ISPRS Journal of Photogrammetry and Remote Sensing 146 (2018) 389–408

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

A framework for SAR-optical stereogrammetry over urban areas

Hossein Bagheri[a], Michael Schmitt[a], Pablo d'Angelo[b], Xiao Xiang Zhu[a,b,*]

[a] Signal Processing in Earth Observation, Technical University of Munich, Munich, Germany
[b] Remote Sensing Technology Institute, German Aerospace Center, Oberpfaffenhofen, Wessling, Germany

ARTICLE INFO

Keywords:
Epipolarity constraint
Multi-sensor block adjustment
Dense image matching
3D reconstruction
SAR-optical stereogrammetry

ABSTRACT

Currently, numerous remote sensing satellites provide a huge volume of diverse earth observation data. As these data show different features regarding resolution, accuracy, coverage, and spectral imaging ability, fusion techniques are required to integrate the different properties of each sensor and produce useful information. For example, synthetic aperture radar (SAR) data can be fused with optical imagery to produce 3D information using stereogrammetric methods. The main focus of this study is to investigate the possibility of applying a stereogrammetry pipeline to very-high-resolution (VHR) SAR-optical image pairs. For this purpose, the applicability of semi-global matching is investigated in this unconventional multi-sensor setting. To support the image matching by reducing the search space and accelerating the identification of correct, reliable matches, the possibility of establishing an epipolarity constraint for VHR SAR-optical image pairs is investigated as well. In addition, it is shown that the absolute geolocation accuracy of VHR optical imagery with respect to VHR SAR imagery such as provided by TerraSAR-X can be improved by a multi-sensor block adjustment formulation based on rational polynomial coefficients. Finally, the feasibility of generating point clouds with a median accuracy of about 2 m is demonstrated and confirms the potential of 3D reconstruction from SAR-optical image pairs over urban areas.

1. Introduction

Three-dimensional reconstruction from remote sensing data has a range of applications across different fields, such as urban 3D modeling and management, environmental studies, and geographic information systems. Manifold high-resolution sensors in space provide the possibility of reconstructing natural and man-made landscapes over large-scale areas. Conventionally, 3D reconstruction in remote sensing is either based on exploiting phase information provided by interferometric SAR, or on space intersection in the frame of photogrammetry with optical images or radargrammetry with SAR image pairs. In all these stereogrammetric approaches, at least two overlapping images are required to extract 3D spatial information. Both photogrammetry and radargrammetry, however, suffer from several drawbacks. Photogrammetry using high-resolution optical imagery is limited by relatively poor absolute localization accuracy and cloud effects, whereas radargrammetry suffers from the difficulty of image matching for severely different oblique viewing angles.

On the other hand, the huge archives of high-resolution SAR images provided by satellites such as TerraSAR-X and the regular availability of new data alongside archives of high-resolution optical imagery provided by sensors such as WorldView provide a great opportunity to investigate data fusion pipelines for producing 3D spatial information (Schmitt and Zhu, 2016). As relatively few studies have dealt with 3D reconstruction from SAR-optical image pairs (Bloom et al., 1988; Raggam et al., 1994; Wegner et al., 2014), there has been no investigation into the feasibility of a dense multi-sensor stereo pipeline as known from photogrammetric computer vision yet. This paper investigates the possibility of implementing such a pipeline, and describes all processing steps required for 3D reconstruction from very-high-resolution (VHR) SAR-optical image pairs.

In detail, this paper discusses both an epipolarity constraint and a bundle adjustment formulation for SAR-optical multi-sensor stereogrammetry first. Regarding the complicated radiometric relationship between SAR and optical imagery, the epipolarity constraint accelerates the matching process and helps to identify reliable and correct conjugate points (Morgan et al., 2004; Scharstein et al., 2001). For this objective, we first demonstrate the existence of an epipolarity constraint for SAR-optical imagery by reconstructing the rigorous geometry models of SAR and optical sensors using both collinearity and range-Doppler relationships. We prove that a SAR-optical epipolarity constraint can be rigorously modeled using the sensor geometries. Subsequently, rational polynomial coefficients (RPCs) are fitted to the SAR sensor geometry to ease further processing steps. Consequently,

* Corresponding author.
E-mail address: xiaoxiang.zhu@dlr.de (X.X. Zhu).

https://doi.org/10.1016/j.isprsjprs.2018.10.003
Received 20 July 2018; Received in revised form 28 September 2018; Accepted 9 October 2018
Available online 28 October 2018

epipolar curves can be established using projection and back-projection from SAR imagery to terrain and then from terrain to optical imagery using RPCs. In addition, the RPCs ease the formulation of multi-sensor block adjustment for SAR-optical imagery.

The block adjustment is used to align the optical imagery with respect to the SAR data. Generally, the absolute geolocalization accuracy of optical satellite imagery is lower than that of modern SAR sensors. Evaluations show that the absolute accuracy of geopositioning using TerraSAR-X imagery is within a single resolution cell in both the azimuth and range directions, and can even go down to the cm-level (Eineder et al., 2011). In contrast, the absolute accuracy of geolocalization using basic WorldView-2 products is generally no better than 3 m (DigitalGlobe, 2018). Consequently, the block adjustment propagates the high geometrical accuracy of SAR data into the final 3D product, thus avoiding the need for external control points.

The main stage of SAR-optical stereogrammetry, however, is a dense matching algorithm for 3D reconstruction (Bagheri et al., 2018). In this study, we use the semi-global matching (SGM) method, which incorporates both mutual information and census, as well as their weighted sum as cost functions, in its core.

The remainder of this paper is organized as follows. First, the modeling of SAR sensor geometries with RPCs is explained in Section 2.1. After briefly introducing the epipolarity constraint and its benefits, a mathematical proof of this constraint for SAR-optical image pairs is presented in Section 2.2. In Section 2.3, the application of multi-sensor block adjustment using RPCs for SAR-optical image pairs is introduced. The principle of the SGM algorithm is recapitulated in Section 2.4. Section 3 summarizes experiments and results of our implementation of the SAR-optical stereogrammetry workflow for TerraSAR-X/WorldView-2 image pairs over two urban study areas. Based on these results, the feasibility of stereogrammetric 3D reconstruction from SAR-optical image pairs over urban areas, as well as its advantages and limitations, are discussed in Section 4. Finally, Section 5 presents the conclusions to this study.

2. SAR-optical stereogrammetry

Fig. 1 shows the general framework of SAR-optical stereogrammetric 3D reconstruction. Similar to optical stereogrammetry, one grayscale optical image and one amplitude SAR image form a stereo image pair that can be processed by suitable matching methods to find all possible conjugate pixels. However, some important pre-processing steps are required before the matching and 3D reconstruction. Currently, most VHR optical images are delivered using RPCs. Thus, the primary step in the SAR-optical stereogrammetry framework is to estimate the RPCs for SAR imagery as well. This process homogenizes the geometry models of both sensors and simplifies the subsequent processes of SAR-optical block adjustment and establishing an epipolarity constraint. The next phase is to carry out multi-sensor block adjustment to align the optical image to the SAR image. This rectifies the RPCs of the optical imagery with respect to the SAR imagery, thus improving the absolute geolocalization of the optical imagery and correcting the positions of the epipolar curves on the optical imagery. A disparity map is then produced in the frame of the reference image via a dense image matching algorithm such as SGM. From this map, the 3D positions of the points can be determined by reconstructing the geometry of the SAR and optical imagery for a particular exposure. However, the success of the aforementioned framework relies on the possibility of establishing an epipolarity constraint for SAR-optical image pairs. Thus, the existence of the epipolarity constraint for SAR-optical image pairs must be investigated. In the following, the details of each step of the SAR-optical stereogrammetry framework are explained and the potential of using an epipolarity constraint for SAR-optical image pairs is investigated.

2.1. Preparation: RPCs for SAR imagery

RPCs are a well-established substitute for the rigorously derived optical imaging model. They are widely used for different purposes such as epipolar curve reconstruction (Oh et al., 2010), block adjustment (Grodecki and Dial, 2003), space resection-intersection and 3D reconstruction (Fraser et al., 2006; Li et al., 2007; Tao and Hu, 2002; Tao et al., 2004; Toutin, 2006) or image rectification (Tao and Hu, 2001). The relation between the image space and the geographic reference system is created by the rational functions (Grodecki et al., 2004)

$$c = \frac{P_1(\lambda, \phi, h)}{P_2(\lambda, \phi, h)} = f(\lambda, \phi, h) \tag{1}$$

and

$$r = \frac{P_3(\lambda, \phi, h)}{P_4(\lambda, \phi, h)} = g(\lambda, \phi, h), \tag{2}$$

where r, c are normalized image coordinates, i.e. normalized rows and columns of points in the scene and ϕ, λ, and h denote the normalized latitude, longitude, and height of the respective ground point. The relationship between normalized and un-normalized coordinates is given by Tao and Hu (2001)

$$X = \frac{X_u - X_o}{S_x}, \tag{3}$$

where X is the normalized coordinate, X_u is the un-normalized value of the coordinate, and X_o, S_x are the offset and scale factors, respectively.

In Eqs. (1) and (2), P_i $(i = 1, \ldots, 4)$ are n-order polynomial functions that are used to model the relationship between the image space and the reference system. They can be written as

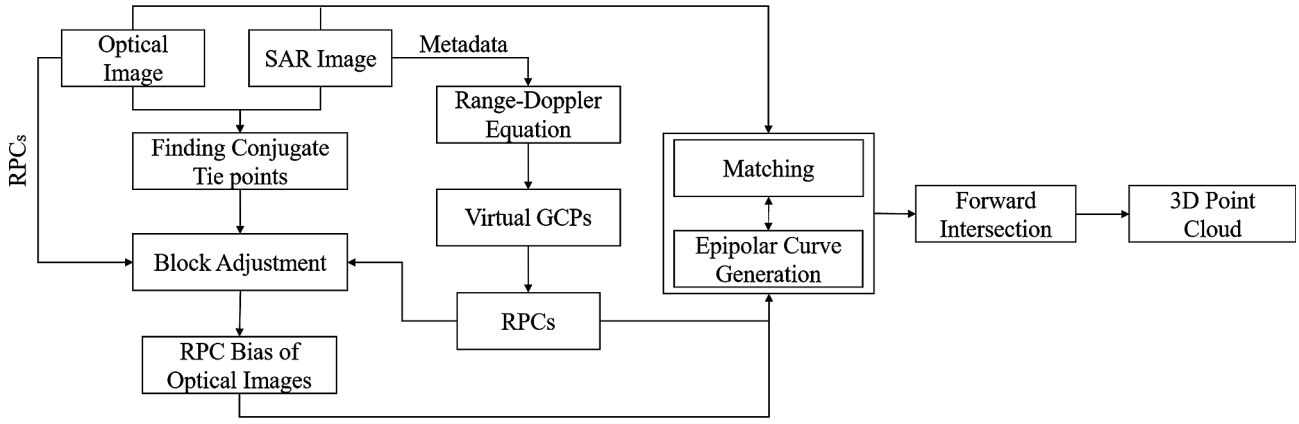


Fig. 1. Framework for stereogrammetric 3D reconstruction from SAR-optical image pairs.

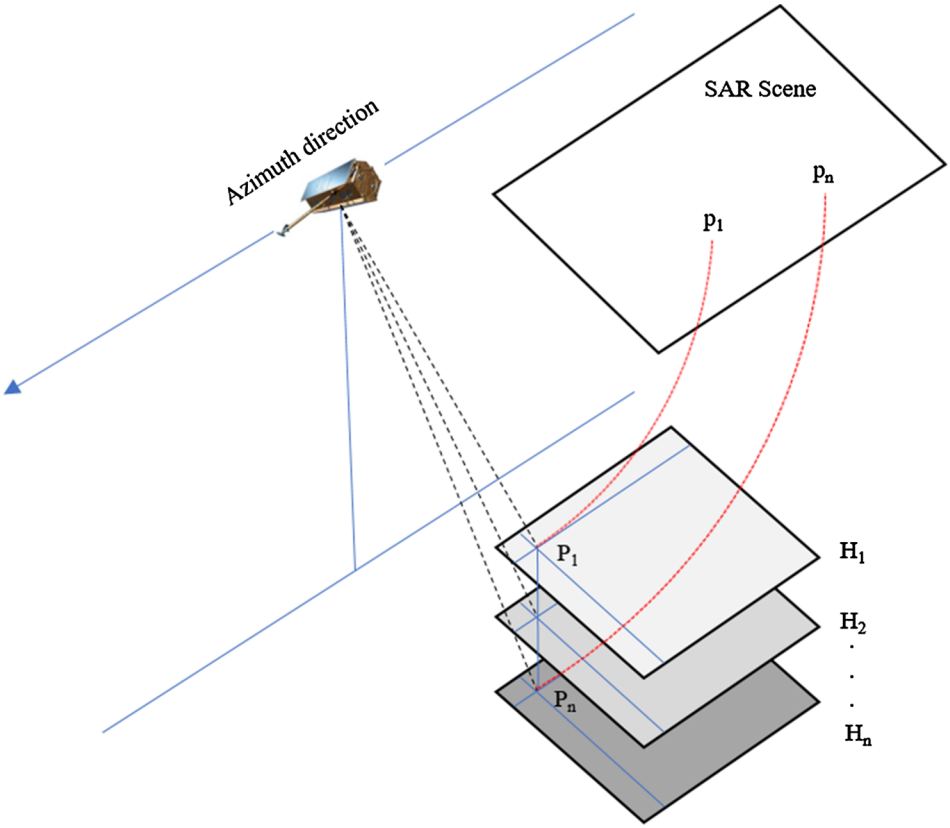


Fig. 2. Procedure of estimating RPCs by terrain-independent approach.

$$P_i = a_{i,0} + a_{i,1}h + a_{i,2}\phi + a_{i,3}\lambda + a_{i,4}h\phi + a_{i,5}h\lambda + a_{i,6}\phi\lambda + a_{i,7}h^2 + a_{i,8}\phi^2 + a_{i,9}\lambda^2 + a_{i,10}h\phi\lambda + a_{i,11}h^2\phi + a_{i,12}h^2\lambda + a_{i,13}\phi^2 h + a_{i,14}\phi^2\lambda + a_{i,15}h\lambda^2 + a_{i,16}\phi\lambda^2 + a_{i,17}h^3 + a_{i,18}\phi^3 + a_{i,19}\lambda^3, \quad (4)$$

where $a_{i,n}$ $(n = 0, 1, \ldots, 19)$ are the polynomial coefficients.

For projection from the image space to terrain, the inverse form of the rational function models is used:

$$\lambda = \frac{P_5(c, r, h)}{P_6(c, r, h)} = f'(c, r, h) \quad (5)$$

and

$$\phi = \frac{P_7(c, r, h)}{P_8(c, r, h)} = g'(c, r, h) \quad (6)$$

For this task, another set of RPCs for inverse projection as well as the terrain height h is needed.

The main reason for using RPCs is to facilitate the computational process of the subsequent processing tasks. Instead of describing the stereogrammetric intersection with a combination of the range-Doppler model for the SAR image and a push-broom model for the optical image, from a mathematical point of view the RPC formulation homogenizes everything to a comparably simple joint model. However, fitting RPCs to a sensor model is challenging in its own way and demands sufficient, well-distributed control points. RPCs are usually calculated with either a terrain-independent or terrain-dependent approach (Tao and Hu, 2001). In the terrain-dependent approach, accurate Ground Control Points (GCPs) are used to estimate the RPCs. Thus, the final accuracy of the RPCs depends on the number, accuracy, and distribution of GCPs.

While the terrain-dependent approach is an expensive way of estimating RPCs (and GCPs may not be available for every study area), the terrain-independent method allows RPCs to be estimated without any GCPs (Tao and Hu, 2001). Instead, a set of virtual GCPs (VGCPs), which are related to the image through the rigorous imaging model of the respective sensor, are used to approximate the RPCs. The VGCPs are arranged in a grid-shape format on planes located at different heights over the study area, such as depicted in Fig. 2. The resulting cube of points is then projected to the image space, and their corresponding image coordinates are determined by reconstructing the rigorous model. The RPCs can subsequently be estimated using a least-squares calculation. Note that, when using higher-order RPCs, the least-squares estimation suffers from an ill-posed configuration that causes the results to deviate from their optimal values. In these circumstances, a regularization approach based on Tikhonovs method can be employed to obtain acceptable solutions (Tikhonov and Arsenin, 1977).

The RPCs for optical sensors are usually delivered by vendors alongside the image files. For SAR sensors – with the exception of the Chinese satellite GaoFen-3 – however usually only ephemerids and orbital parameters are attached to the data. Therefore, Zhang et al. investigated the generation of RPCs for various SAR sensors based on the terrain-independent approach (Zhang et al., 2011). Their results show that RPCs can be used as substitutes for the range-Doppler equations with acceptable accuracy. In this study, we use this terrain-independent approach for SAR RPC generation.

2.2. Epipolarity constraint for SAR-optical image matching

In most stereogrammetric 3D reconstruction scenarios, the epipolarity constraint facilitates the procedure of image matching by reducing the search space from 2D to 1D (Morgan et al., 2004). The epipolarity constraint always exists for optical stereo images captured by frame-type cameras that follow a perspective projection (Cho et al., 1993). This phenomenon is illustrated in Fig. 3.

For a point p in the *left-hand* image, the conjugate point in the corresponding *right-hand* image is located on the so-called epipolar line. This epipolar line lies on the plane passing through both image

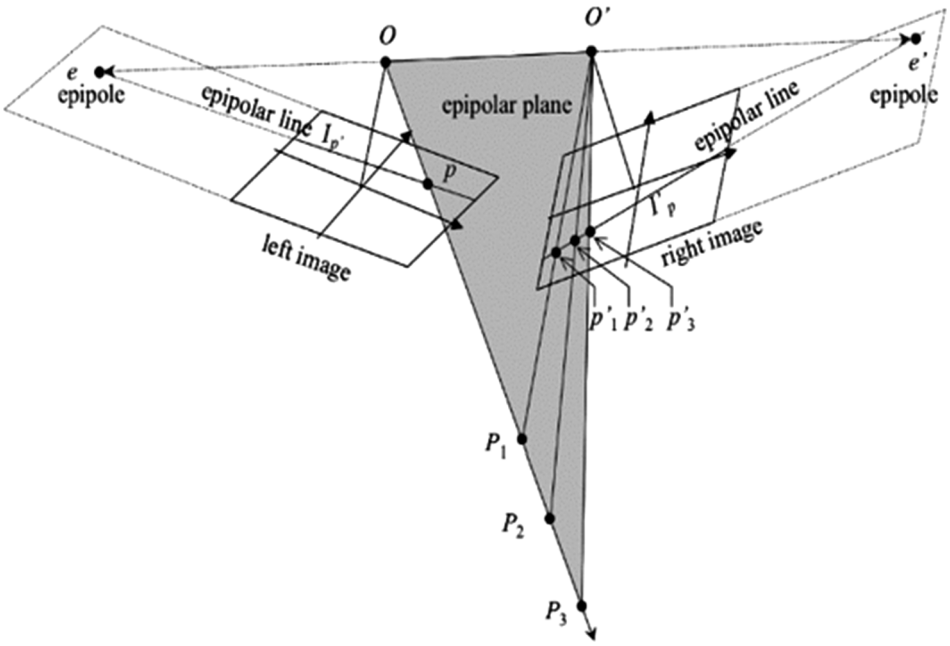


Fig. 3. Epipolarity constraint for frame-type camera (Morgan et al., 2004).

projection centers (O, O') and the image point p. It can also be obtained by changing the depth or height of p in the reference coordinate system. While it is known that epipolar lines exist for images captured from frame-type cameras, straightness cannot be ensured for other sensor types (Cho et al., 1993). We thus refer to epipolar curves instead of epipolar lines to express generality in the remainder of this paper.

With respect to remote sensing, several studies have demonstrated that the epipolar curves for scenes acquired by linear array push-broom sensors are not straight (Gupta and Hartley, 1997; Kim, 2000). For example, Kim (2000) used the model developed by Orun (1994) to prove that the epipolar curves in SPOT scenes looks like hyperbolas. Orun and Natarajan's model assumes that the rotational roll and pitch parameters are constant during the flight, while the yaw can be modeled by quadratic time-dependent polynomials. Morgan et al. (2004) demonstrated that the epipolar curves would not be straight even with uniform motion.

For a SAR sensor, the imaging geometry is completely different from that of optical sensors, as data are collected in a side-looking manner based on the range-Doppler geometry (Curlander, 1982). However, the possibility of establishing the epipolarity constraint in stereo SAR image pairs has been investigated by Gutjahr et al. (2014) and Li and Zhang (2013) for radargrammetric 3D reconstruction. Gutjahr et al. experimentally showed that epipolar curves in SAR image pairs are also not perfectly straight, but can be approximately assumed to be straight for radargrammetric 3D reconstruction tasks through dense matching (Gutjahr et al., 2014).

In this research, we investigate the epipolarity constraint mathematically and experimentally for the unconventional multi-sensor situation of SAR-optical image pairs. In general, epipolar curves in image pairs captured by frame-type cameras (as shown in Fig. 3) can be described as (Hartley and Zisserman, 2004)

$$l_r = \mathbf{F}^T p' \tag{7}$$

where l_r refers to the epipolar curve in the right-hand image associated with the image point p' on the left-hand image. $\mathbf{F}$ is the fundamental matrix, which includes interior and exterior orientation parameters for projecting coordinates between the two images. Similarly, an epipolar curve in the left-hand image can be written as $l_l = \mathbf{F}p''$. For push-broom satellite image pairs, the epipolarity constraint can be verified in a similar way, but linear arrays are substituted for a frame image. Furthermore, the fundamental matrix for push-broom sensors is more complex than that for frame-type sensors. In the following, inspired by the mathematical proof of the epipolarity constraint for stereo optical imagery given in Morgan et al. (2004) and using the epipolar curve equation presented in (7), a rigorous epipolar model for SAR-optical image pairs acquired by space-borne platforms will be constructed. For this task, the optical image is considered as the left-hand image and the SAR image is the right-hand image. Fig. 4 shows the configuration of the SAR-optical stereo case. The points o and s mark the positions of the optical linear array push-broom sensor and the SAR sensor, respectively. Using a collinearity condition, a rigorous model for reconstructing the imaging geometry of linear array push-broom sensors can be expressed as (Kratky, 1989)

$$\begin{pmatrix} x_l = 0 \\ y_l \\ f \end{pmatrix} = \lambda R_{\omega(t)\phi(t)\kappa(t)} \begin{pmatrix} X - X^o(t) \\ Y - Y^o(t) \\ Z - Z^o(t) \end{pmatrix}, \tag{8}$$

where (x_l, y_l) are the coordinates of point p in the linear array coordinate system, f is the focal length, $(X^o(t), Y^o(t), Z^o(t))$ represents the satellite position at time t in the reference coordinate system, (X, Y, Z) are the ground coordinates of the target point T, λ is the scale factor, and $\mathbf{R}_{\omega(t)\phi(t)\kappa(t)}$ is the 3D rotational matrix computed from rotations $\omega(t)$, $\phi(t)$, $\kappa(t)$ along the three dimensions at time t. Note that the aforementioned rotational and translational components are estimated by time-dependent polynomials.

A rigorous model based on the range-Doppler geometry (Curlander and McDonough, 1991) (displayed in Fig. 4 as well) can also be applied to the SAR imagery. In this model, the slant-range equation is first used to describe the range sphere as (Curlander, 1982):

$$R = \|\mathbf{R}_{CT} - \mathbf{R}_{CS}\| \tag{9}$$

where R is the slant-range and $\mathbf{R}_{CT}$, $\mathbf{R}_{CS}$ are the target point and SAR sensor position vectors in the reference coordinate system. C refers to the center of the reference coordinate system.

For a given pixel y_r in the slant-range SAR scene, Eq. (9) can be reformulated as

$$R = c\,t = c\,(t_0 + \frac{y_r}{2f_r}) = c\,t_0 + c\,\frac{y_r}{2f_r} = R_0 + \Gamma\,y_r, \tag{10}$$

where R is the slant-range of the target point, c is the velocity of light,

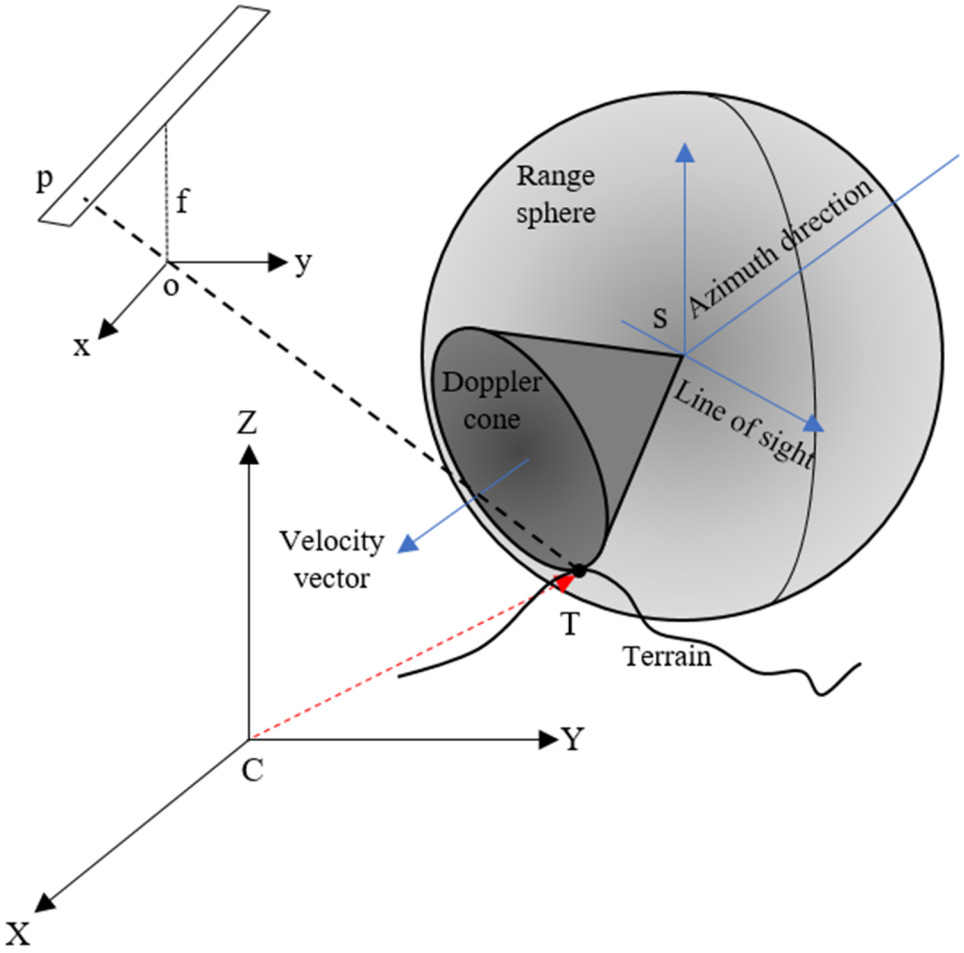


Fig. 4. Imaging geometry for configuration of SAR-optical imagery.

t_0, t are the one-way signal transmission times for the first range pixel and range pixel coordinate y_r, respectively, and f_r is the range sampling rate. R_0 gives the slant-range for the first range pixel and $\Gamma = \frac{c}{2f_r}$.

The second equation describes the geometry of the Doppler cone:

$$f_D = \frac{2}{\lambda_r R}\mathbf{V}\cdot(\mathbf{R}_{CT} - \mathbf{R}_{CS}), \tag{11}$$

where f_D is the Doppler frequency, λ_r is the SAR signal wavelength, $\mathbf{V}$ is the velocity vector and $\cdot$ denotes the inner product operator.

For the epipolarity constraint, we assume that the target point T is imaged at time τ by the push-broom sensor. If we fix the time variable t with τ and consider the corresponding image coordinate in the linear array coordinate system as $(0, y_l, f)$, we can use Eq. (8) to back-project from the linear array coordinate system to the terrestrial reference system as follows:

$$\begin{pmatrix} X - X^o \\ Y - Y^o \\ Z - Z^o \end{pmatrix} = \lambda^{-1} M_{\omega\phi\kappa} \begin{pmatrix} 0 \\ y_l \\ f \end{pmatrix} \tag{12}$$

where $M_{\omega\phi\kappa} = R^T_{\omega\phi\kappa}$. For imaging time t, all time-dependent parameters are estimated under the constraint $t = \tau$. Thus, for this specific instance, the variable index t is eliminated from Eq. (8). By expanding Eq. (12) and removing the scale factor effect, we have:

$$X - X^o = (Z - Z^o)\frac{m_{11}\,0 + m_{12}\,y_l + m_{13}\,f}{m_{31}\,0 + m_{32}\,y_l + m_{33}\,f} = (Z - Z^o)\frac{m_{12}\,y_l + m_{13}\,f}{m_{32}\,y_l + m_{33}\,f} \tag{13}$$

$$Y - Y^o = (Z - Z^o)\frac{m_{21}\,0 + m_{22}\,y_l + m_{23}\,f}{m_{31}\,0 + m_{32}\,y_l + m_{33}\,f} = (Z - Z^o)\frac{m_{22}\,y_l + m_{23}\,f}{m_{32}\,y_l + m_{33}\,f} \tag{14}$$

where m_{ij} are elements of matrix $M_{\omega\phi\kappa}$.

If the velocity vector of the SAR sensor is computed in the zero-Doppler frequency transition, we can reformulate Eq. (11) as:

$$V_x(t)(X - X^s(t)) + V_y(t)(Y - Y^s(t)) + V_z(t)(Z - Z^s(t)) = 0 \tag{15}$$

where $(V_x(t), V_y(t), V_z(t))$ are the components of velocity vector $\mathbf{V}$. Hence:

$$V_x(t)X - V_x(t)X^s(t) + V_y(t)Y - V_y(t)Y^s(t) + V_z(t)Z - V_z(t)Z^s(t) = 0 \tag{16}$$

From Eqs. (13) and (14), we can derive.

$$X - \left(\frac{m_{12}\,y_l + m_{13}\,f}{m_{32}\,y_l + m_{33}\,f}\right)Z = X^o - \left(\frac{m_{12}\,y_l + m_{13}\,f}{m_{32}\,y_l + m_{33}\,f}\right)Z^o \tag{17}$$

$$Y - \left(\frac{m_{22}\,y_l + m_{23}\,f}{m_{32}\,y_l + m_{33}\,f}\right)Z = Y^o - \left(\frac{m_{22}\,y_l + m_{23}\,f}{m_{32}\,y_l + m_{33}\,f}\right)Z^o \tag{18}$$

Multiplying both sides of the Eqs. (17) and (18) by $-V_x(t)$ and $-V_y(t)$, respectively, and then combining them with Eq. (16), Z can be calculated as follows:

$$Z = \frac{(m_{32}\,y_l + m_{33}\,f)\,[V_x(t)(X^s(t) - X^o) + V_y(t)(Y^s(t) - Y^o) + V_z(t)Z^s(t)]}{(m_{12}\,y_l + m_{13}\,f)\,V_x(t) + (m_{22}\,y_l + m_{23}\,f)\,V_y(t) + (m_{32}\,y_l + m_{33}\,f)\,V_z(t)} + \frac{[(m_{12}\,y_l + m_{13}\,f)\,V_x(t) + (m_{22}\,y_l + m_{23}\,f)\,V_y(t)]Z^o}{(m_{12}\,y_l + m_{13}\,f)\,V_x(t) + (m_{22}\,y_l + m_{23}\,f)\,V_y(t) + (m_{32}\,y_l + m_{33}\,f)\,V_z(t)} \tag{19}$$

Changing the position of target point T in the Z direction is equivalent to changing the corresponding image coordinates on the epipolar curve. Consequently, the determination of image coordinates in the SAR scene can be realized by tracking the sensor positions in the respective instances. In fact, in spite of fixing the position of the target point for the optical sensor at time τ, the location components $(X^s(t), Y^s(t), Z^s(t))$ and the velocity components of the SAR sensor $(V_x(t), V_y(t), V_z(t))$ are time-dependent and can be estimated for each instant using time-dependent polynomials. For the sake of simplicity, the SAR sensor trajectory can be approximated by a linear motion model. Thus, the velocity components $(V_x(t), V_y(t), V_z(t))$ will remain constant with time at $(X^s(t), Y^s(t), Z^s(t))$ i.e., the acceleration is 0 and the location components $(X^s(t), Y^s(t), Z^s(t))$ can be calculated using linear time-dependent functions:

$$\begin{aligned} X^s(t_a) &= X_0^s + V_x\,t_a \\ Y^s(t_a) &= Y_0^s + V_y\,t_a \\ Z^s(t_a) &= Z_0^s + V_z\,t_a \end{aligned} \tag{20}$$

where (X_0^s, Y_0^s, Z_0^s) is the position of the SAR sensor at initial time t_0. The time t_a is the azimuthal time, which can be expressed according to the line coordinate x_r in the SAR scene as

$$t_a = \frac{x_r}{PRF} = k\,x_r, \tag{21}$$

where PRF is the pulse repetition frequency in Hz.

Substituting the parameters expressed in Eqs. (20) and (21) into (19), we obtain:

$$Z = \frac{(m_{32}\,y_l + m_{33}\,f)\,[V_x(X_0^s - X^o) + V_y(Y_0^s - Y^o) + V_z Z_0^s]}{(m_{12}\,y_l + m_{13}\,f)\,V_x + (m_{22}\,y_l + m_{23}\,f)\,V_y + (m_{32}\,y_l + m_{33}\,f)\,V_z} + \frac{[(m_{12}\,y_l + m_{13}\,f)\,V_x + (m_{22}\,y_l + m_{23}\,f)\,V_y]Z^o}{(m_{12}\,y_l + m_{13}\,f)\,V_x + (m_{22}\,y_l + m_{23}\,f)\,V_y + (m_{32}\,y_l + m_{33}\,f)\,V_z} + k\,x_r \left(\frac{V_x^2 + V_y^2 + V_z^2}{(m_{12}\,y_l + m_{13}\,f)\,V_x + (m_{22}\,y_l + m_{23}\,f)\,V_y + (m_{32}\,y_l + m_{33}\,f)\,V_z}\right) \tag{22}$$

Eq. (22) can be simplified to:

$$Z = c_0 + c_1\,x_r \tag{23}$$

and substituting (23) into (17) and (18) gives:

$$\begin{aligned} X &= X^o - \left(\frac{m_{12}\,y_l + m_{13}\,f}{m_{32}\,y_l + m_{33}\,f}\right)Z^o + \left(\frac{m_{12}\,y_l + m_{13}\,f}{m_{32}\,y_l + m_{33}\,f}\right)(c_0 + c_1\,x_r) \\ &= X^o - \left(\frac{m_{12}\,y_l + m_{13}\,f}{m_{32}\,y_l + m_{33}\,f}\right)Z^o + \left(\frac{m_{12}\,y_l + m_{13}\,f}{m_{32}\,y_l + m_{33}\,f}\right)c_0 \\ &\quad + \left(\frac{m_{12}\,y_l + m_{13}\,f}{m_{32}\,y_l + m_{33}\,f}\right)c_1\,x_r \end{aligned} \tag{24}$$

$$\begin{aligned} Y &= Y^o - \left(\frac{m_{22}\,y_l + m_{23}\,f}{m_{32}\,y_l + m_{33}\,f}\right)Z^o + \left(\frac{m_{22}\,y_l + m_{23}\,f}{m_{32}\,y_l + m_{33}\,f}\right)(c_0 + c_1\,x_r) \\ &= Y^o - \left(\frac{m_{22}\,y_l + m_{23}\,f}{m_{32}\,y_l + m_{33}\,f}\right)Z^o + \left(\frac{m_{22}\,y_l + m_{23}\,f}{m_{32}\,y_l + m_{33}\,f}\right)c_0 \\ &\quad + \left(\frac{m_{22}\,y_l + m_{23}\,f}{m_{32}\,y_l + m_{33}\,f}\right)c_1\,x_r \end{aligned} \tag{25}$$

Eqs. (24) and (25) can be written in the form:

$$X = a_0 + a_1\,x_r \tag{26}$$

$$Y = b_0 + b_1\,x_r \tag{27}$$

The slant-range represented by Eq. (10) can be reformulated as:

$$(X - X^s)^2 + (Y - Y^s)^2 + (Z - Z^s)^2 = (R_0 + \Gamma\,y_r)^2 \tag{28}$$

Substituting (26), (27), (23) and (20) into (28), we have:

$$(a_0 + a_1\,x_r - X_0^s - V_x\,k\,x_r)^2 + (b_0 + b_1\,x_r - Y_0^s - V_y\,k\,x_r)^2 + (c_0 + c_1\,x_r - Z_0^s - V_z\,k\,x_r)^2 = (R_0 + \Gamma\,y_r)^2 \tag{29}$$

For simplicity, setting $A_0 = a_0 - X_0^s$, $A_1 = a_1 - V_x\,k$, $B_0 = b_0 - Y_0^s$, $B_1 = b_1 - V_y\,k$, $C_0 = c_0 - Z_0^s$, $C_1 = c_1 - V_z\,k$ gives:

$$(A_0 + A_1\,x_r)^2 + (B_0 + B_1\,x_r)^2 + (C_0 + C_1\,x_r)^2 = (R_0 + \Gamma\,y_r)^2 \tag{30}$$

which can be expanded to yield:

$$(A_0^2 + B_0^2 + C_0^2) + 2(A_0\,A_1 + B_0\,B_1 + C_0\,C_1)\,x_r + (A_1^2 + B_1^2 + C_1^2)\,x_r^2 = (R_0 + \Gamma\,y_r)^2 \tag{31}$$

If $A_0^2 + B_0^2 + C_0^2 = F_0$, $2(A_0A_1 + B_0B_1 + C_0C_1) = F_1$, and $A_1^2 + B_1^2 + C_1^2 = F_2$, this can be rewritten as:

$$\Gamma\,y_r = \sqrt{F_2\,x_r^2 + F_1\,x_r + F_0} - R_0 \tag{32}$$

Eq. (32) is a general rigorous model representing the epipolarity constraint for SAR-optical image pairs based on their imaging parameters contained in F_0, F_1, F_2, Γ and R_0. This shows that an epipolarity-like constraint can be established for SAR-optical image pairs. However, the non-linear relation between y_r and x_r in Eq. (32) shows that SAR-optical epipolar curves are not straight, even under the assumption of linear motion for the SAR system. In Section 3.3, this epipolarity constraint will be experimentally investigated for an RPC-based imaging model.

2.3. SAR-optical multi-sensor block adjustment

As illustrated in Fig. 1, the main step before implementing dense image matching is to align the optical image to the SAR image. This process is performed using a multi-sensor block adjustment which is based on RPCs instead of rigorous sensor models as proposed in Grodecki and Dial (2003). The block adjustment process improves the relative orientation between both images fixed to the more accurate SAR image orientation parameters. Through the block adjustment, the bias components induced by attitude, ephemeris, and drift errors in the optical image are compensated (d'Angelo and Reinartz, 2012).

The main bias compensation for the RPCs of the optical image involves translating the locations of the epipolar curves to accurate positions using the SAR geopositioning accuracy. Generally, designing an appropriate function for modeling the existing bias in the RPCs given by the optical image depends on the sensor properties (Tong et al., 2010), but for most sensors an affine model can be applied (Fraser and Hanley, 2005). Even for the current generation of VHR linear push-broom array sensors such as WorldView-2, employing only the shift parameters will be sufficient. The affine model for RPC bias compensation can be formulated as

$$\begin{aligned} \Delta x &= m_0 + m_1\,x_o + m_2\,y_o \\ \Delta y &= n_0 + n_1\,x_o + n_2\,y_o, \end{aligned} \tag{33}$$

where x_o, y_o represent column and row of tie points in the optical images and m_i and n_i $(i = 0, 1, 2)$ are unknown affine parameters to be estimated through the block adjustment procedure. Note that tie points are the common points between the SAR and optical images, and can be obtained by manual or automatic sparse matching between two images. Since the automation of the tie point generation process is not the focus of this study, we refer the reader to possible solutions described in Suri and Reinartz (2010), Perko et al. (2011), Merkle et al. (2017).

The geographic coordinates of the tie points in the SAR image are calculated by the inverse rational functions computed for the SAR

Fig. 5. Display the location of SAR-optical image pairs of the Munich study area. The red and blue rectangles identify areas covered by the WorldView-2 and TerraSAR-X images, respectively, and the black rectangle displays the study subset selected for stereogrammetrix 3D reconstruction over Munich. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

imagery as described in Section 2.1:

$$\lambda^i = f'_s(x_s^i, y_s^i, H) \tag{34}$$

and

$$\phi^i = g'_s(x_s^i, y_s^i, H) \tag{35}$$

where λ^i and ϕ^i are the normalized longitude and latitude of tie point i with normalized image coordinates x_s^i and y_s^i (index s denotes the SAR scene), and here, H is a constant, e.g., the mean height of the study area. f'_s and g'_s are inverse rational functions computed for the SAR sensor to project the tie points from the SAR image to the reference system. The output is a collection of GCPs that can be applied for the RPC rectification of the optical imagery. The resulting GCPs are then projected by the rational function associated with the optical images to give the image coordinates of the GCPs:

$$c_o^i = f_o(\lambda^i, \phi^i, H) \tag{36}$$

and

$$r_o^i = g_o(\lambda^i, \phi^i, H) \tag{37}$$

where c_o^i, r_o^i are the normalized image coordinates of tie point i computed by the forward rational functions of the optical sensor, f_o and g_o.

From Eqs. (33), (36), and (37), the primary equations for SAR-optical block adjustment are formed as:

$$x_o^i = c_{ou}^i + \Delta x^i + v_x^i \tag{38}$$

and

$$y_o^i = r_{ou}^i + \Delta y^i + v_y^i \tag{39}$$

where, x_o^i and y_o^i denote the column and row of tie point i in the optical scene, and c_{ou}^i and r_{ou}^i are the un-normalized coordinates of the tie point after projection and back-projection using the RPCs. The block adjustment equations can then be written as:

$$F_x^i = -x_o^i + c_{ou}^i + \Delta x^i + v_x^i = 0, \tag{40}$$

and

$$F_y^i = -y_o^i + r_{ou}^i + \Delta y^i + v_y^i = 0. \tag{41}$$

Finally, through an iterative least-squares adjustment (Grodecki and Dial, 2003), the unknown parameters m_i and n_i are estimated and the affine model can be formed. This affine model is added to the rational functions of the optical image to improve the geolocation accuracy to that of the SAR image.

2.4. SGM for dense multi-sensor image matching

The core step in a stereogrammetric 3D reconstruction workflow is the dense image matching algorithm to obtain the disparity map, which can then be transformed into the desired 3D point cloud. Generally, there are two different dense matching rationales that can be used according to whether local or global optimization is more important (Brown et al., 2003). For the case of global optimization, an energy functional consisting of two terms is established to find the optimal disparity map (Scharstein et al., 2001):

$$E(d) = E_{data}(d) + \lambda E_{smooth}(d) \tag{42}$$

where $E_{data}(d)$ is a fidelity term that makes the computed disparity map consistent with the input image pairs, $E_{smooth}(d)$ considers the smoothness condition for the disparity map, and λ is a regularization parameter that balances the fidelity and smoothness terms.

For a given image pair, the disparity map is calculated by minimizing the energy functional in (42). The main advantage of global dense matching over local matching methods is greater robustness against noise (Brown et al., 2003), although most existing algorithms for global dense image matching have a greater computational cost

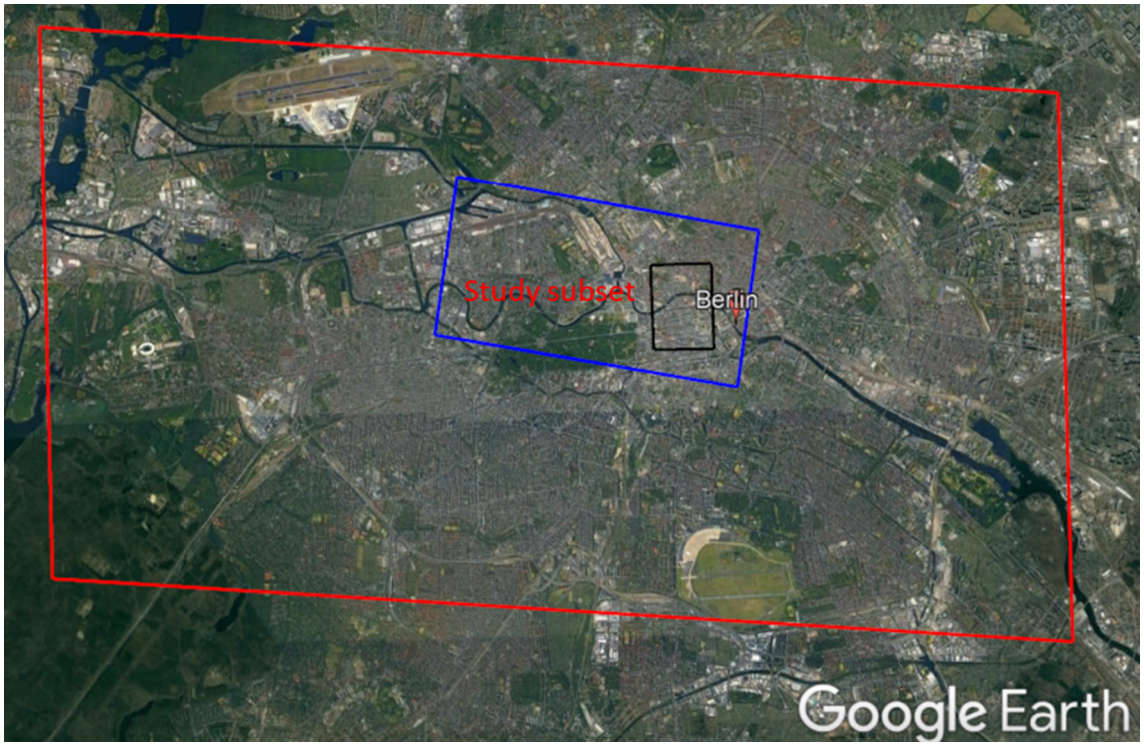


Fig. 6. Display the location of SAR-optical image pairs of the Berlin study area. The red and blue rectangles identify areas covered by the WorldView-2 and TerraSAR-X images, respectively, and the black rectangle displays the study subset selected for stereogrammetrix 3D reconstruction over Berlin. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Specifications of the TerraSAR-X and WorldView-2 images used for dense matching.

Area	Sensor	Acquisition Mode	Off-Nadir Angle (°)	Ground Pixel Spacing (m)	Acquisition date
Munich	TerraSAR-X	Spotlight	22.99	0.85 × 0.45	03.2015
	WorldView-2	Panchromatic	5.2	0.5 × 0.5	07.2010
Berlin	TerraSAR-X	Staring Spotlight	36.11	0.17 × 0.45	04.2016
	WorldView-2	Panchromatic	29.1	0.5 × 0.5	05.2013

(Hirschmüller, 2008).

For the experiments presented in this paper, we use the well-known SGM method (Hirschmüller, 2008), which offers acceptable computational cost and high efficiency, and performs very similarly to global dense image matching.

2.4.1. Cost functions used in SGM

In this study, the ability of performing dense image matching for SAR-optical image pairs using SGM is investigated. For this purpose, two different cost functions, namely Mutual Information (MI) and Census, as well as their weighted sum, are examined for the dense matching of high-resolution SAR and optical imagery. Typically, the similarity measures employed in the cost function are either signal-based or feature-based metrics (Hassaballah et al., 2016). Classically, signal-based similarity measures such as Normalized Cross Correlation (NCC) and MI are preferable to feature-based similarity measures when used in dense image matching algorithms because of faster calculation.

Among the signal-based matching measures, MI was recommended for SGM as it is known to perform well for images with complicated illumination relationship, such as SAR-optical image pairs (Suri and Reinartz, 2010).

Another similarity measure used in the SGM cost function is Census, which actually acts as a nonparametric transformation. The weighted sum of MI and Census is beneficial for 3D reconstruction in urban areas, especially for reconstructing the footprints of buildings to produce sharper and clearer images (Zhu et al., 2011). The weighted similarity measure can be defined as

$$SM = \alpha\, MI + (1 - \alpha)\,\text{Census}, \tag{43}$$

where α changes from 0 to 1 to weigh the effect of Census cost in relation to MI.

2.4.2. SGM settings for efficient SAR-optical dense matching

To increase the efficiency of the SGM performance, some important settings for the dense matching of SAR-optical image pairs must be considered. The basic principle of 3D reconstruction by dense matching is to use the epipolarity constraint to limit the search space. Usually, before dense matching, normal images are created by resampling the original images according to epipolar geometry (Morgan et al., 2004; Oh et al., 2010). In this study, we use the RPC model to realize the SAR-optical epipolar geometry implicitly without the need to generate normal images. This is done by implementing projections and back-projections from the reference image to the ground and back to the corresponding image, respectively, for a specified height range using rational functions. Then, the search for computing disparities can be performed along the thus-created epipolar curves.

In addition, the minimum and maximum disparity values should be selected to restrict the length of the search space along the epipolar

Fig. 7. Display of SAR-optical sub-scenes extracted from Munich study areas (the left-hand image is from WorldView-2, right-hand image is from TerraSAR-X).

Fig. 8. Display of SAR-optical sub-scenes extracted from Berlin study areas (the left-hand image is from WorldView-2, right-hand image is from TerraSAR-X).

curves. In general, there is more flexibility regarding the selection of this disparity interval for optical image pairs than for SAR-optical image pairs, and using unsuitable values will result in more outliers. The minimum and maximum disparity values can be determined using external data such as the SRTM digital elevation model, which is available for most land surfaces around the world (USGS, 2000). For sake of exploiting the simplicity offered by comparably flat study scenes, we just add and subtract 20-m height differences to the mean terrain heights of the study scenes to obtain the disparity thresholds.

The next setting is to switch off the *minimum region size* option in the

Table 2
Accuracy (standard deviation: STD) of RPCs fitted on SAR sensor model (units: m).

Area	Virtual GCPs		Check points	
	Row	Column	Row	Column
Munich	0.00026	0.00114	0.00025	0.00031
Berlin	0.00024	0.00027	0.00026	0.00118

SGM algorithm, which is usually used to decrease the noise level in the stereogrammetric 3D reconstruction of optical image pairs by eliminating isolated patches from the disparity map based on their small size. Experimental results show that, for SAR-optical image pairs, the complex illumination relationship between the images and the different imaging effects (especially for urban areas) make the *minimum region size* criterion useless, as connectivity cannot be ensured in the disparity map.

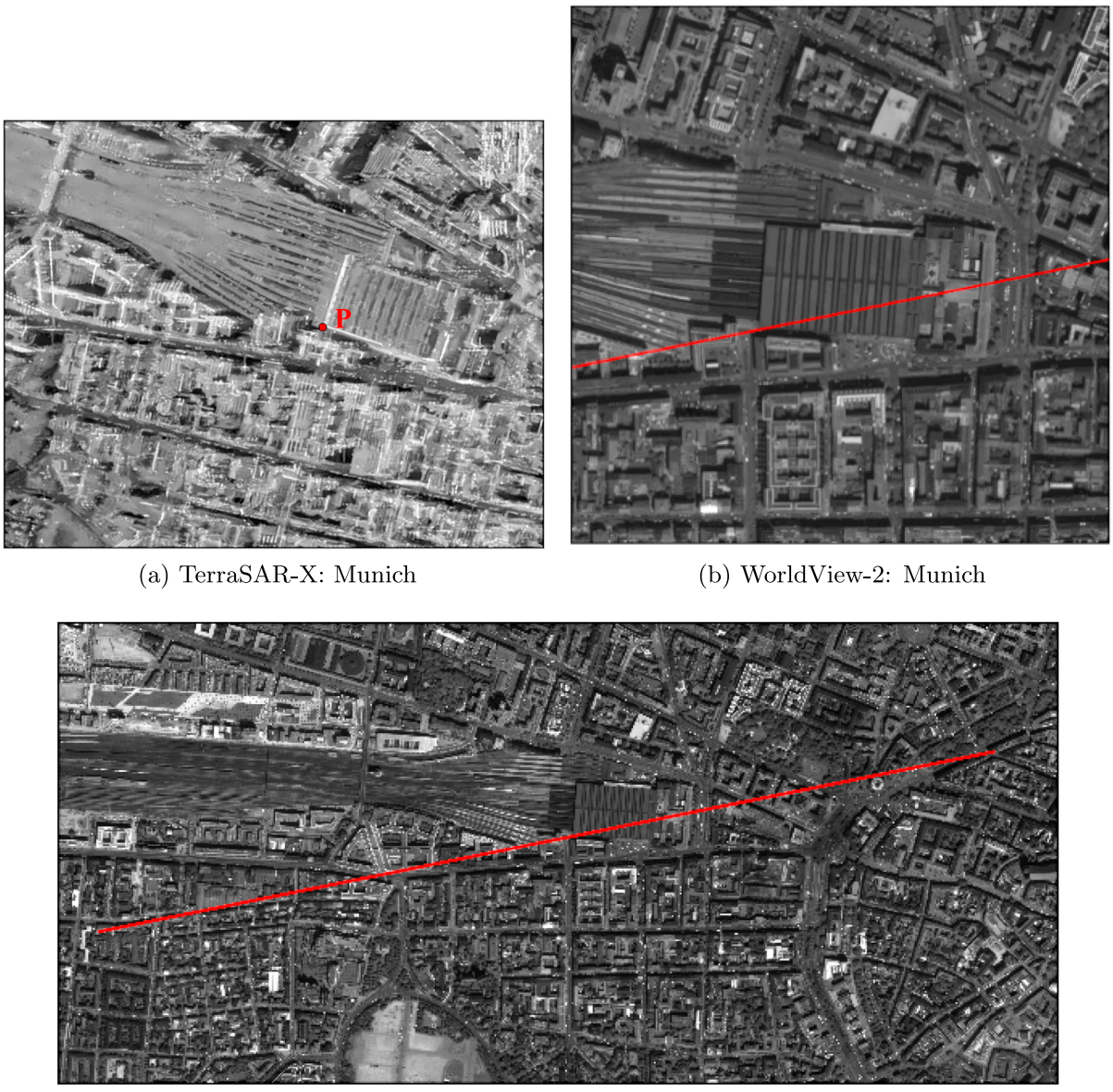
Similar to other dense matching cases, we use the LR (Left-Right) check to investigate binocular half-occlusions (Egnal and Wildes, 2002). This strategy changes the reference images from left to right, and consequently produces two disparity maps that can be checked against each other. To reach sub-pixel accuracy, the disparity in each point is estimated by a quadratic interpolation of neighboring disparities.

In this study, SGM is implemented at four hierarchy levels and the aggregated cost is calculated along 16 directions around each point.

3. Experiments and results

3.1. Study areas and datasets

We selected two study areas, one in Berlin and one in Munich (both located in Germany), to investigate the potential for 3D reconstruction from high-resolution SAR-optical image pairs over urban areas. The locations of TerraSAR-X and WorldView-2 images are displayed in Figs. 5 and 6. The properties of the image pairs for each study area are



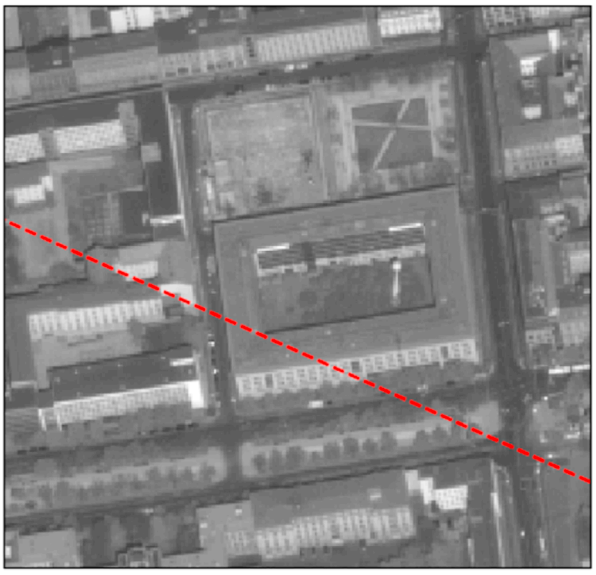
(a) TerraSAR-X: Munich

(b) WorldView-2: Munich

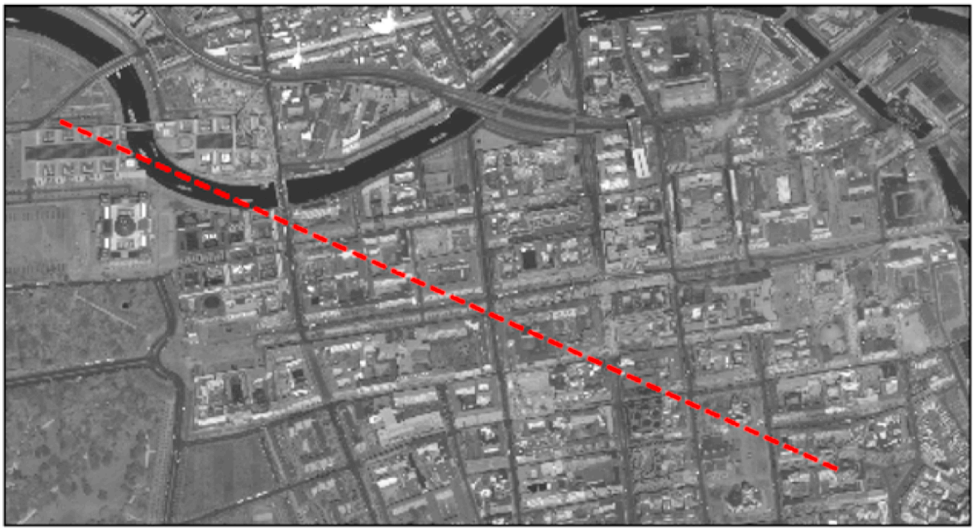
(c) WorldView-2: Munich

Fig. 9. Epipolar curves for the WorldView-2 image (of Munich) given by changing the heights of point p (located at the corner of the Munich central train station in the TerraSAR-X scene) for all possible height values in the image scene. The epipolar curves look like straight, but are not actually straight.

(a) TerraSAR-X: Berlin



(b) WorldView-2: Berlin



(c) WorldView-2: Berlin

Fig. 10. Epipolar curves for the WorldView-2 image (of Berlin) given by changing the heights of point p (a distinct corner of a building in the Berlin TerraSAR-X scene) for all possible height values in the image scene. The epipolar curves look like straight, but are not actually straight.

presented in Table 1. In order to enhance the general image similarity and facilitate the matching process, all images were resampled to 1 m × 1 m pixel spacing and the SAR images were filtered with a non-local speckle filter. After implementing bundle adjustment for both datasets, two sub-scenes (with a size of 1000 × 1500 pixels each) from overlapped parts of the study areas were cropped. These sub-scenes are displayed in Figs. 7 and 8.

3.2. Validation of RPCs to model SAR sensor geometry

As described in Section 2.1, RPCs can be used as a substitute for the rigorous range-Doppler model, similar to the standard RPCs delivered with optical imagery. This step is performed to simplify the multi-sensor block adjustment and epipolarity constraint construction. The accuracy of the RPCs can be estimated using independent virtual checkpoints that are produced in a similar way to VGCPs using the range-Doppler model. The word independent implies that the virtual checkpoints are never used in the RPC fitting, i.e., they are located in different positions respective to the VGCPs. The accuracy of the fitted RPCs for the TerraSAR-X data in each study area is listed in Table 2. Analysis was performed based on the residuals of the rows and columns, given by the differences $row_{RPC} - row_{DR}$ and $col_{RPC} - col_{DR}$, i.e., the differences between image coordinates computed by RPCs and range-Doppler. The analysis results confirm that the RPCs can model the range-Doppler geometry for TerraSAR-X data to within a millimeter, and can thus well be used in the 3D reconstruction process.

3.3. Validity of the epipolarity constraint

A general model that proves the epipolarity constraint for SAR-optical image pairs was described in Section 2.2. It was also concluded that epipolar curves are usually not straight. Experimentally, the epipolarity constraint for SAR-optical image pairs can also be modeled based on RPCs. In this paper, we evaluate the epipolarity constraint for

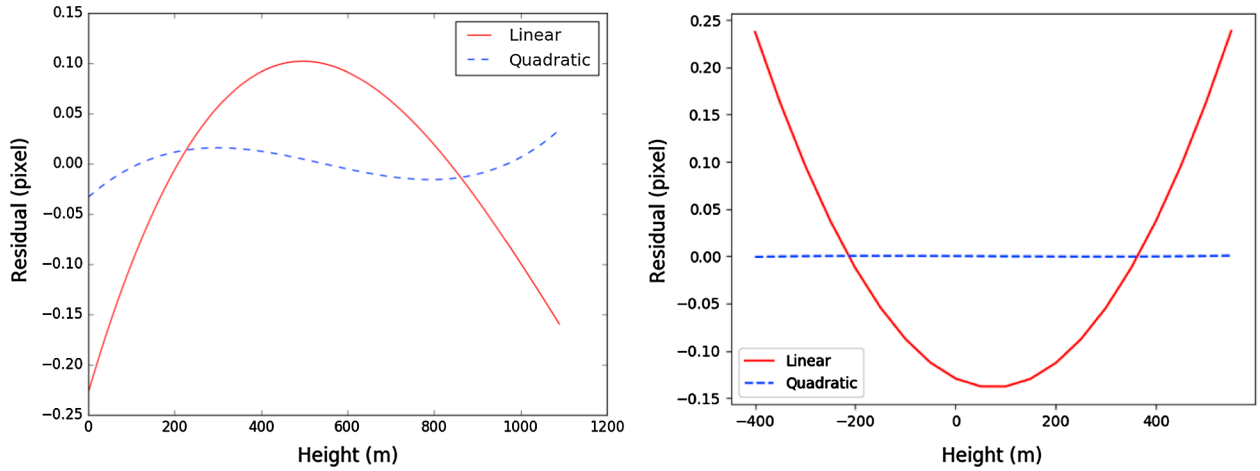


Fig. 11. Linear and quadratic polynomials fitted on the epipolar curves in the WorldView-2 images.

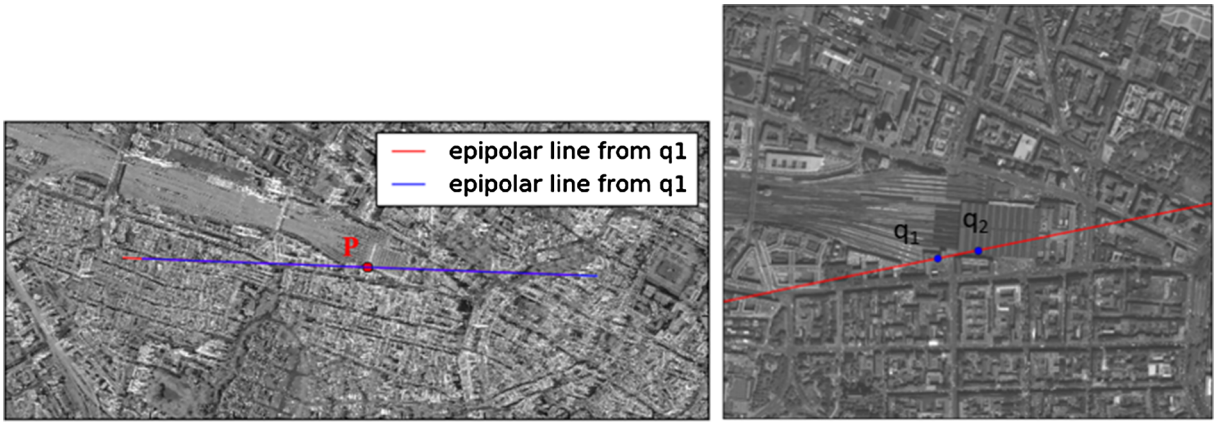


Fig. 12. Corresponding epipolar curves in the Munich TerraSAR-X image (left) derived from two points, q_1 and q_2 on the epipolar curve of the Munich WorldView-2 image (right).

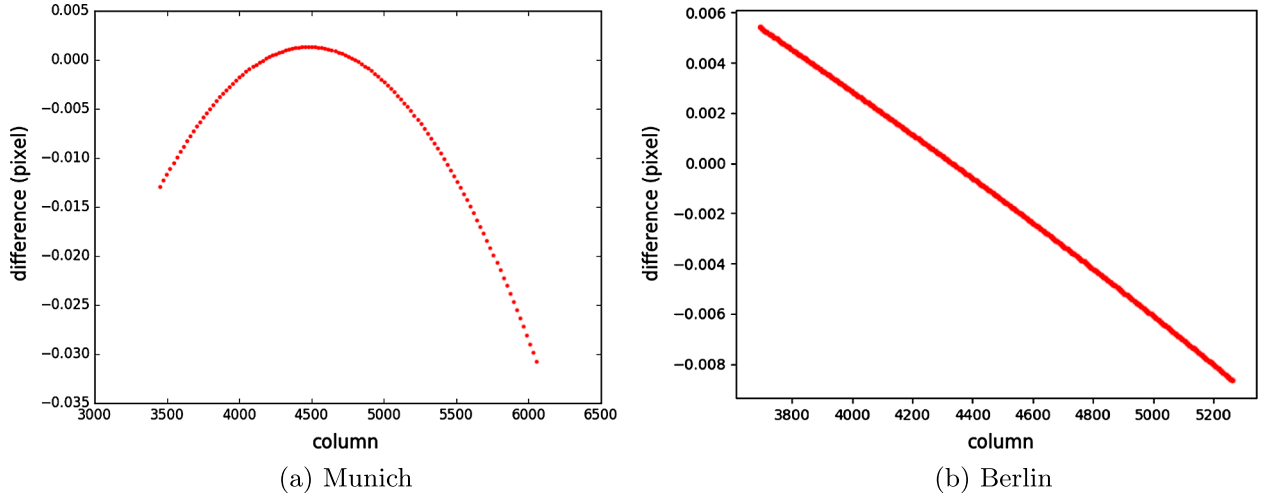


Fig. 13. Difference of two corresponding epipolar curves over the column direction. The maximum difference between the two epipolar curves is less than one pixel.

TerraSAR-X and WorldView-2 image pairs acquired over the two study areas.

3.3.1. Existence of epipolar curves

We analyzed the validity of the derived SAR-optical epipolarity constraint for an exemplary point located at the corner of the Munich central train station building (p). This point was projected to the terrain space by changing the heights in specific steps, e.g., 10 m, starting from the lowest possible height and proceeding to the highest possible height in the scene (for this experiment we used the interval [0 m, 1200 m]). The output will be an ensemble of points with different heights, such as depicted in Fig. 9(c). All these points were then back-projected to the

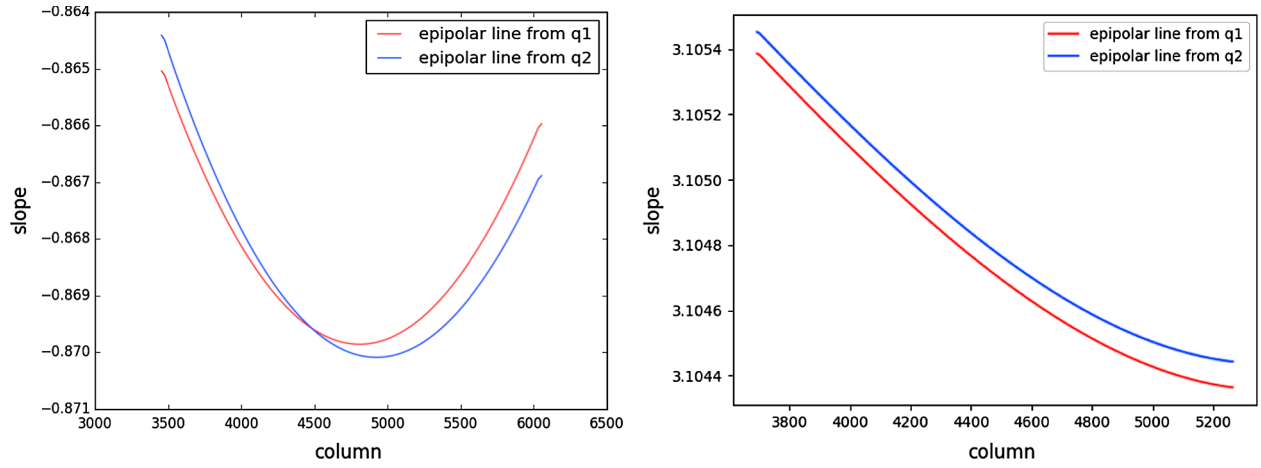


Fig. 14. Gradients of two epipolar curves constructed from q1 and q2 in TerraSAR-X.

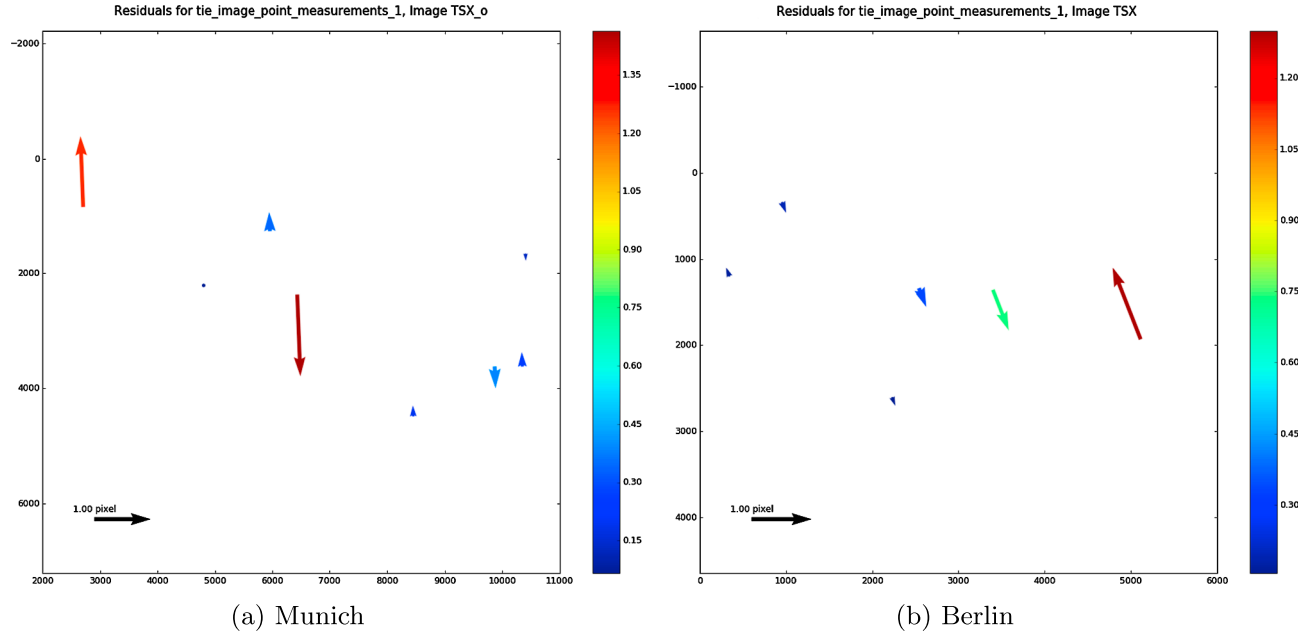


Fig. 15. Residuals of tie points after full multi-sensor block adjustment in the TerraSAR-X image space.

Table 3
Block adjustment results (units: m).

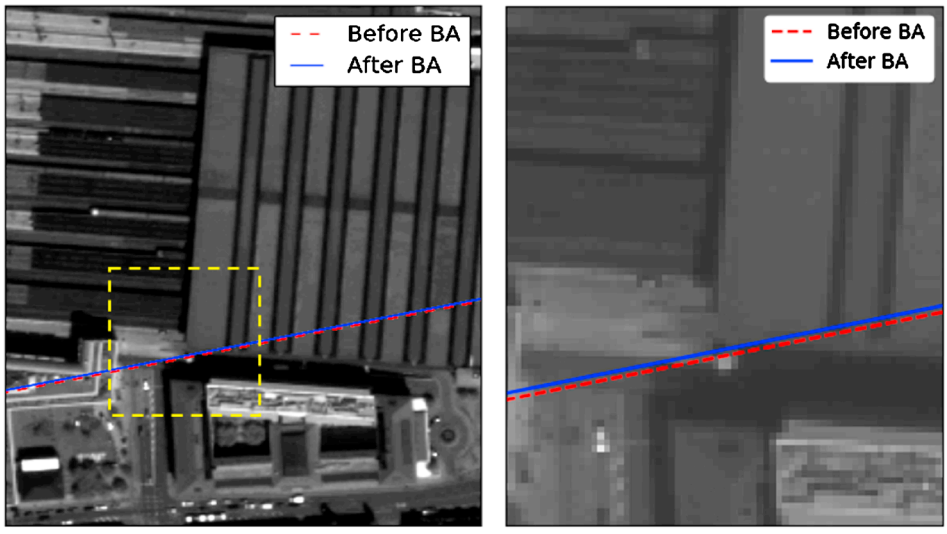
Area	Sensor	Bias Coefficients		STD	MAD	Min	Max	No. of Tie Points
Munich	WorldView-2	−2.47	−0.53	0.50	0.14	0.07	1.46	8
	TerraSAR-X	0	0	0.50	0.14	0.07	1.46	
Berlin	WorldView-2	−0.73	0.28	0.51	0.13	0.19	1.59	6
	TerraSAR-X	0	0	0.42	0.11	0.16	1.30	

WorldView-2 image space using RPCs. The corresponding epipolar curve for all possible heights in the study area is constructed by connecting the image points obtained in this way, as shown in Fig. 9(c). Although the epipolar curve appears to be straight, more analysis is required to determine whether this is the case. By expanding the image, it can be seen in Fig. 9(b) that the epipolar curve nearly passes through the conjugate point of p in the WorldView-2 image. Similarly, another experiment was carried out using the Berlin dataset. The epipolar curve was constructed for an exemplary point located on the corner of a building and for all possible heights in the scene. Fig. 10(d) displays the position of the selected point on the SAR image as well as the corresponding epipolar curve in the WorldView-2 imagery (Figs. 10(e) and (f)).

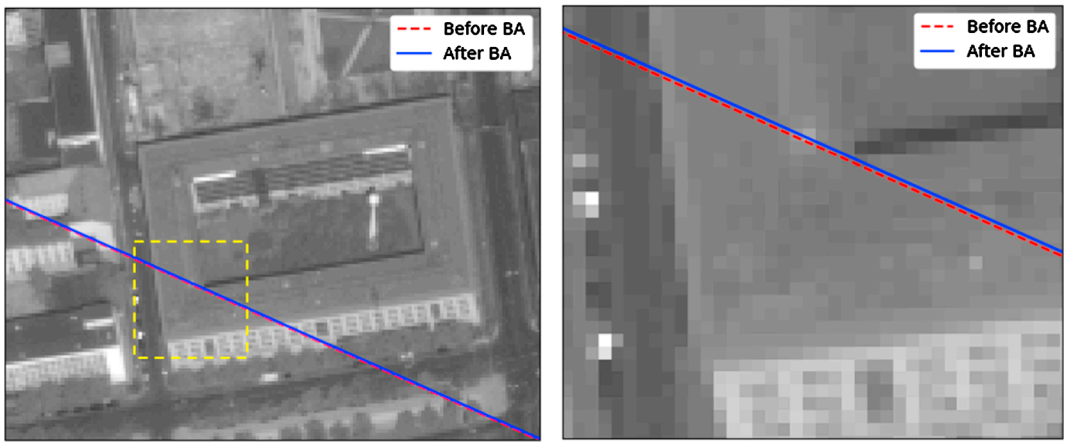
3.3.2. Straightness of epipolar curves

To clarify the straightness of the epipolar curve constructed for point p, linear and quadratic polynomials were fitted to the image points of the epipolar curve. Figs. 11(a) and (b) represent the least-squares residuals with respect to the point heights for the epipolar curves created in both study subsets. The residuals of the linear fit for the epipolar curve established for the Munich WorldView-2 image range from −0.25 to 0.1 pixels (i.e. meters), whereas the residuals of the quadratic fit are close to zero.

Similar results were given by fitting linear and quadratic polynomials to the image points of the epipolar curve established in the WorldView-2 image of Berlin. Fig. 11(b) clearly shows that the residuals of the epipolar points fitted to the quadratic model are zero, whereas those of the linear fit vary between −0.15 m and 0.25 m. Both analyses illustrate that the constructed epipolar curves are not straight.



(a) Munich



(b) Berlin

Fig. 16. Displacement of epipolar curves after block adjustment by RPCs. Left images show the epipolar curve positions before and after the bundle adjustment and right images display the selected patch (identified by dashed yellow rectangles) in an enlarged image. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Residuals (unit: m) of the control points for block adjustment validation. *Original* indicates the residuals before the adjustment, *modified* those after adjustment.

Area	RPCs	Mean	Max	RMSE	No. of Points
Munich	Original	1.301	2.359	1.364	31
	Modified	0.666	2.142	0.923	
Berlin	Original	0.775	1.736	0.866	32
	Modified	0.600	1.600	0.752	

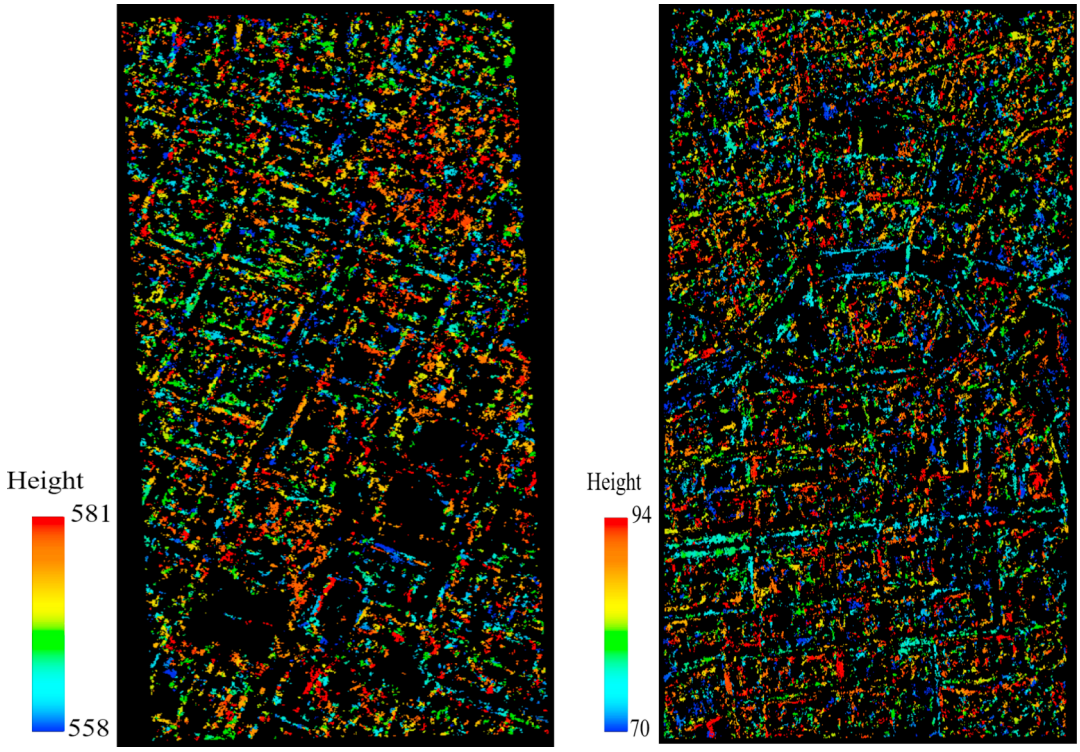
However, their curvatures do not exceed more than one pixel over the whole possible range of heights in these scenes.

3.3.3. Conjugacy of epipolar curves

To investigate the conjugacy of the SAR-optical epipolar curves, two distinct points (q_1, q_2) were selected from the epipolar curve in the WorldView-2 image. From each of these points, the corresponding epipolar curves were constructed in the TerraSAR-X image for all possible heights as in the experiments before. Fig. 12 displays the corresponding epipolar curves in TerraSAR-X given by q_1 and q_2 located in the WorldView-2 image. The epipolar curve appears to pass through point p located in the SAR image. Further analysis clarifies that the differences in the column direction between the two epipolar curves passing through point p are less than one pixel, allowing the matching of the two epipolar curves (Fig. 13(a)). In addition, Fig. 14(a) shows that the gradients of the two epipolar curves change at each point, as illustrated by the column index, whereas the maximum difference is less than 0.001, i.e., 0.1%. This indicates that the epipolar curves can be assumed to be parallel. The gradient changes at each point also confirm that the epipolar curves in TerraSAR-X are not perfectly straight.

In a similar manner, the conjugacy of the epipolar curves was evaluated for the Berlin dataset. Fig. 13(b) shows that the difference between the epipolar curves is less than one pixel, so these lines can be paired. Similarly, the maximum difference between the slopes of the epipolar curves is less than 0.2%, which confirms the possibility of epipolar curve conjugacy (Fig. 14(b)).

From the above investigations and discussions, it is clear that the epipolarity constraint can be established for SAR-optical image pairs such as those from TerraSAR-X and WorldView-2 data. As expected, the epipolar curves are not perfectly straight and there are tiny differences between the epipolar curve in one image produced from points on the epipolar curve in the other image. However, analyses show that the epipolar curves can be approximated as straight lines without

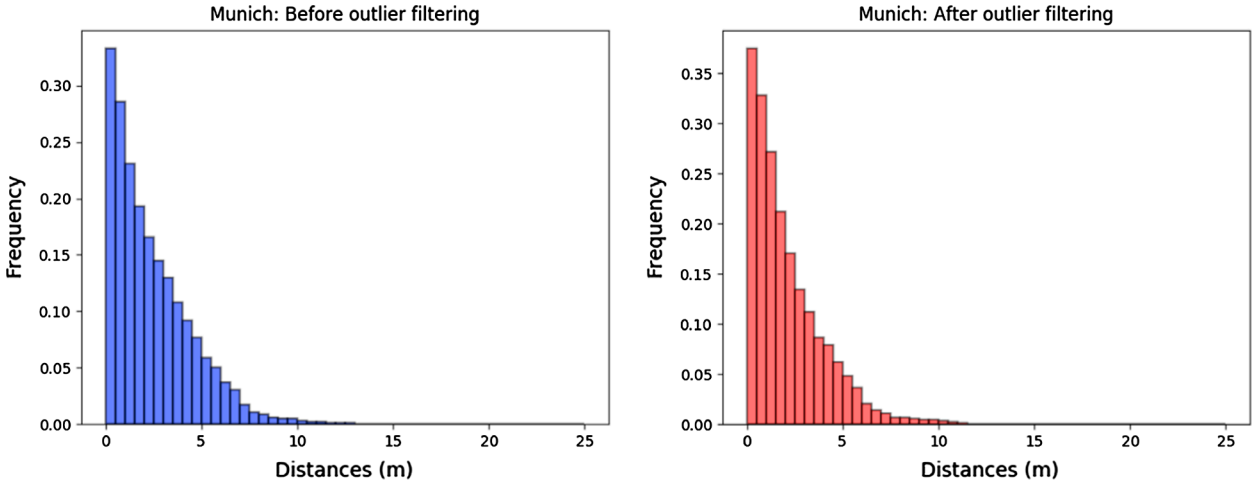


(a) Munich (b) Berlin

Fig. 17. Point clouds reconstructed from Munich and Berlin sub-scenes.

Table 5
Accuracy assessment of reconstructed point clouds with respect to LiDAR reference.

Area	Mean (m)			STD (m)			RMSE (m)		
	X	Y	Z	X	Y	Z	X	Y	Z
Munich	−0.003	0.025	0.080	1.285	1.350	2.652	1.285	1.351	2.653
Berlin	0.000	−0.041	0.273	1.566	1.692	3.091	1.566	1.693	3.103



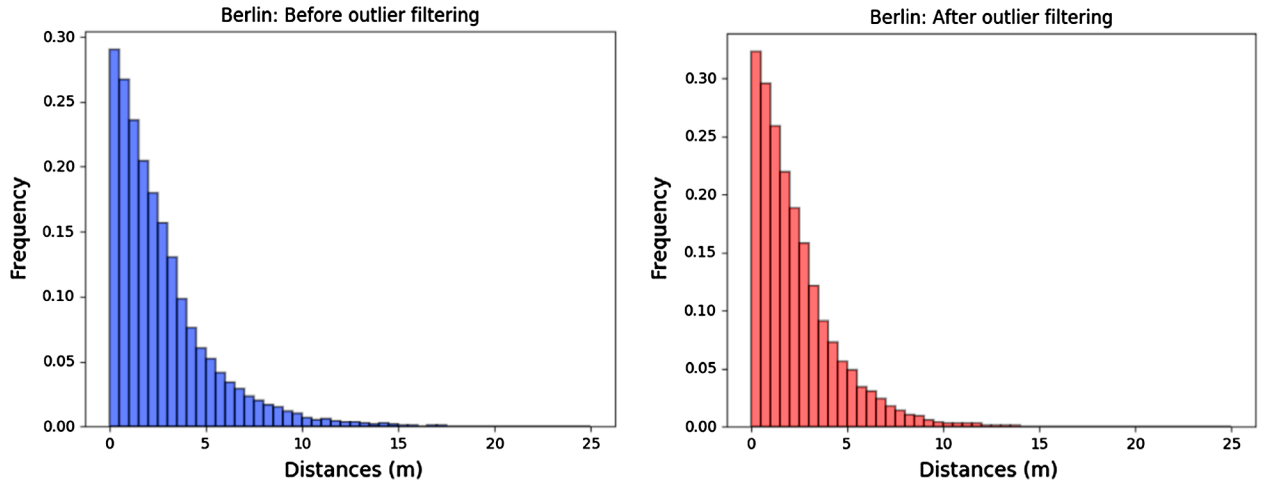
(a) before outlier removal (b) after outlier removal

Fig. 18. Euclidian distances between reconstructed points and reference planes for Munich.

sacrificing too much, and that they can be paired together well. This means that the epipolarity constraint can be used to ease the subsequent stereogrammetric matching process.

3.4. Use of block adjustment

As discussed in Section 3.3 and mathematically proved in Section 2.2, the epipolarity constraint can be established for a SAR-optical



(a) before outlier removal (b) after outlier removal

Fig. 19. Euclidian distances between reconstructed points and reference planes for Berlin.

Table 6
Accuracy assessment of point clouds after SRTM-based outlier removal.

Area	Point Cloud	25%-quantile	50%-quantile	75%-quantile	Mean (m)
Munich	original	0.77	1.89	3.58	2.44
	filtered	0.67	1.56	3.04	2.12
	SRTM	*0.73*	*1.64*	*3.25*	*2.21*
Berlin	original	0.89	2.01	3.67	2.75
	filtered	0.79	1.76	3.22	2.35
	SRTM	*0.86*	*1.93*	*3.63*	*2.65*

image pair. However, the positions of epipolar curves in the optical image can be placed in a more accurate position by exploiting the high geolocalization accuracy of the SAR image through a multi-sensor block adjustment. The experiments described in Section 3.3 demonstrate that, for the case of WorldView-2 imagery, the curvature of the epipolar curves does not exceed one pixel, and using only two bias terms as shifts in the column and row directions suffice to modify the position of the epipolar curves.

Implementing block adjustment using RPCs requires some conjugate points to be assigned as common tie points between the target (WorldView-2) and the reference (TerraSAR-X) images. Theoretically, just one tie point would be sufficient to estimate the bias in the least-squares adjustment based on Eq. (33) (two unknowns and two equations), but using more redundancy and incorporating more tie points allows for more accurate estimations of the bias parameter. For this experiment, eight and six tie points were selected to match the WorldView-2 images to the TerraSAR-X images in the Munich and Berlin study areas, respectively. The block adjustment equations were then established as described in Section 2.3. During the iterative least-squares adjustment, tie points with residuals exceeding a threshold were removed from the full adjustment process. Fig. 15(a) and (b) show the residuals of the full multi-sensor block adjustment for each tie point. The results demonstrate that the residuals of most points are less than one pixel in both experiments, which indicates a successful implementation of SAR-optical block adjustment. Table 3 presents the bias of the row and column components resulting from the block adjustment of WorldView-2 and TerraSAR-X image pairs for both study areas. As the SAR image was selected as the reference to which the optical imagery was aligned, the bias components for the reference SAR imagery are zero. The quality of block adjustment has been evaluated by calculating the positional errors of tie points under projection and re-projection from the SAR image to terrain and from terrain to the optical image, and vice versa. Statistical metrics such as the standard deviation (STD), median absolute deviation (MAD), and minimum (Min)/maximum (Max) errors were calculated. The results illustrate that the major bias component is in the row direction in both study areas and that the epipolar curves will mostly shift in this direction after block adjustment. Fig. 16(a) and (b) display the locations of the epipolar curves before and after adjustment. By enlarging the images, it is possible to confirm that the displacement of the epipolar curves is minimal, yet noticeable.

To verify the success of SAR-optical block adjustment, we require highly accurate GCPs, which are not available for the study areas. However, we evaluated the accuracy of block adjustment in sub-scenes of the two study areas with the assistance of available LiDAR point clouds. First, we manually found some matched points that had been measured in the SAR and optical sub-scenes, similar to the tie point selection step. Next, the measured points located in the optical imagery were projected to the terrain using the corresponding reverse rational polynomial functions ($f_o'(c, r, h)$ and $g_o'(c, r, h)$). To ensure exact back-projection, the height h of each point was extracted from the available high-resolution LiDAR point clouds of the target sub-scenes. To overcome the noise in the LiDAR data, we considered neighboring points around the selected measured point, and the final height of the target point was selected based on the mode of the heights in the considered neighborhood. The resulting ground points ($f_o'(c, r, h)$, $g_o'(c, r, h)$, h) were then back-projected to the SAR scene using the forward RPCs fitted to the SAR imagery. Finally, comparing the image coordinates of the measured points on the SAR imagery (from manual matching) with their coordinates derived by projection from the optical to the SAR imagery using RPCs and LiDAR data provides an evaluation of the SAR-optical block adjustment performance. The residuals can be calculated as:

$$\begin{aligned} dc &= f_s(f_o'(c, r, h), g_o'(c, r, h), h) - c_s^m \\ dr &= g_s(f_o'(c, r, h), g_o'(c, r, h), h) - r_s^m \end{aligned} \tag{44}$$

where (dc, dr) is the column and row difference between the measured point (c_s^m, r_s^m) located on the SAR imagery and the corresponding coordinates given by the projection from the optical to the SAR imagery using RPCs.

Table 4 presents some statistical analysis on residuals calculated according to Eq. (44) for two states: using the original WorldView-2 RPCs for the projections and using the WorldView-2 RPCs modified with respect to the TerraSAR-X SAR imagery. The results demonstrate the successful implementation of RPC-based multi-sensor block adjustment for SAR-optical image pairs. This means that the existing bias in the RPCs of optical imagery such as WorldView-2 can be modified

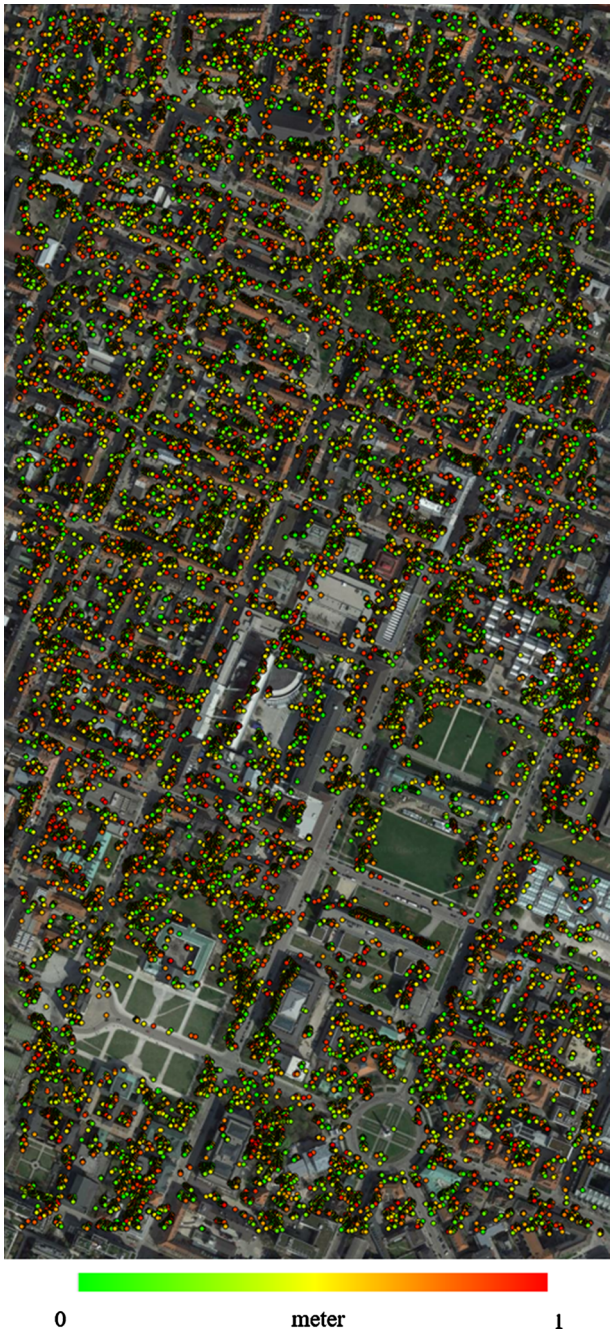


Fig. 20. Position of points with subpixel accuracy (1 m) achieved by dense matching over Munich city.

according to high-resolution SAR imagery such as TerraSAR-X to improve the absolute geolocalization accuracy of the optical imagery, and consequently the modification of epipolar curves in stereo cases.

3.5. Dense matching results

The output of dense matching by SGM is a disparity map that is calculated in the frame of the reference sensor geometry. This disparity map should be transferred from the reference sensor geometry to a terrestrial reference coordinate system such as UTM. The difference of SAR and optical observation geometries and the lack of jointly visible scene parts means that stereogrammetric 3D reconstruction leads to sparse rather than dense point clouds over urban areas. Fig. 17(a) and (b) display the reconstructed point clouds from SAR-optical sub-scenes of Munich and Berlin.

The accuracy of these sparse point clouds was compared to that of reference LiDAR point clouds with densities of 6 and 6.5 points per square meter acquired by airborne sensors over Munich and Berlin, respectively.

Different approaches can be used to assess the accuracy of point clouds. The simplest way is to calculate the Euclidean distance to the nearest-neighbor of each target point in the reference point cloud (Muja and Lowe, 2009). This strategy, however, should only be used when both point clouds are very dense. We therefore used another approach, which is based on fitting a plane to the k (here is 6) nearest neighbors of each target point in the reference point cloud (Mitra et al., 2004). The perpendicular distance from the target point to this plane is then the measured reconstruction error. To speed up the process of point cloud evaluation, an octree data structure is used for the binary partitioning of both reconstructed and reference point clouds (Schnabel and Klein, 2006). The measured distances between both point clouds are decomposed into three components that represent the accuracy of the reconstructed points along the X, Y, and Z directions. Table 5 summarizes the mean, STD, and Root Mean Square Error (RMSE) of the distances along the different axes.

In addition, histograms of the Euclidean distances between reconstructed points and k-nearest neighbors-based reference planes are depicted in Figs. 18 and 19, while the corresponding metrics are summarized in Table 6. In order to also provide an outlier-free accuracy assessment, we additionally show results corresponding to point clouds that were cleaned by removing points deviating from the SRTM model by more than 5 m.

Finally, Figs. 20 and 21 display high-accuracy points (i.e. those with a Euclidean distance of less than 1 m to the reference) achieved by SGM dense matching for TerraSAR-X- WorldView-2 image pairs in Munich and Berlin, respectively.

4. Discussion

4.1. Feasibility of SAR-optical stereogrammetry workflow

The results described in the previous section demonstrate the potential of the proposed SAR-optical stereogrammetry framework. Our analyses show that all primary steps involved in SAR-optical stereogrammetry, such as RPC fitting, epipolar-curve generation, and multi-sensor block adjustment, can be successfully implemented for VHR SAR-optical image pairs. In addition to the mathematical proof of the existence of an epipolarity constraint for arbitrary SAR-optical image pairs in Section 2.2, the experimental results have illustrated the validity of establishing an epipolarity constraint by showing that SAR-optical epipolar curves are approximately straight. Using RPCs for both sensor types paves the way for the implementation of stereogrammetry. As a result, estimating the RPCs for SAR imagery is a prerequisite for SAR-optical stereogrammetry. The RPCs delivered with optical imagery must be improved with respect to the SAR sensor geometry using RPC-based multi-sensor block adjustment. The block adjustment aligns pairs of SAR and optical images and improves the absolute geopositioning accuracy of the optical imagery. This ensures that the epipolar curves pass through the correct positions of conjugate points. Applying a dense matching algorithm such as SGM then produces a disparity map.

4.2. Potential and limitations of SAR-optical stereogrammetry

As discussed in Section 3.5, the dense matching of TerraSAR-X/WorldView-2 imagery produces a sparse point cloud over each of the urban study areas. However, the resulting point clouds are affected by a significant amount of noise because of the difficult radiometric and geometric relationships between the SAR and the optical images. Hence, the SGM algorithm struggles to find the exact conjugate points. On the one hand, this is related to the similarity measures employed in

Fig. 21. Position of points with subpixel accuracy (1 m) achieved by dense matching over Berlin city.

this prototypical study. The influence of similarity measures on the height accuracy of the Munich point cloud is shown in Fig. 22.

The RMSE of the estimated heights decreases when Census and MI are combined and used as a weighted sum cost function, although the number of outliers increases. Identifying the optimum weighting to balance the percentage of outliers against the height accuracy is impractical, because the output disparity maps are rather sparse; a visualization would not be helpful for this task. In stereogrammetric 3D reconstruction using optical image pairs, visualizing the disparity map enables the weight value to be tuned so as to preserve the edges and sharpness of building footprints, whereas in the SAR-optical case, there is no perfectly dense disparity map. Using Census alone produces points with higher accuracy than the MI-only results, but a higher percentage of outliers. In general, a similarity measure specifically designed for SAR-optical matching is required.

On the other hand, the reconstruction suffers from the fact that the SGM search strategy is designed for relatively simple isotropic geometric distortions and was not adapted to the peculiarities of SAR-optical matching yet. Therefore, the differences in the imaging geometries of SAR and optical sensors in terms of their off-nadir and horizontal

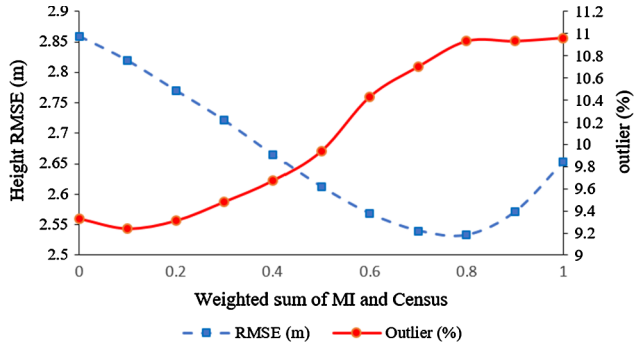


Fig. 22. Performance of MI, Census, and weighted sum of both measures as cost functions in SGM.

viewing angles can further decrease the matching accuracy. For example it is known from previous research that the optimal geometrical condition for SAR-optical stereogrammetry would be an image pair acquired with similar viewing geometries (Qiu et al., 2018). This, however, would make the geometrically induced dissimilarities in the images even larger and render the matching more complicated. If both sensors were at the same position, and thus would share the same viewing angle, the intersection geometry would be perfect (Qiu et al., 2018). However, due to the different imaging geometries, elevated objects would appear to collapse away from the sensor in the optical image, while they would appear to collapse towards the sensor in the SAR image. Thus, the choice of a good stereo geometry will always need to be a trade-of between image similarity and favorable intersection angle in the SAR-optical case.

Last, but not least, many points cannot be sensed by a nadir-looking optical sensor but are well observed by a side-looking SAR sensor, such as points located on building facades. As has been shown before, the joint visibility between SAR and optical VHR images of urban scenes can be as low as about 50% (Hughes et al., 2018). In the present study, the situation was most complicated in the Berlin case, because of differences in both the horizontal viewing directions and the off-nadir angles. The horizontal viewing direction of the WorldView-2 sensor was approximately north-south, whereas that of TerraSAR-X was east-west. This affected the visibility of common points between the two images during 3D reconstruction negatively. Consequently, most of the reconstructed points are located on the flat areas or outlines of buildings that are observed by both sensors (see Figs. 20 and 21).

Finally, some differences in the image pairs may be caused by the interval between the acquisition times of the WorldView-2 and TerraSAR-X data (5 and 3 years for Munich and Berlin, respectively). This can cause the matching process to fail in problematic areas, thus affecting the quality and density of the disparity maps.

In spite of the differences in sensor geometries, acquisition times, and illumination conditions between the two SAR-optical image pairs, the quantitative analysis demonstrated in Section 3.5 shows that 25% of all points are reconstructed with clear sub-pixel accuracy, while the median accuracy lies at about 1.5–2 m. The experiments also show that the results can be further improved by filtering outliers from the reconstructed point clouds. In this study, we employed the globally available SRTM DEM as prior knowledge for outlier removal. As Table 6 shows, discarding points with a height difference to SRTM greater than 5 m improves the results significantly.

Of course, this simple filtering strategy will probably also remove some accurate points that just deviate a lot from the SRTM DEM (e.g. newly built skyscrapers). In conclusion, a more sophisticated algorithm should be developed for removing noise and outliers from derived point clouds in the future. Nevertheless it can be confirmed that the SAR-optical stereo results have the potential to provide both higher accuracy and higher point density than the SRTM data, making SAR-optical stereogrammetry another possible means for 3D reconstruction in remote sensing.

5. Conclusion

In this study, we investigated the possibility of stereogrammetric 3D reconstruction from VHR SAR-optical image pairs by developing a full 3D reconstruction framework based on the classic photogrammetric workflow. First, we analyzed all prerequisites for this task. The main requirement for SAR-optical stereogrammetry is to establish an epipolarity constraint to reduce the search space of the matching process. We mathematically proved that the epipolarity constraint can be established for SAR-optical image pairs. Furthermore, experimental analysis demonstrated that the epipolarity constraint can be employed for SAR-optical image pairs such as those from TerraSAR-X/WorldView-2, and showed that the epipolar curves are sufficiently straight. Because of the limited accuracy of the RPCs delivered with optical data, the relative orientation between both images can be improved with respect to the more accurate SAR orientation parameters using multi-sensor block adjustment. This shifts the epipolar curves toward their correct positions. An SGM-based dense matching algorithm was implemented using the MI and Census similarity measures, as well as their weighted sum. The outputs were sparse point clouds with a median accuracy of about 1.5 to 2 m and the 25%-quantile of best points well in the sub-pixel accuracy domain. Finally, SRTM data were used to remove outliers from the point clouds. This improved the accuracy of the point clouds further. Overall, this study has demonstrated that a 3D reconstruction framework can be designed and implemented for SAR-optical image pairs over urban areas. Future work will have to focus on the development of similarity metrics specific to the multi-sensor matching problem, and on an adaption of the semi-global search strategy that accounts for the anisotropic geometric distortions between SAR and optical images.

Acknowledgments

The authors want to thank everyone, who has provided test data for this research: European Space Imaging for the WorldView-2 image, DLR for the TerraSAR-X images, the Bavarian Surveying Administration for the LiDAR reference data of Munich, and Land Berlin (EU EFRE project) for the LiDAR reference data of Berlin.

This work is jointly supported by the German Research Foundation (DFG) under grant SCHM 3322/1-1, the Helmholtz Association under the framework of the Young Investigators Group SiPEO (VH-NG-1018), and the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement No. ERC-2016-StG-714087, Acronym: So2Sat).

References

Bagheri, H., Schmitt, M., d'Angelo, P., Zhu, X.X., 2018. Exploring the applicability of semi-global matching for SAR-optical stereogrammetry of urban scenes. ISPRS Int. Arch. Photogram. Remote Sens. Spatial Inform. Sci. 42 (2), 43–48.

Bloom, A.L., Fielding, E.J., Fu, X.-Y., 1988. A demonstration of stereophotogrammetry with combined SIR-B and Landsat TM images. Int. J. Remote Sens. 9, 1023–1038.

Brown, M.Z., Burschka, D., Hager, G.D., 2003. Advances in computational stereo. IEEE Trans. Pattern Anal. Mach. Intell. 25, 993–1008.

Cho, W., Schenk, T., Madani, M., 1993. Resampling digital imagery to epipolar geometry. Int. Arch. Photogram. Remote Sens. 29, 404.

Curlander, J.C., 1982. Location of spaceborne SAR imagery. IEEE Trans. Geosci. Remote Sens. 22, 106–112.

Curlander, J.C., McDonough, R.N., 1991. Synthetic Aperture Radar, vol. 396 John Wiley & Sons, New York, NY, USA.

d'Angelo, P., Reinartz, P., 2012. DSM based orientation of large stereo satellite image blocks. ISPRS Int. Arch. Photogram. Remote Sens. Spatial Inform. Sci. 39 (B1), 209–214.

DigitalGlobe, 2018. Accuracy of WorldView products. < https://dg-cms-uploads-production.s3.amazonaws.com/uploads/document/file/38/DG_ACCURACY_WP_V3.pdf > (accessed 03.18).

Egnal, G., Wildes, R.P., 2002. Detecting binocular half-occlusions: empirical comparisons

of five approaches. IEEE Trans. Pattern Anal. Mach. Intell. 24, 1127–1133.
Eineder, M., Minet, C., Steigenberger, P., Cong, X., Fritz, T., 2011. Imaging geodesy–toward centimeter-level ranging accuracy with TerraSAR-X. IEEE Trans. Geosci. Remote Sens. 49, 661–671.
Fraser, C.S., Hanley, H.B., 2005. Bias-compensated RPCs for sensor orientation of high-resolution satellite imagery. Photogram. Eng. Remote Sens. 71, 909–915.
Fraser, C., Dial, G., Grodecki, J., 2006. Sensor orientation via RPCs. ISPRS J. Photogram. Remote Sens. 60, 182–194.
Grodecki, J., Dial, G., 2003. Block adjustment of high-resolution satellite images described by rational polynomials. Photogram. Eng. Remote Sens. 69, 59–68.
Grodecki, J., Dial, G., Lutes, J., 2004. Mathematical model for 3D feature extraction from multiple satellite images described by RPCs. In: ASPRS Annual Conference Proceedings, Denver, Colorado.
Gupta, R., Hartley, R.I., 1997. Linear pushbroom cameras. IEEE Trans. Pattern Anal. Mach. Intell. 19, 963–975.
Gutjahr, K., Perko, R., Raggam, J., Schardt, M., 2014. The epipolarity constraint in stereo-radargrammetric DEM generation. IEEE Trans. Geosci. Remote Sens. 52, 5014–5022.
Hartley, R.I., Zisserman, A., 2004. Multiple View Geometry in Computer Vision, second ed. Cambridge University Press ISBN: 0521540518.
Hassaballah, M., Abdelmgeid, A.A., Alshazly, H.A., 2016. Image features detection, description and matching. In: Awad, A.I., Hassaballah, M. (Eds.), Image Feature Detectors and Descriptors: Foundations and Applications. Springer International Publishing, pp. 11–45.
Hirschmüller, H., 2008. Stereo processing by semiglobal matching and mutual information. IEEE Trans. Pattern Anal. Mach. Intell. 30, 328–341.
Hughes, L.H., Auer, S., Schmitt, M., 2018. Investigation of joint visibility between SAR and optical images of urban environments. ISPRS Ann. Photogram. Remote Sens. Spatial Inform. Sci. 4, 129–136.
Kim, T., 2000. A study on the epipolarity of linear pushbroom images. Photogram. Eng. Remote Sens. 66, 961–966.
Kratky, V., 1989. Rigorous photogrammetric processing of SPOT images at CCM Canada. ISPRS J. Photogram. Remote Sens. 44, 53–71.
Li, D., Zhang, Y., 2013. A rigorous SAR epipolar geometry modeling and application to 3D target reconstruction. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 6, 2316–2323.
Li, R., Zhou, F., Niu, X., Di, K., 2007. Integration of Ikonos and QuickBird imagery for geopositioning accuracy analysis. Photogram. Eng. Remote Sens. 73, 1067.
Merkle, N., Luo, W., Auer, S., Müller, R., Urtasun, R., 2017. Exploiting deep matching and SAR data for the geo-localization accuracy improvement of optical satellite images. Remote Sens. 9, 586.
Mitra, N.J., Nguyen, A., Guibas, L., 2004. Estimating surface normals in noisy point cloud data. Int. J. Comput. Geom. Appl. 14, 261–276.
Morgan, M., Kim, K., Jeong, S., Habib, A., 2004. Epipolar geometry of linear array scanners moving with constant velocity and constant attitude. ISPRS Int. Arch. Photogram. Remote Sens. Spatial Inform. Sci. 35, 508–513.
Morgan, M.F., El-Sheimy, N., Saalfeld, A., 2004. Epipolar resampling of linear array scanner scenes (PhD dissertation). University of Calgary, Department of Geomatics Engineering.
Muja, M., Lowe, D.G., 2009. Fast approximate nearest neighbors with automatic algorithm configuration. In: VISAPP International Conference on Computer Vision Theory and Applications, pp. 331–340.
Oh, J., Lee, W.H., Toth, C.K., Grejner-Brzezinska, D.A., Lee, C., 2010. A piecewise approach to epipolar resampling of pushbroom satellite images based on RPC. Photogram. Eng. Remote Sens. 76, 1353–1363.
Orun, A.B., 1994. A modified bundle adjustment software for SPOT imagery and photography: tradeoff. Photogram. Eng. Remote Sens. 60, 1431–1437.
Perko, R., Raggam, H., Gutjahr, K., Schardt, M., 2011. Using worldwide available TerraSAR-X data to calibrate the geo-location accuracy of optical sensors. In: 2011 IEEE International Geoscience and Remote Sensing Symposium, pp. 2551–2554.
Qiu, C., Schmitt, M., Zhu, X.X., 2018. Towards automatic SAR-optical stereogrammetry over urban areas using very high resolution imagery. ISPRS J. Photogram. Remote Sens. 138, 218–231.
Raggam, J., Almer, A., Strobl, D., 1994. A combination of SAR and optical line scanner imagery for stereoscopic extraction of 3D data. ISPRS J. Photogram. Remote Sens. 49, 11–21.
Scharstein, D., Szeliski, R., Zabih, R., 2001. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In: Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV 2001), pp. 131–140.
Schmitt, M., Zhu, X.X., 2016. Data fusion and remote sensing: an ever-growing relationship. IEEE Geosci. Remote Sens. Mag. 4, 6–23.
Schnabel, R., Klein, R., 2006. Octree-based point-cloud compression. In: Proceedings of the 3rd Eurographics/IEEE VGTC Conference on Point-Based Graphics, SPBG'06. Eurographics Association, Aire-la-Ville, Switzerland, pp. 111–121.
Suri, S., Reinartz, P., 2010. Mutual-information-based registration of TerraSAR-X and Ikonos imagery in urban areas. IEEE Trans. Geosci. Remote Sens. 48, 939–949.
Tao, C., Hu, Y., 2001. Use of the rational function model for image rectification. Can. J. Remote Sens. 27, 593–602.
Tao, C.V., Hu, Y., 2001. A comprehensive study of the rational function model for photogrammetric processing. Photogram. Eng. Remote Sens. 67, 1347–1358.
Tao, C.V., Hu, Y., 2002. 3D reconstruction methods based on the rational function model. Photogram. Eng. Remote Sens. 68, 705–714.
Tao, C.V., Hu, Y., Jiang, W., 2004. Photogrammetric exploitation of Ikonos imagery for mapping applications. Int. J. Remote Sens. 25, 2833–2853.
Tikhonov, A., Arsenin, V.Y., 1977. Methods for Solving Ill-posed Problems. John Wiley and Sons, Inc.
Tong, X., Liu, S., Weng, Q., 2010. Bias-corrected rational polynomial coefficients for high accuracy geo-positioning of QuickBird stereo imagery. ISPRS J. Photogram. Remote Sens. 65, 218–226.
Toutin, T., 2006. Comparison of 3D physical and empirical models for generating DSMs from stereo HR images. Photogram. Eng. Remote Sens. 72, 597–604.
USGS, 2000. Shuttle Radar Topography Mission (SRTM) void filled. < https://lta.cr.usgs.gov/SRTMVF > (accessed 09.17).
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 Trans. Geosci. Remote Sens. 52, 5840–5854.
Zhang, L., He, X., Balz, T., Wei, X., Liao, M., 2011. Rational function modeling for spaceborne SAR datasets. ISPRS J. Photogram. Remote Sens. 66, 133–145.
Zhu, K., d'Angelo, P., Butenuth, M., 2011. A performance study on different stereo matching costs using airborne image sequences and satellite images. In: Stilla, U., Rottensteiner, F., Mayer, H., Jutzi, B., Butenuth, M. (Eds.), Photogrammetric Image Analysis. Springer, Berlin, Heidelberg, pp. 159–170.