

Bound-Constrained Multiple-Image Least-Squares Matching for Multiple-Resolution Images

Han Hu and Bo Wu

Abstract

Satellite images from multiple sources with different resolutions are currently able to observe the same region. Reliable image matching between these images is the first step in their integrated use. Image matching of multiple-resolution images is not trivial because of the large geometric differences among the images, which can cause failure of matching and losses of matching accuracy. This paper presents a bound-constrained, multiple-image, least-squares matching (LSM) method that extends the classical LSM in two ways for better performance. First, the *a priori* metadata of the images, including the georeferencing and scale information, are used for initial matching and to provide bound constraints in the LSM to improve its stability. Second, multiple images are matched in a single optimization rather than the traditional pairwise matching. This brings additional observations in the least-squares optimization, which makes the matching aware of both larger and local context and improves matching quality even with inaccurate initializations for high resolution images. Experimental analysis using multiple-source satellite images with multiple resolutions collected on Mars and in Hong Kong reveals that the proposed method is capable of obtaining reliable multiple-fold matches effectively, even in challenging cases with resolution differences as much as 20-fold. The method proposed in this paper has significance for the synergistic use of multiple-source satellite images in various applications.

Introduction

Recent advances in satellite imaging have provided great opportunities for the same region to be covered by images from multiple sources with multiple resolutions, both on Earth (Elaksher and Alharthy, 2011; Tang *et al.*, 2016) and on other planets (Wu *et al.*, 2014; Tao *et al.*, 2016). These images can be integrated for synergistic use in a variety of applications. For example, images with sub-meter-level resolution are suitable for detailed topographic mapping (Qiao *et al.*, 2010) or hazard monitoring (Corbane *et al.*, 2011; Voigt *et al.*, 2011), and images with relatively coarse resolution feature wide coverage and are suitable for topographic mapping of large areas (Xue *et al.*, 2015; Zhang *et al.*, 2015). However, geometric inconsistencies at different levels among these images are inevitable because they were collected by different sensors with different characteristics onboard different platforms. These multiple-source images must be co-registered before their synergistic use, and reliable image matching between them is a key step in this process.

Han Hu is with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong; and Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, P.R. China.

Bo Wu is with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University (bo.wu@polyu.edu.hk).

Despite the milestone image-matching algorithm of SIFT (Scale Invariant Feature Transform) (Lowe, 2004), matching of multiple-resolution images is still an open problem that is actively being studied (Sedaghat and Ebadi, 2015). The problems are caused by the scale differences between the images, which have two consequences: (a) a failure to find matches, which is limited by the performance of the existing feature detectors and descriptors; and (b) a loss of matching accuracy, which is introduced in the sampling steps when matching images with different scales. A greater difference in scale between the images leads to more serious problems.

Lerma *et al.* (2013) presented a general workflow for multiple-resolution image matching that comprised two steps. The first step is initial matching between the images using SIFT-like descriptors, and the second is to locate initial matches at the sub-pixel level through the well-established least-squares matching (LSM) method (Gruen, 1985). Theoretically, LSM is fully invariant to affine deformations, which is a superset of the scale difference; however, in practice, LSM sometimes suffers from convergent issues when scale difference is larger than 30 percent. Furthermore, the nonlinear nature of its solver means that it requires similar image quality or the solver will diverge (Gruen, 2012). The two drawbacks of LSM have limited its use in the scenario of matching of images with multiple resolutions, in which the scale differences of the images may be as much as 20-fold.

To surmount the widely acknowledged disadvantages of LSM described above for reliable matching of multiple-resolution images, this paper presents a bound-constrained multiple-image LSM method that renovates the traditional LSM in two ways: (a) following previous work on improving the stability of LSM using bound constraints (Hu *et al.*, 2016), with the information of image resolutions, we extend the bound constrained solver into multiple image matching by directly matching in the original image space rather than rectified images. The bound constraints enforce the affine parameters not to exceed a certain range with physical significance and thus increase the reliability of the iterative solver; (b) Instead of traditional pairwise formulation of the LSM, this method combines all of the images (typically four images of two stereo pairs with different resolutions) into a single optimization. Using carefully designed parameter sets and the minimization formulation of LSM, we achieve both matching reliability between images of different resolutions and accuracies between images of the high resolution images at the same time.

The remainder of this paper is organized as follows. The next section presents a literature review of matching between images collected by different sensors and with multiple resolutions; followed by the revisits of the classical LSM and the bound-constrained multiple-image. Next, LSM is then

Photogrammetric Engineering & Remote Sensing
Vol. 83, No. 10, October 2017, pp. 667–677
0099-1112/17/667-677

© 2017 American Society for Photogrammetry
and Remote Sensing
doi: 10.14358/PERS.83.10.667

presented for multiple-resolution image matching. The experimental results and analyses are presented, and a summary with concluding remarks completes the paper.

Related Works

Matching of features on different images is a fundamental topic in both the photogrammetry and computer vision communities, and is generally decomposed into two separate steps: feature detection and feature description (Szeliski, 2011). For example, the classical corner detectors, such as Moravec (Moravec, 1981), Förstner (Förstner and Gülich, 1987), and Harris (Harris and Stephens, 1988), are used for feature detection, and the raw pixel intensities in a template window are used for description and are then compared using normalized cross correlation (NCC) to find possible correspondences. The major consideration for the detection step is repeatability (Remondino, 2006; Zhu *et al.*, 2007), which is the ability to recall the same feature under various geometric and radiometric transformations and possible noise. From this nature, most feature detectors can be classified into two categories: corner detectors that detect by the cross section of contours (Tuytelaars and Mikolajczyk, 2008; Rosten *et al.*, 2010) and blob-like detectors that exploit the salient regions in the image (Lowe, 2004; Mikolajczyk *et al.*, 2005; Bay *et al.*, 2008). In the subsequent matching step, the main concern is the distinctiveness of features (Carneiro and Jepson, 2005; Mikolajczyk and Schmid, 2005), which requires that the best match can be distinguished from the others by a large margin. This clue can also be adopted to find possible matches with the widely adopted ratio match (Lowe, 2004), as used in SIFT.

However, the classical corner detectors (Harris and Stephens, 1988) and template window matches are sensitive to any geometric differences, such as scale differences and projective deformations (Szeliski, 2011). These geometric problems are generally handled in the detection step. The scale space theory was first proposed by Lindeberg (1993) using the Laplacian of Gaussian (LoG); it was extended to detect features that are stable in the scale space (Lindeberg, 1998) and also approximated using Difference of Gaussian (DoG) to accelerate the computation (Lowe, 2004). Both LoG and DoG are exploited in the linear scale space, and the same notion can also be exploited using nonlinear scale construction (Alcantarilla *et al.*, 2012), which turns out to be more repeatable. The projective deformation is caused by the view point difference and can be approximated by the in-plane rotation or affine transformation. The rotation model requires only one parameter, i.e., the rotation angle; thus, the dominant orientation is detected by the gradient information and used to warp the local patch. This strategy is independent of the detection step and is widely used in many feature-matching methods (Lowe, 2004; Bay *et al.*, 2008; Tola *et al.*, 2010; Rublee *et al.*, 2011; Alcantarilla *et al.*, 2012; Ma *et al.*, 2015). The affine model should be approximated by the elliptical model. This means that the detector must obtain an elliptical region (Mikolajczyk and Schmid, 2002; Matas *et al.*, 2004), rather than the general point with scale, so the affine invariant is dependent on the detectors and will sometimes decrease the repeatability (Lowe, 2004; Ma *et al.*, 2015).

The above detectors retrieve repeatable points or regions across different images and can handle the geometric problem. To match the detected features, descriptors must be extracted. Raw pixel intensities are the most straightforward descriptors, and through the multiple-scale oriented patch scheme (Brown *et al.*, 2005); they can be used to handle scale and rotation problems. However, the intensities are generally sensible to noise, and features derived from the local patch are commonly used in subsequent works. Histogram of Oriented Gradients (HOG) (Dalal and Triggs, 2005;

Chandrasekhar *et al.*, 2009) is one of the most widely adopted feature transformation methods (Lowe, 2004; Mikolajczyk and Schmid, 2005; Bay *et al.*, 2008; Tola *et al.*, 2010). HOG is based on gradients of the local patch and thus is invariant to linear intensity variations. Furthermore, HOG uses the accumulated information in a local patch, so it is not sensitive to slight shifts caused by noise (Lowe, 2004). However, as described above, HOG-based descriptors are only invariant to linear intensity variations, and the performance will decrease for more complex illumination changes, such as specular reflections and exposures (Wang *et al.*, 2016).

The methods mentioned above generally do not address the sub-pixel issue. To locate the correspondence at the sub-pixel level, LSM (Gruen, 1985) is the standard approach. However, classical LSM may suffer from problems with reaching convergent solutions. To address this problem, several modifications have been made to the classical LSM. For example, Gruen and Baltsavias (1988) used epipolar information from multiple images, which constrains the image correlation to a local region around a single point intersected from multiple epipolar lines. However, their method requires known orientation parameters of the images, which is not as versatile as general image matching. Remondino (2006) used elliptical geometric information derived from affine invariant detectors in the LSM optimization as weighted observations to increase the stability of the nonlinear optimization. Yang *et al.* (2014) used information from SIFT detectors to establish homography transformation for establishing least squares optimization similar to LSM for improved affine adaption. However, the above method requires specific affine detectors, which have been proven to be inferior to other general detectors (Lowe, 2004; Rosten *et al.*, 2010).

With respect to image matching of multiple images, there are generally two strategies and both constrained by epipolar geometry. The first is directly matching in the image space and the same affine parameter sets are optimized against a template image (Gruen and Baltsavias, 1988). The other strategy warped the images onto the putative object space and rectified the images into the same image size. The matching is then transformed into searching a patch in the object space, which achieved the best matching quality among all the images. To the best of our knowledge, the first work that adopt this strategy is the GC³ (Geometrically Constrained Cross Correlation) (Zhang, 2005), which has been exploited for frame camera in both close-range (Barazzetti *et al.*, 2010; Furukawa and Ponce, 2010) and aerial photogrammetry (Jiang, 2006) and further for general case with RPCs (Zhang *et al.*, 2011; Ling *et al.*, 2016). Despite the robustness of the above multiple image matching strategy, they required known exterior orientation parameters to geometrically constrain the search space and furthermore, because only one template is used for all references, no results on images with large resolution difference, e.g., up to 20 folds in this paper, are reported previously.

The method proposed in this paper takes advantages of SIFT matching and uses it to obtain initial matches. For the subsequent sub-pixel matching, we extend existing bound constrained LSM framework (Hu *et al.*, 2016) to multi-image matching. And in order to achieve both robustness between images of difference scales and accuracies between images of the same scale, we directly matching in the original image space rather than the warped image as the original work (Hu *et al.*, 2016).

Bound-Constrained Multiple-Image LSM for Multiple-Resolution Images

Overview of the Approach

Following previous works on using LSM to improve the matching quality (Barazzetti *et al.*, 2010; Lerma *et al.*, 2013; Hu *et al.*, 2016), this paper also uses SIFT-like features as

initial matches and then exploits LSM to further refine the localization of the matches. However, in the case of multiple-resolution image matching, the scale differences of images may be as high as 20-fold (e.g., the Mars satellite images used in the following experimental analysis). In this case, SIFT-like matching, which is assumed to be scale invariant, will also fail if no extra strategy is adopted. Therefore, a two-step strategy is used in this method. The first step (upper part of Figure 1) is to identify the approximate overlapping region of the images based on the georeferencing information of the images and to down-sample the higher-resolution images to the same resolution as the lower-resolution images. SIFT matching followed by a RANSAC (Random Sample Consensus) (Fischler and Bolles, 1981) filtering process is then carried out to obtain initial matches. To obtain accurate matches, the second step of the bound-constrained multiple-image LSM (bottom part of Figure 1) is then carried out to refine the initial matches so as to obtain accurate sub-pixel matching results.

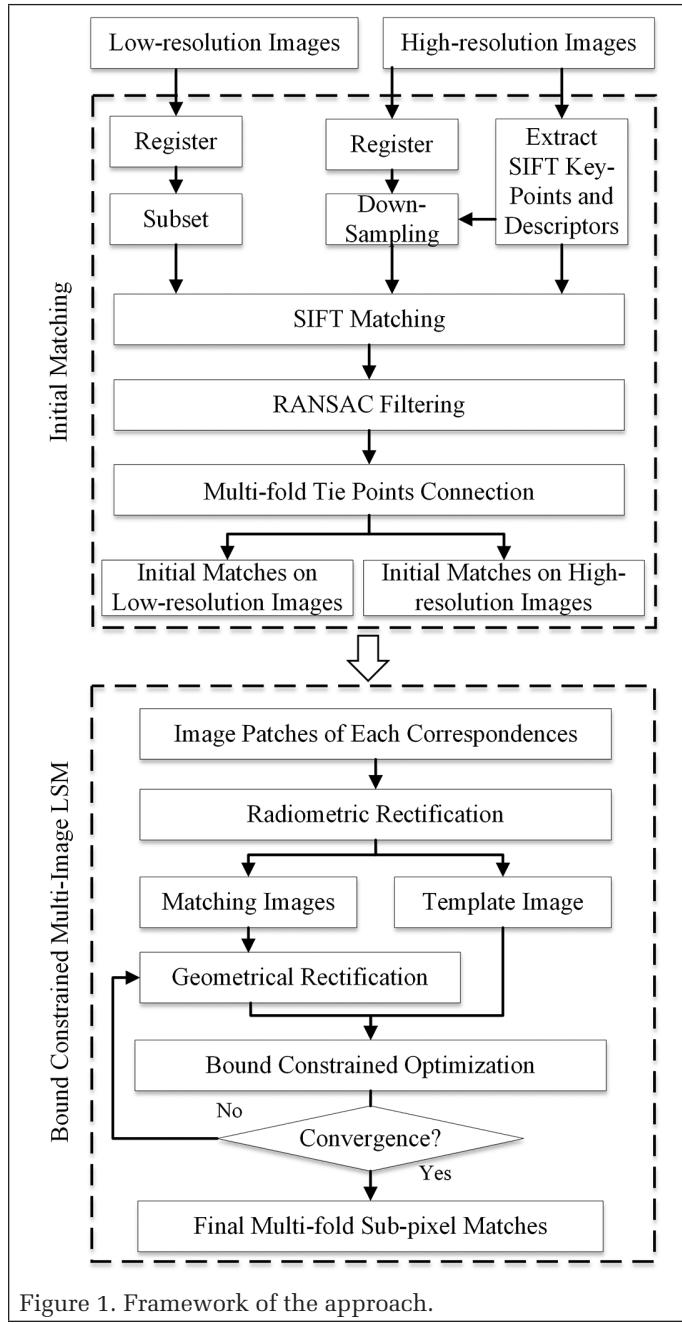


Figure 1. Framework of the approach.

Initial Matching

Two major obstacles should be solved to establish initial matches. One is the scale difference between the images, which causes problems in feature extraction and matching, and the other is the spatial coverage difference between the images, which causes problems in the searching space during the matching process. Figure 2 shows examples of Mars images collected by two different sensors with different resolutions and coverage. However, because both types of images cover the common local region, after registering to the same geolocation, they should reflect the same scale, as shown in Figure 2a. Thus the first step in the initial matching is georeferencing of the respective images.

In general, the satellite images are delivered together with metadata, including Rational Polynomial Coefficients (RPCs) and georeferencing information, from the vendor. The georeferencing information is commonly expressed by certain functional models (e.g., the affine transformation with six parameters) and encoded in the image headers. In this way, a direct relationship between the image coordinates (x, y) and the corresponding geographic coordinates (λ, ϕ) (λ and ϕ denote longitude and latitude, respectively) can be established. Even in some rare cases in which the georeferencing information is not available, an initial georeferencing model can be established by manually selecting a few control points and fitting an affine model. In practice, four control points are sufficient with two redundant observations.

Using the georeferencing information, we could explicitly establish a one-on-one map between the image coordinates of the higher-resolution image (x_h, y_h) and the lower-resolution image (x_l, y_l) from the intermediate geographic coordinates (λ, ϕ). Because the higher-resolution image covers only a portion of the lower-resolution image, the coverage bounding box is estimated as a part of the lower-resolution image for initial matching. The initial georeferencing will not be perfectly accurate, as indicated by the misalignment of the same terrain feature on different images as shown in Figure 2c; therefore, the bounding box is extended to account for the imperfections. For the images with a lower resolution, the subset in the bounding box is then clipped, and the images with a higher resolution will be down-sampled to the same resolution as the former.

After subset and down-sampling, the images from different sources should have similar geometric properties; the SIFT (Lowe, 2004) detectors and descriptors are then extracted from the images, and the two nearest neighbors in the descriptor space are searched using the approximate nearest neighbor method (Muja and Lowe, 2014). Cross checking is conducted to ensure that the correspondences are established uniquely. Because the epipolar geometry is not determined for push-broom satellite images, a more general homograph model is used as the geometric kernel for the RANSAC (Fischler and Bolles, 1981) outlier filtering. In fact, a recent variant of RANSAC, the AC-RANSAC (Moisan *et al.*, 2012), which features automatic threshold determination and better inlier retrieval, is used in this study. After outlier removal, pairwise matches are then connected using a connection component algorithm, from which multiple-fold tie points are obtained among the multiple-source images. In this paper, only the tie points matched on all images are selected, and a grid mask is adopted to ensure that the selected points are evenly distributed. Finally, the image coordinates of the initial matches are inversely transformed to the original image space for further refined matching.

Bound-Constrained Multiple-Image LSM

Pairwise LSM

After obtaining the coordinates of the multiple correspondences, a small patch centered around each point is extracted for LSM refinement. In order to reduce unnecessary parameters

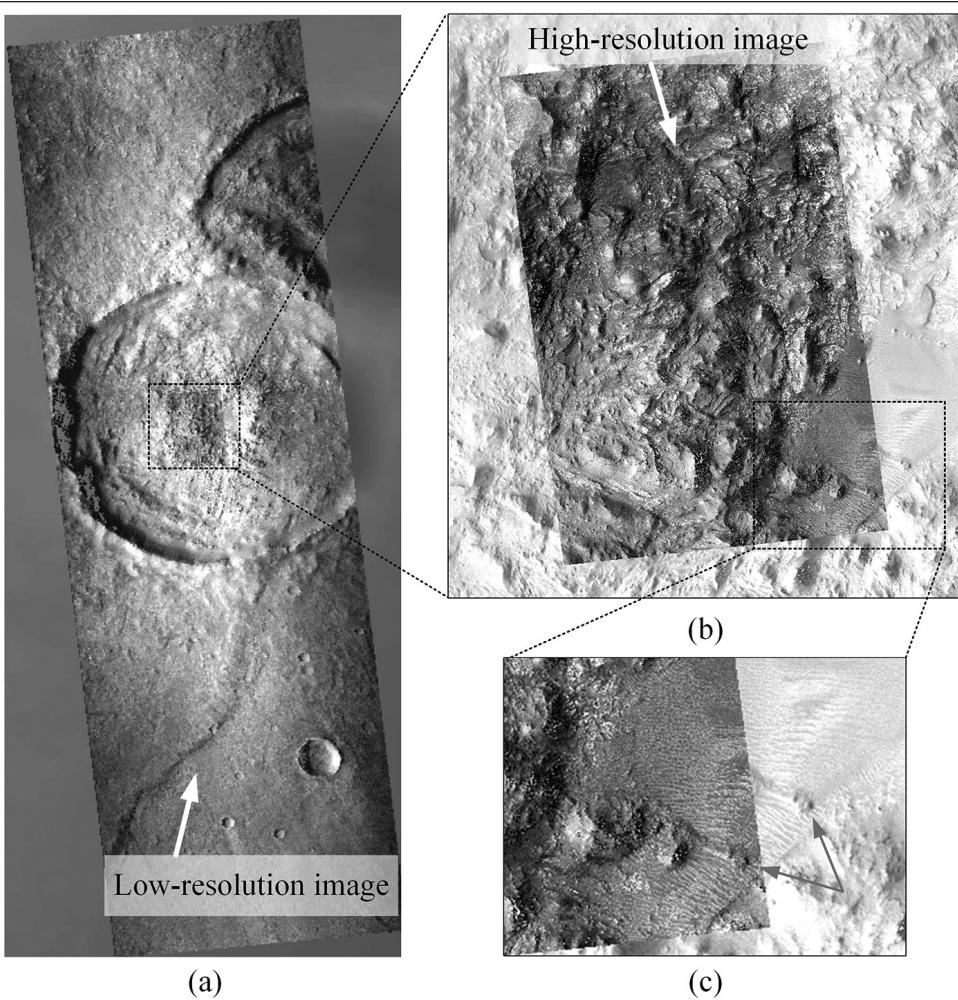


Figure 2. Multiple-resolution images for initial matching: (a) Georeferenced low-resolution image overlapped on a global DEM; (b) georeferenced high-resolution image overlapped on the low-resolution image; and (c) enlarged view of the border area. Note the imperfection of the georeferencing.

and improve the robustness of the least squares solver, the patches are all zero-meaned (Hirschmüller and Scharstein, 2009) before LSM. And in this way, the parameters that control the linear intensity transformation are ignored. Pairwise LSM is then defined as a nonlinear optimization between the template image I and the matching image I' , as the following minimization of Sum of Squared Difference (SSD) (Hu *et al.*, 2016):

$$\min_{\mathbf{x}} \sum_{r,c} [I'(r',c') - I(r,c)]^2 \quad (1)$$

where $\mathbf{x} = [a_1, a_2, a_3, b_1, b_2, b_3]$ is the unknown vector that contains the affine parameters, which is adopted by $r' = a_1c + a_2r + a_3$, $c' = b_1c + b_2r + b_3$, and (r, c) are the pixel coordinates for row and column of the patches. In order to compute the matching cost and Jacobian for the linear solver, the image patches are geometrically rectified in each iteration using the affine parameters. Equation 1 can be linearized in the following matrix form:

$$V = AX - L \quad (2)$$

where V is the residual vector, X is the vector of incremental corrections to the unknowns, and L is the vector of

observations. A is the design matrix or Jacobian matrix, in which each row contains partial derivatives with respect to the unknowns for each observation. Given initial values of the unknowns, the incremental corrections to the unknowns can be solved iteratively, and the unknowns can finally be solved.

Because the observations are relevant to the image space, the partial derivatives are related to the gradients of the image. During the iterative solver, the solution may not be convergent for multiple-resolution image matching, as aforementioned. In the following section, we demonstrate a method to make the iterative solutions more stable and accurate by (a) adoption of multiple images in a single optimization to handle inaccurate initialization for high resolution images, and (b) incorporation of bound-constrained solvers for multi-image LSM.

Multiple-Image Formulation of LSM

For brevity, only the case with four images is described here, which is a common case for two stereo pairs with different resolutions. However, it is simple to extend the strategy to general cases of multiple images. The low-resolution image pair is denoted as I_1 and I_2 , and the high-resolution image pair is denoted as h_1 and h_2 . Similar to pairwise LSM, a template image is required, and the other three are matching images, each with a set of affine parameters. The template image is chosen by identification of the high-resolution image that has higher feature responses, denoted as h_1 in the following description,

which may also vary for different four-folded tie points.

In general, there should be C_n^2 pairs of observations for n images and thus six pairs for the studied two stereo pairs. In order to obtain better matching results between high-resolution images, different weights are assigned and in this case, the observations are separated to two groups balanced by a weight parameter. Given a specific matching window size w , the mathematic model for multiple-image LSM is defined as:

$$\min_{\substack{\mathbf{x}_{h1 \rightarrow h1} \\ \mathbf{x}_{l2 \rightarrow h1} \\ \mathbf{x}_{h2 \rightarrow h1}}} \sum_{r,c} \left[\begin{array}{l} [I_{l2}(r_{l2},c_{l2}) - I_{h1}(r_{h1},c_{h1})]^2 + \\ [I_{h1}(r_{h1},c_{h1}) - I_{h1}(r_{l1},c_{l1})]^2 + \\ [I_{h1}(r_{h1},c_{h2}) - I_{l2}(r_{l2},c_{l2})]^2 + \\ [I_{h2}(r_{h2},c_{h2}) - I_{h1}(r_{l1},c_{l1})]^2 + \\ [I_{h2}(r_{h2},c_{h2}) - I_{l2}(r_{l2},c_{l2})]^2 \end{array} \right] + \lambda \sum_{r,c} [I_{h2}(r_{h2},c_{h2}) - I_{h1}(r_{l1},c_{l1})]^2 \quad (3)$$

where, for the left part, the matching cost involves both low-resolution and high-resolution images, and $\mathbf{x}_{h1 \rightarrow h1}$, $\mathbf{x}_{l2 \rightarrow h1}$, and $\mathbf{x}_{h2 \rightarrow h1}$ denote different sets of affine parameters for I_1 , I_2 , and h_2 , respectively. In fact, the template image also involves a fix parameter set $\mathbf{x}_{h1 \rightarrow h1}$, which is related to the scale difference,

in order to obtain the corresponding coordinates of $I_{h_1}(r_{h_1}, c_{h_1})$. The arrow in the subscript describes the image relative to which the affine parameter is set. The r and c are the image coordinates in the template image, and those with subscripts are the image coordinates in the matching image after affine transformation. A special case is made for $I_{h_2}(r_{h_2}, c_{h_2})$, which is propagated to l_1 using both parameters of $\mathbf{x}_{h_2 \rightarrow h_1}$ and $\mathbf{x}_{h_1 \rightarrow l_1}$.

For the right part in Equation 3, the matching cost is related only to the high-resolution image pair h_1 and h_2 . This part is included to ensure the matching accuracy between the high-resolution image pair in the LSM. In this case, the coordinates of $I_{h_2}(r'_{h_2}, c'_{h_2})$ is only relevant to $\mathbf{x}_{h_2 \rightarrow h_1}$. Another important component is the weight parameter λ , which balances the contributions from the fourth and fifth rows of the left part and the right part. In theory, λ is related to the precision of the observations and thus the image scale difference, but in practice, small variations of the parameters will produce similar results. Therefore, in this study, $\lambda = 10$ is used for all the experiments, which indicates five times importance of the matching cost of the two high-resolution images.

For the general case with N ($N \geq 3$) images, the formulation of multi-image least squares optimization is similar to Equation 3. And the images with the highest resolution could be chosen as the template image. With respect to the unknowns, there should be $N-1$ sets of affine parameters and the parameters should connect incrementally to the lower resolution

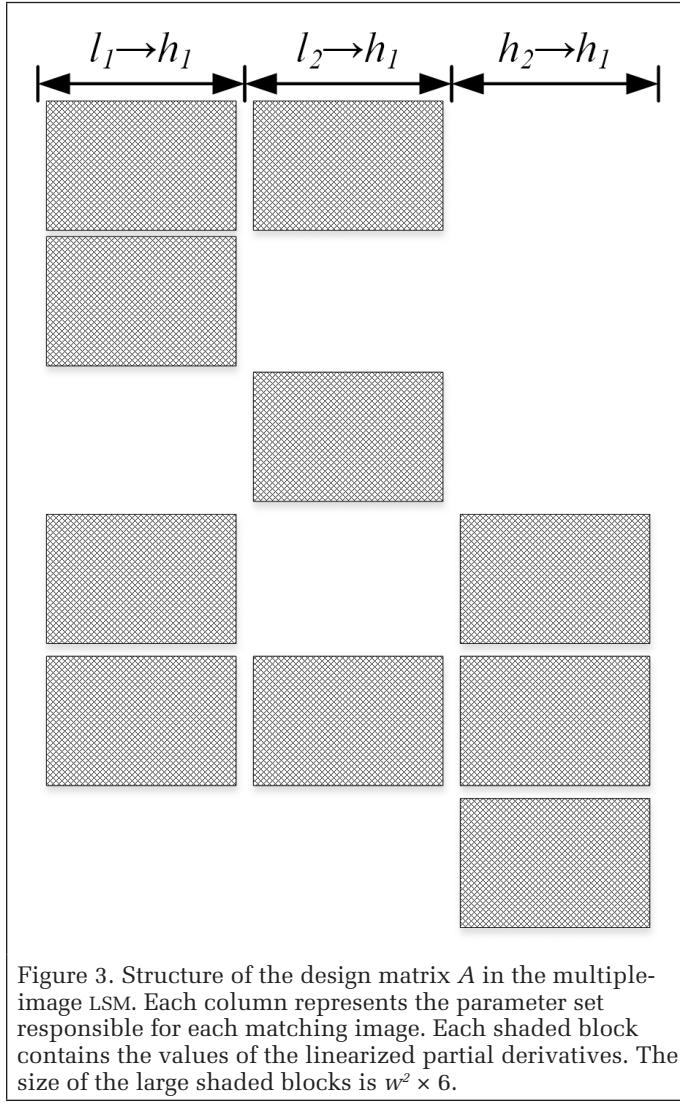


Figure 3. Structure of the design matrix A in the multiple-image LSM. Each column represents the parameter set responsible for each matching image. Each shaded block contains the values of the linearized partial derivatives. The size of the large shaded blocks is $w^2 \times 6$.

images, rather than against the same template image; otherwise the matching between the higher scales are also down sampled. With respect to the observations, C_n^2 pairs of observations could be formulated at most. When matching between images with higher resolutions, the correspondence established in the shifted position according to the affine parameters, as demonstrated in the right part of Equation 3.

The linearization of Equation 3 is similar to the classical LSM, except that the Jacobian matrix A contains more rows and is sparse because not all of the observations are correlated. The values for the partial derivatives for each parameter are the same as with the classical LSM (Gruen, 1985). The structure of matrix A is shown in Figure 3. The top five rows in Figure 3 correspond to the left part of Equation 3. It can be noticed that the second and third observation blocks are relevant to only one parameter set, which is in fact identical to that of classical LSM, and performs in a manner similar to pairwise matching. However, for the remainder of the multiple-image LSM, the parameters are correlated. The first row is responsible for the matching between l_1 and l_2 , which involves two parameter sets. The fourth and fifth rows are responsible for matching h_2 with l_1 and l_2 , respectively. The last row is the matching between h_1 and h_2 and is directly conducted in the original image space of the high-resolution images.

After linearization of Equation 3, least-squares optimization can be solved iteratively after providing initial values for all of the parameters. The selection of initial values are all selected as $[0, 1, 0, 1, 0, 0]$, which assumes no affine and shift transformation. This is because, for matching between a high resolution image and low resolution image, a hidden and fixed parameter set, $\mathbf{x}_{h_1 \rightarrow l_1}$ as $[0, s, 0, s, 0, 0]$, is used to warp the high resolution image to the low resolution image space and s is the scale difference between the images.

Bound-Constrained Optimization

In the multiple-image LSM, the affine transformation models between different image combinations must be optimized for the local image patch of the matching window. In practice, if the unknowns of the LSM are not constrained, the iterative solver will generally converge to a reasonable result with scale differences as high as 15 percent (Hu *et al.*, 2016), and the appearance of the images after down sampling will be quite different to the original low resolution images. Therefore, the bound constraints around the initial values of the affine model are explicitly injected into optimization as follows:

$$\min \sum_{\substack{\mathbf{x}_{h_1 \rightarrow h_1} \\ \mathbf{x}_{l_2 \rightarrow h_1} \\ \mathbf{x}_{h_2 \rightarrow h_1}}} \left[\begin{array}{l} [I_{l_2}(r_{l_2}, c_{l_2}) - I_{l_1}(r_{l_1}, c_{l_1})]^2 + \\ [I_{h_1}(r_{h_1}, c_{h_1}) - I_{l_1}(r_{l_1}, c_{l_1})]^2 + \\ [I_{h_1}(r_{h_1}, c_{h_1}) - I_{l_2}(r_{l_2}, c_{l_2})]^2 + \\ [I_{h_2}(r_{h_2}, c_{h_2}) - I_{l_1}(r_{l_1}, c_{l_1})]^2 + \\ [I_{h_2}(r_{h_2}, c_{h_2}) - I_{l_2}(r_{l_2}, c_{l_2})]^2 \end{array} \right]^2 + \lambda \sum_{r,c} [I_{h_2}(r'_{h_2}, c'_{h_2}) - I_{h_1}(r, c)]^2 \quad (4)$$

$$s.t. |a_1^i - \bar{a}_1^i| < \delta_1, |a_2^i - \bar{a}_2^i| < \delta_2, |b_1^i - \bar{b}_1^i| < \delta_2, |b_2^i - \bar{b}_2^i| < \delta_1$$

$$|a_3^i| < \delta_3, |b_3^i| < \delta_3, i \in \{l_1 \rightarrow h_1, l_2 \rightarrow h_1, h_2 \rightarrow h_1\}$$

where i denotes the parameter set of the matching image i , \bar{a}_j^i and \bar{b}_j^i are the initial values of the affine parameters as described in the previous section, δ_k is the bound threshold. In practice, δ_1 and δ_2 account for affine transformation, and δ_3 accounts for the translation of the affine model.

In our previous study (Hu *et al.*, 2016), we pre-rectified the images, and LSM was only conducted between the rectified images. The unknowns of the affine parameters were thus constrained in a certain range around $[0, 1, 0, 1, 0, 0]$. However, this strategy will not propagate the optimization to the original image space when a large scale difference is encountered.

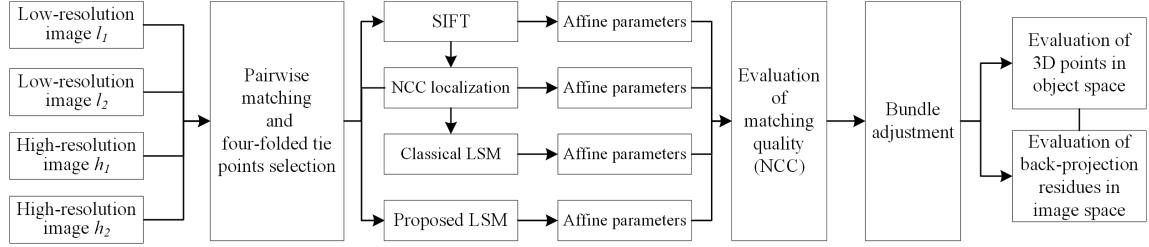


Figure 4. Evaluation workflow of the proposed LSM for multi-resolution image matching.

In this paper, the optimization can reach the original high-resolution images through the affine parameters, which also accounts for the scale difference. With the publicly available Levenberg-Marquardt solver (Lourakis, 2004; Agarwal and Mierle, 2010), the sparse bound-constrained optimization problem can be solved as long as the matching cost and Jacobian matrix are evaluated. The other modification for the solver is to the early termination strategy. Because the numerical thresholds on the cost or gradient require many more iterations after a reasonable result is reached, another strategy is also adopted with regard to the shifts of the template window (Hu *et al.*, 2016).

Experimental Analyses

To evaluate the approach presented, two datasets are used for experimental analysis. The first dataset comprises two stereo pairs of satellite images collected on the surface of Mars at different resolutions. The second dataset comprises two stereo pairs of satellite images collected in Hong Kong. These two datasets represent different types of terrain surfaces; the former is relatively smooth and the latter has the large elevation variations that typify metropolitan regions. For the parameters used in the algorithm, the weight $\lambda = 10$ is used; the bound constraint $\delta_1 = \delta_2 = 0.15$ is enough for a skewness of about 20° ; and $\delta_3 = 1$ pixel is used for translation, considering that the initial matches are filtered through RANSAC. For the termination threshold, the default numerical thresholds in the Levenberg-Marquardt solver are used, and the threshold on the maximum shift of the template window is set to 0.1 pixels.

In this paper, the proposed multi-image LSM is compared to the initial matches using SIFT, matches refined with NCC localization, and classical LSM that sequentially conduct three pairwise matches. The comparisons are both evaluated using a standard similarity measurement, NCC, and through metric accuracies in 3D object space by independent intersections and 2D back-projection residues, after a combined bundle adjustment. The evaluation workflow is illustrated in Figure 4. As shown in the workflow, only four-folded tie points are considered for evaluation, the points are matched and connected in the transformed images and re-projected to the original image space before evaluation. In order to unify the evaluation procedure, all the points are formatted by six parameters representing an affine transformation. It should be noted that for SIFT, it is just identity transformation with zero translation and for NCC localization, only the translation parts are changed. For the matching window size, a larger window will increase robustness, however the sub-pixel accuracy may decrease and increase the risk to cross the object boundaries in the matching window for urban areas. In this paper, a matching window of 21×21 pixels is used for both NCC match and LSM optimization, which is empirically a good balance between robustness and accuracy and also the default window size for typical photogrammetry software systems.

Because the tie points generally only cover quite a small subset of the low-resolution images, for the bundle adjustment, only the translation model in the image space are used as additional parameters for the RPCs, in order to improve the robustness of the adjustment. This simple adjustment model is considered to be enough to correct the inconsistencies for state-of-the-art satellite images (Fraser and Ravanbakhsh, 2009). And because no ground truth data are used, one image has to be fixed during the bundle adjustment, in order to remove the ambiguity of the null space of the solutions. After bundle adjustment, the discrepancies in 3D object space by photogrammetric intersection and the back-projection residues in the 2D image space are evaluated.

Experiments with Multiple-Resolution Mars Satellite Images

Two stereo pairs of images collected by the Context Camera (CTX) and the High-Resolution Imaging Science Experiment (HiRISE) camera, both onboard NASA's Mars Reconnaissance Orbiter (MRO) orbiting Mars, are used for experimental analysis. The CTX image pair has a spatial resolution of 6 m/pixel (Zurek and Smrekar, 2007), and the HiRISE image pair has a spatial resolution of 0.3 m/pixel (Kirk *et al.*, 2008), a 20-fold resolution difference between the two types of images. The intrinsic parameters and exterior orientation (EO) parameters for both CTX and HiRISE images are known and can be retrieved from the Spacecraft, Planet, Instrument, C-Matrix, and Events (SPICE) kernels (Acton, 1996). For both the CTX and HiRISE images, level 2 products are used, which are radiometrical and geometrical calibrated, and should be considered the same location accuracy because they are onboard the same platform.

Figure 5 shows an overview of the Mars satellite images and the matching results. The image-orientation parameters derived from the SPICE kernels are used for georeferencing of the images. Figure 5a shows the left image of the CTX stereo pair overlaid with the matched points. Figure 5b shows the left image of the HiRISE stereo pair overlaid with the corresponding matched points. From the distribution of the matched points, it can be seen that the HiRISE image covers only a small part in the middle of the CTX image. The right images of the CTX and HiRISE stereo pairs are not presented here because they show scenarios similar to those of the left images. Figure 5c shows the corresponding matches between the CTX and HiRISE images. The HiRISE image is scaled to the same resolution as the CTX image for better presentation. As indicated by the pattern of the lines connecting the corresponding matches, no visible outliers are observed.

Figure 6 shows enlarged views of the images and matching results. It can be seen that even though the appearances of the HiRISE and CTX images are quite different, as shown in Figure 6c and 6d, the proposed method can retrieve reasonable matching results in this difficult case. As indicated in Figure 6e, the initial matches on the HiRISE stereo pair (yellow crosses) sometimes do not correspond well to each other; after bound-constrained multiple-image LSM, they have been located to the same feature position (cyan crosses). Although the SIFT key points are detected in the original image space,

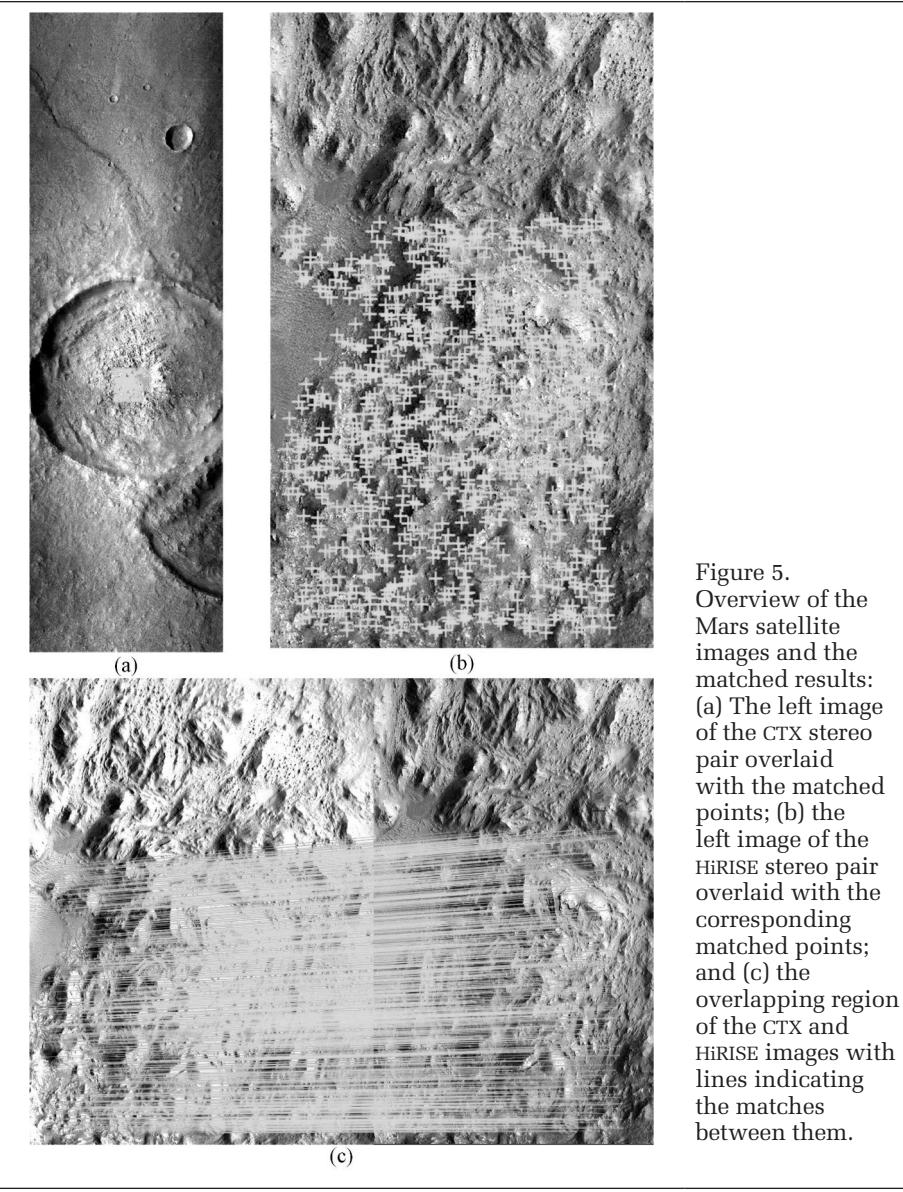


Figure 5. Overview of the Mars satellite images and the matched results: (a) The left image of the CTX stereo pair overlaid with the matched points; (b) the left image of the HiRISE stereo pair overlaid with the corresponding matched points; and (c) the overlapping region of the CTX and HiRISE images with lines indicating the matches between them.

and for the matching between high resolution images, the original SIFT descriptors are compared and matched. From Figure 6d it could be noted that, although some initial features will not move very much during the LSM process; however, because matching between the low resolution images involves a much larger context, it is possible for incorrect key points to be connected to the same track by the connected components of the matching graph, through the edges established between high resolution image and low resolution image, as shown in Figure 6e. However, these erroneous matches have been successfully adjusted to the corresponding location during the multi-image LSM, because the proposed method is both aware of large and small contexts.

Turning to quantitatively evaluate the performance of the proposed method, we compare both the similarity measurements through NCC and metric measurements after bundle adjustment. As shown in Table 1, for the four folded tie points, we compare all the six possible matching pairs, and the average value is also recorded. For the NCC match, we only search a small window (10 pixels) around the initial tie points and for the high resolution image; the scale difference should also be accounted. And the points obtained from NCC localization is used as initial matches for both classical LSM and the proposed LSM as described in Figure 4. For the classical LSM, the three sets of parameters as described in Equations 3 and 4 are refined sequentially, e.g., $x_{l1 \rightarrow h1}$, $x_{l2 \rightarrow h1}$ and $x_{h2 \rightarrow h1}$, and for the proposed method, all the parameters are adjusted in a single optimization. When LSM fails to converge, we resort to previous NCC points rather than ignore them, in order to make a fair comparison between all the methods. It can be noted

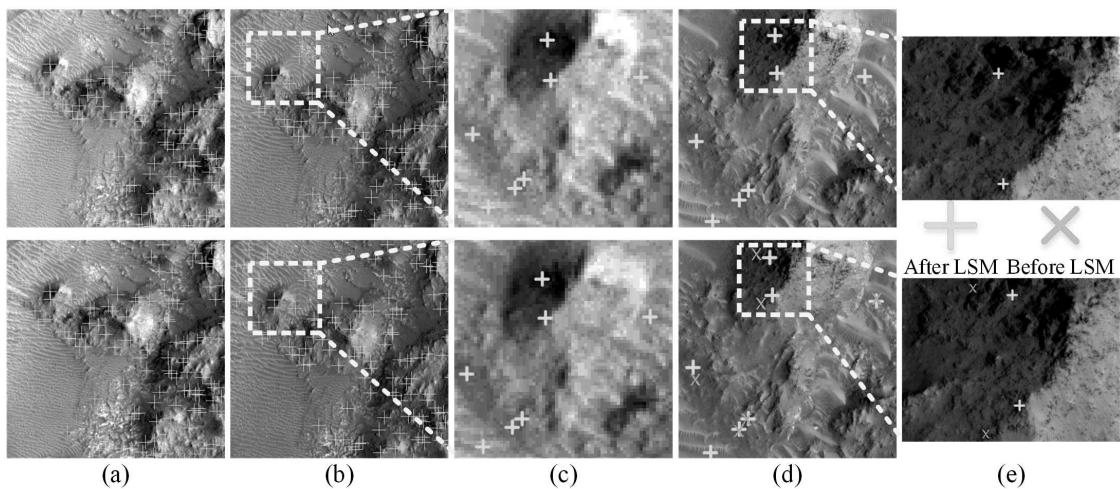


Figure 6. Enlarged views of the images and matched results for Mars dataset: (a) The CTX stereo pair (top left and bottom right images) at the original resolution of 6 m/pixel and the matched points; (b) the HiRISE stereo pair down-sampled to 6 m/pixel and the matched points; (c) a subset of the CTX stereo pair scaled to 0.3 m/pixel; (d) the same subset of the HiRISE stereo pair with the original resolution of 0.3 m/pixel; and (e) comparison of matches before and after LSM.

Table 1. Comparison of similarity measurements by NCC for the four methods of the Mars datasets; the proposed LSM method has the best overall performance; the bold values indicated the best results for each pair.

Pair	SIFT	NCC	Classical LSM	Proposed LSM
$I_1 - I_2$	0.894	0.914	0.967	0.967
$h_1 - I_1$	0.804	0.906	0.939	0.918
$h_1 - I_2$	0.765	0.865	0.932	0.914
$h_2 - I_1$	0.748	0.823	0.789	0.890
$h_2 - I_2$	0.786	0.860	0.806	0.916
$h_1 - h_2$	0.315	0.704	0.670	0.749
Avg.	0.719	0.845	0.851	0.892

that the proposed method has the best overall performance over other competitors, especially for the image pairs that are not directly optimized by the classical LSM, namely $h_2 - I_1$, $h_2 - I_2$. This is because of the faint texture information in the Mars dataset, and the classical LSM between $h_1 - h_2$ may be prone to overfitting, thus resulting in inferior matching quality. This also explained why the NCC values for classical LSM is even worse than that of the NCC localization for the fourth to sixth rows in Table 1. Furthermore, another finding is that for the pair of $h_1 - I_1$ and $h_1 - I_2$, the proposed method is also on-par with the classical LSM, even though that the classical LSM is a directly matching between the two images.

For the evaluation of metric capabilities, the original RPCs for the CTX and HiRISE images are fitted using least squares solver using the exterior orientation parameters and camera intrinsic information obtained from SPICE kernels for the MRO missions. In the bundle adjustment, the *a priori* standard deviations for image observations are 1 pixel and 20 pixels for the CTX and HiRISE images, respectively. In this setting, the standard deviation of unit weight has the same scale on the points measurements of CTX image. Both image reprojection and intersection errors are recorded. Because the reprojection errors do not have the same scale, they are separately recorded in two rows as shown in Table 2. There is a steady improvement through the adoption of NCC localization and LSM, which is consistent with previous works on close-range and aerial photogrammetry (Barazzetti *et al.*, 2010; Lerma *et al.*, 2013; Hu *et al.*, 2016). Although classical LSM may also achieve good internal consistency between the two low resolution images, the inconsistency between the two high resolution images will also influence the final 3D reconstruction accuracy as shown in the right four columns of Table 2.

Experiments with Multiple-Resolution Satellite Images in Hong Kong

To further evaluate the proposed method, two stereo pairs of satellite images of a typical metropolitan region in Hong Kong were also used for experimental analysis. The stereo pair with low resolution was obtained by the camera aboard the Chinese ZY-3 satellite, which is a three-line CCD pushbroom sensor. The stereo images were obtained using the along-track stereo configuration by the forward-looking and backward-looking CCD lines and have a resolution of approximately 3.5 m/pixel for both images. The location accuracy without ground control points is less than 15 m. The convergence angle between the forward-looking and backward-looking images is about 44°. The stereo pair with a high resolution were obtained by the Pleiades-1 satellite. The stereo images were captured using the along-track stereo configuration by tilting the satellite. The stereo images have a resolution of about 0.5 m/pixel for both sets of images, and the convergence angle between the stereo images is about 22°. The location accuracy without ground control points is 3 m according to the vendor.

Figure 7 shows an overview of the Hong Kong satellite images and the matching results. The vendor-provided

Table 2. Comparison of the metric accuracies after bundle adjustment of the two stereo pairs; the top rows for RMS of pixels represent back projection errors on the CTX images and the bottom rows for HiRISE images; The proposed LSM method has the best overall performance; the bold cells denote the best values for each metric.

	σ_0	RMS _x (pixels)	RMS _y (pixels)	RMS (pixels)	RMS X (m)	RMS Y (m)	RMS Z (m)	RMS (m)
SIFT	0.70	0.47	0.55	0.73	5.42	5.43	19.41	20.87
		9.38	14.92	17.62				
NCC	0.31	0.21	0.32	0.38	2.92	2.58	8.02	8.91
		4.22	4.30	6.02				
Classical LSM	0.28	0.19	0.25	0.31	2.37	2.52	5.60	6.58
		3.83	3.98	5.52				
Proposed LSM	0.25	0.17	0.22	0.28	2.10	2.37	5.34	6.21
		3.46	4.03	5.31				

georeferencing information of the four image corners are used for georeferencing in this study. Figure 7a shows the georeferenced backward-looking image of the ZY-3 stereo pair overlaid with the matched points. Figure 7b shows the georeferenced backward-looking image of the Pleiades-1 stereo pair overlaid with the corresponding matched points. The Pleiades-1 mainly covers the island portion of Hong Kong. The forward-looking images of the ZY-3 and Pleiades-1 stereo pairs are not presented here because they show scenarios similar to those of the backward-looking images. Figure 7c shows the corresponding matches between the ZY-3 and Pleiades-1 images, and no visible outliers are observed. It should be noted that the matches are mainly distributed along the coastal area and areas with a lower density of buildings for two reasons: (a) the mountainous areas will result in fewer features for matching because the textures are less distinctive; and (b) in the areas with dense high-rise buildings, feature matching suffers from enormous occlusion problems and other difficulties, so fewer matches are obtained in these areas. Nevertheless, the obtained matches among the ZY-3 and Pleiades-1 images are already sufficient for their integrated use.

Turning to the detailed matching results, Figure 8 shows enlarged views of the images and matching results. It can be seen that the distribution of the matches corresponds well. A small skew difference can be seen from the images, and the affine model in LSM is able to handle the skewness well. Although, most initial matches only have small movements during LSM as shown in the bottom row of Figure 8d, Figure 8e shows an example of incorrect initial matches due to the connection with low resolution images, as described above, that the locations of the matches after LSM on the Pleiades-1 images have been shifted by about ten pixels.

For the evaluation of matching quality, Table 3 shows the NCC values for the four methods. Compared to the Mars dataset, because the scale difference is only seven times, other methods will also obtain relatively better matching quality for pair $h_1 - h_2$ in the original image space due to better initial matching accuracy. And classical LSM is also available to increase the matching quality because of better texture information. And for the matching between different images, because the affine deformation is more severe as shown in Figure 8c and 8d, the NCC values are relatively lower. However, the proposed LSM still obtain the best overall performance with regard to NCC values. For the evaluation of the metric capabilities, the vendor-provided rational polynomial coefficients of the satellite images are used for the bundle adjustment and only the two translation parameters on image space are estimated. In the bundle adjustment, the standard deviation of unit weight (σ_0) is computed with *a priori* weight one pixel for the ZY-3 image and seven pixels for the Pleiades-1 image, respectively. Table 4 summarizes the evaluations after the bundle adjustment, which indicates about 10 percent improvements in the final 3D reconstruction consistencies.

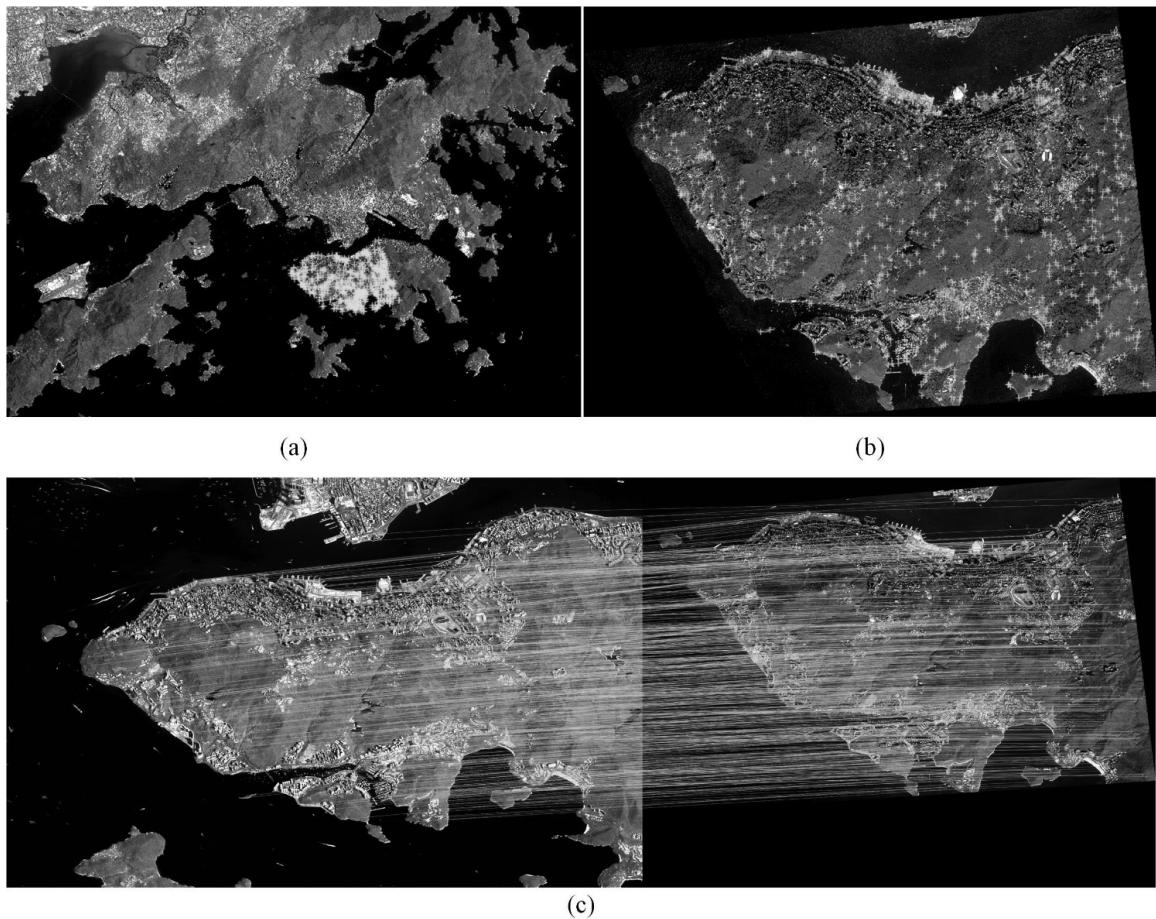


Figure 7. Overview of the satellite images in Hong Kong and the matched results: (a) The backward-looking image of the ZY-3 stereo pair overlaid with the matched points; (b) the backward-looking image of the Pleiades-1 stereo pair overlaid with the corresponding matched points; and (c) the overlapping region of the ZY-3 and Pleiades-1 images with lines indicating the matches between them.

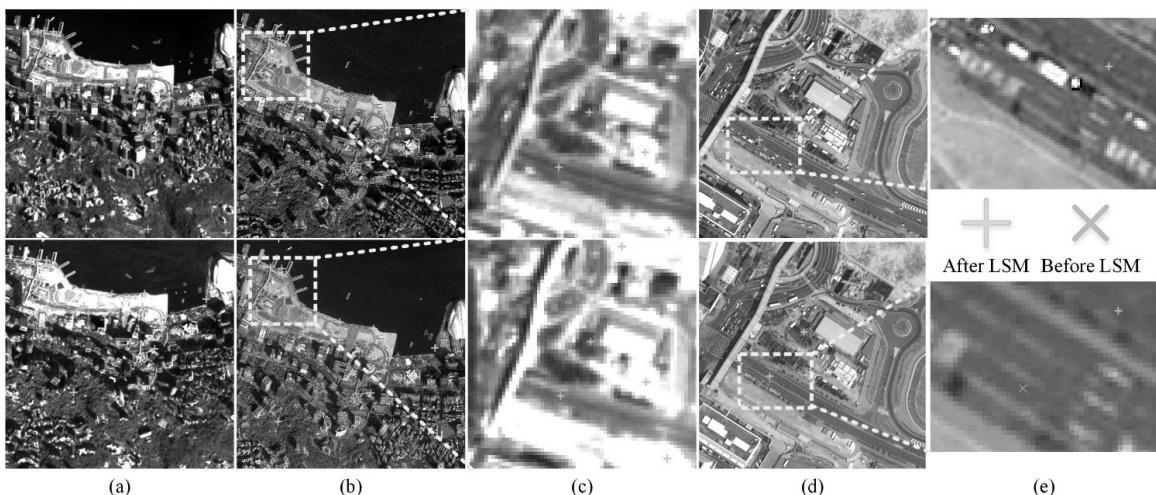


Figure 8. Enlarged views of the images and matched results for the Hong Kong dataset: (a) The ZY-3 stereo pair (backward-looking image on the top and forward-looking image on the bottom) at the original resolution of 3.5 m/pixel and the matched points; (b) the Pleiades-1 stereo pair down-sampled to 3.5 m/pixel and the matched points; (c) a subset of the ZY-3 stereo pair scaled to 0.5 m/pixel; (d) the same subset of the Pleiades-1 stereo pair with the original resolution of 0.5 m/pixel; and (e) comparison of matches before and after LSM for an incorrect initial matches.

Table 3. Comparison of Similarity Measurements by NCC for the four methods of the Hong Kong datasets; the proposed LSM method has the best overall performance; the bold values indicated the best results for each pair.

Pair	SIFT	NCC	Classical LSM	Proposed LSM
$I_1 - I_2$	0.784	0.801	0.887	0.898
$h_1 - I_1$	0.620	0.667	0.820	0.823
$h_1 - I_2$	0.667	0.709	0.870	0.859
$h_2 - I_1$	0.652	0.678	0.768	0.850
$h_2 - I_2$	0.667	0.670	0.758	0.825
$h_1 - h_2$	0.434	0.668	0.803	0.838
Avg.	0.637	0.699	0.818	0.849

Table 4. Comparison of the Metric Accuracies after Bundle Adjustment of the Two Stereo Pairs; The Top Rows for RMS of Pixels Represent Back Projection Errors on the ZY-3 Images and the Bottom Rows for Pleiades-1 Images; The Proposed LSM Method has the Best Overall Performance; The Bold Cells Denote the Best Values for Each Metric

σ_0		RMS _x (pixels)	RMS _y (pixels)	RMS (pixels)	RMS X (m)	RMS Y (m)	RMS Z (m)	RMS (m)
SIFT	0.59	0.55	0.43	0.70	3.15	2.98	5.46	6.97
		3.17	3.14	6.97				
NCC	0.35	0.36	0.23	0.43	1.32	1.69	3.49	4.09
		1.72	1.83	2.51				
Classical LSM	0.20	0.17	0.13	0.21	0.85	0.99	2.95	3.23
		0.73	1.17	1.38				
Proposed LSM	0.16	0.17	0.13	0.21	0.72	0.92	2.70	2.94
		0.76	1.13	1.36				

Conclusions and Discussion

In this paper, a bound-constrained multiple-image LSM method is proposed for the reliable matching of images with different resolutions. The performance of the developed method was evaluated with various types of satellite images collected on Mars and in Hong Kong. The theoretical analysis and experimental validation convey the following conclusions.

1. Although SIFT-like features are assumed to be scale invariant, in practice they may fail if they are used for images with relatively large scale differences, as indicated by the inaccurate initial matches in our experiments. For the standard SIFT and other SIFT-like detectors, when the scale difference of images is large, the scale space of the features may not overlap or only a few keypoints are detected within the overlap scales, which may lead to the failure of matching in a larger set of training descriptors.
2. The proposed LSM method is able to improve the stability and accuracy of multiple-resolution image matching by incorporating the bound constraint and multiple images in an integrated optimization framework, even in challenging cases with resolution differences as high as 20-fold. Misalignments at various levels between the 3D information derived from different satellite images are observed in the experiments in this study, which are further calibrated or rectified using the matched points from the proposed method. The bound-constrained multiple-image LSM method has significance for the synergistic use of multiple-source satellite images in various applications.

Acknowledgments

This work was supported by grants from the Research Grants Council of Hong Kong (Project No. PolyU 5330/12E and Project No. PolyU 152086/15E), grants from the National Natural Science Foundation of China (Project No. 91338110, Project No. 61602392, and Project No. 41631174), and grants from the Hong Kong Polytechnic University (Project No. 1-ZE24, Project No. 1-ZEA4, and Project No. 4-BCBW). The authors thank the Satellite Surveying and Mapping Application Center of the National Administration of Surveying, Mapping and Geoinformation of China for providing the ZY-3 satellite imagery and thank all those who worked on the Planetary Data System archive for the MRO data sets to make the HiRISE and CTX imagery publicly available.

References

- Acton, C.H., 1996. Ancillary data services of NASA's navigation and ancillary information facility, *Planetary and Space Science*, 44(1):65–70.
- Agarwal, S., and K. Mierle, 2010. Ceres Solver, URL: <http://ceres-solver.org/> (last date accessed: 08 August 2017.)
- Furukawa, Y., and J. Ponce, 2010. Accurate, dense, and robust multiview stereopsis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8):1362–1376.
- Gruen, A., 1985. Adaptive least squares correlation: A powerful image matching technique, *South African Journal of Photogrammetry, Remote Sensing and Cartography*, 14(3):175–187.
- Alcantarilla, P.F., A. Bartoli, and A.J. Davison, 2012. KAZE features, *Proceedings of the 12th European Conference on Computer Vision (ECCV2012)*, 07–13 October, Florence, Italy, pp. 214–227.
- Barazzetti, L., M. Scaioni, and F. Remondino, 2010. Orientation and 3D modelling from markerless terrestrial images: Combining accuracy with automation, *The Photogrammetric Record*, 25(132):356–381.
- Bay, H., A. Ess, T. Tuytelaars, and L. Van Gool, 2008. Speeded-up robust features (SURF), *Computer Vision and Image Understanding*, 110(3):346–359.
- Brown, M., R. Szeliski, and S. Winder, 2005. Multi-image matching using multi-scale oriented patches, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2005)*, 20–25 June, San Diego, California, pp. 510–517.
- Carneiro, G., and A.D. Jepson, 2005. The distinctiveness, detectability, and robustness of local image features, *Computer Vision and Pattern Recognition, 2005, CVPR 2005, Proceedings of the IEEE Computer Society Conference*, pp. 296–301.
- Chandrsekhar, V., G. Takacs, D. Chen, S. Tsai, R. Grzeszczuk, and B. Girod, 2009. CHoG: Compressed histogram of gradients a low bit-rate feature descriptor, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR2009)*, 20–25 June, Miami, Florida, pp. 2504–2511.
- Corbane, C., K., Saito, L. Dell Oro, E. Bjorgo, S.P. Gill, B. Emmanuel Piard, C.K. Huyck, T. Kemper, G. Lemoine, and R.J. Spence, 2011. A comprehensive analysis of building damage in the 12 January 2010 MW7 Haiti earthquake using high-resolution satellite and aerial imagery, *Photogrammetric Engineering & Remote Sensing*, 77(10):997–1009.
- Dalal, N., and B. Triggs, 2005. Histograms of oriented gradients for human detection, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2005)*, 20–26 June, San Diego, California, pp. 886–893.
- Elakshner, A.F., and A. Alharthy, 2011. Matching conjugate points between multi-resolution satellite images using geometric and radiometric properties, *International Journal of Remote Sensing*, 32(8):2117–2129.
- Fischler, M.A., and R.C. Bolles, 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM*, 24(6):381–395.
- Förstner, W., and E. Gülich, 1987. A fast operator for detection and precise location of distinct points, corners and centres of circular features, *Proceedings of the ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, 02–04 June, Interlaken, Switzerland, pp. 281–305.
- Fraser, C., and M. Ravanbakhsh, 2009. Georeferencing performance of GEOEYE-1, *Photogrammetric Engineering & Remote Sensing*, 75(6):634–638.

- Gruen, A., 2012. Development and status of image matching in photogrammetry, *The Photogrammetric Record*, 27(137):36–57.
- Gruen, A., and E. Baltsavias, 1988. Geometrically constrained multiphoto matching, *Photogrammetric Engineering & Remote Sensing*, 54(5):633–641.
- Harris, C., and M. Stephens, 1988. A combined corner and edge detector, *Proceedings of the Alvey Vision Conference*, 31 August–02 September, Manchester, UK, pp. 147–152.
- Hirschmüller, H., and D. Scharstein, 2009. Evaluation of stereo matching costs on images with radiometric differences, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9):1582–1599.
- Hu, H., Y. Ding, Q. Zhu, B. Wu, L. Xie, and M. Chen, 2016. Stable least-squares matching for oblique images using bound constrained optimization and a robust loss function, *ISPRS Journal of Photogrammetry and Remote Sensing*, 118:53–67.
- Jiang, W., 2006. Least square multiple image matching based on 3D plane primitive and its application in DSM generation, *Proceedings of the ASPRS Fall Conference*, 06–10 November, San Antonio, Texas.
- Kirk, R.L., E. Howington Kraus, M.R. Rosiek, J.A. Anderson, B.A. Archinal, K.J. Becker, D.A. Cook, D.M. Galuszka, P.E. Geissler, and T.M. Hare, 2008. Ultrahigh resolution topographic mapping of Mars with MRO HiRISE stereo images: Meter-scale slopes of candidate Phoenix landing sites, *Journal of Geophysical Research: Planets*, 113(E3):E00A24.
- Lerma, J.L., S. Navarro, M. Cabrelles, A.E. Seguí, and D. Hernández, 2013. Automatic orientation and 3D modelling from markerless rock art imagery, *ISPRS Journal of Photogrammetry and Remote Sensing*, 76:64–75.
- Lindeberg, T., 1993. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention, *International Journal of Computer Vision*, 11(3):283–318.
- Lindeberg, T., 1998. Feature detection with automatic scale selection, *International Journal of Computer Vision*, 30(2):79–116.
- Ling, X., Y. Zhang, J. Xiong, X. Huang, and Z. Chen, 2016. An image matching algorithm integrating global SRTM and image segmentation for multi-source satellite imagery, *Remote Sensing*, 8(8):672.
- Lourakis, M.I.A., 2004. levmar: Levenberg-Marquardt nonlinear least squares algorithms in C/C++, URL: <http://users.ics.forth.gr/~lourakis/levmar/index.html>, (last date accessed: 08 August 2017).
- Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, 60(2):91–110.
- Ma, X., D. Liu, J. Zhang, and J. Xin, 2015. A fast affine-invariant features for image stitching under large viewpoint changes, *Neurocomputing*, 151(Part 3):1430–1438.
- Matas, J., O. Chum, M. Urban, and T. Pajdla, 2004. Robust wide-baseline stereo from maximally stable extremal regions, *Image and Vision Computing*, 22(10):761–767.
- Mikolajczyk, K., and C. Schmid, 2002. An affine invariant interest point detector, *Proceedings of the European Conference on Computer Vision (ECCV2002)*, 28–31 May, Copenhagen, Denmark, pp. 128–142.
- Mikolajczyk, K., and C. Schmid, 2005. A performance evaluation of local descriptors, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630.
- Mikolajczyk, K., T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, 2005. A comparison of affine region detectors, *International Journal of Computer Vision*, 65(1–2):43–72.
- Moisan, L., P. Moulon, and P. Monasse, 2012. Automatic homographic registration of a pair of images, with a contrario elimination of outliers, *Image Processing On Line*, 2:56–73.
- Moravec, H.P., 1981. Rover visual obstacle avoidance, *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI 1981)*, 24–28 August, Vancouver, Canada, pp. 785–790.
- Muja, M., and D.G. Lowe, 2014. Scalable nearest neighbor algorithms for high dimensional data, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(11):2227–2240.
- Qiao, G., W. Wang, B. Wu, C. Liu, and R. Li, 2010. Assessment of geo-positioning capability of high resolution satellite imagery for densely populated high buildings in metropolitan areas, *Photogrammetric Engineering & Remote Sensing*, 76(8):923–934.
- Remondino, F., 2006. Detectors and descriptors for photogrammetric applications, *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVI(Part3):49–54.
- Rosten, E., R. Porter, and T. Drummond, 2010. Faster and better: A machine learning approach to corner detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(1):105–119.
- Rublee, E., V.Rabaud, K. Konolige, and G. Bradski, 2011. ORB: An efficient alternative to SIFT or SURF, *Proceedings of the IEEE International Conference on Computer Vision (ICCV2011)*, 06–13 November, Barcelona, Spain, pp. 2564–2571.
- Sedaghat, A., and H. Ebadi, 2015. Distinctive order based self-similarity descriptor for multi-sensor remote sensing image matching, *ISPRS Journal of Photogrammetry and Remote Sensing*, 108:62–71.
- Szeliski, R., 2011. *Computer Vision: Algorithms and Applications*, First edition, Springer, London, 812 p.
- Tang, S., B. Wu, and Q. Zhu, 2016. Combined adjustment of multi-resolution satellite imagery for improved geo-positioning accuracy, *ISPRS Journal of Photogrammetry and Remote Sensing*, 114:125–136.
- Tao, Y., J. Muller, and W. Poole, 2016. Automated localisation of Mars rovers using co-registered HiRISE-CTX-HRSC orthorectified images and wide baseline Navcam orthorectified mosaics, *Icarus*, 280:139–157.
- Tola, E., V. Lepetit, and P. Fua, 2010. Daisy: An efficient dense descriptor applied to wide-baseline stereo, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(5):815–830.
- Tuytelaars, T., and K. Mikolajczyk, 2008. Local invariant feature detectors: a survey, *Foundations and Trends in Computer Graphics and Vision*, 3(3):177–280.
- Voigt, S., T. Schneiderhan, A. Twele, M. Gähler, E. Stein, and H. Mehl, 2011. Rapid damage assessment and situation mapping: learning from the 2010 Haiti earthquake, *Photogrammetric Engineering & Remote Sensing*, 77(9):923–931.
- Wang, Z., B. Fan, G. Wang, and F.C. Wu, 2016. Exploring local and overall ordinal information for robust feature description, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(11):2198–2211: DOI 10.1109/TPAMI.2015.2513396.
- Wu, B., F. Li, L. Ye, S. Qiao, J. Huang, X. Wu, and H. Zhang, 2014. Topographic modeling and analysis of the landing site of Chang'E-3 on the Moon, *Earth and Planetary Science Letters*, 405: 257–273.
- Xue, Y., W. Xie, O. Du, and H. Sang, 2015. Performance evaluation of DSM extraction from ZY-3 three-line arrays imagery, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-7(W4):247–253.
- Yang, H., M. Yu, and S. Zhang, 2014. Wide baseline stereo matching based on scale invariant feature transformation with hybrid geometric constraints, *IET Computer Vision*, 8(6):611–619.
- Zhang, G., T. Chen, H. Pan, and W. Jiang, 2011. Patch based Least squares image matching based on rational polynomial coefficients model, *Acta Geodaetica et Cartographica Sinica*, 40(5):592.
- Zhang, L., 2005. *Automatic Digital Surface Model (DSM) Generation From Linear Array Images*, PhD. Thesis, ETH, Zürich, Switzerland, 199 p.
- Zhang, Y., M. Zheng, X. Xiong, and J. Xiong, 2015. Multistrip bundle block adjustment of ZY-3 satellite imagery by rigorous sensor model without ground control points, *IEEE Geoscience and Remote Sensing Letters*, 12(4):865–869.
- Zhu, Q., B. Wu, and N. Wan, 2007. A filtering strategy for interest point detecting to improve repeatability and information content, *Photogrammetric Engineering & Remote Sensing*, 73(5):547–553.
- Zurek, R.W., and S.E. Smrekar, 2007. An overview of the Mars Reconnaissance Orbiter (MRO) science mission, *Journal of Geophysical Research: Planets*, 112(E5).