Ingenieurfakultät Bau Geo Umwelt
Signalverarbeitung in der Erdbeobachtung

Advances in Meter-resolution Multipass Synthetic Aperture Radar Interferometry

Yuanyuan Wang, M.Sc.

Vollständiger Abdruck
der von der Ingenieurfakultät Bau Geo Umwelt
der Technischen Universität München
zur Erlangung des akademischen Grades eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr.-Ing. Florian Seitz

Prüfer der Dissertation:

1. Univ.-Prof. Dr.-Ing. habil. Xiaoxiang Zhu
2. Univ.-Prof. Dr.-Ing. habil. Richard Bamler
3. Univ.-Prof. Dr.-Ing. Uwe Sörgel
   Technische Universität Darmstadt

Die Dissertation wurde am 19.10.2015 bei der Technischen Universität München eingereicht und durch die Ingenieurfakultät Bau Geo Umwelt am 04.11.2015 angenommen.

Die Dissertation ist erstveröfentlicht als Forschungsbericht des Deutschen Zentrums für Luft- und Raumfahrt unter: Wang, Yuanyuan: Advances in Meter-resolution Multipass Synthetic Aperture Radar Interferometry, DLR-Forschungsbericht 2016, Köln, ISRN DLR-FB–2016-45 (zugleich Dissertation Technische Universität München 2015) und parallel online freigeschaltet unter <http://elib.dlr.de/105794/>.

Diese Dissertation ist auf dem Server der Deutschen Geodätischen Kommission unter <http://dgk.badw.de/> sowie auf dem Server der Technischen Universität München unter <http://mediatum.ub.tum.de/> elektronisch publiziert.

Abstract

Synthetic aperture radar (SAR) interferometry is the technique of using SAR as interferometer, to measure the phase difference caused by topography or object displacement. Multipass SAR interferometry (InSAR) is so far the only method for assessing long-term millimetre-level deformation over large areas from space on an imaging basis. Very high resolution multipass InSAR techniques have made substantial development in the last decade in monitoring individual building using persistent scatterer interferometry (PSI) and SAR tomography (TomoSAR), in monitoring nonurban area using small baseline subset (SBAS) and SqueeSAR, and so on.

To prepare for future SAR missions which will have higher resolution, larger coverage, and greater data volume, this thesis addresses the following four aspects of multipass InSAR techniques: computational efficiency, information fusion, contextual awareness, and statistical robustness.

TomoSAR is the most competent InSAR method for urban area monitoring. But it is much more computationally intensive than any other multipass InSAR methods. To this end, an computationally efficient multi-dimensional TomoSAR algorithm for urban area was developed, by ***integrating tomographic SAR inversion and the well-known PSI***. The results of PSI were used for a pre-classification of single and double scatterers, and also used as prior in the tomographic inversion. Real data experiments show the proposed approach obtains results comparable to the one obtained by the most computationally expensive tomographic SAR algorithms (e.g. SL1MMER), and saves computational time by a factor of fifty.

Multi-aspect InSAR point clouds fusion is required for a complete monitoring of entire city, due to the SAR side-looking geometry. In this thesis, a robust algorithm, namely "L-shape detection & matching", is proposed, especially for fusing two point clouds from cross-heading orbits, i.e. ascending and descending. The main idea of this algorithm is finding and matching the theoretically exact point correspondence which is the end positions of façades where the two point clouds close. Practical experiment shows the proposed method achieves sub-meter consistency with the state-of-the-art, and is much more computationally efficient.

The development of ***a semantic urban infrastructure monitoring algorithm by fusing InSAR and optical images*** was followed after the point cloud fusion. The attributes derived from optical images, e.g. colour, classification label, are transferred to the InSAR point cloud for a semantic level analysis of the deformation signal. The key lies on a strict 3-D geometric co-registration of SAR and optical images by reconstructing and matching the 3-D point clouds derived from the two types of images. Examples on bridges and railway monitoring are demonstrated.

Robust InSAR optimization is crucial for future SAR data, because of the unprecedented high resolution brings different observation statistics and much more dynamic interferometric phase. The proposed robust InSAR optimization (RIO) framework answers two open questions in multipass InSAR: (1) How to optimally treat images with a large phase error, e.g., due to unmodeled motion phase, uncompensated atmospheric phase, etc.? And (2) How to estimate the covariance matrix of a non-Gaussian complex InSAR multivariate, particularly those with nonstationary phase signals? For the former question, RIO employs a robust M-estimator to effectively down-weight these images, and for the latter question, a new method — the rank M-Estimator — is proposed. Simulated and real data experiments demonstrated that substantial improvement can be achieved in terms of the variance of estimates, comparing to the state-of-the-art estimators for both persistent and distributed scatterers. The proposed framework can be easily extended to other multipass InSAR techniques.

The abovementioned algorithms were tested using TerraSAR-X data of various test sites, especially for the efficient TomoSAR algorithm and the RIO framework.

Keywords

Synthetic aperture radar interferometry, robust estimation, covariance matrix, distributed scatterer, persistent scatterer, SAR tomography, deformation monitoring, semantic interpretation, InSAR point cloud, data fusion

Zusammenfassung

Synthetik Apertur Radar-Interferometrie ist eine Technologie um die Phase die durch die Topographie oder die Verschiebung von Objekten entsteht zu messen. Multipass-SAR-Interferometrie ist bis heute das einzige bildgebende Verfahren um über weite Gebiete langfristig millimetergenau Deformation mit Satelliten zu messen. Multipass SAR-Techniken für sehr hohe Auflösungen haben in der letzten Dekade eine signifikante Entwicklung erfahren, sei es bei der Beobachtung von einzelnen Gebäuden mit Persistenter Streuer Interferometrie (PSI) und SAR-Tomographie (TomoSAR), bei nichturbanen Gebieten mittels Small Baseline Subset (SBAS) und SqueeSAR, und so weiter.

Um für zukünftige SAR Missionen bereit zu sein, welche eine höhere Auflösung, größere Abdeckung und ein größeres Datenvolumen besitzen, behandelt diese Dissertation die folgenden vier Aspekte von Multipass-InSAR-Techniken: Recheneffizienz, Fusion von Information, Berücksichtigung von Zusammenhängen und robuste Schätzverfahren.

TomoSAR ist die leistungsfähigste InSAR-Methode für die Beobachtung von urbanen Gebieten. Allerdings ist es weitaus rechenintensiver als alle anderen Multipass-InSAR-Methoden. Daher wurde ein effizienter Ansatz für mehrdimensionale tomographische SAR-Inversion für urbane Gebiete entwickelt, indem ***die tomographischer SAR-Inversion in die bekannte PSI integriert*** wurde. Die Ergebnisse der PSI wurden für eine Vorklassifizierung in Einzel- und Doppelstreuer benutzt und fungierten als *a priori*-Wissen für die TomoSAR-Rekonstruktion. Versuchsläufe mit Realdaten zeigen, dass der vorgeschlagene Ansatz Ergebnisse liefert die vergleichbar sind mit denen der rechenintensivsten tomographischen SAR-Algorithmen (z.B. SL1MMER) und die Rechenzeit um den Faktor fünfzig verkürzt.

Die Fusion von Multi-aspect InSAR-Punktwolken ist aufgrund der seitlich schauenden Geometrie notwendig um eine ganze Stadt vollständig zu beobachten. In dieser Dissertation wird ein robuster Algorithmus — L-Form Detektion und Zuordnung — zur Fusion von Punktwolken, vor allem von cross-heading Orbits, das heißt ascending und descending, vorgeschlagen. Die Grundidee dieses Algorithmus ist das Finden und die Zuordnung der theoretisch exakten Punktkorrespondenzen, welche die Endpunkte von Fassaden sind an denen zwei Punktwolken zusammenlaufen. Praktische Experimente zeigen, dass die vorgeschlagene Methode Übereinstimmungen von unter einem Meter mit Methoden auf dem Stand der Technik erreicht, bei gleichzeitig weitaus effizienterer Nutzung der Rechenzeit.

Die Entwicklung eines Algorithmus für die ***semantische Beobachtung von urbaner Infrastruktur mittels Fusion von InSAR und optischen Bildern*** folgte nach der Punktwolken-Fusion. Anschließend werden die vom optischen Bild erhaltenen Attribute — Farbe, Klassifizierung, usw. — in die InSAR Punktwolke für eine Semantikanalyse des Deformationsignals übertragen. Entscheidend ist die strenge 3-D geometrische Koregistrierung von SAR-Bildern und optischen Bildern durch die Rekonstruktion und die Zuordnung der 3-D Punktwolken welche von den zwei Typen von Bildern erzeugt wurden. Der vorgeschlagene Ansatz wird anhand von Beispieldaten mit Brücken und Eisenbahnschienen gezeigt.

Robuste InSAR Optimierung ist entscheidend für zukünftige SAR-Daten, da die bisher noch nie dagewesene hohe Auflösung neue Beobachtungsstastistiken und eine weitaus dynamischere Phase mit sich bringt. Das vorgeschlagene robuste InSAR-Optimierungsframework (RIO) beantwortet zwei offene Fragen für Multipass-InSAR: (1) Wie sollen Bilder mit großen Phasenfehlern optimal behandelt werden, z.B. aufgrund nicht im Modell berücksichtigter durch Bewegung verursachte Phasenänderung, nicht kompensierter atmosphärischer Phase, usw.? (2) Wie kann die Kovarianzmatrix für nicht-Gaussche, komplexe, mehrdi-

mensionale InSAR Zufallsvariable geschätzt werden, insbesondere die mit nicht stationären Phasensignalen? Um die erste Frage zu adressieren setzt RIO einen robusten M-Schätzer ein, welcher Bildern mit großen Phasenfehlern ein geringeres Gewicht gibt, für die zweite Frage wird eine neue Methode — der Rang M-Schätzer — vorgeschlagen, der kein phase flattening benötigt. Versuche zeigen, dass erhebliche Verbesserungen bezüglich der Varianz der Schätzung erreicht werden können, im Vergleich zu Schätzern auf dem Stand der Technik für persistente und verteilte Streuer. Das vorgeschlagene Framework kann leicht für andere InSAR Techniken erweitert werden.

Die oben erwähnten Algorithmen wurden mit TerraSAR-X für mehrere Testgebiete getestet, insbesondere gilt dies für den effizienten TomoSAR-Algorithmus und das RIO Framework.

List of Abbreviations

Abbreviation	Description
2-D	two-dimensional
3-D	three-dimensional
AD	Anderson-Darling
ADI	amplitude dispersion index
AIC	Akaike information criterion
APS	atmospheric phase screen
BIC	Bayesian information criterion
BoW	bag of words
CCG	complex circular Gaussian
CDF	cumulative distribution function
CRLB	Cramér-Rao lower bound
CS	compressive sensing
D-InSAR	differential SAR interferometry
DoF	degree of freedom
DS	distributed scatterer
D-TomoSAR	differential SAR tomography
DEM	digital elevation model
GPSR	gradient projection for sparse reconstruction
HPR	hidden point removal
ICP	iterative closest point
i.i.d.	independent and identical distribution
InSAR	SAR interferometry
KS	Kolmogrov-Smmirnov
LAMBDA	least square ambiguity decorrelation

Abbreviation	Description
LOS	line-of-sight
MAP	maximum a posteriori
MLE	maximum likelihood estimator
MUSIC	multiple signal classification
NLS	nonlinear least squares
PDF	probability density function
PCA	principle component analysis
PS	persistent scatterer
PSI	persistent scatterer interferometry
RANSAC	random sample consensus
RIO	robust InSAR optimization
RME	rank M-estimator
RSS	residual sum of squares
SAR	synthetic aperture radar
SBAS	small baseline subset
SCM	sign covariance matrix
SfM	structure from motion
SL1MMER	scale-down by L_1 norm minimization, model selection, and estimation reconstruction
SLC	single look complex
SNR	signal to noise ratio
SVD	singular value decomposition
SVM	support vector machine
TerraSAR-X	TS-X
TomoSAR	SAR tomography
UTM	universal transverse Mercator

List of Symbols

Symbol	Description
g	single SAR pixel value
I	intensity
A	amplitude
ϕ	unambiguous (absolute) phase
φ	wrapped phase
R	range distance
R_s	nominal range distance
λ	wavelength
θ	incidence angle
$B_\perp$	perpendicular baseline
h	height
v	linear deformation rate
s	elevation
Δs	elevation extent
γ	coherence
$\gamma(s)$	reflectivity profile along elevation direction
$\boldsymbol{\gamma}$	discrete reflectivity profile
L	number of elements in $\boldsymbol{\gamma}$
L_s	number of elements in $\boldsymbol{\gamma}$ required for sparse reconstruction
K	number of scatterers
$K_0(\cdot)$	modified Bessel function of the second kind
Li_2	dilogarithm
N	number of images
M	number of samples
$\mathbf{g}$	multipass / multibaseline SAR pixel values
g_n	nth element of $\mathbf{g}$
$\mathbf{g}_m$	mth neighbouring sample of $\mathbf{g}$
$\hat{\mathbf{r}}$	estimated complex rank of $\mathbf{g}$

Symbol	Description
θ	phase history parameters
$\mathbf{a}$	integer 2π phase ambiguities
$\mathbf{C}$	complex covariance matrix
$\hat{\mathbf{C}}_k$	covariance estimate in the kth iteration
$\mathbf{C}_{\gamma\gamma}$	prior covariance matrix of $\boldsymbol{\gamma}$
$\mathbf{R}$	SAR discrete imaging model matrix
$\boldsymbol{\Phi}(\theta)$	diagonal matrix of the modelled phase
$\mathbf{W}$	diagonal weighting matrix
σ	standard deviation
F	distribution function
x	generic variable
$\boldsymbol{\varepsilon}$	residual vector
ε_i	ith element of $\boldsymbol{\varepsilon}$
α	regularization parameter
ν	degree of freedom
η_{RE}	relative efficiency
H_0	null hypothesis (no scatterer)
$H_{1,2}$	alternative hypothesis (single or double scatterer)
P_0	prior probability of the null hypothesis
$P_{1,2}$	prior probability of the alternative hypothesis
$\rho(\cdot)$	loss function
$\psi(\cdot)$	influence function
$w(\cdot)$	weighting function
$p(\cdot\|\cdot)$	likelihood function
$O(\cdot)$	computational complexity
$\Gamma(\cdot)$	Gamma function
$\mathrm{R}^n[\cdot]$	nth order rank operator
$\mathrm{Re}[\cdot]$	real part
$\mathrm{Im}[\cdot]$	imaginary part
$E[\cdot]$	expectation
$\lvert\cdot\rvert$	(element-wise) absolute value
$\lVert\cdot\rVert_p$	vector L_p norm (default L_2)

Contents

1 Introduction

1.1 Motivations and Objectives

Synthetic aperture radar (SAR) is an important remote sensing means that emits electromagnetic wave to the imaged object and receives the backscattered echoes. SAR interferometry (InSAR) is the techniques of using SAR as interferometer for 3-D reconstruction, and especially for monitoring the deformation of subwavelength level. InSAR plays an irreplaceable role in earth observation, because it has been generally appreciated that spaceborne InSAR is so far the only method to assess millimeter-precision long-term ground deformation on a global scale.

The real beauty of spaceborne InSAR is the repeated long-term monitoring of small deformation over large areas. The corresponding techniques are known as multipass / multibaseline / multitemporal InSAR techniques. They make use of a stack of co-registered SAR images to reconstruct the deformation history, as well as the 3-D position of the scatterer.

The main focus of this thesis is on the algorithm development of multipass InSAR techniques for future SAR systems.

The last several years are the era of meter-resolution spaceborne SAR sensors, such as TerraSAR-X (TS-X), TanDEM-X and COSMO-SkyMed. That opened a whole new perspective of analysing spaceborne SAR data. Substantial development in multipass InSAR techniques has been made, e.g. in monitoring individual buildings, modelling higher order motion, reconstruction of 3-D objects from SAR image stacks, and so on.

In the coming several years, SAR remote sensing will still be one of the most important means of spaceborne earth observation. In particular, more and more high resolution wide swath missions are under development. For example, the L-band SAR mission Tandem-L will provide one meter spatial resolution with a swath of 50km in its high resolution mode; the ALOS-2 provides both meter-resolution images with swath of 25km and 50km; and the TS-X Next Generation that will have decimetre-level spatial resolution. For the medium resolution mission, we can also foresee a data rush of the SAR satellite Sentinel-1, because of the free data access policy.

It is obvious that the future earth observation will, on one hand, have huge demand on large area processing, and on the other hand face data with unprecedentedly high resolution. The ever increasing resolution brought challenges, e.g. exponential growth of data volume, redundancy of data, different observation statistics, and much more dynamic motion. The challenges posed to my research under the context of multipass InSAR algorithms can be summarized as 1) How to efficiently perform tomographic SAR inversion for large data, because it is the most competent means of 3-D or even 4-D SAR imaging, but also the most computationally expensive one. 2) How to fuse meter-resolution SAR tomographic point clouds of urban areas from multiple viewing angles? One must consider the much higher point density compared to other InSAR methods, e.g. from persistent scatterer interferometry (PSI). 3) how to make use of the complementary information in high resolution optical image to help the interpretation of InSAR point clouds of urban areas, since an unstructured InSAR point cloud only gives 3-D position, deformation parameters, and the reflectivity of the scatterers, and lastly 4) how to handle unexpected statistical properties for the SAR images of unprecedentedly high resolution, e.g. the covariance matrix estimation of non-Gaussian scatterers.

Therefore, four important objectives of the thesis have been identified:

- ◇ Computational efficiency of SAR tomographic inversion: develop computationally efficient tomographic SAR inversion algorithms for very high resolution urban monitoring. The algorithm should be able to deliver tomographic reconstruction of an entire city using meter-resolution data in an affordable time.
- ◇ Information fusion: fuse TomoSAR point clouds from multiple viewing angles, in order to have a complete point cloud of the city under monitoring. The algorithm should be able to extract important feature points from the dense TomoSAR point cloud, in order to reduce computational cost and to be robust against large amount of outliers contained in the data, e.g. caused by the complex scattering scenario (Auer et al., 2011).
- ◇ Contextual awareness: fuse optical images and InSAR point cloud/SAR images for a better interpretation of SAR images and the deformation signal. The algorithm should particularly address the problem of 3-D geometric registration of optical and SAR images, because of the complex 3-D urban topography.
- ◇ Statistical robustness: develop statistically robust InSAR algorithms for non-urban area monitoring, taking into account different SAR statistics that may bias the covariance estimation and the much more dynamic deformation signal which can lead to model underfitting.

1.2 Thesis Outline

This is a cumulative dissertation comprising of five full paper peer-reviewed journal papers, which are presented in the appendix of the thesis. Chapter 2 covers the necessary basics for the understanding of the thesis. The contribution of the thesis and the state-of-the-art are summarized in Chapter 3. Chapter 4 and Chapter 5 cover the summary of the five journal papers and the conclusion, respectively. Therefore, advanced readers can directly start from Chapter 4.

The variables in this thesis were written with the following rule: *italic*: scalar; **bold regular**: vector, and **CAPITAL BOLD REGULAR**: matrix. The list of symbols can be found in List of Symbols.

2 Basics

Targeting at the readers who are already in the research field of InSAR, the necessary basics are included in this chapter. The discussions in this chapter are based on co-registered complex-valued SAR image stacks. For the basics of SAR imaging geometry, acquisition, focusing, etc., the readers are kindly directed to (Bamler and Hartl, 1998; Rosen et al., 2000) for an extensive review of these fundamentals.

2.1 Single-look SAR Signal

"Single-look" SAR signal refers to the reflected echoes integrated into one range-azimuth resolution cell, i.e. a range-azimuth SAR pixel. It is also known as single look complex (SLC). And in the rest of this thesis, they will be used interchangeably, unless explicitly stated. For the mathematical convenience, the classical SAR and InSAR theory has restricted the discussion to two extreme cases of scattering object: distributed scatterer and point scatterer.

2.1.1 Point Scatterer

Point scatterer refers to the situation when the signal reflected from a single object is absolutely dominating in the resolution cell. The complex radar return g is modelled as a deterministic signal which is (Bamler and Hartl, 1998)

$$g = A \exp\left(-j\frac{4\pi}{\lambda}R + j\varphi_{scatter}\right) \tag{2.1}$$

where A is the signal amplitude, $-4\pi R/\lambda$ and $\varphi_{scatter}$ are the phase contribution from the distance R to the sensor and an intrinsic scattering phase, respectively. If the reflectance from the point scatterer is coherent throughout the whole stack of SAR images, it is then known as persistent/permanent scatterer (PS). Therefore, the scattering phase $\varphi_{scatter}$ of a PS is usually assumed to be identical throughout the stack.

2.1.2 Distributed Scatterer

Distributed scatterer (DS) or Gaussian scatterer is the superposition of sufficiently many random sub-scatterers within a resolution cell. By assuming none of the sub-scatterer dominates another, the complex backscattering of the resolution cell degenerates into a zero-mean complex circular Gaussian (CCG) random variable according to so-called *central limit theorem*, i.e.:

$$g = x + jy \quad x, y \sim N\left(0, \sigma^2\right) \tag{2.2}$$

where the real and imaginary part x and y both follow a zero-mean Gaussian distribution with variance σ^2. This assumption is fulfilled for natural objects in medium resolution ($\sim$ 10 m) SAR system, however can be violated in urban areas or for high resolution systems.

As both the real and imaginary parts are identically Gaussian distributed, one can derive that the intensity of a DS pixel $I = |g|^2$ follows a χ^2 distribution with two degrees of freedom (DoF). The joint probability density function (PDF) of its intensity and phase is again a χ^2 distribution but only scaled with a factor of $1/2\pi$ (Bamler and Hartl, 1998):

$$f(I, \varphi) = \frac{1}{2\pi\bar{I}} \exp\left(-\frac{I}{\bar{I}}\right) \tag{2.3}$$

where $\bar{I} = \sigma^2$ is the expected intensity. As one can see, the PDF does not depend on the phase, because the scattering phase $\varphi_{scatter}$ of single-look DS has a uniform distribution. As a consequence, the phase seemingly carries no information.

2.1.3 Layover of Scatterers

Layover of scatterers happens when multiple scatterers appear at different elevation in the same resolution cell. The complex single look signal is then the integration of all the contributions along the third SAR coordinate elevation which is orthogonal to range and azimuth. Recall the plane wave approximation of the SAR imaging model of a single-look SAR pixel (after focusing) (Bamler and Hartl, 1998; Zhu, 2008):

$$g = \int_{\Delta s} \gamma(s) \exp\left(-j\frac{4\pi}{\lambda} R(s)\right) \mathrm{d}s \tag{2.4}$$

where $\gamma(s)$ is the complex reflectivity profile along the elevation direction s, Δs is the extend of the elevation direction, and $R(s)$ is the range change along elevation direction. As one can see, an SLC SAR image is only a single tomographic projection of the imaged object/area. It requires multiple acquisitions from slightly different positions in order to reconstruct the full reflectivity profile $\gamma(s)$. This is known as *SAR tomography* (TomoSAR).

In urban areas, we mostly consider the layover of PSs in a resolution cell. Therefore, $\gamma(s)$ can be parametrically modelled as multiple delta functions:

$$\gamma(s)_{PS} = \sum_i A_i \delta(s - s_i) \tag{2.5}$$

where A_i and s_i are the unknown complex amplitude and the elevation position of the ith scatterer, respectively.

2.2 Multipass InSAR

Briefly speaking, InSAR is the technique of using SAR as interferometer. Figure 1 shows the plane wave approximation of the interference of two SAR images. The red and blue lines are the iso-range lines with interval of half wavelength of the two images respectively. Because of the perpendicular baseline $B_\perp$, the iso-range lines intersect each other, creating an interference pattern. The background fringe shows the interferometric phase of the pattern. The topography is mapped into interferometric phase depending on the height of the point.

Since the perpendicular baseline is much smaller than the range of the satellite to the imaged object, every iso-range line defined by R = constant in the two SAR images should cover the same range resolution cell. This is a prerequisite of the coherent nature of SAR imaging, i.e. the two images should observe the same object in their corresponding pixels.

The interferometric phase changes as the baseline $B_\perp$ changes. The unambiguous interferometric phase ϕ is shown to be linearly proportional to the topography height h with the equation:

$$\phi = -\frac{4\pi}{\lambda} \frac{B_\perp}{R_S \sin\theta} h + \phi_{defo} \tag{2.6}$$

where λ is the wavelength, R_S is the nominal range which is the range of the SAR sensor to a zero-elevation point, and θ is the incidence angle. The deformation phase ϕ_{defo} must be taken into account if the two images are acquired at different time. The deformation causes

absolute phase change which is not dependent on the baseline. If the pixel is a DS, then (2.6) is the expected interferometric phase.

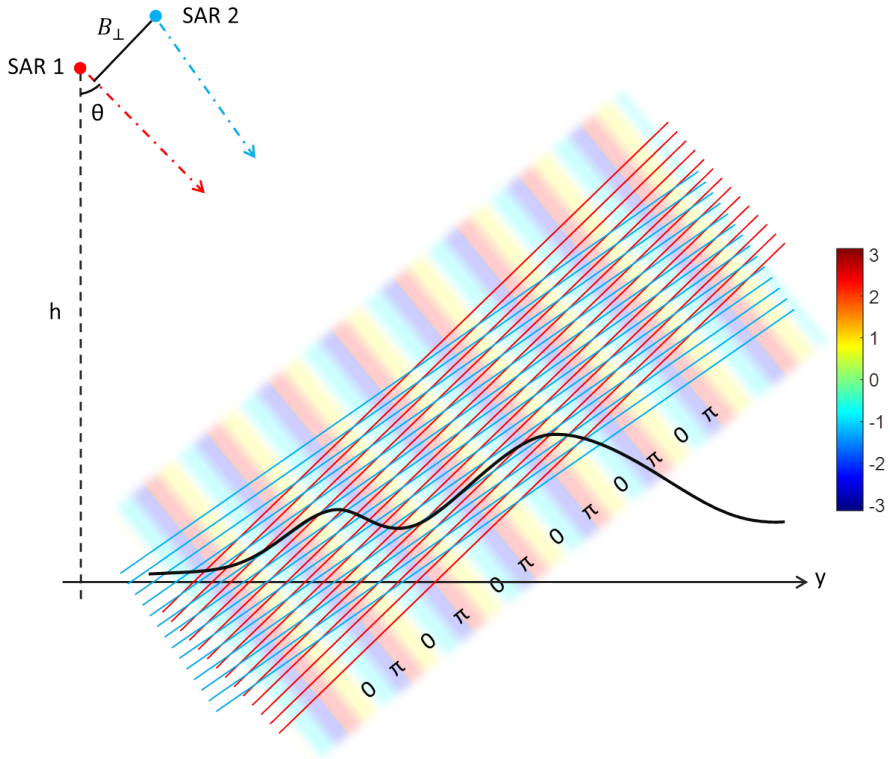


Fig. 2.1. Plane wave approximation of the interference of two SAR images. The red and blue lines are the iso-range lines with interval of half wavelength of the two images respectively. The iso-range lines intersect each other, creating an interference pattern. The background shows the interferometric phase of the pattern. The topography is mapped into interferometric phase depends on the height of the point.

Multipass InSAR is an extension of the single-baseline InSAR. It exploits the phase difference of more than two complex-valued SAR images. It is often employed for long-term deformation monitoring, which is also known as differential InSAR from which dynamic geophysical quantities such as deformations rate (volcanoes, urban underground water, oil reservoir, etc.) can be derived.

As long as only the phase is concerned, the unambiguous interferometric phases $\mathbf{\Phi}$ of the multiple acquisitions with their perpendicular baselines $\mathbf{B}_\perp$ are:

$$\mathbf{\Phi} = -\frac{4\pi}{\lambda}\frac{\mathbf{B}}{R_S \sin\theta}h + \mathbf{\Phi}_{defo}. \tag{2.7}$$

The deformation phase vector $\mathbf{\Phi}_{defo}$ is often modelled as a function $d(\mathbf{t}_n)$ (e.g. linear or periodic) of the acquisition time $\mathbf{t}_n$. The details of popular multipass InSAR techniques are covered in Chapter 3.

2.2.1 Single-look InSAR Statistics

It has been mentioned before that equations (2.6) and (2.7) only apply to PS. Under the assumption of distributed scatterers, they represent the expected interferometric phase(s). The phase variance caused by the decorrelation between the two co-registered SAR pixels g_1 and g_2 is fully characterized by their complex covariance. Normalizing the covariance with the expected intensities of the pixels, we obtain a more commonly used quantity - the

complex coherence γ:

$$\gamma = \frac{E\left[g_1 g_2^*\right]}{\sqrt{E\left[|g_1|^2\right] E\left[|g_2|^2\right]}}. \tag{2.8}$$

The PDF of the interferometric sample $v = g_1 g_2^*$, represented by the joint PDF of its amplitude and the interferometric phase can be shown as follows (Lee et al., 1994; Tough et al., 1995; Bamler and Hartl, 1998)

$$f_{|v|,\phi}(|v|,\phi) = \frac{2|v|}{\pi \bar{I}^2 \left(1-|\gamma|^2\right)} \exp\left(\frac{2|\gamma||v|\cos\left(\phi-\bar{\phi}\right)}{\bar{I}\left(1-|\gamma|^2\right)}\right) K_0\left(\frac{2|v|}{\bar{I}\left(1-|\gamma|^2\right)}\right) \tag{2.9}$$

where $\bar{I} = \sqrt{E\left[|g_1|^2\right] E\left[|g_2|^2\right]}$, $\bar{\phi}$ is the expected interferometric phase, and K_0 is the modified Bessel function of the second kind. Generalizing to multibaseline SAR pixels, the joint PDF of the complex vector $\mathbf{g}$ is a multivariate CCG random variable. The joint PDF can be written as the following compact equation:

$$f_{\mathbf{g}}(\mathbf{g}) = \frac{1}{\pi^N \det(\mathbf{C})} \exp\left(-\mathbf{g}^H \mathbf{C}^{-1} \mathbf{g}\right) \tag{2.10}$$

where N is the number of images, and $\mathbf{C}$ is the complex covariance matrix whose analytical expression is as follows.

$$\mathbf{C} = \begin{bmatrix} \bar{I}_1 & \gamma_{1,2}\sqrt{\bar{I}_1 \bar{I}_2} & \cdots & \gamma_{1,N}\sqrt{\bar{I}_1 \bar{I}_N} \\ \gamma_{1,2}^*\sqrt{\bar{I}_1 \bar{I}_2} & \bar{I}_2 & & \\ \vdots & & \ddots & \vdots \\ \gamma_{1,N}^*\sqrt{\bar{I}_1 \bar{I}_N} & & \cdots & \bar{I}_N \end{bmatrix} \tag{2.11}$$

As one can see in the CCG case, the multivariate $\mathbf{g}$ is fully characterized by the complex covariance matrix $\mathbf{C}$. Therefore the estimation of the covariance matrix becomes an important step in many multipass InSAR techniques. Under the assumption CCG distribution, the maximum likelihood estimator (MLE) of the covariance matrix can be easily derived given a set of M spatially stationary samples $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, ..., \mathbf{g}_M]$:

$$\hat{\mathbf{C}}_{MLE} = \frac{1}{M}\sum_{m=1}^{M} \mathbf{g}_m \mathbf{g}_m^H = \frac{1}{M}\mathbf{G}\mathbf{G}^H \tag{2.12}$$

2.3 The Relativeness of Multipass InSAR Estimates

The InSAR estimates, i.e. the topography and deformation estimates, are relative with respect to a reference point in the image that was selected during the processing. The unknown topography of the reference point shift the estimates of the whole image along the elevation direction of the native radar coordinate system. When the InSAR estimates are transformed to geographical coordinate systems, it will eventually cause the position to shift in all the three dimensions, in addition with a scaling effect due to the different incidence angle at different range.

The following figure explains the aforementioned systematic shift in the Universal Transverse Mercator (UTM) coordinate system. The two black dots indicate the true positions of two scatterers which are on the same height w.r.t. a reference surface. The two scatterers are located at the near range and far range of the SAR image. Their incidence angles θ_{near} and θ_{far} are different, e.g. as many as one degree for TS-X spotlight image. Of course, the difference is exaggerated in the plot for a better visualization. Due to the height Δh of the reference point, the scatterers are shifted to the red points along their elevation directions, respectively. As one can see that the scatterers are equally shifted in height, but scaled in the ground range direction due to their different elevation directions. The difference between the scaled ground range and the true value is

$$\begin{aligned} \Delta R &= \frac{\Delta h}{\tan\theta_{near}} - \frac{\Delta h}{\tan(\theta_{near}+\Delta\theta)} \\ &\approx \Delta h \Delta\theta \quad \text{if} \quad \theta_{near} \gg \Delta\theta \end{aligned} \tag{2.13}$$

where $\theta_{far} = \theta_{near} + \Delta\theta$. Such difference is approximately linearly proportional to the unknown height of the reference point and the diversity of the incidence angle. As the heading angle of the satellite is not parallel to north-south direction, this scaling will also transferred to both north-south and east-west directions.

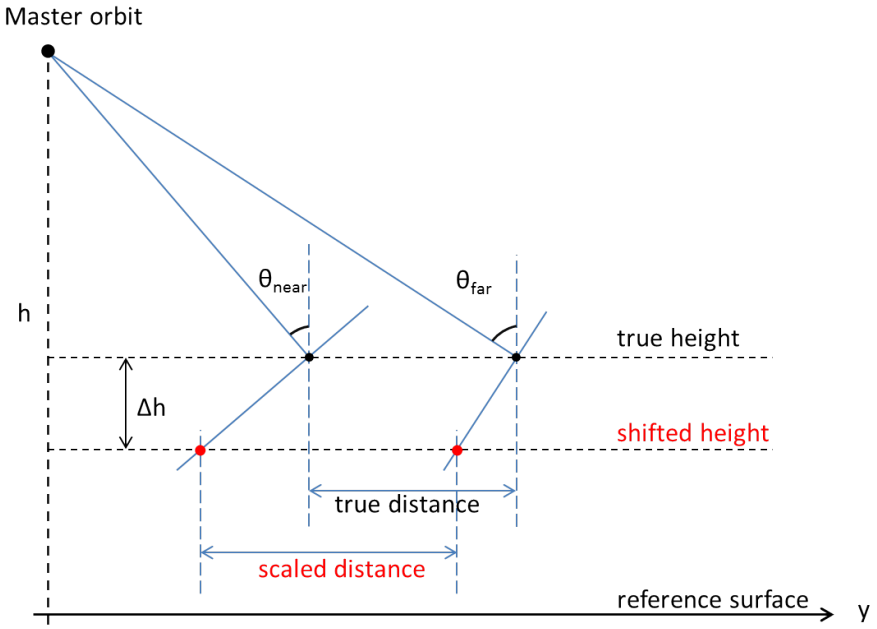


Fig. 2.2. Illustration of systematic shift of the multipass InSAR position estimates in UTM coordinate. The two black dots indicate the true positions of two scatterers which are on the same height w.r.t. a reference surface. The two scatterers are located at the near range and far range of the SAR image, respectively. Their incidence angles θ_{near} and θ_{far} are different, e.g. as many as one degree for TS-X spotlight images. Due to the unknown height of the reference point, the scatterers are shifted to the red points along their elevation directions, respectively. The scatterers are equally shifted in height, but scaled in the ground range direction due to their different elevation directions. This ground range scaling will translates into both north and east direction in the UTM coordinate. Therefore, any relative InSAR position estimate is shifted in three directions, with additional horizontal 2-D scaling.

Figure 2.3 also plots the ground range difference per meter of the height of the reference point as a function of the near range incidence angle and the angle diversity. As one can see, the approximation in (2.13) is only applicable for large incidence angles. For a typical image stack of 40° of incidence angle and 1° diversity, the scaling is 3.5 cm/m/°. However, the scaling will become 15 cm/m/° if the incidence angle becomes 20°, which will cause 1.5m of scaling if the unknown height of the reference point is 10 m.

The shift can be mitigated by choosing a reference point close to the reference digital elevation model (DEM), i.e. $h \approx 0$m. For medium resolution SAR data of non-urban area, such

shift is not easily noticeable, because of the relatively smooth topography and large pixel size. However, they must be taken into account for meter resolution SAR data of urban area.

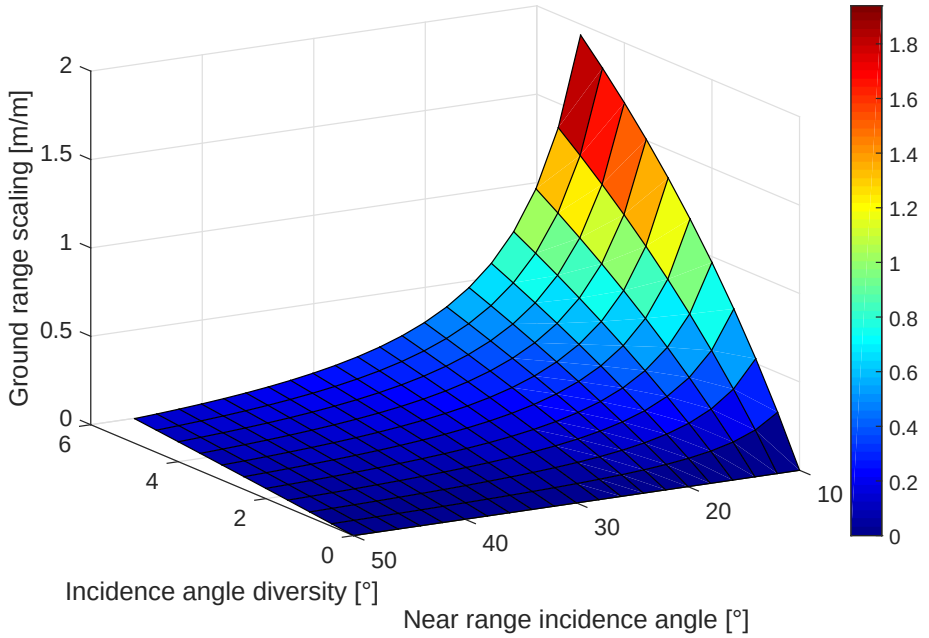


Fig. 2.3. Maximum ground range scaling (per meter of the height of reference point) as a function of the near range incidence angle and the incidence angle diversity between near range and far range. For a typical image stack of 40° incidence angle and 1° diversity, the scaling is 3.5 cm/m/°. However, the scaling will become 15 cm/m/° if the incidence angle becomes 20°, which will cause 1.5 m of scaling if the unknown height of the reference point is 10 m.

An more accurate method to mitigate this shift is to find the absolute 3-D position of the reference point. This concept is the so-called *geodetic InSAR* in which: by means of "SAR imaging geodesy" (Eineder et al., 2011; Cong et al., 2012; Balss et al., 2014) the absolute 2-D (range and azimuth) positions of the targets such as corner reflectors and PSs can be estimated to a centimeter-level accuracy after compensating for the most prominent geodynamic and atmospheric error sources; utilizing the stereo configuration available from multiple viewing angles, the absolute 3-D positions of the selected points can be retrieved using StereoSAR (Gisinger et al., 2015); finally, the InSAR estimates are reconstructed with respect to these points whose absolute positions are known. In this manner, the obtained geodetic InSAR estimates are absolute (Zhu et al., 2015c). However, this strategy relies on the successful identification of exactly identical scattering object in SAR images acquired from different view angles which is very challenging (Gisinger et al., 2015; Zhu et al., 2015c).

2.4 Robust Estimation

The estimator of a parameter is derived based on certain assumptions of the measurements. Given the PDF of the measurement noise, MLE is the best unbiased estimator if no prior knowledge of the parameter is given. In practice, we should also consider the robustness of the estimator. A robust estimator should degrade gracefully with respect to the violation of the assumption. For multipass InSAR techniques, we are interested in the performance of the estimator with respect to different SAR signal and phase statistics.

The performance of an estimator is reflected in the variance and bias of its estimate. An estimator which is unbiased and attains the Cramér-Rao lower bound (CRLB) is said to be *efficient* (Kay, 2001). Therefore, the term *efficiency* in the context of estimator often refers the ratio of the estimator's variance with respect to the CRLB.

In the following, the basics of MLE will be recapped in order to introduce the M-estimator which brings a unified view of weighted averaging.

2.4.1 Maximum Likelihood Estimator (MLE)

An MLE maximizes the goodness-of-fit (likelihood) of the measurements to the parameters to be estimated. It is always with respect to a certain distribution of the measurement statistics. Consider a simple location estimation problem, where the likelihood function of the univariate observation of the location $g \in \mathbb{R}$ is $f_g(g|\mu)$ in which μ is the expected location to be estimated. Let $f_g(g|\mu)$ equal a known PDF $f_x(x)$, where $x = g - \mu$. Without losing generality, let us assume that x is scaled to unit variance. Given multiple independent observations of g, the MLE of μ maximizes the product of the likelihood:

$$\hat{\mu}_{\text{MLE}} = \arg\max_{\mu} \prod_i f_x(x_i) = \arg\min_{\mu} \sum_i -\ln f_x(x_i) \tag{2.14}$$

where x_i is $g_i - \mu$. The above equation is usually done in the logarithmic domain, because $f_x(x)$ is often in certain exponential form, e.g. Gaussian distribution under which (2.14) is simplified to

$$\hat{\mu}_{\text{MLE}} = \arg\min_{\mu} \sum_i (g_i - \mu)^2 = \text{average}(g_i) \tag{2.15}$$

2.4.2 M-estimator

M-estimators are a class of well-known robust estimators that take the form of a minimization problem: $\min \sum_i \rho(x_i)$. The M-estimator permits a minimization of a customized loss function $\rho(x)$ of the residuals instead of only the quadratic form, to resist outliers. The M-estimator of μ is as follows:

$$\hat{\mu}_{\text{M-est}} = \arg\min_{\mu} \sum_i \rho(x_i) \tag{2.16}$$

Therefore, choosing $\rho(x) = -\ln f_x(x)$ is equivalent to the MLE. For example, the MLE of a parameter with Gaussian distributed observations is an M-estimator with $\rho(x) = x^2$.

Equation (2.16) can be solved by finding the zeros of its derivative, i.e.

$$\sum_i \frac{\rho'(x_i)}{x_i} x_i = 0 \tag{2.17}$$

where $\rho'(x)$ is the derivative of $\rho(x)$. $\rho'(x)/x$ is commonly referred to as the weighting function $w(x)$, because substituting $x_i = g_i - \mu$ into the previous equation yields an M-estimator that is actually a weighted average of the observations:

$$\hat{\mu}_{\text{M-est}} = \frac{\sum_i w(g_i - \hat{\mu})\, g_i}{\sum_i w(g_i - \hat{\mu})} \tag{2.18}$$

Equation (2.18) can be solved iteratively with an initial estimate $\hat{\mu}$ being given and the weights being updated at each iteration. Therefore, an M-estimator is actually an iteratively re-weighted averaging, and the weighting function can be directly derived from the PDF

$f_x(x)$ (if exist) of the observations with the following relation

$$w(x) = \frac{-\partial \ln f_x(x)}{x \partial x} \tag{2.19}$$

Typically, the weighting function is designed to have high efficiency for normal distribution. Therefore, a more common form of the weights is $w = \rho'(x/\sigma)/(x/\sigma)$ where σ is the standard deviation of x. Should the PDF be unknown, one may use empirical robust weighting function, such as the Tukey biweight functions, such as the Tukey biweight function (Tukey, 1977). The convergence of (2.18) is only guaranteed for convex ρ functions (Venables and Ripley, 2002). The initial estimate must be carefully estimated should ρ be nonconvex.

To summarize, M-estimator is a trade-off between robustness and efficiency. Nevertheless, it can still maintain satisfactory efficiency under the nominal model by properly choosing the loss function.

2.4.3 Influence Function

The influence function qualitatively measures the robustness of the estimator. It describes the influence of infinitesimal outlier at an arbitrary point on the estimator. Given the nominal distribution F of the observation, the fraction of contamination ε, and the estimate $\hat{g}$, the influence function is defined as follows:

$$IF(x|\hat{g}, F) = \lim_{\varepsilon \to 0^+} \frac{\hat{g}(G) - \hat{g}(F)}{\varepsilon} \tag{2.20}$$

where $G = \varepsilon \Delta_x + (1-\varepsilon)F$ is the contaminated distribution with Δ_x being the point-mass probability.

The influence function of a qualitatively robust estimator should be bounded and continuous, which means a small fraction of outliers can have only limited effect on the estimates, and gradual changes in the data also lead to gradual changes in the estimate (Zoubir et al., 2012).

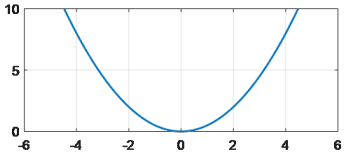
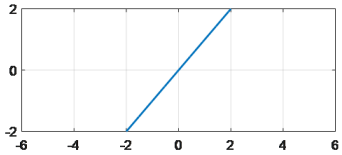
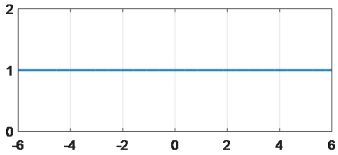
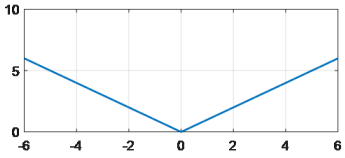
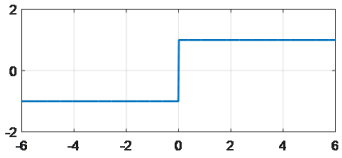
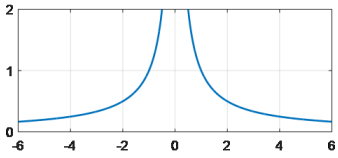
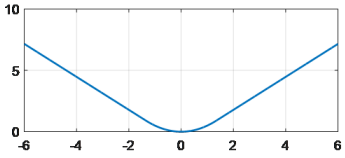
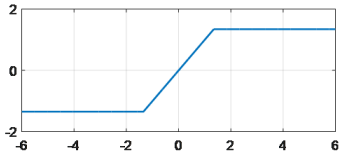
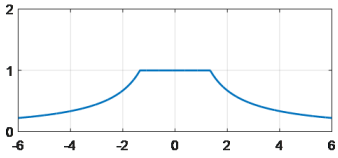
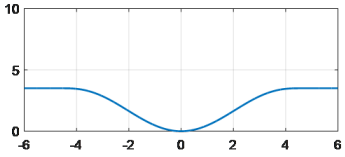
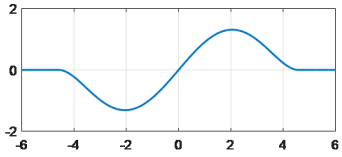
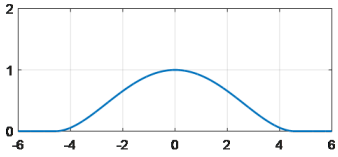
For M-estimators, the influence function with nominal distribution being Gaussian is proportional to the derivative of the loss function, i.e. $IF \propto \psi(x) = \rho'(x)$. This allows us to conveniently study the performance of different robust loss functions. Table 1 shows some typical loss functions, and the corresponding influence and weighting functions. And their plots are also shown in Table 2.

The L_2 loss function corresponds to Gaussian distributed noise. Its influence function is not bounded. One large outlier will heavily bias the estimates. If we generalize to L_p norm minimization, e.g. L_1, its influence function is bounded, meaning the influence of large outliers is limited. Another popular robust loss function is the Huber function named after P.J. Huber — the pioneer in robust estimation (Huber, 1981). Tukey biweight function (Tukey, 1977) is one of the re-descending functions, since its influence function is bounded and descends to zeros as the position of the outlier increases to a certain value. The parameter $c \in \mathbb{R}^+$ in Huber and Tukey functions is to tune the performance of the estimator. The lower the value of c, the more robust but less efficient is the estimator for Gaussian distributed noise, and vice versa. Therefore, c is usually chosen to give a relative efficiency of 95% at normal distribution. This number is 1.345 and 4.586 for Huber and Tukey functions, respectively (Venables and Ripley, 2002).

Table 1. Some typical loss functions, and their corresponding influence and weighting functions.

Function	Loss function $\rho(x)$	Influence function $\psi(x)$	Weighting function $w(x)$
L_2	$x^2/2$	x	1
L_p	$\lvert x\rvert^p/p$	$x\lvert x\rvert^{p-2}$	$\lvert x\rvert^{p-2}$
Huber $\begin{cases}\lvert x\rvert < c\\ \lvert x\rvert \geq c\end{cases}$	$\begin{cases} x^2/2 \\ c\lvert x\rvert - c^2/2\end{cases}$	$\begin{cases} x \\ cx/\lvert x\rvert\end{cases}$	$\begin{cases} 1 \\ c/\lvert x\rvert\end{cases}$
Tukey $\begin{cases}\lvert x\rvert < c\\ \lvert x\rvert \geq c\end{cases}$	$\begin{cases} c^2/6 - \left(c^2 - x^2\right)^3/6c^4 \\ c^2/6\end{cases}$	$\begin{cases} x\left(c^2 - x^2\right)^2/c^4 \\ 0\end{cases}$	$\begin{cases} \left(c^2 - x^2\right)^2/c^4 \\ 0\end{cases}$

Table 2. Plots of the loss functions, and their corresponding influence and weighting functions.

Function	Loss function $\rho(x)$	Influence function $\psi(x)$	Weighting function $w(x)$
L_2			
L_p (p=1)			
Huber (c=1.345)			
Tukey (c=4.685)			

3 State-of-the-art of Multipass SAR Interferometry

This chapter focuses on the estimation aspect of the state-of-the-art multipass InSAR techniques including persistent scatterer interferometry (PSI), small baseline subset (SBAS), SqueeSAR, and TomoSAR.

The discussion is restricted to the two extreme cases of scattering objects: PS and DS. DSs are the most natural scatterers in a SAR image. Because of their scattering mechanism, they experience temporal and geometric decorrelation. To reduce the effect of decorrelation, one can either use interferograms of small spatial and temporal baselines, i.e. SBAS, or average pixels of a same scattering statistics, i.e. SqueeSAR.

PS overcomes the limitation of DS by identifying single scatterers whose phase is stable throughout the SAR image stack. Theoretically, PSs do not suffer from geometrical decorrelation, permitting the use of interferograms with large spatial baseline even beyond the critical one.

TomoSAR is a generalization of the single phase center PS or DS model, by considering the full reflectivity profile along the radar third dimension *elevation*. It solves the layover of scatterers in a range-azimuth resolution cell.

3.1 State-of-the-art

3.1.1 Persistent Scatterer Interferometry

General Remark

PSI was developed for long-term ground deformation monitoring (Ferretti et al., 2001). Due to the long temporal baseline, low resolution, and large spatial baseline of some SAR image stacks, the individual DS pixel often lose their coherence, rendering them not directly useful for long-term deformation monitoring. PSI was invented to exploit the PSs whose scattering phases are stable throughout the whole SAR image stack. Not only are the phases of PSs stable, but also is the dimension of a PS often much smaller than a resolution cell, permitting the use of image with large spatial baseline even beyond the critical one.

There exist several implementations of PSI (Ferretti et al., 2001; Adam et al., 2003; Hooper et al., 2004; Kampes and Adam, 2006) etc., as well as improvements using different motion models (Colesanti et al., 2003; Gernhardt and Bamler, 2012), improvements using dense differential PS pairs (Wang et al., 2014; Costantini et al., 2014), and some variants of PSI that exploit partially coherent PSs (Wegmüller et al., 2010; Perissin and Wang, 2012) which follow the same spirit of SBAS that will be covered in the next section. Generally, PSI estimates the atmospheric phase, the topography, and the motion parameters of each PS. A generally recognized workflow is as follows (Ferretti et al., 2001):

(1) N interferograms are formed from N+1 co-registered SAR images. A master scene is carefully selected to maximize the overall coherence. A reference DEM is subtracted from all the interferograms to flatten the topographic phase.
(2) An initial set of PS candidates is selected, which should represent the set of most stable PSs in the scene. A few methods have been proposed to fulfil this task:
   ◇ using amplitude dispersion analysis for high signal-to-noise-ratio (SNR) PSs (Ferretti et al., 2001). This is mostly applicable to urban areas.

- ◇ using phase analysis for low SNR PSs (Hooper et al., 2004), which was demonstrated to outperform the previous method in non-urban area.
- ◇ and alternatively using signal-to-clutter-ratio to approximate the phase variance (Adam et al., 2004).

(3) The PS candidates are connected to form a reference network. The differential PS pairs, also known as *arcs*, are limited to be within the spatial correlation length of typical atmospheric phase signal, so that the atmospheric phase can be greatly mitigated through differentiation. This length was found to be 2-3 km in (Hanssen, 2006). After that, the differential phase history parameters θ, i.e. the differential topography height and differential motion parameters, of the arcs are estimated.

The differential height estimates are integrated with respect to a reference point, so that their absolute phase contribution can be removed from the interferometric phase, leaving the phase with contributions only from the deformation, atmosphere, and noise. The atmospheric phase is then extracted by low-pass filtering in the spatial domain followed by high-pass in the temporal domain, assuming the atmospheric phase is highly correlated in the spatial domain and temporally nearly uncorrelated for typical temporal baselines of tens of days.

The atmospheric phase is interpolated over the whole scene, so that it can be subtracted from the interferograms.

(4) More PSs are detected after the mitigation of the atmospheric phase. They are linked to the nearest point in the reference network, and their phase history parameters are estimated.

(5) The positions of PSs are geocoded to geographical coordinate system for further analysis.

Maximum Likelihood Estimator

On the estimation point of view, a PS pixel value is modelled as a deterministic complex value with additive CCG noise. For a time series of co-registered PSs, $\mathbf{g}$, its noise is assumed to follow an independent and identical distribution (i.i.d.). Under such assumption, the MLE of the phase history parameters θ is (Rife and Boorstyn, 1974; Wang and Zhu, 2015b)

$$\hat{\theta}_{MLE} = \arg\max_{\theta} \left\{ \left(2\pi\sigma^2\right)^{-N} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{g} - \bar{\mathbf{g}}(\theta)\|_2^2\right)\right\} \tag{3.1}$$

where θ is the vector of phase history parameters comprised of topography height h, linear deformation rate v, etc., $\bar{\mathbf{g}}(\theta)$ is the modelled PS signal, and σ^2 is the variance of the complex noise. Simplifying the above equation leads to the well-known formulation of periodogram or the so called "temporal coherence" (Rife and Boorstyn, 1974; Colesanti et al., 2003; De Maio et al., 2009):

$$\hat{\theta}_{MLE} = \arg\max_{\theta} \left\{ \left| \frac{1}{N} \sum_{n=1}^{N} g_n \exp\left(-j\varphi_n(\theta)\right) \right| \right\} \approx \arg\max_{\theta} \left\{ \left| \frac{1}{N} \sum_{n=1}^{N} \frac{g_n}{|g_n|} \exp\left(-j\varphi_n(\theta)\right) \right| \right\} \tag{3.2}$$

where $\varphi_n(\theta)$ is the modelled phase of the PS in the nth image. Often, the amplitude of $\mathbf{g}$ is dropped in the estimation (Ferretti et al., 2001), since it barely changes the estimates for high SNR PSs.

Equation (3.2) can also be understood as the parameter estimation of a complex signal with single frequency in each dimension of θ, since the modelled phase $\varphi_n(\theta)$ is linearly proportional to θ. It finds the best fit of $\varphi_n(\theta)$ with given N wrapped phases of $\mathbf{g}$.

Least Squares Ambiguity Decorrelation (LAMBDA) Estimator

Equation (3.2) can also be treated as a phase unwrapping problem with parametric phase model. (Kampes and Hanssen, 2004) solved it using the LAMBDA method which is originally developed for the ambiguity resolution of GPS signal.

The LAMBDA method for PS parameter estimation is formulated as $\boldsymbol{\varphi} = \mathbf{A}\mathbf{a} + \mathbf{B}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$, i.e.

$$\begin{bmatrix} \varphi_1 \\ \varphi_1 \\ \vdots \\ \varphi_N \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ -2\pi & & & \\ & -2\pi & & \\ & & \ddots & \vdots \\ 0 & & \cdots & -2\pi \end{bmatrix} \begin{bmatrix} a_2 \\ \vdots \\ a_N \end{bmatrix} + \begin{bmatrix} \xi_1 & \zeta_1 & \cdots \\ \xi_2 & \zeta_2 & \\ \vdots & \vdots & \\ \xi_N & \zeta_N & \cdots \end{bmatrix} \begin{bmatrix} h \\ v \\ \vdots \end{bmatrix} + \boldsymbol{\varepsilon} \tag{3.3}$$

where $\boldsymbol{\varphi} \in \mathbb{R}^N$ is the wrapped phase of $\mathbf{g}$, $\mathbf{A}$ is a diagonal matrix of -2π but discarding the first column (since the ambiguities are relative), $\mathbf{a} \in \mathbb{Z}^{N-1}$ is the integer ambiguity, matrix $\mathbf{B}$ is the model matrix of the phase containing the parameter-to-phase conversion factors (ξ, ζ, etc.), and $\boldsymbol{\varepsilon}$ is the phase noise.

The LAMBDA estimator is an integer least square problem:

$$\hat{\boldsymbol{\theta}}_{LAMBDA} = \arg\min_{\mathbf{a},\boldsymbol{\theta}} \|\boldsymbol{\varphi} - \mathbf{A}\mathbf{a} - \mathbf{B}\boldsymbol{\theta}\|_2^2 \quad \text{s.t. } \mathbf{a} \in \mathbb{Z}^{N-1} \tag{3.4}$$

Equation (3.4) assumes a Gaussian phase noise which is not true for PS with CCG noise, but can be valid for PS of high SNR.

Considering the integer constraint of $\mathbf{a}$, solving (3.4) is a three-step procedure in (Kampes and Hanssen, 2004).

(1) The unconstrained estimates $\hat{\mathbf{a}}$ and $\hat{\boldsymbol{\theta}}$, also known as float solution, are computed by unconstrained least square. Important to note that the linear system of equation (3.3) is underdetermined, i.e. more unknowns than observations. Regularization, such as augmenting the system with pseudo-observations, must be introduced in order to solve the linear equation system.

(2) Given the float solution $\hat{\mathbf{a}}$, its integer-constrained solution is equation (3.5).

$$\check{\mathbf{a}} = \arg\min_{\mathbf{a}} \|\hat{\mathbf{a}} - \mathbf{a}\|^2_{\mathbf{C}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}} \quad s.t.\ \mathbf{a} \in \mathbb{Z}^{N-1} \tag{3.5}$$

If each element of $\hat{\mathbf{a}}$ is uncorrelated, i.e. the covariance matrix $\mathbf{C}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}$ is diagonal, (3.5) is a simple rounding. Otherwise, the correlation among the elements of $\hat{\mathbf{a}}$ should be mitigated by multiplying it with an integer matrix $\mathbf{Z}$ that decorrelates it the most. In another word, given the covariance matrix $\mathbf{C}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}$ of $\hat{\mathbf{a}}$, the covariance matrix of $\hat{\mathbf{z}} = \mathbf{Z}^T\hat{\mathbf{a}}$ which is $\mathbf{C}_{\hat{\mathbf{z}}\hat{\mathbf{z}}} = \mathbf{Z}^T\mathbf{C}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}\mathbf{Z}$ should be as diagonal as possible. In practice, $\mathbf{Z}$ matrix is found by the LDL decomposition (Golub and Loan, 1996) of $\mathbf{C}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}$. To this end, (3.5) is transformed to

$$\check{\mathbf{z}} = \arg\min_{\mathbf{z}} \|\hat{\mathbf{z}} - \mathbf{z}\|_2^2 \quad s.t.\ \mathbf{z} \in \mathbb{Z}^{N-1} \tag{3.6}$$

And the solution search space has been greatly reduced. The diagonalness of $\mathbf{C}_{\hat{\mathbf{z}}\hat{\mathbf{z}}}$ is not always guaranteed. Therefore, a series of bounds for $\mathbf{z}$ is computed, in order to further restrict the search space.

(3) The final integer-constrained solution $\check{\boldsymbol{\theta}}$ is computed using least square after subtracting the 2π ambiguities.

Summary

To conclude, the estimation of the phase history parameters of a PS is rather straightforward, a benefit brought by the simple point scatterer model. The estimation techniques include the MLE (periodogram) and LAMBDA.

The accuracy of the estimates primarily depends on the number of images N and the SNR of the PS. Typically, for PSs with relatively high SNR, e.g. 10dB, it requires around 30 images to achieve a millimeter precision on the yearly linear deformation rate (Ferretti et al., 2001).

Comparing MLE and LAMBDA, they produce nearly identical results as demonstrated in (Gernhardt et al., 2007). Nevertheless, LAMBDA explicitly solves the ambiguity problem, whereas the periodogram does not. But periodogram will outperform LAMBDA for low SNR PSs, since the assumption of Gaussian phase noise in LAMBDA is no longer valid for low SNR PSs. However, LAMBDA is more computationally efficient than MLE (periodogram) because the ambiguity solution space in LAMBDA is greatly reduced, and the phase parameters are solved analytically by least square. Still, PSI is the most computationally efficient one among all state-of-the-art multipass InSAR methods (even periodogram is used). Therefore, it is widely used, and has become the workhorse for deformation monitoring.

3.1.2 Small Baseline Subset

General Remark

SBAS exploits DS instead of PS. As we know, DSs may suffer from temporal decorrelation. The innovation of SBAS (Berardino et al., 2002) is to form interferograms of only small baselines, so that the geometric and temporal decorrelation of DSs can be controlled. The following figure shows an example of the baseline plots of a single master InSAR stack (e.g. PSI) and SBAS. The scattered dots are the perpendicular baseline versus the temporal baseline of each image. PSI links every image to a single master, while SBAS only forms interferograms with small spatial and temporal baselines, resulting in multiple masters.

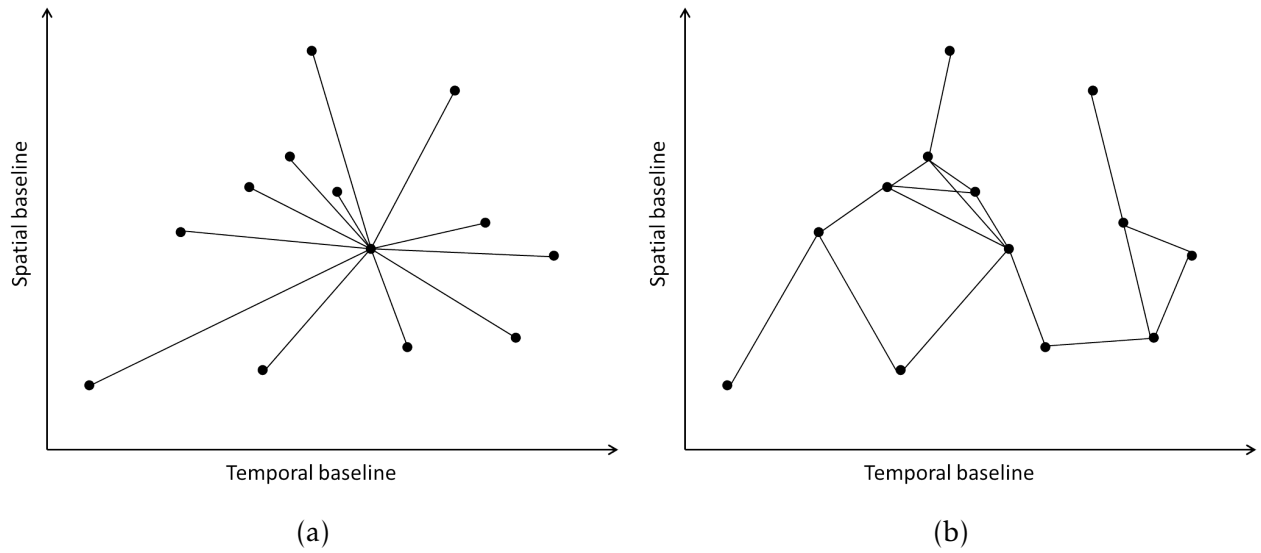


Fig. 3.1. Spatial-temporal baseline plot of (a) single master InSAR stack, e.g. PSI, and (b) SBAS InSAR stack.

SBAS was firstly proposed in (Berardino et al., 2002). Several similar techniques exist, including (Mora et al., 2003; Blanco-Sánchez et al., 2008), an update from the original group (Lanari et al., 2004), as well as (Goel and Adam, 2012a) which make use of adaptive multilooking to further reduce the decorrelation of DSs — an improvement following the innovation of SqueeSAR. Nevertheless, the spirit of SBAS is to use small baseline, which does

not only limit to DS. This has been shown in applying to partially coherent PSs (Wegmüller et al., 2010; Perissin and Wang, 2012).

A generally recognized procedure of SBAS is as follows (Berardino et al., 2002).

(1) Given a stack of N+1 co-registered images, multiple interferograms of small spatial and temporal baselines with individual masters are formed, instead of N interferogram with a single master. The atmospheric phase is assumed to be already removed. Otherwise, the procedure is similar to that explained in PSI.
(2) Coherent pixels are selected by thresholding on the coherence. Note its difference to the PS selection. The coherence explicitly defines the variance of interferometric phase noise by the analytical expression (Tough et al., 1995; Bamler and Hartl, 1998):

$$\sigma_{\varphi}^{2}\left(|\gamma|\right) = \pi^{2}/3 \; - \pi \arcsin\left(|\gamma|\right) + \arcsin^{2}\left(|\gamma|\right) - \mathrm{Li}_{2}\left(|\gamma|^{2}\right)/2 \tag{3.7}$$

where Li_2 is the dilogarithm.
(3) The interferograms are unwrapped in 2-D. The small baselines reduce the topographic phase contribution, and eases the unwrapping step.
(4) The topography height and the low frequency deformation signal (usually a linear deformation model) are estimated. This step is similar to the least square estimation in LAMBDA after the 2π ambiguity been fixed. The phase contribution from topography is subtracted from the unwrapped phase. Considering a possible phase unwrapping error in step 3, the residual phase is usually unwrapped once again.
(5) The remaining interferometric phase has contributions from non-linear motion signal and noise. They are with respect to multiple master images. In order to reconstruct the N absolute phase with respect to a single master, one needs to solve an linear equation system $\Delta\boldsymbol{\varphi} = \mathbf{A}\boldsymbol{\varphi} + \boldsymbol{\varepsilon}$, i.e.:

$$\begin{bmatrix} \varphi_{2-1} \\ \varphi_{3-2} \\ \vdots \\ \\ \\ \end{bmatrix} = \begin{bmatrix} -1 & 1 & \cdots & \\ \vdots & -1 & 1 & \\ & -1 & & 1 \\ & & \ddots & \\ & & -1 & 1 \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_1 \\ \vdots \\ \varphi_N \end{bmatrix} + \boldsymbol{\varepsilon} \tag{3.8}$$

where $\Delta\boldsymbol{\varphi} \in \mathbb{R}^M$ is the unwrapped phase of the small baseline interferograms after removing the topographic phase, $\mathbf{A}$ of dimension $M \times N$ is the design matrix of the small baseline interferograms , $\boldsymbol{\varphi} \in \mathbb{R}^N$ the absolute phase with respect to a single master, and $\boldsymbol{\varepsilon}$ the phase noise.

Equation (3.8) is often overdetermined but ill-posed, due to the fact that not all the interferogram are connected to a single subset, rendering the rank of A to be less than N. Therefore, regularization is required. Singular value decomposition (SVD) has been used for this task (Liebhart et al., 2010).

Summary

The parameter estimation in SBAS is all based on linear least squares. Hence similar to LAMBDA estimator, it also assumes a Gaussian phase noise, which is not true under the strict DS model. Regularization is required if the interferograms do not connect into a single subset. Computational-wise, SBAS requires the unwrapping of a substantial number of interferograms, which is usually much larger than N. Therefore, the phase unwrapping is the most time consuming and risky step in SBAS.

3.1.3 SqueeSAR (Adaptive Multilooking)

General Remark

Same as SBAS, SqueeSAR also exploits the DSs. Instead of restricting the DS decorrelation using interferograms of small baselines, SqueeSAR tackles it by averaging the DS pixels. As we know, the MLE of the interferometric phase given M ergodic DS samples is

$$\hat{\varphi}_{MLE} = \arg\left(\sum_{m=1}^{M} v\,[m]\right) \tag{3.9}$$

This idea is also known as *multilooking*, e.g. the "box car" filter. However, the samples selected in a rectangle window are often not ergodicity, and hence, produce unsatisfactory results. Therefore, the core innovation of SqueeSAR (Ferretti et al., 2011) and its variants (Wang et al., 2012; Goel and Adam, 2012a; Jiang et al., 2015) is adaptive multilooking.

A generally recognized SqueeSAR workflow is as follows (Ferretti et al., 2011):

(1) N interferograms are formed with respect to a single master. This step is the same as PSI.
(2) The neighbourhood (similar pixels) of each pixel is found by comparing their intensity series using a statistical test. The neighbourhood pixels are averaged to produce the filtered interferogram. The statistical test can be either parametric or non-parametric.
   ◇ Non-parametric test

   Kolmogrov-Smirnov (KS) test

   It measures the maximum distance between the empirical cumulative distribution functions (CDFs) of two data sets (Ferretti et al., 2011), i.e.

$$D_{KS} = \max_{x} |F_a(x) - F_b(x)| \tag{3.10}$$

   where $F_a(x)$ and $F_b(x)$ are the empirical CDFs of the two target pixels' intensities, respectively. By setting a threshold on the distance, the null hypothesis that the two pixels belong to the same distribution is accepted or rejected. However, the maximum distance of two CDFs usually happens in the midway of the CDF curve for any distribution. Therefore, the KS test is not sensitive to unknown distributions with variations on the head or tail.

   Anderson-Darling (AD) test

   Likewise, AD test also measures the distance between two CDFs, but it includes weighting according to the shape of the distribution (Parizzi and Brcic, 2011; Wang et al., 2012; Goel and Adam, 2012a), i.e.

$$A^2 = \frac{N}{2} \sum \frac{(F_a(x) - F_b(x))^2}{F_{a,b}(x)(1 - F_{a,b}(x))} \tag{3.11}$$

   where $F_{a,b}(x)$ is the CDF of the combined pixels' intensities. It places more weight on the head and tails of the distributions. AD test is actually a weighted L_2-norm of the difference of two CDFs.

   ◇ Parametric test

   Parametric test compares the intensity series of two pixels through generalized likelihood ratio test conditioned on the null and alternative hypotheses, with a given

parametric model of the intensity distribution. It gives the best performance if the data follows the assumed intensity distribution.

The marginal distribution of a DS intensity series, if not considering the correlation between images, is χ^2 distribution with two DoF (equation (2.3)). Since the expectation value fully characterizes this distribution, the null and alternative hypotheses are H_0: $\bar{I} = \bar{I}_a = \bar{I}_b$ and H_1: $\bar{I}_a \neq \bar{I}_b$ where $\bar{I}_a$, $\bar{I}_b$, and $\bar{I}$ are the expected intensity of the target pixels a, b, and combined. The test statistic is then (Parizzi and Brcic, 2011):

$$T = 2N \ln\left(\hat{\bar{I}}\right) - N \ln\left(\hat{\bar{I}}_a\right) - N \ln\left(\hat{\bar{I}}_b\right) \overset{H_1}{>} 0 \tag{3.12}$$

where $\hat{\bar{I}}_a$, $\hat{\bar{I}}_b$, and $\hat{\bar{I}}$ are the MLE of $\bar{I}_a$, $\bar{I}_b$, and $\bar{I}$, respectively.

An advance of the parametric similarity test is demonstrated in the non-local InSAR framework (Deledalle et al., 2015), where the similarity is compared patch-wisely taken into account the interferometric phase.

(3) The multilooked interferometric phase is not conservative, i.e. $\bar{\varphi}_{1-2} + \bar{\varphi}_{2-3} + \bar{\varphi}_{3-1} \neq 0$. They are "triangulated" by solving a linear equation system similar to the SBAS equation (3.8).

(4) The phase history parameters of the filtered DS are estimated as a PS using periodogram or LAMBDA.

DS MLE

On the estimation point of view, the MLE of the phase of CCG distributed DS can be formulated as follows (De Zan and Rocca, 2005; De Zan, 2008; Guarnieri and Tebaldini, 2008; Ferretti et al., 2011; Wang et al., 2012):

$$\hat{\theta} = \arg\min_{\theta} \left\{ \mathbf{g}^H \mathbf{\Phi}(\theta) \operatorname{abs}\left(\hat{\mathbf{C}}\right)^{-1} \mathbf{\Phi}(\theta)^H \mathbf{g} \right\} \tag{3.13}$$

where $\mathbf{\Phi}(\theta) \in \mathbb{C}^{N \times N}$ is the diagonal matrix containing the modelled phase of $\mathbf{g}$, and $\operatorname{abs}(\cdot)$ is the element-wise absolute value of a matrix. Given a stationary neighbourhood of the target DS, the MLE maximize the joint likelihood:

$$\hat{\theta} = \arg\min_{\theta} \sum_{m=1}^{M} \mathbf{g}_m^H \mathbf{\Phi}(\theta) \operatorname{abs}\left(\hat{\mathbf{C}}\right)^{-1} \mathbf{\Phi}(\theta)^H \mathbf{g}_m \tag{3.14}$$

where m is the index of the neighbouring sample. In (3.14), the same phase parameters θ is assumed for all the observations, i.e. $\mathbf{g}_i$ for $m = 1, ..., M$.

Summary

SqueeSAR is a powerful method to make use of partially decorrelated DSs. The core of SqueeSAR is adaptive averaging. SqueeSAR represents the state-of-the-art in DS processing, since it takes both advantage of interferograms of small baselines, as well as the adaptive multilooking.

The performance of standard SqueeSAR greatly depends on the similarity test. The performance of similarity tests degrades as the number of images decreases (Parizzi and Brcic, 2011). Computational-wise, SqueeSAR is tens of times more expensive than PSI. However, the overall computation is still moderate compared to TomoSAR.

3.1.4 SAR Tomography and Differential SAR Tomography

General Remark

As mentioned in section 2.1.3, TomoSAR solves the layover problem where the full reflectivity profile $\gamma(s)$ along the third dimension *elevation s* is retrieved. The following equation is obtained by subtracting the phase at zero elevation from equation (2.4) (ignore the second order elevation term in the phase) (Reigber and Moreira, 2000; Lombardini et al., 2003; Fornaro et al., 2003; Zhu, 2008):

$$g_n = \int_{\Delta s} \gamma(s) \exp\left(-j\frac{4\pi b_n}{\lambda R}s\right) ds \tag{3.15}$$

where g_n and b_n are the pixel value and the perpendicular baseline of the nth image. In the case of differential TomoSAR (D-TomoSAR), equation (3.15) is extended into higher dimensions (Zhu and Bamler, 2010b, 2011). Write (3.15) in a more compact matrix form, it is

$$\mathbf{g} = \mathbf{R}\boldsymbol{\gamma} \tag{3.16}$$

where $\mathbf{R}$ is basically a discrete Fourier matrix whose frequencies are linearly proportional to the baselines. A generally recognized TomoSAR workflow is as follows (Zhu and Bamler, 2010b):

(1) Interferograms are formed like PSI. The atmospheric phase screen is mitigated.
(2) The reflectivity profile is reconstructed using spectral estimation techniques.
(3) The scatterers' positions and motion parameters are determined by detecting maxima on the reflectivity profile.

Since the first step is the same as a standard PS processing, the following discussion will focus on the estimation of the reflectivity profile.

MLE and Maximum A Posteriori (MAP) Estimator

Given a deterministic $\mathbf{g}$ with i.i.d. complex Gaussian noise, the MLE of $\boldsymbol{\gamma}$ is

$$\hat{\boldsymbol{\gamma}}_{MLE} = \min_{\boldsymbol{\gamma}} \|\mathbf{g} - \mathbf{R}\boldsymbol{\gamma}\|_2^2 \tag{3.17}$$

For stochastic $\mathbf{g}$, the MLE should consider the covariance matrix of the signal (Wax, 1991; Sauer et al., 2009; Tebaldini, 2010). Since the dimension of $\boldsymbol{\gamma}$ is often much larger than $\mathbf{g}$, the equation $\mathbf{g} = \mathbf{R}\boldsymbol{\gamma}$ is an underdetermined system. There are a few methods to regularize the equation system, i.e. MAP estimators:

◇ Nonlinear least square (Stoica and Moses, 2005): Assuming the resolution cell consists of K scatterers with their elevation $\mathbf{s} = [s_1, s_2, \cdots, s_K]$, this system can be reduced to where the support of the signals are. For small number of K which is usually the case, it becomes an overdetermined system. The minimization is reduced to the following form.

$$\hat{\boldsymbol{\gamma}}(\hat{\mathbf{s}})_{NLS} = \min_{\boldsymbol{\gamma}(\mathbf{s})} \|\mathbf{g} - \mathbf{R}(\mathbf{s})\boldsymbol{\gamma}(\mathbf{s})\|_2^2 \tag{3.18}$$

The dimension of $\mathbf{R}(\mathbf{s})$ and $\boldsymbol{\gamma}(\mathbf{s})$ are much smaller than $\mathbf{R}$ and $\boldsymbol{\gamma}$, respectively. However, they depend on the location of $\mathbf{s}$ which needs to be estimated simultaneously. Application in layover separation has been demonstrated in (Ferretti et al., 2005; Zhu, 2008).

◇ Tikhonov regularization: apart from the undermined equation system, the matrix $\mathbf{R}$ can be rank deficient due to the irregularity and small span of the baseline. That is to say, (3.16) can be ill-posed and underdetermined. The Tikhonov regularized MAP is as follows:

$$\hat{\boldsymbol{\gamma}}_{Tikho} = \min_{\boldsymbol{\gamma}} \|\mathbf{g} - \mathbf{R}\boldsymbol{\gamma}\|_2^2 + \alpha_{Tik}\|\boldsymbol{\gamma}\|_2^2 \tag{3.19}$$

which restricts the energy of $\boldsymbol{\gamma}$ by performing L_2-norm regularization. (Zhu and Bamler, 2010b) implemented the MAP using SVD with Wiener weighing on the eigenvalues. Alternatively, (Fornaro et al., 2003) truncated small eigenvalues.

◇ Compressive sensing (CS)-based estimators exploits the sparsity prior of the reflectivity profile, i.e. assuming only small number of scatterers on the reflectivity profile. Instead of the L_2-norm regularization in equation (3.19), sparse reconstruction of $\boldsymbol{\gamma}$ penalize on its L_1-norm:

$$\hat{\boldsymbol{\gamma}}_{CS} = \min_{\boldsymbol{\gamma}} \|\mathbf{g} - \mathbf{R}\boldsymbol{\gamma}\|_2^2 + \alpha_{CS}\|\boldsymbol{\gamma}\|_1 \tag{3.20}$$

The L_2- plus L_1-norm minimization is a sparse reconstruction problem. Its first application in TomoSAR was presented in (Zhu and Bamler, 2009) using real data. The same authors also developed the promising SL1MMER algorithm, which delivered the first CS-based TomoSAR reconstruction of single building (Zhu and Bamler, 2010a). The SL1MMER algorithm was also recently extended to the M-SL1MMER (Zhu et al., 2015b) which exploits the group sparsity in the urban environment. M-SL1MMER achieves comparable result with far less images than SL1MMER and other existing algorithms. Other CS-based TomoSAR with COSMO-Skymed data (Budillon et al., 2011) and airborne millimetre-wave SAR data (Schmitt and Stilla, 2013) were also presented. The CS-based TomoSAR algorithms are much more computationally expensive than the Tikhonov regularization.

Summary

TomoSAR is so far the most competent InSAR method for urban and forest area monitoring. The high computational cost limits it for an extensive use like PSI, especially for the CS-based TomoSAR algorithms.

Since the forward model $\mathbf{R}$ is a discrete Fourier transform, TomoSAR is basically a spectral estimation problem. Many classical algorithms can be applied to TomoSAR, such as beamforming (Pasquali et al., 1995; Lombardini et al., 2003; Fornaro et al., 2009), CAPON (Lombardini et al., 2001), MUSIC (Gini and Lombardini, 2002; Guillaso and Reigber, 2005; Huang et al., 2012; Nannini et al., 2012), etc. Please refer to (Stoica and Moses, 2005; Zhu, 2011) for an extensive review of these classical methods.

3.1.5 InSAR Point Cloud Fusion

General Remark

The abovementioned InSAR techniques deliver 3-D point clouds together with the motion parameters. Since the point clouds are relative to their reference points, they need to be co-registered when considering the results from multiple stacks. Although the general point cloud fusion is a classical topic in the computer vision field, there is actually only a handful of studies on InSAR point cloud fusion, especially for point clouds from image stacks of cross-heading orbits (i.e. ascending and descending), because there is theoretically no corresponding point from such two point clouds.

The first attempt of fusing cross-heading InSAR point clouds in urban area was presented in (Gernhardt and Bamler, 2012). This algorithm is essentially a 3-D surfaces matching. After removing the façade points where the point correspondence is most unlikely to appear, the point correspondences are found by searching closely spaced point pairs on surface. Therefore, this algorithm does not address the exact point correspondence.

The fusion of along-heading (both ascending or both descending) point clouds or point clouds of smooth terrain is less challenging. Classical point cloud co-registration methods such as iterative closest point (ICP) can be directly applied. (Gernhardt et al., 2012) has demonstrated the direct application of ICP on multiple InSAR point clouds of a volcano.

Summary

The major difficulty in cross-heading InSAR point clouds fusion is the lack of exact corresponding point pairs, due to the SAR side-looking imaging geometry. The current methods are all based on inexact match, except a recent advance (Zhu et al., 2014, 2015c) where the authors manually locate dozens of point pairs of street lampposts which are supposed to be the closest match to exact point pair, and performed careful calculation of the absolute position of the lamppost using the concept of StereoSAR (Gisinger et al., 2015), with additionally taking into account the diameter of the lamppost.

3.1.6 Optical and SAR Images Fusion

General Remark

There have been many works regarding the fusion of optical and SAR images. As (Soergel, 2010) mentioned: "SAR inevitably requires an oblique scene illumination resulting in undesired occlusion and layover especially in urban areas", optical images can complement such disadvantage of SAR images. The relevant works follows basically three categories: 1) geometric co-registration, 2) joint height reconstruction, and 3) joint classification.

Geometric co-registration

Geometric co-registration is the basis of most of the higher level fusion algorithms. The current techniques are mostly based on matching of 2-D spatial features, e.g. contours, edges, templates, etc. (Li et al., 1995; Ali et al., 2002; Cheng et al., 2004; Hong and Schowengerdt, 2005; Fan et al., 2013; Palubinskas and Reinartz, 2015). They are good for large scale analysis, but not applicable for the analysis of individual buildings, especially for high resolution data.

A precise 3-D model is crucial in co-registration of SAR and optical images, for example (Auer et al., 2015) used carefully reconstructed 3-D models from photogrammetric measurements; (Tao et al., 2014) used a precise airborne LiDAR point cloud; and (Schack et al., 2015) employed a PS point cloud retrieved from tens of high resolution TS-X spotlight images. Without a precise 3-D model, one needs to assume a strict prior (Brunner et al., 2010), such as rectangle footprint etc. Considering the complex 3-D geometry of urban area, there has not been an effective algorithm of matching SAR and optical image yet.

Building height reconstruction

Several literatures also explored the potential of 3-D reconstruction using a few SAR and optical images. The joint 3-D reconstruction can be said to be contextually or geometrically. Contextually, (Tupin and Roux, 2005) used the optical image as an a priori term in order to regularize the height reconstructed from radargrammetry. Geometrically, (Sportouche et al., 2011) projected the building footprint extracted from the optical image into the SAR image, and solved building height geometrically. Alternatively, (Wegner et al., 2014) retrieved the height of buildings from SAR and optical images independently, and fused them together using least square adjustment. Apart from these, there are also works on building reconstruction from multi-aspect SAR images (Xu and Jin, 2007; Thiele et al., 2007).

Joint classification

Many studies have been dedicated to the joint classification of SAR and optical images. Most of the applications are on urban mapping and classification (Gamba and Houshmand, 2002; Pacifici et al., 2008; Wegner et al., 2009; Amarsaikhan et al., 2010; Poulain et al., 2011), since the optical and SAR image provide complimentary information in urban area. As the focus of the thesis is not on the classification method, the relevant literatures are only briefly listed for the readers' interest.

Summary

Geometric fusion of optical and SAR image is an important step, because many higher level fusion is based on a very precise geometric fusion. So far, there has not been a precise fusion of meter-resolution SAR and optical image in urban area yet.

3.2 Contribution of this Thesis

3.2.1 Computationally Efficient TomoSAR Inversion

The much higher computational cost of TomoSAR is the major factor of preventing it from replacing PSI as the workhorse of InSAR monitoring. Therefore, the first contribution of this thesis is an efficient TomoSAR inversion approach. The proposed method combines PSI and TomoSAR by performing a pre-classification of single and double scatterers instead of the state-of-the-art approaches that always perform full model order selection meaning inevitable full TomoSAR inversion for all the pixels. The computational time is reduced by a factor of more than fifty.

3.2.2 TomoSAR Point Cloud Fusion

The current cross-heading InSAR point cloud fusion methods are based on inexact point match. They are essentially 3-D surfaces matching that relay on removing the façade points where the point correspondence is most unlikely to appear. Therefore, an InSAR point cloud fusion method was proposed making use of the rich façade information in TomoSAR point clouds in order to derive the theoretically exact point correspondence. The proposed method is much more computationally efficient than the state-of-the-art, while achieving consistent results.

3.2.3 InSAR Optical Fusion

InSAR estimates and optical images can complement each other. However, the fusion of the two in urban areas is still a topic under research, because a precise geometric fusion of optical and SAR images of urban areas is still a challenge. Therefore, an effective algorithm of geometrically and semantically fusing InSAR estimates and optical images are developed. The algorithm performs 3-D geometric co-registration, image classification, and eventually semantic level analysis in InSAR estimate.

3.2.4 Robust Estimators

Robustness w.r.t. Large Phase Error

A systematic solution to optimally weigh the InSAR observation with large phase errors in the multipass InSAR algorithms was developed, which has not been addressed so far. Experiments showed that the proposed algorithm increases the efficiency of the estimator by a factor of 7 to 35 compared to state-of-the-art estimators, when 40% of the images are contaminated with large phase errors.

Robustness w.r.t. Non-Gaussian Scatterers

The covariance estimation of non-Gaussian and nonstationary InSAR samples has not been addressed so far. Therefore, a covariance estimator that is robust against non-Gaussian samples and invariant with respect to non-stationary phase was developed. The proposed robust

covariance estimator requires no flattening of the interferometric phase, and performs well even on extremely heavily tailed distributions.

4 Summary of the Work

The objectives of the thesis are addressed in five full paper peer-reviewed journal articles. This chapter gives a brief summary of the articles including the following main contributions.

- ◇ First demonstration of the integration of PSI and TomoSAR for operational processing, which delivered a TomoSAR reconstruction of an entire city using meter-resolution SAR data.
- ◇ Computationally efficient fusion of two cross-heading TomoSAR point clouds.
- ◇ First demonstration of semantic fusion of optical image and TomoSAR of an urban area.
- ◇ First systematic proposal of robust estimators for multipass InSAR techniques, which includes a robust estimator that can handle large phase offsets in the observations, as well as a robust covariance estimator that is invariant to the non-stationary phase.

4.1 Computationally Efficient Tomographic SAR Inversion

Appendix A explains the detailed procedure of combining PSI and TomoSAR for computationally efficient tomographic SAR inversion. It delivers the tomographic reconstruction of an entire city using meter-resolution spaceborne SAR data. The reduction of computational cost credits to the following three points:

(1) Pre-classification of single and double scatterers using non-parametric methods without any full TomoSAR inversion.
(2) Use PS periodogram to solve the detected single scatterers, and pass only the detected double scatterers to TomoSAR algorithms, e.g. SVD-Wiener and SL1MMER.
(3) Use estimates from PSI as the prior for the follow on TomoSAR inversion.

The right subfigure of Figure 4.1 shows the flowchart of the method in Appendix A. The most important is the layer right after the PSI processing, where pixels are pre-classified to no- (H_0), single-(H_1), or double-scatterer (H_2). The left figure shows the algorithm of pre-classification of the scatterers. Firstly, the pixels without scatterer are separated using penalized likelihood ratio test (Wang et al., 2014):

$$\frac{f(\mathbf{g}|H_0)}{f(\mathbf{g}|H_1)\exp(-AIC)}\frac{P_0}{P_{1,2}}\overset{H_0}{>}1 \tag{4.1}$$

where H_0 and H_1 likelihood of each pixel are compared, taking into account the Akaike information criterion (AIC) penalty (Zhu and Bamler, 2010b), and P_0, $P_{1,2}$ are the prior probabilities of pixels with no scatterer and with single or double scatterer(s), respectively. At this stage, H_1 likelihood is used for pixels with single and double scatterers. But the detection of single and double scatterer(s) is almost 100% correct with 0% false alarm rate, because the H_1 likelihood of the pixels with double scatterers are still much higher than their H_0 likelihood.

4.1.1 Pre-classification of Single and Double Scatterers

The pre-classification of pixels with single or double scatterer(s) is carried out using a few easily accessible quantities associated with each pixel. They are the amplitude dispersion index (ADI), the likelihood to single scatterer model (H_1), and the pixel mean intensity with the following three reasons:

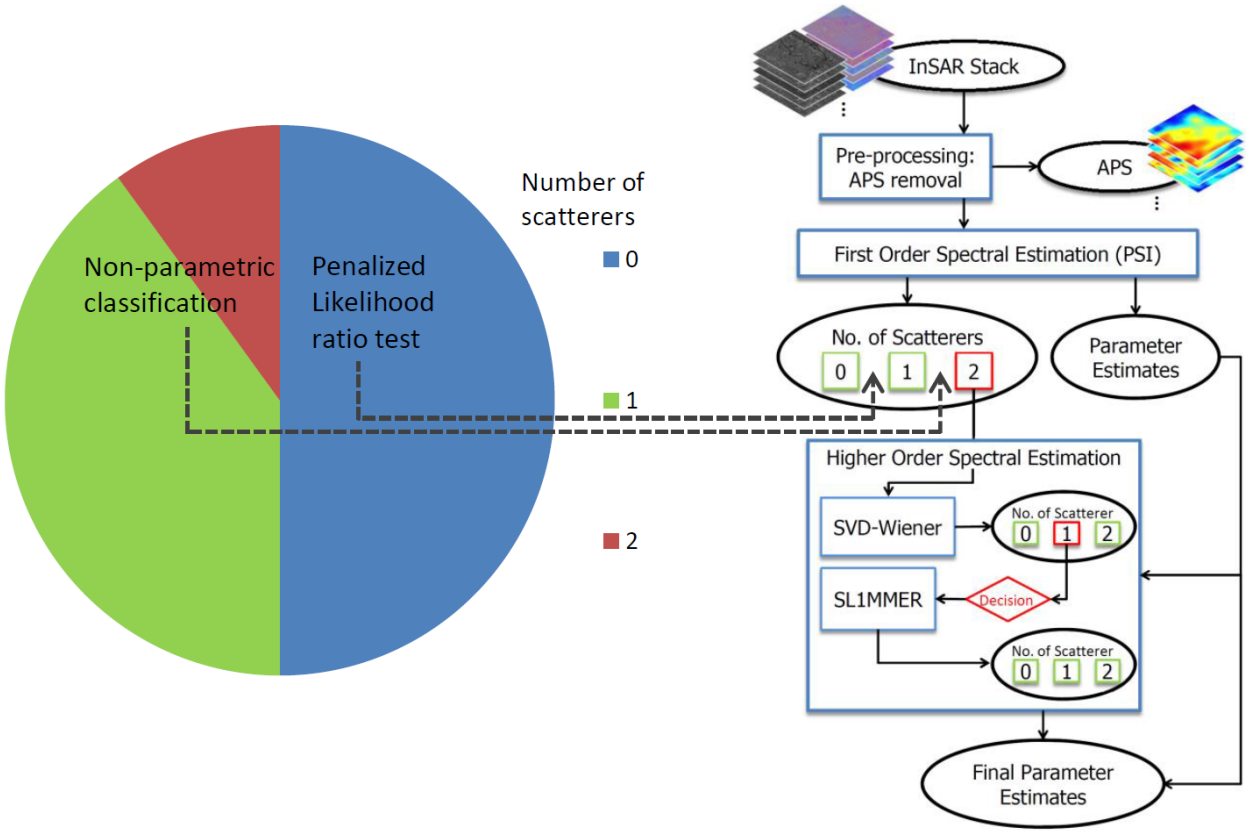


Fig. 4.1. left: Algorithm of pre-classification of non-, single-, and double-scatterers. Non-scatterers are detected by penalized likelihood ratio test. Single and double scatterers are classified using fast non-parametric methods without TomoSAR inversion, e.g. classification using mean intensity, amplitude dispersion index, etc. And right: the flowchart of the efficient TomoSAR inversion algorithm. The most important is the layer right after the PSI processing, where pixels are pre-classified, and PSI estimates are passed to TomoSAR processing as prior knowledge.

(1) pixels with single scatterer experience lower ADI, since the amplitudes of pixels with only a single dominant scatterer do not show pronounced fluctuations as a function of baseline;
(2) the H_1 likelihood tends to be lower if multiple scatterers are present in one resolution cell;
(3) the abovementioned two aspects are based on a reasonable SNR range. Therefore, only relatively bright pixels should be selected as candidates of pixels with double scatterers.

As an example, Figure 4.2 shows the joint distribution of mean intensity and the H_1 log-likelihood of single and double scatterers, respectively. Here, we must note that the real ground truth of single and double scatterers classification is not obtainable. The single and double scatterers in the following plots refer to those detectable ones from the estimation point of view. In Figure 4.2, they are detected by the CS-based SL1MMER algorithm (Zhu and Bamler, 2010a, 2012), which is one of the state-of-the-art algorithms. Based on these, it can be observed that the H_1 log-likelihood (represented by normalized residual sum of squares (RSS)) of the single scatterers shows more linear relationship with the pixel intensity, whereas they are more independent for double scatterers. Double scatterers are in general 20dB brighter than single scatterers. Different behaviour can also be observed in the joint distributions of the other quantities among the three.

Figure 4.3(a) shows the pre-classification result of single and double scatterers using the three quantities, with red representing pixels with double scatterers, green for single scatterers, and blue for no scatterer. A linear support vector machine (SVM) was employed as the classifier. For comparison, Figure 4.3(b) shows the classification result from the CS-based method SL1MMER (Zhu and Bamler, 2010a, 2012) as a reference, which is one of the

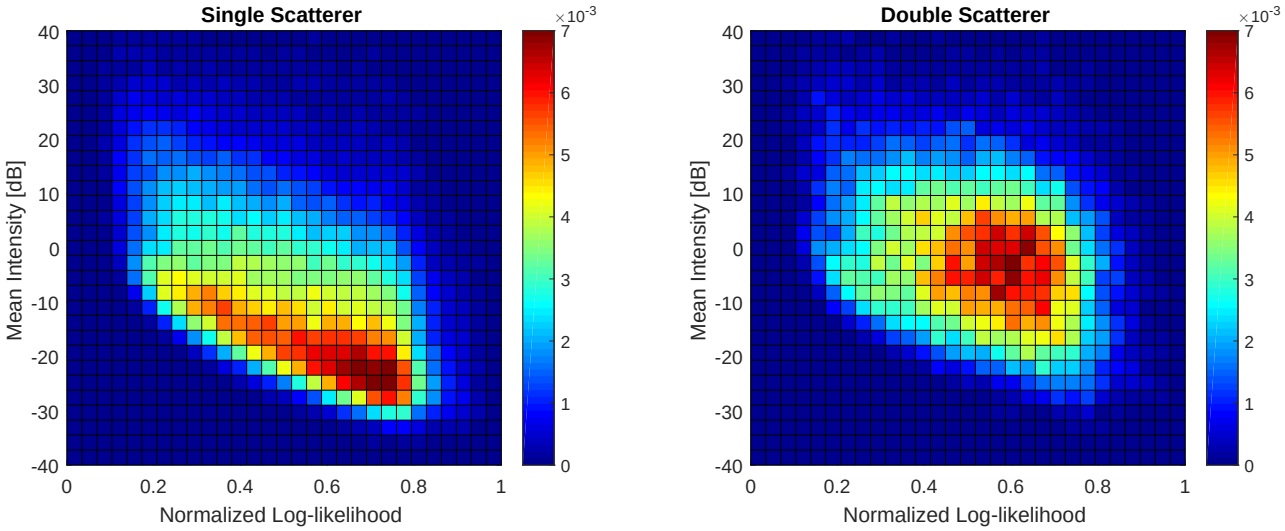


Fig. 4.2. The joint distribution densities of mean intensity and the log-likelihood to the single scatterer model for pixels with left: single scatterer, and right: double scatterer. The log-likelihood is represented by the residual sum of squares normalized to [0 1].

best single and double scatterer discriminations one can achieve so far. Figure 4.3(c) is the final classification after the higher order spectral estimation.

In Figure 4.3(a), 47% of the pixels are discarded as no-scatterers, 36% are detected as single scatterers, and the rest 17% are detected as double scatterers. Taking the SL1MMER result as the reference, the classification of pixel with no scatterer is almost 100% correct, and the detection rate of pixels with double scatterers is 72.4%, with a false alarm rate of 25.7% which is why much more pixels are detected as double scatterers. However, higher detection rate (consequently higher false alarm rate) is still favoured in the processing, because the falsely detected double scatterers can still be corrected in the higher order spectral estimation, but not for the undetected ones. Depending on the processing power, the threshold can be shifted towards higher detection rate, affording higher false alarm rate, and therefore more processing effort by more frequently applying the SVD-Wiener and the SL1MMER.

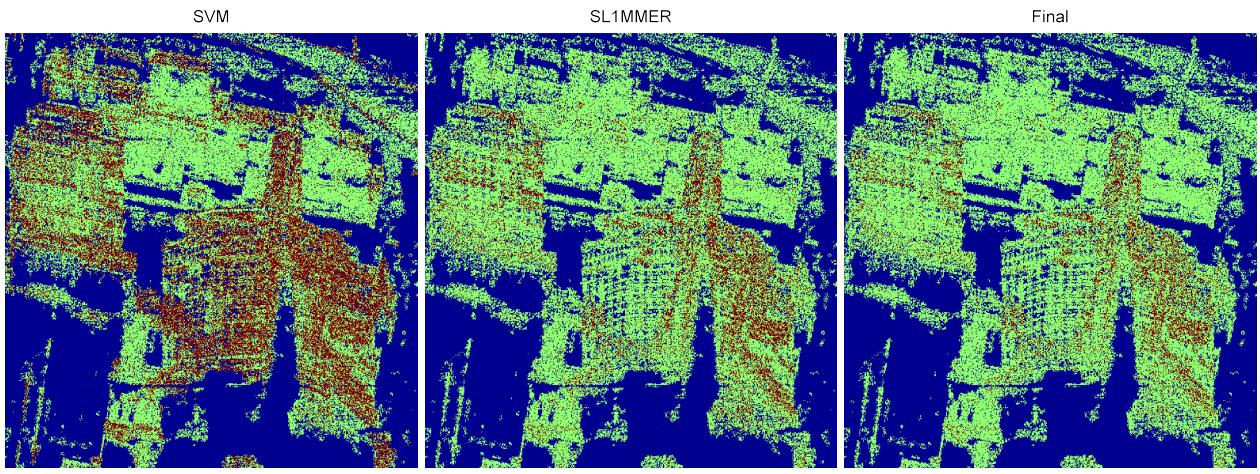


Fig. 4.3. (a) the result of the pre-classification method, (b) the result from the SL1MMER algorithm as a reference, and (c) the final classification after the higher order spectral estimation. The red pixels are detected double scatterers, the green are the single scatterers, and the blue are no scatterer.

Table 3. Overall computational cost of the proposed TomoSAR inversion method (using the test area shown in Figure 4.3 as the example), compared to using SL1MMER or SVD-Wiener only.

Processing Schemes	Percentage of pixels processed by each algorithm			
	Periodogram	SVD-Wiener	SL1MMER (full elevation search range)	SL1MMER (60m elevation search range)
100% SL1MMER (w/o prior)	—	—	100%	—
100% SVD-Wiener (w/o prior)	—	100%	—	—
Proposed method	100%	17%	—	9% (w/o prior)

4.1.2 Computational Cost Analysis

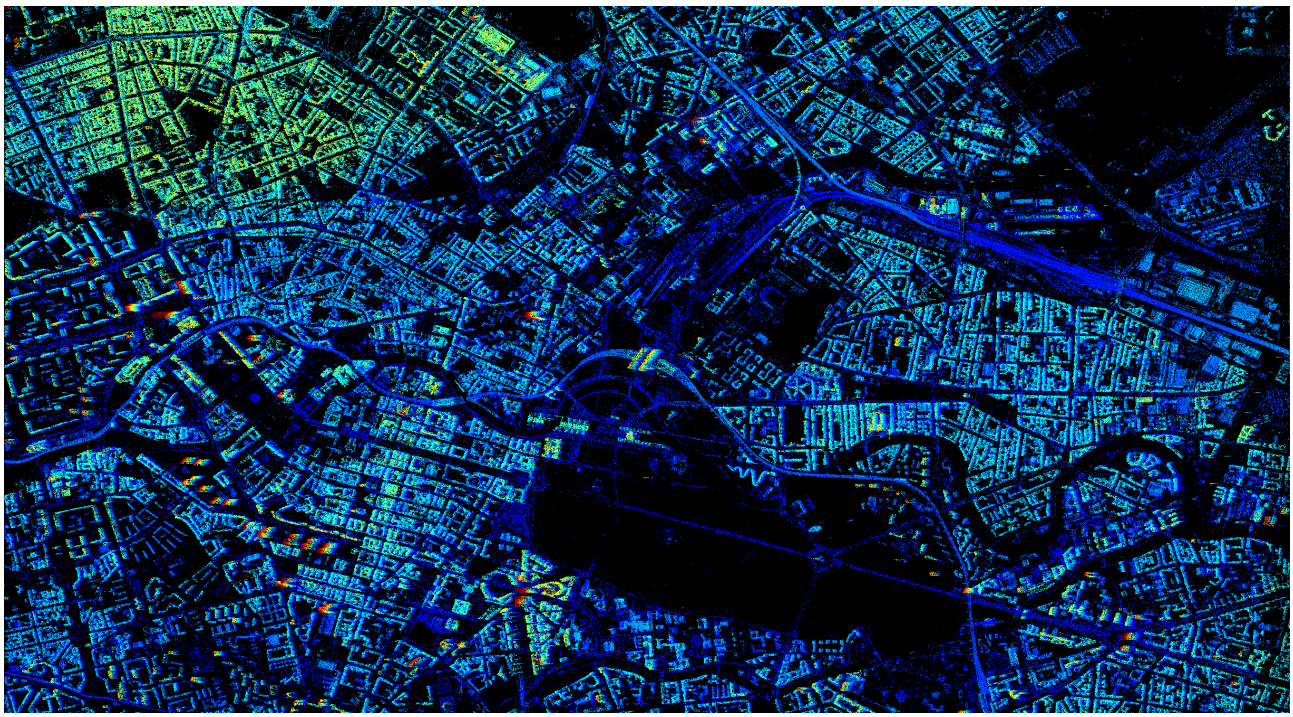
Compare the PS periodogram and SVD-based TomoSAR MAP, the periodogram consists of multiplication of the measurement vector with the modelled phase, and repeats at different elevation positions. Therefore, the computational complexity is $O(NL)$, where N is the number of images (usually 20 - 50) and L is the discretization level in the elevation direction (usually 50 - 200 for TS-X). In the higher order spectral estimation, the SVD-Wiener needs at least $O\left(N^2+NL\right)$ according to the equation (11) in (Zhu and Bamler, 2010b). The CS-based TomoSAR algorithm requires an L_1-L_2 norm optimization which is extremely computationally costly, e.g. the well-known Basic Gradient Projection for Sparse Reconstruction (GPSR-Basic) solver (Figueiredo et al., 2007) requires at least $O\left(KML_s^2\right)$, where K is the number of iterations (approximately 20), M is the number of multiplication of a specific matrix in each iteration (modest, usually several), and L_s is the elevation discretization level required for sparse reconstruction, which is usually much higher than L. Our experience is the SL1MMER algorithm is several hundred times slower than SVD-Wiener.

When prior knowledge of the elevation span is available, the L can be restricted to a specific elevation range, and hence reduce the computational cost. For instance, assuming a uniform distribution of building heights prior over the processing area, the total processing cost can be reduced to half for SVD-Wiener when $L \gg N$, and to one third for SL1MMER. Therefore, it is always more economic to obtain the prior through some low-cost methods before the higher order spectral estimation.

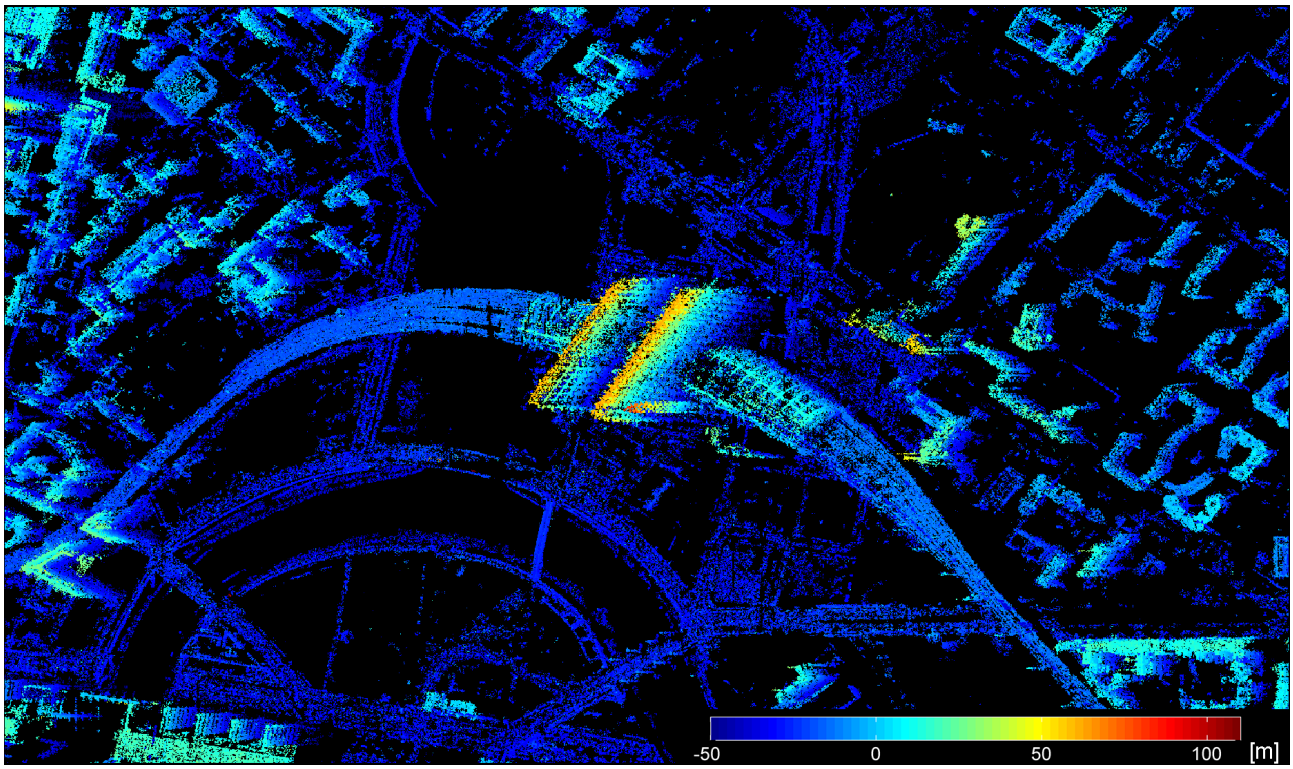
The overall computational cost also depends on the percentage of pixels allocated to each algorithm. As an example, for the test area discussed, 100% of the pixels are first processed by periodogram, 17% are then passed to SVD-Wiener, and only 9% are finally processed by SL1MMER. This small percentage together with the restriction of the search space leads to an overall speed-up of about 50 times compared to SL1MMER-only processing. Table 3 compares the overall computational cost of the integrated inversion approach and the conventional TomoSAR-only inversion.

4.1.3 Tomographic Reconstruction of Entire City

The abovementioned approach is applied on several stacks of TS-X high resolution spotlight images over Berlin. The following figure is the elevation estimates and a close-up view of one of the stacks which comprises 94 images. The proposed processing chain is able to catch most of the structure.



(a)



(b)

Fig. 4.4. (a) Elevation estimates of Berlin using the proposed approach, and (b) a close-up over the Berlin central station.

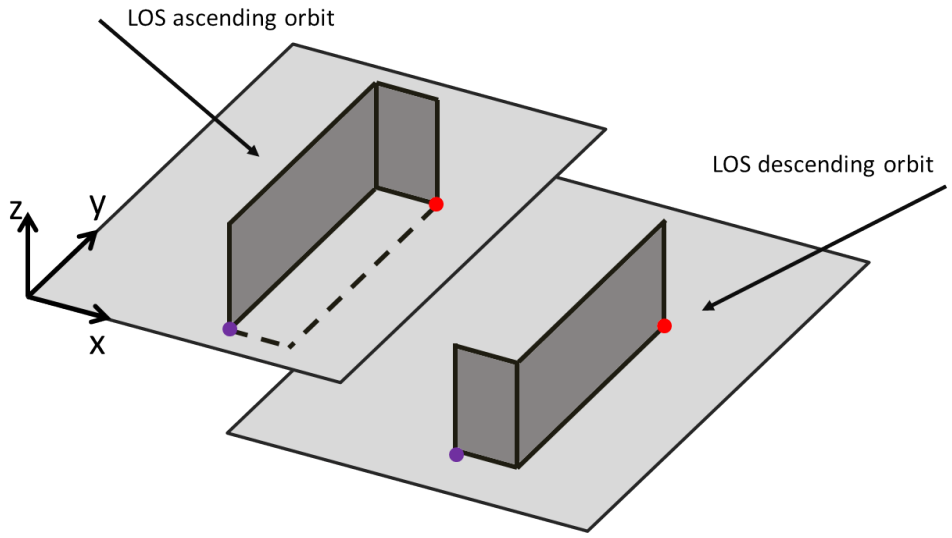


Fig. 4.5. Fusion geometry of the proposed L-shape method. The two building façades are coloured in dark grey representing the InSAR point clouds from cross-heading orbits whose line of sight directions are indicated by the two arrows. The light grey planes denote the corresponding ground level of the two point clouds. The two pairs of red and purple dots are the façade end point position to be estimated. Due to various facts, e.g. material, decorrelation, imaging geometry, etc., these end points may not be exactly from the scatterers located on the true end position of the façades, but they are still the closest match to the exact point correspondence in two cross-heading TomoSAR point clouds.

4.2 Automatic Feature-based TomoSAR Point Cloud Fusion

The previous section provides a solution to retrieve the 3-D TomoSAR point cloud of an entire city. A natural follow-up step is to combine several 3-D point clouds of the same area, especially for two cross-heading TomoSAR point clouds, i.e. derived from image stacks of ascending and descending orbits, as a minimum of two TomoSAR point clouds from cross-heading orbits are required to achieve a complete monitoring over the whole area due to the SAR side-looking imaging geometry.

InSAR point clouds are relative measurements whose reference points are usually unknown. The fusion task is to estimate the unknown 3-D shift. However, challenge arises at no exact point correspondence can be directly found among the natural scatterers in two cross-heading InSAR point clouds.

The contribution of Appendix B is finding and matching the point correspondences in two cross-heading TomoSAR point clouds. It is not difficult to realize that the end positions of façades where the two point clouds converge are the closest match to the exact point correspondences. The proposed point cloud matching algorithm is named as ”L-shape detection and matching”, because building façades often appear as L-shapes in InSAR point cloud. This concept is shown in Figure 4.5. The end points of the L-shapes are indicated by the two pairs of red and purple dots. The two building façades coloured in dark grey represent the point clouds from ascending and descending orbits whose line of sight (LOS) directions are indicated by the two arrows, respectively. And the light grey planes denote the corresponding ground levels of the two point clouds.

Figure 4.6 shows the flowchart of the algorithm. Emphasis is put on the two most important procedures: point density estimation and L-shape detection which are marked as red in the figure.

4.2.1 Point Density Estimation

Because of the SAR side-looking geometry, TomoSAR retrieves dense façade points. The estimation of the point density is for detecting the façade points based on which the locations

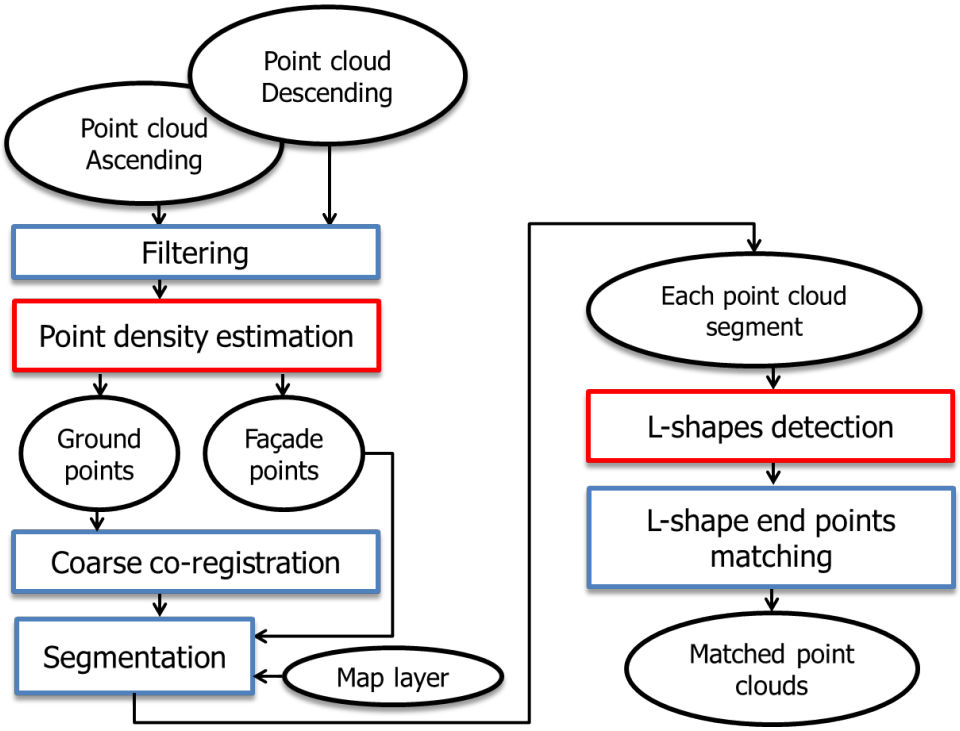


Fig. 4.6. Flowchart of the proposed algorithm. The black ellipses represent input/output data. The blue rectangles are procedures/processes.

of the façade end points are estimated. Therefore, the point density estimation is the most crucial step in the whole algorithm.

We used a directional window for the density estimation, considering that façades are usually vertical and linear. This idea is illustrated in Figure 4.7. It shows the 2-D top view of a point cloud, with the color indicating the point density, and the red dot being the target point whose point density is to be estimated. The estimation follows three steps: firstly, points within a larger area (the red dashed rectangle) centered at the target point are selected; then a straight line segment (the green dashed line) passing through the target point is fitted using a robust algorithm, e.g. random sample consensus (RANSAC), robust principle component analysis (PCA), etc.; and lastly, only the inliers (within the solid red rectangle) of the line segment are counted for the point density, instead of counting all the points in the black rectangle.

The benefit of using adaptive directional window for point density estimation can be clearly seen in Figure 4.8 where subfigures (a) and (b) are the density estimates using directional and rectangular windows, respectively. It is obvious that the directional window emphasizes the façade points, while keeping the point density of non-façade points unchanged. And hence, it improves the detection rate of the façade points. The final façade points are selected by thresholding on the point density.

4.2.2 L-shape End Point Estimation

Based on the detected façade points, a dominant L-shape is found for each building block by a weighted Hough transform (Wang and Zhu, 2015a). Since Hough transform only gives discretized estimates of the L-shape end points, their precise positions must be re-estimated.

The position of L-shape end point includes its horizontal 2-D coordinates and its height. The height is defined as the position where the façade intersects the ground plane, as depicted already in Figure 4.5.

The horizontal coordinates are estimated using a model-based approach. The procedures of

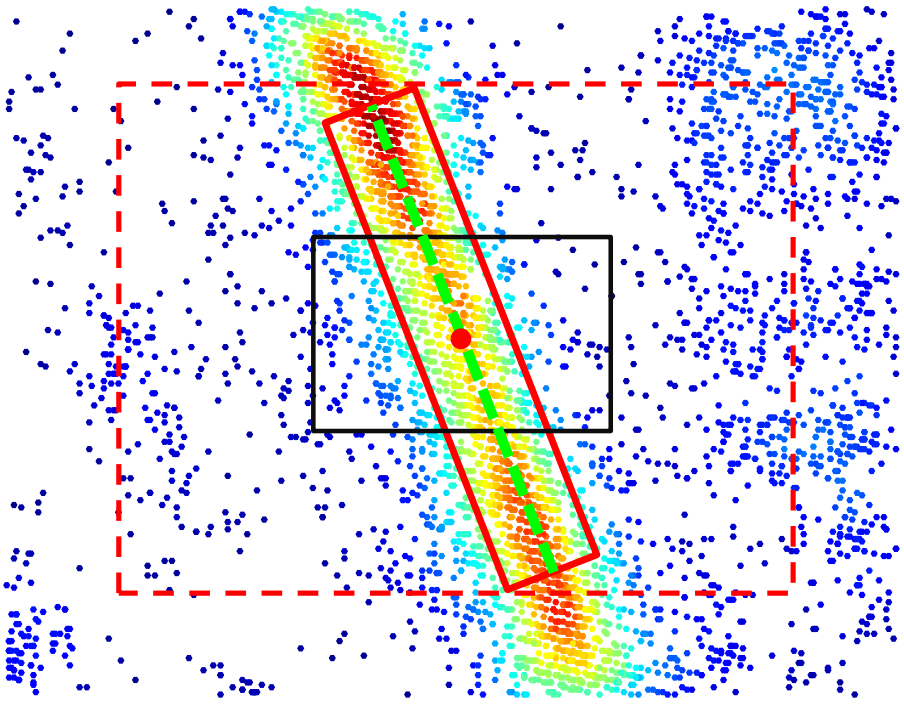


Fig. 4.7. Schematic drawing of point density estimation using directional window. The coloured points are the projection of the 3-D point cloud in horizontal plane, with the colour indicating the point density (blue: low density, red: high density). The green dashed line segment is a robust fit to the 2-D projection in the selected area within the red dashed rectangle. Points within the region bounded by the red solid rectangle are considered in the directional point density estimation, while those within the black rectangle are considered in the box-car density estimation.

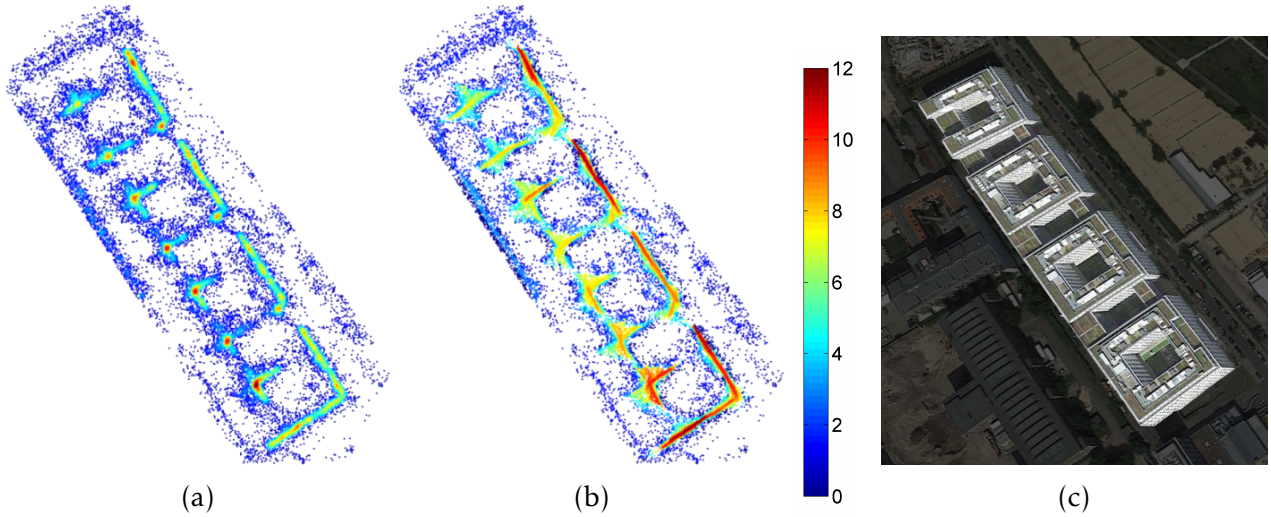


Fig. 4.8. Comparison of point density estimated using (a) directional window, and (b) rectangle window. The colour bar applies for both (a) and (b) with the unit $\left[\mathrm{m}^{-2}\right]$. For reference, (c) is the optical image of the same area from Google. The size of directional and rectangle windows, in terms of area, are identical in this comparison. The directional window emphasizes the façade point, meanwhile keeps the point density of non-façade points unchanged compared to rectangle window. It improves the detection rate of façade points.

the algorithm are briefly described as follows:

(1) The two line segments of the L-shape are separated and rotated according to the façade direction. For example, one line segment of the L-shape with inlier points should look like Figure 4.9(a) where the black line is the estimated façade direction from the weighted Hough transform. Figure 4.9(b) is the rotated coordinates of the points in Figure 4.9(a). The middle section with much higher density belongs to the façade part.
(2) An estimate of the façade point density is obtained by convolving it with a rectangular window along the façade direction. For example, the blue curve in Figure 4.9(c) shows

the point density estimate of the points shown in Figure 4.9(b).

(3) The point density is modelled as a rectangle function plus a constant offset. Hence, the estimated point density is the convolution of two rectangle windows, which is a trapezoid. Therefore, the goal is to robustly deconvolve the trapezoidal façade density curve, which is equivalent to estimate the location of the two slopes of the trapezoid.

(4) A moving line fitting method is developed to estimate the slopes' positions. A straight line is fitted to the points within a moving window with the same width as the rectangular window for density estimation. For any trapezoidal shape, high slopes should be detected at the position of the two slopes. By detecting the position of the two high slopes, one can determine the position of the two sides.

For robustness, the product of the absolute value of the slope and the number of inliers to the fitted line is considered in the detection of the two slopes, because the maximum number of inliers should be reached when the window centered at the left or the right slopes.

For example, Figure 4.9(d) shows the product of the absolute value of the slope and the number of inliers for the point density curve in Figure 4.9(c). Two prominent peaks can be detected. The positions of the peaks are the direct estimates of the center positions of the left and right position of the façade.

After obtaining the 2-D horizontal coordinates of the L-shape end point, we look for its neighbouring points on the ground level. The mean value of these points is an estimate of the height of the end point. A plane should be fitted to these points, if the local topography is not removed.

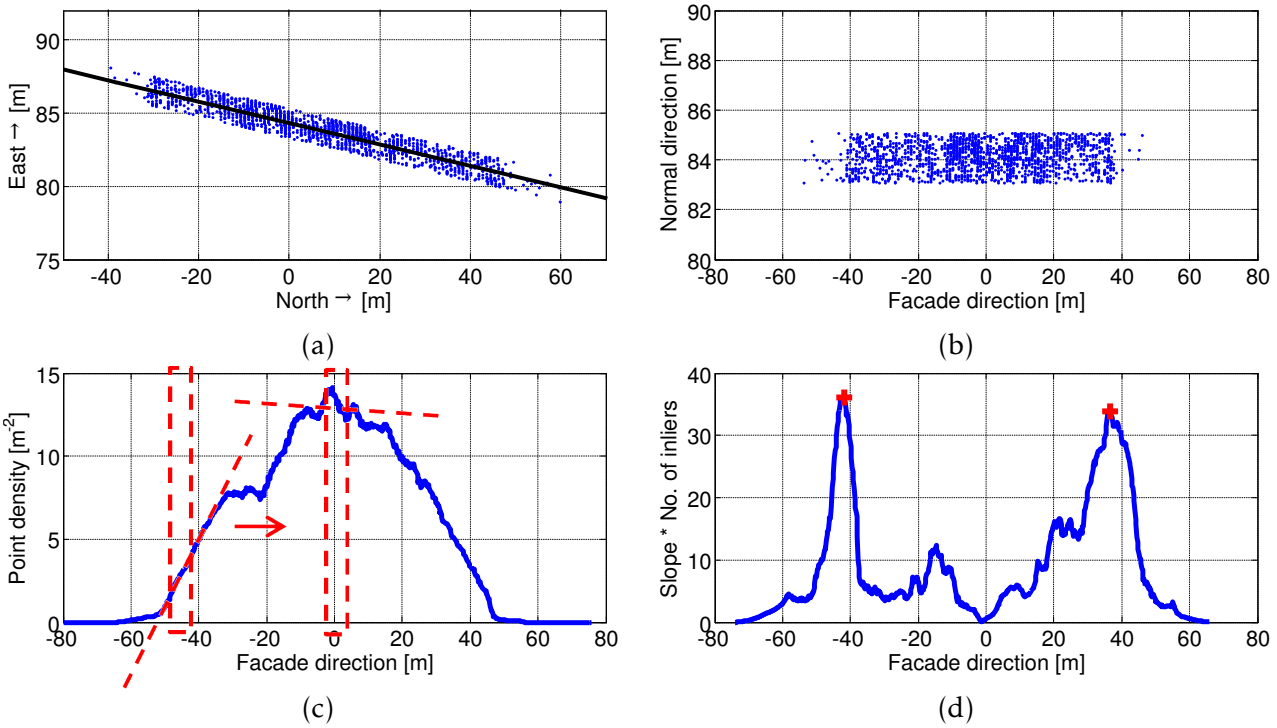
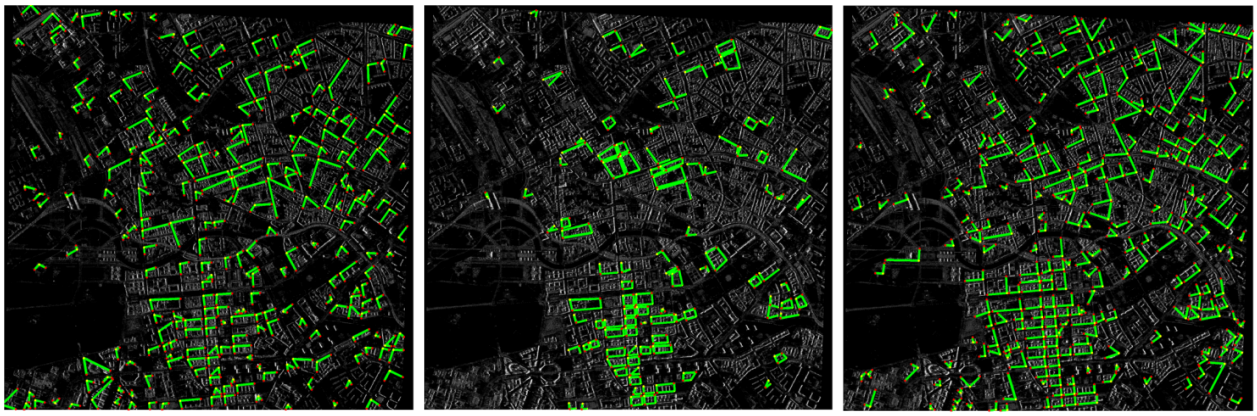


Fig. 4.9. (a) The inlier points of one line segment of an L-shape, with the black line being the façade direction obtained from the weighted Hough transform, (b) the coordinates of the inlier points projected to the façade direction, (c) the blue curve is the estimate of (b)'s point density using a rectangular window of 5m, and the red dashed rectangle and lines demonstrate the moving line fitting method, and (d) the product of the slope and the number of inliers of the fitted line using the points in the moving rectangle window along the point density estimate shown in (c). The two red crosses are the automatically detected façade start and end position at -41.38m and 37.22m, respectively.

4.2.3 Fusion Result

The proposed method was applied on two TomoSAR point clouds of Berlin contains about 20 million points each. Around 500 L-shapes are detected from each point cloud, corresponding to 1000 end points pairs. Figures 4.10(a) and (b) show the detected L-shapes and end points of the two point clouds, overlaying on their grey scale point density images, respectively. The matched L-shapes end points are shown in Figure 4.10(c). Closed quadrilateral implies both pairs of end points in an L-shape are matched. A single L-shape implies only one pair of end points in the L-shape is matched. In total, 150 pairs of end points are matched, which is sufficient to robustly estimate the 3-D shift and the scaling.

This result is compared to the one computed using the method explained in (Gernhardt and Bamler, 2012). A difference of 0.51m in east direction; 0.09m in north direction and almost no difference in height is found. A 3-D view of the first fused TomoSAR point clouds of an entire city is shown in Figure 4.11. The height of the points is color-coded.



(a) (b) (c)

Fig. 4.10. (a) and (b): The detected L-shapes and their end points of the two input point clouds over Berlin, and (c) the matched L-shapes and end points. The L-shapes are plotted in green, and the end points are marked as red. They are overlaid on the grey scale point density image.

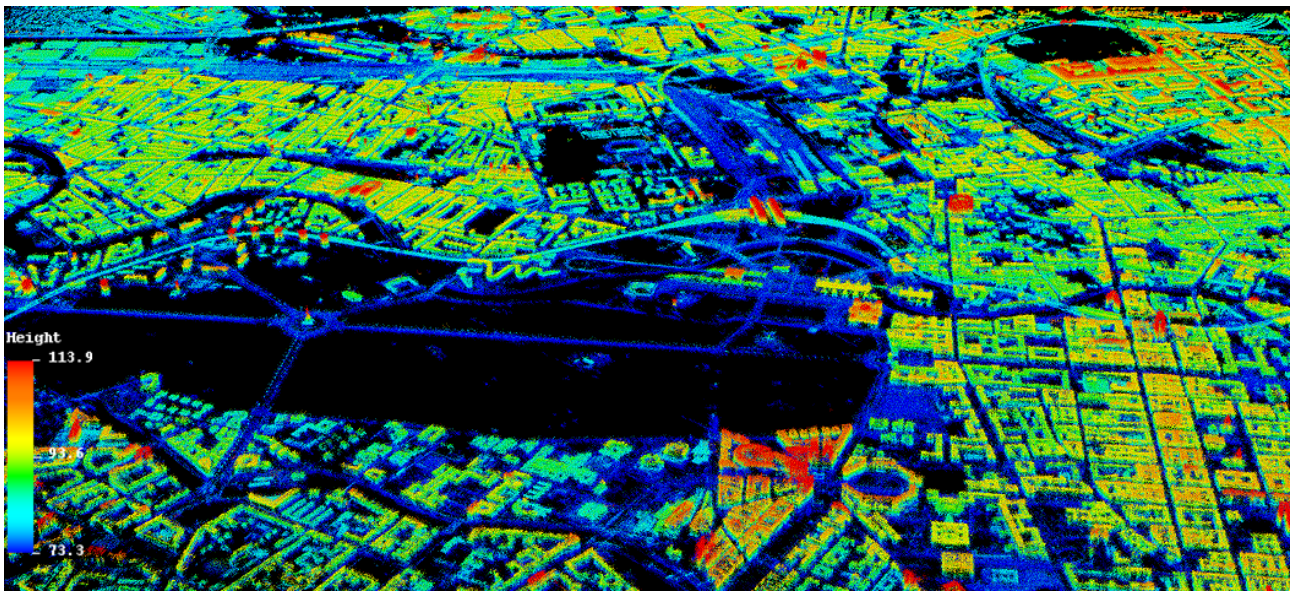


Fig. 4.11. The first fused TomoSAR point cloud of an entire city. The result combines an ascending and a descending point cloud of Berlin. The color indicates height on WGS84 reference surface. The unit of the colobar is meter.

4.3 Semantic Interpretation of InSAR Estimates Using Optical Images

Having reconstructed the 3-D InSAR point cloud of an urban area, the scientific question of the origin of the scatterers in the real world arises. By fusing InSAR point clouds and optical images, we can trace the scatterers in the optical images. The interpretability of optical images also complements the InSAR deformation monitoring, permitting a systematic monitoring of urban areas by the semantics of infrastructures, e.g. monitoring of bridges, roads and railways. The contribution of Appendix C is a method of semantic interpretation of InSAR point cloud using optical images.

The basic idea is by geometrically 3-D co-registering optical images to an InSAR point cloud, we can transfer the semantics derived from the optical image to the InSAR point cloud for further analysis. The flowchart of the algorithm is as follows.

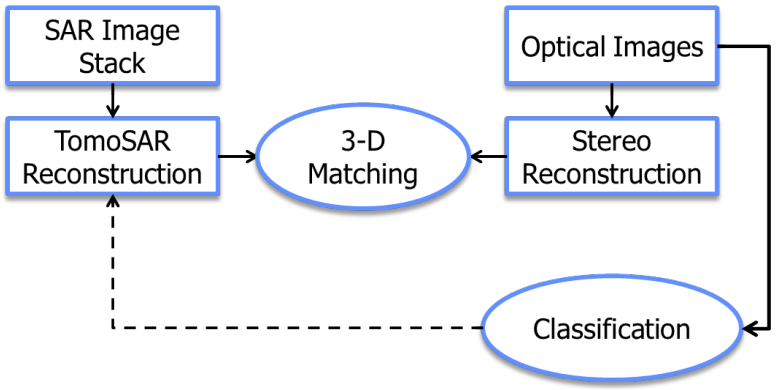


Fig. 4.12. Flowchart of the semantic InSAR point cloud interpretation method. The semantic classification of the InSAR point cloud is achieved by geometric co-registering the InSAR point cloud and the optical image to a reference model.

4.3.1 Geometric Co-registration of Optical Images and InSAR Point Cloud

The first contribution of the proposed algorithm is a strict 3-D geometric fusion of InSAR point cloud and optical images in urban area. SAR and optical images are different 2-D projections of a 3-D scene. Due to the complex urban topography, the co-registration must be done with a precise 3-D model. Figure 4.13 is a comparison of optical and SAR image of an urban area. A lot of details which are visible in the optical image are completely hidden underneath the strong reflection of the façade scatterers in the SAR image.

The proposed co-registration algorithm is as follows

(1) Reconstruct 3-D point cloud from two or more optical images using stereo matching with structure from motion (SfM) if necessary
(2) Reconstruct 3-D point cloud from SAR images using TomoSAR
(3) 3-D co-registration of the optical and InSAR point clouds
(4) Project optical image into the SAR geometry, a step also known as rendering

Since the first two steps are well explained in other literatures, the following text will focus on the co-registraion of optical and InSAR point cloud.

The aerial optical images are usually nadir-looking, in contrast to the side-looking geometry of SAR. In another word, façade points barely appear in optical point cloud while they are prominent in the TomoSAR point cloud. The point cloud derived from aerial optical images

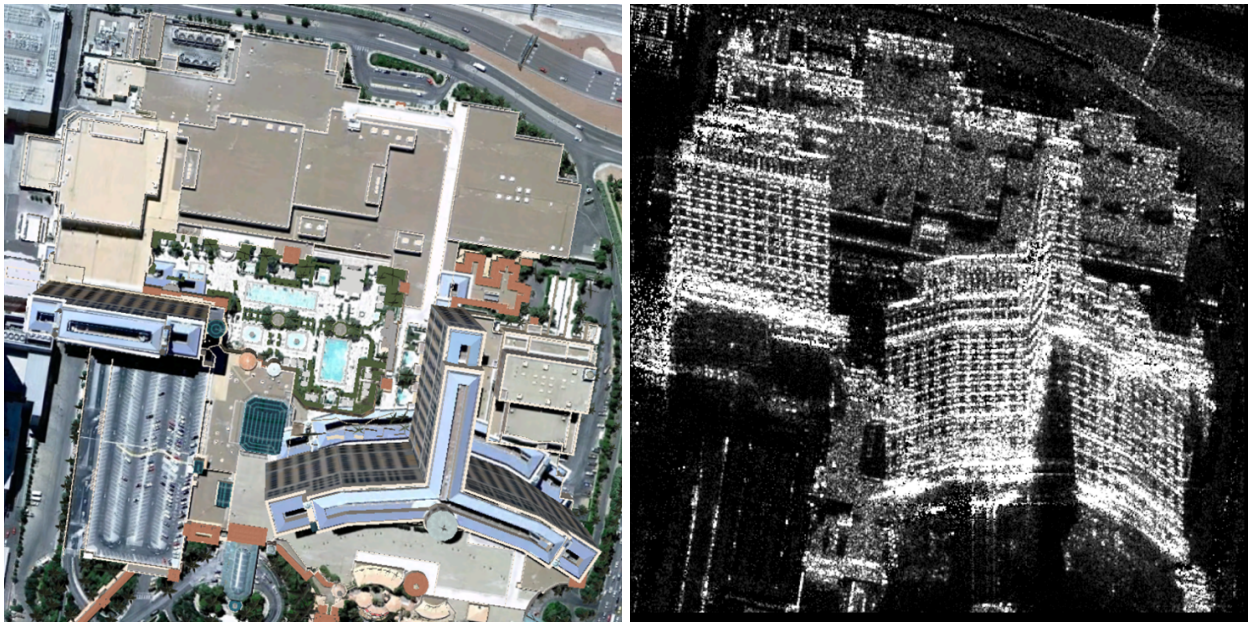
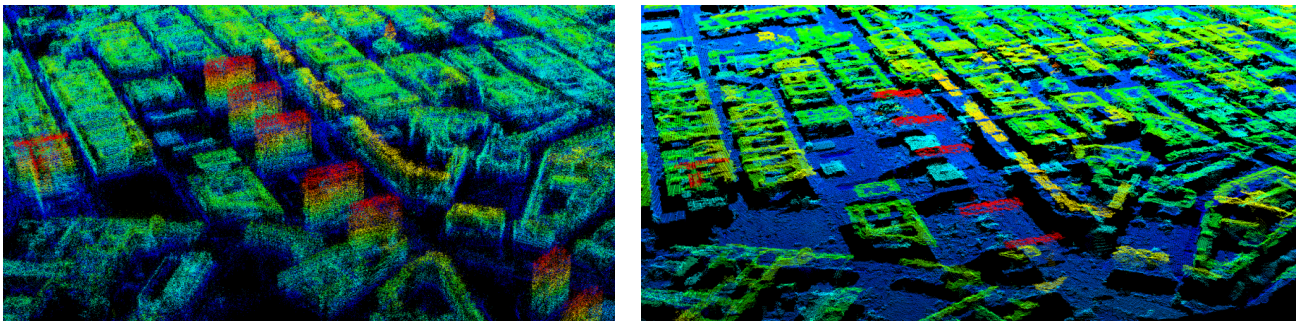


Fig. 4.13. Comparison of Google optical image (left) and TS-X high resolution spotlight image (right) of a same area. A lot of details visible in the optical image are completely hidden underneath the strong reflection of the façade scatterers in the SAR image. Images are from (Zhu and Bamler, 2012).

also in general has better relative positioning accuracy than the spaceborne TomoSAR point cloud. This difference is exemplified in Figure 4.14, where the left and right subfigures correspond to the TomoSAR and optical point clouds of the same area, respectively. These unique modalities have driven our algorithm to be developed in the following way:

(1) Edge extraction
  (a) The optical point cloud is rasterized into a 2-D height image.
  (b) The point density of TomoSAR point cloud is estimated on the rasterized 2-D grid.
  (c) The edges in the optical height image and the TomoSAR point density image are detected, respectively.
(2) Initial alignment
  (a) Horizontally by cross-correlating the two edge images.
  (b) Vertically by cross-correlating the height histogram of the two point clouds.
(3) Refined solution
  (a) The façade points in both point clouds are removed.
  (b) The final solution is obtained using *anisotropic* iterative closest point (ICP) with robustly estimated covariance matrix.



(a) (b)

Fig. 4.14. (a) the TomoSAR point cloud of high-rise buildings, and (b) the optical point cloud of the same area. Building façades are almost invisible in the optical point cloud, while they are prominent in the TomoSAR point cloud.

More detail of the co-registration algorithm can be found in Appendix C. The final co-registration result is demonstrated in Figure 4.15. The top subfigure shows the recon-

structed optical point cloud with colour representing the height, and the middle subfigure is the co-registered point cloud combining the optical one and two TomoSAR point clouds from ascending and descending viewing angles. Successful co-registration can be confirmed by the correct location of the façade points in the bottom of Figure 4.15 which is the top view of the fused point cloud with different colours representing different point clouds.

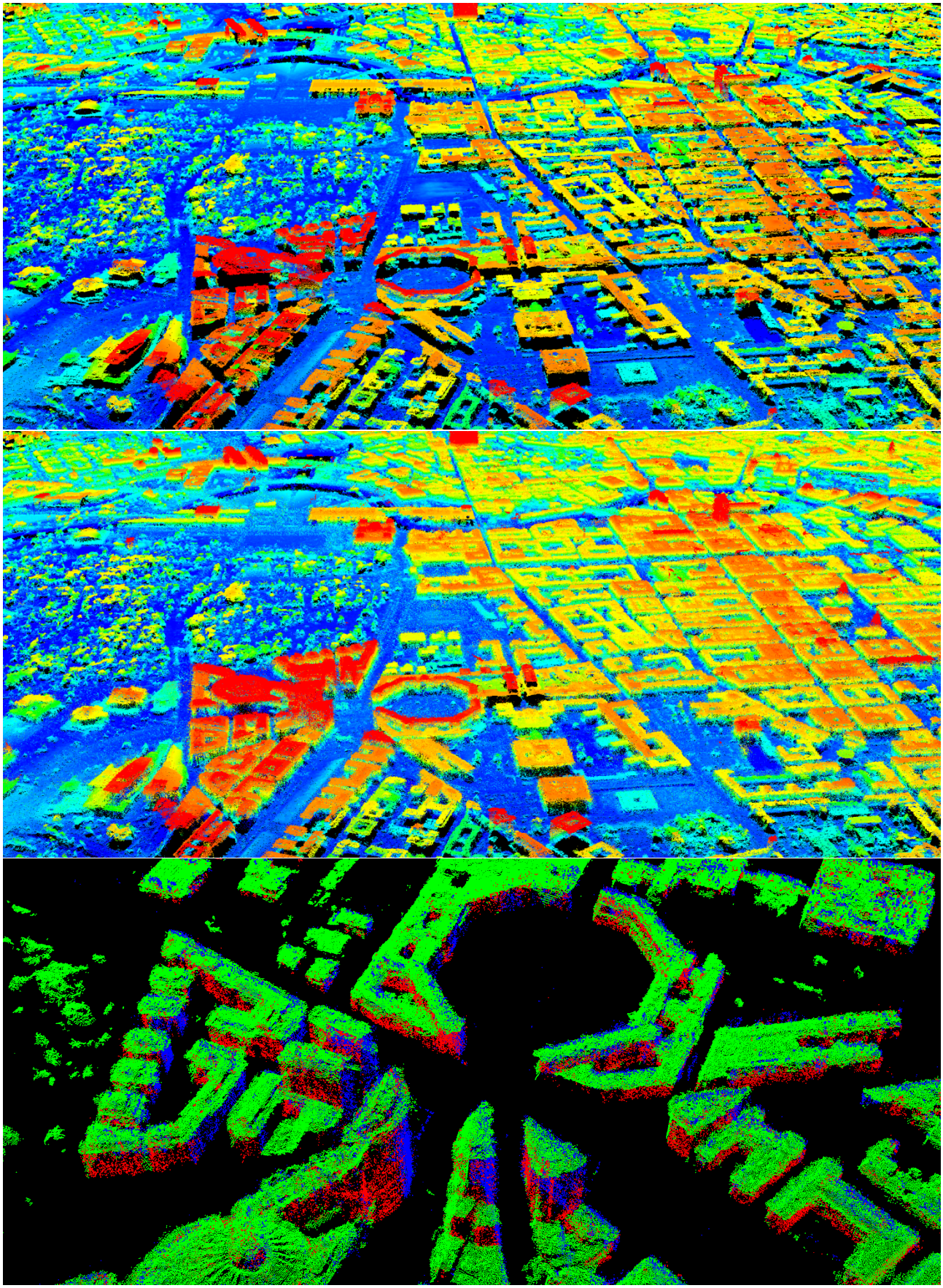


Fig. 4.15. Top: optical point cloud reconstructed using standard structure from motion algorithm with colour representing the height, middle: the fused point cloud combining the optical one and two TomoSAR point clouds from ascending and descending viewing angle, and bottom: a close up view of co-registered point cloud in Berlin Potsdamer Platz, where green, red, blue represent the points from optical, ascending and descending TomoSAR point cloud, respectively. The ground points were removed for a better visualization.

4.3.2 TomoSAR Point Cloud Rendering

Once the 3-D point clouds are co-registered, the relative positions of the optical images with respect to the SAR image are then known. The optical images can be projected into the SAR geometry. This is also known as *rendering*: given a viewpoint, solve the occlusion problem. However, points do not occlude each other except degenerated points, because they are dimensionless. Typically, the rendering of a point cloud requires reconstructing the 3-D surface and testing the depth of each point. But reconstructing a reliable 3-D surface from a spaceborne TomoSAR point cloud is still a challenge task, although there are some pioneer works on reconstructing façades and roofs from spaceborne TomoSAR points (Zhu and Shahzad, 2014; Shahzad and Zhu, 2015). Therefore, the rendering technique should avoid an explicit surface reconstruction.

This challenge has been addressed in Appendix D by an improved hidden point removal operator (a point-based rendering technique). The hidden point removal operator (HPR) (Katz et al., 2007) inverts the original point cloud using certain function, so that the visible points will lie on the convex hull of the inverted point cloud. The most common inverting function is the spherical flipping. Given a point $\mathbf{x}$ of any dimension, the *spherical flipping* function is defined as follows:

$$\hat{\mathbf{x}} = f(\mathbf{x}) = \mathbf{x} + 2(r - \|\mathbf{x}\|)\frac{\mathbf{x}}{\|\mathbf{x}\|} \tag{4.2}$$

where r is radius of a multi-dimensional sphere. The following figure illustrates the transformation function and HPR in 2-D. The blue dots are the original point cloud. And the red dots are the inverted one given a viewpoint $\boldsymbol{O}$. The convex hull of the inverted point cloud is computed and the points lying on the convex hull are marked as visible.

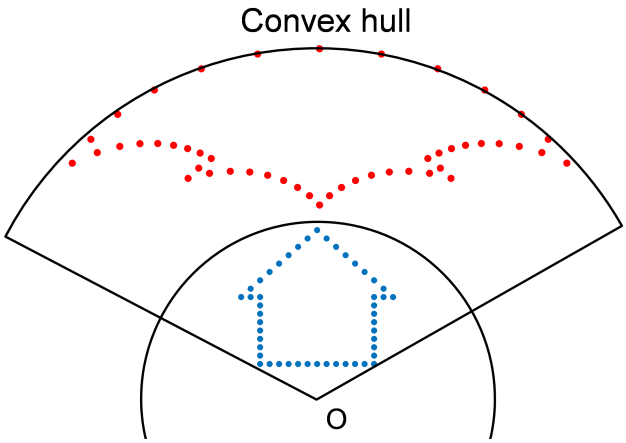


Fig. 4.16. Hidden point removal operator. The blue dots are the original point cloud. And the red dots are the inverted one given a viewpoint $\boldsymbol{O}$.

Considering the much larger and anisotropic noise in spaceborne TomoSAR point clouds than in typical computer graphics problems, the HPR is not directly applicable to TomoSAR point clouds. Inspired by (Mehra et al., 2010), we introduced a visibility probability which is a Gaussian function of the point-to-convex hull distance d:

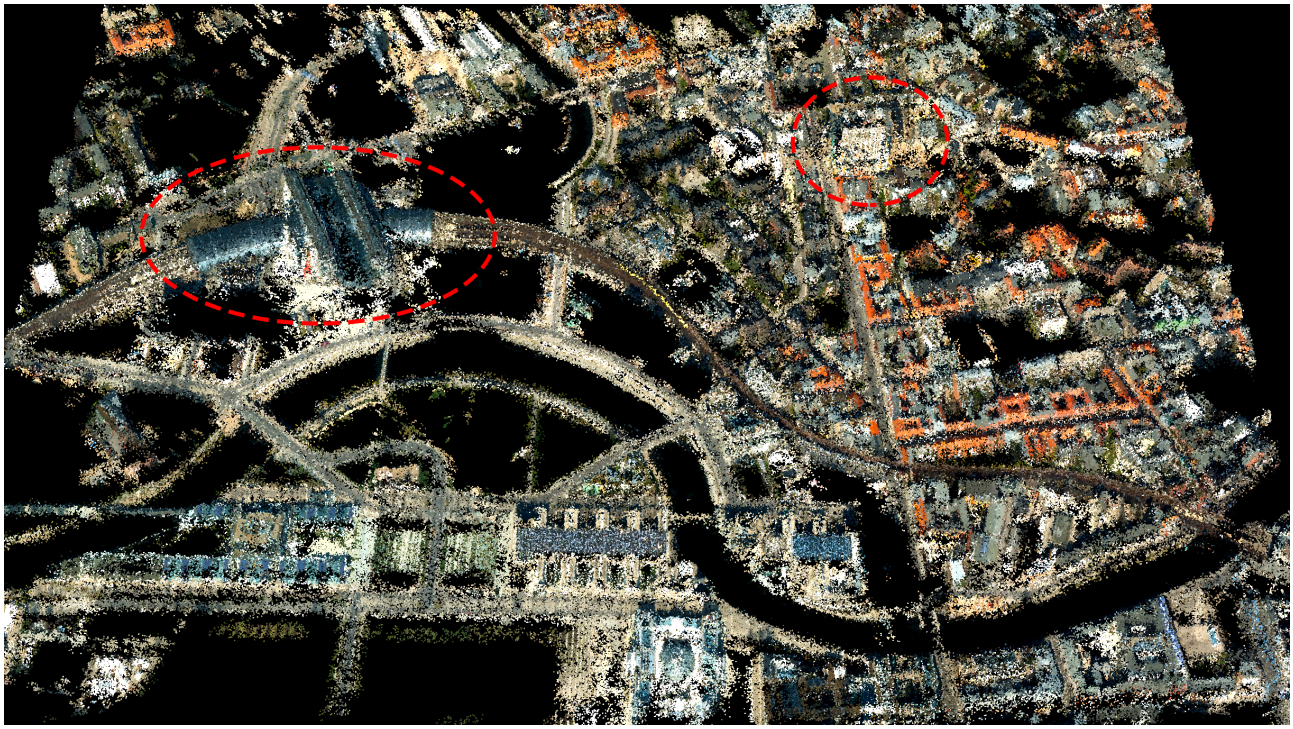
$$p(\hat{\mathbf{x}}) = \exp\left(-\frac{d^2}{2\sigma_{\hat{\mathbf{x}}}^2}\right) \tag{4.3}$$

where $\sigma_{\hat{\mathbf{x}}}$ is the norm of the total noise standard deviations (all three directions) in the inverted point cloud. $\sigma_{\hat{\mathbf{x}}}$ can be directly computed from the noise of the original point cloud.

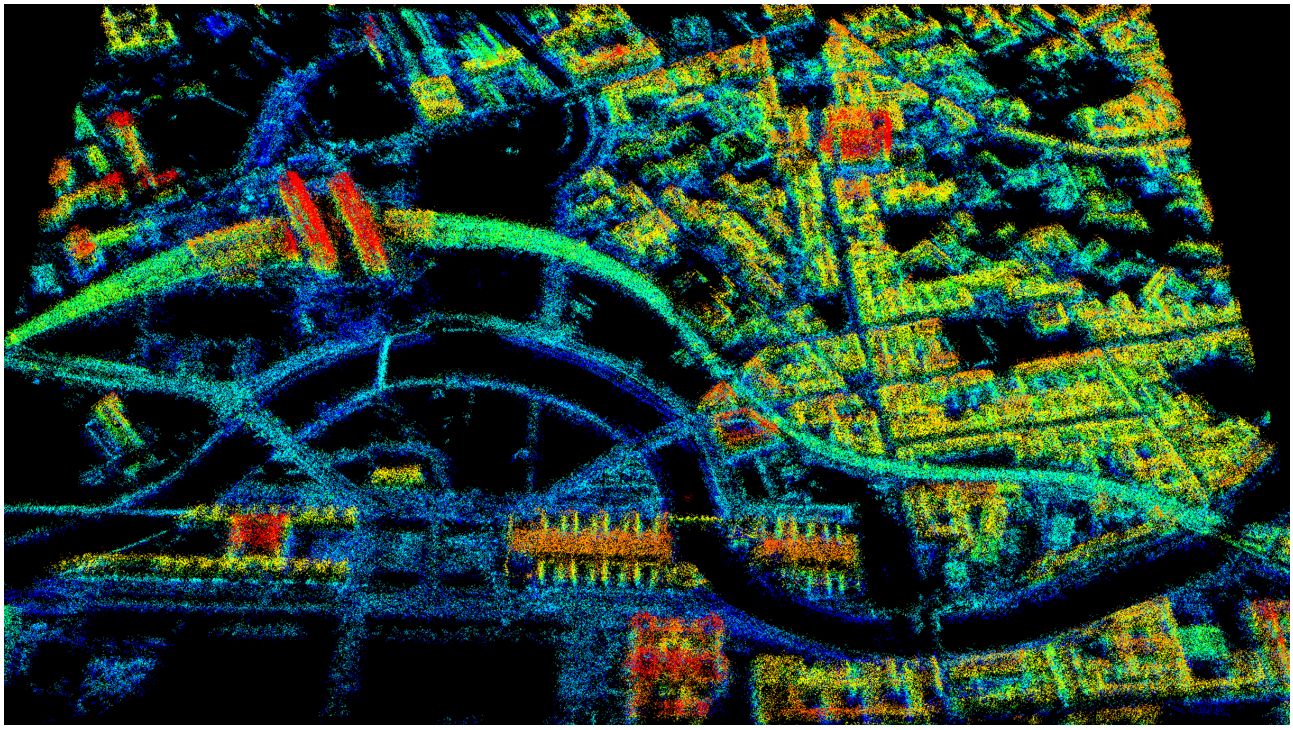
Apart from this, we also include an image source consistency check by weighted majority voting to achieve consistent texturing among neighbouring points in the TomoSAR point cloud.

As a result, Figure 4.17(a) shows the first spaceborne meter-resolution 3-D TomoSAR point cloud of an urban area textured with colour from the optical images. Where no TomoSAR point is available, the black background is filled. The subfigure (b) is the same TomoSAR point cloud, but colour-coded by its height. The dashed ellipses in subfigure (a) mark the Berlin central station and the main building of Universitätsmedizin Berlin. The former consists of some railway segments on the ground level and the station building of about 5 stories tall, and the latter is a high-rise building of about 20 stories tall. The close up views of these areas are shown in subfigures (c) and (d). The boundary of the railway and the building edges are clearly projected. For the more challenging task of texturing high-rise objects, the proposed algorithm also successfully captures the regular window raster on the façade of Universitätsmedizin Berlin. This implies a good estimation of the cameras parameters, the relative positions between the cameras and the SAR sensor, as well as a good estimation of the points visibilities to the different cameras.

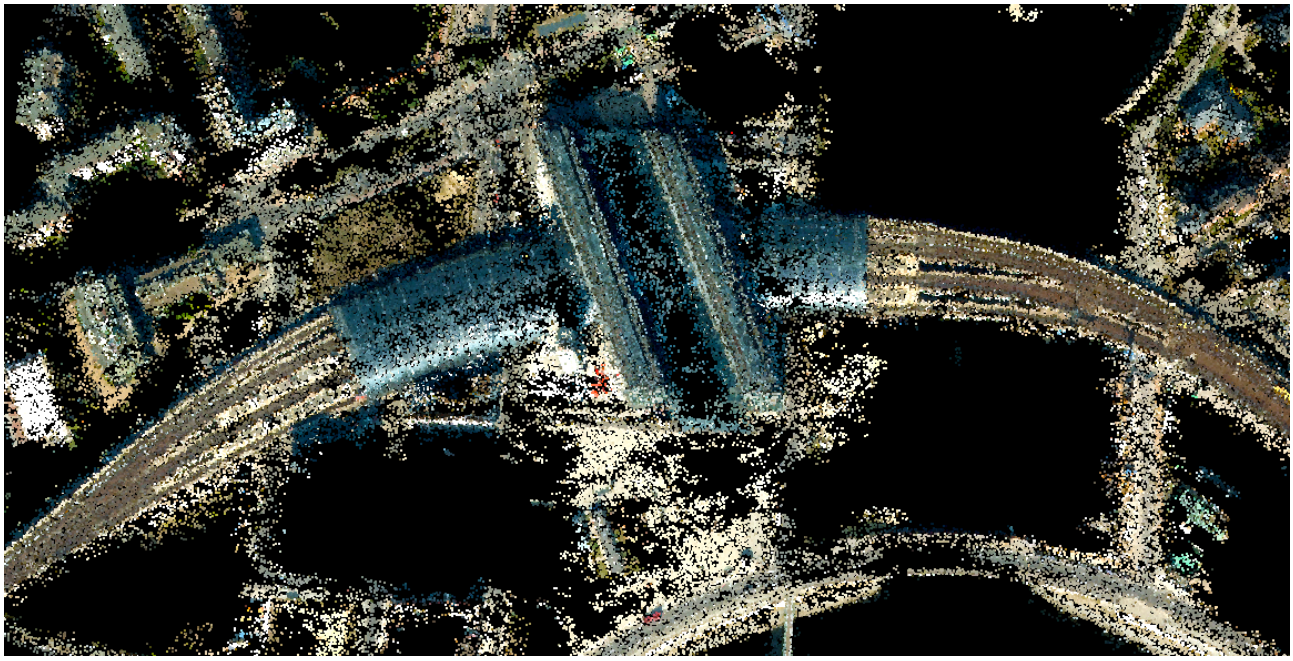
Of course, each point in the point cloud also contains its deformation parameters. By associating the deformation parameters with its optical appearance, we have an intuitive way of analysing the scattering mechanism of different objects. This will be exemplified in the next section.



(a)



(b)



(c)

(d)

Fig. 4.17. (a) The first spaceborne meter-resolution 3-D TomoSAR point cloud of urban area textured with colour from optical image, (b) the corresponding TomoSAR point cloud colour-coded by height, (c) and (d) zoom in to the Berlin central station and UniversitÃ¤tsmedizin Berlin marked in dashed red ellipses in (a), respectively.

4.3.3 Practical Demonstration

InSAR railway monitoring

If classification is also performed on the optical image, the semantic labels can be transferred to the InSAR point cloud for further analysis. The details of the image classification are not presented here, as this is not the focus of this thesis. One can use any classification method. For this study, the Bag of Word (BoW) classification method (Csurka et al., 2004) was employed for classifying the railway and river class of an area near Berlin central station for the purpose of railway and bridge monitoring. The following figure shows the classification map where the railway class and river class are labelled in green and red, respectively.

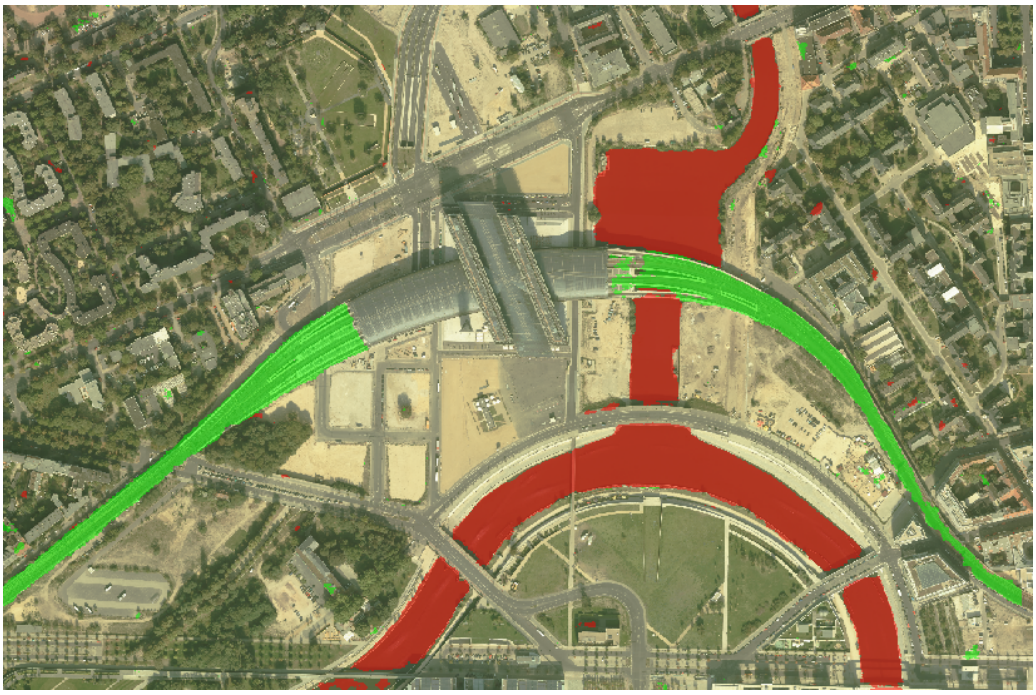


Fig. 4.18. River (red) and railway (green) classified using the BoW method.

Using the semantic label, the corresponding TomoSAR points can be extracted. Figure 4.19(a) shows the extracted continuous railway points overlaid on the optical image. The colour shows the amplitude of seasonal motion caused by thermal dilation. It can be easily observed that the railway deformation experiences certain discontinuity. We suspect it is because of the mechanical clearance between different railway segments. Therefore, we would like to detect and monitor these gaps.

However, the raw TomoSAR estimates are relatively noisy due to the relative low SNR of some scatterers. To filter out the inaccurate estimates, a special prior was employed, i.e. the thermal dilation of a steel beam is linearly proportional to its length at a first approximation (Kerr, 1978). That is to say, the railway deformation estimates are piecewise linear, and the scatterers on the same cross-section of the railway undergo similar deformation. Therefore, the deformation estimates can be filtered by minimizing its total generalized variation (with second order derivative which favours piecewise linear prior):

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}} \left\{ \frac{1}{2} \|\mathbf{f} - \mathbf{y}\|_2^2 + \lambda \|\Delta \mathbf{f}\|_1 \right\} \tag{4.4}$$

where $\mathbf{y}$ is the deformation estimates along the 1-D railway direction (i.e. stretch the railway to a straight line), $\mathbf{f}$ is the noise-free version of $\mathbf{y}$, and $\Delta\mathbf{f}$ is the second order derivative of $\mathbf{f}$. As shown in literatures of total generalized variation (Bredies et al., 2010; Knoll et al., 2011), the L_1 norm of second order derivative is convex and lower semi-continuous, one can solve it using convex optimization solvers.

The filtered deformation estimates are shown in Figure 4.19(b), in parallel with the unfiltered estimates. The piecewise linear deformation signal is well revealed. To have a quan-

titative comparison, the upper subfigure of Figure 4.20 plots the original (in blue) and the denoised (in red) deformation estimates as a function of the 1-D railway direction. The original estimates have a standard deviation about 1 - 2 mm which concealed some edges where the magnitude of deformation is low. The filtering successfully reconstructed these edges.

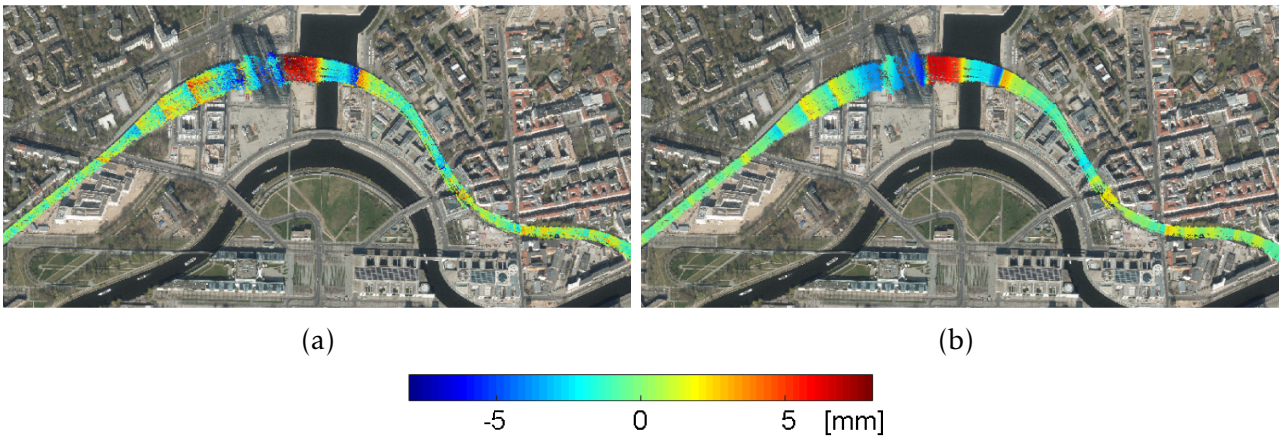


Fig. 4.19. (a) Railway points extracted from the original TomoSAR point cloud, and (b) the railway points filtered using total generalized variation. The colour shows the amplitude of seasonal motion due to the thermal expansion of the railway.

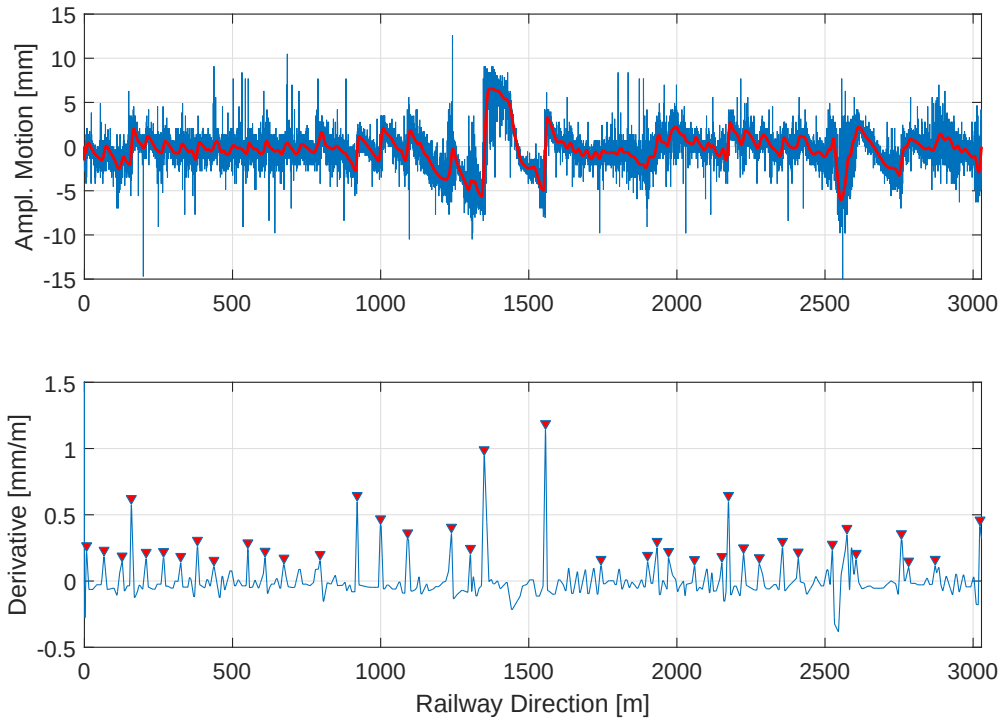


Fig. 4.20. Upper: the original and the filtered estimates of amplitude of seasonal deformation in blue and red, respectively, and lower: peaks detected in the derivative of the denoised deformation function along the railway direction.

By detecting the peaks in the derivative of the deformation function, the discontinuities in the deformation signal can be detected. Constraint was put on the minimum distance between two peaks. The detected peaks in the deformation's derivative can be seen in the lower subfigure of Figure 4.20. The positions of the peaks on the railway are shown as the green dots in Figure 4.21. Each green dot represents the midpoint of the cross-section. In the lower subfigure of 4.21, we provide the close up view of the two joints in the optical image. After carefully inspecting the optical images, we found that the several green dots close to the central station are indeed physical railway joints, while no physical mechanical joint is

observed for the others green dots. Hence, the real cause of the deformation discontinuity still needs to be investigated.

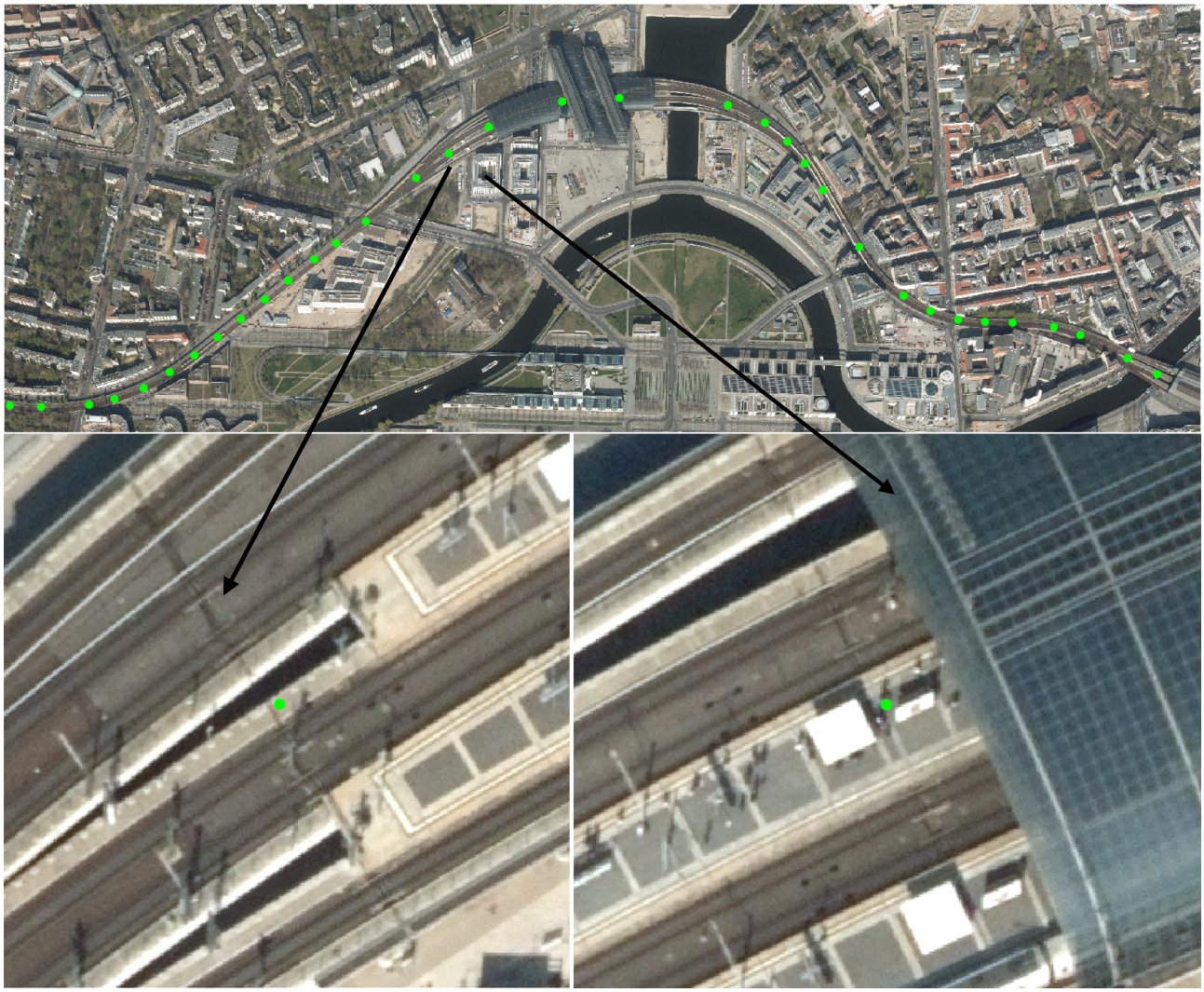


Fig. 4.21. Upper: the midpoint of the detected railway joint cross-section marked in green, and lower: close up view of the railway joint. The background optical image has a ground spacing of 20cm.

4.4 Robust Estimators for Multipass SAR Interferometry

Robust estimation has not been systematically addressed for multipass InSAR techniques. Appendix D and E address the robust estimation of two fundamental problems in the multipass InSAR algorithms. They are:

- ◇ covariance matrix estimation for non-Gaussian and nonstationary samples
- ◇ phase history parameters estimation for observations with large unmodelled phase errors.

The integrated new framework is named *robust InSAR optimization* (RIO). Table 4 summarizes the comparison of the state-of-the-art and RIO.

4.4.1 Robust Covariance Matrix Estimators

Non-Gaussian samples

The MLE for stationary CCG samples is the sample covariance matrix (2.12). RIO robustifies it using the M-estimator of the covariance which is basically an iterative reweighted sample

Table 4. Comparison of RIO framework with the state-of-the-art

	State-of-the-art	**RIO**
	Covariance estimation	
Stationary DS Nonstationary DS	Sample covariance matrix: $\hat{\mathbf{C}}_{MLE}$	M-estimator: $\hat{\mathbf{C}}_{M-est}$ Rank M-estimator: $\hat{\mathbf{C}}_{RME}$
	Phase history parameters estimation	
PS	Periodogram: $\hat{\theta}_{MLE}$	M-estimator: $\hat{\theta}_{M-est}$
Stationary DS	Periodogram on samples mean: $\hat{\theta}_{MLE}$	M-estimator on samples mean: $\hat{\theta}_{M-est}\left(\hat{\mathbf{C}}_{M-est}\right)$
Nonstationary DS	—	M-estimator pixel-wise inversion: $\hat{\theta}_{M-est}\left(\hat{\mathbf{C}}_{RME}\right)$

covariance matrix (Ollila and Koivunen, 2003a; Zoubir et al., 2012):

$$\hat{\mathbf{C}}_{k+1} = \frac{1}{M}\sum_{m=1}^{M} w\left(\mathbf{g}_m^H \hat{\mathbf{C}}_k^{-1} \mathbf{g}_m\right)\mathbf{g}_m \mathbf{g}_m^H \tag{4.5}$$

where m and k are the sample index and the iteration index, respectively, and $w(x)$ is a weighting function of the whitened squared residual $\mathbf{g}_m^H \hat{\mathbf{C}}_k^{-1} \mathbf{g}_m$. The weighting function down-weights highly deviated samples whose whitened residual is large in magnitude.

To derive a meaningful weighting function for some heavily tailed distributions, RIO models the non-Gaussian SAR pixel as a complex t-distribution (Tan, 1973; Kotz and Nadarajah, 2004):

$$f_{\mathbf{g}}(\mathbf{g}) = \frac{\Gamma(\nu + N)}{(\pi\nu)^N \Gamma(\nu) \det(\mathbf{C})^N}\left(1 + \frac{1}{\nu}\mathbf{g}^H \mathbf{C}^{-1} \mathbf{g}\right)^{-(\nu+N)} \tag{4.6}$$

where $\Gamma(\cdot)$ is the Gamma function, and $\nu \in \mathbb{R}^+$ is the DoF. The complex t-distribution is a realistic model for multipass SAR observations, as the DoF allows a quantitative definition of its deviation from CCG. This distribution approaches CCG as the DoF approaches $+\infty$, and becomes more heavily tailed as it decreases toward zero. In practical terms, this provides flexibility for handling different levels of outliers. The corresponding weighting function is derived according to equation (2.19):

$$w(x) = \frac{2N + \nu}{\nu + 2x} \tag{4.7}$$

For a good safeguard against outliers, a small value of ν (<5) is preferred (Ollila and Koivunen, 2003b). For the best safeguard against outliers, one can choose $\nu = 1$. This particular complex t-distribution with $\nu = 1$ is also known as the complex Cauchy distribution, which has been thoroughly investigated in many studies (Feller, 1971). Besides, (Schmitt and Stilla, 2014) chose $\nu = 2.5 - 5$ for centimeter resolution millimeter-wave SAR data of urban areas. This implies that the DoF is dependent on the object as well as the sensor. The simulation of different DoF is presented later in "quantitative comparison".

Equation (4.5) is solved iteratively. An approximation to drop the iteration is the sign covariance matrix (SCM) (Visuri et al., 2000; Croux et al., 2002). Extend it to complex number,

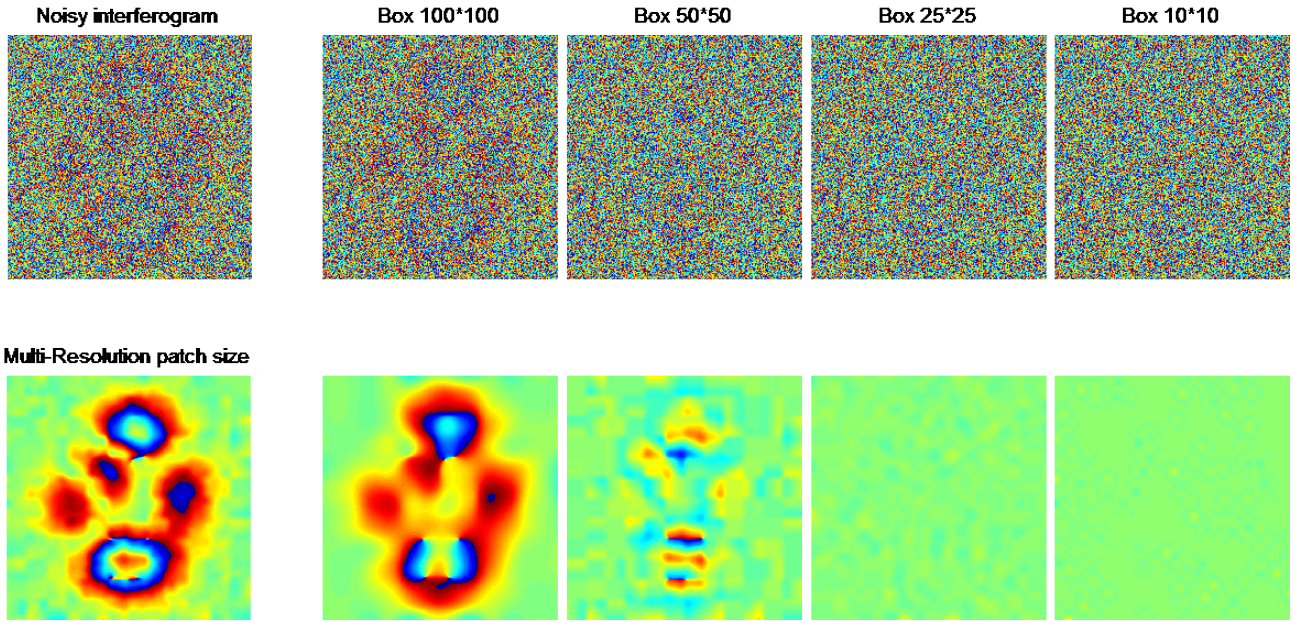


Fig. 4.22. Left: input noisy interferogram (upper), and the final estimated fringe (lower), and right: the input noisy interferogram (upper row) and estimated fringes (lower row) at each resolution level. The estimated fringes at each resolution level are subtracted from the noisy interferogram. The residual phase is taken as the input for the next resolution level with smaller window size.

it is

$$\hat{\mathbf{C}}_{SCM} = \frac{N\,\bar{I}}{M}\sum_{m=1}^{M}\|\mathbf{g}_m\|^{-2}\mathbf{g}_m\mathbf{g}_m^H \tag{4.8}$$

Non-Gaussian and nonstationary samples

To complete the theory, we must consider the situation where samples are neither complex Gaussian nor stationary. The nonstationarity we consider is the uncompensated deterministic part of the interferometric phase, e.g. topography, motion etc. In this case, a joint estimation of the expected interferometric phase and the covariance (De Zan and Rocca, 2005) is required, i.e.,

$$\hat{\mathbf{C}}_{MLE} = \frac{1}{M}\sum_{m=1}^{M}\hat{\mathbf{\Phi}}_m^H\mathbf{g}_m\mathbf{g}_m^H\hat{\mathbf{\Phi}}_m \tag{4.9}$$

where $\hat{\mathbf{\Phi}}_m$ is the complex diagonal matrix containing the estimated phase values of $\mathbf{g}_m$. Usually, the estimation of the interferometric phase is performed by spatial filtering. Its performance greatly depends on the filter itself, which can be very challenging in urban areas.

In Appendix E, a multi-resolution defringe algorithm was developed to tackle this challenge. Figure 4.22 is a demonstration of the defringe algorithm. The left two figures are the simulated interferogram with the coherence of 30% of the pixels equal 0.3 and the rest equal to 0 (up), and the final estimated interferometric phase (bottom), respectively. The right eight subfigures show the input and estimated interferometric phase at each resolution level. Starting from the largest patch size, the fringe frequency is estimated using only the coherent DSs in a patch. The interferogram is then demodulated, and passed to the next resolution level. In the test, four resolution levels were set: 100×100, 50×50, 25×25, and 10×10 pixels.

Nevertheless, the phase estimation requires additional effort, and poor estimates significantly affect the covariance matrix estimation. Therefore, a new covariance estimator — *Rank M-estimator* (RME) — for complex multivariate is proposed, which is invariant to the

multiplicative phase signals. The RME is derived by replacing the multivariate $\mathbf{g}$ with its rank $\mathbf{r}$ in equation (4.5):

$$\hat{\mathbf{C}}_{RME,k+1} = \frac{1}{M}\sum_{m=1}^{M} w\left(x_m\left(\hat{\mathbf{C}}_{RME,k}\right)\right)\hat{\mathbf{r}}_m\hat{\mathbf{r}}_m^H \tag{4.10}$$

The complex rank vector $\mathbf{r}$, being analogous to its real number version (Croux et al., 2002), is defined as follows:

$$\hat{\mathbf{r}}_m = \frac{1}{J}\sum_{j=1}^{J}\frac{\mathbf{g}_m \odot \mathbf{g}_j^*}{\left\|\mathbf{g}_m \odot \mathbf{g}_j^*\right\|} \tag{4.11}$$

where $\mathbf{g}_j$ is the direct neighbourhood of $\mathbf{g}_m$, and $\odot$ denotes the element-wise multiplication. The subtraction of neighbouring samples in the original definition is replaced by multiplication of the complex conjugate. The deterministic phase of $\mathbf{g}_m$ is reduced through this operation. The RME is a fourth-order descriptor of the sample statistics. It can be proven that the element-wise square root of $\left|\hat{\mathbf{C}}_{RME}\right|$ approaches $\left|\hat{\mathbf{C}}_{MLE}\right|$ asymptotically under CCG distribution when calculating the rank using one neighbourhood sample. Due to the multiplication operation, the variance of the square-root $\left|\hat{\mathbf{C}}_{RME}\right|$ estimate is doubled under a CCG distribution compared to $\left|\hat{\mathbf{C}}_{MLE}\right|$. Therefore, the number of samples should be doubled when using RME in order to keep a comparable variance as MLE. For an effective phase cancellation in equation (4.11), the neighbourhood should be kept small. We usually restrict $\mathbf{g}_j$ to be the only four adjacent pixels of $\mathbf{g}_m$. Hereafter, we assume that the element-wise square root on $\left|\hat{\mathbf{C}}_{RME}\right|$ has been performed.

Qualitative comparison

We compared the original MLE, the M-Estimator with complex t-distribution weighting, and the RME under three different scenarios: (1) multivariate CCG, (2) multivariate complex t-distribution with one DoF, and (3) nonstationary multivariate complex t-distribution with one DoF. The data are simulated to have ten acquisitions, with a neighbourhood of 1000 samples. Each ten-variate sample vector is simulated by de-whitening a ten-variate vector of zero-mean i.i.d. CCG distribution (for scenario 1) or complex t-distribution (for scenarios 2 and 3) with a predefined coherence matrix. In the last scenario, linear fringes with ten different fringe frequencies randomly picked within $[0\ \pi/10]$ were added to the phases of the ten acquisitions.

In addition, two different coherence matrices were predefined; one exponentially decaying and the other with a constant 0.5 coherence between acquisitions. The results are shown in Figure 4.23 where the left subfigure corresponds to the exponential coherence matrix, and the right corresponds to the constant matrix. Each row represents one of the aforementioned three scenarios, and each column represents one of the three covariance estimators. The top left subplots in both subfigures are regarded as their reference coherence matrices, respectively, because the MLE is the optimal estimator under CCG and is asymptotically unbiased.

All three estimators can preserve the correct shape of the covariance matrix under CCG. The MLE fails in the second scenario, where the samples are contaminated by outliers. The coherence is usually overestimated because of the large amplitude of the outliers. Nevertheless, the M-estimator and RME remain as correct estimations. In particular, this M-estimator is the MLE for the simulated data. Finally, the M-estimator is not capable of dealing with nonstationary samples. Heavy underestimation occurs because of the summation of the complex numbers with nonstationary phases. The degree of underestimation depends on the local fringe frequency and the spatial extend of the samples. The larger the fringe frequency and

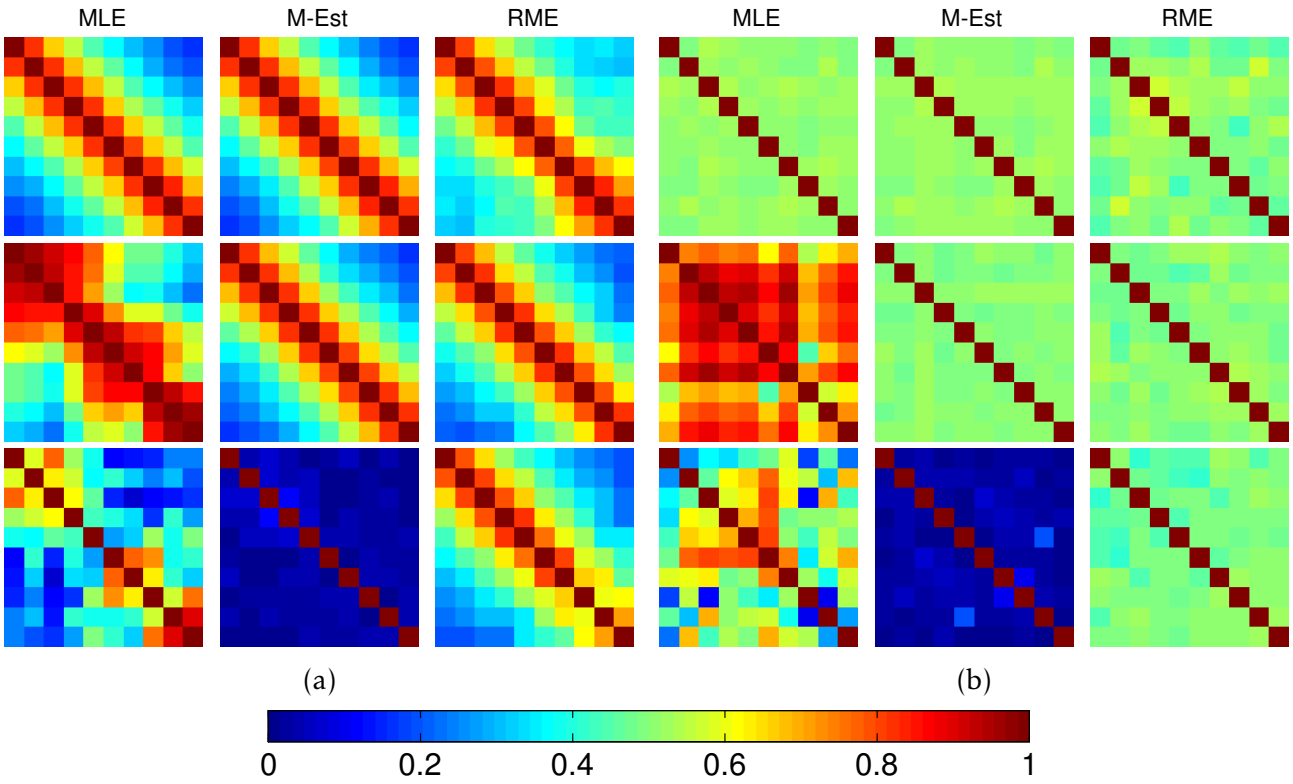


Fig. 4.23. Comparison of three covariance matrix estimators under three different observation cases; 1st row: complex circular Gaussian, 2nd row: complex t-distribution with one degree of freedom, and 3rd row: nonstationary complex t-distribution with one degree of freedom. (a) and (b) show the results of the exponential and constant coherence matrices, respectively. In both (a) and (b), 1st column: MLE (under Gaussian), 2nd column: M-estimator with t-distribution weighting, and 3rd column: rank M-estimator with t-distribution weighting.

the spatial extend of samples, the severer the underestimation. The RME is mean invariant, and maintains good performance in all conditions.

$\left|\hat{\mathbf{C}}_{RME}\right|$ shows higher bias at low coherence (as observed in Figure 4.23(a)), and also experiences higher variance (as observed in Figure 4.23(b)), due to the multiplication of two samples. This can be solved by using more samples in the estimation. The efficiencies of the M-estimator and RME are always lower than 100% under CCG distribution. This is a trade-off between robustness and efficiency.

Quantitative comparison

The abovementioned example demonstrates the effectiveness of the robust estimator at the worst case, i.e. complex Cauchy distribution whose variance is infinite. For a systematic analysis of the robust covariance estimator at different levels of outlier percentage, the DoF of the complex t-distribution was varied from 2 to 5. As it was mentioned before, complex t-distribution approaches complex Gaussian as the DoF increases. Therefore, the advantage of the robust estimator diminishes as the DoF increases. Figure 4.24 shows the variance (upper) and bias (lower) of coherence estimates. The left side is for data with complex t-distribution of DoF=2.5, and the right side is DoF=5. The advantage of robust estimator is much less at DoF=5 than DoF=2.5, because the distribution of the former is more Gaussian. In this experiment, the robust estimator always outperforms the Gaussian MLE, because the robust weighting function is t-distribution weight. Should some empirical robust function be employed, the Gaussian MLE will outperform the robust estimator at certain point as the DoF increases. Since we are particularly interested in scatterers of low coherence, the variance and bias as a function of the DoF for true coherence = 0.2 were also plotted in Figure 4.25. As we can see, the performance of the robust estimator is very stable for different level of outliers.

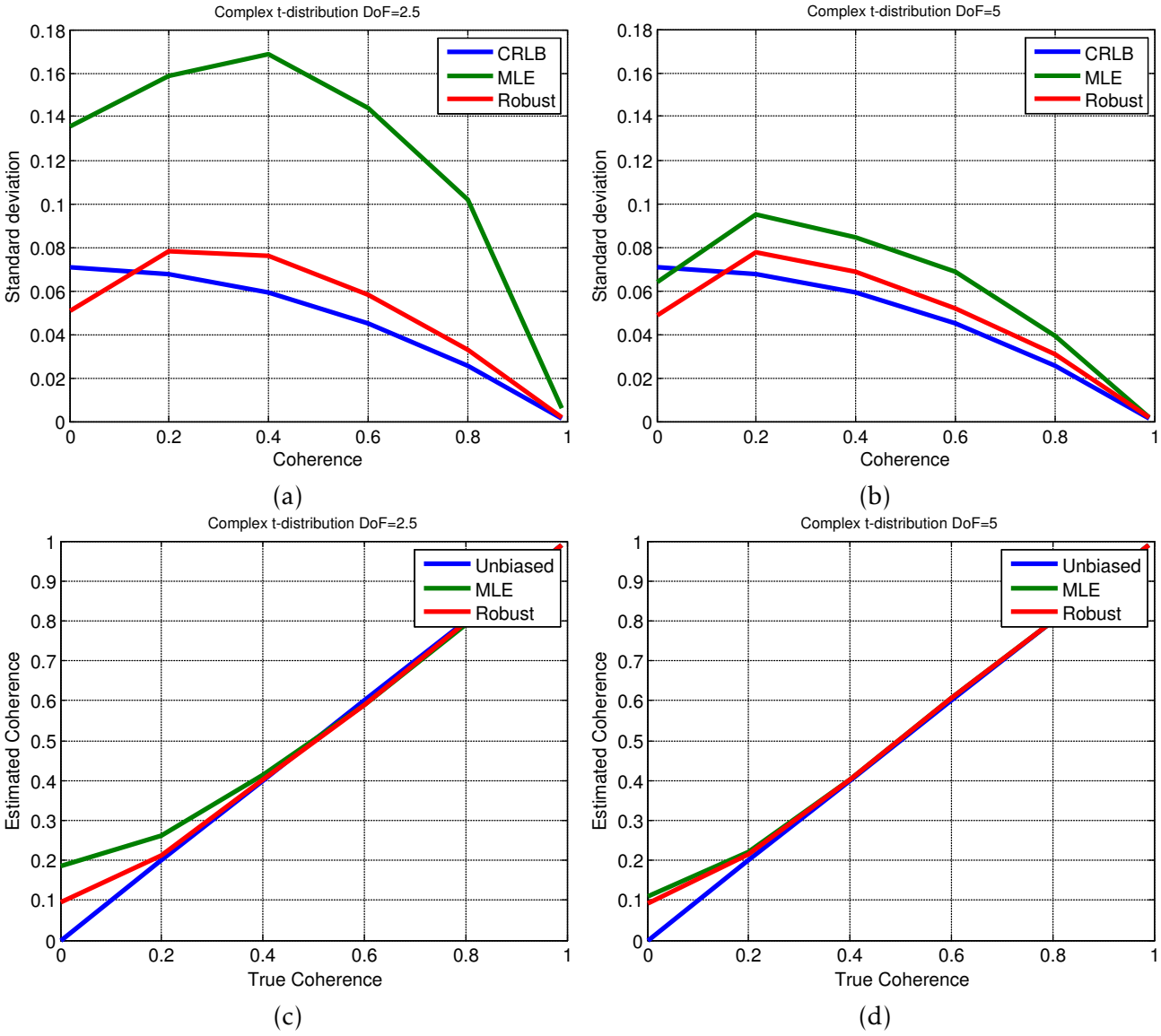


Fig. 4.24. (a) and (b): standard deviation of the coherence estimates of Gaussian MLE and the proposed robust estimator for complex t-distributed data with degree of freedom of 2.5 and 5, respectively, (c) and (d): bias of the coherence estimates. 100 looks is used in the coherence estimation.

4.4.2 Generalized Rank Covariance Estimator

The abovementioned covariance RME is a robust estimator based on the first order phase derivative. It can be generalized to any order, so that the estimator will be invariant to any spatially continuous interferometric phase signal.

Consider a SLC observation g_j with image index j, its first order basic rank operator $\mathrm{R}^1[\cdot]$ is defined as the multiplication of itself with the complex conjugate of its neighbour pixel:

$$r_j = \mathrm{R}^1\left[g_j\right] = g_j g_{j+1}^* \tag{4.12}$$

Assuming g_j being a Gaussian scatterer, and g_j being independent from g_{j-1}, the PDF of r is the product of two identical and independent CCG distribution. It is equivalent to the PDF of interferometric pixel with zero coherence, which was derived to be:

$$f\left(|r|,\phi\right) = \frac{2|r|}{\pi\bar{I}^2} K_0\left(\frac{2|r|}{\bar{I}}\right) \tag{4.13}$$

which does not depends on the phase ϕ of r. Given a coherence of γ between two single look

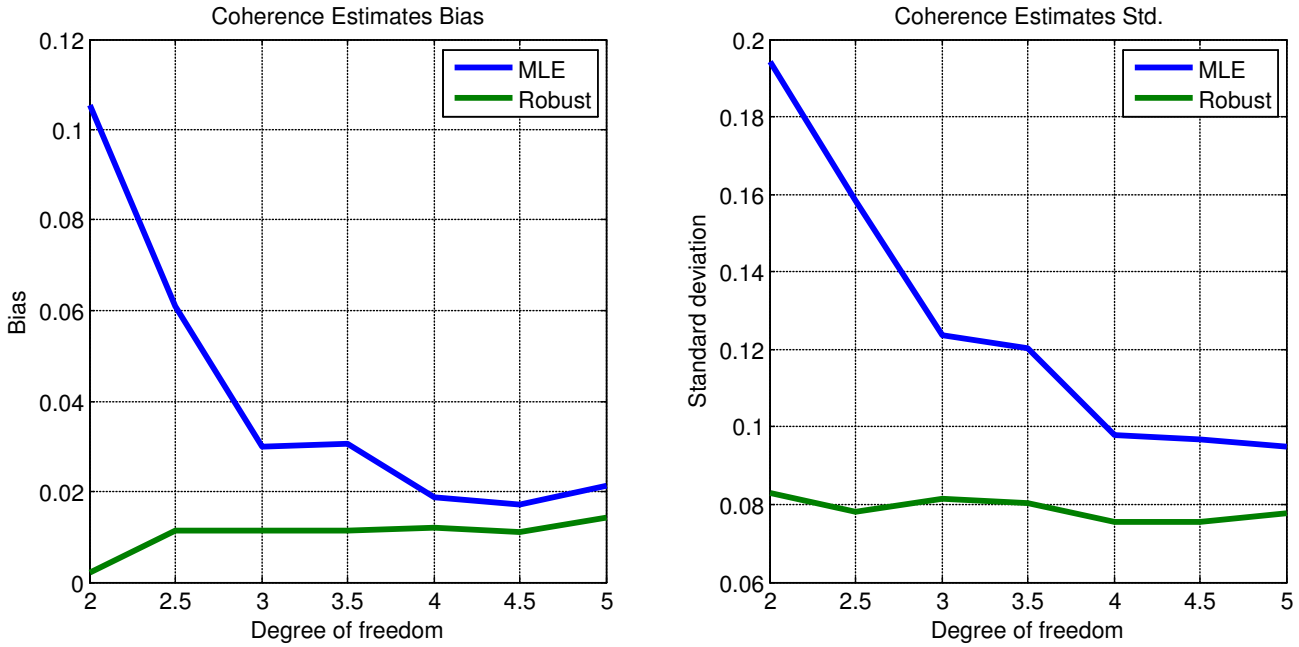


Fig. 4.25. Bias (left) and standard deviation (right) of the coherence estimates of Gaussian MLE and the proposed robust estimator as functions of degree of freedom of the complex t-distribution, for true coherence = 0.2. 100 looks is used in the coherence estimation.

observations, the second order moment of their first order basic rank, let's call it *first order rank coherence*, can be proofed to be γ^2 by substituting equation (4.12) into the coherence equation (2.8).

Now the n-th order rank can be defined as the product of two (n-1)-th order rank pixel:

$$\mathrm{R}^n\left[g_j\right] = \mathrm{R}^{n-1}\left[g_j\right]\mathrm{R}^{n-1}\left[g_{j+2^{n-1}}\right]^* \quad n \in \mathbb{N} \tag{4.14}$$

Please note that the multiplication is not performed with the pixel $j+1$, but $j+2^{n-1}$. This makes sure that an nth order rank pixel is the product of 2^n independent SLC pixels. Therefore, the corresponding nth order rank coherence is

$$\gamma_n = \gamma^{2^n} \tag{4.15}$$

The variance of γ_n also increases accordingly, presumably by a factor of 2^n. Therefore, the number of samples used in the estimation should also be increased by the same factor. The following figure compares the performance of the coherence estimation using different orders of rank. An interferogram is simulated with a constant coherence of 0.5 and the phase from the famous optical image of "Lena". The first subfigure shows the interferometric phase. The results from rank 0 to 2 are plotted in order. Particularly, rank 0 refers to the results using the original SLC pixels. It is obvious that heavy underestimation occurs if the original SLC is used. The degree of underestimation depends on the local fringe frequency. Rank 1 greatly reduced the underestimation. Rank 2 is able to eliminate all the underestimation. But the variance of the coherence estimates also increases as the rank increases.

To conclude, the generalized rank covariance estimator can be invariant to any differentiable spatial phase signal, if higher orders of rank are employed. However, more samples must be used for higher orders of rank, due to the fact that high variance is introduced by the rank operator. Rank 1 is a good balance between phase invariance and estimates variance.

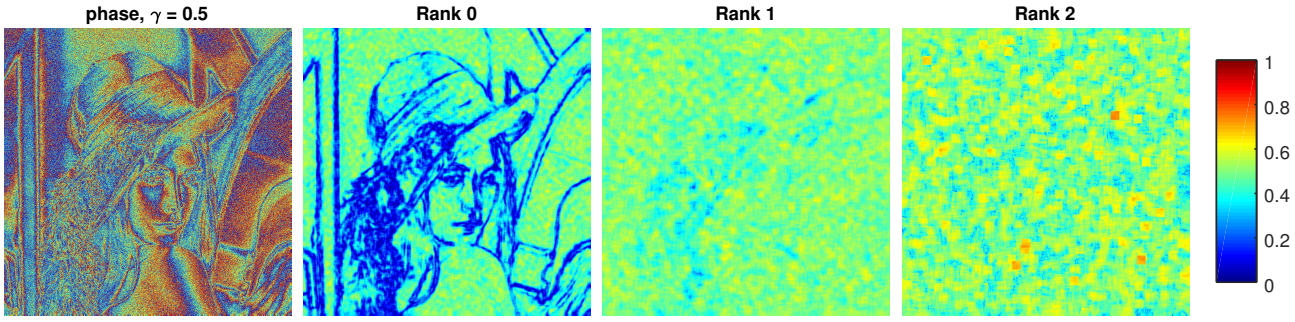


Fig. 4.26. Comparison of the coherence estimates using different orders of rank. The left subfigure shows the simulated interferometric phase with a constant coherence of 0.5. The other subfigures are the coherence estimates using rank 0 to 2. Rank 0 refers to using the original single look complex pixels. The colour bar indicates the coherence. Rank 0 heavily underestimate the coherence. Rank 1 greatly reduced the underestimation. Rank 2 is able to eliminate all the underestimation, but the variance of the coherence estimates also increases as the rank increases.

4.4.3 Robust Phase History Parameters Estimators

This section describes how to robustly estimate the phase history parameters, i.e. elevation, and motion parameters, given a multipass InSAR observation $\mathbf{g} \in \mathbb{C}^N$ with unmodelled phase, e.g. uncompensated atmospheric phase, unmodelled motion phase. The proposed methods were demonstrated using robust PS and DS phase history parameters estimators. This method can also be applied to other InSAR techniques, such as TomoSAR.

Robust PS Estimator

Recall equation (3.1) of PS MLE, which can be simplified as

$$\hat{\theta}_{MLE} = \arg\min_{\theta} \|\mathbf{g} - \bar{\mathbf{g}}(\theta)\|_2^2 \tag{4.16}$$

where $\bar{\mathbf{g}}(\theta)$ is the modelled PS signal. Similar to the robust covariance estimator, it can be robustified by an M-estimator:

$$\hat{\theta}_{M-est} = \arg\min_{\theta} \sum_{i=1}^{N} \rho\left(\mathrm{Re}\left[\varepsilon_i(\theta)\right]/\sigma_R\right) + \rho\left(\mathrm{Im}\left[\varepsilon_i(\theta)\right]/\sigma_I\right) \tag{4.17}$$

where the residual $\varepsilon_i(\theta) = g_i - \bar{g}_i(\theta)$, and $\mathrm{Re}[\cdot]$, $\mathrm{Im}[\cdot]$ are the real and imaginary parts of a complex number, and σ_R and σ_I are the standard deviations of real and imaginary parts, respectively.

The $\rho(x)$ function can be derived from the PDF of the contaminated distribution of $\mathbf{g}$, if it is known or can be modelled. However, it is usually unknown in practice. We shall use stable empirical functions instead, e.g. the Tukey biweight function introduced in section 2.4.3.

Equation (4.17) is solved iteratively, with σ_R and σ_I being updated at each iteration. Their initial values are obtained from the initial estimate $\hat{\theta}_0$. This is not critical for convex $\rho(x)$. However, a robust initial estimation is strongly recommended for nonconvex $\rho(x)$. The initial estimation need not to be efficient. To give one example, the least trimmed square that minimizes the sum of q smallest squared residuals (Rousseeuw, 1984) is as follows:

$$\hat{\theta}_0 = \arg\min_{\theta} \sum_{i=1}^{q} |\varepsilon_i(\theta)|^2 \tag{4.18}$$

The initial estimates of standard deviation, as well as those at each iteration, can be obtained using a robust estimator such as the median absolute deviation:

$$\hat{\sigma} = 1.483 \cdot \mathrm{median}\left(|\varepsilon - \mathrm{median}(\varepsilon)|\right) \tag{4.19}$$

The factor 1.483 is a normalization constant, which gives an expectation of unity for a normal distribution. With this combination, convergence is achieved in just a few iterations; therefore, the computational cost is only approximately three to four times more than that for the standard estimator. The same procedures can be easily extended to the LAMBDA estimator (Kampes and Hanssen, 2004) for PS phase parameter estimations.

Robust DS Estimator

If stationarity is assumed for a DS and its neighbourhood, one can treat the DSs as a single PS by averaging them, as proposed in SqueeSAR. Then, the robustified DS estimator is identical to equation (4.17).

However, stationarity is not assumed in our considerations. The objective is a full inversion of individual single-look DS observation (DS pixel without averaging). The robustified estimator for DS phase history parameters is derived to be:

$$\hat{\theta} = \arg\min_{\theta} \left\{ \varepsilon^H(\theta)\, \mathbf{W}(\bar{\varepsilon})\, \varepsilon(\theta) \right\} \tag{4.20}$$

The residuals $\varepsilon(\theta)$ are whitened by a robust covariance matrix estimate, e.g. $\hat{\mathbf{C}}_{RME}$:

$$\varepsilon(\theta) = \left| \hat{\mathbf{C}}_{RME} \right|^{-0.5} \mathbf{\Phi}(\theta)^H \mathbf{g} \tag{4.21}$$

$\mathbf{W} \in \mathbb{R}^{N \times N}$ is the diagonal weighting matrix computed from each element of the expected residual $\bar{\varepsilon}$.

One should be aware that the marginal distribution of the phase of the residual is uniform, as it is in its whitened version $\varepsilon(\theta)$. The unexpected large phase error, presumably also uniformly distributed over time, applies no change to the DS observations statistically. In other words, the robust loss function is blind to phase contamination in single-look DS observations. Therefore, the correct weighting should be determined by the expected residuals $\bar{\varepsilon}$ of the neighbourhood ensemble. The expected residuals should be calculated using a robust estimator:

$$\bar{\varepsilon} = \sum_{m=1}^{M} w^m \varepsilon^m / \sum_{m=1}^{M} w^m \tag{4.22}$$

where the superscript m denotes the sample number in the neighbourhood, and w^m is a robust weight.

To summarize, equation (4.20) is a joint estimation of the phase parameters of individual single-look DS observations in a neighbourhood. It is solved iteratively. Its computation should begin with initial estimates of each sample in the neighbourhood (assumed to be the same), which jointly determine the initial weighting matrix. The same weighting matrix is used to retrieve the parameters of each single-look DS in the neighbourhood, and is updated on the basis of all the estimates upon finishing one iteration. The pseudocode of the proposed DS estimator is as stated in the following table.

Efficiencies of the Robust Estimators

The efficiencies of the proposed robust PS and DS estimators and their original MLEs are quantitatively assessed using Monte Carlo simulations. The sensor related parameters in the simulation were set to be similar to those of TS-X data, where the wavelength is 0.031 m, the range distance R = 700 km, the acquisitions number N = 20, the spatial baselines $\mathbf{b}_\perp$ were uniformly distributed between -100 m and +100 m, and the temporal baselines $\mathbf{t}$ were evenly spaced from -1 year to +1 year with uniform distribution.

Table 5. Algorithm of the robust estimator for single-look DS observations

Get initial estimate $\hat{\theta}_0$, which can be the same for all DS samples in the same neighbourhood

Calculate initial weighting matrix $\mathbf{W}_0$

for iteration $k = 1$ until convergence

  for DS sample $m = 1$ to M

    Estimate $\hat{\theta}^m$ by solving equation (4.20) using $\mathbf{W}_k$

  end

  Compute the expected residuals $\bar{\varepsilon}$ by a weighted average of ε^m following equation (4.22).

  Compute $\mathbf{W}_{k+1}$ based on $\bar{\varepsilon}$

end

The contamination-free PS and DS observations were simulated firstly. Then contamination is introduced in two ways: (1) random phase is added to eight (40%) of the 20 acquisitions. The phase was uniformly distributed between $[-\pi, \pi]$ for both PS and DS. For DS, sufficient spatial samples were also simulated. The same phase is added to the samples of the same acquisition to simulate the temporally random but spatially stationary atmospheric phase, and (2) a complex t-distribution with one DoF, instead of CCG, was used in the DS simulation to simulate outliers in the DS samples. More details of the simulation can be found in Appendix E.

The proposed robust estimators and their original MLEs were applied to both the contamination-free and contaminated observations. The integral of Tukey's biweight function was chosen as the loss function for both robust PS and DS estimators. The parameter c was set to 4.586, and the Monte Carlo simulation was repeated 1000 times.

Figure 4.27 shows the standard deviation of the estimates of linear deformation rate with respect to different SNRs. The left and right subfigures correspond to the robust PS and DS estimators, respectively. The SNR of the DS refers to the correlated signal w.r.t. the decorrelated part, and can be directly related to the coherence by the formula $\gamma = \left(1 + SNR^{-1}\right)^{-1}$ (Zebker and Villasenor, 1992; Bamler and Hartl, 1998). In each subfigure, the green curves are the results of the proposed robust estimators, and the blue curves are the original MLEs; the dashed curves are the estimators performing on observations with CCG distribution noise, and the solid curves are results for contaminated data. The blue dashed curves in all the subfigures are the MLEs under CCG; hence, they are the optimal estimates.

Figure 4.27 conveys two important messages:

(1) The results suggest that the robust estimators significantly outperform the original MLEs when large amount of the observations are contaminated. The robust estimators improve the efficiency by a factor of 7 to 35 (the squared ratio of the solid curves).
(2) Still, the efficiency of the robust estimator under the nominal distribution (i.e. CCG) is close to that of the original MLE. The relative efficiency of the robust estimator and the MLE under a nominal distribution, defined as

$$\eta_{RE} = \frac{\sigma^2_{MLE}}{\sigma^2_{Robust}} \tag{4.23}$$

is approximately 70% for the proposed robust PS estimator and 80% for the DS estimator. Therefore, the robust estimators are also capable of handling uncontaminated

observations without too much loss of efficiency.

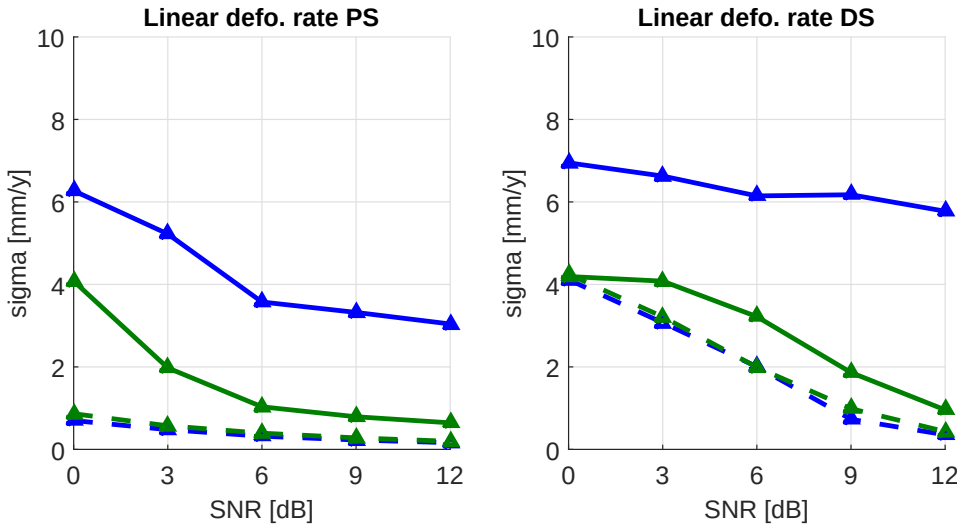


Fig. 4.27. Comparison of the standard deviation of the linear deformation rates estimated by the robust estimators and the ordinary MLEs. The left and right subfigures correspond to the PS and DS estimators, respectively. In both subfigure, the green curves are the results of the proposed robust estimators, and the blue curves are those of the original MLEs; the dashed curves are the estimators performing on contamination-free observations, and the solid curves are the results for contaminated data. The blue dashed curves in all the subfigures are the MLEs under a complex circular Gaussian distribution; hence, they are the optimal estimates. The robust estimators improve the efficiency by a factor of 7 to 35 for contaminated data (the squared ratio of the solid curves) while still keep a relative efficiency of 70% - 80% for contamination-free data (the squared ratio of the dashed curves).

4.4.4 Practical Demonstration of the Robust InSAR Optimization

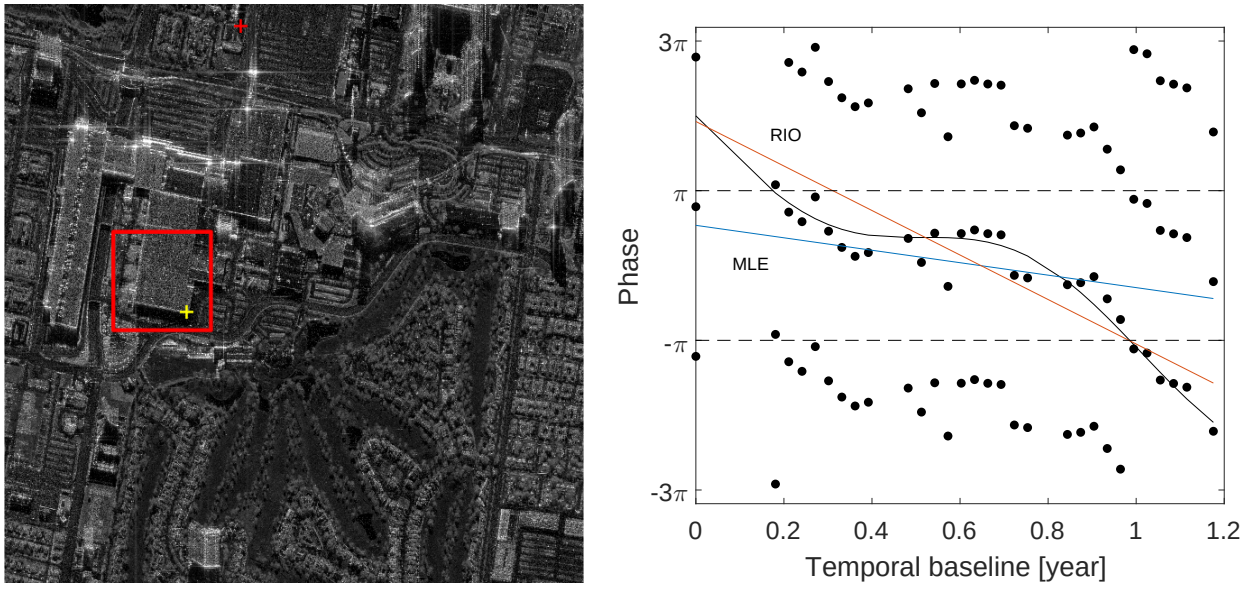
Against unmodelled phase

To demonstrate the robustness against large phase errors, an area in Las Vegas with significant non-linear motion was selected, in a stack of 30 TS-X high resolution spotlight images. The test area is marked by the red rectangle in Figure 4.28(a). The reference point marked by the red cross is a few hundred meters to the area. The correct deformation phase history of a sample pixel (marked by yellow cross) considering both linear and non-linear motion models is shown in Figure 4.28(b). As we can see, not only the motion is complex but also the magnitude of motion is very large. If only a linear motion model is considered, the unmodeled motion phase is equivalent to a large phase error. Non-robust estimators will give biased estimates.

Figure 4.29(a) compares the result of the proposed robust estimator and the ordinary MLE when only the linear model is considered. The robust estimates correctly reconstruct the subsiding bowl, whereas the MLE estimates are heavily biased by the unmodeled motion of the building roof. The advantage of the robust estimator is clearer in Figure 4.29(b) which is the bias of the estimates with respect to the correct linear deformation rate.

Against non-Gaussian scatterer

Figure 4.30 shows the difference between the linear deformation rate of the DSs estimated using the classical sample covariance matrix $\hat{\mathbf{C}}_{MLE}$ (left column) and that using the proposed $\hat{\mathbf{C}}_{RME}$ (right column). Identical samples were used for estimating these two covariance matrices. They were adaptively selected with the KS test using ten amplitude images, which were expected to introduce a non-negligible number of outliers in the selected samples because of the low detection rate from the low number of images. The same ordinary DS MLE was employed to estimate the linear deformation rate in both cases. Therefore, any improvement was solely credited to the use of a more robust covariance matrix estimate.



(a) (b)

Fig. 4.28. (a) the test area with significant non-linear motion is marked by the red rectangle. The red cross is the reference point. (b) the deformation phase history of the pixel marked by yellow cross in (a), where the black dots are the wrapped deformation phases and its duplicates by adding and subtracting 2π, the black curve is a non-linear fit to the phase history, and the red and blue lines are linear fittings using RIO and MLE, respectively.

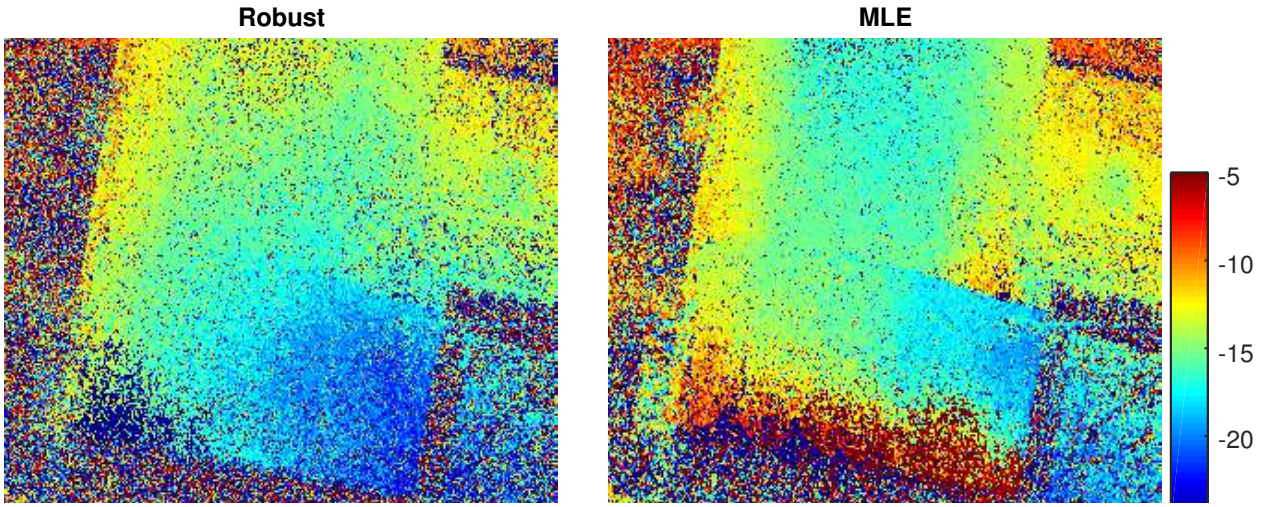
The improvement is clearly demonstrated. The test area is mainly vegetation, except a road in the center which usually appears as a DS in the X-band images. Homogeneous deformation rates are expected, as the spans of both areas were roughly one hundred meters. However, many estimates appear in subfigure (a) as salt-and-pepper noise. We believe this is due to the low detection rate of the KS test, and the nonstationarity of the samples. In contrast, homogenous deformation rate of the road is shown in subfigure (b).

For further comparison, the histograms of the linear deformation rates of the road in Figure 4.30(a) and (b) were plotted in Figure 4.30(c) and (d) respectively. When using $\hat{\mathbf{C}}_{MLE}$, many local peaks appear in the histogram which should not correspond to the deformation signal. With $\hat{\mathbf{C}}_{RME}$, results are considerably more homogenous, and thus more reasonable.

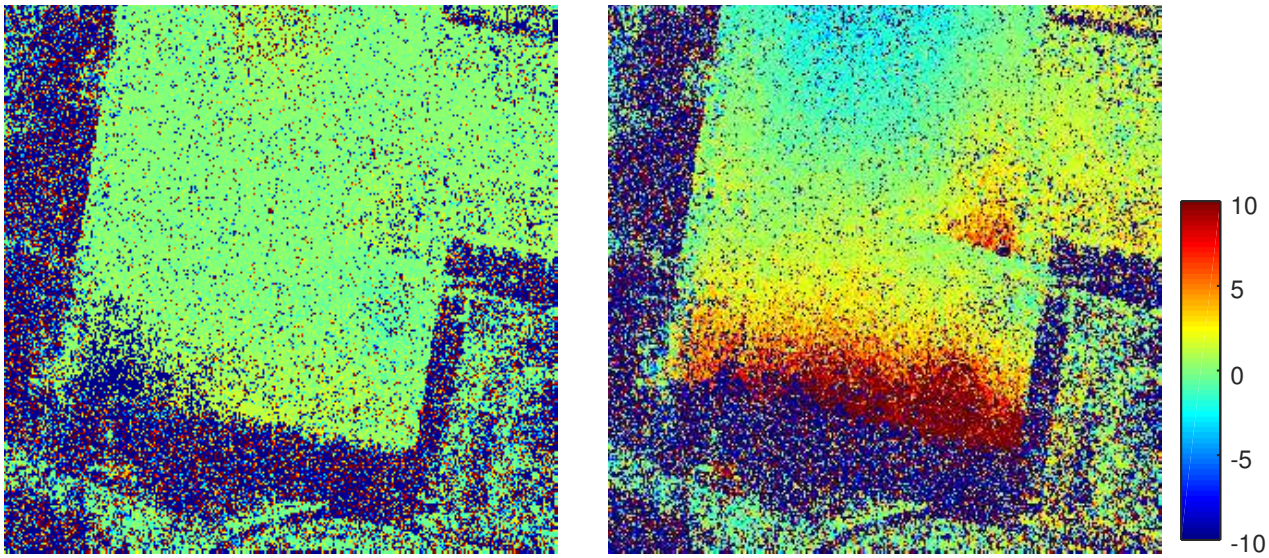
Big Area Demonstration

The RIO framework was applied on a stack of 33 TS-X high resolution spotlight images of the super volcano Campi Flegrei. Figure 4.31 shows the results of the linear deformation rate using the RIO framework, in which Tukey's biweight loss function is used for both robust PS and DS estimator. We specifically chose the SCM for the robust covariance estimate because of the absence of significant fringes in the interferogram.

RIO retrieved 15 times more scatterers than that using only PS. This number should also be valid for standard PS+DS methods, e.g. SqueeSAR. However, RIO greatly improves the overall robustness of the current multipass InSAR estimators. This can be seen in the small area comparison (Figure 4.32) of the two random areas marked by red rectangles in Figure 4.31, where RIO greatly reduced the salt and pepper noise which is probably caused by unmodeled phase and bad covariance estimation.



(a)



(b)

Fig. 4.29. (a) linear deformation rate estimates by robust estimator and ordinary MLE, and (b) the bias of the robust and non-robust estimates to the ”ground truth” (linear deformation rate estimated considering the non-linear motion model). Units: [mm/year].

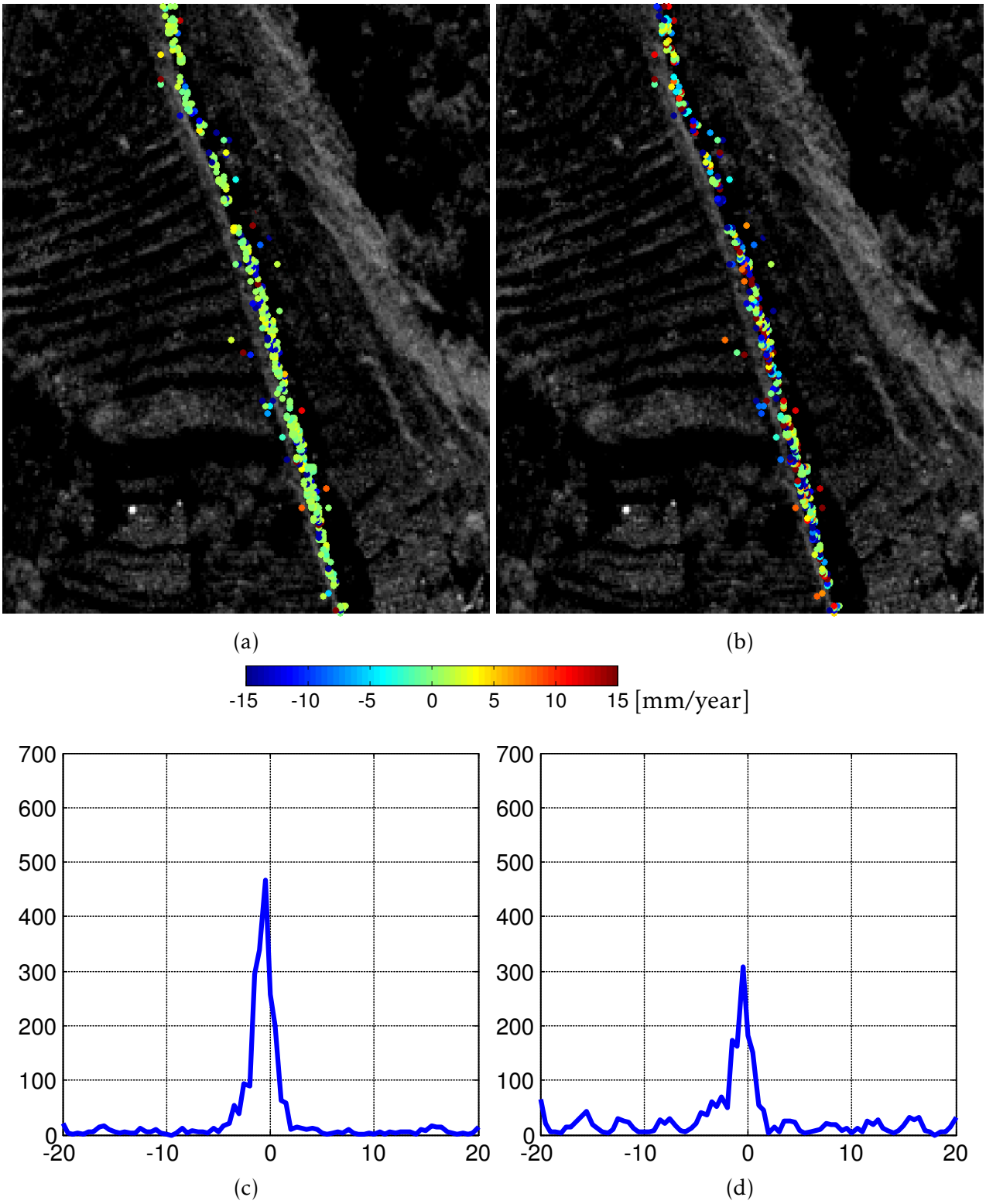


Fig. 4.30. (a) and (b) comparison of the linear deformation rate of the two test sites estimated using the proposed $\hat{\mathbf{C}}_{RME}$ and the ordinary $\hat{\mathbf{C}}_{MLE}$, respectively. Identical samples were used for estimating both covariance matrices. They were adaptively selected (KS test) using ten images. This is expected to introduce a non-negligible number of outliers in the selected samples because of the low detection rate from the low number of images. Results (overlaid on the SAR intensity images) are specific to a local reference point. A homogenous deformation rate is expected because of the small size of the test area. The same ordinary DS MLE was employed for the linear deformation rate estimations in both cases. And (c) and (d) the corresponding histograms of the linear deformation rates in (a) and (b), respectively. In (d), many local peaks appear almost uniformly in the search range, most of which should not correspond to the deformation signals.

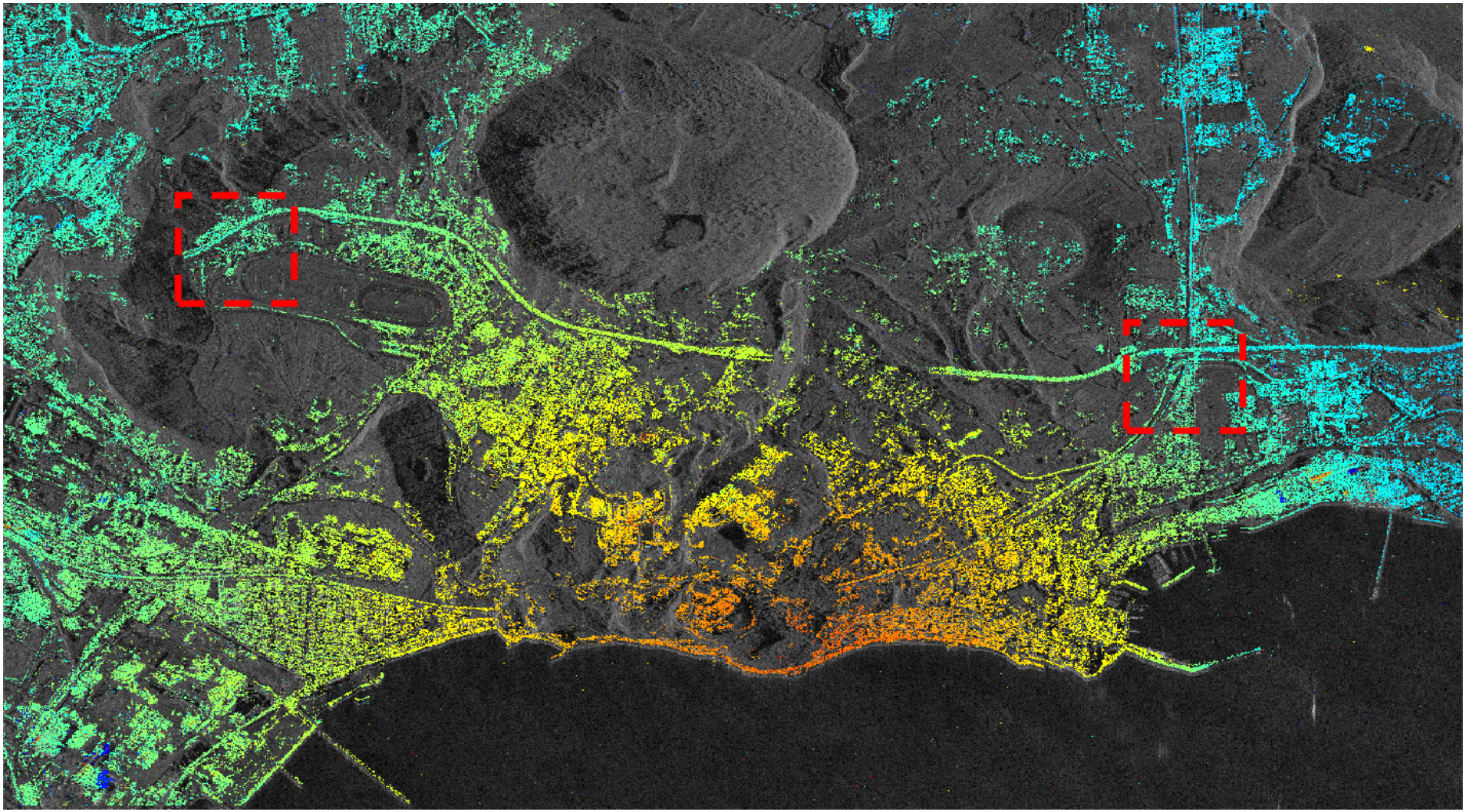


Fig. 4.31. Linear deformation rate of the super volcano Campi Flegrei, as retrieved using the proposed robust InSAR optimization framework that combines both PS and DS in a stack of 33 TS-X high-resolution spotlight images. Tukey's biweight function is set to be the robust loss function, and the sign covariance matrix is employed as the robust covariance matrix estimate. RIO retrieves 15 times more scatterers than those using only PS only. This number should also be valid for standard PS+DS methods, e.g. SqueeSAR. However, RIO greatly reduced the salt and pepper noise probably caused by unmodeled phase and bad covariance estimation. This can be seen in the small area comparison in which plots the results for the two areas marked by red rectangles.

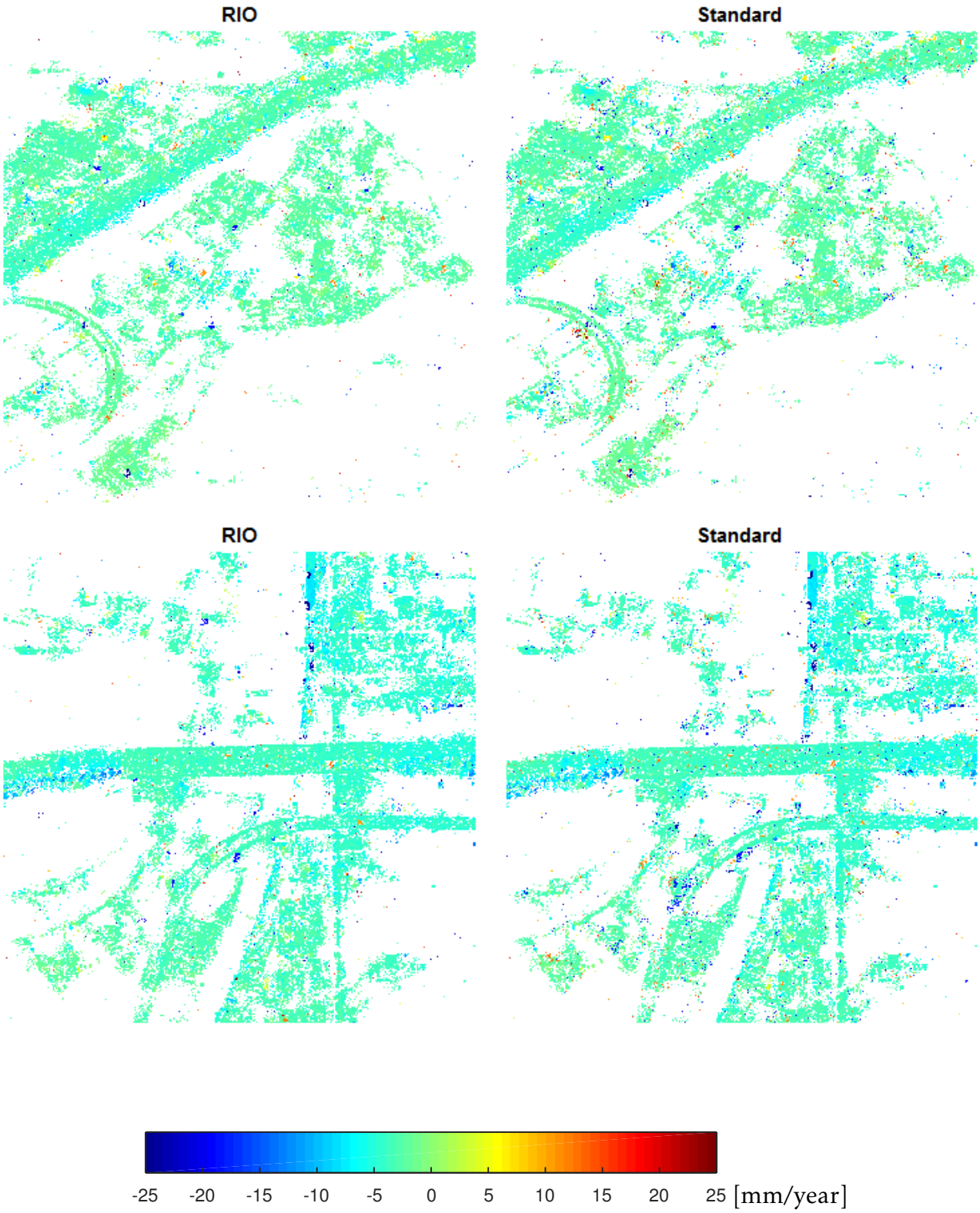


Fig. 4.32. Zoom in of the two areas marked by red rectangles in 4.31.

5 Conclusion and Outlook

5.1 Conclusion

This dissertation is to develop advanced multipass InSAR algorithms, with the aim of computational efficiency, information fusion, contextual awareness, and statistical robustness. The objectives are fulfilled with the work presented in this thesis. The following contributions can be drawn.

- The high computational cost of tomographic SAR inversion compared to PSI requires an efficient approach for operational processing. The computational cost can be greatly reduced by
  - Pre-classifying single and double scatterers without performing TomoSAR inversion on every pixel. The classification should use simple quantities. We used pixel mean intensity, amplitude dispersion index, and the likelihood to single scatterer model for the pre-classification. The current implementation can achieve 70% double scatterer detection rate with 25% false alarm rate. The falsely detected double scatterer can be discriminated in the full tomographic SAR inversion afterwards.
  - using the estimates from PSI as a prior in the TomoSAR inversion of double scatterers.
  - and pass only the necessary pixels to the most expensive superresolution algorithms, e.g. SL1MMER. In our approach, only the single scatterers detected from the non-superresolution algorithm (e.g. SVD-Wiener) whose elevation estimates are less than 1.5 times of the Rayleigh resolution were passed.

  The computational cost of a large area can be reduced by a factor of 50.
- The interferometric phase often contains unmodeled contributions, e.g. unmodeled deformation and uncompensated atmospheric phase. And this is particularly true for very high resolution data of urban area, due to its complex motion and atmospheric phase with possibly high spatial frequency. Robust estimation is recommended for such case. The proposed robust estimator can improve the variance of the deformation estimates by a factor of 7 to 35, while still maintain a relative efficiency of 70% to 80% w.r.t. the MLE for uncontaminated data.
- The expected interferometric phase must be removed in prior of the covariance estimation. It is commonly called defringe, and commonly realized by spatial filtering. The filtering can be replaced by a simple subtraction of the phase of neighbouring pixel, i.e. multiply the complex conjugate of the neighbouring pixel. The augmented observation is called rank observation. The rank covariance estimator is invariant w.r.t. very large fringe frequency. The rank covariance can be generalized to any order, and be theoretically invariant to any fringe frequency.
- The performance of adaptive sample selection in SAR image stack decreases as the number of images decreases. Robust estimation of the covariance is necessary. The M-estimator of covariance provides a general framework. Among them, the non-iterative sign covariance matrix is shown to be a promising method for fast covariance estimation of low coherent signals.
- SAR images as well as InSAR estimate are difficult to interpret, especially for high resolution data of urban area. By geometrically fusing the optical image to the 3-D InSAR point cloud, we can intuitively understand the properties of the scatterers in visible spectrum. In this thesis, the first meter-resolution TomoSAR point cloud textured with optical colour was demonstrated.
- The current multipass InSAR techniques are very much pixel-based inversion and phase history analysis. By transferring the semantic classification label of the optical image to the InSAR point cloud, one can perform a systematic analysis of the InSAR point cloud

based on semantics, e.g. automatic railway monitoring.

5.2 Outlook

The future SAR and InSAR algorithms will become more intelligent and user friendly, which can be seen in the following aspects.

Optical and SAR Image Joint Reconstruction

Some pioneer studies on joint pixel processing are emerging. For example, joint iso-height pixels TomoSAR (Zhu et al., 2015a,b; Ge et al., 2015), contextual phase unwrapping (Baselice et al., 2014), etc. Still, there is much information not fully exploited in SAR images. One obvious way is to use the rich spatial information in optical images to help the InSAR inversion. The following research directions are foreseen:

- Efficient and precise co-registration algorithm for very high resolution optical and SAR images in urban area, instead of full 3-D reconstruction from SAR and optical image introduced in this thesis.
- How to use the features in the optical image, e.g. the colour, classification labels, etc., to help the InSAR reconstruction. One way is to set different prior for different classes. Recall the TomoSAR MAP estimator in equation (3.19), better results can be achieved by introducing different prior covariance matrix $\mathbf{C}_{\gamma\gamma}$ for the prior of every class:

$$\hat{\gamma}_{MAP} = \min_{\gamma} \|\mathbf{g} - \mathbf{R}\gamma\|_2^2 + \lambda \|\gamma\|_{\mathbf{C}_{\gamma\gamma}}^2 \tag{5.1}$$

  Another direction is to find similar pixels in optical image, and perform joint reconstruction of the similar pixels with InSAR.
- Fusion of social media images with InSAR.

3-D Reconstruction from Images of Diverse Angles

Multipass InSAR is based on the concept of coherent imaging, i.e. small angle diversity between images. As the evolution of spaceborne SAR technology, abundant data are available from different viewing angles. How can we break through the limitation of coherent imaging, and perform 3-D/4-D reconstruction using images from very different angle? Two paths in this research are foreseen:

- Real SAR tomography, i.e. by back projection. This probably requires precise pixel-based matching.
- Radargrammetry with multiple images dense matching, i.e. SAR *"structure from motion"*. For urban area, there have been researches on matching of images with orthogonal incidence angles (Soergel et al., 2009), and point scatterer matching (Goel and Adam, 2012b). Focus can be put on better features for dense matching. For non-urban area, there have been study on SIFT descriptor for SAR (Dellinger et al., 2015), dense matching of two images (Rodriguez Gonzalez et al., 2015), and three images matching (Raggam et al., 2010). Research is still requires on multiple images matching.

Efficient Big Data Processing

We expect an unprecedented demand of InSAR processing in the future. Therefore, there is a necessity of developing intelligent InSAR algorithm that can select/prioritize the processing tasks. Most of the pixels in a SAR image do not undergo any significant deformation. Intelligent algorithms which can determine the hotspot of deformation in an image stack prior to the InSAR processing is very useful. The algorithms should use any easily obtainable quantities and other prior knowledge of the deformation to infer the probability of

possible large deformation in the image. One can also exploit the sparsity in spatial and temporal deformational signal, in order to reduce the number of pixels in the processing.

Advanced Deformation Analysis

The current InSAR deformation estimates are still subjected to the following few points: 1) millimeter-precision of linear deformation rate over a large area requires a fairly large stack and good SNR of the scatterers; and 2) InSAR estimates are relative, although there are pioneer studies of absolute positioning and deformation, e.g. SAR imaging geodesy (Eineder et al., 2011; Cong et al., 2012; Balss et al., 2014), using active transponders (Mahapatra et al., 2014), and geodetic TomoSAR (Zhu et al., 2015c). Future research can be focused on:

- ◇ Geodetic InSAR combining absolute geodetic measurement and SAR interferometric estimates.
- ◇ Combining subsurface measurements such as seismic data with InSAR, for modelling the geophysical behaviour of volcanic region from inside and outside.
- ◇ Retrieval higher level information from the phase history, e.g. classification of different motion behaviour, and predict risk level by combining the prior knowledge of the scattering object from optical image.

References

Adam, N., Kampes, B., Eineder, M., 2004. Development of a scientific permanent scatterer system: Modifications for mixed ERS/ENVISAT time series. In: ERS-ENVISAT Symposium.

Adam, N., Kampes, B., Eineder, M., Worawattanamateekul, J., Kircher, M., 2003. The development of a scientific permanent scatterer system. In: ISPRS Workshop High Resolution Mapping from Space, Hannover, Germany, Vol. 2003, 6.

Ali, M., Clausi, D., others, 2002. Automatic registration of SAR and visible band remote sensing images. In: Geoscience and Remote Sensing Symposium, 2002. IGARSS'02. 2002 IEEE International, Vol. 3, IEEE, 1331–1333.

Amarsaikhan, D., Blotevogel, H. H., Genderen, J. L. v., Ganzorig, M., Gantuya, R., Nergui, B., 2010. Fusing high-resolution SAR and optical imagery for improved urban land cover study and classification. International Journal of Image and Data Fusion 1 (1): 83–97.

Auer, S., Gernhardt, S., Bamler, R., 2011. Ghost Persistent Scatterers Related to Multiple Signal Reflections. IEEE Geoscience and Remote Sensing Letters 8 (5): 919–923.

Auer, S., Gisinger, C., Tao, J., 2015. Characterization of Facade Regularities in High-Resolution SAR Images. Geoscience and Remote Sensing, IEEE Transactions on 53 (5): 2727–2737.

Balss, U., Breit, H., Fritz, T., Steinbrecher, U., Gisinger, C., Eineder, M., 2014. Analysis of Internal Timings and Clock Rates of TerraSAR-X. In: Proceedings of the 2014 IEEE International Geoscience & Remote Sensing Symposium, IEEE Xplore, Quebec, Kanada, 2671–2674.

Bamler, R., Hartl, P., 1998. Synthetic aperture radar interferometry. Inverse problems 14 (4): R1.

Baselice, F., Ferraioli, G., Pascazio, V., Schirinzi, G., 2014. Contextual Information-Based Multichannel Synthetic Aperture Radar Interferometry: Addressing DEM reconstruction using contextual information. IEEE Signal Processing Magazine 31 (4): 59–68.

Berardino, P., Fornaro, G., Lanari, R., Sansosti, E., 2002. A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms. IEEE Transactions on Geoscience and Remote Sensing 40 (11): 2375–2383.

Blanco-Sánchez, P., Mallorquí, J. J., Duque, S., Monells, D., 2008. The Coherent Pixels Technique (CPT): An Advanced DInSAR Technique for Nonlinear Deformation Monitoring. Pure and Applied Geophysics 165 (6): 1167–1193.

Bredies, K., Kunisch, K., Pock, T., 2010. Total Generalized Variation. SIAM J. Imaging Sci. 3 (3): 492–526.

Brunner, D., Lemoine, G., Bruzzone, L., 2010. Earthquake Damage Assessment of Buildings Using VHR Optical and SAR Imagery. IEEE Transactions on Geoscience and Remote Sensing 48 (5): 2403–2420.

Budillon, A., Evangelista, A., Schirinzi, G., 2011. Three-Dimensional SAR Focusing From Multipass Signals Using Compressive Sampling. Geoscience and Remote Sensing, IEEE Transactions on 49 (1): 488–499.

Cheng, H., Zheng, S., Yu, Q., Tian, J., Liu, J., 2004. Matching of SAR images and optical images based on edge feature extracted via SVM. In: Signal Processing, 2004. Proceedings. ICSP'04. 2004 7th International Conference on, Vol. 2, IEEE, 930–933.

Colesanti, C., Ferretti, A., Novali, F., Prati, C., Rocca, F., 2003. SAR monitoring of progressive and seasonal ground deformation using the permanent scatterers technique. IEEE Transactions on Geoscience and Remote Sensing 41 (7): 1685–1701.

Cong, X., Balss, U., Eineder, M., Fritz, T., 2012. Imaging Geodesy — Centimeter-Level Ranging Accuracy With TerraSAR-X: An Update. IEEE Geoscience and Remote Sensing Letters 9 (5): 948–952.

Costantini, M., Falco, S., Malvarosa, F., Minati, F., Trillo, F., Vecchioli, F., 2014. Persistent Scatterer Pair Interferometry: Approach and Application to COSMO-SkyMed SAR Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7 (7): 2869–2879, 00007.

Croux, C., Ollila, E., Oja, H., 2002. Sign and rank covariance matrices: statistical properties and application to principal components analysis. In: Statistical data analysis based on the L1-norm and related methods. Springer, 257–269.

Csurka, G., Dance, C., Fan, L., Willamowski, J., Bray, C., 2004. Visual categorization with bags of keypoints. In: Workshop on statistical learning in computer vision, ECCV, Vol. 1, 1–2.

De Maio, A., Fornaro, G., Pauciullo, A., 2009. Detection of Single Scatterers in Multidimensional SAR Imaging. Geoscience and Remote Sensing, IEEE Transactions on 47 (7): 2284–2297.

De Zan, F., 2008. Optimizing SAR interferometry for decorrelating scatterers. Ph.D. thesis.

De Zan, F., Rocca, F., 2005. Coherent processing of long series of SAR images. In: Geoscience and Remote Sensing Symposium, 2005. IGARSS '05. Proceedings. 2005 IEEE International, Vol. 3, 1987–1990.

Deledalle, C.-A., Denis, L., Tupin, F., Reigber, A., Jager, M., 2015. NL-SAR: A Unified Nonlocal Framework for Resolution-Preserving (Pol)(In)SAR Denoising. IEEE Transactions on Geoscience and Remote Sensing 53 (4): 2021–2038.

Dellinger, F., Delon, J., Gousseau, Y., Michel, J., Tupin, F., 2015. SAR-SIFT: A SIFT-Like Algorithm for SAR Images. IEEE Transactions on Geoscience and Remote Sensing 53 (1): 453–466.

Eineder, M., Minet, C., Steigenberger, P., Cong, X., Fritz, T., 2011. Imaging Geodesy - Toward Centimeter-Level Ranging Accuracy With TerraSAR-X. IEEE Transactions on Geoscience and Remote Sensing 49 (2): 661–671.

Fan, B., Huo, C., Pan, C., Kong, Q., 2013. Registration of Optical and SAR Satellite Images by Exploring the Spatial Relationship of the Improved SIFT. IEEE Geoscience and Remote Sensing Letters 10 (4): 657–661.

Feller, W., 1971. An introduction to probability theory and its applications. Wiley.

Ferretti, A., Bianchi, M., Prati, C., Rocca, F., 2005. Higher-order permanent scatterers analysis. EURASIP Journal on Advances in Signal Processing 2005 (20): 3231–3242.

Ferretti, A., Fumagalli, A., Novali, F., Prati, C., Rocca, F., Rucci, A., 2011. A New Algorithm for Processing Interferometric Data-Stacks: SqueeSAR. IEEE Transactions on Geoscience and Remote Sensing 49 (9): 3460–3470.

Ferretti, A., Prati, C., Rocca, F., 2001. Permanent scatterers in SAR interferometry. IEEE Transactions on Geoscience and Remote Sensing 39 (1): 8–20.

Figueiredo, M. A. T., Nowak, R. D., Wright, S. J., 2007. Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems. IEEE Journal of Selected Topics in Signal Processing 1 (4): 586–597.

Fornaro, G., Reale, D., Serafino, F., 2009. Four-Dimensional SAR Imaging for Height Estimation and Monitoring of Single and Double Scatterers. IEEE Transactions on Geoscience and Remote Sensing 47 (1): 224–237.

Fornaro, G., Serafino, F., Soldovieri, F., 2003. Three-dimensional focusing with multipass SAR data. IEEE Transactions on Geoscience and Remote Sensing 41 (3): 507–517.

Gamba, P., Houshmand, B., 2002. Joint analysis of SAR, LIDAR and aerial imagery for simultaneous extraction of land cover, DTM and 3d shape of buildings. International Journal of Remote Sensing 23 (20): 4439–4450.

Ge, N., Zhu, X., Shahzad, M., 2015. Exploiting group sparsity in sar tomography. In: CoSeRa 2015, IEEE Xplore, Pisa, Italy.

Gernhardt, S., Bamler, R., 2012. Deformation monitoring of single buildings using meter-resolution SAR data in PSI. ISPRS Journal of Photogrammetry and Remote Sensing 73: 68–79.

Gernhardt, S., Cong, X., Eineder, M., Hinz, S., Bamler, R., 2012. Geometrical Fusion of Multitrack PS Point Clouds. IEEE Geoscience and Remote Sensing Letters 9 (1): 38–42.

Gernhardt, S., Meyer, F., Bamler, R., Adam, N., 2007. A stability analysis of the lambda estimator for solving the ambiguity problem in persistent scatterer interferometry. In: Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International, IEEE, 2082–2085.

Gini, F., Lombardini, F., 2002. Multilook APES for multibaseline SAR interferometry. Signal Processing, IEEE Transactions on 50 (7): 1800–1803.

Gisinger, C., Balss, U., Pail, R., Zhu, X., Montazeri, S., Gernhardt, S., Eineder, M., 2015. Precise Three-Dimensional Stereo Localization of Corner Reflectors and Persistent Scatterers With TerraSAR-X. IEEE Transactions on Geoscience and Remote Sensing 53 (4): 1782–1802.

Goel, K., Adam, N., 2012a. An advanced algorithm for deformation estimation in non-urban areas. ISPRS Journal of Photogrammetry and Remote Sensing 73: 100–110.

Goel, K., Adam, N., 2012b. Three-Dimensional Positioning of Point Scatterers Based on Radargrammetry. IEEE Transactions on Geoscience and Remote Sensing 50 (6): 2355–2363.

Golub, G. H., Loan, C. F. V., 1996. Matrix Computations. JHU Press, 51949.

Guarnieri, A. M., Tebaldini, S., 2008. On the Exploitation of Target Statistics for SAR Interferometry Applications. IEEE Transactions on Geoscience and Remote Sensing 46 (11): 3436–3443.

Guillaso, S., Reigber, A., 2005. Scatterer characterisation using polarimetric SAR tomography. International Geoscience and Remote Sensing Symposium 4: 2685.

Hanssen, R. F., 2006. Radar Interferometry: Data Interpretation and Error Analysis. Springer Science & Business Media.

Hong, T. D., Schowengerdt, R. A., 2005. A robust technique for precise registration of radar and optical satellite images. Photogrammetric Engineering & Remote Sensing 71 (5): 585–593.

Hooper, A., Zebker, H., Segall, P., Kampes, B., 2004. A new method for measuring deformation on volcanoes and other natural terrains using InSAR persistent scatterers: A NEW PERSISTENT SCATTERERS METHOD. Geophysical Research Letters 31 (23): n/a–n/a.

Huang, Y., Ferro-Famil, L., Reigber, A., 2012. Under-Foliage Object Imaging Using SAR Tomography and Polarimetric Spectral Estimators. IEEE Transactions on Geoscience and Remote Sensing 50 (6): 2213–2225.

Huber, P. J., 1981. Robust Statistics. John Wiley & Sons.

Jiang, M., Ding, X., Hanssen, R. F., Malhotra, R., Chang, L., 2015. Fast Statistically Homogeneous Pixel Selection for Covariance Matrix Estimation for Multitemporal InSAR. IEEE Transactions on Geoscience and Remote Sensing 53 (3): 1213–1224.

Kampes, B., Hanssen, R., 2004. Ambiguity resolution for permanent scatterer interferometry. IEEE Transactions on Geoscience and Remote Sensing 42 (11): 2446–2453.

Kampes, B. M., Adam, N., 2006. The STUN algorithm for persistent scatterer interferometry. In: Fringe 2005 Workshop, Vol. 610, 16.

Katz, S., Tal, A., Basri, R., 2007. Direct visibility of point sets. ACM Transactions on Graphics (TOG) 26 (3): 24.

Kay, S., 2001. Fundamentals Of Statistical Signal Processing: Estimation Theory. Prentice Hall.

Kerr, A. D., 1978. Analysis of thermal track buckling in the lateral plane. Acta Mechanica 30 (1-2): 17–50.

Knoll, F., Bredies, K., Pock, T., Stollberger, R., 2011. Second order total generalized variation (TGV) for MRI. Magnetic resonance in medicine 65 (2): 480–491.

Kotz, S., Nadarajah, S., 2004. Multivariate t distributions and their applications. Cambridge University Press, Cambridge; New York.

Lanari, R., Mora, O., Manunta, M., Mallorqui, J., Berardino, P., Sansosti, E., 2004. A small-baseline approach for investigating deformations on full-resolution differential SAR interferograms. IEEE Transactions on Geoscience and Remote Sensing 42 (7): 1377–1386.

Lee, J.-S., Hoppel, K. W., Mango, S. A., Miller, A. R., 1994. Intensity and phase statistics of multilook polarimetric and interferometric SAR imagery. Geoscience and Remote Sensing, IEEE Transactions on 32 (5): 1017–1028.

Li, H., Manjunath, B. S., Mitra, S. K., 1995. A contour-based approach to multisensor image registration. Image Processing, IEEE Transactions on 4 (3): 320–334.

Liebhart, W., Adam, N., Parizzi, A., 2010. Least Squares Estimation of PSI Networks for Large Scenes with Multithreaded Singular Value Decomposition. In: Synthetic Aperture Radar (EUSAR), 2010 8th European Conference on, 1–4.

Lombardini, F., Gini, F., Matteucci, P., 2001. Application of array processing techniques to multibaseline InSAR for layover solution. In: Radar Conference, 2001. Proceedings of the 2001 IEEE, 210–215.

Lombardini, F., Montanari, M., Gini, F., 2003. Reflectivity estimation for multibaseline interferometric radar imaging of layover extended sources. IEEE Transactions on Signal Processing 51 (6):

1508–1519.

Mahapatra, P. S., Samiei-Esfahany, S., van der Marel, H., Hanssen, R. F., 2014. On the Use of Transponders as Coherent Radar Targets for SAR Interferometry. IEEE Transactions on Geoscience and Remote Sensing 52 (3): 1869–1878.

Mehra, R., Tripathi, P., Sheffer, A., Mitra, N. J., 2010. Visibility of noisy point cloud data. Computers & Graphics 34 (3): 219–230.

Mora, O., Mallorqui, J. J., Broquetas, A., 2003. Linear and nonlinear terrain deformation maps from a reduced set of interferometric SAR images. Geoscience and Remote Sensing, IEEE Transactions on 41 (10): 2243–2253.

Nannini, M., Scheiber, R., Horn, R., Moreira, A., 2012. First 3-D Reconstructions of Targets Hidden Beneath Foliage by Means of Polarimetric SAR Tomography. IEEE Geoscience and Remote Sensing Letters 9 (1): 60–64.

Ollila, E., Koivunen, V., 2003a. Influence functions for array covariance matrix estimators. In: Statistical Signal Processing, 2003 IEEE Workshop on, IEEE, 462–465.

Ollila, E., Koivunen, V., 2003b. Robust antenna array processing using M-estimators of pseudo-covariance. In: Personal, Indoor and Mobile Radio Communications, 2003. PIMRC 2003. 14th IEEE Proceedings on, Vol. 3, IEEE, 2659–2663.

Pacifici, F., Del Frate, F., Emery, W., Gamba, P., Chanussot, J., 2008. Urban Mapping Using Coarse SAR and Optical Data: Outcome of the 2007 GRSS Data Fusion Contest. IEEE Geoscience and Remote Sensing Letters 5 (3): 331–335.

Palubinskas, G., Reinartz, P., 2015. Template based matching of optical and SAR Imagery. In: Joint Urban Remote Sensing Event (JURSE) 2015, IEEE Xplore, Lausanne, Switzerland, 1–4.

Parizzi, A., Brcic, R., 2011. Adaptive InSAR Stack Multilooking Exploiting Amplitude Statistics: A Comparison Between Different Techniques and Practical Results. Geoscience and Remote Sensing Letters, IEEE 8 (3): 441–445.

Pasquali, P., Prati, C., Rocca, F., Seymour, M., Fortuny, J., Ohlmer, E., Sieber, A., 1995. A 3-D SAR experiment with EMSL data. In: Geoscience and Remote Sensing Symposium, 1995. IGARSS '95. 'Quantitative Remote Sensing for Science and Applications', International, Vol. 1, 784–786 vol.1.

Perissin, D., Wang, T., 2012. Repeat-Pass SAR Interferometry With Partially Coherent Targets. IEEE Transactions on Geoscience and Remote Sensing 50 (1): 271–280.

Poulain, V., Inglada, J., Spigai, M., Tourneret, J.-Y., Marthon, P., 2011. High-Resolution Optical and SAR Image Fusion for Building Database Updating. IEEE Transactions on Geoscience and Remote Sensing 49 (8): 2900–2910.

Raggam, H., Gutjahr, K., Perko, R., Schardt, M., 2010. Assessment of the Stereo-Radargrammetric Mapping Potential of TerraSAR-X Multibeam Spotlight Data. IEEE Transactions on Geoscience and Remote Sensing 48 (2): 971–977, 00045.

Reigber, A., Moreira, A., 2000. First demonstration of airborne SAR tomography using multibaseline L-band data. IEEE Transactions on Geoscience and Remote Sensing 38 (5): 2142–2152.

Rife, D., Boorstyn, R. R., 1974. Single tone parameter estimation from discrete-time observations. Information Theory, IEEE Transactions on 20 (5): 591–598.

Rodriguez Gonzalez, F., Eineder, M., Bamler, R., 2015. TerraSAR-X Stereo-Radargrammetric Forest DEM Generation. In: IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2015, IEEE, Milan, Italy.

Rosen, P. A., Hensley, S., Joughin, I. R., Li, F. K., Madsen, S. N., Rodriguez, E., Goldstein, R. M., 2000. Synthetic aperture radar interferometry. Proceedings of the IEEE 88 (3): 333–382.

Rousseeuw, P. J., 1984. Least Median of Squares Regression. Journal of the American Statistical Association 79 (388): 871.

Sauer, S., Ferro-Famil, L., Reigber, A., Pottier, E., 2009. Polarimetric Dual-Baseline InSAR Building Height Estimation at L-Band. IEEE Geoscience and Remote Sensing Letters 6 (3): 408–412.

Schack, L., Soergel, U., Heipke, C., 2015. Persistent Scatterer Aided Facade Lattice Extraction in Single Airborne Optical Oblique Images. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences II-3/W4: 197–205.

Schmitt, M., Stilla, U., 2013. Compressive Sensing Based Layover Separation in Airborne Single-Pass Multi-Baseline InSAR Data. IEEE Geoscience and Remote Sensing Letters 10 (2): 313–317.

Schmitt, M., Stilla, U., 2014. Adaptive Multilooking of Airborne Single-Pass Multi-Baseline InSAR Stacks. IEEE Transactions on Geoscience and Remote Sensing 52 (1): 305–312.

Shahzad, M., Zhu, X., 2015. Robust Reconstruction of Building Facades for Large Areas Using Spaceborne TomoSAR Point Clouds. IEEE Transactions on Geoscience and Remote Sensing 53 (2): 752–769.

Soergel, U. (Hrsg.), 2010. Radar Remote Sensing of Urban Areas. Vol. 15 of Remote Sensing and Digital Image Processing. Springer Netherlands, Dordrecht.

Soergel, U., Michaelsen, E., Thiele, A., Cadario, E., Thoennessen, U., 2009. Stereo analysis of high-resolution SAR images for building height estimation in cases of orthogonal aspect directions. ISPRS Journal of Photogrammetry and Remote Sensing 64 (5): 490–500, 00032.

Sportouche, H., Tupin, F., Denise, L., 2011. Extraction and Three-Dimensional Reconstruction of Isolated Buildings in Urban Scenes From High-Resolution Optical and SAR Spaceborne Images. IEEE Transactions on Geoscience and Remote Sensing 49 (10): 3932–3946.

Stoica, P., Moses, R. L., 2005. Spectral analysis of signals. Pearson/Prentice Hall, Upper Saddle River, N.J.

Tan, W., 1973. On the complex analogue of bayesian estimation of a multivariate regression model. Annals of the Institute of Statistical Mathematics 25 (1): 135–152.

Tao, J., Auer, S., Palubinskas, G., Reinartz, P., Bamler, R., 2014. Automatic SAR Simulation Technique for Object Identification in Complex Urban Scenarios. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7 (3): 994–1003.

Tebaldini, S., 2010. Single and Multipolarimetric SAR Tomography of Forested Areas: A Parametric Approach. IEEE Transactions on Geoscience and Remote Sensing 48 (5): 2375–2387.

Thiele, A., Cadario, E., Schulz, K., Thonnessen, U., Soergel, U., 2007. Building Recognition From Multi-Aspect High-Resolution InSAR Data in Urban Areas. IEEE Transactions on Geoscience and Remote Sensing 45 (11): 3583–3593.

Tough, R. J. A., Blacknell, D., Quegan, S., 1995. A statistical description of polarimetric and interferometric synthetic aperture radar data. Proceedings of the Royal Society of London. Series A: mathematical and Physical Sciences 449 (1937): 567–589.

Tukey, J. W., 1977. Exploratory data analysis, 1st Edition. Behavioral Science. Pearson.

Tupin, F., Roux, M., 2005. Markov random field on region adjacency graph for the fusion of SAR and optical data in radargrammetric applications. IEEE Transactions on Geoscience and Remote Sensing 43 (8): 1920–1928.

Venables, W. N., Ripley, B. D., 2002. Modern applied statistics with S. Springer.

Visuri, S., Koivunen, V., Oja, H., 2000. Sign and rank covariance matrices. Journal of Statistical Planning and Inference 91 (2): 557–575.

Wang, Y., Zhu, X., 2015a. Automatic Feature-based Geometric Fusion of Multi-view TomoSAR Point Clouds in Urban Area. IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing 8 (3): 953 – 965.

Wang, Y., Zhu, X., 2015b. Robust Estimators in Multipass SAR Interferometry. IEEE Transactions on Geoscience and Remote Sensing 54 (2).

Wang, Y., Zhu, X., Bamler, R., 2012. Retrieval of Phase History Parameters from Distributed Scatterers in Urban Areas Using Very High Resolution SAR Data. ISPRS Journal of Photogrammetry and Remote Sensing 73: 89–99.

Wang, Y., Zhu, X., Bamler, R., 2014. An Efficient Tomographic Inversion Approach for Urban Mapping Using Meter Resolution SAR Image Stacks. IEEE Geoscience and Remote Sensing Letters 11 (7): 1250–1254.

Wax, M., 1991. Detection and localization of multiple sources via the stochastic signals model. Signal Processing, IEEE Transactions on 39 (11): 2450–2456.

Wegmüller, U., Walter, D., Spreckels, V., Werner, C., 2010. Nonuniform Ground Motion Monitoring With TerraSAR-X Persistent Scatterer Interferometry. IEEE Transactions on Geoscience and

Remote Sensing 48 (2): 895–904.

Wegner, J. D., Thiele, A., Soergel, U., 2009. Fusion of optical and InSAR features for building recognition in urban areas. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 38 (Part 3): W4.

Wegner, J. D., Ziehn, J. R., Soergel, U., 2014. Combining High-Resolution Optical and InSAR Features for Height Estimation of Buildings With Flat Roofs. IEEE Transactions on Geoscience and Remote Sensing 52 (9): 5840–5854.

Xu, F., Jin, Y.-Q., 2007. Automatic Reconstruction of Building Objects From Multiaspect Meter-Resolution SAR Images. IEEE Transactions on Geoscience and Remote Sensing 45 (7): 2336–2353.

Zebker, H. A., Villasenor, J., 1992. Decorrelation in interferometric radar echoes. Geoscience and Remote Sensing, IEEE Transactions on 30 (5): 950–959.

Zhu, X., 2008. Spectral Estimation for Synthetic Aperture Radar Tomography. Ph.D. thesis, Lehrstuhl für Methodik der Fernerkundung, TU München, Munich, Germany.

Zhu, X., 2011. Very High Resolution Tomographic SAR Inversion for Urban Infrastructure Monitoring: A Sparse and Nonlinear Tour. Vol. 666 of Verlag der Bayerischen Akademie der Wissenschaften. Deutsche Geodätische Kommission.

Zhu, X., Bamler, R., 2009. Very high Resolution SAR tomography via Compressive Sensing. In: Proceedings of Fringe 2009, Frascati, Italy, 1–7.

Zhu, X., Bamler, R., 2010a. Tomographic SAR Inversion by L1-Norm Regularization – The Compressive Sensing Approach. IEEE Transactions on Geoscience and Remote Sensing 48 (10): 3839–3846.

Zhu, X., Bamler, R., 2010b. Very High Resolution Spaceborne SAR Tomography in Urban Environment. IEEE Transactions on Geoscience and Remote Sensing 48 (12): 4296–4308.

Zhu, X., Bamler, R., 2011. Let's Do the Time Warp: Multicomponent Nonlinear Motion Estimation in Differential SAR Tomography. IEEE Geoscience and Remote Sensing Letters 8 (4): 735–739.

Zhu, X., Bamler, R., 2012. Demonstration of Super-Resolution for Tomographic SAR Imaging in Urban Environment. IEEE Transactions on Geoscience and Remote Sensing 50 (8): 3150–3157.

Zhu, X., Ge, N., Shahzad, M., 2015a. Group Sparsity in SAR Tomography - Experiments on TanDEM-X data Stacks. In: International Radar Symposium 2015, Dresden, Germany.

Zhu, X., Ge, N., Shahzad, M., 2015b. Group sparsity in sar tomography for urban mapping. IEEE Journal of Selected Topics in Signal Processing Accepted.

Zhu, X., Montazeri, S., Gisinger, C., Hanssen, R., Bamler, R., 2014. Geodetic TomoSAR - Fusion of SAR Imaging Geodesy and TomoSAR for 3d Absolute Scatterer Positioning. In: Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International, Quebec, Canada.

Zhu, X., Montazeri, S., Gisinger, C., Hanssen, R., Bamler, R., 2015c. Geodetic SAR Tomography. IEEE Transactions on Geoscience and Remote Sensing .

Zhu, X., Shahzad, M., 2014. Facade Reconstruction Using Multiview Spaceborne TomoSAR Point Clouds. IEEE Transactions on Geoscience and Remote Sensing 52 (6): 3541–3552.

Zoubir, A., Koivunen, V., Chakhchoukh, Y., Muma, M., 2012. Robust Estimation in Signal Processing: A Tutorial-Style Treatment of Fundamental Concepts. IEEE Signal Processing Magazine 29 (4): 61–80.

List of Figures

List of Tables

Acknowledgement

This work is absolutely not possible without the help of many people. First of all, my greatest gratitude goes to my supervisor Prof. Xiaoxiang Zhu, for not only but especially, sharing her comprehensive knowledge of multiple disciplines and providing unparalleled supervision. Her standard on research quality has influenced and changed many aspects of my life. Her previous excellent researches have been a perfect platform without which my work will not be visible to the community. I am grateful to her trust and support without which I could not finish this thesis after the scholarship funding period. She stresses not only scientific research but also the integrity of the team. She carefully distributes topics among the team members. I made good friends with all the team members, and feel the sense of belonging to the SiPEO team. She has also been a friend, a mentor, and a role model. As a friend, she showed great support to many aspects of my life in Germany; as a mentor, she guided me and made me a better person; and as a role model, she is a visionary and a practitioner. I am still being encouraged by her everyday passion of conducting hands-on research herself.

The same sincere gratitude goes to my supervisor Prof. Richard Bamler, for not only the same quality as Prof. Zhu but especially, his capacity of understanding and generousness. This is not only limited to scientific discussions. He does not only hear what you said, but also to what you have not said. He is a great researcher. His article "SAR Interferometry" is the sanctuary of my research in SAR signal processing, because all sorts of help and emergency assistant can be found. He has also played many other important roles. He brought laughter to us with his well-integrated science-in-life joke; he searched cheap solutions for our health insurance; and he urged me on my research from time to time with gentle force, and so on. I remember the conversation in the kitchen "none of us are tall" "but together we are strong".

Sincere gratitude also gives to Prof. Uwe Sörgel from TU Darmstadt for being the third reviewer of this thesis. His fast response to the request also plays an important role in finishing this thesis. He is a significant figure in the SAR community. His contributions in urban remote sensing have been remarkable. His scientific work together with his continuous attention in our works have encouraged me.

This thesis is also absolutely not possible without the great support from the colleagues of DLR-IMF, TUM-LMF, TUM-SiPEO, especially the SAR department led by Prof. Michael Eineder. I benefited a lot from the DLR-IMF's PSI-GENESIS, and now the IWAP system. Therefore, the whole InSAR team deserves a warm thanks, including the team leader Nico Adam, and Fernando Rodriguez, Christian Minet, Alessandro Parizzi, Dr. Ramon Brcic, Franz Ulmer, Dr. Kanika Goel, Robert Shau, and so on.

I also thank other individuals who provided support and data, including Dr. Stefan Gernhardt, Dr. Stefan Auer, Dr. Helko Breit, Gerald Baier, Dr. Michael Bäßler, Dr. Pablo d'Angelo, Dr. Ruppert Müller, Dr. Jiaojiao Tian, Dr. Jian Xu, Dr. Andrian Doicu, Dr. Xiaoying Cong, Nan Ge, Muhammad Shahzad, Claas Grohnfeldt, Dr. Heiko Hirschmüller, and so on.

Last but not least, my gratitude to my family is beyond any language. Every bit of my life right now traces back to every bit of love they offered me on my path.

Appendices

A Wang Y, Zhu X, Bamler R (2014) An Efficient Tomographic Inversion Approach for Urban Mapping Using Meter Resolution SAR Image Stacks. IEEE Geoscience and Remote Sensing Letters, 11 (7): 1250 - 1254.

http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6714382

An Efficient Tomographic Inversion Approach for Urban Mapping Using Meter Resolution SAR Image Stacks

Yuanyuan Wang, Xiao Xiang Zhu, and Richard Bamler

***Abstract*—This letter describes an efficient approach of multidimensional synthetic aperture radar (SAR) imaging for urban mapping. The proposed approach is an integration of tomographic SAR inversion and the well-known persistent scatterer interferometry (PSI). It consists of three steps: first, a global estimation of the topography and motion parameters using efficient algorithms such as PSI; second, a single and double scatterer discrimination step based on the results of the first step; finally, a tomographic SAR inversion, which is performed on the preclassified double scatterers, using the prior knowledge obtained in the first step, retrieving the topography and motion parameters of both scatterers. The proposed approach has been tested on a dozen of TerraSAR-X high-resolution spotlight image stacks. In this letter, examples from Las Vegas and Berlin are presented. The results are comparable with the one obtained by the most computationally expensive tomographic SAR algorithms (e.g., SL1MMER) only and saves computational time by a factor of 50.**

***Index Terms*—SAR tomography (TomoSAR), SL1MMER, synthetic aperture radar (SAR), TerraSAR-X, urban mapping.**

I. Introduction

A. TomoSAR and PSI

TOMOGRAPHIC synthetic aperture radar (SAR) inversion includes SAR tomography (TomoSAR) [1] and differential TomoSAR (D-TomoSAR) [2]–[4]. Because of their layover resolving capability, they are often used for urban 3-D topographic reconstruction, forest structure imaging, and ground or building deformation monitoring. As an extension of TomoSAR, D-TomoSAR introduces an additional motion term in the TomoSAR system model, allowing the retrieval of the motion parameters of each scatterer in a resolution cell. The D-TomoSAR system model can be expressed as follows:

$$g_n = \int_{\Delta s} \gamma(s) \exp\left(-j2\pi\left(\xi_n s + \frac{2d(s,t_n)}{\lambda}\right)\right) ds \quad (1)$$

Manuscript received May 9, 2013; revised July 15, 2013 and September 3, 2013; accepted September 26, 2013. This work was supported in part by the International Graduate School of Science and Engineering of Technische Universität München under Project 6.08 4D City, by the Helmholtz Association under the framework of the Young Investigators Group "SiPEO," and by the German Research Foundation (DFG) under Förderkennzeichen BA2033/3-1.

Y. Wang is with the Lehrstuhl für Methodik der Fernerkundung, Technische Universität München, 80333 Munich, Germany (e-mail: wang@bv.tum.de).

X. X. Zhu and R. Bamler are with the Lehrstuhl für Methodik der Fernerkundung, Technische Universität München, 80333 Munich, Germany, and also with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234 Weßling, Germany (e-mail: xiao.zhu@dlr.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2013.2290833

where g_n is the complex pixel value in the nth single-look complex SAR image, s represents the *elevation* direction, $d(s,t_n)$ is the displacement as a function of the elevation and the acquisition time t_n, λ is the wavelength of the SAR system, $\xi_n = -2b_n/(\lambda r)$, which is the so-called "spatial frequency" and is proportional to the baseline b_n and range r, and $\gamma(s)$ is the complex reflectivity profile along the elevation direction of that pixel. Solving the reflectivity $\gamma(s)$ is essentially a spectral estimation problem. In case of solving it parametrically, the solution is, in addition, subject to a multidimensional search in the elevation and motion parameters solution space.

Persistent scatterer interferometry (PSI) [5] assumes a maximum single dominant scatterer in a pixel. Under such assumption, $\gamma(s)$ in (1) is modeled as a delta function with a constant amplitude. Assuming *linear* and *seasonal* motion models (common in urban area), the PSI system model can be simplified as

$$g_n = A \exp\left(-j2\pi(\xi_n s + \eta_n v + \omega_n a)\right) \quad (2)$$

where A is the amplitude of the point scatterer, which is usually dropped in the estimation, $\eta_n = 2t_n/\lambda$ is the temporal frequency modeling linear movement, and $\omega_n = 2\sin(2\pi(t_n - t_0))/\lambda$ is the temporal frequency modeling seasonal motion. v and a are the velocity of linear motion and the amplitude of periodic motion, respectively. Solving the unknown elevation and the motion parameters is to match the phase on both sides of (2), which can be efficiently solved by maximizing the ensemble coherence (periodogram) [6]. Because of the simplification in the system model, PSI is an efficient method for deformation monitoring of large areas.

B. Integration of TomoSAR and PSI

As the resolution of SAR images goes higher, it provides us enormous details on one hand, and on the other hand, it increases the computational burden. This is particularly true for monitoring an entire urban area using D-TomoSAR because it is a spectral estimation problem whose computational efficiency is mainly restricted by the following: 1) dimension of the spectral estimation problem for each pixel. Usually, 3-D is applied (elevation, linear motion, and periodic motion) [7]; 2) extent of the parameter solution range, i.e., the search range in each dimension of the solution space; and 3) the spectral estimation algorithms being applied, such as periodogram (economic), SVD-Wiener [4] (a little bit more expensive) and SL1MMER [8] (much more expensive).

To reduce the computation, we propose to integrate PSI into the D-TomoSAR inversion, with these considerations. 1) PSI

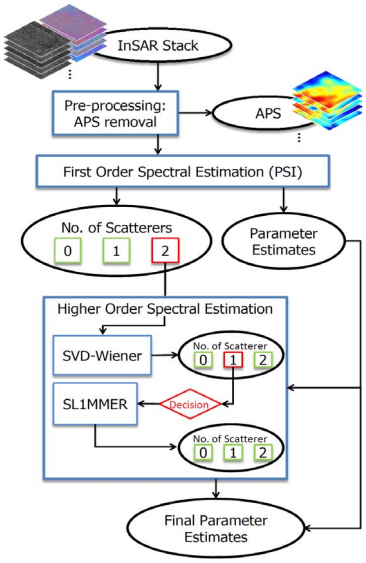


Fig. 1. Processing flowchart. Black ellipses represent input/output data. Blue rectangles represent processes/algorithms. The red diamond represents decision. The green rectangles in the black ellipses represent accepted results, whereas those red ones represent results that need to be passed to the next step.

needs to provide a reasonable atmospheric phase screen (APS) estimation; 2) PSI should provide the prior knowledge about the dimension of the problem (i.e., the motion model) and the search range in each dimension of the solution space; 3) More expensive methods such as SVD-Wiener and SL1MMER shall only be applied to the pixels that probably contain more than one scatterer; and 4) Algorithms with superresolving capability such as SL1MMER shall only be applied when superresolution is required.

C. Summary

The proposed TomoSAR inversion approach is an integration of PSI and D-TomoSAR. The processing flowchart is shown in Fig. 1. It follows three steps: first-order spectral estimation, single and double scatterer(s) discrimination, and higher order spectral estimation. A near-operational processor is implemented by combining the PSI, SVD-Wiener, and SL1MMER developed in the Remote Sensing Technology Institute of the German Aerospace Center (DLR).

The rest of the letter will describe each step in this approach. The details of PSI, SVD-Wiener, and SL1MMER will not be addressed and can be found in [4], [5], and [8]. Focus will be put on the integration of PSI and D-TomoSAR, particularly the discrimination of pixels with single or double scatterer(s). Results using TerraSAR-X high-resolution spotlight data are presented in Section III.

II. TomoSAR Processing Chain

The proposed approach is summarized in Fig. 1. First, APS is removed from the coregistered image stack in the PSI-type preprocessing; then, first-order spectral estimation (i.e., single-scatterer model) is applied on all the pixels, obtaining the estimates of the model parameters. Based on these estimates, amplitude dispersion index (ADI), and pixel brightness, a prediscrimination of the number of scatterers in each pixel is then obtained. As we limit the maximum number of scatterers to two, all the pixels shall fall into one of the three classes: no, single, or double scatterer(s). The prediscrimination follows two steps: The Bayesian decision rule is first applied to discard pixels with no scatterer and followed by a support vector machine (SVM) to separate pixels with single scatterer and double scatterers. The pixels detected with double scatterers (upper red box in Fig. 1) are passed to higher order spectral estimation. They are first processed with SVD-Wiener. Some of them are confirmed as pixels containing double scatterers, and the rest are rejected as having either single or no scatterer. Because SVD-Wiener has almost no superresolution capability, among those pixels rejected as single scatterers (lower red box in Fig. 1), some actually contain double scatterers. Therefore, after a decision (red diamond in Fig. 1) on the elevation distance, some of them are passed to the SL1MMER algorithm with superresolution capability. The following text in this section will explain each step in detail.

A. Preprocessing

The purpose of preprocessing is to estimate and further remove the APS from a coregistered complex SAR image stack. The image stacking is done by DLR's PSI-GENESIS system [9]. The APS is removed using a similar approach as that in standard PSI processing and the "spatial difference" method mentioned in [10]. Since the APS is spatially smooth varying, preprocessing is performed on the downsampled image stacks to reduce computational cost. Note that the pixel selection and parameter estimates in the preprocessing do not affect/replace any procedures in the following D-TomoSAR inversion steps. The basic preprocessing procedures are listed as follows.

1) Downsample images in azimuth and range direction.
2) Build redundant network from spatially differential measurement in the downsampled image stacks.
3) Estimate model parameters of the differential measurements in the network, assuming a single-scatterer model with linear and seasonal motion.
4) Integrate differential topography estimates and remove the corresponding phase contribution from the original interferometric phase.
5) Unwrap the remaining residual phase in azimuth and range direction.
6) Apply a temporal high-pass filter and a spatial low-pass filter to the unwrapped phase stack. This results to the APS in the downsampled image size.
7) Upsample the APS to the original image size and remove from the original phase.

B. First-Order Spectral Estimation

The first-order spectral estimation refers to a special case of the D-TomoSAR problem considering only a single scatterer in each range–azimuth pixel. In the proposed processing chain, it is solved by maximizing the periodogram [11] of the amplitude-normalized measurements [5], i.e.,

$$\{\hat{s},\hat{v},\hat{a}\}=\arg\max\left\{\left|\frac{1}{N}\sum_{n=1}^{N}e^{-j2\pi(\xi_n s+\eta_n v+\omega_n a)}\frac{g_n^*}{|g_n|}\right|\right\} \quad (3)$$

where N is the total number of interferograms, and $*$ stands for the complex conjugate operator.

After applying the estimation on every pixel, we obtain the estimates of the elevation and motion parameters related to a reference point that we assume stable. Median and maximum filters are applied on the range–azimuth image of the estimates to obtain the prior for higher order spectral estimation. In addition, the following quantities are also recorded: 1) pixel brightness; 2) ADI; and 3) likelihoods of the single- and no-scatterer models w.r.t. the measurements g_n.

Under a circular Gaussian assumption of the measurements with covariance ε, the log likelihood of the single-scatterer model, i.e., H_1, is simply the sum of squared residuals [4], i.e.,

$$\ln p(\mathbf{g}|H_1) = \ln\left(\frac{1}{\pi^N|\varepsilon|^{2N}}\right) - \frac{1}{|\varepsilon|^2}\sum_{i=1}^{N}\left|g_n - \hat{A}\exp\left(-j2\pi(\xi_n\hat{s}+\eta_n\hat{v}+\omega_n\hat{a})\right)\right|^2 \quad (4)$$

and for the no-scatterer model, i.e., H_0, it is the sum of squared measurements, i.e.,

$$\ln p(\mathbf{g}|H_0) = \ln\left(\frac{1}{\pi^N|\varepsilon|^{2N}}\right) - \frac{1}{|\varepsilon|^2}\sum_{i=1}^{N}|g_n|^2. \quad (5)$$

C. Number of Scatterers Discrimination

Discriminating pixels with different numbers of scatterers is a classification problem. One of the optimum detection strategies is the likelihood ratio test, if the assumption of the underlying models is correct. However, such detection strategy requires the likelihood to the double-scatterer model, i.e., H_2, and hence is not feasible before the TomoSAR inversion. Therefore, in order to prediscriminate the pixels before a TomoSAR inversion, our method is a two-step approach, comprising both theoretical and empirical parts.

First, pixels without scatterer are discriminated following the Bayesian decision rule with penalized likelihood criterion, i.e.,

$$\frac{p(\mathbf{g}|H_0)}{p(\mathbf{g}|H_1)\exp(-\mathrm{AIC})}\frac{P_0}{P_{1,2}} \overset{H_0}{>} 1. \quad (6)$$

The H_0 and H_1 likelihoods of each pixel are compared, taking into account the Akaike information criterion (AIC) [4]. P_0 and $P_{1,2}$ are the prior probabilities of pixels with no scatterer and with single or double scatterer(s), respectively. Close to 100% detection rate and 0% false alarm rate can be achieved in this stage because the H_1 likelihoods of the pixels with double scatterers are much higher than their H_0 likelihoods. This is proven in the real data experiment presented in Section III-A.

The remaining pixels need to be further classified into single and double scatterer(s). Since the H_2 likelihood is not available at this stage, a few other easily accessible features associated with each pixel are used for the classification. They are ADI, H_1 likelihood, and pixel mean intensity, with these three reasons.

1) Pixels with single scatterer experience lower ADI since the amplitudes of pixels subjected to only a single dominant scatterer do not show pronounced fluctuations as a function of baseline.
2) The H_1 likelihood tends to be lower if multiple scatterers are present.
3) The aforementioned two aspects are based on a reasonable signal-to-noise ratio range. Therefore, only relatively bright pixels should be selected as candidates of pixels with double scatterers.

Without the analytical model of the joint probability density function of the aforementioned three quantities, we employ the SVM to find a separation of these two classes. Considering our situation of missing classification ground truth, the usage of the SVM basically follows these points.

— The training data of the SVM classifier are the classification results obtained by SL1MMER (theoretically, the closest to ground truth) [12] on a small and representative test area. The SVM classifier takes into account three features: ADI, H_1 likelihood, and pixel mean intensity.
— A nonlinear SVM classifier is applied. A Gaussian kernel is used, which is suggested in [13] as a good first try.
— The optimum parameter setting is determined by the following.
  1) Train the classifier using 50% of the pixels in the small test area.
  2) Classify all the pixels in the test area using the trained classifier.
  3) Repeat this with a geometric sequence of the parameters and find the parameters that give the highest detection rate at a constant false alarm rate.
— The trained SVM classifier is then used to classify all the pixels in the whole scene, based on the three features of each pixel.

D. Higher Order Spectral Estimation (SVD-Wiener, SL1MMER)

All pixels classified as double scatterers are processed using SVD-Wiener first, taking into account the prior knowledge of the estimates obtained in the first-order spectral estimation. In the SVD-Wiener processing, these pixels are tested through model order selection and are classified as no, single, or double scatterer(s) again. Two layers of elevation and motion parameter estimates are retrieved from the pixels confirmed with double scatterers. The lower layer of elevation estimates is interpolated in range–azimuth dimension to get a ground surface height s_0 for each pixel (the index of the pixel is ignored in s_0).

Since SVD-Wiener has almost no superresolution capability, double scatterers with elevation distance shorter than the Rayleigh resolution $\rho_s = \lambda r/2\Delta b$ [1] (r: slant range; Δb: elevation aperture size) will be misclassified as single scatterer. Therefore, among the detected pixels with single scatterer, those whose elevation estimates with respect to its corresponding ground surface height are smaller than ρ_s (in practice, we allow some buffer and set it to be $1.5\rho_s$) are potential pixels that would require superresolution. In this case, SL1MMER is applied. This decision is marked as the red diamond in the processing flowchart in Fig. 1. In the SL1MMER processing, the search range is restricted to $[s_0 \quad s_0 + 1.5\rho_s]$.

III. Validation With Real Data

A. Experiments on Number of Scatterers Discrimination

This section will evaluate the performance of the scatterer discriminator over a test area in downtown Las Vegas shown in Fig. 2. This stack contains 25 TerraSAR-X high-resolution spotlight images. Since the missing of ground truth, the SL1MMER result is taken as reference for comparison.

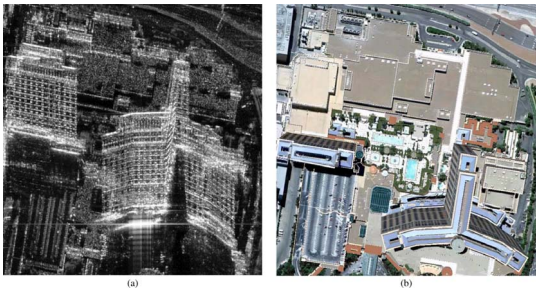


Fig. 2. (a) Amplitude image of the test area in downtown Las Vegas. (b) Optical image from Google Earth of the same area.

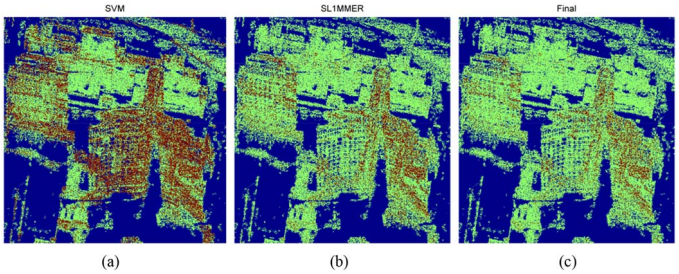


Fig. 3. (a) Preclassification result of the proposed method (likelihood ratio test + SVM). (b) Result from the SL1MMER algorithm. (c) Final classification after the higher order spectral estimation. Red represents pixels with double scatterers, green represents pixels with single scatterer, and blue represents pixels without coherent scatterer.

First, the pixels without scatterer are determined according to (6). The prior probabilities P_0 and $P_{1,2}$ are set to be both 50% in the experiment. Then, the remaining pixels are separated using an SVM into containing either single or double scatterer(s). Fig. 3(a) and (b) shows a comparison of the SVM and SL1MMER classification results, with red representing pixels with double scatterers, green representing pixels with single scatterer, and blue representing pixels with no scatterer.

In Fig. 3(a), 47% of the pixels are detected as no scatterers, 36% are detected as single scatterers, and the remaining 17% are detected as double scatterers. The classification of pixels with no scatterer is almost 100% correct compared with the SL1MMER result. In the rest of the pixels, the detection rate of pixels with double scatterers is 72.4%, with a false alarm rate of 25.7%, which is why much more pixels are detected as double scatterers. However, higher detection rate (consequently, higher false alarm rate) is still favored in the processing because the falsely detected pixels with double scatterers can still be corrected in the higher order spectral estimation, but not for the misdetected pixels with double scatterers. In order to fully characterize the proposed single and double scatterer(s) detector, its receiver operating characteristic (ROC) curve is plotted in Fig. 4. The blue dots are the experimental results, and the green curve is an analytical fit. Depending on the processing power, the threshold can be shifted toward higher detection rate, affording higher false alarm rate, and, therefore, more processing effort by more frequently applying SVD-Winer and SL1MMER.

Next, SVD-Wiener is applied on the red pixels shown in Fig. 3(a). Most of them are confirmed as double scatterers, and others are rejected as no or single scatterer. Among those rejected as single scatterers, those having small elevation estimates (60 m in our case $= 1.5$ times of ρ_s) w.r.t the ground surface height are handed over to SL1MMER for reprocessing.

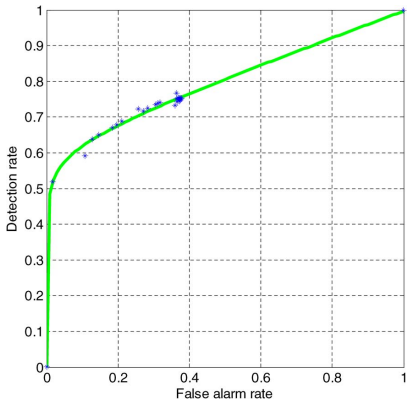


Fig. 4. ROC curve of the proposed SVM double scatterers detector. The blue dots are the best detection rates achieved in the experiments at a specific false alarm rate. The green curve is an analytical fit to the blue dots.

They account for 9% of the total pixel number in the test area. Then, the final corrected classification map of the test area can be obtained, which is shown in Fig. 3(c).

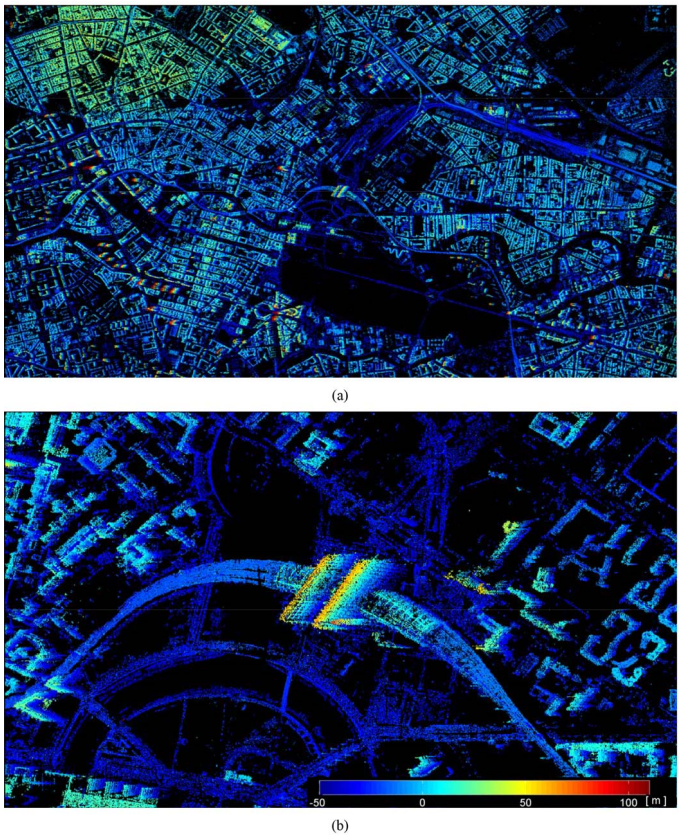
For a quantitative measure of the final classification, 98% of the pixels are correctly classified (the SL1MMER results in Fig. 3(b) as reference). The major classification error comes from the misdetection of pixels with double scatterers. This can be improved by allowing higher detection rates in the double scatterer detection. However, computation effort will also increase correspondingly since more pixels will be processed by higher order spectral estimation.

B. Analysis of the Computational Complexity

This section will discuss the computational complexity of the proposed approach. We define $O(1)$ to be the computation time for one multiplication. The discussion will be focused on the simple case without any motion model. To extend to a single- or a multicomponent motion model, the computational complexity exponentially increases with the number of dimensions in the system model of the problem.

In the first-order spectral estimation, the periodogram consists of multiplication of the measurement vector with the modeled phase and repeats at different elevation positions. Therefore, the computational complexity is $O(NL)$, where N is the number of images (usually 20–50), and L is the discretization level in the elevation direction (usually 50–200 for TerraSAR-X). In the higher order spectral estimation, SVD-Wiener needs at least $O(N^2 + NL)$ according to [4, eq. (11)]. The SL1MMER algorithm requires a L_1-L_2 norm optimization, which is extremely computationally costly, e.g., the well-known Basic Gradient Projection for Sparse Reconstruction solver [14] requires at least $O(KML_s^2)$, where K is the number of iterations (approximately 20), M is the number of multiplication of a specific matrix in each iteration (modest, usually several), and L_s is the elevation discretization level required for sparse reconstruction, which is about ten times greater than L. Our experience is that the SL1MMER algorithm is several hundred times slower than SVD-Wiener.

When prior knowledge of the elevation span is available, L can be restricted to a specific elevation range and hence reduce the computational cost. For instance, assuming a uniform distribution of building heights prior over the processing area, the total processing cost can be reduced to half for SVD-Wiener



(a)

(b)

Fig. 5. (a) Elevation estimates of Berlin using the proposed approach. (b) Close-up view over the Berlin central station.

when $L \gg N$ and to one third for SL1MMER. Therefore, it is always more economic to obtain the prior through some low-cost methods before the higher order spectral estimation.

The overall computational cost also depends on the percentage of pixels allocated to each algorithm. As an example, for the test area discussed in Section III-A, 100% of the pixels are first processed by a periodogram, 17% are then passed to SVD-Wiener, and only 9% are finally processed by SL1MMER. This small percentage together with the restriction of the search space leads to an overall speedup of about 50 compared with SL1MMER-only processing. The computational cost of the pixel preclassification using the SVM is negligible compared with the PSI plus TomoSAR processing and, hence, ignored in this analysis. Quicker processing can be achieved if no superresolution is required.

C. Large-Area Processing: Berlin

The proposed processing chain is applied on several stacks of TerraSAR-X high-resolution spotlight images over Berlin. Fig. 5 shows the elevation estimates and a close-up view of one of the stacks, which is composed of 94 images. In the close-up view of the Berlin central station, the proposed processing chain is able to catch most of the structure.

IV. Conclusion and Outlook

This letter has proposed an efficient TomoSAR inversion approach for meter resolution SAR image stacks of urban areas. It is the integration of three spectral estimation algorithms: periodogram, SVD-Wiener, and SL1MMER. It has been tested on a number of image stacks of urban areas. Reliable results are obtained, and a speedup factor of over 50 is confirmed. The computational cost is saved in two aspects: prior knowledge of the solution search range is obtained through low-cost periodogram, and only pixels with more than one strong scatterer are processed using SVD-Wiener, and SL1MMER is only applied to pixels that require superresolution. The resulting very high resolution (VHR) spaceborne TomoSAR point clouds can be used for multiple applications. For example, individual building monitoring over the whole city, fusion of point cloud from different angles [15], and city 3-D model reconstruction from the point clouds [16].

Acknowledgment

The authors thank the Gauss Centre for Supercomputing (GCS) e.V. for funding this project by providing computing time on the GCS Supercomputer SuperMUC at the Leibniz Supercomputing Centre (LRZ).

References

[1] 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.

[2] F. Lombardini, "Differential tomography: A new framework for SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 1, pp. 37–44, Jan. 2005.

[3] 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.

[4] 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.

[5] 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.

[6] B. Kampes, *Radar Interferometry: Persistent Scatterers Technique*. Dordrecht, The Netherlands: Springer-Verlag, 2006.

[7] 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, Jul. 2011.

[8] 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, Oct. 2010.

[9] N. Adam, B. Kampes, M. Eineder, J. Worawattanamateekul, and M. Kircher, "The development of a scientific permanent scatterer system," in *Proc. ISPRS Hannover Workshop*, Hannover, Germany, 2003, pp. 1–6.

[10] G. Fornaro, A. Pauciullo, and F. Serafino, "Deformation monitoring over large areas with multipass differential SAR interferometry: A new approach based on the use of spatial differences," *Int. J. Remote Sens.*, vol. 30, no. 6, pp. 1455–1478, Jan. 2009.

[11] A. De Maio, G. Fornaro, and A. Pauciullo, "Detection of single scatterers in multidimensional SAR imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2284–2297, Jul. 2009.

[12] X. Zhu and R. Bamler, "Demonstration of super-resolution for tomographic SAR imaging in urban environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 3150–3157, Aug. 2012.

[13] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, *A Practical Guide to Support Vector Classification*, 2010. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf

[14] M. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–597, Dec. 2007.

[15] Y. Wang and X. Zhu, "Feature-based fusion of TomoSAR point clouds from multi-view TerraSAR-X data stacks," presented at the Proceedings International Geoscience And Remote Sensing Symposium, Melbourne, Australia, Jul. 2013, Paper MO4.T04.4.

[16] X. Zhu and M. Shahzad, "Facade reconstruction using multiview spaceborne TomoSAR point clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, Jul. 2014. [Online]. Available: http://ieeexplore.ieee.org, to be published.

Appendices

B Wang Y, Zhu X (2015a) Automatic Feature-based Geometric Fusion of Multi-view TomoSAR Point Clouds in Urban Area. IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing, 8 (3): 953 - 965.

http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6942160

Automatic Feature-Based Geometric Fusion of Multiview TomoSAR Point Clouds in Urban Area

Yuanyuan Wang, *Student Member, IEEE*, and Xiao Xiang Zhu, *Senior Member, IEEE*

***Abstract*—Interferometric synthetic aperture radar (InSAR) techniques, such as persistent scatterer interferometry (PSI) or SAR tomography (TomoSAR), deliver three-dimensional (3-D) point clouds of the scatterers' positions together with their motion information relative to a reference point. Due to the SAR side-looking geometry, minimum of two point clouds from cross-heading orbits, i.e., ascending and descending, are required to achieve a complete monitoring over an urban area. However, these two point clouds are usually not coregistered due to their different reference points with unknown 3-D positions. In general, no exact identical points from the same physical object can be found in such two point clouds. This article describes a robust algorithm for fusing such two point clouds of urban areas. The contribution of this paper is finding the theoretically exact point correspondence, which is the end positions of façades, where the two point clouds close. We explicitly define this algorithm as "L-shape detection and matching," in this paper, because the façades commonly appear as L-shapes in InSAR point cloud. This algorithm introduces a few important features for a reliable result, including point density estimation using adaptive directional window for better façade points detection and L-shape extraction using weighed Hough transform. The algorithm is fully automatic. Its accuracy is evaluated using simulated data. Furthermore, the proposed method is applied on two TomoSAR point clouds over Berlin with ascending and descending geometry. The result is compared with the first PSI point cloud fusion method (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, 2012.) for urban area. Submeter consistency is achieved.**

***Index Terms*—Point cloud fusion, SAR tomography (TomoSAR), synthetic aperture radar (SAR), TerraSAR-X.**

Manuscript received May 30, 2014; revised August 22, 2014; accepted September 18, 2014. This work was supported in part by the International Graduate School of Science and Engineering, Technische Universität München (Project 6.08: "4D City"), in part by the Helmholtz Association under the framework of the Young Investigators Group "SiPEO" (VH-NG-1018, www.sipeo.bgu.tum.de), in part by the German Research Foundation (DFG, Förderkennzeichen BA2033/3-1), and in part by the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) by providing computing time on the GCS Supercomputer SuperMUC at Leibniz Supercomputing Centre (LRZ, www.lrz.de).

Y. Wang is with the Helmholtz Young Investigators Group "SiPEO", Technische Universität München, Munich 80333, Germany (e-mail: wang@bv.tum.de).

X. Zhu is with the Helmholtz Young Investigators Group "SiPEO", Technische Universität München, Munich 80333, Germany, and also with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Weßling 82234, Germany (e-mail: xiao.zhu@dlr.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2014.2361430

I. Introduction

A. TomoSAR

TOMOGRAPHIC SAR inversion (TomoSAR) [2]–[6] is an interferometric synthetic aperture radar (InSAR) technique that tracks down the long-term or partially coherent pixels in a stack of coregistered SAR images, which is similar to other InSAR techniques such as persistent scatterer interferometry (PSI) [7]–[10], small baseline subset (SBAS) [11]–[13], SqueeSAR [14], [15], and CEASAR [16]. They all aim at retrieving the three-dimensional (3-D) position and the parameters of the undergoing motion of point and/or volumetric scatterers.

However, fundamentally different, TomoSAR exploits possible multiple scatterers in the third dimension *elevation* of a rang-azimuth pixel—a phenomenon called *layover* in radar jargon that happens when two or more scatterers are at the same range (distance) to the sensor. For example, the closely packed vertical structures in urban areas cause severe layover. TomoSAR retrieves the elevations and motion parameters of multiple scatterers by means of spectral estimation. This is not possible with the above-mentioned single scatterer model-based methods, e.g., PSI. Therefore, TomoSAR is so far the most competent InSAR method for long-term urban monitoring [17].

Therefore, the 3-D TomoSAR point clouds derived from meter-resolution SAR image stacks are the basis of the work of this paper. The point clouds we used are obtained by *Tomo-GENESIS* [18]—a TomoSAR software of the German Aerospace Center (DLR) for large urban areas monitoring. It is developed based on the work of [5], [6], [17], [19]–[21]. It delivers, for the first time, a 5-D point cloud—3-D position plus linear deformation rate and amplitude of seasonal motion—from a stack of SAR images. The point density is up to $10^6/\mathrm{km}^2$, when using TerraSAR-X high-resolution spotlight data. Such density is comparable with some light detection and ranging (LiDAR) product and enables the retrieval of the features of individual building.

B. Fusion Geometry of Cross-Heading InSAR Point Clouds

Cross-heading orbits refer to satellite's ascending and descending flying trajectories. Due to the side-looking geometry of SAR, minimum of two TomoSAR point clouds from cross-heading orbits are required to achieve a complete monitoring over the whole area. As demonstrated in Fig. 1, point cloud derived from SAR image stacks of ascending or descending orbits provides either the front or the rear side of the same building. Such two point clouds have basically no *exact point*

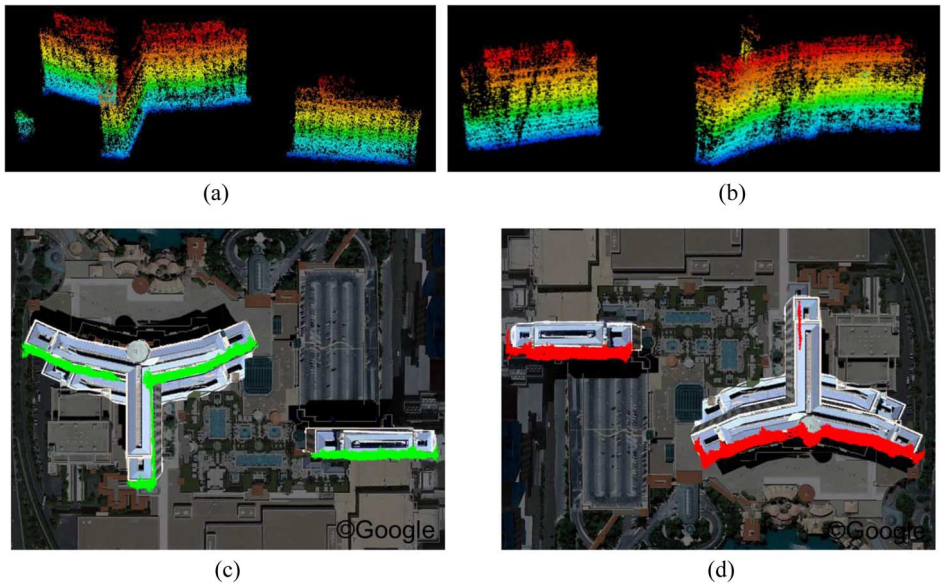


Fig. 1. TomoSAR point cloud derived from (a) an ascending image stack in Las Vegas and (b) a descending image stack of the same area. Color represents the height. (c) and (d) are the corresponding footprint of (a) and (b) on Google Map satellite images.

correspondence, which is referred to as the identical geolocation of the sources of two corresponding points (in the two point clouds, respectively). Examples are man-made objects like cylindrical reflector or 3-D corner reflector that could reflect radar signal in all directions. Not difficult to imagine, fusion of such point clouds is as tricky as coregistering two optical images of front and rear side of an opaque object.

These two point clouds are always relative to two different reference points with unknown heights. The uncompensated height of the reference point causes the point cloud to shift along the elevation direction. Therefore, the fusion task is to find the unknown heights of the two reference points. A schematic drawing of fusion in Universal Transverse Mercator (UTM) coordinate is illustrated in Fig. 2 (modified from [1]), where the left and right satellites indicate the ascending and descending master orbits at a fixed *east-up* plane, respectively. Perpendicular toward the reader's screen/paper is the *north* direction. The flying direction of the two satellites is indicated by the cross (fly away from reader) and dot (coming toward reader) on the satellites. They are not parallel to the north direction. The angle between the flying direction and the meridian (north) is the so-called heading angle. R^a and R^b are the range distances of the two orbits. The two black dots are the corresponding scatterers assumed to have identical position, but geocoded into different positions $\mathbf{P}^a_{xyz}$ and $\mathbf{P}^b_{xyz}$ due to their different and unknown reference height Δz^a and Δz^b. This also causes horizontal shifts Δxy^a and Δxy^b, which are functions of the incidence angles θ^a and θ^b, i.e., the horizontal shift is not constant throughout the whole point cloud. To achieve a fusion, the two scatterers have to be shifted along their elevation directions by $\Delta S^a = \Delta z^a / \sin\theta^a$ and $\Delta S^b = \Delta z^b / \sin\theta^b$, respectively. The fusion model in UTM coordinate system is then [1]

$$\mathbf{P}^a_{xyz} + \frac{\Delta z^a}{\sin\theta^a}\mathbf{s}^a = \mathbf{P}^b_{xyz} + \frac{\Delta z^b}{\sin\theta^b}\mathbf{s}^b \tag{1}$$

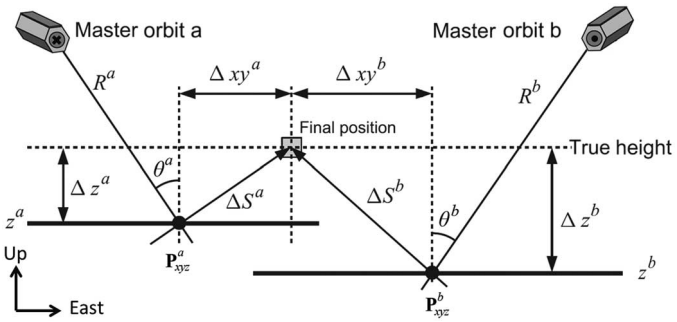


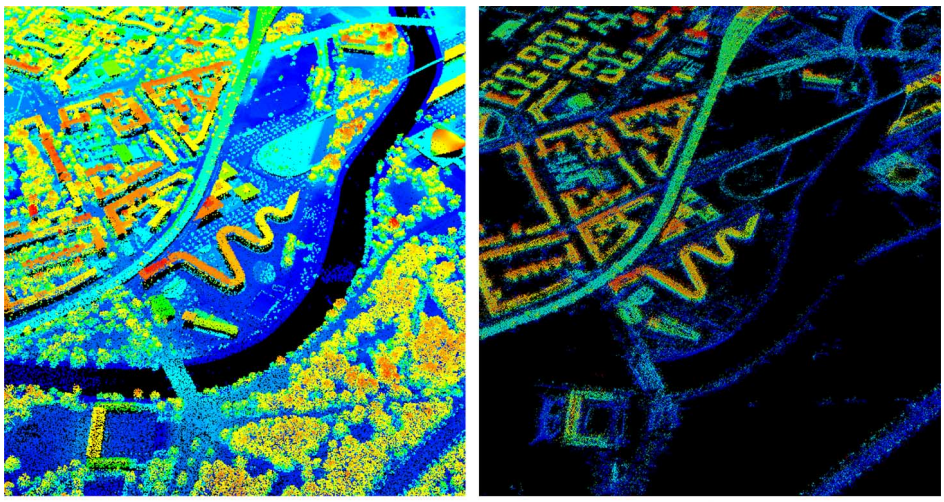
Fig. 2. Schematic representation (modified figure from [1]) of the InSAR point cloud shift in UTM coordinate due to different reference points. The reader's screen/paper is the east-up plane, and perpendicular to it away from the reader is the north direction. The two satellites indicate the ascending and descending master orbits positions. Their flying directions are indicated by the cross (fly away from reader) and dot (coming toward reader) on the satellites. They are not parallel to the north direction. The angle between the flying direction and the meridian (north) is the so-called heading angle. The two black dots are the two scatterers assumed to be identical in UTM coordinate system, but shifted in the two point clouds due to the heights Δz^a and Δz^b of their reference points.

where bold letters indicate vectors, and $\mathbf{s}^a$ and $\mathbf{s}^b$ are the unit vectors of the elevation directions of the two scatterers, which depends on the incidence angle θ, and heading angle t

$$\mathbf{s} = \begin{bmatrix} \cos t \cos\theta \\ -\sin t \cos\theta \\ \sin\theta \end{bmatrix}. \tag{2}$$

In another words, three equations can be written for each pair of corresponding points $\left(\mathbf{P}^a_{xyz}, \mathbf{P}^b_{xyz}\right)$

$$\begin{aligned} x^a - x^b + \Delta z^a \cos t^a \cot\theta^a - \Delta z^b \cos t^b \cot\theta^b = 0 \\ y^a - y^b - \Delta z^a \sin t^a \cot\theta^a + \Delta z^b \sin t^b \cot\theta^b = 0 \\ z^a - z^b + \Delta z^a - \Delta z^b = 0 \end{aligned} \tag{3}$$



(a) (b)

Fig. 3. Comparison of (a) LiDAR point cloud [22] and (b) TomoSAR point cloud of the same area in downtown Berlin. The color represents the height. LiDAR point cloud has, in general, better quality in terms of 3-D positioning accuracy.

where Δz^a and Δz^b are the only unknowns. However, the incidence and heading angles θ and t are direct functions of the point current position. As described in [1], this problem is the so-called *Gauß-Helmert* model. It needs to be solved iteratively, where $\mathbf{P}_{xyz}^{\{a,b\}}$, $\theta^{\{a,b\}}$, and $t^{\{a,b\}}$ are updated at each iteration. The estimation is robust with many pairs of points available.

C. Related Work and Motivation

In computer vision, the 3-D rigid point cloud coregistration methods are very well studied. For a successful coregistration, overlapping of the two point clouds is required, so that one can find common points or features that describe a common location in the point clouds. The basic approach of such task follows the procedures of feature detection, description, and matching, where so do the researches focus. The feature detection and description are usually combined. The most well-known ones are:

1) gradient-based features, e.g., scale invariant feature transform (SIFT) [23] and speeded up robust features (SURF) [24];
2) histogram-based features, e.g., histogram of gradients (HoG) [25] and point feature histogram (PFH) [26]; and
3) surface normal-based features, e.g., spin-image [27] and normal-aligned radial feature (NARF) [28].

Here, we just name a few. Many of their variants are also highly used. The matching algorithms, according to the optimization methods used, can be classified as either global or local. As pointed out in [26], the most famous algorithms of the first type are generic and evolutionary algorithms [29], [30]. Not surprisingly, most of the study is the second type, due to the computational complexity of the first type. The most well-known one is the iterative closest point (ICP) [31], [32]. Other highly cited works are mostly variants or improvements in the original ICP. For instance, to find better point correspondence using feature descriptors [26], [33], [34], to find better initial transformation [34], [35], and to use nonlinear optimization methods in ICP [36], [37].

In the remote sensing community, these techniques are employed for coregistration of LiDAR or optical image derived point clouds [38]–[42]. However, no attempt has been made for fusion of PSI or TomoSAR-derived point clouds of urban areas. This is especially true for two point clouds generated from data stacks of cross-heading orbits, due to the lack of exact point correspondence. Another reason prevents the above-mentioned methods from being directly applicable is the relatively worse accuracy of InSAR point clouds compared with LiDAR or optical image-derived ones. An example is made in Fig. 3, which is the comparison of LiDAR and TomoSAR point cloud of the same area in Berlin.

The first attempt of such task is described in [1] using PSI point clouds. It is essentially a 3-D surfaces matching that relays on removing the façade points, where the point correspondence is most unlikely to appear. The point correspondences are found by searching closely space point pairs on surface. Its accuracy is limited, since the exact point correspondence is not addressed. The computational efficiency is compromised when applying to TomoSAR point clouds [43]. Practically, only 10% of the non-façade points are used, since a TomoSAR point cloud is usually much denser then PSIs'.

Therefore, the initiative of the proposed method is to make use of the rich façade information in TomoSAR point clouds in order to derive the theoretically exact point correspondence and to develop a computationally efficient algorithm that can handle large point clouds. Finally, we can obtain a 3-D point cloud associated with the movement information covering an entire city.

As no exact point correspondence can be found among the points of the two point clouds, we will have to look for higher level features. Studies such as [44]–[46] have already suggested that quadrilateral buildings appear as "L-shapes" in SAR images. As a result, the end points of the L-shape, i.e., the building edges where the two point clouds close, can be used as the

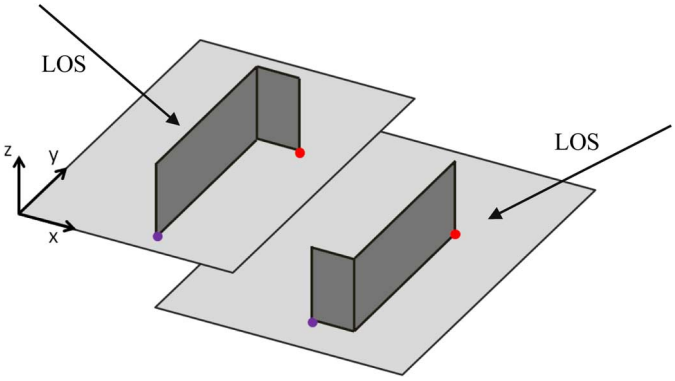


Fig. 4. Fusion geometry of the proposed L-shape method. The two building façades are colored in dark gray representing the geocoded point clouds from cross-heading orbits whose LOS directions are indicated by the two arrows. The light gray planes denote the corresponding ground level of the two point clouds. The two pairs red and purple dots are the façade end point position to be estimated.

exact point correspondence. Described in Fig. 4, the L-shape end points are indicated by the two pairs of red and purple dots. The two building façades colored in dark gray represent the point clouds from ascending and descending orbits whose line of sight (LOS) directions are indicated by the two arrows, respectively. The light gray planes denote the corresponding ground levels of the two point clouds.

The coregistration is then broken down to the task of detecting the L-shapes' end points, estimating their 3-D positions as the descriptor, and finally matching the two reduced point clouds. To achieve this, some robust procedures are designed to extract façade points, estimate the positions of the façade edges, and then find correspondence in the two point clouds.

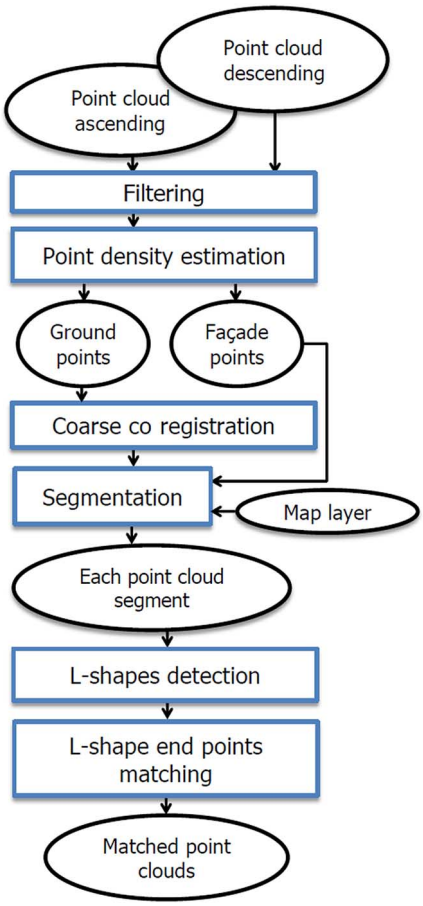


Fig. 5. Flowchart of the proposed algorithm. The black ellipses represent input/output data. The blue rectangles are procedures/processes.

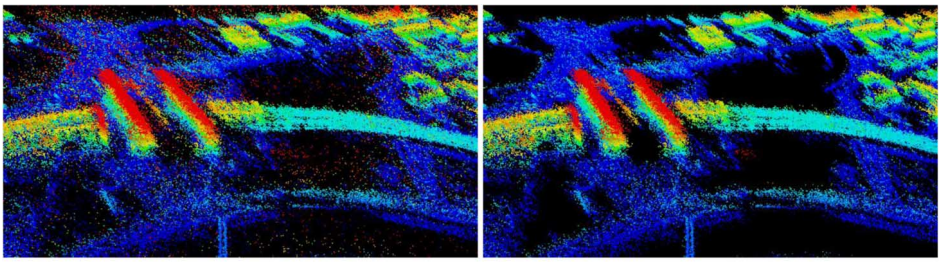
II. Feature-Based InSAR Point Cloud Fusion

The flowchart of the algorithm is plotted in Fig. 5. Starting from the raw TomoSAR point clouds, the outliers are filtered out using neighborhood analysis. The façade points are then extracted using point density estimation. Subsequently, each point cloud can be split into the façade part and the non-façade part. By cross-correlating their non-façade parts, the two point clouds can be coarsely coregistered within an accuracy of several meters, which gives an initial guess of the final shifting vector. The façade parts of the two point clouds are segmented into building blocks assisted by a two-dimensional (2-D) building shape layer from online map provider such as OpenStreetMap. After obtaining the segmented point clouds, L-shape is searched for each segment, and the positions of the end points of the L-shape are estimated. Therefore, the fusion of two complete point clouds are reduced to the fusion of two much sparser point clouds comprised of the virtual façade end points. Thus, the same random sample consensus (RANSAC) point correspondence matching procedures introduced in [1] can be applied. The final shifting vectors are estimated using (3).

A. Point Cloud Filtering

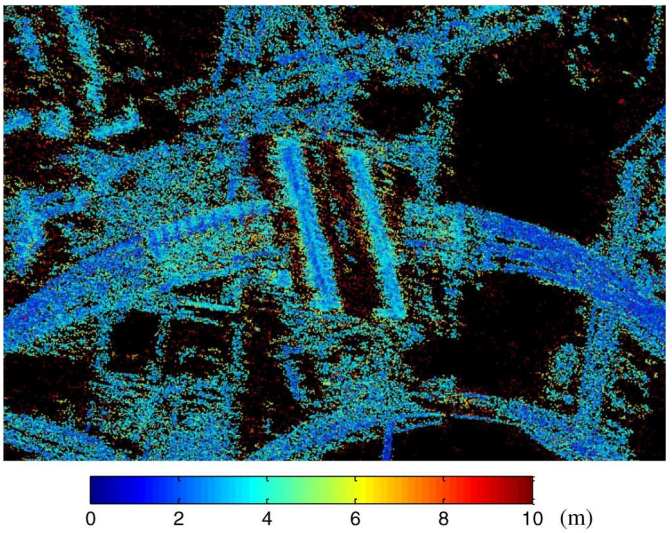
As mentioned in Section I-C, the overall geolocalization quality of a raw TomoSAR point cloud is subject to the high dynamic range of the SAR image pixels' SNR. The criteria in model order selection for rejecting noncoherent pixels are usually set to be very soft [5], in order to preserve as much information as possible from the expensive data stacks. This leads to some amount of outliers in the resulting raw TomoSAR point clouds. Fig. 6(a) is an example of a raw TomoSAR point cloud of Berlin central station. Outliers appear as single point or small clusters elevated from the main cluster nearby. They usually have less neighboring points and experiencing larger distances to their nearest neighbor.

Therefore, the outlier filtering can be done based on the neighborhood analysis of each point in the raw point cloud. We use the mean distance to k-nearest neighbors as the single criteria, i.e., for each point, we look for k points that are closest to it, and calculate the mean distance. Points with mean distance larger than a certain threshold are discarded. For instance, in Fig. 7(a), the mean distances of the points in Fig. 6(a) to their 20 nearest neighbors are plotted as a top view. The mean distance is color coded according to the color bar. Closely clustered points have small distances showing in blue, while outliers show larger distances and are usually displayed in red. Fig. 7(b) is the histogram of the log10 of this mean distance. The histogram shows two major distributions: 1) the left one refers to the main structures, and 2) the right one corresponds to points away from the main structures. The outliers can be removed

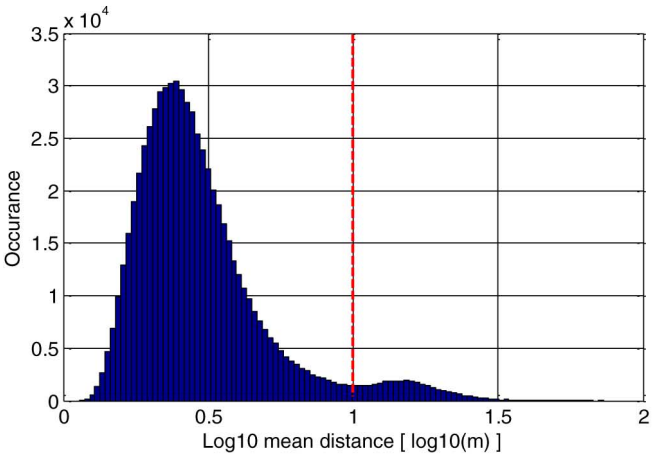


(a) (b)

Fig. 6. (a) Raw point cloud of Berlin central station and (b) filtered point cloud. Color represents the height.



(a)



(b)

Fig. 7. (a) Mean distances of the points in Fig. 6(a) to their 20 nearest neighbors and (b) the histogram of the log 10 of this mean distance. The red dashed line indicates the cutoff distance which is 10 m.

by thresholding (indicated by the red dashed line) on the mean distance. As shown in the filtered point cloud in Fig. 6(b), most of the outliers are effectively removed. We found a threshold at 10 m and using 20 nearest neighbors are effective for TomoSAR point cloud derived from TerraSAR-X high-resolution spotlight image stacks.

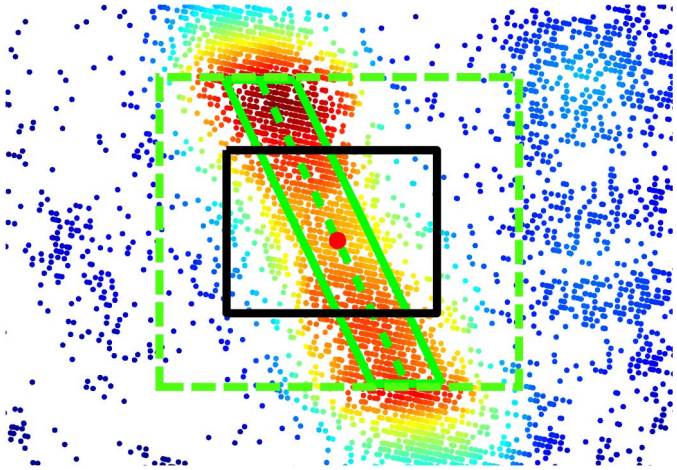


Fig. 8. Schematic drawing of point density estimation using directional window. The colored points are the projection of the 3-D point cloud in *east–north* plane, with the color indicating the point density. Blue indicates low density and red indicates high density. The color bar is not explicitly given for simplicity. The green-dashed line segment is a robust fit to the 2-D projection in a selected area. Points within the region bounded by the solid green lines are considered in the directional point density estimation, while those within the black box are considered in the box-car density estimation.

B. Point Density Estimation Using Directional Window

The point density per square meter is employed for façade point detection. Because of the vertical accumulation of the points on the façade, their point density is usually much higher over other areas. In order to better distinguish the façade points from other, we introduce a point density estimation method using an adaptive directional window, instead a moving box.

The approach is illustrated in Fig. 8. It shows the projection of a point cloud in the horizontal plane, with the color indicating the point density, and the red dot being the target point whose point density is going to be estimated. The estimation follows three steps: 1) points within an area (the green-dashed rectangle) centered at the target point is selected; 2) a straight line segment (the green-dashed line) passing through the target point is fitted using robust algorithm, e.g., RANSAC [47], robust principle component analysis [48], etc.; and 3) only the inliers (within the solid green parallelogram) of the line segment are counted for the point density, instead of counting all the points in the black rectangle.

For comparison, Fig. 9(a) and (b) shows the point density result using directional and rectangle window, respectively, for

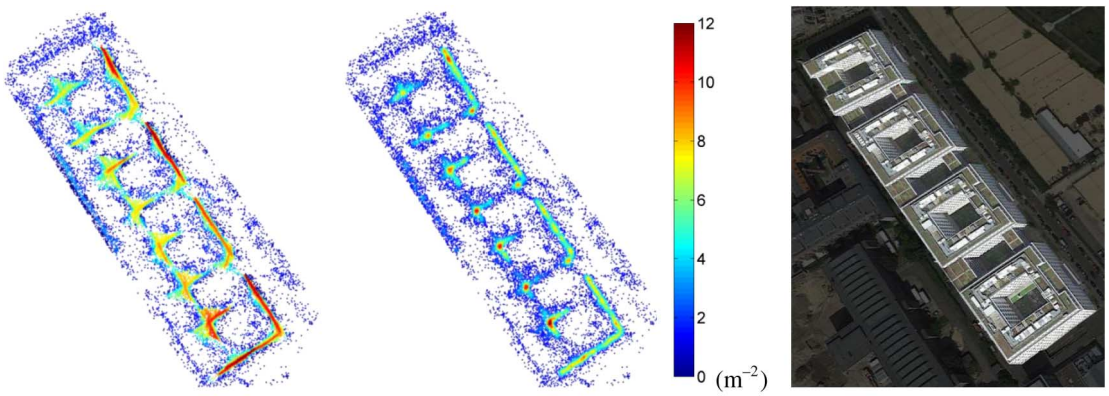


Fig. 9. Comparison of point density estimated using (a) directional window and (b) rectangle window. The color bar applies for both (a) and (b). For reference, (c) is the optical image of the same area from Google. The area of directional and rectangle windows are identical.

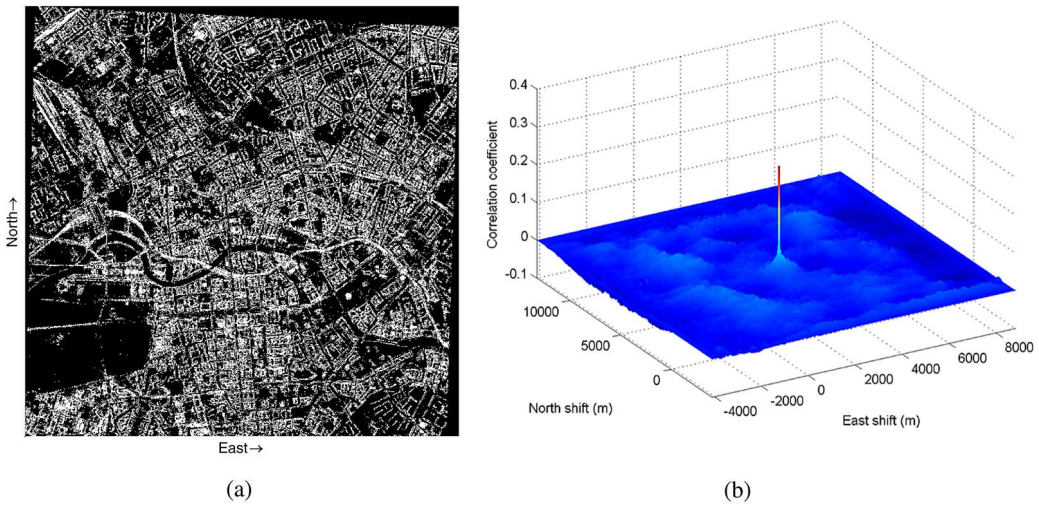


Fig. 10. (a) Binary image of the 2-D footprint of ascending point cloud of Berlin and (b) the 2-D cross correlation coefficient of the footprints of ascending and descending point clouds. The position of the peak value is the estimated coarse shift between the point cloud and the map layer.

a test area near Berlin central station. Since the same color bar is applied to the two figures, it is easy to tell that the directional window emphasizes the façade point, meanwhile keeps the point density of non-façade points unchanged compared with rectangle window. Since the urban buildings mostly consist of linear facades, our method works quite well in emphasizing façade points. The final façade points are selected by thresholding on the point density.

C. Coarse Coregistration

The point clouds needs to be coarsely coregistered to obtain an initial solution. This is a common practice in point cloud coregistration, which limits the search space in the refine coregistration step.

The coarse 3-D coregistration is done separately in horizontal and vertical directions. Horizontally, we cross-correlate the 2-D footprints of the non-façade parts of the two point clouds; and vertically, by simply aligning the mean height of the two point clouds. The 2-D footprints are rasterized binary images with a sampling spacing of a few meters. For example, Fig. 10(a) is the binary image of the ascending point cloud of Berlin. The descending one will look similar. Their 2-D cross-correlation coefficient is in Fig. 10(b), with brighter pixel being higher correlation. The position of the peak value is the coarse shift amount.

D. Point Cloud Segmentation

Since L-shape is a local feature of each building, the two point clouds are segmented to building blocks. Since the segmentation only needs to distinguish different building blocks, and it is also not the main focus of the algorithm, we make use of the building shape layer [49] from OpenStreetMap for assistance.

The building shape layer is a binary mask of the buildings [see Fig. 11(a)]. It is naturally not aligned with the point clouds footprints due to the shifted position of the point clouds, and sometimes the incorrect geolocation of the map itself. The alignment of the building shape layer with the point clouds 2-D footprint is done automatically using the same 2-D cross correlating technique explained in the previous section. After the

(a) (b) (c)

Fig. 11. (a) Building shape layer extracted from OpenStreetMap, (b) labeled individual buildings, and (c) segmentation mask derived from (b). The color assigned to each segment is random.

alignment, each building in the mask is labeled by checking their connectivity, yielding Fig. 11(b), where a random color is assigned to each building for better visualization. The unlabeled areas, i.e., the white region in Fig. 11(b), are assigned to their nearest building, giving the final segmentation mask in Fig. 11(c).

E. L-Shape Detection Using Weighted Hough Transform

For each point cloud segment, the façade part is selected, and one or zero L-shape is detected from its projection in the *east–north* plane. The detection is achieved by catching two interconnected line segments using weighted Hough transform. The proposed weighted Hough transform sums up the density of the points inside a Hough bin $\Delta\theta$ and Δr, instead of only counting the number of points, i.e., $H = \sum_{i=1}^{N} d_i$, with H being the Hough transform value, N the number of points in the Hough bin, and d_i the point density of each point. This concept is also illustrated in Fig. 12, where the green lines bound the Hough bin, and the size of the points inside the Hough bin denotes their point density. Both $\Delta\theta$ and Δr are exaggerated for visualization purpose. Standard Hough transform gives the same value for the two bins, but the proposed weighted Hough transform favors the right bin, since its point density is generally higher.

After the weighted Hough transform, the brightest pixel in the Hough matrix is first extracted. Its corresponding line is the most prominent one in the point cloud segment. It is taken as the first line of the L-shape. The neighborhood value of this pixel is suppressed in order to keep a constraint on the minimum angle of the L-shape. Among the rest, a few pixels with the highest amplitude are selected as the candidates of the second line. The second line shall be the one that connects the first one and forms the longest continuous contour. To determine the length of the two line segments, their intersection points are calculated using their equations, and the two end points are selected by locating the last point (with threshold density) passing by the lines, respectively. Nevertheless, such end points position estimation only gives a preliminary estimation of the length of the line segments. Therefore, it needs to be refined for better

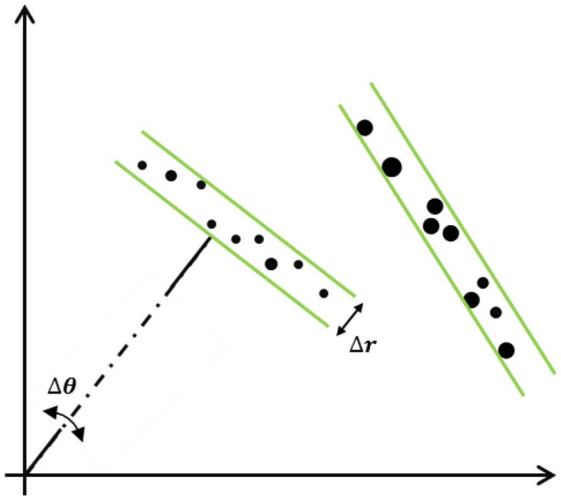


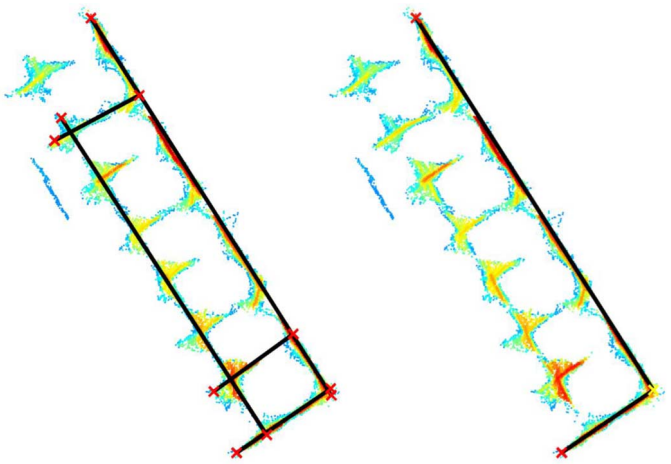
Fig. 12. Illustration of weighted Hough transform. The green arcs bounds the region of a Hough bin with spacing of $\Delta\theta$ and Δr, which is exaggerated for visualization. The dots inside the Hough bins are the point cloud projection in the horizontal plane. The size of the dot represents its point density, with larger the size higher the density. Both Hough bins have nine points, which gives the same value in Hough transform. However, the right bin has higher value in weighted Hough transform.

accuracy, which is explained in the Section III. The L-shape detection is applied on each segment of the two point clouds.

Global constraints are also put on

1) the minimum line segment length $l_{\min}$, according to the knowledge of minimum façade length;
2) the minimum Hough value $H_{\min}$, which can be calculated as $H_{\min} = \Delta r\, l_{\min}\, d_{\min}^2$, where $d_{\min}$ is the minimum point density of an identifiable façade in the point cloud; and
3) the opening direction of the L-shape depending on the orbit.

As an example, Fig. 13 shows the detected line segments and the final L-shape of the same point cloud used in Fig. 9. The detected lines are overlaid on the point density. In Fig. 13(a), the longest line segment is fixed as the first line of the L-shape, with the rest being the candidates of the second line segment.



(a) (b)

Fig. 13. Example of (a) detected line segments and (b) the final L-shape, using the same point cloud as in Fig. 9. The black lines are detected line segment. The line segments are overlaid on the point density. The roughly estimated end points are marked as red crosses. The estimation of their precise positions is explained in the next section.

Among them, the proposed algorithm automatically determines the second line by checking the connectivity with the first line and the total length of the L-shape. In the final detected L-shape, the intersection point is marked as yellow cross and the end points are marked as green.

F. Precise L-Shape End Points Position Estimation

The position of L-shape end point includes its horizontal 2-D coordinates and its height. The height is defined as the position where the façade intersects the ground plane, as depicted already in Fig. 4.

The horizontal coordinates are estimated using a model-based approach. Based on the detected L-shapes, the inlier points of the L-shape are projected to their corresponding line segments, i.e., façade direction. That is to say, we select only the points within a certain perpendicular distance to the L-shape and separate the two arms of the L-shape. One arm of the L-shape with inlier point should look like Fig. 14(a), where the black line is the estimated façade direction from the weighted Hough transform. For each arm of the L-shape, a projection should be obtained. The projection is basically the one-dimensional (1-D) coordinate of the façade points together with the non-façade points in the façade direction. Fig. 14(b) is the projected coordinates of the inlier points in Fig. 14(a). The middle section with much higher density belongs to the façade part. Therefore, its point density can be modeled as a rectangle function plus a constant offset for most of the urban façade.

An estimate of its point density can be obtained by convolving it with a rectangle window. For example, Fig. 14(c) shows the point density estimate of the points shown in Fig. 14(b). Its system model can be written as

$$f(x) = \int \left(A \operatorname{rect}\left(\frac{\tau - a}{2b}\right) + c \right) \operatorname{rect}\left(\frac{x - \tau}{2W}\right) d\tau \quad (4)$$

where $f(x)$ is the estimated point density as a function of the position x in the façade direction; A, a, b, and c are the unknown parameter of the exact point density model, where A models the façade point density, a is the façade center position, b is the half façade length, and c is the point density of the non-façade part. Finally, W is the known half-width of the rectangle filter. Since the convolution of two rectangle functions is a trapezoid, (4) can be written as

$$f(x) = \begin{cases} 2cW & x \le a - b - W \\ A(x - a + b + W) + 2cW & a - b - W < x \le a - b + W \\ 2(A + c)W & a - b + W < x \le a + b - W \\ -A(x - a - b - W) + 2cW & a + b - W < x \le a + b + W \\ 2cW & x > a + b + W. \end{cases} \quad (5)$$

Equation (5) shows that the left and right sides of the trapezoid can fully characterize the unknown parameters A, a, and b. Therefore, a moving fitting method is introduced to robustly fix the equation of the two sides. We use a moving window with the same width as the rectangle filter. A straight line is fitted to the points within the window. For any trapezoidal shape, high slopes should be detected at the position of the left and right sides. By detecting the position of the two high slopes, one can determine the position of the two sides. For robustness, the product of the absolute value of the slope and the number of inliers is considered in the detection of the two sides, because the maximum number of inliers should be reached when the window centered at the left or the right side. Fig. 14(d) shows the product of the absolute value of the slope and the number of inliers for the point density curve in Fig. 14(c). Two prominent peaks can be detected. The positions of the peaks are the direct estimates of the center positions of the left and right sides that are equal to $a - b$ and $a + b$. Hence, they are also the estimates of the start and end position of the rectangle function that models the façade.

To ensure reliable estimation, two regularizations are performed: 1) the coarse façade length obtained from the weighted Hough transform is used as a prior to regulate the estimation; and 2) the difference in the absolute value of the two slopes is restricted, because they must be both close to A according to the model.

After obtaining the 2-D horizontal coordinates of the L-shape end point, we look for the neighboring points of the 2-D position in the ground level. The mean value of these points is an estimate of the intersection of the 2-D coordinates with the ground plane. It is taken as the height of the L-shape end point. A plane should be fitted to these points, if the local topography is not removed.

G. End Points Matching and Final Shift Estimation

At this step, a number of L-shape end points are found for both point clouds. Therefore, the fusion of two complete point clouds are reduced to the fusion of two much sparser 3-D point clouds comprised of only the façade end points. For robustness, RANSAC is employed to find the maximum number of matched point pairs, which is also used in [1]. More details can be found in [50]. Briefly speaking, at each iteration of

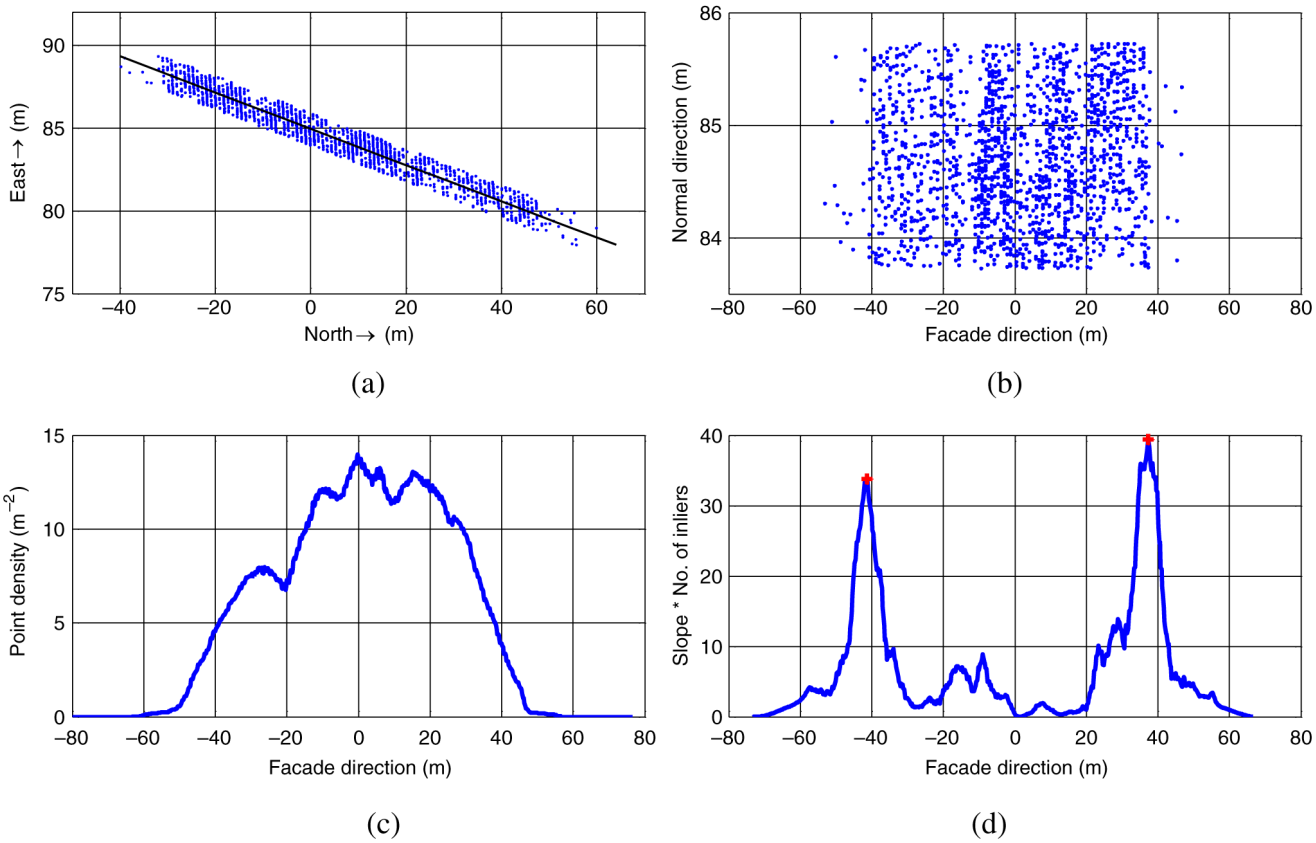


Fig. 14. (a) The inlier points of one arm of an L-shape, with the black line being the façade direction obtained from the weighted Hough transform, (b) the coordinates of the inlier points projected to the façade direction, (c) the estimate of (b)'s point density using a rectangle window, and (d) the product of the slope and the number of inliers of the fitted line using the points in the moving rectangle window along the point density estimate shown in (c). The two red crosses are the automatically detected façade start and end position at -41.38 and 37.22 m, respectively.

RANSAC, a pair of façade end points are selected from the two sparse point clouds, least square adjustment following (3) are performed, obtaining the solution of two height offsets Δz^a and Δz^b. The two sparse point clouds are then shifted along their elevation directions. The number of matched pairs, i.e., pairs of points at close range, is then counted. The final shift is given by the one with the maximum number of matched pairs.

III. Discussion

A. Fusion Accuracy

The final fusion accuracy directly depends on the number of matched pairs N_{pairs} of façade end points, and the accuracy σ_{point} of these end point estimates, i.e., $\sigma_{\text{final}} = \sqrt{2}\sigma_{\text{point}}/\sqrt{N_{\text{pairs}}}$, where $\sqrt{2}\sigma_{\text{point}}$ is the accuracy of one pair of façade end points, according to the fusion model.

N_{pairs} depends on the quality of the façade end point position estimates which are jointly affected by the accuracy of input TomoSAR point clouds, the segmentation performance, the L-shape extraction accuracy, and the number of quadrilateral buildings in the scene. According to our experience using high-resolution TerraSAR-X data, the number of quadrilateral buildings in the scene plays the most crucial role. Their typical range is a few hundreds for European cities like Berlin. We will exemplify this using real data in Section IV-B.

To analyze accuracy σ_{point} of the façade end points estimation algorithm explained in Section II-F, we test it using simulated data. It is generated by distributing points long the façade direction, with a constant point density on the façade part, and also a constant, but much lower point density on the non-façade part. The simulated data should look like Fig. 14(b). Important to note that the elevation accuracy σ_s (projected on façade direction) of the TomoSAR point cloud will affect the spatial distribution of the points. In the extreme case, when the TomoSAR point cloud is noise free, the distribution of the points should be evenly spaced in the façade and non-façade part, respectively.

We set total data length to be 40 m, façade length to be 20 m spanning from 12.00 to 32.00 m. The façade points are simulated five times, with the density being 5, 10, 15, 20, and $25/\text{m}^2$, respectively. The point density of non-façade area is always kept at $1/\text{m}^2$, and the rectangle filter size to be 5 m. Gaussian noise are then added to these point to simulate the inherent elevation accuracy σ_s from the TomoSAR processing, with σ_s being 1 and 5 m, respectively. Therefore, total 10 Monte Carlo simulations of 10 000 realizations each were performed. The simulation result is shown in Fig. 15, which shows the end point accuracy improves with respect to increasing point density. The end point location accuracy σ_{point} is below 30 cm in the façade direction for a typical five-story high building, i.e., with façade point density being about $15/\text{m}^2$ in the case

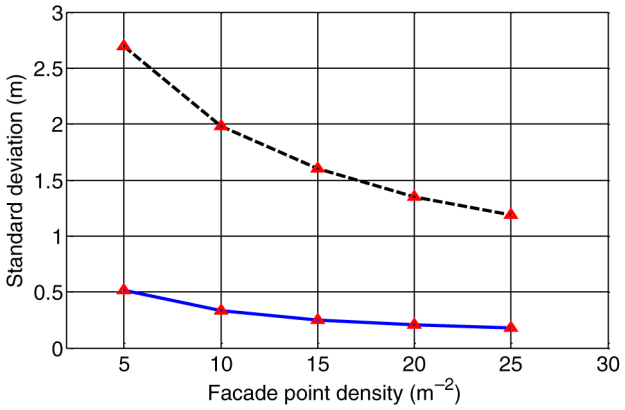


Fig. 15. Standard deviation of the end point estimates versus the façade point density. The non-façade point density is kept at $1/\mathrm{m}^2$. The black-dashed line corresponds to a point cloud accuracy of 5 m in the façade direction, and the blue solid line corresponds to the accuracy of 1 m.

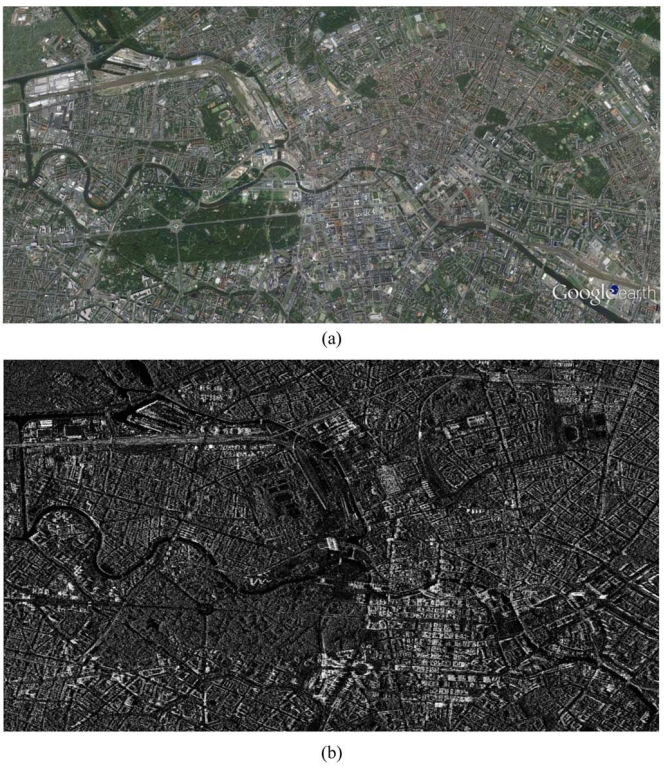
of meter resolution SAR data. The estimation accuracy also improves as the façade point density increases. For a typical urban area, the façade point density ranges from 5 to $25/\mathrm{m}^2$, which corresponds to an accuracy of 50–20 cm. The final fusion accuracy will be typically varying from 5 to 30 cm, if we set N_{pairs} to be 100.

One should be aware of that the 3-D coordinates of a real TomoSAR point cloud are always discretized, due to the range, azimuth, and elevation sampling of 3-D SAR imaging. Therefore, the discretization in the façade direction also depends on the same facts, but in addition is subject to the heading angle of the satellite and the direction of the façade which is random. Consequently, the discretization along the façade direction is different for each façade. A simulation is performed for a discretization level of 0.5 m using the same aforementioned setting. No significant degradation of estimation accuracy is found.

B. Parameters Settings

The proposed algorithm involves several parameters setting, including the point cloud filtering, and the weighted Hough transform for L-shape detection. Although these parameters seems to be data dependent, they are actually very much according to the inherent scale of the urban structures, e.g., the street width, number of building floors, minimum building size, etc. They can be reasonably derived with some prior knowledge of the urban structure. In the following, we give the parameter setting for TomoSAR point clouds derived from TerraSAR-X high-resolution spotlight data.

1) *Point cloud filtering*: we found using 20–50 nearest points, and a threshold at 10–20 m are suitable.
2) *2-D cross correlation*: the sampling distance of raster image should be less than half of the typical street width. We use 3 m in our algorithm.
3) *Point density estimation*: linear filter length is 10 m, which is a reasonable value for a shortest façade length, and width is set to be close to data resolution, i.e., 1 m.
4) *L-shape detection*: $l_{\min}$ is set to 10 m, which is the shortest façade length, and $d_{\min}$ is set to $2/\mathrm{m}^2$, which corresponds to the point density of a one to two-story high building in our TomoSAR point cloud. This leads to $H_{\min} = 40$ ($\Delta r = 1\mathrm{m}$).
5) *Façade end points position estimation*: the rectangle filter size to be 5 m that corresponds to half of the shortest façade length.



(a)

(b)

Fig. 16. (a) Google optical image of the test area and (b) the incoherent average SAR amplitude of the test area.

IV. Application on Real Data

A. Dataset

The proposed algorithm was tested on two TomoSAR point clouds generated from an ascending and a descending stack of high resolution TerraSAR-X spotlight data of Berlin. The ascending stack comprised of 79 interferograms and the descending stack has 94. The InSAR stacking and TomoSAR processing were done by the *PSI-GENESIS* [8], [51] and *TomoGENESIS* [18] systems, the PSI and TomoSAR processing system of Remote Sensing Technology Institute of DLR. Fig. 16 is the optical image from Google Earth, and an incoherent average of SAR amplitude of the ascending stack.

B. Fusion Result

Each point cloud contains about 20 million points. Around 500 L-shapes are detected from each point cloud, corresponding to 1000 end points. Fig. 17(a) and (b) shows the detected L-shapes and end points of the two point clouds, overlaying on their gray-scale point density images, respectively. The matched L-shapes end points are shown in Fig. 17(c).

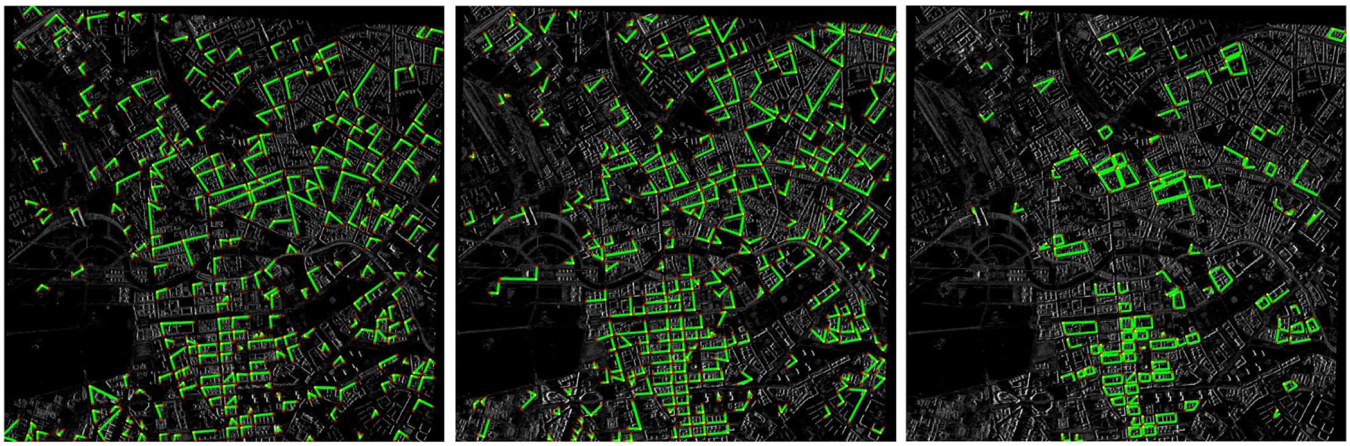


Fig. 17. (a) and (b) shows the detected L-shapes and their end points of the two input point clouds over Berlin and (c) the matched L-shapes and end points. The L-shapes are plotted in green, and the end points are marked as red. They are overlaid on the gray scale point density image.

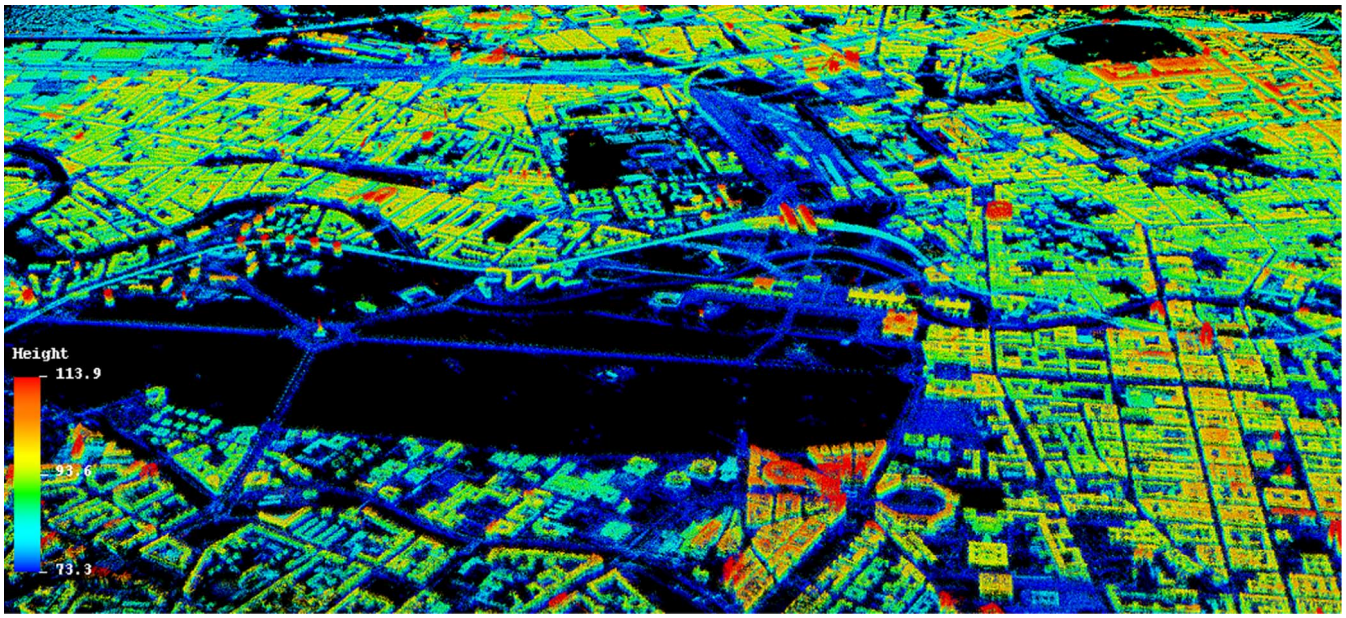


Fig. 18. Fusion result of ascending and descending point cloud of Berlin. The color indicates height on WGS84 reference surface. The unit of the color bar is meter.

Closed quadrilateral implies both pairs of end points are identified, a single L-shape implies only one pair of end points is identified. In total, 150 pairs of end points are matched, which is sufficient to robustly estimate the two unknowns Δz^a and Δz^b.

This result is compared with the result computed using the method explained in [1]. A difference of 0.51 m in *east* direction; 0.09 m in *north* direction, and almost no difference in height is found. A 3-D view of the complete fusion result is shown in Fig. 18. The height of the points is color-coded. The fused point cloud provides a Google-Street-View like visual sensation of the study area with very high level of detail. Such shadow-free TomoSAR point clouds are very useful for dynamic city model reconstruction [52] and scientific visualization [53], [54].

V. Conclusion and Outlook

This paper introduced a robust method to fuse two TomoSAR point clouds derived from meter-resolution SAR data stacks of cross-heading orbits, i.e., ascending and descending orbits. The fusion is done by finding the L-shaped buildings from the two point cloud, and matching the end points of the two L-shapes belong to the same building.

The proposed algorithm was successfully tested on TomoSAR point clouds obtained from two stacks of high-resolution TerraSAR-X spotlight data. It is also compared with the fusion result using the only existing method [1] for fusing PSI point clouds derived from meter-resolution SAR data stacks of urban areas. Submeter consistency is achieved. Compared with [1], the computational speed has been greatly improved,

due to the reduced point cloud in the final RANSAC matching step. The fusion accuracy is theoretically better, since the matched pairs of points are theoretically the exact correspondence. The simulation shows the estimation accuracy of façade end point ranges from 20 cm to a few meters, depends on the façade point density (i.e., building height) and the quality of the TomoSAR point cloud itself.

The proposed method can be applied to point cloud with point density up to 10 times lower than the TerraSAR-X high resolution data, e.g., stripmap data. In order to extract sufficient number of buildings, instead of setting a threshold on point density only, a soft threshold should be set first, and then use surface normal information of the points to further classify them. This method is explained and successfully applied in [55].

The proposed algorithm requires a point cloud segmentation step, which requires either a sophisticated method or external data source, so the authors employ GIS data for assistance. The final fusion accuracy also depends on the abundance of quadrilateral buildings in the scene. Therefore, to improve the proposed method, the following work is already undergoing: 1) automatic point cloud segmentation and 2) modeling building façades using piecewise line segment instead of using L-shape only.

ACKNOWLEDGMENT

The authors would like to thank L. Ding from TUM-LFK for providing the source of Berlin building shape layer.

REFERENCES

[1] 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.

[2] G. Fornaro, F. Serafino, and F. Soldovieri, "Three-dimensional focusing with multipass SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 3, pp. 507–517, Mar. 2003.

[3] F. Lombardini, "Differential tomography: A new framework for SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 1, pp. 37–44, Jan. 2005.

[4] 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.

[5] 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.

[6] 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, Jul. 2011.

[7] 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.

[8] N. Adam, B. Kampes, M. Eineder, J. Worawattanamateekul, and M. Kircher, "The development of a scientific permanent scatterer system," in *Proc. ISPRS Workshop High Resolut. Mapping Space*, Hannover, Germany, 2003, vol. 2003, p. 6.

[9] B. M. Kampes, *Radar Interferometry—Persistent Scatterer Technique*, vol. 12. Dordrecht, The Netherlands: Springer, 2006.

[10] S. Gernhardt, N. Adam, M. Eineder, and R. Bamler, "Potential of very high resolution SAR for persistent scatterer interferometry in urban areas," *Ann. GIS*, vol. 16, no. 2010–06, pp. 103–111, 2010.

[11] P. Berardino, G. Fornaro, R. Lanari, and E. Sansosti, "A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2375–2383, Nov. 2002.

[12] R. Lanari *et al.*, "A small-baseline approach for investigating deformations on full-resolution differential SAR interferograms," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1377–1386, Jul. 2004.

[13] 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.

[14] A. Ferretti *et al.*, "A new algorithm for processing interferometric datastacks: SqueeSAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 9, pp. 3460–3470, Sep. 2011.

[15] Y. Wang, 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.

[16] G. Fornaro, A. Pauciullo, D. Reale, and S. Verde, "Improving SAR tomography urban area imaging and monitoring with CAESAR," in *Proc. 10th Eur. Conf. Synth. Aperture Radar (EUSAR'14)*, 2014, pp. 1–4.

[17] X. Zhu, *Very High Resolution Tomographic SAR Inversion for Urban Infrastructure Monitoring—A Sparse and Nonlinear Tour*. Munich, Germany: Deutsche Geodätische Kommission, Verlag der Bayerischen Akademie der Wissenschaften, 2011.

[18] X. Zhu, Y. Wang, S. Gernhardt, and R. Bamler, "Tomo-GENESIS: DLR's Tomographic SAR processing system," in *Proc. Joint Urban Remote Sens. Event (JURSE)*, 2013, pp. 159–162.

[19] 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, Oct. 2010.

[20] X. Zhu and R. Bamler, "Demonstration of super-resolution for tomographic SAR imaging in urban environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 3150–3157, Aug. 2012.

[21] 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. 1–5, Jul. 2014.

[22] Data provided by "Land Berlin" and "Business Location Service," supported by "Europäischer Fonds für Regionale Entwicklung."

[23] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, 1999, vol. 2, pp. 1150–1157.

[24] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up Robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

[25] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR'05)*, 2005, vol. 1, pp. 886–893.

[26] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA'09)*, 2009, pp. 3212–3217.

[27] S. Ruiz-Correa, L. G. Shapiro, and M. Melia, "A new signature-based method for efficient 3-D object recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR'01)*, 2001, vol. 1, pp. I-769–I-776.

[28] B. Steder, R. B. Rusu, K. Konolige, and W. Burgard, "Point feature extraction on 3D range scans taking into account object boundaries," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2011, pp. 2601–2608.

[29] L. Silva, O. R. P. Bellon, and K. L. Boyer, "Precision range image registration using a robust surface interpenetration measure and enhanced genetic algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 762–776, May 2005.

[30] O. Cordón, S. Damas, and J. Santamaría, "A fast and accurate approach for 3D image registration using the scatter search evolutionary algorithm," *Pattern Recognit. Lett.*, vol. 27, no. 11, pp. 1191–1200, Aug. 2006.

[31] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.

[32] Z. Zhang, "Iterative point matching for registration of free-form curves and surfaces," *Int. J. Comput. Vis.*, vol. 13, no. 2, pp. 119–152, 1994.

[33] G. C. Sharp, S.-W. Lee, and D. K. Wehe, "ICP registration using invariant features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 1, pp. 90–102, Jan. 2002.

[34] N. Gelfand, N. J. Mitra, L. J. Guibas, and H. Pottmann, "Robust global registration," in *Proc. Symp. Geom. Process.*, 2005, vol. 2, p. 5.

[35] A. Makadia, A. I. Patterson, and K. Daniilidis, "Fully automatic registration of 3D point clouds," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Washington, DC, USA, 2006, vol. 1, pp. 1297–1304.

[36] A. W. Fitzgibbon, "Robust registration of 2D and 3D point sets," *Image Vis. Comput.*, vol. 21, no. 13–14, pp. 1145–1153, Dec. 2003.

[37] A. Gruen and D. Akca, "Least squares 3D surface and curve matching," *ISPRS J. Photogramm. Remote Sens.*, vol. 59, no. 3, pp. 151–174, May 2005.
[38] P. Gamba, F. Dell'acqua, and B. Houshmand, "Comparison and fusion of LIDAR and InSAR digital elevation models over urban areas," *Int. J. Remote Sens.*, vol. 24, no. 22, pp. 4289–4300, 2003.
[39] L. Bornaz, A. Lingua, and F. Rinaudo, "Multiple scan registration in LIDAR close range applications," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 34, pp. 72–77, 2003.
[40] W. Von Hansen, H. Gross, and U. Thoennessen, "Line-based registration of terrestrial and airborne LIDAR data," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 37, pp. 161–166, 2008.
[41] J.-J. Jaw and T.-Y. Chuang, "Registration of ground-based LiDAR point clouds by means of 3D line features," *J. Chin. Inst. Eng.*, vol. 31, no. 6, pp. 1031–1045, 2008.
[42] A. Gressin, C. Mallet, J. Demantké, and N. David, "Towards 3D lidar point cloud registration improvement using optimal neighborhood knowledge," *ISPRS J. Photogramm. Remote Sens.*, vol. 79, pp. 240–251, 2013.
[43] Y. Wang, X. Zhu, Y. Shi, and R. Bamler, "Operational TomoSAR processing using TerraSAR-X high resolution spotlight stacks from multiple view angles," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2012, pp. 7047–7050.
[44] E. Simonetto, H. Oriot, and R. Garello, "Rectangular building extraction from stereoscopic airborne Radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2386–2395, Oct. 2005.
[45] F. Zhang, Y. Shao, X. Zhang, and T. Balz, "Building L-shape footprint extraction from high resolution SAR image," in *Proc. Joint Urban Remote Sens. Event (JURSE)*, 2011, pp. 273–276.
[46] F. Zhang, L. Liu, and Y. Shao, "Building footprint extraction using dual-aspect high-resolution synthetic aperture radar images in urban areas," *J. Appl. Remote Sens.*, vol. 6, no. 1, pp. 063599–063599, 2012.
[47] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.
[48] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 11:1–11:37, Jun. 2011.
[49] OpenStreetMap shape layer provided for free by Geofabrik GmbH [Online]. Available: http://download.geofabrik.de/europe/germany, accessed on Oct. 29, 2012.
[50] 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, Jan. 2012.
[51] N. Adam, M. Eineder, N. Yague-Martinez, and R. Bamler, "High resolution interferometric stacking with TerraSAR-X," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS'08)*, 2008, vol. 2, pp. II-117–II-120.
[52] X. Zhu and M. Shahzad, "Façade reconstruction using multiview spaceborne TomoSAR point clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3541–3552, Jun. 2014.
[53] L. Ding, X. Zhu, and L. Meng, "Visual analysis of large amounts of 4-D building deformation data," in *Proc. 26th Int. Cartographic Conf.* (ISBN: 978-1-907075-06-3), Dresden, Germany, 2013, pp. 287–297.
[54] D. Liang, T. Balz, Z. Wang, L. Wei, and M. Liao, "Web-based interactive visualization of PS-InSAR point clouds for education and training," *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. II-6, pp. 7–12, Apr. 2014. Demo website [Online]. Available: http://sarviz.org/Psi/View/LasVegas
[55] M. Shahzad and X. Zhu, "Robust reconstruction of building facades for large areas using spaceborne TomoSAR point clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 752–769, Feb. 2015.

Yuanyuan Wang (S'11) received his B.Eng. (Hons.) degree in electrical engineering from The Hong Kong Polytechnic University, Hung Hom, Hong Kong, in 2008, and the M.Sc. degree in Earth Oriented Space Science and Technology (ESPACE) from Technische Universität München (TUM), München, Germany, in 2010. Since January 2011, he has been pursuing the Doctoral degree with the TUM-IGSSE research team "4D City."

Currently, he is with the Helmholtz Young Investigators Group "SiPEO." In June and July of 2014, he was a Guest Scientist at the Institute of Visual Computing, ETH Zürich, Switzerland. His research interests include optimal parameters estimation in InSAR techniques, multisensor fusion algorithms of SAR/optical data, and the applications of these techniques in urban and volcanic areas.

Xiao Xiang Zhu (S'10–M'12–SM'14) received the Bachelor's degree in space engineering from the National University of Defense Technology (NUDT), Changsha, China, in 2006, the M.Sc., Dr.-Ing. degrees, and the "Habilitation" in the field of signal processing from Technische Universität München (TUM), Munich, Germany, in 2008, 2011, and 2013, respectively.

During October/November 2009, she was a Guest Scientist with the Italian National Research Council (CNR), Institute for Electromagnetic Sensing of the Environment (IREA), Naples, Italy. Since May 2011, she has been a Scientist with the Remote Sensing Technology Institute, German Aerospace Center (DLR), Oberpfaffenhofen, Weßling, Germany, where she is the Head of the Team Signal Analysis, and with the Chair of Remote Sensing Technology, TUM. Since September 2013, she has been leading the Helmholtz Young Investigators Group "SiPEO" and is appointed as TUM Junior Fellow. Her research interests include modern signal processing, including innovative algorithms such as compressive sensing and sparse reconstruction, with applications in the field of remote sensing such as multi/hyperspectral image analysis; advanced InSAR techniques such as high-dimensional tomographic SAR imaging and SqueeSAR; and computer vision in remote sensing including object reconstruction and multidimensional data visualization.

Appendices

C Wang Y, Zhu X, Zeisl B, Pollefeys M (2015b) Fusing Meter-Resolution 4-D InSAR Point Clouds and Optical Images for Semantic Urban Infrastructure Monitoring. IEEE Transactions on Geoscience and Remote Sensing. in press.

Fusing Meter-Resolution 4-D InSAR Point Clouds and Optical Images for Semantic Urban Infrastructure Monitoring

Yuanyuan Wang [a], Xiao Xiang Zhu [a, b, *], Bernhard Zeisl [c], Marc Pollefeys [c]

[a] Signal Processing in Earth Observation (SiPEO) Technische Universität München, Arcisstraße 21, 80333 Munich, Germany.

wang@bv.tum.de

[b] Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, 82234 Weßling, Germany.

xiao.zhu@dlr.de

[c] Institute for Visual Computing, Department of Computer Science, ETH Zurich, CH-8092 Zurich, Switzerland.

marc.pollefeys@inf.ethz.ch

KEY WORDS: optical InSAR fusion, semantic classification, InSAR, SAR, railway monitoring, bridge monitoring

ABSTRACT:

Using synthetic aperture radar (SAR) interferometry to monitor long-term millimeter-level deformation of urban infrastructures, such as individual buildings and bridges, is an emerging and important field in remote sensing. In the state-of-the-art, deformation parameters are retrieved and monitored on a pixel-basis solely in the SAR image domain. But the inevitable side-looking imaging geometry of SAR results in undesired occlusion and layover in urban area, rendering the current method less competent for a semantic-level monitoring of different urban infrastructures.

This paper presents a framework of a semantic-level deformation monitoring by linking the precise deformation estimates of SAR interferometry and the semantic classification labels of optical images via a 3-D geometric fusion and semantic texturing. The proposed approach provides the first "SARptical" point cloud of an urban area, which is the TomoSAR point cloud textured with

* Corresponding author

attributes from optical images. This opens a new perspective of InSAR deformation monitoring. Interesting examples on bridge and railway monitoring are demonstrated.

1. INTRODUCTION

Monitoring long-term millimeter-level deformation of urban infrastructures is an important field in remote sensing. Synthetic aperture radar (SAR) interferometry is so far the only imaging-based method to assess the deformation in such accuracy. Credited to the launch of meter-resolution spaceborne SAR sensors [1], as well as the development of multipass interferometric SAR (InSAR) techniques such as persistent scatterer interferometry (PSI) and SAR tomography (TomoSAR), it is now also feasible to monitor individual buildings over a large area in high detail.

The state-of-the-art of multipass InSAR mainly leans towards the optimal parameters retrieval on a pixel-basis, i.e. the estimation of the 3-D position and the deformation parameters of scatterers. Moreover, the inevitable SAR side-looking imaging geometry results in undesired occlusion and layover especially in urban areas [2], rendering the SAR images difficult to interpret. For example, Figure 1 shows a comparison of the optical and SAR images of a same area (Potsdamer Platz, Berlin). The rich details in the optical image are buried underneath the strong reflection of the façade in the SAR image, i.e. superposition of the reflections from multiple sources in one pixel. This renders the current approach of pixel-based deformation analysis and manual identification of the region of interest less competent for understanding the semantics of the object being analyzed. Only until recently, the first investigation of semantic-level urban deformation analysis in InSAR has been presented in [3], [4].

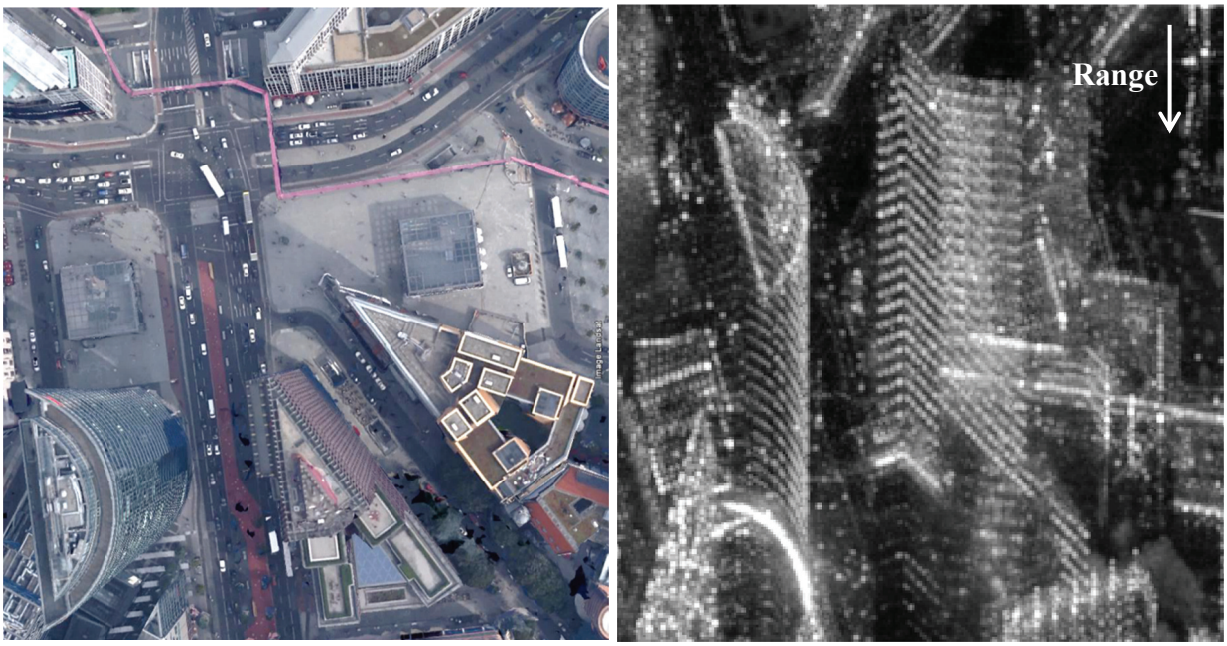


Figure 1. Comparison of optical image (left ©Google) and SAR image (right, TerraSAR-X) of the Potsdamer platz in Berlin. The rich details in the optical image are buried underneath the strong reflection of the façade in the SAR image, causing the so-called "layover", i.e. superposition of multiple reflection sources in one pixel.

This paper proposes a framework for semantic urban infrastructure monitoring using multipass InSAR and optical images. We aim to combine InSAR and optical images by complementing the high precision deformation estimates of InSAR with the high interpretability of optical images. The focus of this paper is put on linking the optical image to the SAR image by 3-D matching and projection.

The proposed framework leads to a few contributions to the remote sensing community. We addressed the 3-D co-registration of side- and nadir-looking point clouds in urban areas, e.g. nadir-looking optical and InSAR point clouds. We also addressed the texturing of InSAR point cloud with optical semantics without explicit reconstruction of 3-D model. Last but not least, the joint deformation analysis using optical and SAR images leads to new perspective in multipass InSAR.

2. METHDOLOGY OVERVIEW

The general framework of the proposed approach is shown in Figure 2. The framework applies to a stack of SAR images and a pair of (or more) optical images. The focus of this paper is put on linking the attributes from optical image to the SAR image by 3-D matching and projection. The basic idea is to match the 3-D models derived from SAR and optical images respectively. As a result, the 2-D SAR and optical images will also be matched. Only then, the following semantic label texturing and joint deformation analysis can be conducted. The detailed procedures of the proposed framework are as follows.

1. **3-D reconstructions**
    a. Retrieve the 3-D positions and deformation parameters of the scatterers from the SAR image stacks. Since urban area is of our main interest, tomographic SAR inversion (TomoSAR), including SAR tomography and differential SAR tomography, is employed in order to resolve a substantial amount of layovered scatterers.

       TomoSAR is the most computationally expensive step in the framework. In addition, TomoSAR and other multipass InSAR algorithms typically requires a fairly large SAR image stack (>20 images). The computational and image resource are the main limitation for this step.

    b. Retrieve the 3-D positions of points from the optical images using stereo matching with structure from motion (SfM) if necessary. For covering large urban area, aerial or spaceborne images are preferred. This step also calibrates the camera parameters.

       Stereo matching and SfM are well studied topics. Many matured algorithms and software are readily available.
2. **3-D matching**: Co-register the TomoSAR point cloud and the optical point cloud.

The main challenges present in this step are the different modalities of optical and TomoSAR

point clouds, i.e. nadir-looking and side-looking, as well as the relatively large anisotropic noise in the TomoSAR point cloud. However, considering the large amount of points compared to the few co-registration parameters to be estimated, the co-registration accuracy is expected to be high enough for the following steps.

If only a single optical image is available, which means no optical point cloud can be reconstructed, one can try to match higher level features of the 2-D optical image with the 3-D InSAR point cloud, e.g. matching 2-D and 3-D lines [5]. However, this has not been implemented in the proposed framework.

3. **Optical image classification**: applying semantic classification to the optical images.

   This part is not the focus of this paper. Depending on the application, different classification algorithms can be applied.

4. **Semantic texturing**: Texture the InSAR point cloud with the attributes derived from optical images, e.g. RGB color, semantic classification label, etc.

   The main challenge of this step is to project the optical image to TomoSAR point cloud without explicit 3-D surface reconstruction in the TomoSAR point cloud. Therefore, we choose point-based rendering technique.

   The main limitation of this step is the relatively poor positioning accuracy (1 to 10m) of spaceborne TomoSAR point cloud. This error will directly translate to the projection accuracy of the TomoSAR points in optical image.

5. **Deformation analysis**: Perform further analysis on semantic-level in the InSAR point cloud based on their semantic class.

   The main bottleneck of this step remains at the positioning accuracy of TomoSAR points, which limits the joint deformation analysis to be at an object level instead of at a point level.

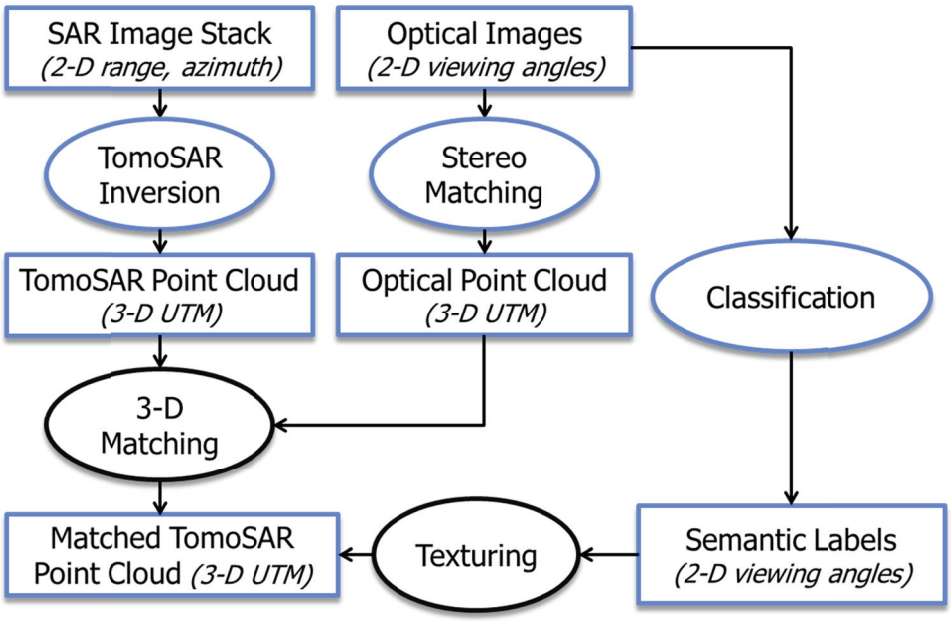


Figure 2. Flowchart of the proposed framework. It matches the 3-D TomoSAR and optical point clouds, and then transfers the semantic classification label from optical images to the TomoSAR point cloud via a point-based texturing. This paper focuses on the 3-D matching and texturing which are shown in black ellipses. The coordinate system of each dataset in the flowchart is indicated by the italic text in the bracket.

3. RELATED WORK

3.1 Differential SAR Tomography

Regarding urban area monitoring using multipass InSAR, TomoSAR in its differential formulation [6]–[9] is the most competent method because of its capability of separating layovered scatterers in one pixel and estimating their motion parameters. Figure 3 shows the SAR imaging geometry a range-azimuth pixel, where the reflections of multiple sources are layovered in one range-azimuth-elevation resolution cell.

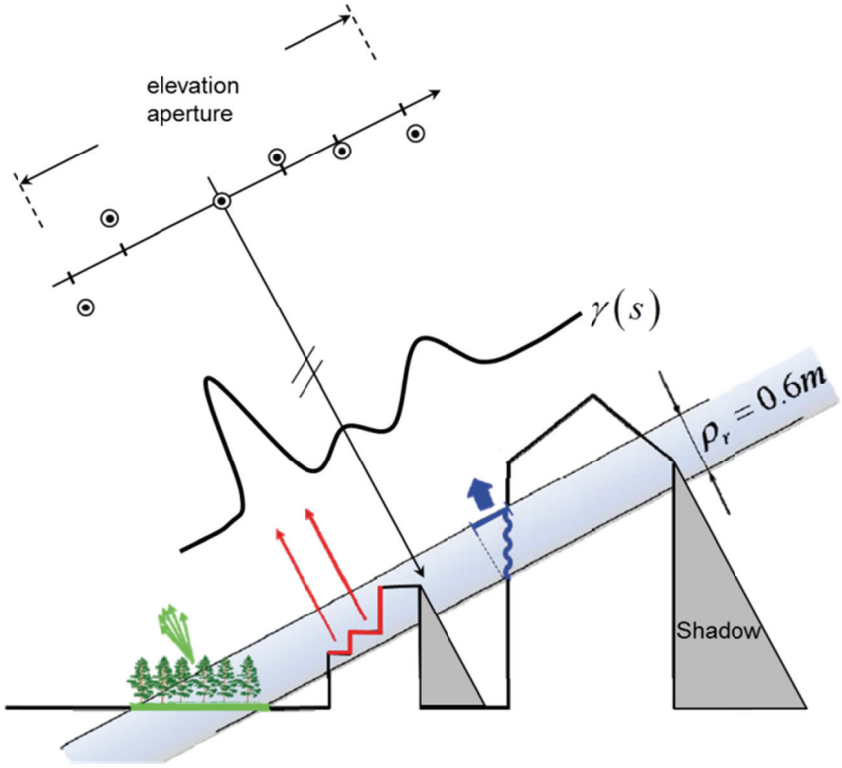


Figure 3. SAR imaging geometry (modified after [8]). Multiple scatterers are layovered in one resolution cell, which became one range-azimuth pixel. TomoSAR retrieves the whole reflectivity profile $\gamma(s)$ using multiple images acquired at slightly different antenna positions.

The reflectivity profile $\gamma(s)$ along the third dimension *elevation* s is integrated into one range-azimuth pixel, as shown by the following equation [9]–[11]:

$$g_n = \int_{\Delta s} \gamma(s) \exp(-j2\pi\xi_n s) \mathrm{d}s \tag{1}$$

where g_n is the complex SAR pixel value in the nth image, Δs is the extent of the elevation direction, and $\xi_n = 2b_n/(\lambda R)$ is the spatial frequency which is related to the baseline b_n of the nth image, the radar carrier wavelength λ, and the radar range R.

TomoSAR retrieves the full reflectivity profile $\gamma(s)$ from multiple observations. Equation (1) can be discretized as

$$\mathbf{g} = \mathbf{R}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{2}$$

where $\mathbf{g} \in \mathbb{C}^N$ is the observation vector, $\mathbf{R} \in \mathbb{C}^{N \times L}$ is the so-called steering matrix which is a discrete Fourier transform according to ξ_n and the discretization level L in the elevation direction, and $\boldsymbol{\varepsilon} \in \mathbb{C}^N$ is the noise assumed to be complex circular Gaussian distributed. The discretization level L is usually much larger than the number of images N. Therefore, equation (2) is often underdetermined, and can be even ill-posed due to the irregularity and small span of the baseline. Hence, regularization is required. The Tikhonov regularization of the estimator is equivalent to the maximum a posteriori (MAP) estimator with white noise and white prior, which is usually realized using singular value decomposition-based methods [8], [10].

From the reconstructed reflectivity profile, multiple scatterers can be separated, and hence the full 3-D (range, azimuth, and elevation) reflectivity distribution is obtained. The motion of scatterers can also be modelled and estimated by introducing new dimensions into the TomoSAR system model attributing to the motion of the scatterers [6]. For a motion model with M parameters, the extended TomoSAR system model is a M+1 dimensional discrete Fourier transform [12]:

$$\begin{aligned} g_n = & \int_{\Delta p_M} \cdots \int_{\Delta p_1} \int_{\Delta s} \gamma(s)\delta\left(p_1 - P_1(s), \cdots, p_M - P_M(s)\right) \\ & \exp\left(-j2\pi\left(\xi_n s + \eta_{1,n} p_1 + \cdots + \eta_{M,n} p_M\right)\right) \mathrm{d}s \mathrm{d}p_1 \cdots \mathrm{d}p_M \end{aligned} \tag{3}$$

where $p_1, \cdots, p_M$ are the M motion dimensions, and $P_1(s), \cdots P_M(s)$ are the motion dimensions as functions of the elevation s. $\eta_{1,n}, \cdots, \eta_{M,n}$ are the M motion models, e.g. $\eta_n = 2t_n/\lambda$ for linear motion model and $\eta_n = 2\sin\left(2\pi\left(t_n - t_0\right)\right)/\lambda$ for seasonal motion model, with t_n being the acquisition time in unit of year of the nth image w.r.t. the master image, and t_0 the initial offset also in unit of year.

3.2 Multi-view 3-D Reconstruction from Optical Images

Dense 3-D reconstruction from optical images is a classical topic in photogrammetry that is usually carried out by stereo matching. On the other hand, many algorithms were developed over the last decade to exploit the available unstructured image data. At the core of all algorithms are local image features which are matched to each other among the set of captured images. These relations allow recovering both, the camera motion (camera parameters) and the scene structure (3-D geometry), which is why these algorithms are referred to as *Structure-from-Motion* (SfM) [13]–[17]. Tremendous progress has been made and led to impressive results, even modelling whole cities from optical images only [13]–[15]. In the state-of-the-art, SfM is typically followed by a dense reconstructions step, e.g. semi-global matching [18], for 3-D reconstruction of urban area.

The principle of SfM is shown in Figure 4. If all the red dots in the image are known to correspond to the same 3-D point, given its 3-D position and the cameras parameters, one can predict its positions in the images. Consequently, one can estimate its 3-D position and the cameras parameters by adjusting their values so that the difference between the predicted positions of the 3-D points in the images w.r.t. the observed ones is minimized. For example, Camera C has a large reprojection error; hence, the 3-D point estimates and the camera parameters should be adjusted accordingly.

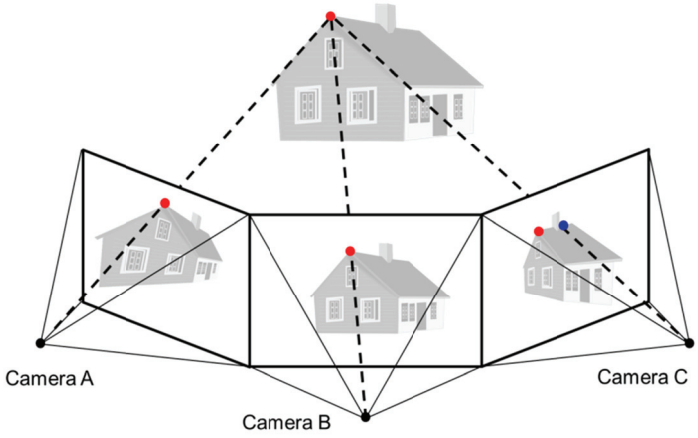


Figure 4. 2-D projections of a 3-D house in three different camera positions and orientations. The red dots are a known correspondence point between the three images. SfM estimates the unknown 3-D position of the point and the camera parameters which minimize the reprojection error (the predicted point location in image w.r.t the observed one, e.g. the blue point in Image C is the predicted location. Camera C has a large reprojection error.)

Therefore, the SfM system model can be formulated as follows. Given a set of corresponding image points, where we use $\tilde{\mathbf{x}}_{ij}$ to denote the *i*th points in image *j*, the projection from their real 3-D coordinates $\mathbf{x}_i$ is [13]

$$\tilde{\mathbf{x}}_{ij} = f_j \prod\left(\mathbf{R}_j\left(\mathbf{x}_i - \mathbf{t}_j\right)\right) \tag{4}$$

where f is the scalar camera focal length in number of pixels, $\mathbf{R}$, $\mathbf{t}$ are the camera orientation matrix and translation vector respectively, and $\prod$ is the projection function: $\prod(x,y,z) = (x/z, y/z)$. SfM is to estimate the unknown 3-D position and the camera parameters from the observations $\tilde{\mathbf{x}}_{ij}$. Assuming a Gaussian distributed reprojection noise, the maximum likelihood estimator of the parameters is [13]

$$\left\{\hat{\mathbf{x}}_i, \hat{\mathbf{R}}_j, \hat{\mathbf{t}}_j, \hat{f}_j\right\} = \arg\min_{\mathbf{x}_i, \mathbf{R}_j, \mathbf{t}_j, f_j} \sum_{i,j} \left\| \tilde{\mathbf{x}}_{ij} - f_j \prod\left(\mathbf{R}_j\left(\mathbf{x}_i - \mathbf{t}_j\right)\right)\right\|_2^2 \tag{5}$$

Sometimes, the camera radial and tangential distortions should also be taken into account, and should be estimated [16], [19].

3.3 Optical and SAR Images Co-registration

Geometric co-registration is the basis of many higher level fusion algorithms. The current techniques are mostly based on matching of 2-D spatial features, e.g. contours, edges, templates, etc. [20]–[25]. They are good for large scale analysis, but not applicable for the analysis of individual building, especially for high resolution data.

Therefore, for the co-registration of optical and SAR images of urban area including many studies of their joint analysis, a precise 3-D model is crucial. For example, [26] used 3-D model reconstructed from photogrammetric measurements; [27] used precise airborne light detection and ranging (LiDAR) point cloud; and [28] used PS point cloud retrieved from tens of high resolution TerraSAR-X spotlight images. Without a precise 3-D model, one needs to assume a strict prior [29], such as rectangle footprint etc.

Considering the complex 3-D topography of urban area, there has not been an effective algorithm of matching SAR and optical image yet. The method presented in this paper transform the 2-D images matching to 3-D point clouds co-registration. Handful of papers have explained the co-registration of very high resolution InSAR point clouds [30], [31]. In a broader context of 3-D point clouds co-registration, one of the most successfully methods is the iterative closest point (ICP) [32]. As its name implies, ICP solves the following equation at each iteration:

$$\left\{\hat{\mathbf{R}},\hat{\mathbf{t}}\right\} = \arg\min_{\mathbf{R},\mathbf{t}} \sum_i \left\|\mathbf{x}_i - \mathbf{R}\mathbf{p}_i - \mathbf{t}\right\|_2^2 \tag{6}$$

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

anisotropic noises in the point clouds, [33], [34] take into account the covariance matrix $\mathbf{C}_{\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i}$ of the residual $\boldsymbol{\varepsilon}_i = \mathbf{x}_i - \mathbf{R}\mathbf{p}_i - \mathbf{t}$:

$$\left\{\hat{\mathbf{R}},\hat{\mathbf{t}}\right\} = \arg\min_{\mathbf{R},\mathbf{t}} \sum_i \left(\mathbf{x}_i - \mathbf{R}\mathbf{p}_i - \mathbf{t}\right)^T \mathbf{C}_{\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i}^{-1} \left(\mathbf{x}_i - \mathbf{R}\mathbf{p}_i - \mathbf{t}\right) \tag{7}$$

4. CO-REGISTRATION OF TOMOSAR AND OPTICAL POINT CLOUDS

The 2-D SAR and optical image matching is transferred to 3-D point clouds matching. Any InSAR methods, including TomoSAR, deliver relative estimates with respect to a reference point whose height is usually unknown. Such differential measurement is always performed in multipass InSAR processing in order to mitigate some common error in each image, e.g. atmospheric effect or orbit errors. As a consequence, while being geo-coded into Universal Transverse Mercator (UTM) coordinate, the TomoSAR point cloud is shifted from its true position by an unknown amount due to the unknown height of the reference point. The co-registration is then the estimation of the translation between two rigid point clouds, subject to a certain tolerance on rotation and scaling. A good matching is the key to find the camera position w.r.t. the SAR image.

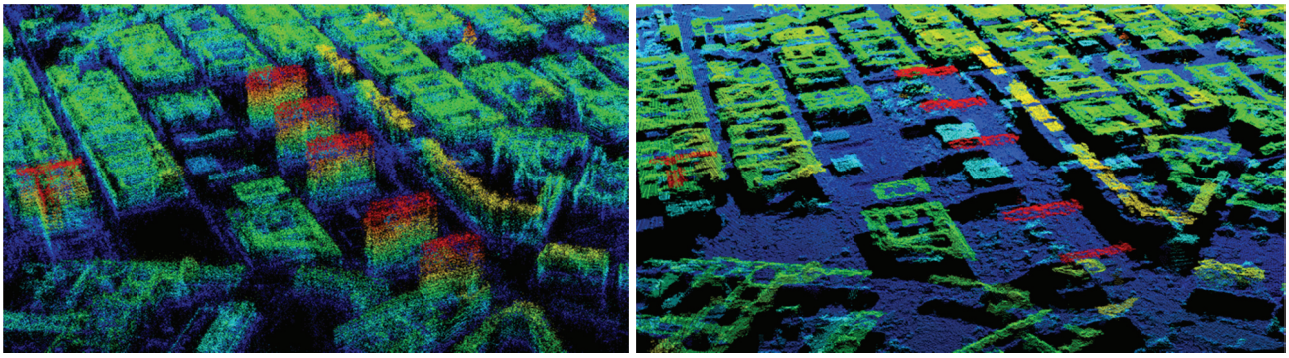
Considering the different image characteristics of side-looking SAR and nadir-looking aerial images used in our experiment, such co-registration problem can be generalized to the matching of side-looking and nadir-looking point clouds. The following issues must be taken into account:

- Due to the nadir-looking geometry of the optical images, façade points barely appear in the optical point cloud while they are prominent in the TomoSAR point cloud. This difference is exemplified in Figure 5, where the left and the right subfigures correspond to the TomoSAR and the optical point clouds of a same area in Berlin, respectively.

- The noise of the TomoSAR point cloud is extremely anisotropic. The Cramér–Rao lower bound (CRLB) of the elevation estimate for pixels of high signal-to-noise-ratio is [1], [8]

$$\sigma_{\hat{s}} = \frac{\lambda R}{4\pi\sigma_b\sqrt{N}\sqrt{2SNR}} \tag{8}$$

where σ_b is the standard deviation of the baseline, N is the number of images, and SNR is the signal-to-noise-ratio in linear scale. For a typical stack of 50 TerraSAR-X images, the CRLB of a 10dB pixel is about 1m which represents almost the best situation. In the contrary, the range and azimuth positions can be as accurate as 1 cm [35]–[37]. Considering the high dynamic range of the SNR in the scene, the elevation estimates are at least *one to two orders of magnitude* worse than those of range and azimuth. In UTM coordinate, the error in elevation is mostly transferred to the east and up directions for TerraSAR-X data, since the satellite is roughly on a polar orbit.



(a) (b)

Figure 5. (a) TomoSAR point cloud of high-rise buildings, and (b) the optical point cloud of the same area. Points are color-coded by height. Building façades are almost invisible in the optical point cloud, while they are prominent in the TomoSAR point cloud. The coordinate is UTM.

4.1 Co-registration algorithm

Typical point cloud co-registration problem employs ICP algorithm. However, ICP requires a good initial alignment. The unique modalities of optical and TomoSAR point clouds have driven our algorithm to be developed in the following way:

1 Edge extraction

   a. The optical point cloud is rasterized into a 2-D height image.

   b. The point density of TomoSAR point cloud is estimated on the rasterized 2-D grid.

   c. The edges in the optical height image and in the TomoSAR point density image are detected. These edges both correspond to the façade locations in the point clouds.

2 Initial alignment

   a. Horizontally by cross-correlating the two edge images.

   b. Vertically by cross-correlating the height histogram of the two point clouds.

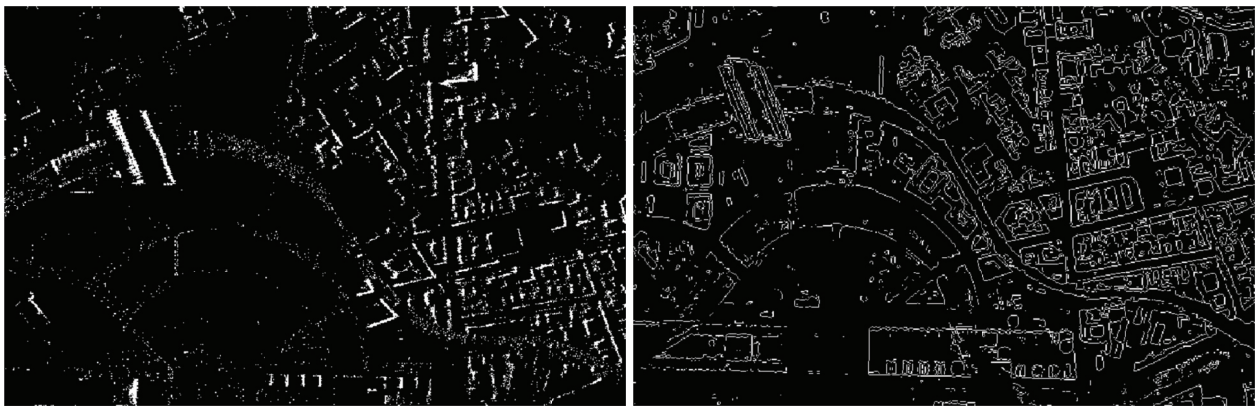
3 Refined solution

   a. The façade points in the TomoSAR point clouds are removed, because the optical point cloud contains nearly no façade point.

   b. To refine the co-registration of the two point clouds, an *anisotropic* ICP (AICP) with robustly estimated covariance matrices was applied. The covariance is estimated using a robust M-estimator.

2-D Edge Extraction

The rasterized optical height image and TomoSAR point density image are produced by computing the mean height and counting the number of points, inside each grid cell, respectively. The edges can be extracted from these two images using an edge detector, such as Sobel filter [38]. The thresholds in the edge detector are decided adaptively, so that the numbers of edge pixels in the two edge images are on the same order. Figure 6 is a comparison of the binary edge images of TomoSAR and optical point clouds. These edges correspond to the position of façades.



(a) (b)

Figure 6. (a) An example of the binary edge image of the TomoSAR point cloud in downtown Berlin, and (b) the edge image of the optical point cloud at the same area.

Coarse Alignment

The coarse alignment provides an initial solution to the ICP algorithm which is known to suffer from finding a local minimum. The coarse alignment in the horizontal and the vertical directions are carried out independently. The horizontal shift is found by cross-correlating the binary edge images of the two point clouds. The larger the area, the more prominent and unambiguous the correlation peak will be. Figure 7 shows the 2-D correlation of two edge images, where a single prominent peak is found. The vertical shift is found by cross-correlating the height histograms of the two point clouds, which is shown in Figure 8. The accuracy of the coarse shift is of course limited by the discretization level in the edge image and the height histogram. Nevertheless, this is sufficient for our application.

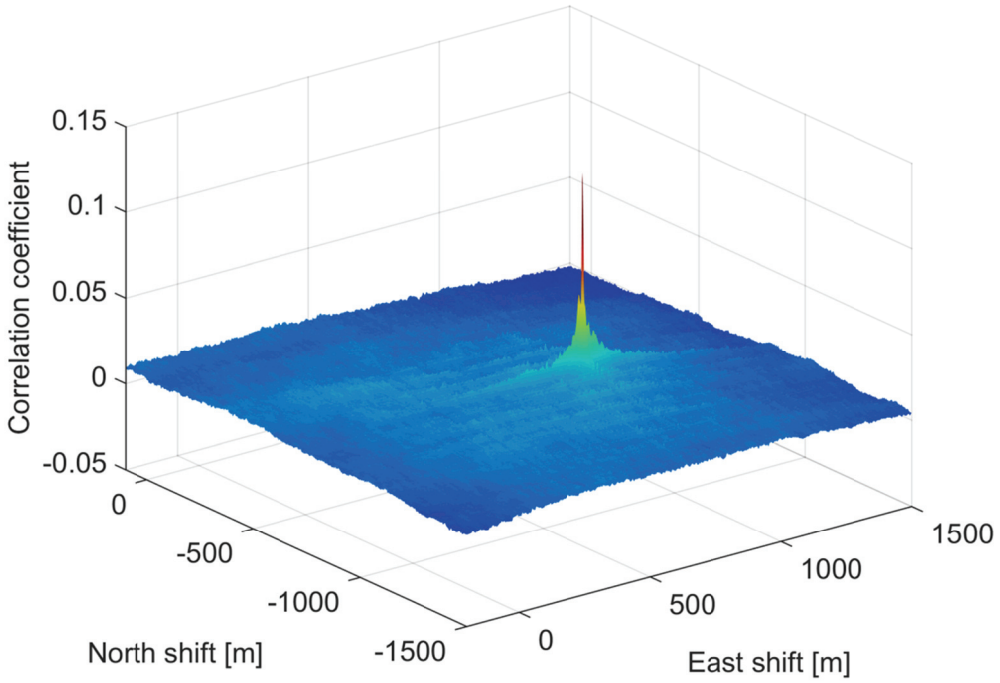
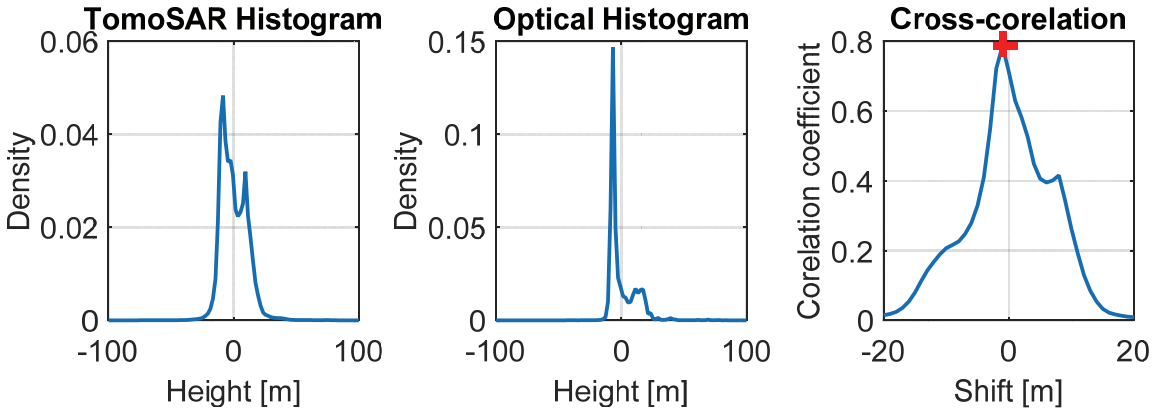


Figure 7. 2D cross-correlation of the edge images of TomoSAR and optical point clouds. The larger the area of the edge images, the more prominent and unambiguous the correlation peak will be. For the size of an entire city, a single prominent peak can usually be found.



(a) (b) (c)

Figure 8. (a) The height histogram of the TomoSAR point cloud, (b) the height histogram of the optical point cloud, and (c) the correlation of (a) and (b), where the red cross marks the peak position indicating the vertical shift.

Robust Anisotropic ICP

The final solution is obtained using an anisotropic ICP algorithm based on the coarse alignment. The key of the anisotropic ICP lies on the covariance matrix $\mathbf{C}_{\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i}$ of each point pair $\mathbf{x}_i$ (optical) and $\mathbf{p}_i$ (TomoSAR). Assuming the noise of optical point cloud is not correlated with that of the TomoSAR point cloud, $\mathbf{C}_{\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i}$ can be computed by [33]

$$\mathbf{C}_{\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i} = \mathbf{R}\mathbf{C}_{\mathbf{p}_i\mathbf{p}_i}\mathbf{R}^T + \mathbf{C}_{\mathbf{x}_i\mathbf{x}_i} \tag{9}$$

where $\mathbf{C}_{\mathbf{p}_i\mathbf{p}_i}$ and $\mathbf{C}_{\mathbf{x}_i\mathbf{x}_i}$ are the covariance matrix of the TomoSAR and the optical points, respectively.

As mentioned before, the dynamic range of the SNR of TomoSAR points is very high. In particular, closely space pixels may have dramatic SNR difference. This renders their positioning error distribution non-ergodic when selected for covariance estimation. To this end, a robust estimation of the covariance matrix — a topic has not been addressed in any previous ICP algorithms — is necessary.

To this end, we introduce an M-estimator [39] to robustly estimate the covariance matrix. Given a set of M samples $\mathbf{X} = \left[\mathbf{x}_1, \mathbf{x}_2 ...\right] \in \mathbb{R}$, M-estimator is an iteratively re-weighted sample covariance matrix [40]–[42],

$$\hat{\mathbf{C}}_k = \frac{1}{M}\sum_{i=1}^{M} w\left(\boldsymbol{\varepsilon}_i^T \hat{\mathbf{C}}_{k-1}^{-1}\boldsymbol{\varepsilon}_i\right)\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^T \tag{10}$$

where k is the iteration index, $\boldsymbol{\varepsilon}_i = \mathbf{x}_i - \hat{\bar{\mathbf{x}}}$ with $\hat{\bar{\mathbf{x}}}$ being the estimated sample mean, and $w(x)$ is a robust weighting function. The weighting function downweights heavily deviated samples whose whitened residual $\boldsymbol{\varepsilon}_i^T \hat{\mathbf{C}}_k^{-1}\boldsymbol{\varepsilon}_i$ is large. We choose the t-distribution weighting function which can handle heavily tailed distributions [41]:

$$w(x) = \frac{3+\nu}{\nu+x} \tag{11}$$

where $\nu \in \mathbb{R}^+$ is the degree of freedom (DoF) of the t-distribution. The lower the ν, the more robust but less efficient (under Gaussian distributed noise) of the estimator, and *vice versa*. ν is typically set to 2~5.

5. TOMOSAR POINT CLOUD TEXTURING

Texturing the TomoSAR point cloud is the step to link the semantics from the optical images to the deformation parameters of the TomoSAR point cloud, which can only be done upon obtaining the position and the orientation of the cameras w.r.t. the TomoSAR point cloud. This is equivalent to solving the visibility/occlusion problem given a camera viewpoint. Figure 9 illustrates the occlusion of TomoSAR points where certain points are only visible to certain images.

For solving the occlusion in a spaceborne TomoSAR point cloud, the following two issues should be taken into consideration.

- Typically, solving the occlusion of a point cloud requires reconstructing the 3-D surface and testing the depth of each point. This is known as the z-buffer algorithm. But reconstructing a reliable 3-D surface from TomoSAR point cloud still remains a challenging task, although there are some pioneer works on reconstructing façades and roofs [43]–[45]. Therefore, the employed technique should avoid surface reconstruction.
- Secondly, one point can be visible from multiple images. Strategy should be developed to choose the appropriate textures from multiple optical images. The textured point cloud should be visually consistent.

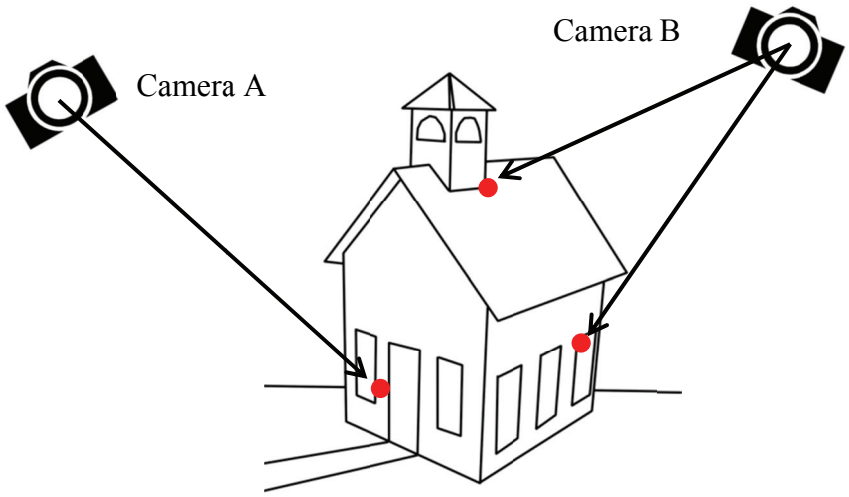


Figure 9. Illustration of TomoSAR point cloud texturing. The red dots represent TomoSAR points. The leftmost point is only visible to Camera A, whereas the right ones are only visible to Camera B.

5.1 Point-based Rendering

Hidden point removal (HPR) operator was introduced in [46], which is an elegant solution to avoid reconstructing surface or even surface normal in solving the occlusion problem. HPR inverts the point cloud using certain functions, so that all the visible points will lie on the convex hull of the inverted point cloud. And hence, finding the visible points becomes equivalent to compute the convex hull of the inverted point cloud.

The most common inverting function is the *spherical flipping*. Given a point $\mathbf{x}$, the spherical flipping function is defined as follows [46]:

$$\hat{\mathbf{x}} = f(\mathbf{x}) = \mathbf{x} + 2\left(r - \|\mathbf{x}\|\right)\frac{\mathbf{x}}{\|\mathbf{x}\|} \tag{12}$$

where r is radius of a multi-dimensional sphere. Figure 10 illustrates the transformation function and HPR in 2-D. The blue dots are the original noisy point cloud. And the red dots are the inverted

one given the camera viewpoint *O*. The bottom of the blue point cloud is visible to the camera. After inversion, these points lie closely to the convex hull of the inverted point cloud.

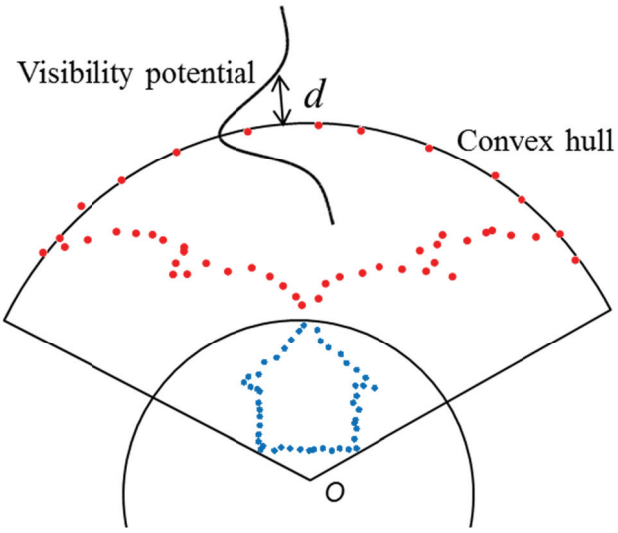


Figure 10. The improved hidden point removal operator. The blue dots are the original point cloud. Some noise was added to indicate the noise of the input TomoSAR point cloud. And the red dots are the inverted point cloud given the camera viewpoint *O*. The visible points will lie closely to the convex hull. Considering the noise of the point cloud, the visibility potential gives a soft threshold to the point visibility. *d* is the point to convex hull distance.

However, as mentioned in [47], the original HPR performs poorly in presence of noise. Considering the much larger noise in the spaceborne TomoSAR point cloud than in a typical computer vision problem, the HPR is not directly applicable.

[47] suggested to include a buffer zone around the convex hull, so that all the point in the buffer zone can be marked as visible. Inspired by this, we introduced a *visibility potential* which is a Gaussian function of the point-to-convex hull distance *d*:

$$p(\hat{\mathbf{x}}) = \exp\left(-\frac{d^2}{2\sigma_{\hat{\mathbf{x}}}^2}\right) \tag{13}$$

where $\sigma_{\hat{\mathbf{x}}}$ is the norm of the total noise standard deviations (all three directions) in the inverted point cloud. $\sigma_{\hat{\mathbf{x}}}$ is related to original point cloud. Consider the very anisotropic noise of TomoSAR point cloud, we define the standard deviation in the native SAR coordinate range, azimuth, and elevation being $\begin{bmatrix}\sigma & \sigma & k\sigma\end{bmatrix}$, where $k>1$. As mentioned in Chapter 4, this factor is about 10 to 100. Hence, the noise in range and azimuth can be neglected. Transforming to UTM coordinate, the standard deviation for east (x), north (y), and up (z) is approximately $\begin{bmatrix}k\sigma/\sqrt{2} & \sigma & k\sigma/\sqrt{2}\end{bmatrix}$ (assuming a near polar orbit, and 45° of incidence angle). The maximum $\sigma_{\hat{\mathbf{x}}}$ is derived to be:

$$\sigma_{\hat{\mathbf{x}}} = \frac{4rk\sigma}{\|\mathbf{x}\|_{\min} - k\sigma} - k\sigma \tag{14}$$

where $\|\mathbf{x}\|_{\min}$ is the minimum distance of the original point cloud to the viewpoint.

5.2 Consistency Check

Many points are visible to multiple images. Selecting the most appropriate image for texturing should consider the visual consistency of the textured point cloud, as well as the optical quality of the texture, since pixels located far away from the center of an image often have oblique angle with poor resolution.

We use the information of the neighbouring points to check the consistency of the texturing, so that closely spaced points will select texture from a same optical image. Firstly, for each point, a score similar to [5] is obtained w.r.t. each image as follows:

$$c_j = p_j d_j \tag{15}$$

where j is the image index, p_j is the visibility potential (equation (13)), and d_j is the inverse of the distance to the viewpoint of camera j. d_j can be regarded as the resolution of the image at the point.

Equation (15) basically favors image with greater visibility and shorter distance. The image with the largest score is selected as the initial texture source image.

Secondly, to decide the final texture source image for each point, we search the *k* nearest neighbors of each point, and perform a weighted multi-class majority voting. The weight is a Gaussian kernel as a function of the distance of the neighbor points to the center point. Largest weight is assigned to the point itself, and lower weights are assigned to the points further away. The image with the highest votes is selected as the final texture source.

6. PRACTICAL DEMONSTRATION

This section demonstrates the proposed framework by examples of railway and bridge monitoring, which are important ground infrastructures that require systematic monitoring. Monitoring large façades is also a potential application, given oblique aerial images which are unfortunately not available in this study.

6.1 Datasets

SAR

The SAR datasets consist of two stacks of TerraSAR-X high resolution spotlight images of Berlin with spatial resolution of about 1m. The two image stacks were acquired from Feb. 2008 to Mar. 2013 with about 100 images each. They were acquired at ascending and descending orbits respectively, which provide a full coverage of the whole city. More information of the two SAR image stacks can be found in Table 1.

Table 1. Information of the SAR image stacks

Beam	Orbit	Incidence angle	No. of images	Temporal span	Baseline span
42	Descending	36°	109	5.3 year	363m
57	Ascending	42°	102	5.3 year	840m

Optical

The optical images are 9 UltraCam aerial images of Berlin acquired in March 2014 at an altitude of about 4000 m. The ground spacing is approximately 20cm per pixel. The camera positions and orientations are measured by onboard GPS and inertial measurement unit. After SfM adjustment, their standard deviations are about 5cm and 5×10^{-4} degree, respectively.

6.2 3-D Reconstructions

TomoSAR

The InSAR stacking was performed by the German Aerospace Center (DLR)'s *IWAP* processor—the integrated wide area processor [48] based on its predecessor *PSI-GENESIS* [49]. The atmospheric phase correction and the following D-TomoSAR processing was performed by *Tomo-GENESIS* [50], [51] – a TomoSAR software of the DLR for large urban areas monitoring. It is developed based on the work of [8], [9], [12], [51]–[53]. It delivers a 5-D point cloud – 3-D position plus linear deformation rate and amplitude of seasonal motion - from a stack of SAR images of urban area.

Figure 11 is the TomoSAR point clouds from the two image stacks reconstructed by the SVD-Wiener [8] (Fourier-based) algorithm in the Tomo-GENESIS. The two point clouds are co-registered using a feature-based matching algorithm which estimates and matches common building edges in the two point clouds [31]. The combined TomoSAR point cloud has about 40 million points. The point density is about $10^6/km^2$ which is even comparable to some LiDAR product. It enables the retrieval of the features of individual building. Behind each of these points, the linear deformation rate and the amplitude of seasonal deformation were also estimated.

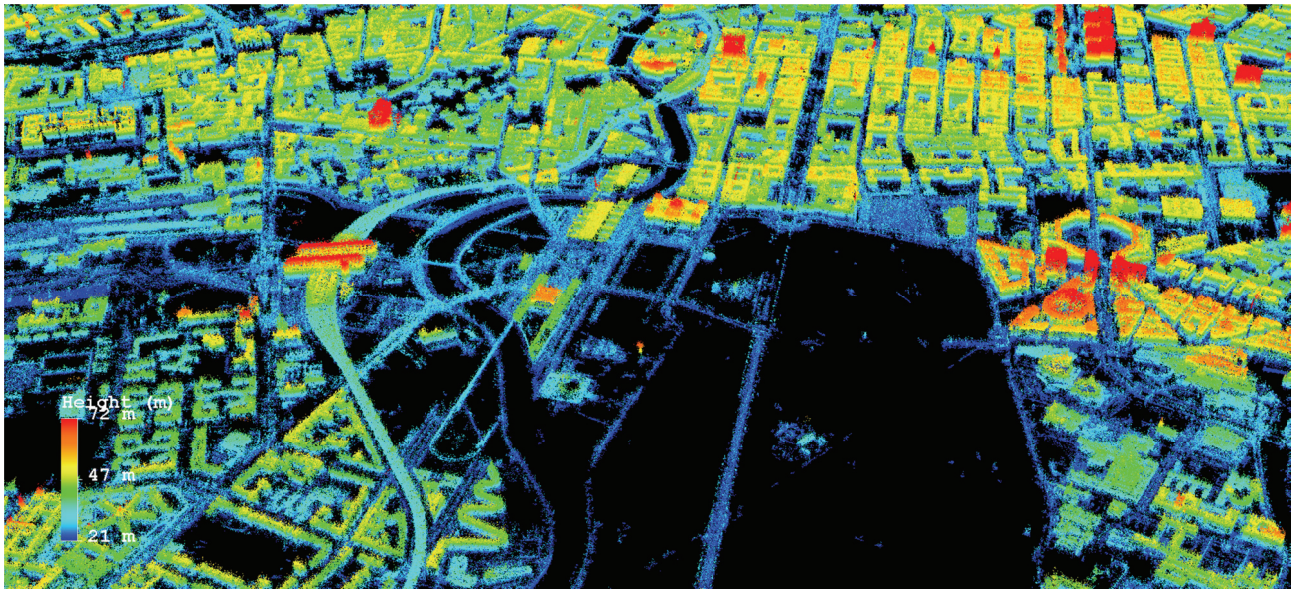


Figure 11. The fused TomoSAR point cloud of Berlin, which combines the result from an ascending stack and a descending stack. The height is color-coded. The coordinate is UTM.

Optical

The 3-D reconstruction in optical images was carried out by the commercial software Pix4D [54]. It calibrates the camera model, as well as the positions and orientations of the cameras. Figure 12 is the reconstructed 3-D optical point cloud showing the identical area as that in Figure 11. The point cloud contains 35 million points. However, basically no façade point exist which contradicts the rich façade points in the TomoSAR point cloud.

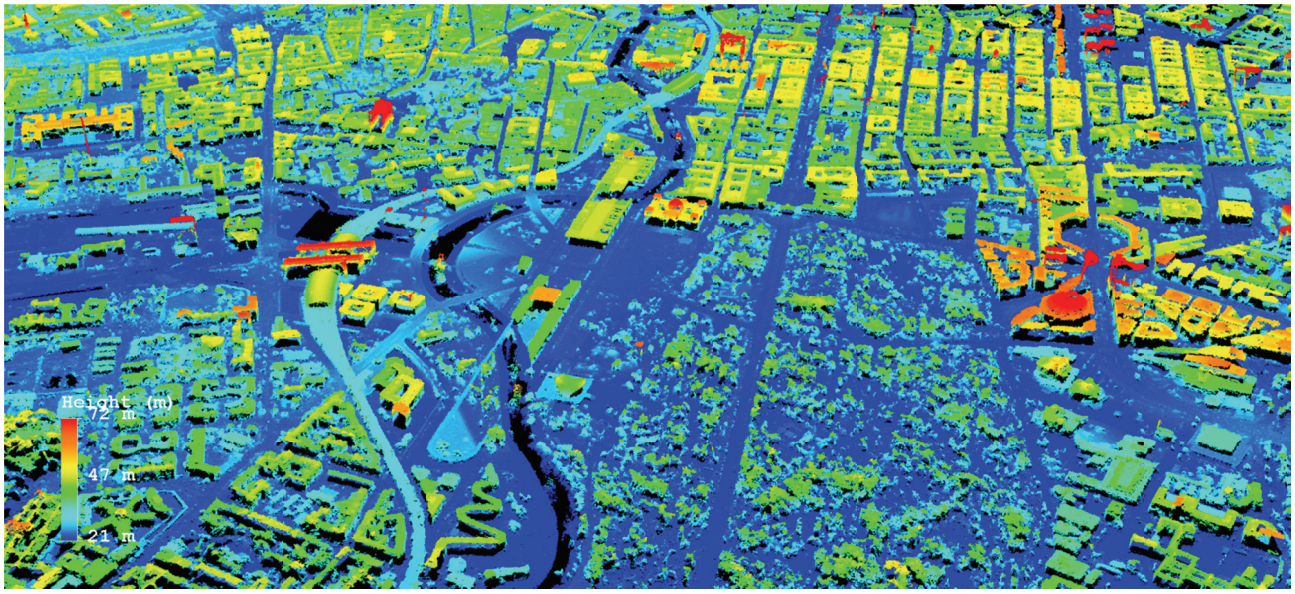


Figure 12. Optical point cloud of Berlin reconstructed using structure from motion and dense matching. The height is color-coded. The colorbar and area are identical to Figure 11. The coordinate is UTM.

6.3 Point Clouds Accuracies

TomoSAR point cloud

InSAR methods, like TomoSAR, always provide *relative estimates*, i.e. positioning and deformation estimates are referred to a selected reference point. The relative positioning accuracy, i.e., relative to the reference point, depends on the SNR and the number of images (equation (8)). For the TomoSAR point clouds in this paper, the accuracy ranges from 4m to 4cm, which correspond to SNR of -10dB to 10dB.

Absolute TomoSAR/InSAR, also referred to as geodetic TomoSAR/InSAR, is still a very new and open research topic [55]. Preliminary study on several manually selected points with an extremely high SNR suggests a positioning accuracy of up to 20cm [55], [56].

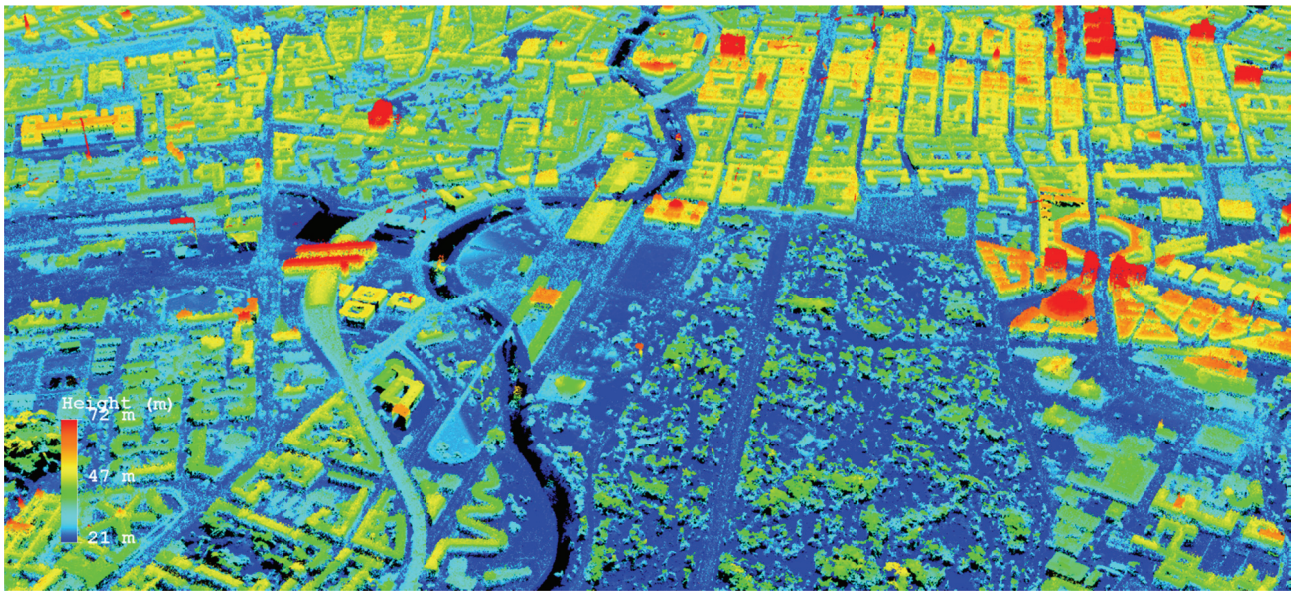
Optical point cloud

The global localization accuracy of the optical point cloud is given by the accuracy of the exterior parameters of the camera, which are about 5cm in position and 5×10^{-4} degree in orientation after adjustment. This translates to about 10cm at 4000m altitude where the images were taken.

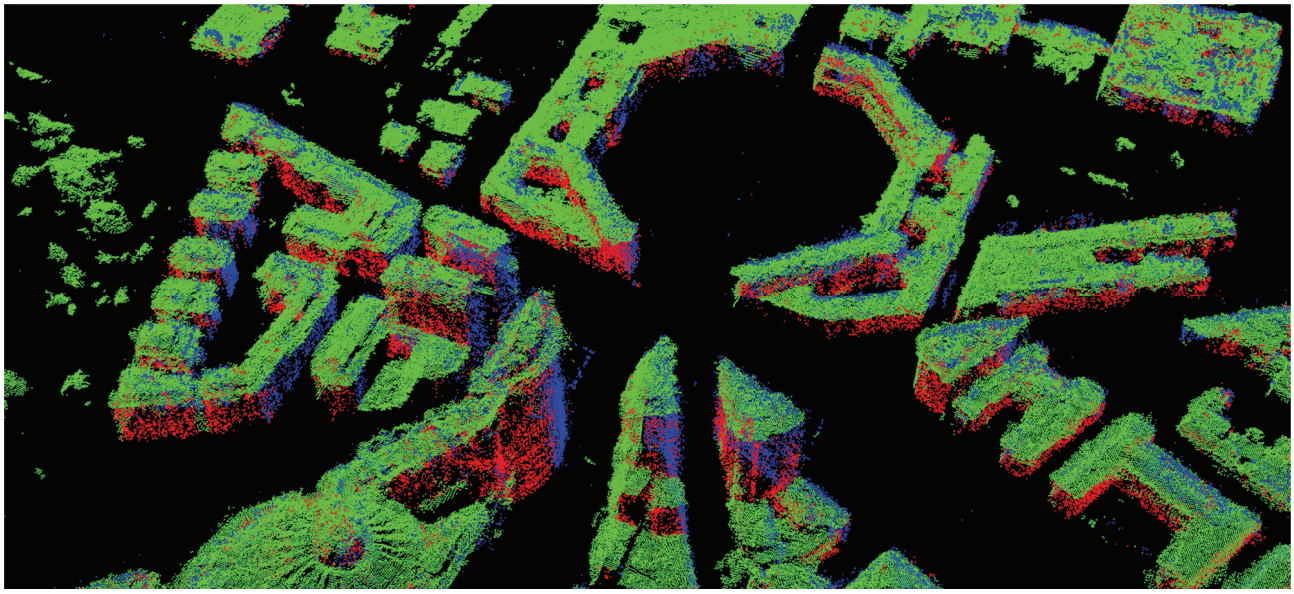
The relative accuracy of the optical point cloud depends on the stereo matching algorithm, e.g. semi-global matching. Since we used a commercial software, the algorithm behind it is unknown. But at the best, the standard deviation of the 84 ground control points is 3cm. Also by checking some flat surfaces in the optical point cloud, the accuracy in height is about 20cm.

6.4 Point Clouds Co-registration

Since the optical point cloud is quite accurate, and ground control points were used, the scaling and rotation are therefore not considered in the co-registration. Figure 13(a) demonstrates the co-registered point cloud combining the optical and two TomoSAR point clouds. The figure shows the identical area as that in Figure 11 and Figure 12. Successful co-registration can be confirmed by seeing the correct location of the façade points in Figure 13(b) which shows a close-up view of the matched point cloud with different colors representing different point clouds. The ground points in Figure 13(b) were removed for a better visualization.



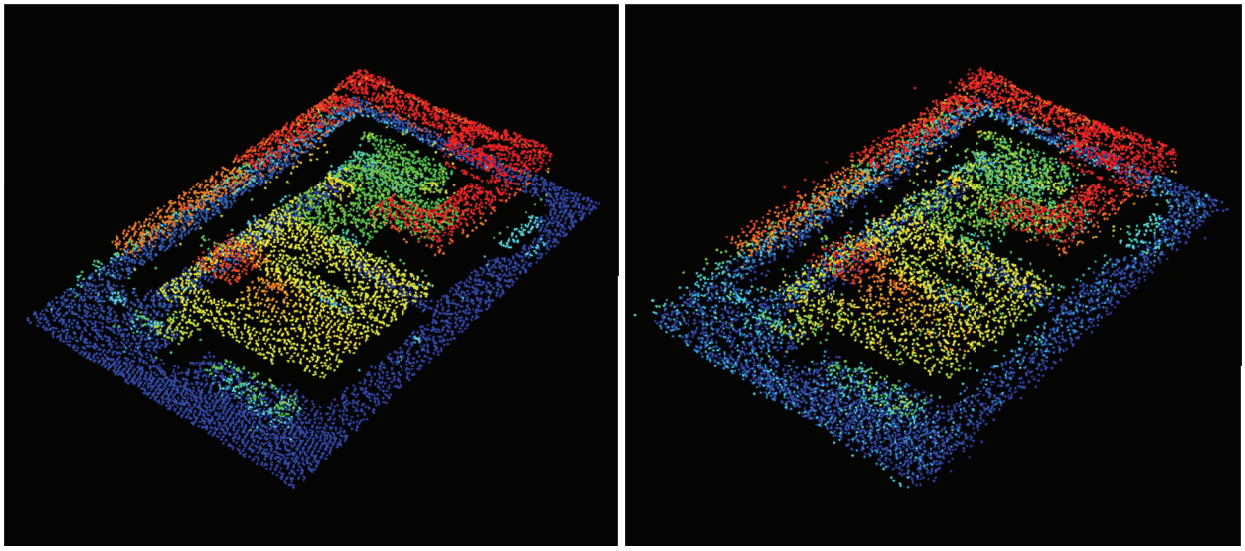
(a)



(b)

Figure 13. (a) The fused point cloud combining the optical point cloud and two TomoSAR point clouds from ascending and descending viewing angle. The height is color-coded. The colorbar and the area are identical to those in Figure 11 and Figure 12, and (b) a close up view of the co-registered point cloud in Berlin Potsdamer Platz, where green, red, blue represent the points from optical, ascending and descending TomoSAR point cloud, respectively. The ground points were removed for a better visualization.

As no ground truth of the co-registration of the optical and TomoSAR point clouds is available, we used a simulation to assess the performance of the robust AICP. A LiDAR point cloud of a building in Berlin about 10,000 points with centimeter accuracy was used as the reference point cloud, i.e. optical point cloud. We simulated non-ergodic noise for the reference point cloud with SNR distributed from -10 to 10dB, which well simulated the noise in the TomoSAR point cloud. The noisy LiDAR point cloud is then translated to simulate the target point cloud, i.e. TomoSAR point cloud. Figure 14 shows the original and the noisy LiDAR point clouds.



(a) (b)

Figure 14. (a) original LiDAR point cloud with centimeter accuracy, and (b) simulated TomoSAR point cloud by adding non-ergodic noise to the LiDAR point cloud.

Figure 15 compares the translation error of the original ICP, the AICP, and AICP with M-estimate of covariance matrix (MAICP) at each iteration. All three algorithms converge. Comparison shows that MAIPC outperforms the other methods in terms of co-registration accuracy and convergence rate. In theory, AICP should also outperform ICP. But the incorrect covariance matrix estimation due to non-ergodic noise leads to its worse performance than the original ICP. This implies a robust estimation of the covariance matrices is necessary in such case. The same experiment was also performed on several other typical benchmark point clouds, such as the Stanford bunny. Consistent results were obtained.

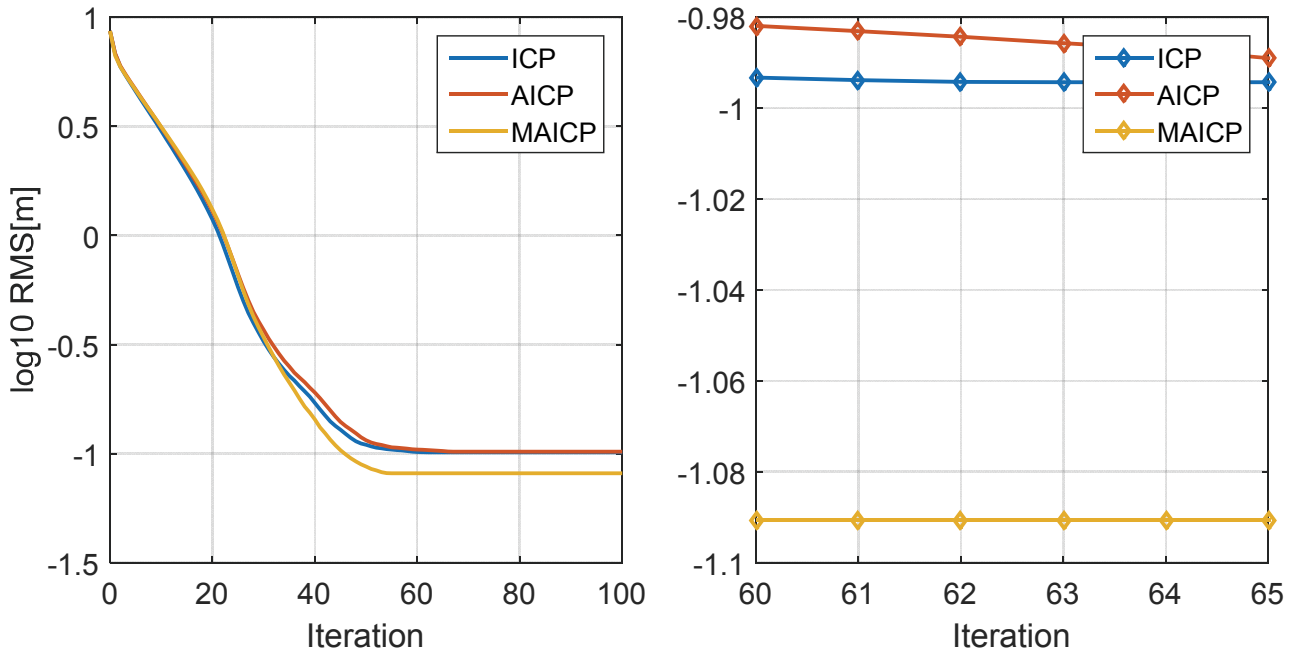


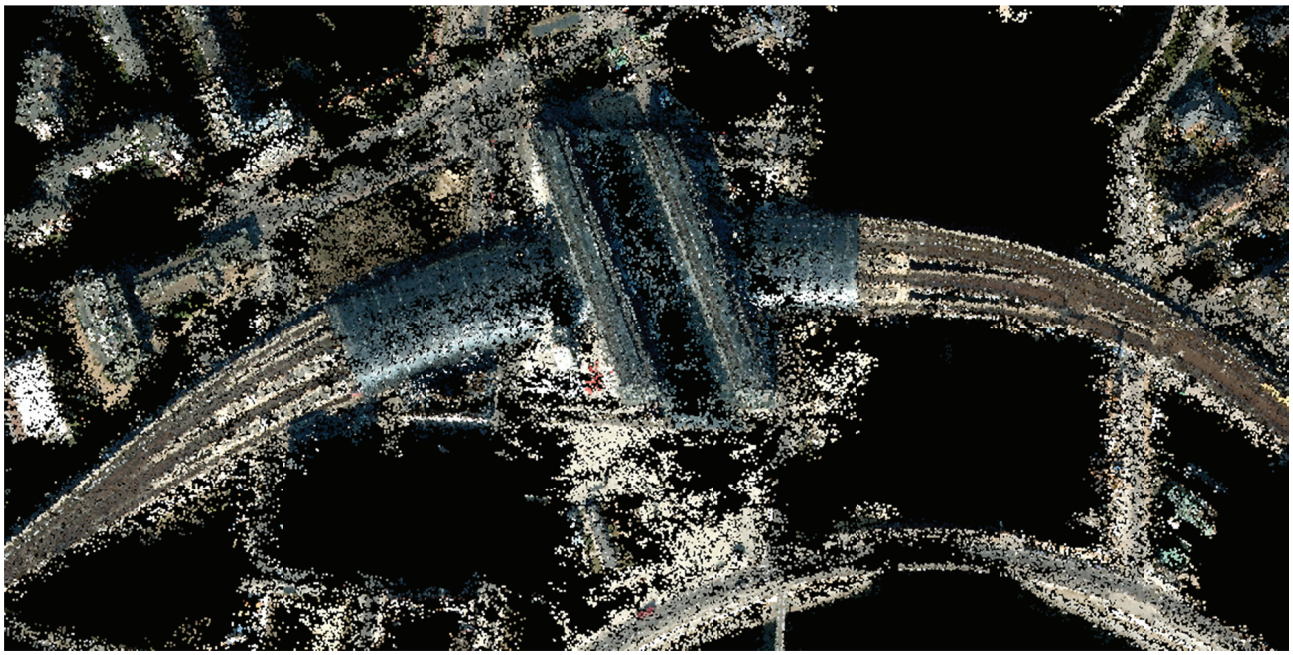
Figure 15. Comparison of the translation error of original ICP, anisotropic ICP, and anisotropic ICP with robust covariance matrix estimate (M-estimate) at each iteration of the algorithms. The scale is log10 [m]. The right subfigure is a close up view of iteration 60 to 65.

6.5 TomoSAR Point Cloud Texturing

By applying the texturing algorithm, Figure 16 shows the first meter-resolution spaceborne TomoSAR point cloud of an urban area textured with optical color. The subfigure (b) and (c) are zoom in to the Berlin central station and the high-rise building of Universitätsmedizin Berlin which are indicated by the red arrows in subfigure (a). The black parts are where no TomoSAR point was reconstructed. The proposed framework successfully projects optical images to urban TomoSAR point cloud, including points on large façades, as well as points on ground objects.

On the other hand, it is also possible to project the 2-D optical image to the 2-D SAR image geometry. Figure 17 shows the SAR amplitude image and the optical image projected in SAR range-azimuth geometry. This is also the first high resolution urban optical image projected in SAR geometry. The Berlin central station and the hospital are also indicated by the red arrows in Figure 17.

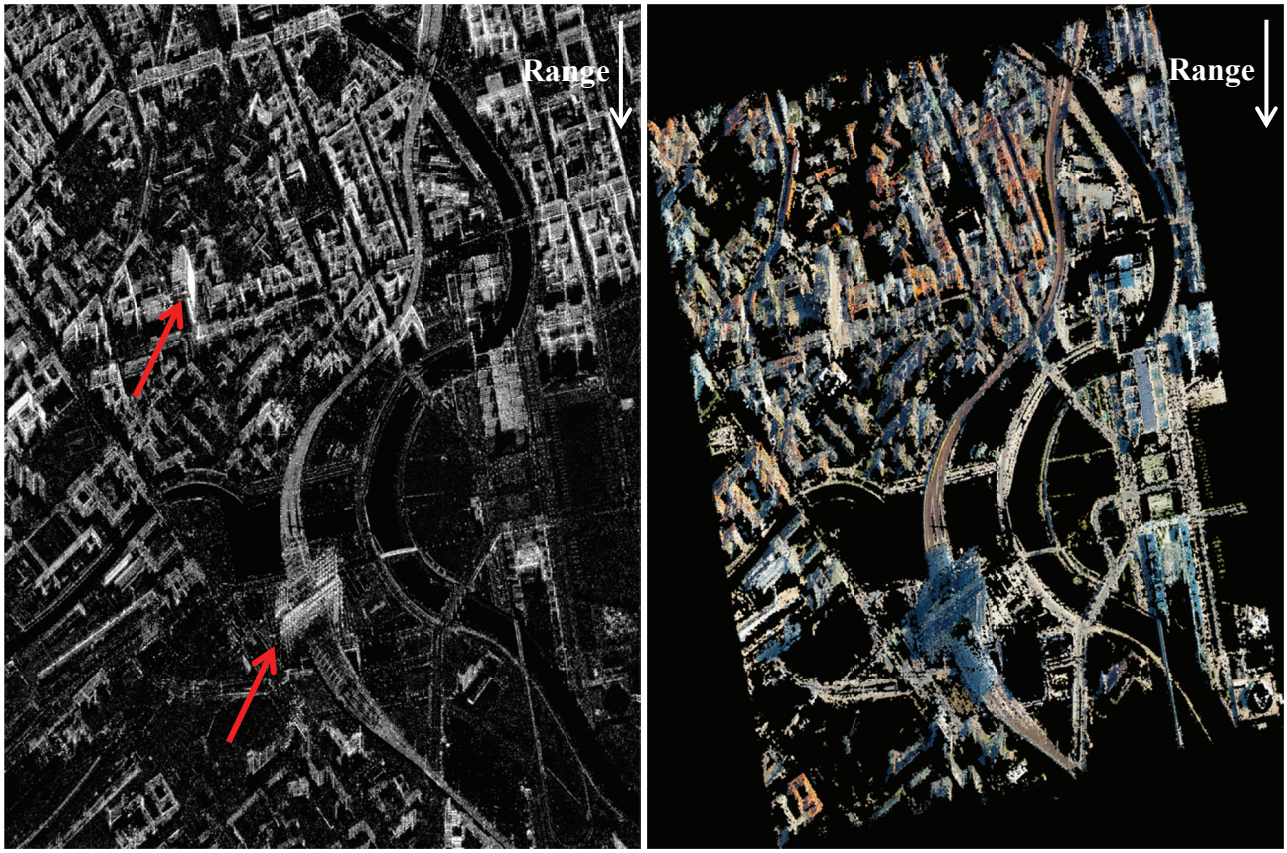
(a)



(b)

(c)

Figure 16. (a) First TomoSAR point cloud textured with optical image, and (b) zoom in to the Berlin central station, and (c) zoom in to the high-rise building of Universitätsmedizin Berlin. The coordinate is 3-D UTM. The proposed framework successfully projects optical images to urban TomoSAR point cloud, including points on large façades, as well as points on ground objects.



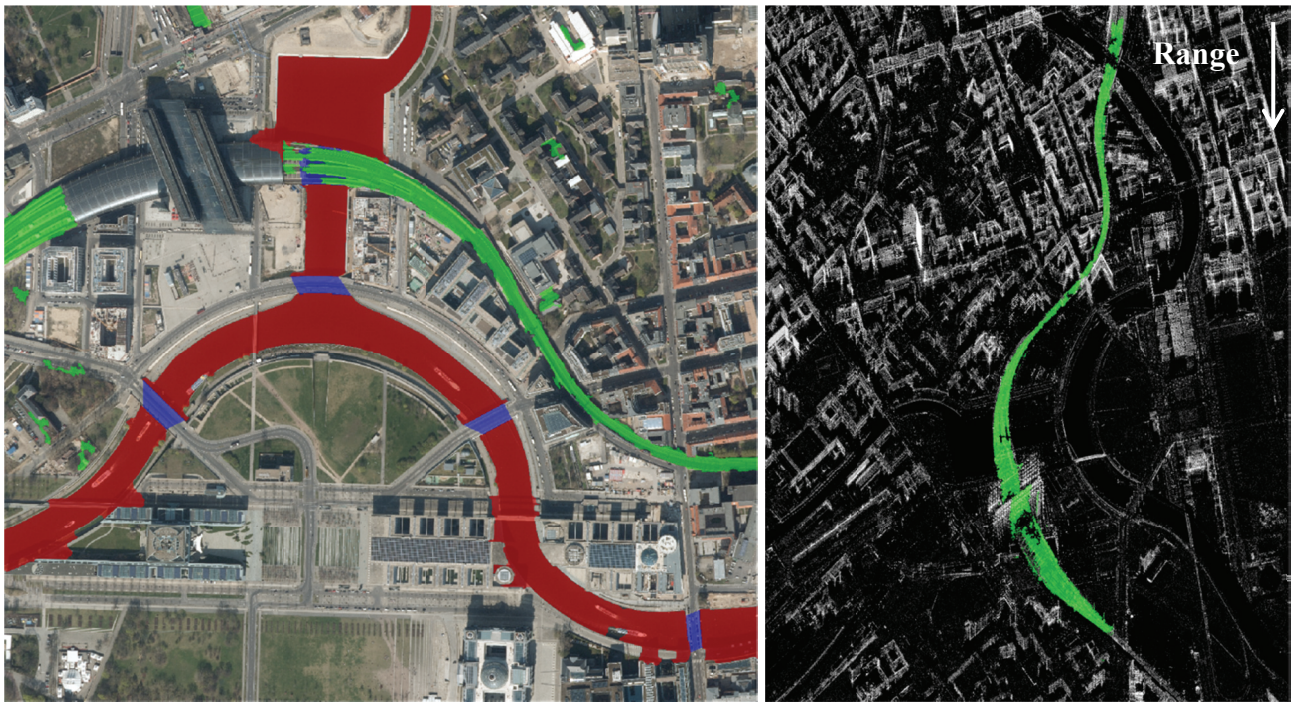
(a) (b)

Figure 17. (a) the SAR amplitude image, and (b) the optical image in SAR image geometry. Both images are in SAR azimuth and range coordinate. The red arrows indicate the locations of Berlin central station and the high-rise building of Universitätsmedizin Berlin.

6.6 Optical Image Classification

The optical images were classified patch-wisely using the Bag of Words (BoW) [57] method which is a well-known technique in the computer vision community. As this is not the focus of this article, the readers are kindly referred to the original literature for more information. For the deformation analysis in this paper, we classified the railway and the river class, since we found interesting deformation pattern on the railway and bridges. Classification on large façades is also a potential application, which of course requires oblique optical images.

The classification is supervised. Training patches were manually selected for each class. The classification is done patch-wisely in the large aerial image. The feature used is simply the RGB value in a 3×3 sliding window in the patch. The classifier is a linear SVM [58] implemented in an open source library VLFeat [59]. Figure 18(a) shows the classification result, where red is river, green is railway, and blue is bridge. The bridge label is derived by filling the gap between the river segments. Small clusters do appear as false alarm. But they can be easily removed by post-processing. These labels can be projected to the SAR image or the TomoSAR point cloud, in order to perform further analysis. For example, Figure 18(b) is the railway label (after post-processing, explained in section 6.7) projected to the SAR image geometry identical to that of Figure 17.



(a) (b)

Figure 18. (a) River (red), railway (green) and bridge (blue) classified using the BoW method. The blue label is derived by filling the gap between the river segments, which indicates the bridge locations. Some false alarm appeared as small clusters. They can be removed using post-processing. (b) The railway label (after post-processing explained in section 6.7) projected to SAR image geometry which is identical to that in Figure 17.

6.7 Railway Monitoring

Based on the classification, the corresponding points in the TomoSAR point cloud can be extracted by projecting the classification label to the point cloud. A smooth spline function was fitted to the east-north coordinates of the railway points to connect separated segments. For example, the railway label is disconnected due to the presence of the Berlin central station. Figure 19(a) shows the extracted continuous railway points overlaid on the optical image. The color shows the amplitude of seasonal motion.

This motion is mostly caused by the thermal dilation of the steel railway because of the seasonal temperature change. It can be observed that the railway deformation experiences certain periodic discontinuity. It is suspected to be caused by the mechanical joints between different railway segments.

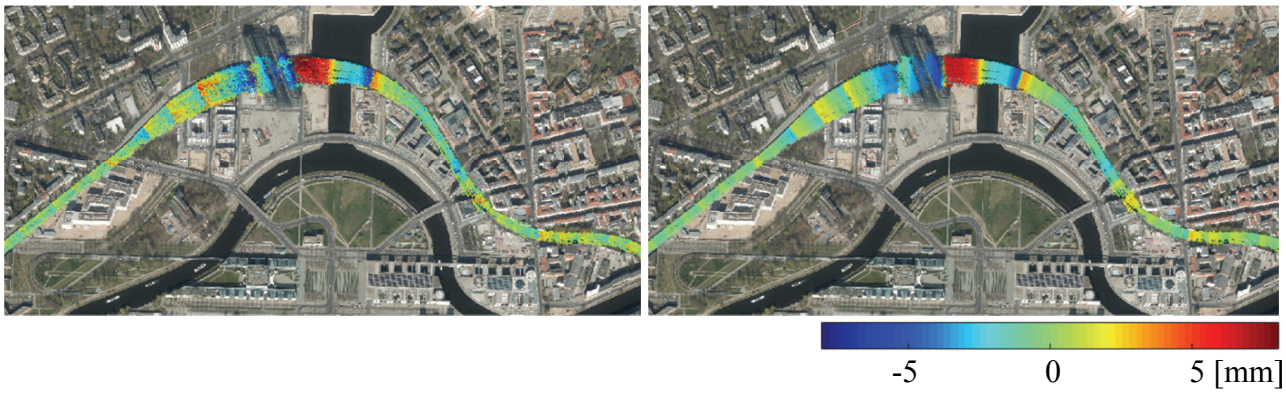
For monitoring these discontinuities, the raw TomoSAR estimates are relatively noisy due to the relative low SNR of some scatterers. To filter out the noise, a special prior was employed, i.e. the thermal dilation of a steel beam is linearly proportional to its length at its first approximation [60]. That is to say, the railway deformation parameters is piecewise linear, and the scatterers on the same cross-section of the railway undergo similar deformation. Therefore, the deformation estimates can be filtered by minimizing its total generalized variation along the railway direction (with second order derivative which favors piecewise linear structure):

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}} \left\{ \frac{1}{2} \left\| \mathbf{f} - \mathbf{a} \right\|_2^2 + \lambda \left\| \Delta \mathbf{f} \right\|_1 \right\} \tag{16}$$

where $\mathbf{a}$ is the deformation estimates along the 1-D railway direction (i.e. unfold the railway to a straight line), $\mathbf{f}$ is the filtered version of $\mathbf{a}$, and $\Delta\mathbf{f}$ is the second order derivative of $\mathbf{f}$. As shown

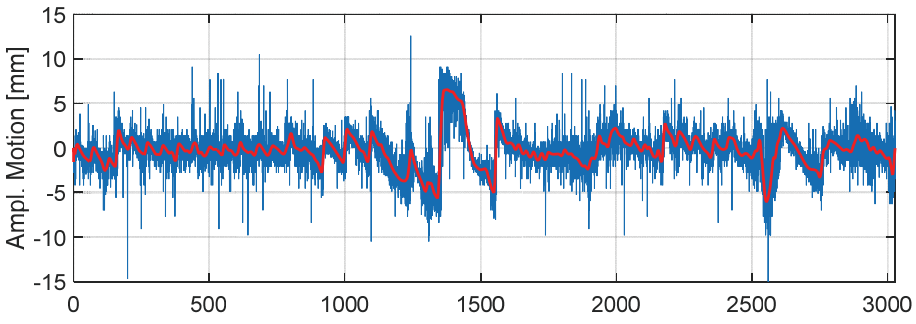
in literatures of total generalized variation [61], [62], the L_1 norm of second order derivative is convex and lower semi-continuous, one can solve it using convex optimization solvers.

The filtered deformation estimates are shown in Figure 19(b), in comparison with the unfiltered estimates Figure 19(a). The noise has been greatly reduced while the discontinuity is well preserved. To have a quantitative comparison, Figure 20(a) plots the original (in blue) and the denoised (in red) deformation estimates as a function of the 1-D railway direction. The original estimates have a standard deviation about 1~2 mm which concealed some edges where the magnitude of the deformation is low. The filtering successfully reconstructs the edges.

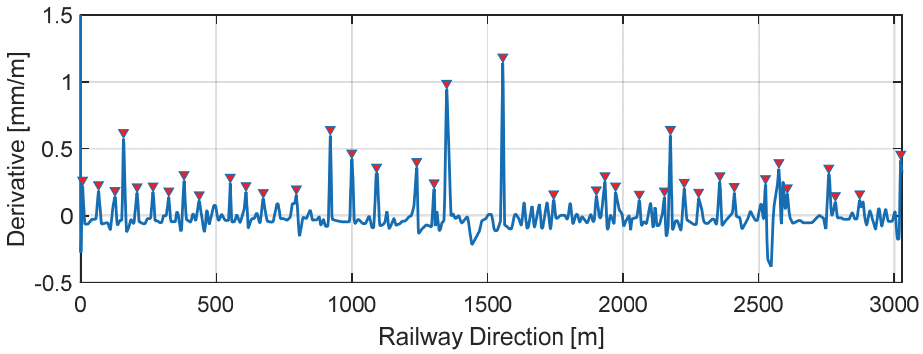


(a) (b)

Figure 19. (a) the railway points extracted from the original TomoSAR point cloud, and (b) the railway points filtered using total generalized variation. The color shows the amplitude of seasonal motion due to the thermal expansion of the railway.



(a)



(b)

Figure 20. (a) the original and the filtered estimates of amplitude of seasonal deformation in blue and red, respectively, and (b) peaks detected in the derivative of the filtered deformation along the railway direction.

By detecting the peaks in the derivative of the deformation function, the location of the discontinuities can be detected. Constraint was put on the minimum distance between two peaks representing the minimum length of a railway segment. The detected peaks in the deformation's derivative can be seen in Figure 20(b). The positions of the peaks on the railway are shown as the green dots in Figure 21. Each green dot represents the midpoint of its railway cross-section. In the middle subfigures of Figure 21, the close up view of the two joints in the optical image is provided. As the optical image has limited resolution, we also provide a higher resolution one (7cm ground spacing) in Figure 22. It can be clearly observed that the railway joint shown up as dark lines in the optical image.

However, further investigation shows that such large mechanical joints only appear close to railway stations. Visual inspection of the high resolution image of other parts of the railway does not find obvious mechanical joints. Therefore, a more general cause of the discontinuities in the deformation still requires further investigation.

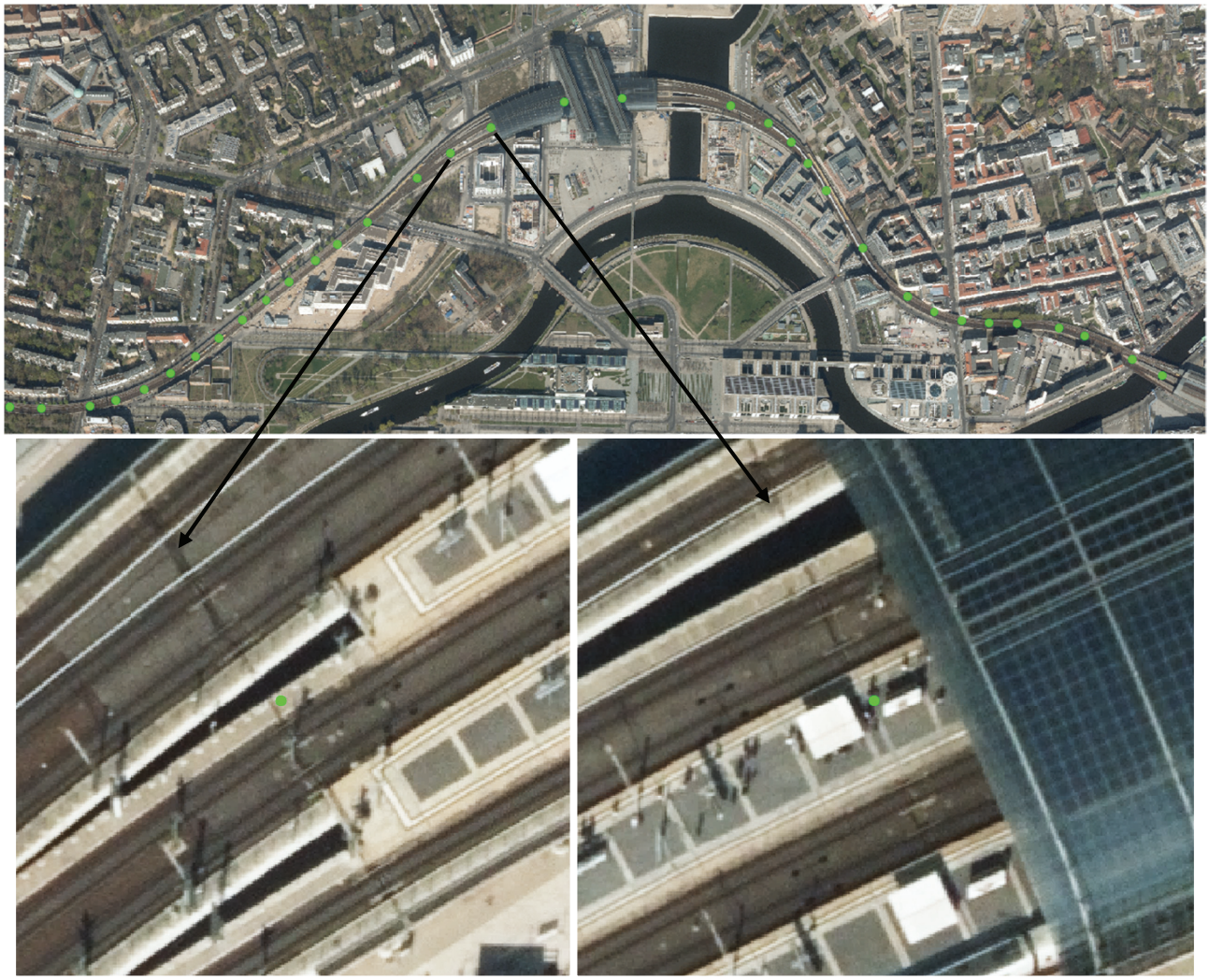


Figure 21. Upper: the midpoint of the detected railway joint cross-section marked in green, and lower: close up view of the railway joints. The background optical image has a ground spacing of 20cm.

Figure 22. A higher resolution image (7cm ground spacing) of the lower left subfigure of Figure 21.

6.8 Bridges Monitoring

By analyzing the gaps of the river segmentation and assuming the gaps are caused by bridges, the bridges' positions can be detected. Without going into too much detail, the extracted bridges points overlaid on the optical image are shown in Figure 23 where the color represents the amplitude of seasonal deformation. The upper bridge belongs to a segment of the railway which is known to have thermal expansion. The middle bridge undergoes a 5mm seasonal motion on its west end and 2mm at the east end. This suggests a more rigid connection of the bridge to the foundation at its east end. The two lower bridges are stable according to the motion estimates.

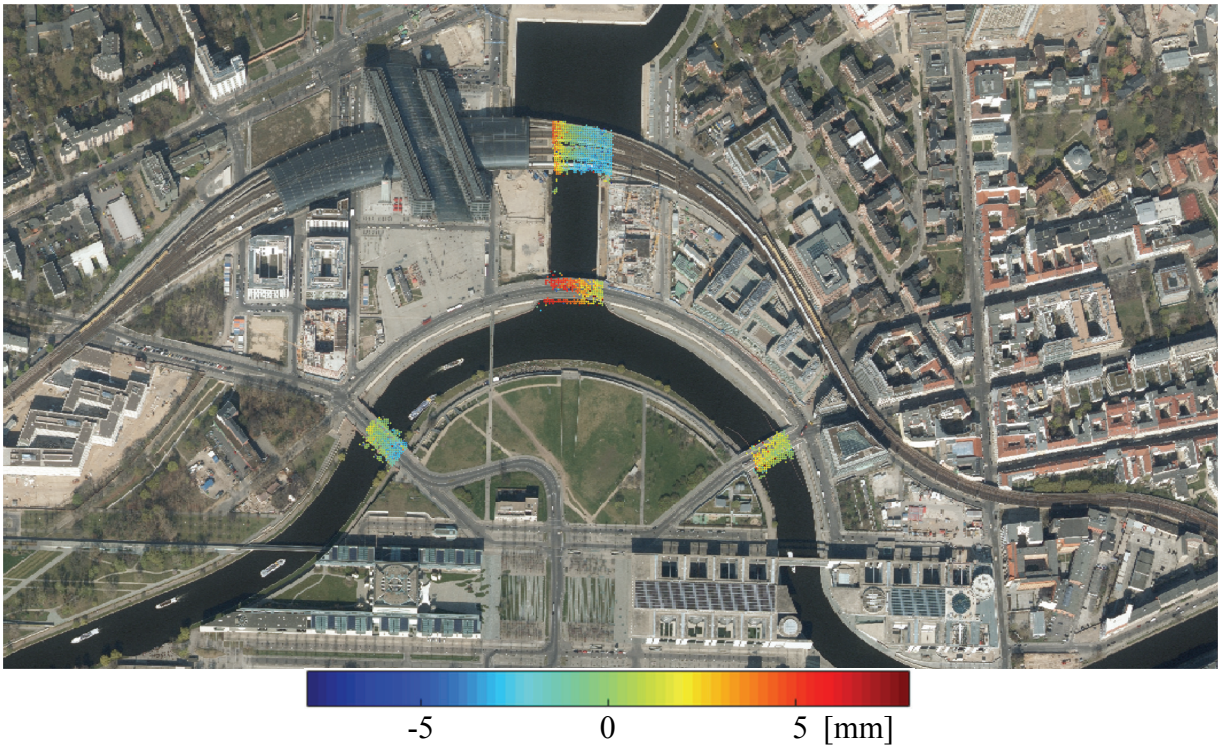


Figure 23. The amplitude of seasonal motion of the bridges extracted from the TomoSAR point cloud overlaid on the optical image.

CONCLUSION AND OUTLOOK

This paper presents the first joint semantic analysis of optical images and meter-resolution spaceborne InSAR point cloud in urban area. Through a 3-D geometric fusion of optical images and InSAR point cloud, the attributes from the optical image, e.g. color, semantic labels, can be projected to the InSAR point cloud. We presented the first urban "SARptical" image which is the optical image projected to SAR image geometry.

Experiments show that a robust estimation of point covariance matrix is mandatory in applying anisotropic ICP algorithm to TomoSAR point cloud. Example on railway monitoring discovers the periodic deformation pattern along the railway in Berlin is caused by the mechanical joints between railway segments only for several locations close to railway stations. No visible mechanical joint is identified for other location, which leads to some further investigation of the cause of deformation.

This work also opens some new perspectives in high resolution urban remote sensing. For example, *systematic monitoring of large façades*, which requires high resolution oblique optical images and the development of façade classification algorithm; and *joint classification of high resolution optical image and InSAR point cloud/SAR image*, which can utilize InSAR attributes such as 3-D position and deformation parameters as additional features in the classification. For all these methods, the main challenge remains at the positioning accuracy of spaceborne InSAR point cloud, which is in the order of 1 to 10 m for TerraSAR-X with its three imaging modes. Future work on improving the positioning accuracy of TomoSAR, e.g. by exploiting joint sparsity [63] or object based TomoSAR inversion [64], is the key in the joint analysis of high resolution urban optical and SAR images.

ACKNOWLEDGEMENT

This work was supported by the Helmholtz Association under the framework of the Young Investigators Group "SiPEO" (VH-NG-1018, www.sipeo.bgu.tum.de), International Graduate School of Science and Engineering, Technische Universität München (Project 6.08: "4D City") and the German Aerospace Center (DLR, Förderkennzeichen 50EE1417). The authors would like to thank Dr. H. Hirschmüller of DLR-RM for providing the optical data, as well as the reviewers for their valuable suggestions.

REFERENCES

[1] R. Bamler, M. Eineder, N. Adam, X. Zhu, and S. Gernhardt, "Interferometric Potential of High Resolution Spaceborne SAR," *Photogramm. - Fernerkund. - Geoinformation*, vol. 2009, no. 5, pp. 407–419, Nov. 2009.

[2] U. Soergel, Ed., *Radar Remote Sensing of Urban Areas*, vol. 15. Dordrecht: Springer Netherlands, 2010.

[3] Y. Wang and X. X. Zhu, "InSAR Forensics: Tracing InSAR Scatterers in High Resolution Optical Image," presented at the Fringe 2015, 2015.

[4] Y. Wang and X. X. Zhu, "Semantic Interpretation of InSAR Point Cloud," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2015*, Milan, Italy, 2015.

[5] 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.

[6] F. Lombardini, “Differential tomography: a new framework for SAR interferometry,” *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 1, pp. 37–44, Jan. 2005.
[7] 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.
[8] 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.
[9] 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.
[10] G. Fornaro, F. Serafino, and F. Soldovieri, “Three-dimensional focusing with multipass SAR data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 3, pp. 507–517, Mar. 2003.
[11] 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.
[12] 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.
[13] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski, “Building Rome in a day,” *Commun. ACM*, vol. 54, no. 10, p. 105, Oct. 2011.
[14] M. Pollefeys, D. Nistér, J.-M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S.-J. Kim, P. Merrell, and others, “Detailed real-time urban 3d reconstruction from video,” *Int. J. Comput. Vis.*, vol. 78, no. 2–3, pp. 143–167, 2008.
[15] M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch, “Visual Modeling with a Hand-Held Camera,” *Int J Comput Vis.*, vol. 59, no. 3, pp. 207–232, Sep. 2004.
[16] 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.
[17] C. Wu, “Towards linear-time incremental structure from motion,” in *3D Vision-3DV 2013, 2013 International Conference on*, 2013, pp. 127–134.
[18] 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.
[19] M. Pollefeys, R. Koch, and L. Van Gool, “Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters,” *Int. J. Comput. Vis.*, vol. 32, no. 1, pp. 7–25, 1999.
[20] M. Ali, D. Clausi, and others, “Automatic registration of SAR and visible band remote sensing images,” in *Geoscience and Remote Sensing Symposium, 2002. IGARSS’02. 2002 IEEE International*, 2002, vol. 3, pp. 1331–1333.
[21] H. Cheng, S. Zheng, Q. Yu, J. Tian, and J. Liu, “Matching of SAR images and optical images based on edge feature extracted via SVM,” in *Signal Processing, 2004. Proceedings. ICSP’04. 2004 7th International Conference on*, 2004, vol. 2, pp. 930–933.
[22] B. Fan, C. Huo, C. Pan, and Q. Kong, “Registration of Optical and SAR Satellite Images by Exploring the Spatial Relationship of the Improved SIFT,” *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 4, pp. 657–661, Jul. 2013.
[23] T. D. Hong and R. A. Schowengerdt, “A robust technique for precise registration of radar and optical satellite images,” *Photogramm. Eng. Remote Sens.*, vol. 71, no. 5, pp. 585–593, 2005.
[24] H. Li, B. S. Manjunath, and S. K. Mitra, “A contour-based approach to multisensor image registration,” *Image Process. IEEE Trans. On*, vol. 4, no. 3, pp. 320–334, 1995.
[25] G. Palubinskas and P. Reinartz, “Template based matching of optical and SAR Imagery,” in *Joint Urban Remote Sensing Event (JURSE) 2015*, Lausanne, Switzerland, 2015, pp. 1–4.

[26] S. Auer, C. Gisinger, and J. Tao, "Characterization of Facade Regularities in High-Resolution SAR Images," *Geosci. Remote Sens. IEEE Trans. On*, vol. 53, no. 5, pp. 2727–2737, May 2015.
[27] 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.
[28] L. Schack, U. Soergel, and C. Heipke, "Persistent Scatterer Aided Facade Lattice Extraction in Single Airborne Optical Oblique Images," *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. II–3/W4, pp. 197–205, Mar. 2015.
[29] D. Brunner, G. Lemoine, and L. Bruzzone, "Earthquake Damage Assessment of Buildings Using VHR Optical and SAR Imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2403–2420, May 2010.
[30] 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.
[31] Y. Wang and X. Zhu, "Automatic Feature-based Geometric Fusion of Multi-view TomoSAR Point Clouds in Urban Area," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 8, no. 3, pp. 953 – 965, 2015.
[32] Z. Zhang, "Iterative point matching for registration of free-form curves and surfaces," *Int. J. Comput. Vis.*, vol. 13, no. 2, pp. 119–152, 1994.
[33] 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.
[34] 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.
[35] 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.
[36] X. Cong, U. Balss, 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.
[37] M. Eineder, U. Balss, C. Gisinger, S. Hackel, X. Cong, F. G. Ulmer, and T. Fritz, "TerraSAR-X pixel localization accuracy: Approaching the centimeter level," in *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*, 2014, pp. 2669–2670.
[38] I. Sobel, "An Isotropic 3x3 Image Gradient Operator," *Present. Stanf. AI Proj. 1968*, 1968.
[39] P. J. Huber, *Robust Statistics*. John Wiley & Sons, 1981.
[40] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust Estimation in Signal Processing: A Tutorial-Style Treatment of Fundamental Concepts," *IEEE Signal Process. Mag.*, vol. 29, no. 4, pp. 61–80, Jul. 2012.
[41] Y. Wang and X. X. Zhu, "Robust Estimators for Multipass SAR Interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 2, pp. 968–980, Feb. 2016.
[42] E. Ollila and V. Koivunen, "Influence functions for array covariance matrix estimators," in *Statistical Signal Processing, 2003 IEEE Workshop on*, 2003, pp. 462–465.
[43] M. Shahzad and X. X. Zhu, "Robust Reconstruction of Building Facades for Large Areas Using Spaceborne TomoSAR Point Clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 752–769, Feb. 2015.
[44] X. Zhu and M. Shahzad, "Façade Reconstruction Using Multiview Spaceborne TomoSAR Point Clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3541–3552, Jun. 2014.

[45] M. Shahzad and X. X. Zhu, “Automatic Detection and Reconstruction of 2-D/3-D Building Shapes From Spaceborne TomoSAR Point Clouds,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1292–1310, Mar. 2016.
[46] S. Katz, A. Tal, and R. Basri, “Direct visibility of point sets,” *ACM Trans. Graph. TOG*, vol. 26, no. 3, p. 24, 2007.
[47] 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.
[48] 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.
[49] 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.
[50] 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.
[51] X. Zhu, *Very High Resolution Tomographic SAR Inversion for Urban Infrastructure Monitoring: A Sparse and Nonlinear Tour*, vol. 666. Deutsche Geodätische Kommission, 2011.
[52] 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.
[53] 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.
[54] “https://www.pix4d.com/.” 2015.
[55] 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, Jan. 2016.
[56] C. Gisinger, U. Balss, R. Pail, X. X. Zhu, S. Montazeri, S. Gernhardt, and M. Eineder, “Precise Three-Dimensional Stereo Localization of Corner Reflectors and Persistent Scatterers With TerraSAR-X,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1782–1802, Apr. 2015.
[57] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, “Visual categorization with bags of keypoints,” in *Workshop on statistical learning in computer vision, ECCV*, 2004, vol. 1, pp. 1–2.
[58] C. Cortes and V. Vapnik, “Support-vector networks,” *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.
[59] B. F. Andrea Vedaldi, “VLFeat: an open and portable library of computer vision algorithms,” in *Proceedings of the 18th International Conference on Multimedea 2010*, Firenze, Italy, 2010, pp. 1469–1472.
[60] A. D. Kerr, “Analysis of thermal track buckling in the lateral plane,” *Acta Mech.*, vol. 30, no. 1–2, pp. 17–50, 1978.
[61] K. Bredies, K. Kunisch, and T. Pock, “Total Generalized Variation,” *SIAM J. Imaging Sci.*, vol. 3, no. 3, pp. 492–526, Jan. 2010.
[62] F. Knoll, K. Bredies, T. Pock, and R. Stollberger, “Second order total generalized variation (TGV) for MRI,” *Magn. Reson. Med.*, vol. 65, no. 2, pp. 480–491, 2011.
[63] X. Zhu, N. Ge, and M. Shahzad, “Joint Sparsity in SAR Tomography for Urban Mapping,” *IEEE J. Sel. Top. Signal Process.*, vol. PP, no. 99, pp. 1–12, 2015.
[64] J. Kang, Y. Wang, M. Köner, and X. X. Zhu, “Object-based InSAR Deformation Reconstruction with Application to Bridge Monitoring,” in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2016*, Beijing, China.

Appendices

D Wang Y, Zhu X, Bamler R (2012) Retrieval of Phase History Parameters from Distributed Scatterers in Urban Areas Using Very High Resolution SAR Data. ISPRS Journal of Photogrammetry and Remote Sensing, 73: 89 - 99.

http://www.sciencedirect.com/science/article/pii/S0924271612001219

ISPRS Journal of Photogrammetry and Remote Sensing 73 (2012) 89–99

Contents lists available at SciVerse ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Retrieval of phase history parameters from distributed scatterers in urban areas using very high resolution SAR data

Yuanyuan Wang [a,*], Xiao Xiang Zhu [a,b], Richard Bamler [a,b]

[a] Lehrstuhl für Methodik der Fernerkundung, Technische Universität München, Arcisstraße 21, 80333 Munich, Germany
[b] Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, 82234 Wessling, Germany

ARTICLE INFO

Article history:
Available online 1 August 2012

Keywords:
Phase history
Distributed scatterer
Covariance matrix
SAR
PSI
TerraSAR-X

ABSTRACT

In a recent contribution Ferretti and co-workers (Ferretti, A., Fumagalli, A., Novali, F., Prati, C., Rocca, F., Rucci, A., 2011. A new algorithm for processing interferometric data-stacks: SqueeSAR IEEE Transactions on Geoscience and Remote Sensing 49(9), pp. 3460–3470) have proposed the SqueeSAR method, a way to exploit temporally coherent distributed scatterers in coherent SAR data stacks. Elevation and deformation or subsidence estimates are obtained with accuracy similar as in the well known persistent scatterer interferometry (PSI).

In this paper we propose an alternative approach and provide a first demonstration of the optimal estimation of distributed scatterers' phase histories in urban areas. Different to SqueeSAR, we derive phase histories for each distributed scatterer pixel rather than for groups of pixels. We use the Anderson–Darling statistical test to identify neighboring samples of the same distribution. Prior to covariance matrix estimation required for maximum likelihood estimation we apply a multi-resolution defringe technique. By using TerraSAR-X high resolution spotlight data, it is demonstrated that we are able to retrieve reliable phase histories and motion parameter estimates from distributed scatterers with signal-to-noise-ratio far below the common range.

1. Introduction

Persistent scatterer interferometry (PSI) (Ferretti et al., 2001; Kampes, 2006) exploits the phase history of strong and long term stable scatterers, i.e. the so-called persistent scatterers (PSs), over the entire period of monitoring. Commonly, PSs are selected according to their amplitude dispersion index or their signal-to-clutter ratio (SCR). They are usually seen as bright pixels in synthetic aperture radar (SAR) images. However these criteria are only indications but not necessarily measurements of the decorrelation. Common PSs in real world are stable objects such as window sills on building façades, building corners, and exposed rocks. By proper modeling the phase histories of PSs, the elevation and the time series of the deformation of each PS can be estimated. PSI can provide both large-scale and long-term deformation monitoring with an accuracy of up to mm/year. This is particularly true for urban areas where the PS density is high. As a price for its efficiency, first, PSI considers only a single dominant target within a SAR resolution cell for each PS which forbids layover separation; and second, only permanently reliable scatterers are exploited. This makes PSI an opportunistic approach and functionally limits the density of the useful scatterers.

The following development of SAR interferometry (InSAR) techniques basically targets at understanding the interaction of the SAR signal within a resolution cell, and at optimally exploring coherent objects other than detected PSs, such as layovered scatterers and distributed scatterers (DSs).

In the first aspect, the authors of (Ferretti et al., 2001) extended PSI by considering phase models for multiple scatterers within a pixel (Ferretti et al., 2005). Or to be more general, differential SAR tomography (D-TomoSAR) (Lombardini, 2005; Fornaro et al., 2009; Zhu and Bamler, 2010, 2011) provides a general solution for such case and avoids the pre-selection of coherent scatterers. D-TomoSAR can separate multiple scatterers at different elevation with possible motion inside a pixel by means of spectral estimation. D-TomoSAR is able to retrieve topography, deformation parameters, or even the entire phase history of multiple scatterers within a resolution cell. Yet compared to PSI, D-TomoSAR is relatively computationally expensive. In the second aspect, the small baseline subset (SBAS) technique (Casu et al., 2005) overcomes the second disadvantage of PSI by further exploiting "temporary" PSs. That is to say, instead of looking for permanently coherent scatterers with respect to a single master acquisition, SBAS takes

* Corresponding author. Tel.: +49 8153 28 2386.
E-mail address: yuanyuan.wang@dlr.de (Y. Wang).

0924-2716/$ - see front matter © 2012 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS) Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.isprsjprs.2012.06.007

advantage of high coherence in pairs of acquisitions with small baselines (both temporal and spatial). In a popular notion, SBAS builds up interferograms between "good" pairs of acquisitions. Then the processing of each set of interferograms with a common master follows standard PSI procedures. Based on these differential results from numbers of interferograms, an inversion of an over-determined linear equation system gives a complete time series of the phase history of a scatterer. SBAS theoretically increases the density of useful scatterers, and decreases the number of images required (Casu et al., 2005).

However, in non-urban environment where surface or volume scattering dominates, PSs density can be very low and DSs are the majority. Even in urban areas, asphalt roads and concrete surfaces seen as DS in X-band also make up a significant percentage. The question then comes to optimally harvest the information from a DS, i.e. how to retrieve its phase history, and model parameters such as elevation and deformation velocity. The challenges are the following. DSs tend to decorrelate temporally, i.e. interferometric DS pixels have usually much higher phase noise than PSs. Also decorrelation may be different for each of the interferograms in the stack.

The SqueeSAR technique (Ferretti et al., 2011a, 2011b) provides a possible solution for these problems. It uses the fact that a DS pixel is usually surrounded by many pixels from the same distribution, also called "brother pixels" in (Ferretti et al., 2011a). SqueeSAR detects these brotherhood segments, uses them for estimation of the covariance matrix and obtains the maximum likelihood estimation (MLE) of the phase history of the segment (De Zan and Rocca, 2005; Seymour and Cumming, 1994). Since SqueeSAR treats each brotherhood segment as a single object, spatial resolution of the estimated phase history is compromised.

This article presents the first demonstration of exploiting DS for urban monitoring using very high resolution SAR image stacks. We estimate elevation and motion history of each DS pixel in an optimum way. Following the basic concept of SqueeSAR, we introduce several new aspects:

- Instead of the brotherhood segments as single objects we retrieve the phase history for each DS pixel.
- For identifying the brotherhood pixels we use a more robust statistical test, i.e. the Anderson–Darlington (AD) test.
- For covariance matrix estimation we introduce a multi-resolution interferometric phase flattening (defringe) algorithm to determine the underlying topographic and motion-induced phase components.

The remainder of this article starts with the explanation of our covariance matrix estimation procedures, followed by phase history retrieval and its model parameters estimation by MLE. Results using TerraSAR-X (TS-X) high resolution spotlight data are present as last.

2. Covariance matrix estimation

The covariance between two single-look complex (SLC) measurements i_n and i_k in image n and k is defined as the expectation of the product of one with the complex conjugate of the other:

$$c_{n,k} = E(i_n i_k^*) \tag{1}$$

Normalizing the covariance by its standard deviation yields the complex correlation coefficients, usually referred to as coherence (Bamler and Hartl, 1998):

$$\gamma_{n,k} = \frac{E(i_n i_k^*)}{E(|i_n|)E(|i_k|)} \tag{2}$$

The MLE of the covariance is obtained by assuming ergodicity and replacing the expectation with averaging M independent samples:

$$\hat{c}_{n,k} = \frac{1}{\mathrm{M}}\sum_m i_n(m) i_k^*(m) \tag{3}$$

and for coherence:

$$\hat{\gamma}_{n,k} = \frac{\sum_m i_n(m) i_k^*(m)}{\sqrt{\sum_m |i_n(m)|^2 \sum_m |i_k(m)|^2}} \tag{4}$$

The covariance matrix $\mathbf{C}$ containing all the covariances is then

$$\mathbf{C} = \mathbf{Z}\mathbf{Z}^{\mathrm{H}} \tag{5}$$

where

$$\mathbf{Z} = \begin{bmatrix} i_1(1) & i_1(2) & \cdots & i_1(\mathrm{M}) \\ i_2(1) & i_2(2) & \cdots & i_2(\mathrm{M}) \\ \vdots & & \ddots & \vdots \\ i_{\mathrm{N}}(1) & i_{\mathrm{N}}(2) & \cdots & i_{\mathrm{N}}(\mathrm{M}) \end{bmatrix} \tag{6}$$

However, this estimator is biased at low magnitudes of coherence for small M (Touzi et al., 1999; Bamler and Hartl, 1998) According to the Reed-Mallett-Brennan (RMB) detection loss theory (Reed et al., 1974), to maintain a loss of likelihood (see Eq. (7)) of the estimates of less than 3 dB, M should be larger than $2N$, where N is the number of measurements:

$$\mathrm{loss} = -10\log_{10}[(M+2-N)/(M+1)] \tag{7}$$

In our experiment, we try to ensure $M = 3N$ as much as possible. For low coherence bias correction, we adopt a method explained in (Zebker and Chen, 2005) which makes use of an empirical function of estimated coherence with respect to the theoretical one.

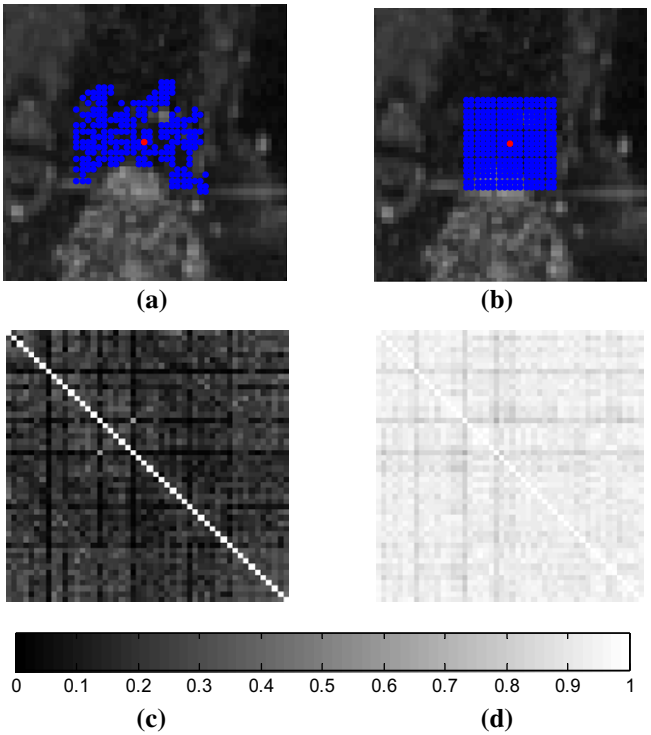


Fig. 1. (a) and (b): Samples selected by adaptive and rectangle sample selection strategy, respectively. The red dot is the target pixel, while those in blue are the selected pixels, i.e. The neighboring pixels sharing the same statistical distribution as the target pixel. (c) and (d): Corresponding coherence matrices.

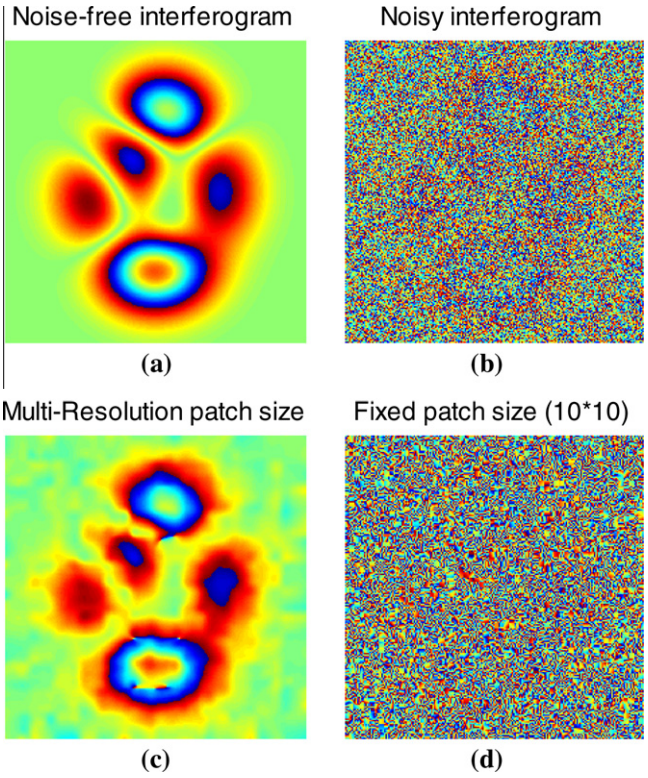


Fig. 2. (a) Simulated noise-free interferogram, (b) simulated noisy interferogram. 30% Of the pixels are DS with coherence = 0.3 corresponding to a SNR of −3.7 dB in the two SLC images, and the rest 70% are noisy pixels with coherence close to 0, (c) fringes estimated by adaptive multi-resolution method, and (d) fringes estimated using single fixed patch size (10×10 pixels).

Other than this systematic drawback of the covariance estimator, it still faces two major practical issues. Especially in high resolution SAR images of urban areas, the ergodicity of the selected samples is hardly ensured. This jeopardizes the covariance estimation in two ways. First, carelessly selecting these samples, such as using a simple rectangular estimation window (commonly known as box-car method) brings the risk of averaging pixels of different distributions, and hence different objects, yielding a wrong estimate of the covariance matrix. And second, the spatially varying phase within the estimation window caused by topography or motion diminishes the summation in Eqs. (3) and (4), leading to an underestimation of the covariance or coherence. Therefore, this phase needs to be removed before the covariance estimation. The covariance estimator in Eqn. (3) then becomes:

$$\hat{c}_{n,k} = \frac{1}{M}\sum_{m} i_n(m) i_k^*(m) \exp(-j(\bar{\phi}_n(m) + \bar{\phi}_k(m))) \tag{8}$$

where $\bar{\phi}_n$ and $\bar{\phi}_k$ are the expected phases of i_n and i_k.

In Eq. (3), the phase of $c_{n,k}$ is the expected phase difference of i_n and i_k. On one hand, we must remove the deterministic part of the phase of i_n and i_k for better covariance estimation, and on the other hand we are exploring this interferometric phase from $c_{n,k}$ for model parameters estimation.

To summarize, the covariance estimation procedure is a two-step approach: (1) adaptive sample selection by AD test on the amplitudes of the samples, and (2) multi-resolution defringe to remove the underlying expectation of the phases. These two steps will be discussed in detail in the following two sections.

2.1. Adaptive sample selection

The complex response of a DS in a SAR image is commonly modeled as circular Gaussian variate that widely holds for medium or lower resolution SAR data. However, due to limited number of natural scatterers in a resolution cell in high resolution SAR systems, not fully developed speckle results an imperfect circular Gaussian distribution. Yet we still favor the Gaussian model, as the best asymptotical one, for conveniences in the estimation (i.e. only the covariance is needed).

The MLE of each DS pixel value in the stack is found via inversion of its covariance matrix. As mentioned before, estimation of the covariance matrix elements requires an ensemble of samples of the same distribution. Selecting samples using a rectangular window is justified sometimes, if the rectangle size is small enough and loss of resolution is affordable. However, this is never the case for high resolution images of urban areas, e.g. any PS in the estimation window will dominate the covariance or the coherence estimate of the surrounding pixels. Not only will the high resolution

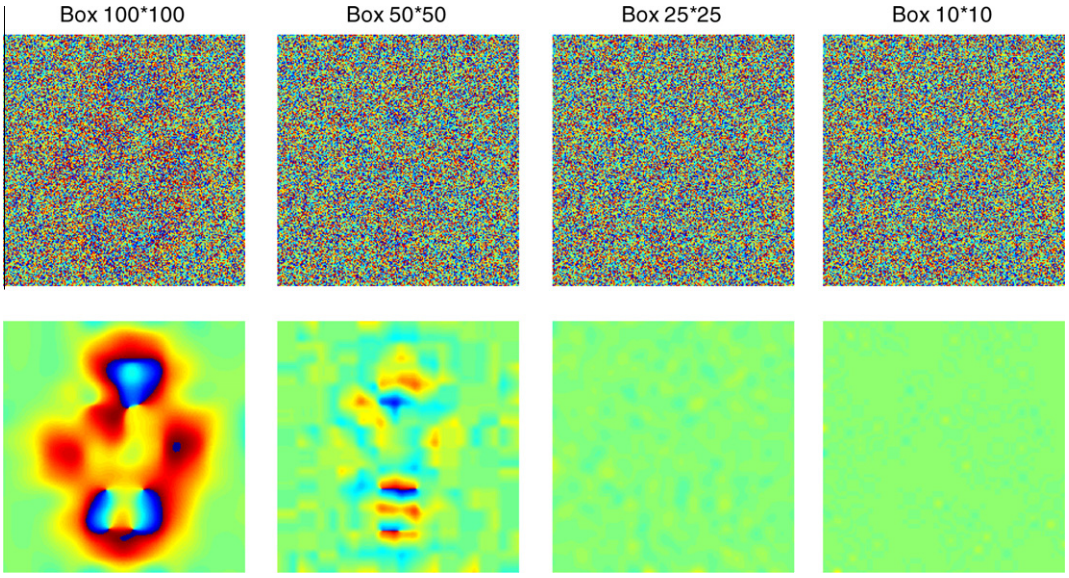


Fig. 3. Input noisy interferogram (first row) and estimated fringes (second row) at each resolution level. The estimated fringes are then subtracted from the noisy interferogram. The residual phase is taken as the input for the next resolution level with smaller window size.

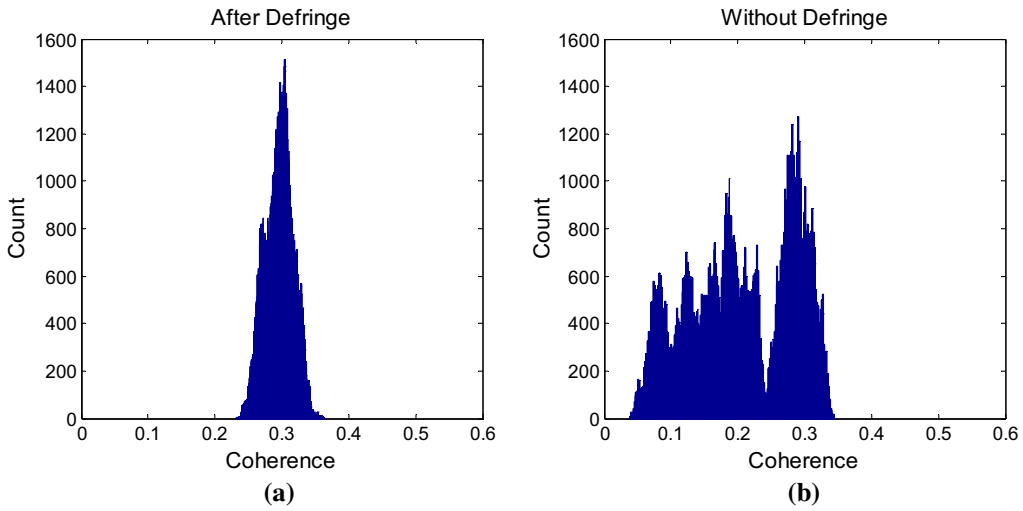


Fig. 4. Histograms of coherence estimates of total 107,133 DSs in the simulated interferogram (a) with defringe performed which stick close to the preset value 0.3, and (b) without defringe performed which severely underestimates the coherence.

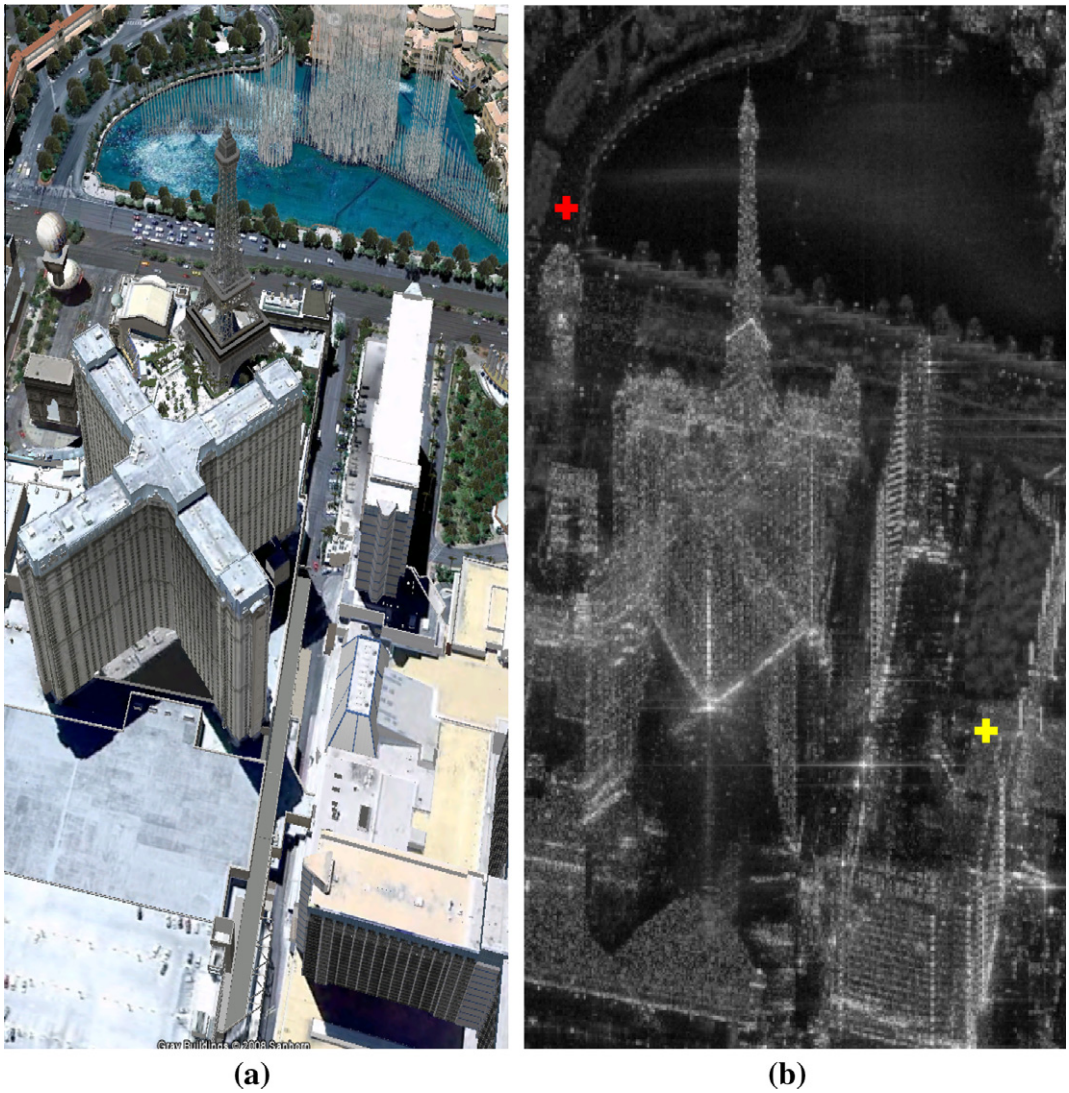


Fig. 5. (a) Optical image with 3D building model from Google Earth™, (b) mean SAR amplitude of the test site. The red and yellow crosses mark the analyzed pixels in Fig. 8.

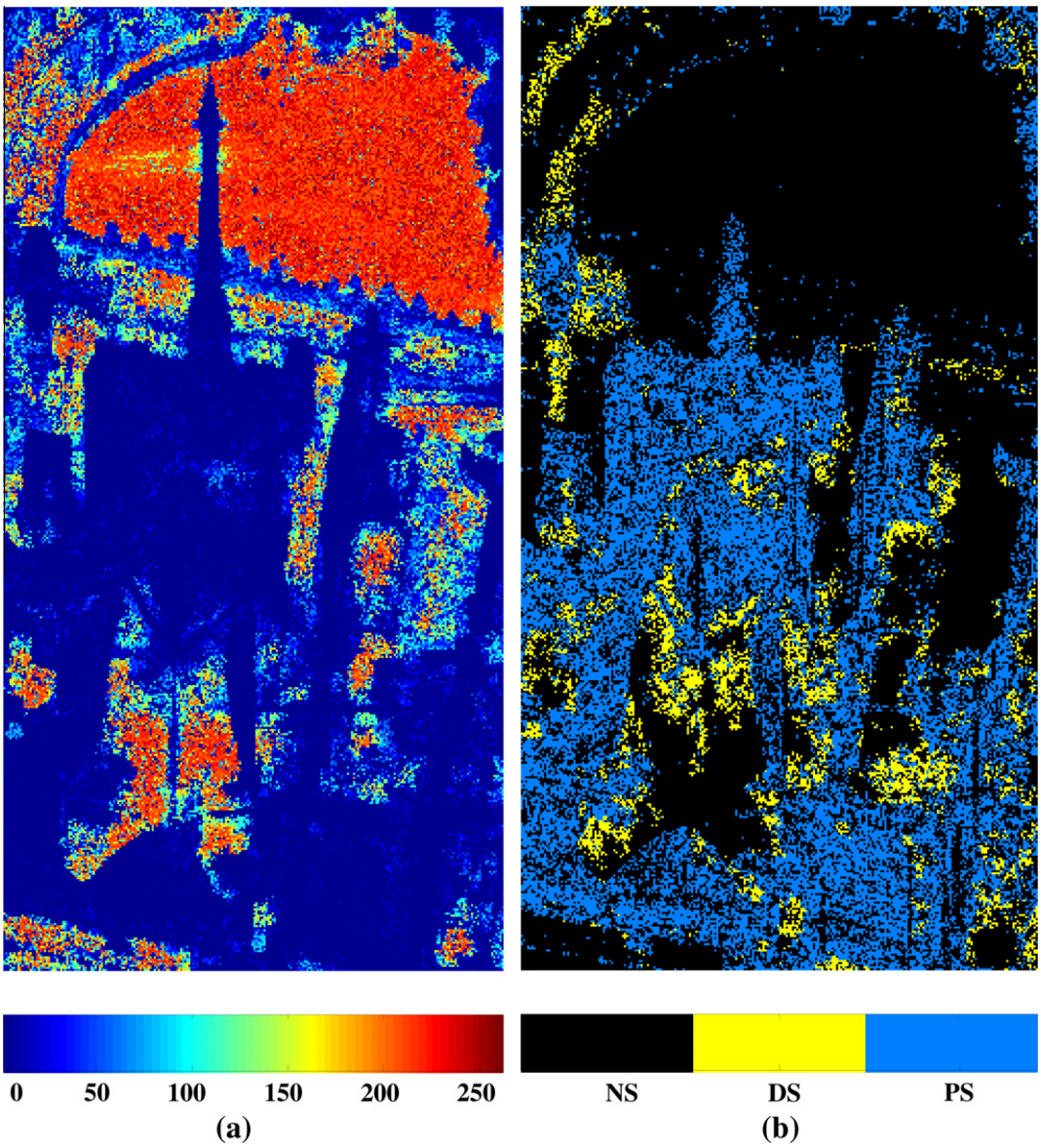


Fig. 6. (a) Number of brother pixels detected within a 20 × 20 pixel box for each pixel of the area shown in Fig. 5. The number is color coded. Red indicates a large number of samples and blue indicates a small number, (b) pixel classification map. The blue pixels indicate the PSs, the yellow the DSs, and the black the NS.

be compromised, but the covariance or coherence of a DS close to a PS will also be severely overestimated.

In (Ferretti et al., 2011a) this problem is tackled by introducing the *DespecKS* method, which detects the samples adaptively by performing the Kolmogorov–Smirnov (KS) test. The KS test compares the maximum distance between the empirical distribution functions (EDFs) of two data sets which, in our case, are the intensities of one pixel throughout the whole stack with another pixel. By setting a threshold on the distance, the hypothesis that the two pixels belong to the same distribution is accepted or rejected. However, the maximum distance of two EDFs usually happens in the midway of the curve, regardless of the shape of the distribution. Therefore, the KS test is not sensitive to unknown distributions with variations on the tails.

We adopt this idea of amplitude based test. Further considering the fact mentioned in the beginning of this section that fully developed speckle can be compromised in high resolution image and Rayleigh distribution is therefore often not satisfied (Jakeman and Pusey, 1976; Jao, 1984), more sophisticated non-parametric tests are more appropriate for our purpose. In (Parizzi and Brcic, 2010), different tests are evaluated. Among them, the AD test offers the best detection rate at a constant false alarm rate. Therefore, we replace the KS test by the AD test.

The AD test belongs to the class of quadratic EDF statistics. To compare the probability distributions of pixels a and b, we first compute the EDFs $F_a(x)$ and $F_b(x)$ of the amplitude values of the pixels through the stack. Then the AD test is a weighted L2-norm of the difference between the two EDFs:

$$A^2 = \frac{N}{2}\sum \frac{[F_a(x) - F_b(x)]^2}{F_{a\cup b}(x)[1 - F_{a\cup b}(x)]} \tag{9}$$

where $F_{a\cup b}(x)$ is the EDF of the combined pixel values. N again represents the number of images in our stack. The denominator $F_{a\cup b}(x)[1 - F_{a\cup b}(x)]$ places more weight on the head and tails of the distributions.

The test is performed for every pixel to identify neighboring pixels of the same distribution, i.e. the brother pixels. These form

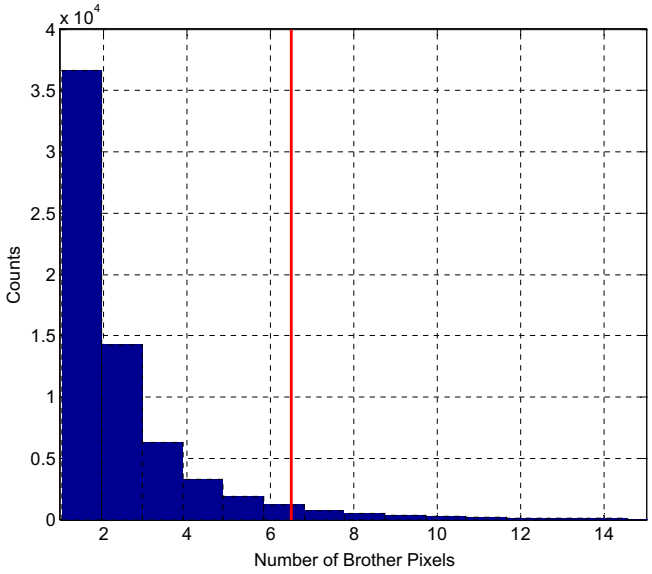


Fig. 7. Histogram of the number of brother pixels of PSs selected by thresholding the amplitude dispersion index. 95% of the PSs have less than 6 brother pixels. The red line indicates is the cut-off of 95% of all the PSs.

the estimation window. The window size and shape can vary from one pixel to another. This is unlike segmentation where each pixel belongs to a unique class. In our algorithm one pixel is included in different brotherhoods.

Fig. 1a is an example of adaptive sample selection using the AD test on a stack of high resolution TS-X images of downtown Las Vegas. The test area is near the red cross marked in Fig. 5b. The red dots in Fig. 1a and b are our target DS pixel located on the asphalt road between two structures, and the blue ones are the detected brother pixels. As a comparison, samples selected by a rectangular window are marked in Fig. 1b. The adaptive window follows the structure boundary well, while using a rectangle includes all the pixels from both the structures and the road, regardless of their varying amplitudes. The corresponding coherence matrices estimated using the selected samples are shown in the second row of Fig. 1. Since the target pixel is on an asphalt surface having a very low SNR, its coherence is expected to be low. Yet, using the rectangular window, the coherence is heavily overestimated (>0.7), as shown in Fig. 1d. Obviously, we are more confident with the results using the adaptive selection window.

2.2. Adaptive multi-resolution defringe

Defringe, also known as phase flattening, is to remove an estimate of the expectation, i.e. a low-pass version, of the interferometric phase before covariance estimation (Eq. (8)). In most of the cases a constant local fringe frequency in range and azimuth directions is assumed. The fringe frequency is found by simply searching the maximum coefficient in the Fourier transform of local patches (Bamler and Hartl, 1998; Zebker and Chen, 2005).

Yet the assumption of a constant frequency is not always valid, especially in urban areas with fast varying topography. And the local fringe frequency cannot be correctly estimated at areas where decorrelation noise dominants. We make use an adaptive multi-resolution defringe algorithm, firstly introduced for phase unwrapping (Davidson and Bamler, 1999). It overcomes the aforementioned two problems by on one hand demodulating the interferogram multiple times by estimating local fringes in patches of difference sizes, and on the other hand taking into account only the pixels with sufficiently high coherence.

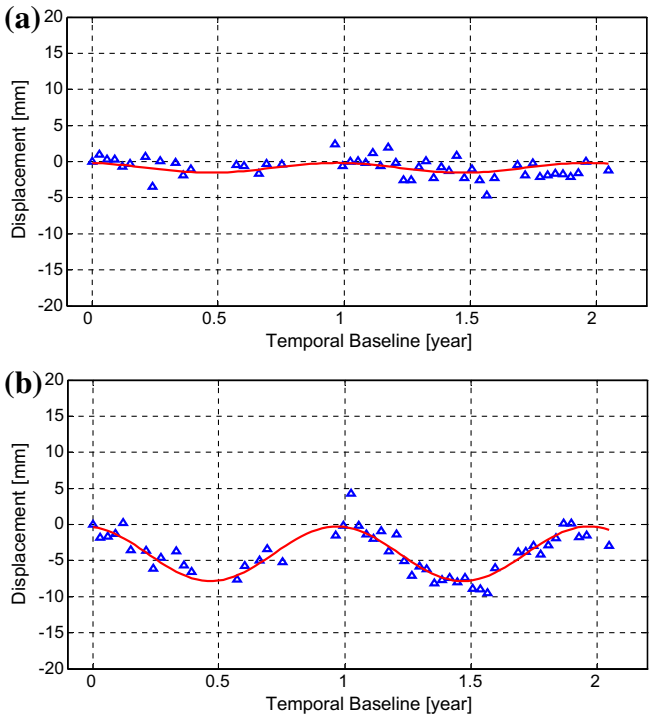


Fig. 8. (a) Phase history of one DS located on the ground, indicated by the red cross in Fig. 5 (b). Its coherence equals 0.16. This pixel experiences no significant deformation, (b) phase history of another DS located on the roof of a building, which undergoes a seasonal motion with 4.5 mm in amplitude. Its coherence is 0.29. This pixel is indicated as the yellow cross in Fig. 5 (b).

We test this algorithm on a simulated interferogram of 600×600 pixels. The simulated interferometric phase is shown in Fig. 2b. The pixels in the interferogram are either DS with coherence equal to 0.3, i.e. SNR = -3.7 dB in both SLC images, or noise (NS) with coherence close to 0. We uniformly distribute 30% of the pixels as DS, and the rest 70% to be NS. This setting simulates the worst real case scenario where no high coherence scatterer is available, but only low coherence targets are sparsely scattered in the scene. The noise-free and noisy interferogram are collocated in Fig. 2a and b. The fringe pattern is hardly distinguishable in the noisy interferogram.

We apply the multi-resolution algorithm to the noisy interferogram. Fig. 3 shows the input and estimated fringes at each resolution level. Starting from the largest patch size, the fringe frequency is estimated using only the DSs in a patch. The interferogram is then demodulated, and passed to the next resolution level. In our test, we set four resolution levels: 100×100, 50×50, 25×25, and 10×10 pixels. The summation of the phase contributions at all the resolution levels gives the final result in Fig. 2c. For comparison, Fig. 2d shows the result for the usual fixed window size and without employing the pre-selection of coherent pixels. This method completely fails to reconstruct the original interferometric phase, while the multi-resolution algorithm reduces the estimation error tremendously and retains higher correlation with the noise-free interferogram.

After defringing, the coherences of all the DSs in the interferogram are evaluated. The results show consistent and accurate solutions with respect to the preset coherence of 0.3. 99% of the results lie in the interval of 0.3 ± 0.05. This result is compared to the coherence estimated without defringe. Fig. 4 plots the histogram of the estimated coherence, with defringe and without defringe, of total 107,133 DSs. It is obvious that the coherence is severely underestimated if defringe is not performed.

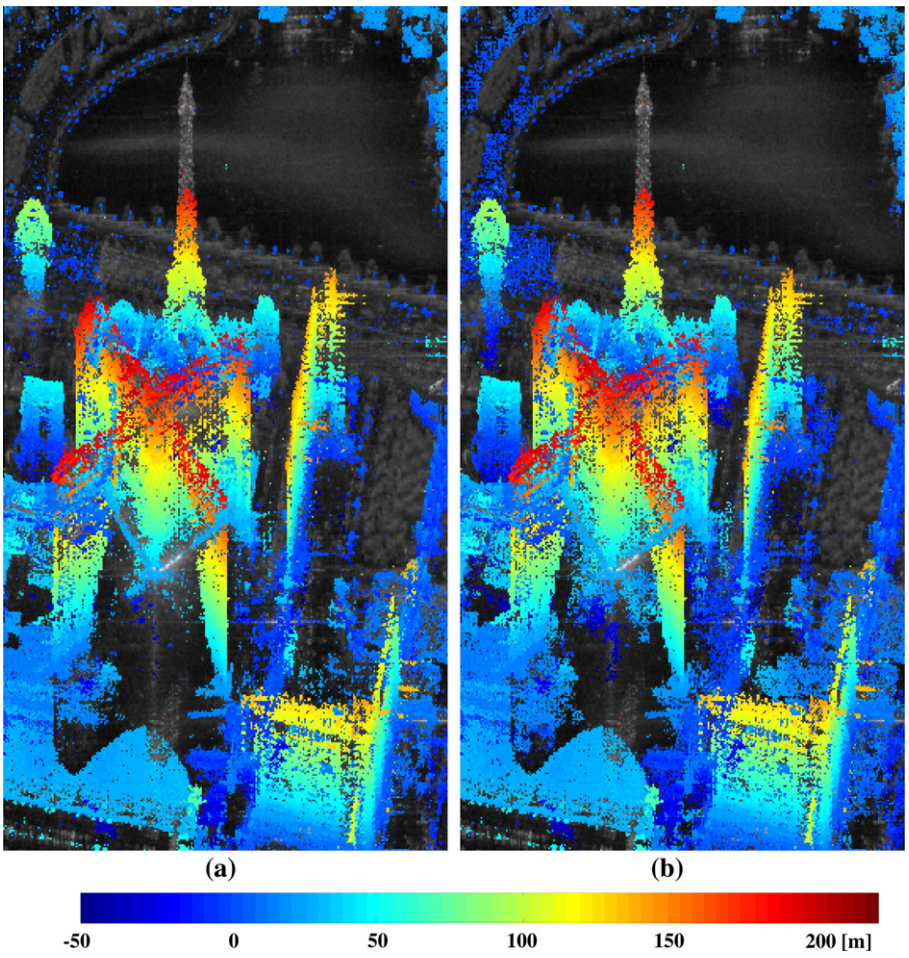


Fig. 9. Estimated elevation at (a) only PSs, and (b) both PSs and DSs.

3. Parameters estimation

3.1. Maximum likelihood estimator

We use MLE for parameter estimation for each DS pixel. We make use of the objective function developed in (De Zan and Rocca, 2005):

$$\{\hat{h},\hat{a}\} = \arg\max\left\{\frac{\exp(-\mathbf{i}^{\mathrm{H}}\boldsymbol{\Phi}(h,a)\mathbf{C}_0^{-1}\boldsymbol{\Phi}(h,a)^{\mathrm{H}}\mathbf{i})}{\det|\mathbf{C}_0|\pi^N}\right\} \tag{10}$$

where i is the complex measurements vector; $\boldsymbol{\Phi}$ is diagonal matrix containing the modeled phase; $\mathbf{C}_0$ is absolute value of the covariance matrix; h, a is height above reference surface, line-of-sight (LOS) motion parameter such as linear deformation velocity and seasonal displacement amplitude and N is the number of images

In the likelihood function, the covariance matrix $\mathbf{C}$ is decomposed to $\boldsymbol{\Phi}(h,a)\mathbf{C}_0^{-1}\boldsymbol{\Phi}(h,a)^{\mathrm{H}}$. The diagonal matrix $\boldsymbol{\Phi}$ contains the phase of the model. The denominator $\det|\mathbf{C}_0|\pi^N$ can be ignored: it is a constant for each pixel.

4. Experiments

4.1. Test site 1 – Paris Las Vegas hotel

We apply the proposed algorithm to a test site about 500 × 500 m centered at Paris Las Vegas hotel in downtown Las Vegas cropped from a stack of co-registered 50 TS-X high resolution spotlight images. Fig. 5a and b are the optical image from Google Earth™ and the mean amplitude image of the test area, respectively. In this test case, the area consists of different types of scattering objects, i.e. partly building façades (probable PS candidates), partly asphalt roads and rough surface (probable DS candidates), and a large fraction of totally decorrelated scatterers (e.g. the fountain on the top of the scene, and vegetated areas).

The pre-processing, including amplitude calibration and atmospheric phase screen (APS) correction, is performed using the PSI-GENESIS system from the German Aerospace Center (DLR) (Kampes, 2006; Adam et al., 2003). We firstly classify all the pixels to PS, DS, or NS, by jointly considering the coherence and the number of brother pixels. Based on the classification, we process PSs using the standard PSI technique, and DSs using the proposed algorithm.

4.2. Pixel classification

To classify PS, DS and NS, we firstly identify the PSs by making use of the property explained in (Ferretti et al., 2011a) that most of the PSs possess very few brother pixels. Showing in Fig. 6, the number of the brother pixel map, most of the building structures show up in blue which indicates a small number of brother pixels, in contrast to the orange color for the rest of the image. Statistics tell that 95% of the PSs (selected by thresholding on amplitude dispersion index) have less than 6 samples (see Fig. 7). Secondly, the DSs are distinguished from the rest of the pixels by thresholding on the coherence. The threshold is set at the unbiased lower limit of the coherence estimator in Eq. (4) using 200 looks. The final classifica-

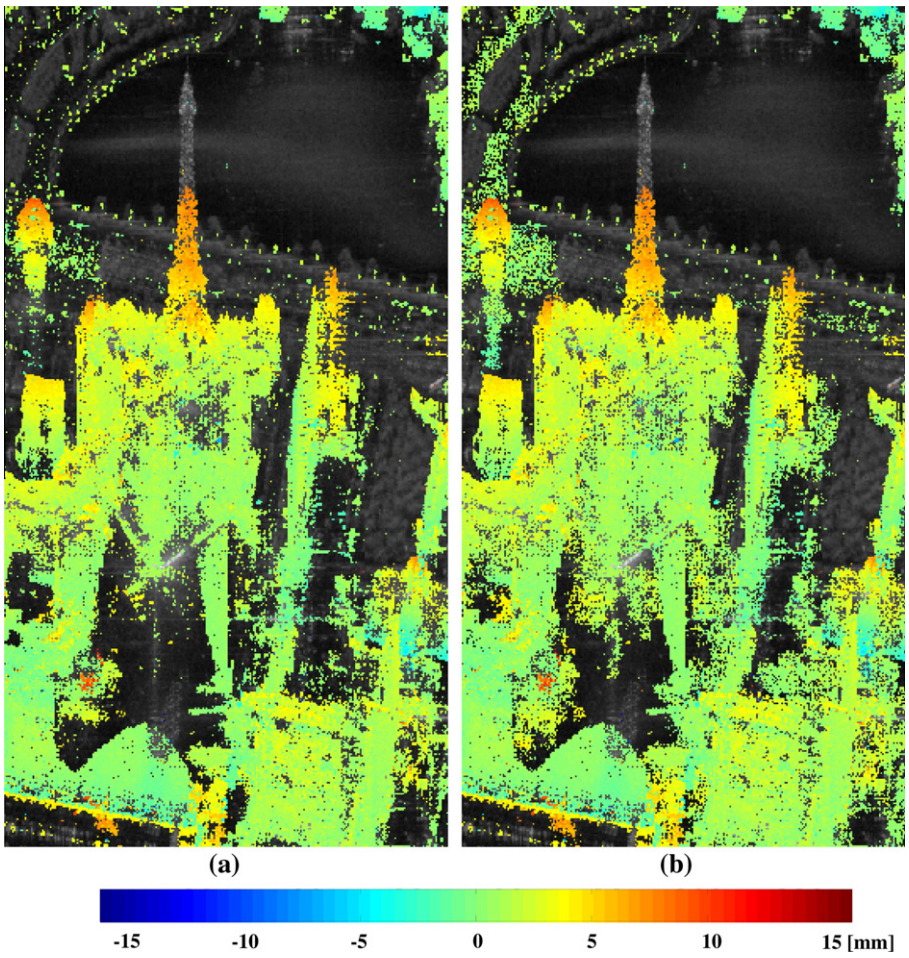


Fig. 10. Estimated seasonal deformation amplitude at (a) only PSs, and (b) both PSs and DSs.

tion is shown in Fig. 6b, where PS, DS, and NS are marked as different colors. In this test site, the number of DSs is about 20% of PSs.

4.3. Retrieved phase history

According to the classification map, we estimate the covariance matrix for each DS after sorting all the interferograms along the temporal baseline. The phases of the covariance matrix's first column represent the expected interferometric phases referring to the first image. Since defringe was applied prior to the covariance estimation, the pre-removed phase was added back to get the final phase history of each pixel. Fig. 8 plots the phase histories of two DSs (indicated as red and yellow crosses in Fig. 5b). Seen from that, the first one is located on the asphalt road at ground level, with presumably no significant deformation. The reflectivity of such material is generally very low, as we can see the surrounding pixels are consistently dark. For this pixel, the coherence is 0.16, equivalent to an average SNR of −7.1 dB. Its phase history shows almost no deformation throughout the two years' time span. The second one is on the roof of a building, possibly subject to thermal expansion. It has a better coherence of 0.29, corresponding to an average SNR of −4 dB. It undergoes a seasonal deformation of 4.5 mm in amplitude. The red curves in the figures are the fitting of the sinusoidal deformation model.

4.4. Estimated parameters

Following the retrieved phase history, we estimate the elevation and seasonal motion amplitude for each DS. The result is combined with the estimates from PSs using the standard PSI procedures. All the estimates are referred to a single reference point. Figs. 9a and 10a give the estimates from PSs only, and Figs. 9b and 10b are PSs and DSs combined. We see a linear gradient from the base of the buildings to the top in Fig. 9, with blue indicating low elevation, and red for high. A similar gradient can also be observed in the seasonal motion amplitude estimate in Fig. 10. The seasonal motion amplitude becomes more positive as the elevation gets higher representing displacement towards the sensor. This is mostly caused by the natural expansion and contraction of the building itself, but also subject to the effect of height dependent APS to a maximum of about 1.5 mm per 100 m in vertical height (Cong and Eineder, 2012).

The PSs already provide very good coverage over most parts of the building façades, except for the upper half of the tower due to its very complex motion. The DSs give extra information over the rest of the image. The contributions of DSs are scattered over the whole area, including areas on the upper left corner and area close to the center of the image. Most of the DSs are on the ground level, indicated by the blue color. Their deformation amplitudes are generally small (green). It is worthy to point out that seemingly reliable estimates also found in the shadow area. They are "ghost scatterers" caused by mirror reflections of the façade at the ground (Auer et al., 2011). The variances of these DS estimates are comparable with those of PS estimates, although the coherences of these DSs are much lower. Nevertheless, the MLE of the phase history of DS is obtained by covariance matrix estimation based on many samples. The residual of the phase history to the sinusoidal model is about 1–2 mm, thus providing a reliable estimation of the model parameters.

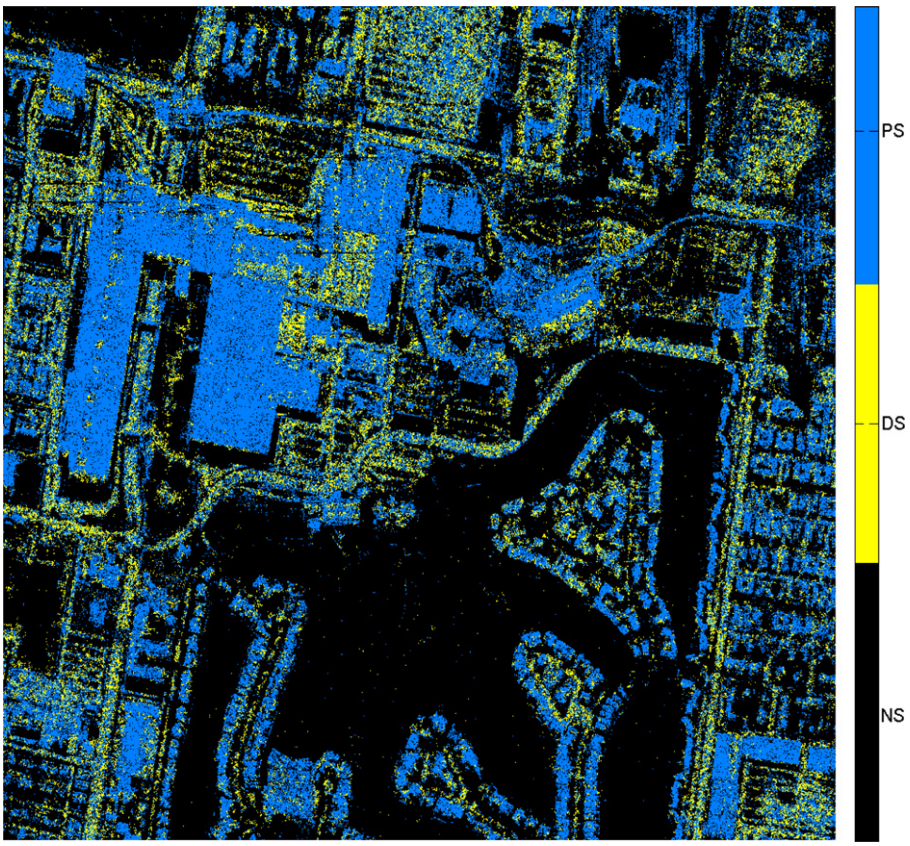


Fig. 11. Pixel classification map. The blue pixels indicate the PSs, the yellow the DSs, and the black the NS.

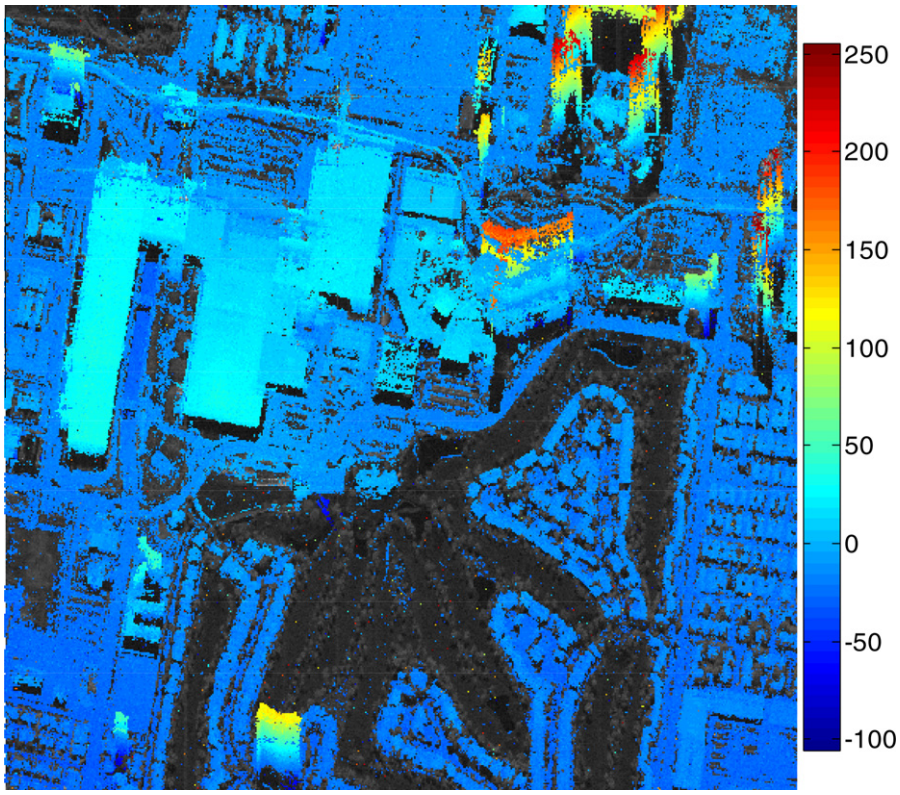


Fig. 12. Estimated elevation of both PSs and DSs of the second test site. The unit is in meter.

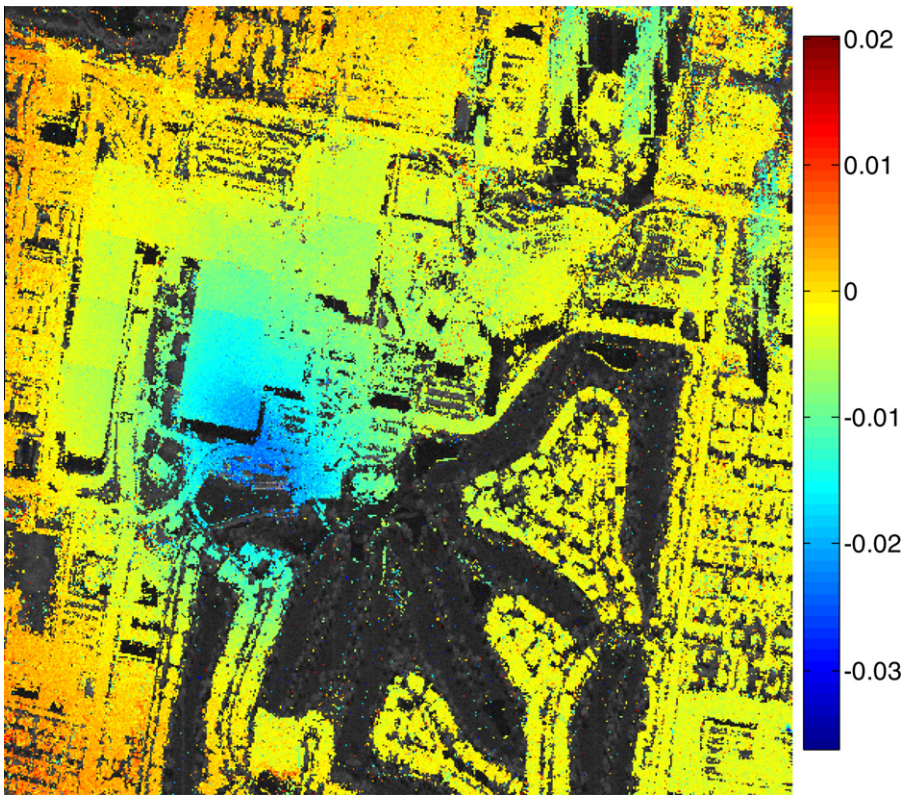


Fig. 13. Estimated linear deformation rate of both PSs and DSs. The unit is in m/year.

4.5. Test site 2 – Las Vegas convention center

A larger area about 1 × 2 km around the Las Vegas convention center is chosen as the second test site due to a known subsidence discovered in (Zhu and Bamler, 2011). The same processing procedures are applied. The pixel classification is shown in Fig. 11. Blue color (PS) fills up most of the building structures, while yellow (DS) appears mostly on the road. The park located on the lower half of the image is almost black due to temporal decorrelation of vegetation. In Figs. 12 and 13, the estimation results of PSs and DSs combined are demonstrated. The elevation estimates look consistent and reliable for both PS and DS, except part of the building façades due to the layover problem. The linear deformation rate estimates show a subsidence pattern centered at Las Vegas convention center with a maximum of about 15 mm/year.

5. Conclusion and outlook

This article presents optimal DS phase history and model parameters retrieval in urban areas using very high resolution TS-X spotlight data. To preserve the high spatial resolution, particular focus is put on the phase history retrieval for each DS pixel. We emphasize the critical problem of accurate covariance matrix estimation. A solution is provided by using adaptive sample selection and a multi-resolution defringe method. The adaptive sample selection method is proven to be an essential procedure in the covariance matrix estimation. The AD test performs well in distinguishing targets of different scattering characteristics. The adaptive multi-resolution defringe algorithm outperforms the method with a fixed window size in terms of fringe estimation accuracy and robustness to noise. The effectiveness of the proposed method is tested using simulated and real TS-X data. Periodic seasonal motion histories of sample DS have been derived. The residuals to the sinusoidal model were in the order of 1–2 mm.

In the future, as a more general solution, we will improve the DS covariance matrix estimation using a "fuzzy brotherhood" concept, i.e. using surrounding pixels with weightings derived from a statistical test instead of using a fixed threshold. The improved algorithm will be applied to non-urban area, and possible mixed single- and repeat-pass data stacks.

Acknowledgements

This work is supported by International Graduate School of Science and Engineering (IGSSE), Technische Universität München (IGSSE Project 6.08: "4D City"), and the German Research Foundation (DFG, Förderkennzeichen BA2033/3-1). The authors would like to thank Alessandro Parizzi from DLR for valuable discussions and one of the reviewers who pointed out the possible misinterpretation of height dependent APS as seasonal motion.

References

Adam, N., Kampes, B., Eineder, M., Worawattanamateekul, J., Kircher, M., 2003. The development of a scientific permanent scatterer system. In: Proc. ISPRS Hannover Workshop, Hannover, Germany, 6–8 October, pp. 1–6 (on CD-ROM).

Auer, S., Gernhardt, S., Bamler, R., 2011. Ghost persistent scatterers related to multiple signal reflections. IEEE Geoscience and Remote Sensing Letters 8 (5), 919–923.

Bamler, R., Hartl, P., 1998. Synthetic aperture radar interferometry. Inverse Problems 14 (4), R1–R54.

Casu, F., Manzo, M., Berardino, P., Fornaro, G., 2005. The SBAS approach for earth surface deformation analysis. Atti della Fondazione Ronchi, Anno LX, no. 4, pp. 611–618.

Cong, X., Eineder, M., 2012. Volcano deformation measurement using persistent scatterer interferometry with atmospheric delay corrections. In: Proc. EUSAR Conference 2012, Nuremberg, Germany, 23–26 April (on CD-ROM).

Davidson, G.W., Bamler, R., 1999. Multiresolution phase unwrapping for SAR interferometry. IEEE Transactions on Geoscience and Remote Sensing 37 (1), 163–174.

De Zan, F., Rocca, F., 2005. Coherent processing of long series of SAR images. In: Proc. IGARSS 2005, Seoul, Korea, 25–29 July, pp. 1987–1990.

Ferretti, A., Fumagalli, A., Novali, F., Prati, C., Rocca, F., Rucci, A., 2011a. A new algorithm for processing interferometric data-stacks: SqueeSAR. IEEE Transactions on Geoscience and Remote Sensing 49 (9), 3460–3470.

Ferretti, A., Fumagalli, A., Novali, F., Prati, C., Rocca, F., Rucci, A., 2011b. Exploitation of temporary coherent scatterers in SqueeSAR analyses. In: Proc. IGARSS 2011, Vancouver, Canada, 24–29 July (on CD-ROM).

Ferretti, A., Prati, C., Rocca, F., 2001. Permanent scatterers in SAR interferometry. IEEE Transactions on Geoscience and Remote Sensing 39 (1), 8–20.

Ferretti, A., Bianchi, M., Prati, C., Rocca, F., 2005. Higher-order permanent scatterers analysis. EURASIP Journal on Applied Signal Processing 2005 (20), 3231–3242.

Fornaro, G., Reale, D., Serafino, F., 2009. Four-dimensional SAR imaging for height estimation and monitoring of single and double scatterers. IEEE Transactions on Geoscience and Remote Sensing 47 (1), 224–237.

Jakeman, E., Pusey, P.N., 1976. A model for non-Rayleigh sea echo. IEEE Transactions on Antennas and Propagation 24 (6), 806–814.

Jao, J.K., 1984. Amplitude distribution of composite terrain radar clutter and the K-distribution. IEEE Transactions on Antennas and Propagation 32 (10), 1049–1062.

Kampes, B., 2006. Radar Interferometry: Persistent Scatterers Technique, first ed. Springer Netherlands, Dordrecht.

Lombardini, F., 2005. Differential tomography: a new framework for SAR interferometry. IEEE Transactions on Geoscience and Remote Sensing 43 (1), 37–44.

Parizzi, A., Brcic, R., 2010. Adaptive InSAR stacks multi-looking exploiting amplitudes statistics – a comparison between different techniques and practical results. IEEE Geoscience and Remote Sensing Letters 8 (3), 441–445.

Reed, I.S., Mallett, J.D., Brennan, L.E., 1974. Rapid convergence rate in adaptive arrays. IEEE Transactions on Aerospace and Electronic Systems 10 (6), 853–863.

Seymour, M., Cumming, I., 1994. Maximum likelihood estimation for SAR interferometry. In: Proc. IGARSS 1994, Pasadena, USA, 8–12 August, pp. 2272–2275.

Touzi, R., Lopes, A., Bruniquel, J., Vachon, P.W., 1999. Coherence estimation for SAR imagery. IEEE Transactions on Geoscience and Remote Sensing 37 (1), 135–149.

Zebker, H.A., Chen, K., 2005. Accurate estimation of correlation in InSAR observations. IEEE Geoscience and Remote Sensing Letters 2 (2), 124–127.

Zhu, X., Bamler, R., 2010. Very high resolution spaceborne SAR tomography in urban environment. IEEE Transactions on Geoscience and Remote Sensing 48 (12), 4296–4308.

Zhu, X., Bamler, R., 2011. Let's do the time warp: multicomponent nonlinear motion estimation in differential SAR tomography. IEEE Geoscience and Remote Sensing Letters 8 (4), 735–739.

Appendices

E Wang Y, Zhu X (2015c) Robust Estimators in Multipass SAR Interferometry. IEEE Transactions on Geoscience and Remote Sensing, 54 (2): 1 - 13.

http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7265054

Robust Estimators for Multipass SAR Interferometry

Yuanyuan Wang, *Student Member, IEEE*, and Xiao Xiang Zhu, *Senior Member, IEEE*

***Abstract*—This paper introduces a framework for robust parameter estimation in multipass interferometric synthetic aperture radar (InSAR), such as persistent scatterer interferometry, SAR tomography, small baseline subset, and SqueeSAR. These techniques involve estimation of phase history parameters with or without covariance matrix estimation. Typically, their optimal estimators are derived on the assumption of stationary complex Gaussian-distributed observations. However, their statistical robustness has not been addressed with respect to observations with nonergodic and non-Gaussian multivariate distributions. The proposed robust InSAR optimization (RIO) framework answers two fundamental questions in multipass InSAR: 1) how to optimally treat images with a large phase error, e.g., due to unmolded motion phase, uncompensated atmospheric phase, etc.; and 2) how to estimate the covariance matrix of a non-Gaussian complex InSAR multivariate, particularly those with nonstationary phase signals. For the former question, RIO employs a robust M-estimator to effectively downweight these images; and for the latter, we propose a new method, i.e., the *rank M-estimator*, which is robust against non-Gaussian distribution. Furthermore, it can work without the assumption of sample stationarity, which is a topic that has not previously been addressed. We demonstrate the advantages of the proposed framework for data with large phase error and heavily tailed distribution, by comparing it with state-of-the-art estimators for persistent and distributed scatterers. Substantial improvement can be achieved in terms of the variance of estimates. The proposed framework can be easily extended to other multipass InSAR techniques, particularly to those where covariance matrix estimation is vital.**

***Index Terms*—Differential interferometric synthetic aperture radar (D-InSAR), SAR interferometry (InSAR), M-estimator, rank covariance matrix, robust InSAR optimization (RIO), robust estimation.**

I. Introduction

A. State-of-the-Art Multipass InSAR Techniques

MULTIPASS interferometric synthetic aperture radar (InSAR) techniques exploit a time series signal of scatterers. For spaceborne SAR data, they are commonly known as differential InSAR (D-InSAR) as multiple acquisitions are usually acquired by repeat passes. Well-known techniques include persistent scatterer interferometry (PSI) [1]–[3], differential SAR tomography [4]–[6], small baseline subset [7], [8], SqueeSAR [9], [10], CAESAR [11], and so on. They aim at retrieving the parameters of interest, such as the scatterer's 3-D position and deformation rate, by utilizing persistent scatterers (PSs) and/or distributed scatterers (DSs). They are widely employed for ground deformation monitoring, as well as individual building and urban 3-D infrastructure reconstruction. In general, millimeter accuracy of the yearly linear deformation rate and meter accuracy of the scatterer's 3-D position can be achieved with respect to a reference point [1], [12], [13].

However, these accuracy levels refer to the optimal estimators derived based on the assumption of Gaussian-distributed data. PS is modeled as a deterministic signal with additive zero-mean white complex circular Gaussian (CCG) noise [1], whereas DS is modeled as correlated zero-mean CCG [9], [14], [15]. Deviation from the assumption will greatly compromise the performance of the estimator. It boils down to two fundamental problems in the parameter estimation of multipass InSAR.

- Unmodeled phase such as residual atmospheric/orbit phase errors, unmodeled motion phase, etc: Such phase errors impose non-Gaussian noise on PS and DS.
- Non-Gaussian scatterers that bias the covariance matrix estimation of DS: Their spatial stationarity cannot be guaranteed because of the spatially varying phase.

For the first type, there has not been a systematic method to optimally weight images with large phase error. The performance of current techniques depends on the modeling of the interferometric phase, which includes the deformation model, the preprocessing step of atmospheric/orbit error phase mitigation, and so on. This is true for any multipass InSAR processor that targets large area processing. The common deformation models are usually linear and periodic models. The atmospheric phase mitigation methods are mainly based on PS and its variants [1], [16]–[18]. Their accuracy depends on the density of the network, the number of images, and the validity of the assumed deformation model. For meter-resolution data, particularly in the urban areas, the atmospheric phase can be usually well mitigated. Other methods are available based on weather data [19]–[21], which can overcome some disadvantages of the PS-based method, e.g., the topography-correlated atmospheric phase. In [20] and [22], it was concluded that centimeter accuracy can be achieved for controlled locations on the ground. However, achieving centimeter accuracy for a network of dense points on a city scale for InSAR deformation monitoring purposes is not yet feasible at present. Only until recently that such concept and preliminary result have been presented as the *Geodetic TomoSAR* [23].

For the second type, the prevailing approach is to make amplitude-based adaptive sample selections using statistical

Manuscript received April 16, 2015; revised July 20, 2015; accepted August 14, 2015. Date of publication September 14, 2015; date of current version January 19, 2016. This work was supported in part by the Helmholtz Association under the framework of the Young Investigators Group "SiPEO" under Grant VH-NG-1018, by the International Graduate School of Science and Engineering, Technische Universität München under Project 6.08: "4-D City," and by the German Aerospace Center (DLR) under Förderkennzeichen 50EE1417. *(Corresponding author: Xiao Xiang Zhu.)*

Y. Wang was with the International Graduate School of Science and Engineering, 85748 Garching, Germany. He is now with the Signal Processing in Earth Observation, Technische Universität München, 80333 Munich, Germany (e-mail: wang@bv.tum.de).

X. X. Zhu is with The Remote Sensing Technology Institute, German Aerospace Center (DLR), 82234 Weßling, Germany, and also with the Signal Processing in Earth Observation, Technische Universität München, 80333 Munich, Germany (e-mail: xiao.zhu@dlr.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2015.2471303

tests such as Kolmogorov–Smirnov (KS) or Anderson–Darling tests [8]–[10], [24]. However, estimation results with these methods degrade as the number of images decrease. According to [25], assuming a Rayleigh-distributed amplitude time series of *ten* observations, the detection rate is only 20% (at a constant 5% false alarm rate) for distinguishing two DSs with an expected intensity ratio of 3 dB using the KS test. As detection rate decreases, pixels of different distributions, including those non-Gaussian, are included in the covariance matrix estimation. This in turn affects the final parameter estimation.

B. RIO Framework

The aim of this paper is to introduce robust estimation for multipass InSAR techniques when the underlying statistical model of the observations is not fulfilled, which was briefly mentioned in [26] and [27]. Let $f_{\mathbf{g}}(\mathbf{g}|\boldsymbol{\theta}, F(\hat{\mathbf{C}}))$ be a generic likelihood function of the observation time series $\mathbf{g}$, the parameter vector $\boldsymbol{\theta}$ to be retrieved, and the system model $F(\cdot)$ (with the estimated covariance matrix $\hat{\mathbf{C}}$, if used). The maximum-likelihood estimator (MLE) is $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} f_{\mathbf{g}}(\mathbf{g}|\boldsymbol{\theta}, F(\hat{\mathbf{C}}))$. Assuming a CCG-distributed $\mathbf{g}$, the MLE can be simplified to the minimization of the sum of the squared residuals, i.e.,

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \ln f_{\mathbf{g}}\left(\mathbf{g}|\boldsymbol{\theta}, F(\hat{\mathbf{C}})\right) = \arg\min_{\boldsymbol{\theta}} \|\boldsymbol{\varepsilon}\|_2^2 \quad (1)$$

where $\boldsymbol{\varepsilon}$ is the residual vector, and $\|\cdot\|_2$ denotes the Euclidean norm.

The proposed robust method introduces the following aspects into the current techniques.

- It replaces the MLE with an M-estimator [28], which minimizes the sum of a customized function $\rho(x)$ of the residuals, i.e.,

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_i \rho(\varepsilon_i). \quad (2)$$

- Should DS be used, the covariance matrix estimator is also replaced with its robustified version. Depending on the stationarity of the samples' phases, different estimators are proposed for:
- Stationary samples: M-estimator of covariance that is an amplitude-based weighting of the sample covariance matrix;
- Nonstationary samples: *rank M-estimator* (RME) that is an M-estimator using the *rank* of the complex observations $\mathbf{g}$.

Table I summarizes a comparison of the current state-of-the-art approaches with the improvements proposed by the RIO framework.

C. Readers' Guide

In this paper, Section II covers the necessary basics of the M-estimator and the multivariate rank. Sections III and IV introduce and provide detailed information about the robust estimators used in InSAR techniques, followed by experiments using simulated and real data in Section V.

Notations of some frequently used variables and their dimensions are listed in the following. Italic small letters denote scalar value, bold small letters are vectors, and bold capital letters denote a matrix.

TABLE I
COMPARISON OF RIO FRAMEWORK WITH THE CURRENT APPROACH

		State-of-the-art	RIO
		Covariance estimation	
DS	Stationary	Sample covariance matrix: $\hat{\mathbf{C}}_{MLE}$	M-estimator: $\hat{\mathbf{C}}_{M\text{-}est}$
DS	Nonstationary		Rank M-estimator: $\hat{\mathbf{C}}_{RME}$
		Phase-history parameters estimation	
PS		Periodogram: $\hat{\boldsymbol{\theta}}_{MLE}$	M-estimator: $\hat{\boldsymbol{\theta}}_{M\text{-}est}$
DS	Stationary	Periodogram on samples mean: $\hat{\boldsymbol{\theta}}_{MLE}$	M-estimator on samples mean: $\hat{\boldsymbol{\theta}}_{M\text{-}est}\left(\hat{\mathbf{C}}_{M\text{-}est}\right)$
DS	Nonstationary	–	M-estimator pixel-wise inversion: $\hat{\boldsymbol{\theta}}_{M\text{-}est}\left(\hat{\mathbf{C}}_{RME}\right)$

$N \in \mathbb{N}$	number of observations;
$M \in \mathbb{N}$	number of samples;
$\mathbf{g} \in \mathbb{C}^N$	complex InSAR observations;
$\mathbf{r} \in \mathbb{C}^N$	rank of the observations;
$\boldsymbol{\theta} \in \mathbb{R}^L$	L parameters of the phase model;
$\mathbf{C} \in \mathbb{C}^{N\times N}$	complex covariance matrix;
$\boldsymbol{\Phi} \in \mathbb{C}^{N\times N}$	diagonal matrix of the modeled phase;
$\mathbf{W} \in \mathbb{R}^{N\times N}$	diagonal weighting matrix;
$\boldsymbol{\varepsilon} \in \mathbb{C}^N$	residuals;
$\rho(x)$	loss function;
$\lvert\cdot\rvert$	on scalar: absolute value; on vector and matrix: element-wise absolute value
$\|\cdot\|$	vector norm;
$\det(\cdot)$	matrix determinant.

II. BACKGROUND INFORMATION

A. M-Estimator Basics

M-estimators are a class of well-known robust estimators that take the form of a minimization problem: $\min \sum_i \rho(x_i)$. The M-estimator permits a minimization of a customized *loss function* $\rho(x)$ of the residuals to resist outliers without preprocessing the data, e.g., by outlier trimming.

The derivation of the M-estimator can be described as follows. Consider a simple location estimation problem, where the likelihood function of the univariate observation of the location $g \in \mathbb{R}$ is $f_g(g|\mu)$ in which μ is the expected location to be estimated. $f_g(g|\mu)$ equals a probability density function (pdf) $f_x(x)$, where $x = g - \mu$. Without losing generality, let us assume that x is scaled to unit variance. Maximizing the sum of log-likelihood is the MLE of μ, given multiple observations of g, i.e.,

$$\hat{\mu}_{\mathrm{MLE}} = \arg\min \sum_i -\ln f(x_i). \quad (3)$$

The M-estimator of μ is as follows:

$$\hat{\mu}_{\mathrm{M-est}} = \arg\min \sum_i \rho(x_i). \tag{4}$$

Therefore, choosing $\rho(x) = -\ln f_x(x)$ yields the ordinary MLE. For example, the MLE, when assuming a normal distribution, corresponds to an M-estimator with $\rho(x) = x^2$.

Solving (4) is equivalent to finding the zero position of its derivative, which can be written as

$$\sum_i \frac{\rho'(x_i)}{x_i} x_i = 0 \tag{5}$$

where $\rho'(x)$ is the derivative of $\rho(x)$. $\rho'(x)/x$ is commonly referred to as the weighting function $w(x)$ because substituting $x_i = g_i - \mu$ into the previous equation yields an M-estimator that is actually a weighted average, i.e.,

$$\hat{\mu}_{\mathrm{M-est}} = \frac{\sum_i w(g_i - \hat{\mu}) g_i}{\sum_i w(g_i - \hat{\mu})}. \tag{6}$$

The weighting function of a given pdf $f_x(x)$ can be derived with the following relationship:

$$w(x) = \frac{-\partial \ln f_x(x)}{x \partial x}. \tag{7}$$

Equation (6) is solved iteratively with an initial μ estimate being given and the weights being updated at each iteration. Typically, the weighting function is designed to have high efficiency at normal distribution. Therefore, if x is not scaled to unit variance, the weights are $w = \rho'(x/\sigma)/(x/\sigma)$, where σ is the standard deviation of x. Convergence is only guaranteed for convex ρ functions [29]. The initial estimate must be carefully estimated should ρ be nonconvex.

The M-estimator is a tradeoff between robustness and efficiency, where the efficiency of an estimator refers to the closeness of the variance of its estimates to the possible lower bound, e.g., the Cramér–Rao lower bound for an unbiased estimator [30]. Nevertheless, it can still maintain high efficiency under the nominal distribution by properly choosing the loss function.

B. Multivariate Rank

Another important tool in robust estimation is the *rank* of random variables. Using the ranks of the samples instead of the samples themselves reveals their robustness when assessing the statistical dependence between two random variables, and more importantly, it is invariance to the nonlinear relation between the two random variables. For example, the Spearman's rank correlation coefficient is the ordinary correlation coefficient of the samples' ranks. It is robust against nonstationary samples, i.e., the nonlinear transformation between two random variables [31].

In univariate real random variables, the rank simply refers to the integer ranking of each realization of this variable according to its value. In [32], the centered rank of the sample x_i in the univariate data set $\{x_i, x_2, \ldots, x_M\}$ is defined as $\hat{r}_i = (1/M) \sum_{j=1}^{M} \mathrm{sign}(x_i - x_j)$, where $\mathrm{sign}(x) = -1, 0, 1$ for $x < 0$, $x = 0$, and $x > 0$, respectively. This is simply a ranking algorithm that can be easily extended to real multivariate case, i.e.,

$$\hat{\mathbf{r}}_i = \frac{1}{M} \sum_{j=1}^{M} \mathrm{sign}(\mathbf{x}_i - \mathbf{x}_j) = \frac{1}{M} \sum_{j=1}^{M} \frac{\mathbf{x}_i - \mathbf{x}_j}{\|\mathbf{x}_i - \mathbf{x}_j\|} \tag{8}$$

for the sample set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\}$, where $\mathbf{x}_i \in \mathbb{R}^N$ is a vector of observations, such as a time series of pixels in multipass InSAR techniques, despite the fact that it is not a complex number in this case.

III. Robust Covariance Estimation

The covariance estimate of a target pixel requires a selected neighborhood of samples. The widely used sample covariance matrix is only suitable for samples following a stationary CCG distribution. Otherwise, a robust estimator is necessary. In the following, we demonstrate the robust covariance estimators for both stationary and nonstationary samples.

A. Sample Covariance Matrix

If a complex SAR pixel value g obeys the classic model of a zero-mean CCG distribution [14], then multiple acquisitions $\mathbf{g} \in \mathbb{C}^N$ at the location of the same pixel lead to a multivariate CCG distribution, i.e.,

$$f_{\mathbf{g}}(\mathbf{g}) = \frac{1}{\pi^N \det(\mathbf{C})} \exp(-\mathbf{g}^H \mathbf{C}^{-1} \mathbf{g}) \tag{9}$$

which is fully characterized by its covariance matrix $\mathbf{C}$, where N is the number of acquisitions. Given a set of M spatially stationary samples $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_M]$, the MLE of $\mathbf{C}$ is one that maximizes the likelihood, i.e.,

$$\hat{\mathbf{C}}_{\mathrm{MLE}} = \arg\min_{\mathbf{C}} \sum_{m=1}^{M} \mathbf{g}_m^H \mathbf{C}^{-1} \mathbf{g}_m. \tag{10}$$

This yields the well-known equation for the sample covariance matrix:

$$\hat{\mathbf{C}}_{\mathrm{MLE}} = \frac{1}{M} \sum_{m=1}^{M} \mathbf{g}_m \mathbf{g}_m^H. \tag{11}$$

$\hat{\mathbf{C}}_{\mathrm{MLE}}$ is not robust against any type of outlier. A single outlier in the samples will considerably bias the estimates. Hereafter, we designate the MLE as the MLE for CCG, unless explicitly stated otherwise.

B. Robust Estimation for Stationary Non-Gaussian Samples

If the selected samples are non-Gaussian, caused by either a non-Gaussian scattering mechanism or a poor sample selection, the covariance estimation can be robustified using an M-estimator. Here, we assume the samples' stationarity.

The M-estimator of a covariance matrix is expressed analogously with (4) by its extension to multivariate variables, i.e.,

$$\hat{\mathbf{C}}_{\mathrm{M-est}} = \arg\min_{\mathbf{C}} \sum_{m=1}^{M} \rho\left(x_m(\mathbf{C})\right) \tag{12}$$

where the squared residual $x_m^2(\mathbf{C}) \in \mathbb{R}$ of the multivariate $\mathbf{g}_m$ is $\mathbf{g}_m^H \mathbf{C}^{-1} \mathbf{g}_m$. It can be solved using an iteratively reweighted approach [33], [34], i.e.,

$$\hat{\mathbf{C}}_{k+1} = \frac{1}{M} \sum_{m=1}^{M} w\left(x_m(\hat{\mathbf{C}}_k)\right) \mathbf{g}_m \mathbf{g}_m^H. \tag{13}$$

The real-valued weighting function $w(x)$ is $\rho'(x)/x$, and k denotes the iteration index. The weighting function downweights highly deviating samples whose whitened version $x_m(\hat{\mathbf{C}}_k) = \|\hat{\mathbf{C}}_k^{-1/2} \mathbf{g}_m\|$ is large in magnitude, which greatly depends on the intensities of $\mathbf{g}_m$.

To derive a meaningful weighting function for some heavily tailed distributions, we model the non-Gaussian SAR pixel distribution as a complex t-distribution [35], [36], i.e.,

$$f_{\mathbf{g}}(\mathbf{g}) = \frac{\Gamma(\nu + N)}{(\pi\nu)^N \Gamma(\nu) \det(\mathbf{C})^N} \left(1 + \frac{1}{\nu} \mathbf{g}^H \mathbf{C}^{-1} \mathbf{g}\right)^{-(\nu+N)} \tag{14}$$

where $\Gamma(\cdot)$ is the gamma function, and $\nu \in \mathbb{R}^+$ is the degree of freedom (DoF). The complex t-distribution is a realistic model for SAR observations as the DoF allows a quantitative definition of its deviation from CCG. This distribution approaches CCG as the DoF approaches $+\infty$ and becomes more heavily tailed as it decreases toward zero. In practical terms, this provides flexibility for handling different levels of outliers.

The weighting function can be derived according to (7). However, (14) must be transformed into its equivalent real-valued version [33] to perform the derivation. Therefore, we form a new multivariate $\tilde{\mathbf{g}} = [\mathrm{Re}(\mathbf{g}); \mathrm{Im}(\mathbf{g})]$ ($\mathrm{Re}[\cdot]$ and $\mathrm{Im}[\cdot]$ are the real and imaginary parts of a complex number; semicolon denotes vertical concatenation), which follows a $2N$-variate real-valued t-distribution [33]. The corresponding weighting function can be derived thereafter as follows (see also Appendix A):

$$w(x) = \frac{2N + \nu}{\nu + 2x^2}. \tag{15}$$

For good overall robustness, a small $\nu(< 5)$ is preferred [37]. For example, in [38] ν is chosen from 2.5 to 3.9 for centimeter-resolution millimeter-wave SAR data of urban areas. The particular complex t-distribution considering $\nu = 1$ is also known as the complex Cauchy distribution, which has been thoroughly investigated in many studies.

To best safeguard against unknown heavily tailed distributions, ν can be set to zero. This leads to a weighting function of $w(x) = N/(\mathbf{g}_m^H \mathbf{C}^{-1} \mathbf{g}_m)$. Further, by assuming $\mathbf{C}$ equals $\bar{I}\ \mathbf{I}$, where $\bar{I}$ is the expected intensity, and $\mathbf{I}$ is the identity matrix, the iterative process can be dropped. This leads to the interesting weighting function $w = \bar{I}\|\mathbf{g}_m\|^{-2}$. The corresponding covariance matrix estimate

$$\hat{\mathbf{C}}_{\mathrm{SCM}} = \frac{N\bar{I}}{M} \sum_{m=1}^{M} \|\mathbf{g}_m\|^{-2} \mathbf{g}_m \mathbf{g}_m^H \tag{16}$$

is also known as the *sign covariance matrix* (SCM) [32], [39], where only the direction of each multivariate sample is considered. The real covariance is lost as a result. Should the exact covariance be retrieved, one could always estimate it from the eigenvectors of the SCM and the samples. SCM does not require iteration and is robust against outliers.

SCM is an engineering solution for fast processing under the general M-estimator's framework. There also exist other methods such as amplitude clipping and dropping, etc. The choice among the methods depends on the data and application. For DSs with low coherence, simulation shows SCM outperforms amplitude clipping and dropping both in terms of variance and bias.

C. Robust Estimation for Nonstationary Non-Gaussian Samples

To complete the theory, we must consider the situation in which samples are neither stationary nor complex Gaussian. The most common nonstationarity found in an SAR interferogram is the uncompensated rapidly varying interferometric phase among the spatial samples. For nonstationary samples, a joint estimation of the expected interferometric phase and the covariance [15] is required, i.e.,

$$\hat{\mathbf{C}}_{\mathrm{MLE}} = \frac{1}{M} \sum_{m=1}^{M} \hat{\mathbf{\Phi}}_m^H \mathbf{g}_m \mathbf{g}_m^H \hat{\mathbf{\Phi}}_m \tag{17}$$

where $\hat{\mathbf{\Phi}}_m$ is the complex diagonal matrix containing the estimated phase values of $\mathbf{g}_m$.

Usually, the estimation of interferometric phase and the covariance matrix is separated. The former is usually achieved by common band filtering by either spatial averaging or spectral trimming. Its performance greatly depends on the filter itself; additionally, estimations can be challenging in urban areas. In [10], a multiresolution defringe algorithm to manage this challenge is used. In any case, the phase estimation requires additional effort, and poor estimates significantly degrade the covariance matrix estimation.

This calls for an estimator that is invariant for multiplicative phase signals. As demonstrated in [32] and [39], the rank covariance matrix for real multivariate is invariant with regard to the additive deterministic signal. Inspired by this fact, we introduce a new estimator, i.e., RME, for complex multivariate with a multiplicative phase signal. By replacing the multivariate $\mathbf{g}$ with its rank $\mathbf{r}$ in (13), we obtain the following:

$$\hat{\mathbf{C}}_{\mathrm{RME},k+1} = \frac{1}{M} \sum_{m=1}^{M} w\left(x_m(\hat{\mathbf{C}}_{\mathrm{RME},k})\right) \hat{\mathbf{r}}_m \hat{\mathbf{r}}_m^H. \tag{18}$$

We define the complex rank vector, being analogous to (8), as follows:

$$\hat{\mathbf{r}}_m = \frac{1}{J} \sum_{j=1}^{J} \frac{\mathbf{g}_m \cdot \mathbf{g}_j^*}{\left\|\mathbf{g}_m \cdot \mathbf{g}_j^*\right\|} \tag{19}$$

where $\mathbf{g}_j$ is the direct neighborhood of $\mathbf{g}_m$, and $\cdot$ denotes the element-wise multiplication. Note its difference from the original definition, where subtraction is replaced by multiplication of the complex conjugate. The deterministic phase has been reduced through this operation, and it approaches zero through

averaging. To maintain effective phase cancelation, the direct neighborhood should be kept small. We usually restrict it to only four adjacent pixels. The RME is a fourth-order descriptor of the sample statistics. It can be proven (Appendix B) that the elementwise square root of $|\hat{\mathbf{C}}_{\mathrm{RME}}|$ approaches $|\hat{\mathbf{C}}_{\mathrm{MLE}}|$ asymptotically under a CCG distribution when using a one-neighborhood sample. The variance of the square root $|\hat{\mathbf{C}}_{\mathrm{RME}}|$ is also higher than that of $|\hat{\mathbf{C}}_{\mathrm{MLE}}|$ because of the multiplication of two uncorrelated complex-valued random variables. Therefore, the number of samples should be doubled when using RME. Hereafter, we assume that the elementwise square root on $|\hat{\mathbf{C}}_{\mathrm{RME}}|$ has been performed.

IV. Retrieval of Robust Phase History Parameters

A. PS Phase History Parameters

One of the commonly used estimators for PS phase history parameters $\boldsymbol{\theta}$ is the periodogram [40] (often with amplitude-dropped measurements [1], [3]), i.e.,

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{ \left| \frac{1}{N} \sum_{n=1}^{N} g_n \exp\left(-j\varphi_n(\boldsymbol{\theta})\right) \right| \right\} \tag{20}$$

where g_n and φ_n are the complex pixel value and the modeled phase of the nth image, respectively. Usually, the amplitude of g_n is dropped in the estimation, as it barely changes estimates at high SNR. Equation (20) is the MLE when assuming additive independent and identically distributed (i.i.d.) CCG noise [41], i.e.,

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{ (2\pi\sigma^2)^{-\mathrm{N}} \exp\left(-\frac{1}{2\sigma^2} \left(\mathbf{g} - \overline{\mathbf{g}}(\boldsymbol{\theta})\right)^H \left(\mathbf{g} - \overline{\mathbf{g}}(\boldsymbol{\theta})\right) \right) \right\} \tag{21}$$

where σ and $\bar{\mathbf{g}}(\boldsymbol{\theta})$ are the standard deviation of the CCG distribution and the modeled PS pixel values, respectively.

Based on the aforementioned MLE, we propose the following robust estimator of PS phase history parameters to deal with possible large phase errors, particularly the residual atmospheric phase and unmodeled motion phase, i.e.,

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \rho\left(\frac{\mathrm{Re}\left[\varepsilon_i(\boldsymbol{\theta})\right]}{\sigma_R} \right) + \rho\left(\frac{\mathrm{Im}\left[\varepsilon_i(\boldsymbol{\theta})\right]}{\sigma_I} \right) \tag{22}$$

where the residual $\varepsilon_i(\boldsymbol{\theta})$ is $g_i - \bar{g}_i(\boldsymbol{\theta})$; and σ_R and σ_I are the standard deviations of real and imaginary parts, respectively. The most appropriate $\rho(x)$ function should be chosen according to the pdf of the PS noise, which depends on the performance of the atmospheric phase correction, the choice of motion model, etc. If such information is not available, a good alternative is Tukey's biweight loss function, i.e.,

$$\rho(x) = \begin{cases} -\frac{(c^2 - x^2)^3}{6c^4} + \frac{c^2}{6} & |x| < c \\ \frac{c^2}{6} & \text{elsewhere} \end{cases} \tag{23}$$

where $c \in \mathbb{R}^+$ is a tuning parameter beyond which the loss is constant. The lower the value of c, the more robust but less efficient is the estimator, and *vice versa*. c is usually chosen to be 4.586, which is said to give an efficiency of 95% at normal distribution [29]. For an overview of different robust loss functions, readers are referred to [29] and [42].

Equation (22) should be solved iteratively, where σ_R and σ_I are updated at each iteration. Their initial values are obtained from the initial estimate $\hat{\boldsymbol{\theta}}_0$; this is not critical for convex $\rho(x)$. However, a robust initial estimation is strongly recommended for nonconvex $\rho(x)$, such as the Tukey's loss function. Fortunately, the initial estimation need not be efficient. To give one example, the least trimmed square that minimizes the sum of q smallest squared residuals [43] is as follows:

$$\hat{\boldsymbol{\theta}}_0 = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{q} \left|\varepsilon_i(\boldsymbol{\theta})\right|^2. \tag{24}$$

The initial estimates of standard deviation, as well as those at each iteration, can be obtained using a robust estimator such as the median absolute deviation, i.e.,

$$\hat{\sigma} = 1.483 \cdot \mathrm{median}\left(\left|\boldsymbol{\varepsilon} - \mathrm{median}(\boldsymbol{\varepsilon})\right|\right). \tag{25}$$

The factor 1.483 is a normalization constant, which gives an expectation of unity for a normal distribution. With this combination, convergence can be achieved in just a few iterations; therefore, the computational cost is approximately three to four times more than that for the standard estimator. The same procedures can be easily extended to the LAMBDA estimator [44] for PS phase parameter estimations, which has a substantial number of users.

B. DS Phase History Parameters

Because of the same artifacts caused by the phase calibration mentioned earlier, the estimator of DS phase history parameters must also be robustified. If stationarity is assumed for a DS and its neighborhood, one can treat DSs as a single PS by averaging them, as proposed in SqueeSAR, and then the robustified estimator is identical to (22).

However, we do not assume stationarity in our considerations. Our objective is a full inversion of individual single-look DS observation (DS pixels without averaging). The DS MLE introduced in [15] is recalled here as follows:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{ \mathbf{g}^H \boldsymbol{\Phi}(\boldsymbol{\theta}) |\hat{\mathbf{C}}|^{-1} \boldsymbol{\Phi}(\boldsymbol{\theta})^H \mathbf{g} \right\} \tag{26}$$

where $\boldsymbol{\Phi}(\boldsymbol{\theta}) \in \mathbb{C}^{N \times N}$ is the diagonal matrix containing the modeled phase of $\mathbf{g}$. Seemingly, the robustified estimator for DS phase history parameters takes an expression similar to (22), but with the residuals $\boldsymbol{\varepsilon}(\boldsymbol{\theta})$ whitened by a robust covariance matrix estimate, e.g., $\hat{\mathbf{C}}_{\mathrm{RME}}$, i.e.,

$$\boldsymbol{\varepsilon}(\boldsymbol{\theta}) = |\hat{\mathbf{C}}_{\mathrm{RME}}|^{-0.5} \boldsymbol{\Phi}(\boldsymbol{\theta})^H \mathbf{g}. \tag{27}$$

The robustified DS estimator can be written into an iteratively reweighted expression with k being the iteration index, i.e.,

$$\hat{\boldsymbol{\theta}}_{k+1} = \arg\min_{\boldsymbol{\theta}} \left\{ \boldsymbol{\varepsilon}^H(\boldsymbol{\theta}) \mathbf{W}_k \boldsymbol{\varepsilon}(\boldsymbol{\theta}) \right\} \tag{28}$$

TABLE II
PSEUDOCODE OF THE ROBUST ESTIMATOR FOR SINGLE-PIXEL DS OBSERVATIONS

Get initial estimate for $\hat{\boldsymbol{\theta}}$ which can be the same for all DS samples in the same neighborhood

Calculate initial $\mathbf{W}_0$

for iteration $k = 1$ until convergence

for DS sample $m = 1$ to M

Estimate $\hat{\boldsymbol{\theta}}^m$ by solving (28) using $\mathbf{W}_k$

end

Compute the robust mean residuals $\bar{\boldsymbol{\varepsilon}}$ of the neighborhood following (29).

Update $\mathbf{W}_k$ based on the mean residual

end

where $\mathbf{W} \in \mathbb{R}^{N\times N}$ is the diagonal weighting matrix whose element is calculated using each element of $\boldsymbol{\varepsilon}(\boldsymbol{\theta})$ based on $w_i = \rho'(\varepsilon_i)/\varepsilon_i$.

However, we should be aware that the marginal distribution of the phase of zero-mean CCG multivariate is uniform as it is in its whitened version. The atmospheric phase, presumably also uniformly distributed over time, applies no change to the DS observations statistically. In other words, the robust loss function is blind to phase contamination from single-look multivariate DS observations. Therefore, correct weighting of the contaminated observations must be introduced to the estimator. It should be determined by the expected residuals $\bar{\boldsymbol{\varepsilon}}$ of the neighborhood ensemble. The expected residuals can be calculated by simple averaging. However, because of possible outliers in the neighborhood, it should be estimated robustly as follows:

$$\overline{\boldsymbol{\varepsilon}} = \sum_{m=1}^{M} w^m \boldsymbol{\varepsilon}^m(\hat{\boldsymbol{\theta}}^m) \Big/ \sum_{m=1}^{M} w^m \tag{29}$$

where the superscript m denotes the sample number in the neighborhood, and w^m is a robust weight.

To summarize, (28) is a joint estimation of the phase parameters of individual single-look DS observations in a neighborhood. Its computation should begin with initial estimates of each sample in the neighborhood (assumed to be the same), which jointly determine the initial weighting matrix. This matrix is used to retrieve the parameters of each single-look DS observation time series in the neighborhood and is updated based on all the estimates upon finishing one iteration. The pseudocode of the proposed DS estimator is as stated in Table II.

One practical issue when dealing with real data is the positiveness of the covariance matrix. The inverse square root operation requires that the matrix be positive definite, i.e., for any nonzero vector $\mathbf{y} \in \mathbb{R}^N$ that $\mathbf{y}^T|\hat{\mathbf{C}}|\mathbf{y} > 0$. This criterion is not fulfilled when the rank of the selected samples is less than N, i.e., when they are spatially correlated, which results in some eigenvalues of $|\hat{\mathbf{C}}|$ being numerically close to zero. On rare occasions, the covariance matrix must be regularized. In our implementation, this is performed by adaptively adding an appropriately small quantity to its diagonal. The pseudocode is as stated in Table III.

TABLE III
PSEUDOCODE OF THE COVARIANCE MATRIX REGULARIZATION

$\mathbf{I}$ = identity matrix

$e \in \mathbb{R}^+$ is a small number, e.g., 1e-6

while $\left|\hat{\mathbf{C}}_{RME}\right|$ is NOT positive definite

$\left|\hat{\mathbf{C}}_{RME}\right| = \left|\hat{\mathbf{C}}_{RME}\right| + e\mathbf{I}$

$e = e \times 2$

end

V. ROBUSTNESS ANALYSIS

A. Comparison of Robust Covariance Estimators

We compared the original MLE, the M-estimator with t-distribution weighting, and the RME under three different scenarios: 1) multivariate CCG; 2) multivariate complex t-distribution with one DoF; and 3) nonstationary multivariate t-distribution with one DoF. The data are simulated to have ten acquisitions, with a neighborhood of 1000 samples. Each sample vector is simulated by dewhitening a ten-variate vector of zero-mean i.i.d. CCG distribution (for scenario 1) or complex t-distribution (for scenarios 2 and 3) with a predefined coherence matrix. In the last scenario, we added linear fringes with ten different fringe frequencies randomly picked within $[0\ \pi/10]$ to the phases of the ten acquisitions.

We also predefined two different coherence matrices: one exponentially decaying and the other with a constant 0.5 coherence between acquisitions. The results are shown in Fig. 1, where the left subfigure corresponds to the exponential coherence matrix, and the right corresponds to the constant matrix. Each row represents one of the aforementioned three scenarios, and each column represents one of the three covariance estimators. The subplots (1, 1) in the two subfigures are regarded as the reference coherence matrices because the MLE is the optimal estimator under CCG and is asymptotically unbiased.

All three estimators can preserve the correct shape of the covariance matrix under CCG. The MLE fails in the second scenario, where the samples are contaminated by outliers. The coherence is usually overestimated because of the large amplitude of the outliers. The M-estimator and RME remain as correct estimations. In particular, this M-estimator is the MLE for the simulated data in the second scenario. Finally, the M-estimator is not capable of dealing with nonstationary samples. Heavy underestimation occurs because of the summation of the complex numbers with nonstationary phases. The RME is mean invariant and maintains good performance in all conditions. However, $\hat{\mathbf{C}}_{\mathrm{RME}}$ experiences higher bias at low coherence [as observed in (a)] and higher variance [as observed in (b)], due to the doubling of the phase variance of the rank observation. Doubling the number of looks should be used in the estimation of $\hat{\mathbf{C}}_{\mathrm{RME}}$. This has been mentioned in Section III-C. The

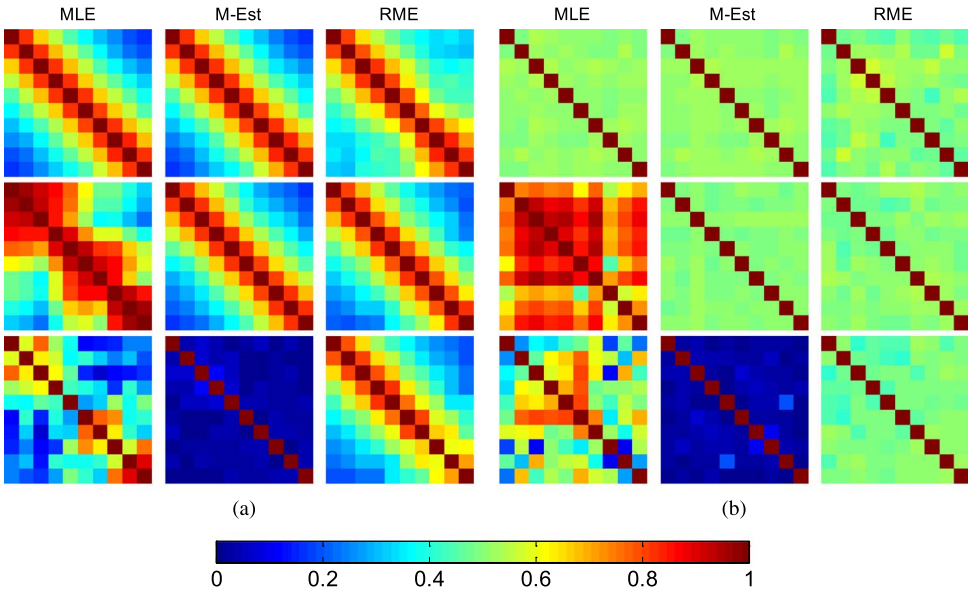


Fig. 1. Comparison of three covariance matrix estimators under three different observation cases. First row: CCG. Second row: Complex t-distribution with one DoF. Third row: Nonstationary complex t-distribution with one DoF. (a) and (b) show the results of the exponential and constant coherence matrices, respectively. In both (a) and (b), first column: MLE (under Gaussian); second column: M-estimator with t-distribution weighting; and third column: RME with t-distribution weighting.

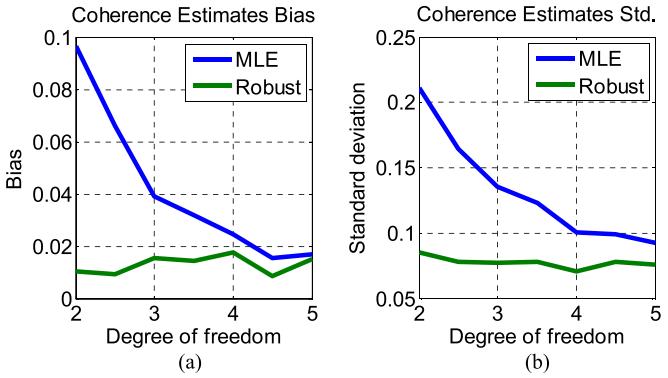


Fig. 2. (a) Bias and (b) standard deviation of the coherence estimates of MLE and the proposed robust estimator for true coherence $= 0.2$. One hundred looks are used in the coherence estimation.

efficiency of the M-estimator and that of RME are always lower than 100% at CCG distribution. This illustrates the tradeoff between robustness and efficiency.

To quantitatively analyze the performance of the robust estimator at different outlier levels, we have performed experiments using less tailed distribution by changing the DoF in the complex t-distribution. The complex t-distribution approaches complex Gaussian as the DoF increases. Fig. 2 plots the variance and bias of the robust estimator and MLE as a function of the DoF for true coherence $= 0.2$. As one can see, the robust estimator outperforms the Gaussian MLE. However, as we expected, the advantage of robust estimator becomes less as the DoF increases because the distribution becomes more Gaussian.

In practice, the number of samples should be sufficiently large to prevent any significant bias in the covariance estimates and to reduce the variance of the estimates. We found that 200 samples are more than sufficient for a stack of 30 images based on the Reed–Mallett–Brennan detection rule [45], which measures the loss of the signal's log-likelihood. When using RME, the number of samples should be doubled to achieve a comparable variance with that of the MLE.

B. *Efficiencies of Robust PS and DS Estimators*

Next, we used a Monte Carlo simulation to quantitatively compare the efficiencies of the proposed PS and DS estimators with their original MLEs. We set the common parameters for the simulation to be similar to those of the TerraSAR-X data, where the wavelength $\lambda = 0.031$ m, the range distance $R = 700$ km, the acquisitions number $N = 20$, and the spatial baselines $\mathbf{b}_\perp$ were randomly sampled from -100 m to $+100$ m with uniform distribution, and the temporal baselines $\mathbf{t}$ were evenly spaced from -1 year to $+1$ year.

The PS observation vector is simulated by adding zero-mean i.i.d. CCG noise with a certain SNR to the deterministic PS signal $\mathbf{g} = A\exp(-j4\pi/\lambda(s\mathbf{b}_\perp/R + v\mathbf{t}))$, where A is a complex amplitude. The same SNR was maintained for each acquisition. A DS vector was simulated using the same procedure described earlier, i.e., by dewhitening a multivariate zero-mean i.i.d. CCG using a predefined coherence matrix. The coherence was maintained constant between any two acquisitions. To complete the DS simulation, we added the interferometric phase $-j4\pi/\lambda(s\mathbf{b}_\perp/R + v\mathbf{t})$ to the phases of the DS vector. We generated 1000 samples of this DS vector for the covariance matrix estimation. These represent the contamination-free PS and DS observations.

We then introduced contamination in two ways. First, we added a random phase to eight (40%) of the 20 acquisitions. The phase was uniformly distributed between $[-\pi, \pi]$ for both PS and DS. For DS, we added the same phase to the samples of the same acquisition to simulate the temporally random but

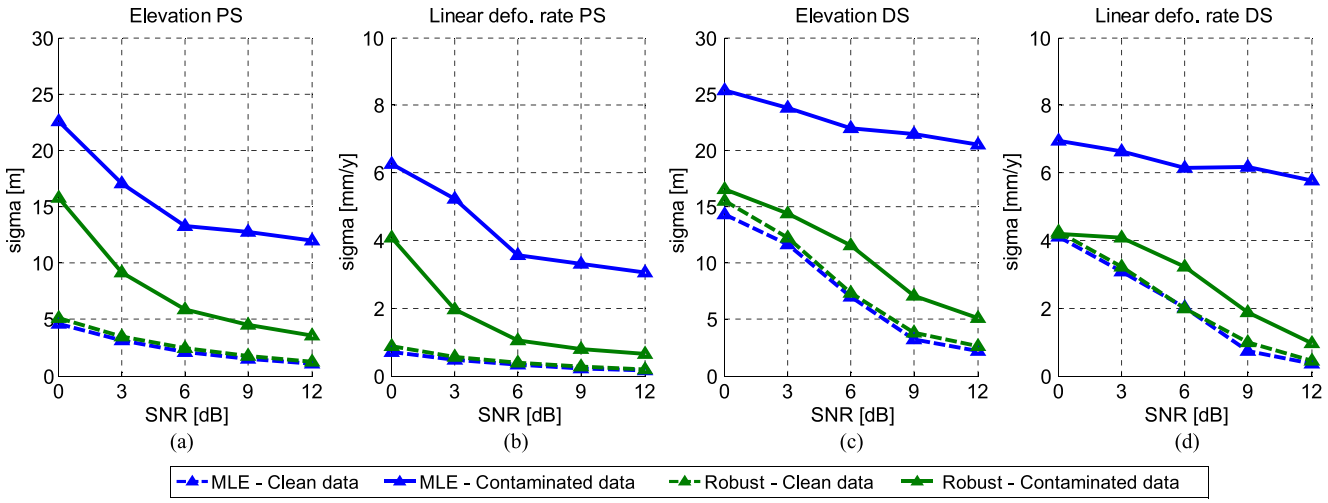


Fig. 3. Comparison of estimates of the standard deviation of elevation and linear deformation rate by the robust estimators and the ordinary MLEs. (a) and (b) refer to the results of the PS estimators, and (c) and (d) refer to the DS estimators. In each subfigure, the green curves are the results of the proposed robust estimators, and the blue curves are those of the original MLE; the dashed curves are the estimators performing on observations with a CCG distribution, and the solid curves are the results for contaminated data, which we generated by adding a uniformly distributed random phase to eight of the 20 acquisitions. The blue dashed curves in all the subfigures are the MLEs under a CCG distribution; hence, they are the optimal estimates.

spatially stationary atmospheric phase. Second, we used an i.i.d. complex t-distribution with one DoF in the DS simulation to simulate outliers in the DS samples.

Both contamination-free and contaminated observations were used in the proposed robust estimators and their original counterparts. We used the Tukey's loss function as the robust kernel for both robust estimators of PS and DS. Parameter c was set to 4.586, and we repeated the Monte Carlo simulation 1000 times. Fig. 3 shows the standard deviation of the estimates of elevation and linear deformation rate with respect to different SNRs. Fig. 3(a) and (b) correspond to the PS estimators, and Fig. 3(c) and (d) correspond to the DS estimators. The SNR of the DS refers to the correlated signal w.r.t. the decorrelated part and can be directly related to the coherence using the formula $\gamma = (1 + \mathrm{SNR}^{-1})^{-1}$ [14], [46]. In each subfigure, the green curves are the results of the proposed robust estimators, and the blue curves are the original MLEs; the dashed curves are the estimators performing on observations with CCG distribution noise, and the solid curves are contaminated data results. The blue dashed curves in all the subfigures are the MLEs under CCG; hence, they are the optimal estimates. All the simulation results suggest that the robust estimators significantly outperform the original MLEs when up to 40% of the observations are contaminated. The robust estimators improve the efficiency by a factor of 7–35 for contaminated data (the squared ratio of the solid curves). Still, the efficiency of the robust estimator at the nominal distribution is close to that of the original MLE. The *relative efficiency* of the robust estimator and the MLE, defined as

$$\eta_{\mathrm{RE}} = \frac{\sigma^2_{\mathrm{MLE}}}{\sigma^2_{\mathrm{Robust}}} \tag{30}$$

at the nominal distribution of the MLE, is approximately 70% for the robust PS estimator and 80% for the DS estimator (the squared ratio of the dashed curves). Therefore, the robust estimators are also capable of handling uncontaminated observations without too much loss of efficiency.

C. Real Data Performance

We selected two test sites from a stack of 33 TS-X high-resolution spotlight interferograms of the super volcano Campi Flegrei, Italy. The acquisition time spans from 2009 to 2012. The coregistration and atmospheric phase correction were performed by the German Aerospace Center (DLR)'s *IWAP* processor, i.e., the integrated wide-area processor [47] based on its predecessor *PSI-GENESIS* [2]. Independent experiments were conducted by comparing the linear deformation estimates for evaluating the performance of the robust covariance estimator, the performance of the robust loss function against the uncompensated phase error, and the overall performance over a large area.

1) Robustness of the RME of Covariance: Fig. 4 shows the difference between the linear deformation rate of the DSs estimated using the classical sample covariance matrix $\hat{\mathbf{C}}_{\mathrm{MLE}}$ (left column) and that using the proposed $\hat{\mathbf{C}}_{\mathrm{RME}}$ (right column). Identical samples were used for estimating these two covariance matrices. They were adaptively selected with the KS test using *ten* amplitude images, which we expected to introduce a nonnegligible number of outliers in the selected samples because of the low detection rate from the low number of images. We employed the same classical DS MLE to estimate the deformation rate in both cases. Therefore, any improvement was solely due to the use of a more robust covariance matrix estimate.

The test area is mainly vegetation, except in the center where there is a road that usually appears as a DS in the X-band images. We had expected homogeneous deformation rates as the spans of the test area was roughly 100 m. However, many estimates appear in the left-hand figures as salt-and-pepper

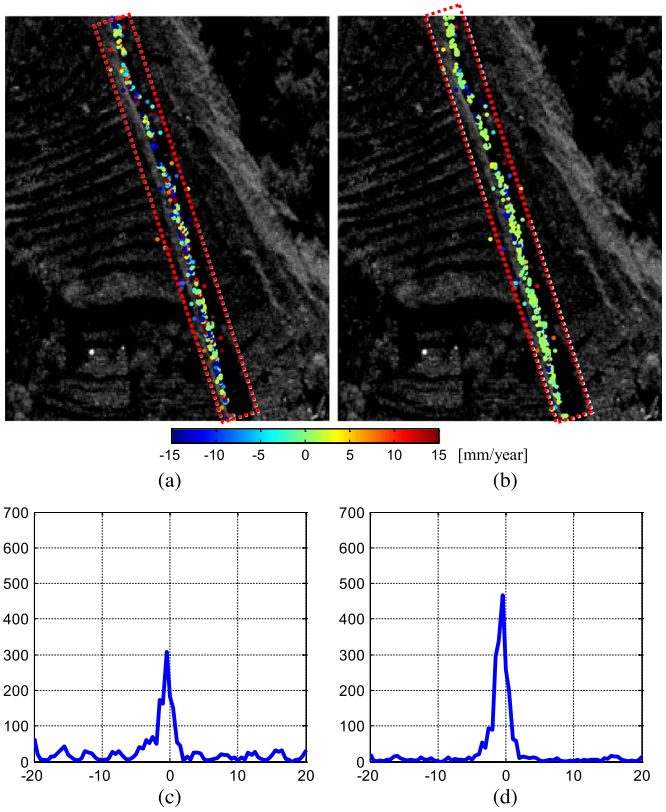


Fig. 4. Comparison of the linear deformation rate of the two test sites estimated using $\hat{\mathbf{C}}_{\mathrm{MLE}}$ and $\hat{\mathbf{C}}_{\mathrm{RME}}$. Results (overlaid on the SAR intensity images) are specific to a local reference point. A homogeneous deformation rate is expected because of the small size of the area. (a) Results using the classical $\hat{\mathbf{C}}_{\mathrm{MLE}}$. (b) Results using the proposed $\hat{\mathbf{C}}_{\mathrm{RME}}$. The samples were identical for both covariance matrix estimations. They were adaptively selected (KS test) using *ten* images, which was expected to introduce a nonnegligible number of outliers in the selected samples because of the low detection rate from the low number of images. The same classical DS MLE was employed for the deformation rate estimations in both cases. (c) and (d) plot the corresponding histograms of the linear deformation rates enclosed in the two dashed red rectangles. In (c), many local peaks appear in the search range, most of which should not correspond to the deformation signals.

noise. We believe this is due to the low detection rate of the KS test and the nonstationarity of the samples. The improvement was clearly demonstrated. $\hat{\mathbf{C}}_{\mathrm{RME}}$ significantly outperforms the widely used sample covariance matrix.

For further comparison, we plotted the histograms of the deformation rates enclosed in the two dashed red rectangles of Fig. 4(a) and (b). When using $\hat{\mathbf{C}}_{\mathrm{MLE}}$, many local peaks in the deformation rate appear in the search range. These peaks should not correspond to the deformation signal. With $\hat{\mathbf{C}}_{\mathrm{RME}}$, results are considerably more homogeneous, and thus more reasonable.

2) Robustness Against Unmodeled Phase: To evaluate the improvements by the robust loss function, we reintroduced the atmospheric phase of *ten* images randomly selected out of 33 back to the corrected phase to simulate realistic atmospheric phase error in 30% of the images. The results estimated by classical PS and DS phase parameter estimators are compared with those using their robust versions. The loss function of the robust estimators was chosen to be the Tukey's loss function with the c parameter set to 4.586. We also compare the difference of using sample covariance matrix $\hat{\mathbf{C}}_{\mathrm{MLE}}$ and rank covariance matrix $\hat{\mathbf{C}}_{\mathrm{RME}}$ in the robust DS estimator.

Fig. 5(a) and (b) compare the linear deformation rates of the combined PSs and DSs in a different test area, which are estimated using the standard and robust estimators, respectively. The same sample covariance matrix $\hat{\mathbf{C}}_{\mathrm{MLE}}$ was used in the estimation. The robust estimators outperform the classical estimators in regard of the variance of the estimates because, in this small area, we expect a homogeneous deformation rate. The improvement is solely due to the use of the robust loss function. In Fig. 5(c), we replaced the $\hat{\mathbf{C}}_{\mathrm{MLE}}$ used in Fig. 5(b) with the proposed $\hat{\mathbf{C}}_{\mathrm{RME}}$. More improvement can be observed. It demonstrates that both phase error and poor sample selection can significantly degrade the performance of classical PS and DS MLEs. However, the results can be significantly improved by introducing a robust loss function and using a better covariance matrix estimate.

For reference, Fig. 5(d) shows the results using standard PS+DS estimators with the atmospheric phase of all image compensated. Under 30% of phase contamination, the robust estimators are able to achieve comparable result to the standard estimators with no phase contamination.

3) Large Area Processing: To assess overall performance, we applied our robust estimators over the entire image area of super volcano Campi Flegrei. Fig. 6(a) shows the combined results of the linear deformation rate using the proposed RIO framework, in which Tukey's loss function is used for both robust PS and DS estimator. We specifically chose the SCM for the robust covariance estimate because of the absence of significant fringes in the interferogram. With this combination, the proposed method retrieved 15 times more scatterers than that using only PS. This number should also be valid for standard PS+DS methods, e.g., SqueeSAR. However, RIO improves the overall robustness. This can be seen in the small area comparison [see Fig. 6(b)] of the two areas marked by red rectangles in Fig. 6(a), where RIO greatly reduced the salt-and-pepper noise probably caused by unmodeled phase and bad covariance estimation.

VI. Conclusion and Outlook

The proposed RIO framework answers two open questions with respect to current multipass InSAR techniques: 1) how to downweight images with large phase error, such as the uncompensated atmospheric phase; and 2) how to correctly estimate the covariance matrix from nonstationary samples. We have provided corresponding solutions by introducing a robust loss function in the estimation of the phase history parameters and by using the *rank* of the complex multivariate together with robust weights (RME) in the covariance matrix estimation.

We demonstrate the benefits of the proposed framework by applying it to state-of-the-art estimators of PS and DS phase history parameters. Based on a simulation with 40% of its images being random-phase contaminated, the robust loss function increases the efficiency of the estimator by a factor of 7–35 compared with standard version estimators. The rank covariance M-estimator also tackles nonstationary samples to

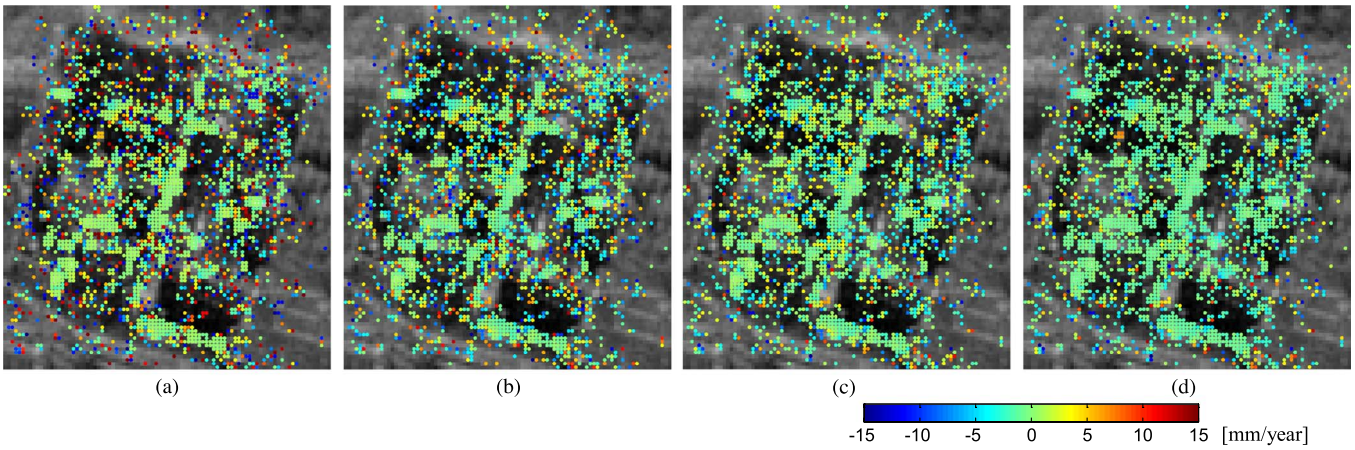


Fig. 5. Comparison of the linear deformation rate estimated using (a) standard PS and DS estimators with the sample covariance matrix $\hat{\mathbf{C}}_{\mathrm{MLE}}$ used in the DS estimator, (b) the proposed robust PS and DS phase parameter estimators using $\hat{\mathbf{C}}_{\mathrm{MLE}}$, and (c) the proposed robust PS and DS phase parameter estimators using the rank of covariance matrix $\hat{\mathbf{C}}_{\mathrm{RME}}$. The atmospheric phases of ten images out of 33 are not compensated, in order to simulate certain phase error. The loss functions of both the robust PS and DS estimators were chosen to be the Tukey's loss function with the c parameter set to 4.586. (d) shows the results using standard PS and DS estimators with all the atmospheric phase compensated. Under 30% of phase contamination, the proposed RIO method are able to achieve comparable result to the standard estimators with no phase contamination.

achieve results well beyond those of existing methods. For stationary samples, we propose alternative covariance estimators, i.e., an iterative M-estimator of covariance and a noniterative version for fast processing: sign covariance matrix.

The proposed framework can be easily extended to other multipass InSAR techniques without requiring excessive modification to the current programming. The computational effort will be approximately 3 to 4 times more than that required by standard processing, largely due to using the iterative method for solving the robust loss function, which usually converges within several iterations.

APPENDIX

A. Expectation for Rank Covariance of the SAR Acquisitions

Let us assume g_1 and g_2 are two stationary CCG random variables with covariance γ, and g_3 and g_4 are their direct neighborhood samples that have the same covariance γ, respectively. For simplicity, we suppose that all the random variables are normalized, i.e., $\sigma^2 = 1$, so that $\gamma \in [0, 1]$, which is the coherence commonly used in SAR.

Then, g_1 and g_2 can be decomposed into three mutually uncorrelated stationary circular white Gaussian processes a, b, and c, of equal variance, i.e.,

$$\begin{aligned} g_1 &= \sqrt{1-\gamma}a + \sqrt{\gamma}c \\ g_2 &= \sqrt{1-\gamma}b + \sqrt{\gamma}c. \end{aligned} \tag{31}$$

The same holds for g_3 and g_4, i.e.,

$$\begin{aligned} g_3 &= \sqrt{1-\gamma}d + \sqrt{\gamma}f \\ g_4 &= \sqrt{1-\gamma}e + \sqrt{\gamma}f. \end{aligned} \tag{32}$$

The rank covariance/coherence of g_1 and g_2, as defined in this paper (using one neighborhood sample), is then given as follows:

$$E\left\{(g_1 g_3^*)\,(g_2 g_4^*)^*\right\} = E\left\{(\gamma c f^*)(\gamma f c^*)\right\} = \gamma^2. \tag{33}$$

B. Weighting Function of Complex Multivariate t-Distribution

In general, the weighting function can be obtained by following (7). However, this method requires the transformation of the complex t-distribution into the real domain so that the derivation can be performed. Therefore, we formed a new real multivariate $\tilde{\mathbf{g}}$ using the real and imaginary parts of the complex multivariate, i.e.,

$$\begin{aligned} \tilde{\mathbf{g}} &= \begin{bmatrix} \mathrm{Re}(\mathbf{g}) \\ \mathrm{Im}(\mathbf{g}) \end{bmatrix} \in \mathbb{R}^{2N} \\ \mathbf{C}_{\tilde{\mathbf{g}}\tilde{\mathbf{g}}} &= \frac{1}{2}\begin{bmatrix} \mathrm{Re}(\mathbf{C}_{\mathbf{gg}}) & -\mathrm{Im}(\mathbf{C}_{\mathbf{gg}}) \\ \mathrm{Im}(\mathbf{C}_{\mathbf{gg}}) & \mathrm{Re}(\mathbf{C}_{\mathbf{gg}}) \end{bmatrix} \in \mathbb{R}^{2N\times 2N} \end{aligned} \tag{34}$$

$\tilde{\mathbf{g}}$ follows a $2N$-variate real t-distribution with the covariance matrix being $\mathbf{C}_{\tilde{\mathbf{g}}\tilde{\mathbf{g}}}$, i.e.,

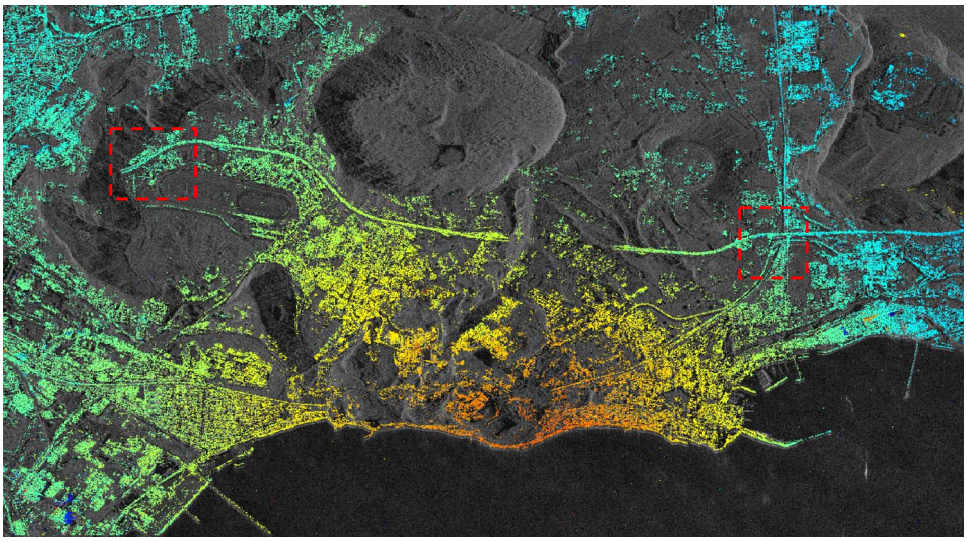
$$f_{\tilde{\mathbf{g}}}(\tilde{\mathbf{g}}) = \frac{\Gamma\left(\frac{\nu+2N}{2}\right)}{(\pi\nu)^N \Gamma\left(\frac{\nu}{2}\right)\det(\mathbf{C}_{\tilde{\mathbf{g}}\tilde{\mathbf{g}}})^{\frac{1}{2}}}\left(1+\frac{1}{\nu}\tilde{\mathbf{g}}^T\mathbf{C}_{\tilde{\mathbf{g}}\tilde{\mathbf{g}}}^{-1}\tilde{\mathbf{g}}\right)^{\frac{\nu+2N}{2}} \tag{35}$$

which, according to [33], is equivalent to the complex t-distribution.

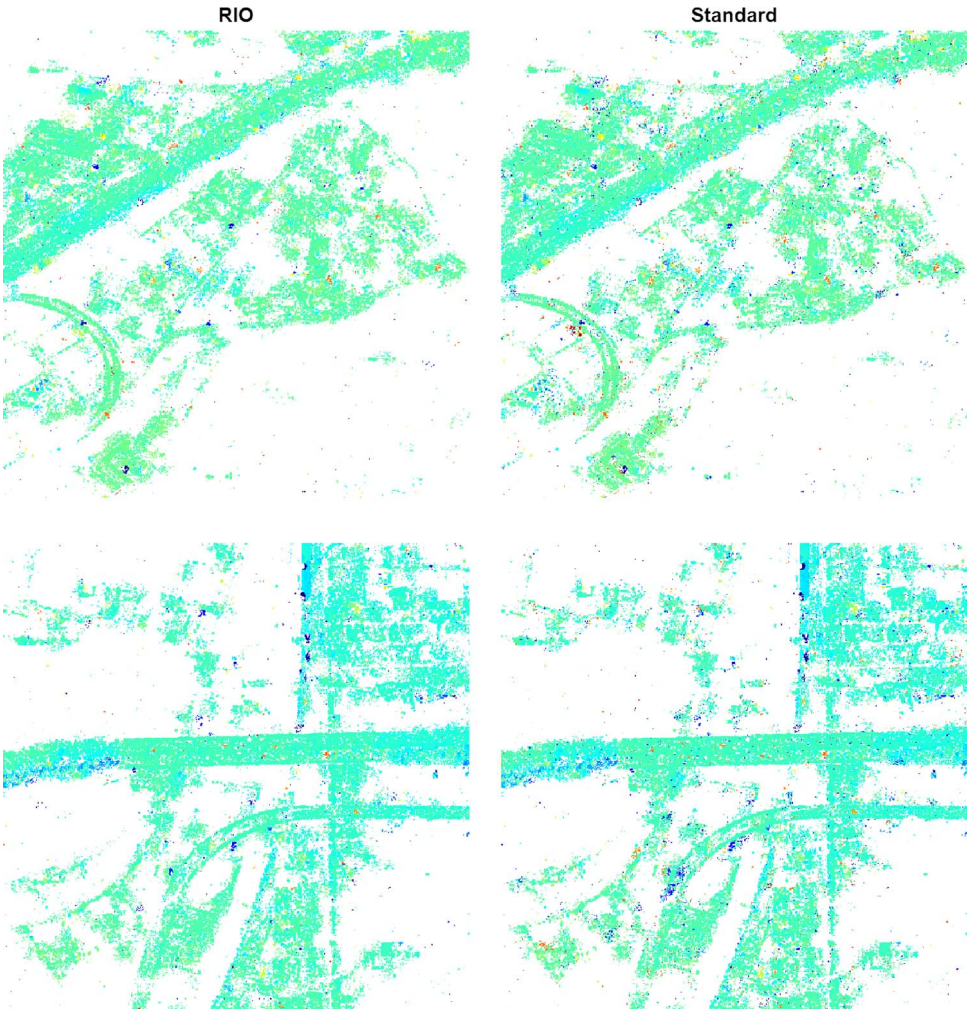
Let the squared residual $\tilde{x}^2 = \tilde{\mathbf{g}}^T\mathbf{C}_{\tilde{\mathbf{g}}\tilde{\mathbf{g}}}^{-1}\tilde{\mathbf{g}}$. As $\tilde{x}$ is whitened, it follows a univariate t-distribution, i.e.,

$$f_{\tilde{x}}(\tilde{x}) = C(1+\nu^{-1}\tilde{x}^2)^{-(\nu+2N)/2} \tag{36}$$

where C is the lengthy constant in the beginning of (34). C will disappear after considering the derivative. Now, the weighting function can be derived following (7), which is $(2N+\nu)/(\nu+\tilde{x}^2)$. Transforming it back to the complex domain, we introduce a factor of 2, as $\tilde{\mathbf{g}}^T\mathbf{C}_{\tilde{\mathbf{g}}\tilde{\mathbf{g}}}^{-1}\tilde{\mathbf{g}} = 2\mathbf{g}^H\mathbf{C}_{\mathbf{gg}}^{-1}\mathbf{g} \Rightarrow \tilde{x}^2 = 2x^2$. Hence, the corresponding weighting function for complex t-distribution is $(2N+\nu)/(\nu+2x^2)$.



(a)



(b)

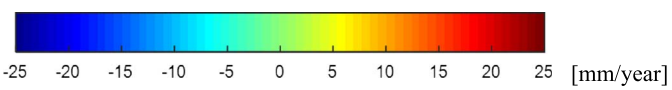


Fig. 6. (a) Linear deformation rate of the super volcano Campi Flegrei, as retrieved using the proposed RIO framework that combines both PS and DS in a stack of 33 TS-X high-resolution spotlight images. Tukey's biweight function is set to be the robust loss function, and the sign covariance matrix is employed as the robust covariance matrix estimate. It retrieves 15 times more scatterers than those using only PS only. This number should also be valid for standard PS+DS methods, e.g., SqueeSAR. However, RIO greatly reduced the salt-and-pepper noise probably caused by unmodeled phase and bad covariance estimation. This can be seen in the small area comparison in (b), which plots the results for the two areas marked by red rectangles in (a).

ACKNOWLEDGMENT

The authors would like to thank Prof. R. Bamler for valuable discussion.

REFERENCES

[1] 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.

[2] N. Adam, B. Kampes, M. Eineder, J. Worawattanamateekul, and M. Kircher, "The development of a scientific permanent scatterer system," in *Proc. ISPRS Workshop High Resolution Mapping Space*, Hannover, Germany, 2003, vol. 2003, pp. 1–6.

[3] B. M. Kampes, *Radar Interferometry—Persistent Scatterer Technique*, vol. 12. Dordrecht, The Netherlands: Springer-Verlag, 2006.

[4] F. Lombardini, "Differential tomography: A new framework for SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 1, pp. 37–44, Jan. 2005.

[5] 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.

[6] X. Zhu, *Very High Resolution Tomographic SAR Inversion for Urban Infrastructure Monitoring: A Sparse and Nonlinear Tour*, vol. 666. München, Germany: Deutsche Geodätische Kommission, 2011.

[7] P. Berardino, G. Fornaro, R. Lanari, and E. Sansosti, "A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2375–2383, Nov. 2002.

[8] 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.

[9] A. Ferretti *et al.*, "A new algorithm for processing interferometric data-stacks: SqueeSAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 9, pp. 3460–3470, Sep. 2011.

[10] Y. Wang, 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.

[11] G. Fornaro, S. Verde, D. Reale, and A. Pauciullo, "CAESAR: An approach based on covariance matrix decomposition to improve multibaseline-multitemporal interferometric SAR processing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2050–2065, Apr. 2015.

[12] R. Bamler, M. Eineder, N. Adam, X. Zhu, and S. Gernhardt, "Interferometric potential of high resolution spaceborne SAR," *Photogramm.—Fernerkund.—Geoinform.*, vol. 2009, no. 5, pp. 407–419, Nov. 2009.

[13] 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.

[14] R. Bamler and P. Hartl, "Synthetic aperture radar interferometry," *Inv. Probl.*, vol. 14, no. 4, pp. R1–R4, 1998.

[15] F. De Zan and F. Rocca, "Coherent processing of long series of SAR images," in *Proc. IEEE IGARSS*, 2005, vol. 3, pp. 1987–1990.

[16] G. Fornaro, A. Pauciullo, and F. Serafino, "Deformation monitoring over large areas with multipass differential SAR interferometry: A new approach based on the use of spatial differences," *Int. J. Remote Sens.*, vol. 30, no. 6, pp. 1455–1478, Mar. 2009.

[17] D. Yang and S. M. Buckley, "Estimating high-resolution atmospheric phase screens from radar interferometry data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 8, pp. 3117–3128, Aug. 2011.

[18] 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, Jul. 2014.

[19] M. Jehle, D. Perler, D. Small, A. Schubert, and E. Meier, "Estimation of atmospheric path delays in TerraSAR-X data using models vs. measurements," *Sensors*, vol. 8, no. 12, pp. 8479–8491, Dec. 2008.

[20] X. Cong, U. Balss, 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.

[21] F. Ulmer, N. Adam, and M. Eineder, "Ensemble based atmospheric phase screen estimation using least squares," in *Proc. IEEE IGARSS*, 2012, pp. 5606–5609.

[22] C. Gisinger *et al.*, "Precise three-dimensional stereo localization of corner reflectors and persistent scatterers with TerraSAR-X," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1782–1802, Apr. 2015.

[23] X. X. Zhu, S. Montazeri, C. Gisinger, R. Hanssen, and R. Bamler, "Geodetic SAR tomography," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 18–35, Jan. 2016.

[24] M. Jiang, X. Ding, R. F. Hanssen, R. Malhotra, and L. Chang, "Fast statistically homogeneous pixel selection for covariance matrix estimation for multitemporal InSAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1213–1224, Mar. 2015.

[25] A. Parizzi and R. Brcic, "Adaptive InSAR stack multilooking exploiting amplitude statistics: A comparison between different techniques and practical results," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 441–445, May 2011.

[26] Y. Wang, X. X. Zhu, and R. Bamler, "Robust covariance matrix estimation with application to volcano monitoring using SAR image stacks," in *Proc. 10th EUSAR* , 2014, pp. 1–4.

[27] Y. Wang and X. X. Zhu, "The robust InSAR optimization framework with application to monitoring cities on volcanoes," presented at the Joint Urban Remote Sensing Event, Lausanne, Switzerland, Mar. 30–Apr. 1, 2015.

[28] P. J. Huber, *Robust Statistics*. Hoboken, NJ, USA: Wiley, 1981.

[29] W. N. Venables and B. D. Ripley, *Modern Applied Statistics With S.* New York, NY, USA: Springer-Verlag, 2002.

[30] S. Kay, *Fundamentals of Statistical Signal Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 2001.

[31] J. L. Myers, A. D. Well, and R. F. Lorch, Jr., *Research Design and Statistical Analysis: Third Edition*, 3rd ed. London, U.K.: Routledge, 2010.

[32] C. Croux, E. Ollila, and H. Oja, "Sign and rank covariance matrices: Statistical properties and application to principal components analysis," in *Statistical Data Analysis Based on the L1-Norm and Related Methods*. Basel, Switzerland: Springer-Verlag, 2002, pp. 257–269.

[33] E. Ollila and V. Koivunen, "Influence functions for array covariance matrix estimators," in *Proc. IEEE Workshop Statist. Signal Process.*, 2003, pp. 462–465.

[34] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts," *IEEE Signal Process. Mag.*, vol. 29, no. 4, pp. 61–80, Jul. 2012.

[35] W. Y. Tan, "On the complex analogue of Bayesian estimation of a multivariate regression model," *Ann. Inst. Statist. Math.*, vol. 25, no. 1, pp. 135–152, 1973.

[36] S. Kotz and S. Nadarajah, *Multivariate t Distributions and Their Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[37] E. Ollila and V. Koivunen, "Robust antenna array processing using M-estimators of pseudo-covariance," in *Proc. 14th IEEE PIMRC*, 2003, vol. 3, pp. 2659–2663.

[38] M. Schmitt, J. L. Schonberger, and U. Stilla, "Adaptive covariance matrix estimation for multi-baseline InSAR data stacks," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 1–11, Nov. 2014.

[39] S. Visuri, V. Koivunen, and H. Oja, "Sign and rank covariance matrices," *J. Stat. Plan. Inference*, vol. 91, no. 2, pp. 557–575, Dec. 2000.

[40] A. De Maio, G. Fornaro, and A. Pauciullo, "Detection of single scatterers in multidimensional SAR imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2284–2297, Jul. 2009.

[41] D. Rife and R. R. Boorstyn, "Single tone parameter estimation from discrete-time observations," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 5, pp. 591–598, Sep. 1974.

[42] W. Rey, *Introduction to Robust and Quasi-Robust Statistical Methods*. Berlin, Germany: Springer-Verlag, 1983.

[43] P. J. Rousseeuw, "Least median of squares regression," *J. Amer. Stat. Assoc.*, vol. 79, no. 388, pp. 871–880, Dec. 1984.

[44] B. M. Kampes and R. F. Hanssen, "Ambiguity resolution for permanent scatterer interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 11, pp. 2446–2453, Nov. 2004.

[45] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-10, no. 6, pp. 853–863, Nov. 1974.

[46] H. A. Zebker and J. Villasenor, "Decorrelation in interferometric radar echoes," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 5, pp. 950–959, Sep. 1992.

[47] 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 Symp., Edinburgh, U.K., Sep. 9–13, 2013, Paper ESA SP-722.

Yuanyuan Wang (S'10) received the B.Eng. (Hons) degree in electrical engineering from The Hong Kong Polytechnic University, Hong Kong, in 2008 and the M.Sc. degree in Earth oriented space science and technology from Technische Universität München (TUM), Munich, Germany, in 2010. Since 2011, he has been pursuing the Doctoral degree with the International Graduate School of Science and Engineering, TUM, and currently with the project "Joy of Long Baseline" of German Aerospace Center.

In June and July 2014, he was a Guest Scientist with the Institute of Visual Computing, ETH Zürich, Zürich, Switzerland. He is currently with Signal Processing in Earth Observation (SiPEO), TUM. His research interests include optimal parameters estimation in interferometric SAR techniques, multisensor fusion algorithms of SAR/optical data, and the applications of these techniques in urban and volcanic areas.

Xiao Xiang Zhu (S'10–M'12–SM'14) received the Bachelor's degree in space engineering from the National University of Defense Technology, Changsha, China, in 2006 and the M.Sc. degree, Dr.-Ing. degree, and Habilitation in signal processing from Technische Universität München (TUM), Munich, Germany, in 2008, 2011, and 2013, respectively.

She was a Guest Scientist or a Visiting Professor with the Italian National Research Council (CNR-IREA), Naples, Italy, in 2009; with Fudan University, Shanghai, China, in 2014; and with the University of Tokyo, Tokyo, Japan in 2015.

Since 2011, she has been a Scientist with the Remote Sensing Technology Institute, German Aerospace Center (DLR), Oberpfaffenhofen, where she is currently the Head of the Signal Analysis Team. Since 2013, she has also been a Helmholtz Young Investigator Group Leader and appointed as a TUM Junior Fellow. She is currently a Professor of signal processing in earth observation with TUM. Her main research interests include advanced interferometric SAR techniques such as high-dimensional tomographic SAR imaging and SqueeSAR; computer vision in remote sensing such as object reconstruction and multidimensional data visualization; and modern signal processing, such as innovative algorithms (e.g., compressive sensing and sparse reconstruction), with applications in the field of remote sensing such as multi/hyperspectral image analysis.

Dr. Zhu serves as an Associate Editor for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

Appendices

F Publishers Copyright Policies

For Authors

Journal authors' home
Author Rights
Ethics
Funding body agreements
Open access
Author services
Journal performance
Early career researchers
Authors' update
Book authors' home
Getting your paper noticed

Author Rights

Elsevier supports the need for authors to share, disseminate and maximize the impact of their research. We take our responsibility as stewards of the online record seriously, and work to ensure our policies and procedures help to protect the integrity of scholarly works.

Author's rights to reuse and post their own articles published by Elsevier are defined by Elsevier's copyright policy. For our proprietary titles, the type of copyright agreement used depends on the author's choice of publication:

For subscription articles: These rights are determined by a copyright transfer, where authors retain scholarly rights to post and use their articles.

For open access articles: These rights are determined by an exclusive license agreement, which applies to all our open access content.

In both cases, the fundamental rights needed to publish and distribute an article remain the same and Elsevier authors will be able to use their articles for a wide range of scholarly purposes.

Details on how authors can reuse and post their own articles are provided below.

Help and support
For reuse and posting not detailed below, please see our posting policy, or for authors who would like to:

- Include material from other sources in your work being published by Elsevier, please visit: Permission seeking guidelines for Elsevier authors.
- Obtain permission to re-use material from Elsevier books, journals, databases, or other products, please visit: Obtaining permission to reuse Elsevier material
- Or if you are an Elsevier author and are contacted by a requestor who wishes to re-use all or part of your article or chapter, please also refer them to our Obtaining Permission to Re-Use Elsevier Material page.
- See our FAQ on posting and copyright queries.
- Contact us directly, please email our Permissions Help Desk.

Author Use | Author Posting | Definitions

How authors can use their own journal articles

Authors can use their articles for a wide range of scholarly, non-commercial purposes as outlined below. These rights apply for all Elsevier authors who publish their article as either a subscription article or an open access article.

We require that all Elsevier authors always include a full acknowledgement and, if appropriate, a link to the final published version hosted on Science Direct.

For open access articles these rights are separate from how readers can reuse your article as defined by the author's choice of Creative Commons user license options.

Authors can use either their accepted author manuscript or final published article for:	
✓	Use at a conference, meeting or for teaching purposes
✓	Internal training by their company
✓	Sharing individual articles with colleagues for their research use* (also known as 'scholarly sharing')
✓	Use in a subsequent compilation of the author's works
✓	Inclusion in a thesis or dissertation
✓	Reuse of portions or extracts from the article in other works
✓	Preparation of derivative works (other than for commercial purposes)

*Please note this excludes any systematic or organized distribution of published articles.

IEEE COPYRIGHT AND CONSENT FORM

To ensure uniformity of treatment among all contributors, other forms may not be substituted for this form, nor may any wording of the form be changed. This form is intended for original material submitted to the IEEE and must accompany any such material in order to be published by the IEEE. Please read the form carefully and keep a copy for your files.

TITLE OF PAPER/ARTICLE/REPORT, INCLUDING ALL CONTENT IN ANY FORM, FORMAT, OR MEDIA (hereinafter, "the Work"):

COMPLETE LIST OF AUTHORS:

IEEE PUBLICATION TITLE (Journal, Magazine, Conference, Book):

COPYRIGHT TRANSFER

1. The undersigned hereby assigns to The Institute of Electrical and Electronics Engineers, Incorporated (the "IEEE") all rights under copyright that may exist in and to: (a) the above Work, including any revised or expanded derivative works submitted to the IEEE by the undersigned based on the Work; and (b) any associated written or multimedia components or other enhancements accompanying the Work.

CONSENT AND RELEASE

2. In the event the undersigned makes a presentation based upon the Work at a conference hosted or sponsored in whole or in part by the IEEE, the undersigned, in consideration for his/her participation in the conference, hereby grants the IEEE the unlimited, worldwide, irrevocable permission to use, distribute, publish, license, exhibit, record, digitize, broadcast, reproduce and archive, in any format or medium, whether now known or hereafter developed: (a) his/her presentation and comments at the conference; (b) any written materials or multimedia files used in connection with his/her presentation; and (c) any recorded interviews of him/her (collectively, the "Presentation"). The permission granted includes the transcription and reproduction of the Presentation for inclusion in products sold or distributed by IEEE and live or recorded broadcast of the Presentation during or after the conference.

3. In connection with the permission granted in Section 2, the undersigned hereby grants IEEE the unlimited, worldwide, irrevocable right to use his/her name, picture, likeness, voice and biographical information as part of the advertisement, distribution and sale of products incorporating the Work or Presentation, and releases IEEE from any claim based on right of privacy or publicity.

4. The undersigned hereby warrants that the Work and Presentation (collectively, the "Materials") are original and that he/she is the author of the Materials. To the extent the Materials incorporate text passages, figures, data or other material from the works of others, the undersigned has obtained any necessary permissions. Where necessary, the undersigned has obtained all third party permissions and consents to grant the license above and has provided copies of such permissions and consents to IEEE.

☐ Please check this box if you do not wish to have video/audio recordings made of your conference presentation.

See reverse side for Retained Rights/Terms and Conditions, and Author Responsibilities.

GENERAL TERMS

- The undersigned represents that he/she has the power and authority to make and execute this assignment.
- The undersigned agrees to indemnify and hold harmless the IEEE from any damage or expense that may arise in the event of a breach of any of the warranties set forth above.
- In the event the above work is not accepted and published by the IEEE or is withdrawn by the author(s) before acceptance by the IEEE, the foregoing copyright transfer shall become null and void and all materials embodying the Work submitted to the IEEE will be destroyed.
- For jointly authored Works, all joint authors should sign, or one of the authors should sign as authorized agent for the others.

(1)_____ _____

Author/Authorized Agent for Joint Authors **Date**

U.S. GOVERNMENT EMPLOYEE CERTIFICATION (WHERE APPLICABLE)

This will certify that all authors of the Work are U.S. government employees and prepared the Work on a subject within the scope of their official duties. As such, the Work is not subject to U.S. copyright protection.

(2)_____ _____

Authorized Signature **Date**

(Authors who are U.S. government employees should also sign signature line (1) above to enable the IEEE to claim and protect its copyright in international jurisdictions.)

CROWN COPYRIGHT CERTIFICATION (WHERE APPLICABLE)

This will certify that all authors of the Work are employees of the British or British Commonwealth Government and prepared the Work in connection with their official duties. As such, the Work is subject to Crown Copyright and is not assigned to the IEEE as set forth in the first sentence of the Copyright Transfer Section above. The undersigned acknowledges, however, that the IEEE has the right to publish, distribute and reprint the Work in all forms and media.

(3)_____ _____

Authorized Signature **Date**

(Authors who are British or British Commonwealth Government employees should also sign line (1) above to indicate their acceptance of all terms other than the copyright transfer.)

rev. 020711

IEEE COPYRIGHT FORM *(continued)*

RETAINED RIGHTS/TERMS AND CONDITIONS

General

1. Authors/employers retain all proprietary rights in any process, procedure, or article of manufacture described in the Work.
2. Authors/employers may reproduce or authorize others to reproduce the Work, material extracted verbatim from the Work, or derivative works for the author's personal use or for company use, provided that the source and the IEEE copyright notice are indicated, the copies are not used in any way that implies IEEE endorsement of a product or service of any employer, and the copies themselves are not offered for sale.
3. In the case of a Work performed under a U.S. Government contract or grant, the IEEE recognizes that the U.S. Government has royalty-free permission to reproduce all or portions of the Work, and to authorize others to do so, for official U.S. Government purposes only, if the contract/grant so requires.
4. Although authors are permitted to re-use all or portions of the Work in other works, this does not include granting third-party requests for reprinting, republishing, or other types of re-use. The IEEE Intellectual Property Rights office must handle all such third-party requests.
5. Authors whose work was performed under a grant from a government funding agency are free to fulfill any deposit mandates from that funding agency.

Author Online Use

6. Personal Servers. Authors and/or their employers shall have the right to post the accepted version of IEEE-copyrighted articles on their own personal servers or the servers of their institutions or employers without permission from IEEE, provided that the posted version includes a prominently displayed IEEE copyright notice and, when published, a full citation to the original IEEE publication, including a link to the article abstract in IEEE Xplore. Authors shall not post the final, published versions of their papers.
7. Classroom or Internal Training Use. An author is expressly permitted to post any portion of the accepted version of his/her own IEEE-copyrighted articles on the author's personal web site or the servers of the author's institution or company in connection with the author's teaching, training, or work responsibilities, provided that the appropriate copyright, credit, and reuse notices appear prominently with the posted material. Examples of permitted uses are lecture materials, course packs, e-reserves, conference presentations, or in-house training courses.
8. Electronic Preprints. Before submitting an article to an IEEE publication, authors frequently post their manuscripts to their own web site, their employer's site, or to another server that invites constructive comment from colleagues. Upon submission of an article to IEEE, an author is required to transfer copyright in the article to IEEE, and the author must update any previously posted version of the article with a prominently displayed IEEE copyright notice. Upon publication of an article by the IEEE, the author must replace any previously posted electronic versions of the article with either (1) the full citation to the IEEE work with a Digital Object Identifier (DOI) or link to the article abstract in IEEE Xplore, or (2) the accepted version only (not the IEEE-published version), including the IEEE copyright notice and full citation, with a link to the final, published article in IEEE Xplore.

INFORMATION FOR AUTHORS

Author Responsibilities

The IEEE distributes its technical publications throughout the world and wants to ensure that the material submitted to its publications is properly available to the readership of those publications. Authors must ensure that their Work meets the requirements as stated in section 8.2.1 of the IEEE PSPB Operations Manual, including provisions covering originality, authorship, author responsibilities and author misconduct. More information on IEEE's publishing policies may be found at http://www.ieee.org/publications_standards/publications/rights/pub_tools_policies.html. Authors are advised especially of IEEE PSPB Operations Manual section 8.2.1.B12: "It is the responsibility of the authors, not the IEEE, to determine whether disclosure of their material requires the prior consent of other parties and, if so, to obtain it." Authors are also advised of IEEE PSPB Operations Manual section 8.1.1B: "Statements and opinions given in work published by the IEEE are the expression of the authors."

Author/Employer Rights

If you are employed and prepared the Work on a subject within the scope of your employment, the copyright in the Work belongs to your employer as a work-for-hire. In that case, the IEEE assumes that when you sign this Form, you are authorized to do so by your employer and that your employer has consented to the transfer of copyright, to the representation and warranty of publication rights, and to all other terms and conditions of this Form. If such authorization and consent has not been given to you, an authorized representative of your employer should sign this Form as the Author.

IEEE Copyright Ownership

It is the formal policy of the IEEE to own the copyrights to all copyrightable material in its technical publications and to the individual contributions contained therein, in order to protect the interests of the IEEE, its authors and their employers, and, at the same time, to facilitate the appropriate re-use of this material by others. The IEEE distributes its technical publications throughout the world and does so by various means such as hard copy, microfiche, microfilm, and electronic media. It also abstracts and may translate its publications, and articles contained therein, for inclusion in various compendiums, collective works, databases and similar publications.

THIS FORM MUST ACCOMPANY THE SUBMISSION OF THE AUTHOR'S MANUSCRIPT.
Questions about the submission of the form or manuscript must be sent to the publication's editor.
Please direct all questions about IEEE copyright policy to:
IEEE Intellectual Property Rights Office, copyrights@ieee.org, +1-732-562-3966 (telephone)