric Data-Stacks: SqueeSAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 9, pp. 3460–3470, Sep. 2011.
- [16] Y. Wang, X. X. Zhu, and R. Bamler, "Retrieval of Phase History Parameters from Distributed Scatterers in Urban Areas Using Very High Resolution SAR Data," *ISPRS J. Photogramm. Remote Sens.*, vol. 73, pp. 89–99, Sep. 2012.
- [17] K. Goel and N. Adam, "An advanced algorithm for deformation estimation in non-urban areas," *ISPRS J. Photogramm. Remote Sens.*, vol. 73, pp. 100–110, Sep. 2012.
- [18] S. Samiei-Esfahany and R. F. Hanssen, "New algorithm for InSAR stack phase triangulation using integer least squares estimation," in *Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International*, 2013, pp. 884–887.
- [19] A. Reigber and A. Moreira, "First demonstration of airborne SAR tomography using multibaseline L-band data," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 5, pp. 2142–2152, Sep. 2000.
- [20] F. Lombardini, "Differential tomography: a new framework for SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 1, pp. 37–44, Jan. 2005.
- [21] 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 Sens.*, vol. 47, no. 1, pp. 224–237, Jan. 2009.
- [22] 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, 2010.
- [23] X. Zhu and R. Bamler, "Tomographic SAR Inversion by L1-Norm Regularization -- The Compressive Sensing Approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3839–3846, 2010.
- [24] Y. Wang, 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, 2014.



- [25] P. Ma and H. Lin, "Robust Detection of Single and Double Persistent Scatterers in Urban Built Environments," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2124–2139, Apr. 2016.
- [26] M. A. Siddique, U. Wegmüller, I. Hajnsek, and O. Frey, "Single-Look SAR Tomography as an Add-On to PSI for Improved Deformation Analysis in Urban Areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6119–6137, Oct. 2016.
- [27] M. Eineder, C. Minet, P. Steigenberger, X. Cong, and T. Fritz, "Imaging Geodesy - Toward Centimeter-Level Ranging Accuracy With TerraSAR-X," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 661–671, Feb. 2011.
- [28] X. X. Zhu, S. Montazeri, C. Gisinger, R. F. Hanssen, and R. Bamler, "Geodetic SAR Tomography," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 18–35, 2015.
- [29] H. Hirschmuller, "Stereo Processing by Semiglobal Matching and Mutual Information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.
- [30] M. Pollefeys *et al.*, "Visual Modeling with a Hand-Held Camera," *Int J Comput. Vis.*, vol. 59, no. 3, pp. 207–232, Sep. 2004.
- [31] M. Pollefeys *et al.*, "Detailed real-time urban 3d reconstruction from video," *Int. J. Comput. Vis.*, vol. 78, no. 2–3, pp. 143–167, 2008.
- [32] N. Snavely, S. M. Seitz, and R. Szeliski, "Modeling the World from Internet Photo Collections," *Int. J. Comput. Vis.*, vol. 80, no. 2, pp. 189–210, Nov. 2008.
- [33] S. Agarwal *et al.*, "Building Rome in a day," *Commun. ACM*, vol. 54, no. 10, p. 105, Oct. 2011.
- [34] C. Wu, "Towards linear-time incremental structure from motion," in *3DTV-Conference, 2013 International Conference on*, 2013, pp. 127–134.
- [35] X. Cong, U. Bals, M. Eineder, and T. Fritz, "Imaging Geodesy — Centimeter-Level Ranging Accuracy With TerraSAR-X: An Update," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 5, pp. 948–952, Sep. 2012.
- [36] R. S. J. Estépar, A. Brun, and C.-F. Westin, "Robust generalized total least squares iterative closest point registration," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2004*, Springer, 2004, pp. 234–241.
- [37] L. Maier-Hein, T. R. dos Santos, A. M. Franz, and H.-P. Meinzer, "Iterative Closest Point Algorithm in the Presence of Anisotropic Noise.," *Bildverarb. Für Med.*, vol. 2010, pp. 231–235, 2010.
- [38] S. Katz, A. Tal, and R. Basri, "Direct visibility of point sets," *ACM Trans. Graph. TOG*, vol. 26, no. 3, p. 24, 2007.
- [39] R. Mehra, P. Tripathi, A. Sheffer, and N. J. Mitra, "Visibility of noisy point cloud data," *Comput. Graph.*, vol. 34, no. 3, pp. 219–230, 2010.
- [40] C. Frueh, R. Sammon, and A. Zakhor, "Automated texture mapping of 3D city models with oblique aerial imagery," in *2nd International Symposium on 3D Data Processing, Visualization and Transmission, 2004. 3DPVT 2004. Proceedings*, 2004, pp. 396–403.
- [41] N. Adam, F. Rodriguez Gonzalez, A. Parizzi, and W. Liebhart, "Wide area persistent scatterer interferometry," in *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, 2011, pp. 1481–1484.
- [42] F. Rodriguez Gonzalez, N. Adam, A. Parizzi, and R. Brcic, "The Integrated Wide Area Processor (IWAP): A Processor for Wide Area Persistent Scatterer Interferometry," presented at the ESA Living Planet Symposium, Edinburgh, UK, 2013.
- [43] X. Zhu, Y. Wang, S. Gernhardt, and R. Bamler, "Tomo-GENESIS: DLR's Tomographic SAR Processing System," in *Urban Remote Sensing Event (JURSE), 2013 Joint*, 2013, pp. 159–162.
- [44] X. X. Zhu and R. Bamler, "Let's Do the Time Warp: Multicomponent Nonlinear Motion Estimation in Differential SAR Tomography," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 735–739, 2011.
- [45] X. Zhu, *Very High Resolution Tomographic SAR Inversion for Urban Infrastructure Monitoring: A Sparse and Nonlinear Tour*, vol. 666. Deutsche Geodätische Kommission, 2011.
- [46] 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, 2012.
- [47] S. Gernhardt, X. Cong, M. Eineder, S. Hinz, and R. Bamler, "Geometrical Fusion of Multitrack PS Point Clouds," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 1, pp. 38–42, 2012.
- [48] Y. Furukawa and J. Ponce, "Accurate, Dense, and Robust Multi-View Stereopsis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1362–1376, 2010.