ISPRS Journal of Photogrammetry and Remote Sensing 118 (2016) 53–67

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Stable least-squares matching for oblique images using bound constrained optimization and a robust loss function

Han Hu [a,b], Yulin Ding [a,*], Qing Zhu [a,c,d], Bo Wu [b], Linfu Xie [d], Min Chen [a]

[a] Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, PR China
[b] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
[c] National-local Joint Engineering Laboratory of Spatial Information Technology for High-speed Railway Running Safety, Southwest Jiaotong University, PR China
[d] Collaborative Innovation Center for Geospatial Technology, Wuhan, PR China

ARTICLE INFO

Article history:
Received 13 July 2015
Received in revised form 23 March 2016
Accepted 31 March 2016

Keywords:
Least-squares matching
Sub-pixel image matching
Oblique images
Bound constrained optimization

ABSTRACT

Least-squares matching is a standard procedure in photogrammetric applications for obtaining sub-pixel accuracies of image correspondences. However, least-squares matching has also been criticized for its instability, which is primarily reflected by the requests for the initial correspondence and favorable image quality. In image matching between oblique images, due to the blur, illumination differences and other effects, the image attributes of different views are notably different, which results in a more severe convergence problem. Aiming at improving the convergence rate and robustness of least-squares matching of oblique images, we incorporated prior geometric knowledge in the optimization process, which is reflected as the bounded constraints on the optimizing parameters that constrain the search for a solution to a reasonable region. Furthermore, to be resilient to outliers, we substituted the square loss with a robust loss function. To solve the composite problem, we reformulated the least-squares matching problem as a bound constrained optimization problem, which can be solved with bounds constrained Levenberg–Marquardt solver. Experimental results consisting of images from two different penta-view oblique camera systems confirmed that the proposed method shows guaranteed final convergences in various scenarios compared to the approximately 20–50% convergence rate of classical least-squares matching.

1. Introduction

Currently, aerial oblique images are becoming the new mapping standard for digital city modeling due to their capabilities for viewing building facades from different angles, such as the penta-view camera system with "Maltese Cross" configuration (Petrie, 2009; Xiao et al., 2012; Rupnik et al., 2014; Xiong et al., 2014; Lemmens, 2014a). The fundamental step in 3D applications with aerial oblique images is determining the exterior orientation (EO) parameters for all the camera heads. Due to possible synchronization problems among the camera heads and the rigidity of camera platform (Wiedemann and Moré, 2012; Lemmens, 2014b; Hu et al., 2015; Rupnik et al., 2015), all images should be oriented through a combined block adjustment; otherwise, if only nadir images are oriented, the inaccurate EO parameters for oblique images obtained from calibrated platform parameters will lead to obvious systematic errors in the object space. This has triggered new challenges for classical photogrammetric software, due to the significant geometrical and radiometric differences between oblique images. The problems are especially severe for image matching (Hu et al., 2015), which are explicitly affected by the geometrical deformations and radiometric changes between oblique images.

Least-squares matching (LSM) is a well-established image matching method providing feature correspondences in sub-pixel level (Förstner, 1982; Ackermann, 1984; Gruen, 1985). LSM minimizes the gray level differences between the template and the matching window whereby the position and the shape of the matching window are parameters to be determined in the adjustment process. LSM is a nonlinear optimization problem that needs to be linearized and solved iteratively to a convergent solution. To obtain the final convergence for LSM, two criteria should be satisfied (Gruen, 2012): (a) accurate initial correspondence must be established and (b) the image quality in both of the images in the LSM window should be similar. For traditional nadir images, the degree of geometrical deformations and radiometric

* Corresponding author. Tel.: +86 13437119857.
E-mail address: rainforests@126.com (Y. Ding).

http://dx.doi.org/10.1016/j.isprsjprs.2016.03.019
0924-2716/

differences are guaranteed by the systematic aerial flight and small varying viewing directions; therefore, the problem relies only on the initial match position. However, when processing oblique images, we have occasionally encountered significant blur and light condition differences due to lack of forward motion compensation in some oblique camera systems, as shown in Fig. 1. Furthermore, the tilting angles between oblique and nadir images will also cause obvious differences in appearance. In this case, accurate initial values for the unknown parameters sometimes cannot ensure that LSM will be convergent.

In each iteration in LSM, the method will generate a shift vector, which consists of the incremental values for the provisional parameters. Because the shift vector is the result of a first order approximation using the Taylor series, it is directly related to the grayscale values and gradients of the images in the LSM window. Incorrect shift vectors due to significant differences between image qualities may cause the problem of zigzagging (Moré and Thuente, 1994) and reach the maximum number of iterations set by the algorithm. Furthermore, in the case of similar textures or patterns, LSM may converge to another false local optimal value, which also causes incorrect results.

It should be noted that all of the parameters of LSM have physical significance, including affine transformation and image illumination differences; the ranges of the parameters will absolutely lie in a reasonable region, e.g., the lower and upper bounds. For example, the initial EO parameters for all images can be obtained from the GNSS/IMU system and the pre-calibrated installation rotation and translation parameters in a laboratory environment. Although large perspective deformations exist between vertical and oblique images, with the aid of the initial EO parameters, a homography transformation (Hartley and Zisserman, 2004) can be estimated, which will quantitatively portray the deformation; using the transformation, all of the images can be warped geometrically to alleviate the deformations, and the warped images will present almost no scale and rotation differences, as detailed in our previous work (Hu et al., 2015). Furthermore, the contrast and brightness differences between the aerial images will lie in a reasonable region.

Aiming at improving the quality and robustness of LSM for oblique images with regard to the significant differences in appearance caused by blur and diverse illumination conditions, this paper proposes to impose constraints on parameters to force LSM to iterate with unknowns in reasonable ranges using *a priori* information. Explicitly, we reformulated the LSM problem into a nonlinear optimization form subjected to bounded constraints (*c.f.* Section 3.2), which defined the lower and upper bounds of the parameters. Then, the bounded problem could be solved efficiently and simply using publicly available solvers as long as the cost function and gradient for the parameters were evaluated in each iteration (*c.f.* Section 3.4). By assigning limits of the parameters, the shift vectors incremented to the unknowns would not stray from the final convergence, and the search space for the unknowns was significantly reduced. Furthermore robust loss function, e.g., Huber loss, was proposed to handle the potential outliers of pixel values (*c.f.* Section 3.3). Huber loss (Hastie et al., 2009) would behave the same as the square loss in the inlier region, but the influence would only grow linearly rather than quadratically in the outlier region. Thus, robust and accurate sub-pixel locations could be achieved. Due to the limited search space for convergent solutions and the robust loss functions, we could guarantee final convergence of LSM.

The remaining parts of this paper are organized as follows. Section 2 summarizes previous work on matching and locating feature points with sub-pixel accuracies. Section 3 describes the reformulated LSM problem and present solutions for the constrained problem. Then, experimental evaluations and analyses are demonstrated in Section 4. Finally, the concluding remarks are presented.

2. Related works

In the community of photogrammetry, where accuracy and quality control sometimes receive first priority, establishing correspondences at sub-pixel locations has been extensively studied (Förstner, 1982), and numerous methods have been proposed for this process. Generally, two strategies exist for achieving sub-pixel correspondences: localization of feature points or matching two images in a small window. The first strategy seeks to obtain the maximum responses of interest points in the two images independently; the latter strategy commonly fixes one point as the reference and optimizes the location of the other image at the sub-pixel level.

In close range photogrammetry, precise and automatic methods to orient the image blocks have been developed, which matured approximately two decades ago, using the artificial coded targets (Fraser, 1997). The core technique for developing such automatic systems generally involves two independent steps: recognition and localization of the target point (Wong et al., 1988). Because the targets are designed to be the peak or trough responses in images, they can be easily detected through dynamically and adaptively thresholding. Then, the location of the targets can be precisely obtained by several methods, including eclipse fitting with the edges and centroiding (and its variants) or Gaussian distribution fitting with the pixel/ binary values in the target area

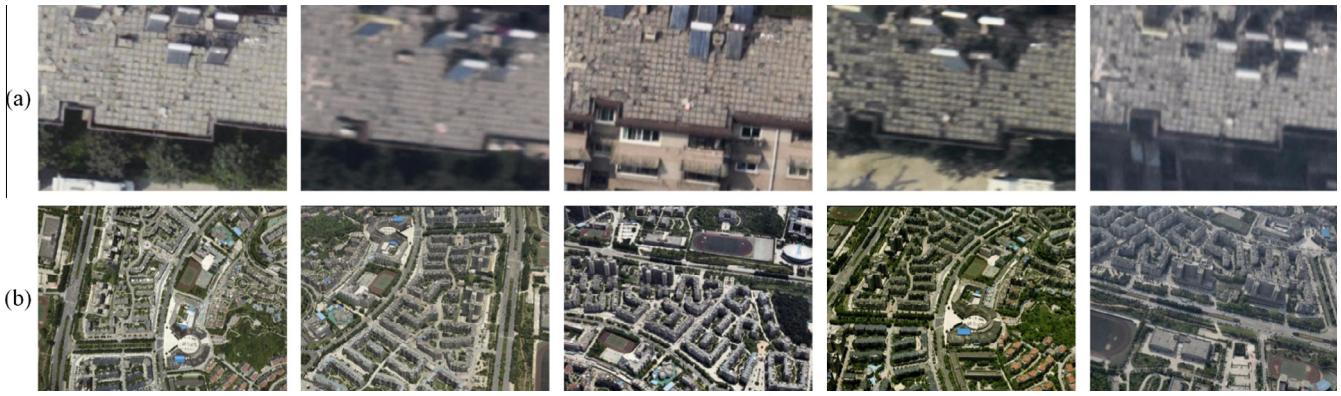

Fig. 1. Differences of appearance caused by diverse angles of incidence of the nadir cameras (first column) and four oblique cameras (other columns), including: (a) blur and (b) illumination differences. It should be noted that we pre-rectified the images in the first row for a more clear presentation. Lack of forward motion compensation in some of the aerial oblique camera systems and variation of ground sample distance (GSD) in the oblique views may lead to the blur effect. And the illumination differences may come from different sun-light directions and tuning of different camera sensors and lenses.

(Shortis et al., 1994). In the laboratory environment, using retro-reflected targets, the accuracies achieved by these target centering algorithms can reach one hundredth of a pixel or even greater (Shortis et al., 1995; Yu et al., 2002).

These localization methods used in close range photogrammetry require artificial targets placed in the scene, which are not suitable for markerless images. In the case of nature scenes, a common strategy shared by some interest detectors is to fit a continuous 3D surface with certain corner responses and optimize for the maximum at the sub-pixel level. For example, the Scale Invariant Feature Transform (SIFT) (Lowe, 2004) fits a bivariate function to the surface of the Difference-of-Gaussian (DoG) function convolved with the images in a 3×3 window. The location of the extremum is determined by a second order Taylor series of the scale-space function. By setting the derivatives to zero, this function has a closed-form solution. Furthermore, the SIFT will eliminate the feature points in which the DoG responses and edge responses are less than a certain threshold, borrowing from the Harris detector (Harris and Stephens, 1988). Furthermore, a more recent feature detector, the Speed-Up Robust Features (SURF) (Bay et al., 2008) also adopts a localization strategy similar to the SIFT. In our previous study (Zhu et al., 2007), we fit a quadratic paraboloid with the Harris interest strength in a small window centered on a point. Then, the least-squares fitting method using Gaussian weights was used to iteratively refine the paraboloid to obtain the final sub-pixel location.

The methods described above localize the feature points in the two segregated matching images. When the maximum corner responses are essentially present in different locations due to image differences *per se*, losses in matching accuracies are expected (Jazayeri and Fraser, 2010; Gruen, 2012). In this situation, localizing the correspondences by matching is more appropriate. A simple and efficient strategy is to correlate a point in one image with a small window in the other image and then fit the obtained Normalized Correlation Coefficient (NCC) with a bivariate unimodal paraboloid with respect to the two axes of the images. The location of the maximal NCC is the matched sub-pixel point (Tian and Huhns, 1986; Wang, 1990). Another widely adopted method for sub-pixel matching is to use upsampling. Althof et al. (1997) enlarged an 11×11 pixel region 16-fold using bilinear interpolation; then, a matched filter was used to determine the peak location, which was designed to improve the signal-to-noise ratio. Szeliski and Scharstein (2004) proposed a symmetric matching method that upsamples images using a bilinear or bicubic interpolant. Then, the matching cost is calculated in the upsampled images and reduced to the original resolution by averaging with a moving window. The authors argued that this method will relieve the bias problem produced by paraboloid fitting as described above. A more recent investigation (Xu et al., 2012) has systematically, theoretically and empirically examined the upsampling method. The upsampled and original images are both matched and then fused by optimizing the unified energy with multiple cues to reduce the outliers. Furthermore, Debella-Gilo and Kääb (2011) provided a comprehensive comparison of the two strategies described above using various interpolants during measurement of surface displacement. However, these methods require that matching costs (NCC or other pixel-wise costs (Birchfield and Tomasi, 1998)) are calculated in a region surrounding the point and thus are more suitable for dense image matching rather than optimizing sparse feature matching points.

Another promising algorithm for sub-pixel matching is phase correlation, which was originally used to estimate the translation shifts in image registration (Foroosh et al., 2002). The basic steps for calculating the phase correlation include: (1) applying a discrete 2D Fourier transform after preprocessing with a window function in the two images; (2) calculating the cross-power spectrum in the frequency space (Morgan et al., 2010); and (3) obtaining a normalized cross-correlation through an inverse Fourier transform; then, similar techniques to those described above can be used to locate the peak responses. However, Stone et al. (2001) noticed that the sub-pixel location of the peak responses can be directly achieved by analyses of the phase characteristics in the frequency space without the third step and, therefore, greatly boost the sub-pixel accuracy to the one-hundredth of a pixel level. The incipient method can only account for translation and can be extended to handle in-plane rotation and zoom (Chen et al., 1994) for image registration. Furthermore, Morgan et al. (2010) adopted this method for stereo matching using nearly identical images with a very narrow baseline. The disparity in accuracies is improved by an order of magnitude compared to that obtained by using the NCC as the matching cost. However, except for the extreme registration accuracies obtained by phase correlation, this method can only handle simple or almost identical scenes, such as translation only cases and those with small baseline-height ratios (only 0.003 in the study by Morgan et al. (2010)). This disadvantage has impeded the use of this method for complex images obtained by penta-view oblique camera systems.

To the best of our knowledge, the most well-known and widely used sub-pixel matching method is LSM, which was firstly proposed by Förstner (1982) and Ackermann (1984) and formally defined by Gruen (1985). LSM optimizes for two groups of parameters, including 6 parameters that address the affine image

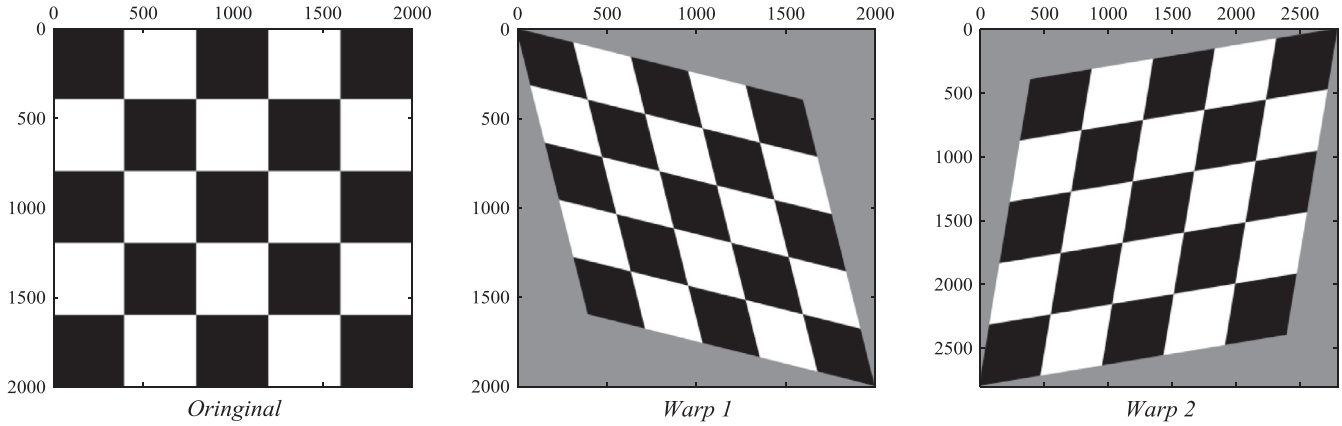

Fig. 2. Illustrations of different affine transformations. Left: original image, middle: $A = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$ and right: $A = \begin{bmatrix} 1.2 & -0.2 \\ -0.2 & 1.2 \end{bmatrix}$. The skewness is approximately 15° for both cases.

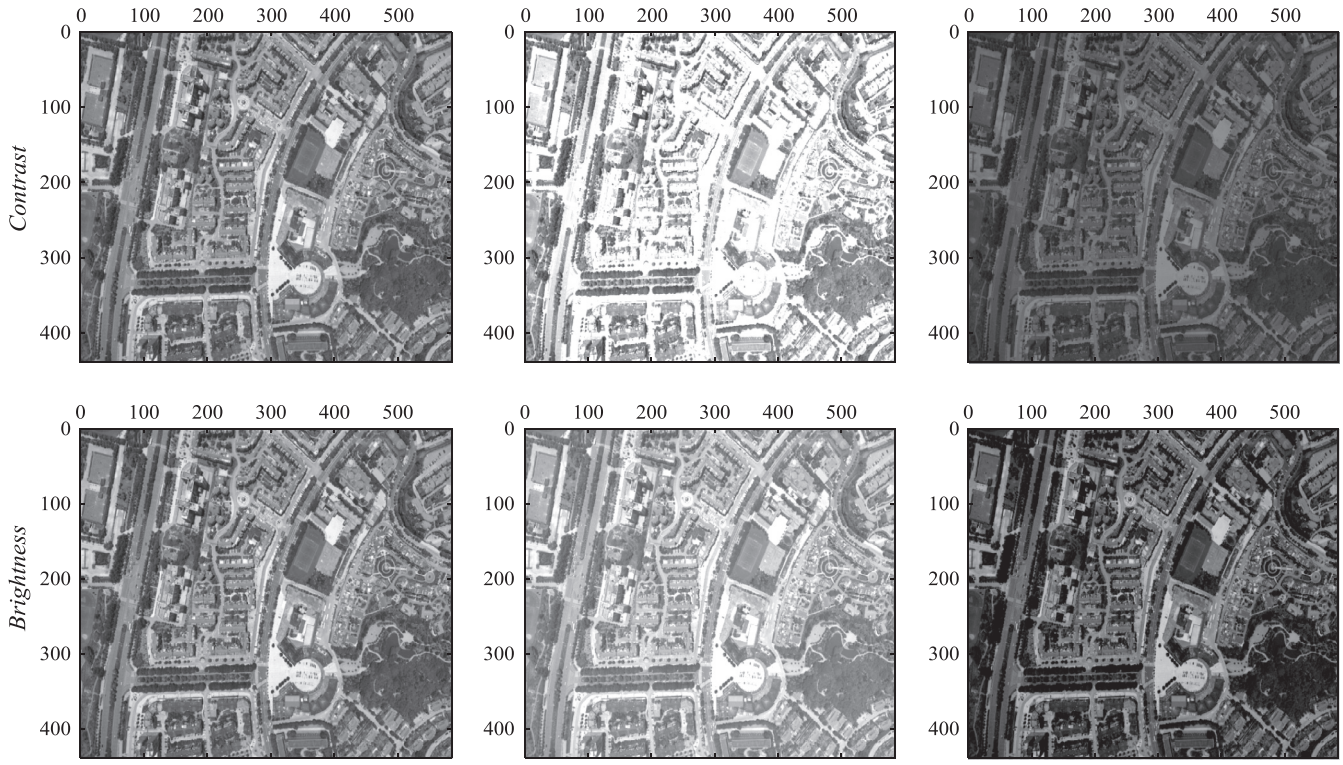

Fig. 3. Illustrations of different illumination conditions in the nadir image. Contrast changes: left: original image, middle: $k_1 = 2.0$, right: $k_1 = 0.5$ and brightness changes: left: original image, middle: $k_2 = 50$, and right: $k_2 = -50$. It can be noted that these ranges for the contrast and brightness are sufficient to address the illumination differences in practical applications.

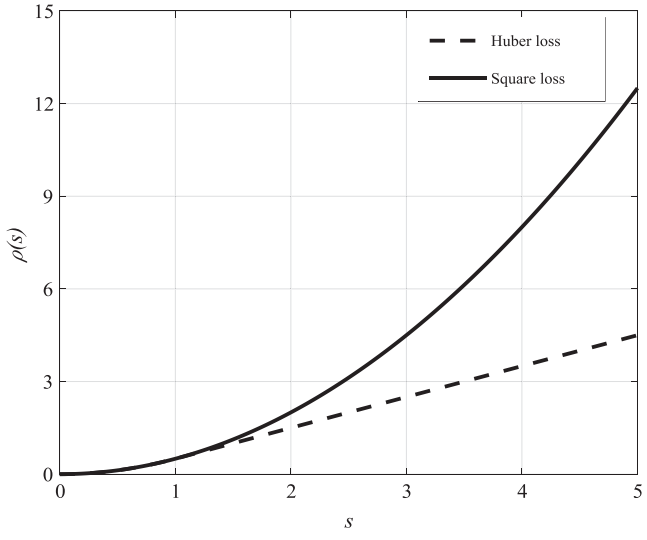

Fig. 4. Plots of Huber loss and square loss, where $a = 1$ as in Eq. (7). When the cost is less than the threshold, Huber loss is equivalent to the square loss, and in larger residual region, Huber loss will suppress the influences of outliers.

transformation and 2 that remedy radiometric differences, by minimizing the Sum of Squared pixel Differences (SSD) through the non-linear least-squares method. LSM has been extended to handle multiple overlapping images with geometric constraints, which are available through known EO parameters by pre-calibration or ground control points (Gruen and Baltsavias, 1988). Further extensions that have attempted to improve LSM include using adaptive cross-correlation instead of the SSD (Zheltov and Sibiryakov, 1997) and using the gradient instead of the pixel difference (Campbell and Wu, 2008). However, both authors have proven the equivalence of their methods with LSM because these substitutions are essentially closely or even linearly related to LSM.

Because LSM is a nonlinear optimization problem and the function is explicitly related to the image pixel values, it is sometimes not stable and relies on the initial positions and image qualities to a large extent. Otherwise, the iterative refinement may not converge (Wang, 1990; Gruen, 2012), especially in the case of oblique image matching. Although a geometrically constrained version (Gruen and Baltsavias, 1988) will significantly improve the stability of LSM, this improvement occurs at the cost of the requirement for known EO parameters and multiple overlaps and is not suitable for general case pairwise feature point matching. Another good strategy to improve stability is the weighted least square approach (Remondino, 2006), which uses elliptical regions from the affine invariant detectors (Mikolajczyk et al., 2005) as approximations and weighted observations in the solver. However, it is dependent on the affine detectors and needs sophisticated strategy to determine the weights. This paper presents a LSM approach for oblique images with two novel aspects: (1) incorporating bound constraints in the optimization for reliable solutions, and (2) employing the Huber loss function in the algorithm for robust matching when outliers are present.

3. Bound constrained least-squares matching

As described above, standard LSM is currently criticized for its instability in some circumstances. To make LSM robust under different conditions, we used a more stable solver rather than modifying the original cost function of LSM (Zheltov and Sibiryakov, 1997; Campbell and Wu, 2008) or adding constraints, which require more rigorous input conditions and limit the use of LSM in general cases (Gruen and Baltsavias, 1988). First, we redefined

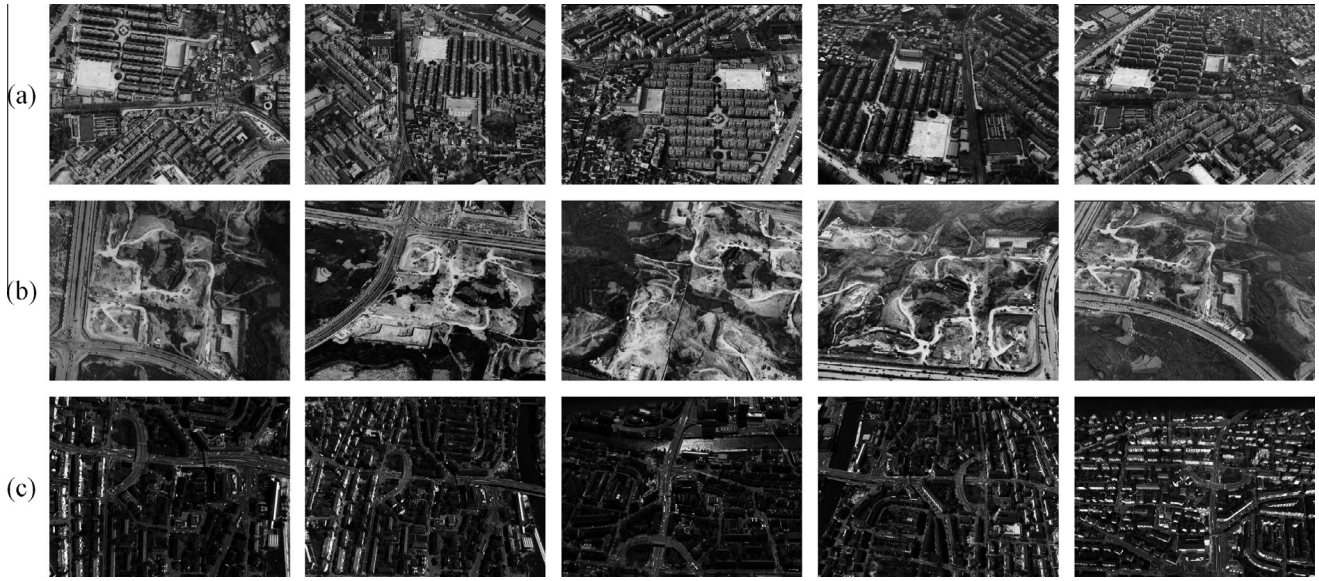

Fig. 5. Overview of the three experimental areas in Jinyang and Zurich including (a) built-up areas with various building types in Jinyang, (b) rural areas featuring bare earth, vegetation and roads in Jinyang and (c) urban area in Zurich, featuring complex combinations. The first column in each row represents the nadir image, denoted as *CamE*, and the other four columns represent the oblique images pointing from east to north, denoted as *CamA* to *CamD*.

Table 1
Comparison of the Jinyang and Zurich datasets. The meanings of the abbreviations in the parenthesis are: "*N*" for nadir images and "*O*" for oblique images.

	Jinyang dataset	Zurich dataset
Camera	SWDC-5	RCD-30
Tilt angles	45°	35°
Focal length (*N*/*O*)	50 mm/80 mm	53 mm/53 mm
Forward motion compensation	N.A.	2 axis mechanical methods
GSD (*N*/*O*)	8 cm/6–20 cm	6 cm/6–13 cm
Image size	8176 × 6132	9000 × 6732

the LSM problem using a general bounds constrained optimization form without changing the original merit of LSM. In this manner, we could easily incorporate constraints to limit the search space of the unknowns and thus a better posed problem. Then, we discuss how to solve the combined optimization problem under these constraints by using and modifying existing solvers.

3.1. Least-squares matching

Before illustrating the proposed problem definition, we provide a recap of classical LSM (Gruen, 1985) for completeness. For the image pixels $I_1(r, c)$ and $I_2(r, c)$ of two images in a square window (normally 21 × 21 as used in many aerial applications), the fundamental idea of LSM is to determine a set of parameters to satisfy the following observation condition,

$$I_1(r,c) = k_1 I_2(r',c') + k_2 + n(r,c) \quad (1)$$

$$r' = a_{11}r + a_{12}c + a_{13}, c' = a_{21}r + a_{22}c + a_{23} \quad (2)$$

where a_{11} through a_{23} are used to compensate for the affine deformations using 6 unknown parameters; k_1 and k_2 control the contrast and brightness differences, respectively, which are also known as the gain and bias of signal processing; and $n(r, c)$ is the random noise, which is assumed to have a Gaussian distribution. There are a total of 8 unknown parameters in LSM in Eq. (1), as $\boldsymbol{x} = [a_{11}, a_{12}, a_{13}, a_{21}, a_{22}, a_{23}, k_1, k_2]^T$.

After expanding the nonlinear form with a first order Taylor series and assigning an appropriate initial value for the parameter $\boldsymbol{x}_0$, the conditions in Eq. (1) lead to the following residual formulation,

$$V = Ad\boldsymbol{x} - \boldsymbol{b} \quad (3)$$

where each row v_i in the residual vector V denotes the signed differences for each pixel; A is a Jacobian or coefficient matrix that records the partial derivatives with respect to each unknown parameter; $d\boldsymbol{x}$ is the incremental vector. After updating the unknowns by $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + d\boldsymbol{x}$ ($k = 0, 1, \ldots, t$), the procedure iterates until it reaches a certain termination criteria as detailed in the work by Gruen (1985). The initial values are chosen as $\boldsymbol{x}_0 = [1, 0, 0, 0, 1, 0, 1, 0]$. Finally, the translation terms (a_{13}, a_{23}) in the 6 affine parameters are used to determine the sub-pixel location of the feature correspondences.

3.2. Bound constraints

The least-squares problem defined in Eq. (3) is equivalent to the following nonlinear optimization problem (Björck, 1996),

$$\min_x \frac{1}{2}\sum_{r,c}[I_1(r,c) - k_1 I_2(r',c') - k_2]^2 \quad (4)$$

which optimizes for the unknown vector $\boldsymbol{x}$ to minimize the SSD in the matching windows, as described above. In fact, Gruen (1985) has also discussed the convergence quality of the solutions of LSM with regard to oscillations and a false local optimum. In the iteration procedure, the parameters for affine deformations and the illuminations may stray from the correct values or even exceed the regions of common knowledge.

Currently, many GNSS/IMU providers exist, and modern digital aerial camera systems have these systems onboard aerial vehicles to provide initial position and attitude information simultaneously during flight. The absolute accuracy for position is at the centimeter to decimeter level, and the accuracy of the attitude at the level of a few thousandths of a degree can be reached (Ip et al., 2007), which has triggered interest for direct georeferencing (Mostafa and Schwarz, 2001; Zhang and Shen, 2013). Thus, with an 8-parameter perspective transformation using this initial

Table 2
Matching statistics for the test areas in both Jinyang and Zurich, where *E* denotes the nadir image, and *A* to *D* denote the oblique images oriented from East to North, respectively.

Pair	#Cross check	#NCC	Avg. NCC	LSM			Bound constrained LSM		
				#Convergences	Avg. NCC	#Filtered	#Convergences	Avg. NCC	#Filtered
(a) Tests in the urban area of Jinyang									
E–A	21,243	13,490	0.907	4117	0.961	3687	13,490	0.931	10,468
E–B	22,811	14,845	0.911	7018	0.963	6439	14,845	0.941	11,668
E–C	21,571	13,098	0.899	4503	0.955	4017	13,098	0.927	9407
E–D	26,695	17,866	0.910	7028	0.962	6388	17,866	0.936	13,865
A–C	10,548	6401	0.902	1843	0.961	1404	6401	0.931	2806
B–D	15,682	9359	0.907	3191	0.961	2665	9359	0.932	5641
E–E	79,303	70,972	0.939	58,369	0.965	57,390	70,972	0.964	68,114
(b) Tests in the rural area of Jinyang									
E–A	33,492	18,273	0.892	6710	0.952	6226	18,273	0.921	16,317
E–B	23,861	13,832	0.892	5310	0.949	5116	13,832	0.924	12,934
E–C	34,132	17,978	0.882	8513	0.944	8213	17,978	0.927	16,541
E–D	29,239	17,564	0.890	8018	0.949	7734	17,564	0.924	16,151
A–C	8677	3785	0.881	1215	0.949	884	3785	0.919	1714
B–D	4151	1952	0.886	726	0.951	584	1952	0.925	1093
E–E	10,3903	100,130	0.932	85,793	0.964	85,065	100,130	0.961	99,023
(c) Tests in the Zurich dataset									
E–A	15,599	12,032	0.933	6162	0.962	5488	12,032	0.956	8883
E–B	13,235	9832	0.932	5243	0.969	4646	9832	0.956	7188
E–C	10,730	7976	0.932	3758	0.964	3284	7976	0.953	5193
E–D	11,937	8882	0.928	4221	0.962	3588	8882	0.952	5913
A–C	7212	4810	0.913	1495	0.965	820	4810	0.938	1905
B–D	6279	3627	0.907	1133	0.960	830	3627	0.934	1636
E–E	36,847	31,920	0.951	24,434	0.986	23,339	31,920	0.983	29,148

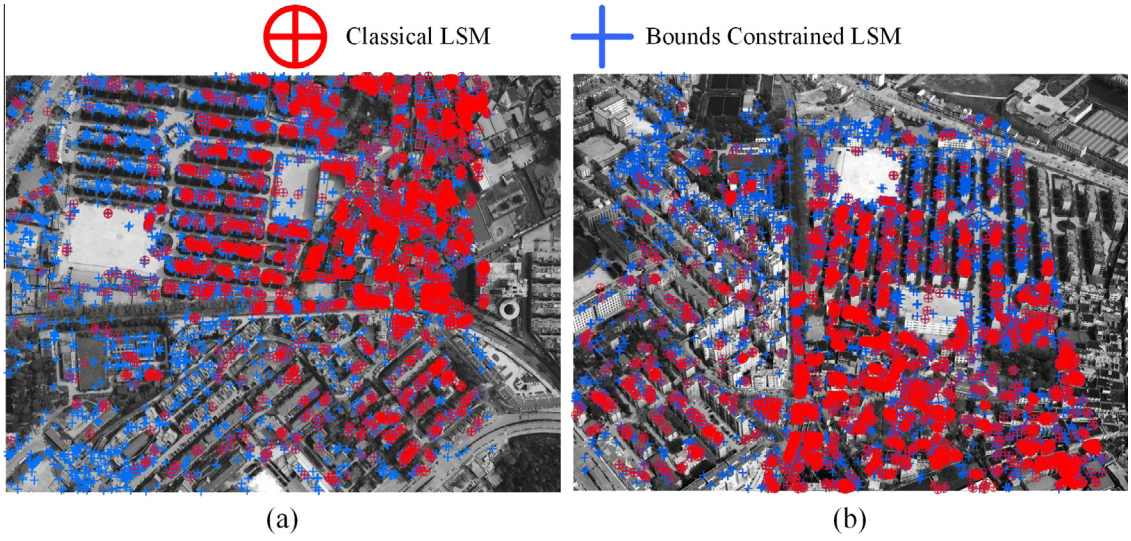

Fig. 6. Overview of the matching results after outlier removal with classical LSM and the proposed method in the urban area of Jinyang. (a) The nadir view and (b) an oblique view are shown. To show the improvements compared to the classical LSM, the red circle are overlaid over the results of the proposed method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

information, we can assume that, in the horizontal direction, the scale and rotation differences can be largely relieved if not eliminated, as discussed in our previous work (Hu et al., 2015). Even without the GNSS/IMU system, as long as the oblique images are obtained in a systematic manner with regular flight lines, there is only small attitude change between consecutive images for typical airborne camera installations. In the standard airborne environments, the altitude differences between consecutive images should not exceed ±5 degrees and even in extreme conditions, variations of more than 15 degrees are very rare in our experience. In order to give an intuitive understanding of the influences of the attitude difference to the image space, the affine transformation with corresponding skewness is illustrated in Fig. 2. If accurate GNSS/IMU system is not available and the acquired images are unordered, such as oblique UAV images and terrestrial images that share similar convergent network (Nex et al., 2015), the ranges of the affine transformation parameters can also be approximated through hierarchical bundle adjustment strategies; the EO parameters and sparse points from the upper level can be used to determine the perspective transformation. Therefore, the ranges of the parameters a_{11}, a_{12}, a_{21}, and a_{22} can be determined with a certain level of confidence.

Furthermore, to determine the initial positions of the correspondences, a commonly used strategy is to select the integer pixel

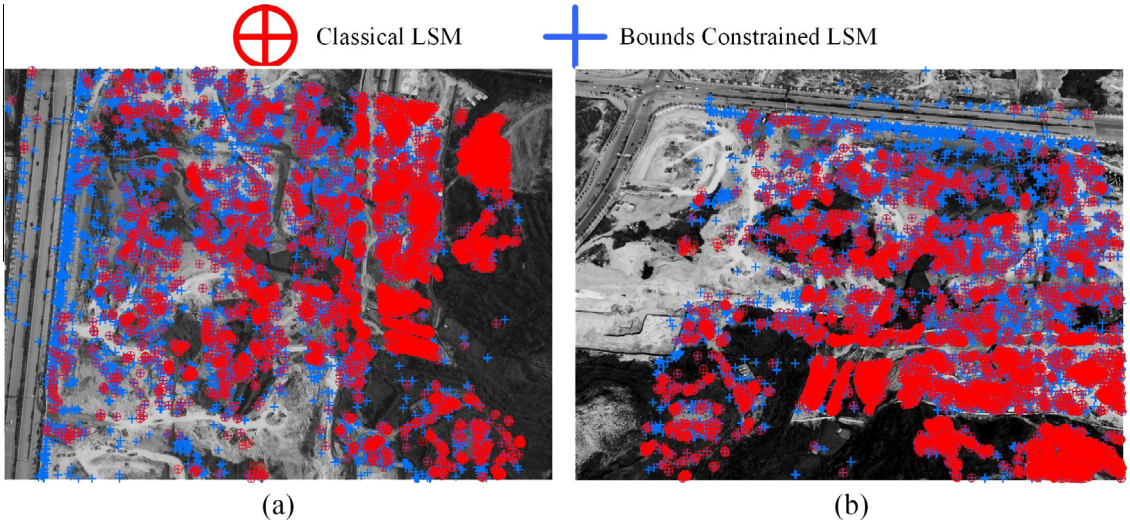

Fig. 7. Overview of the matching results after outlier removal with classical LSM and the proposed method in the rural area of Jinyang. (a) The nadir view and (b) an oblique view are shown. To show the improvements compared to the classical LSM, the red circle are overlaid over the results of the proposed method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

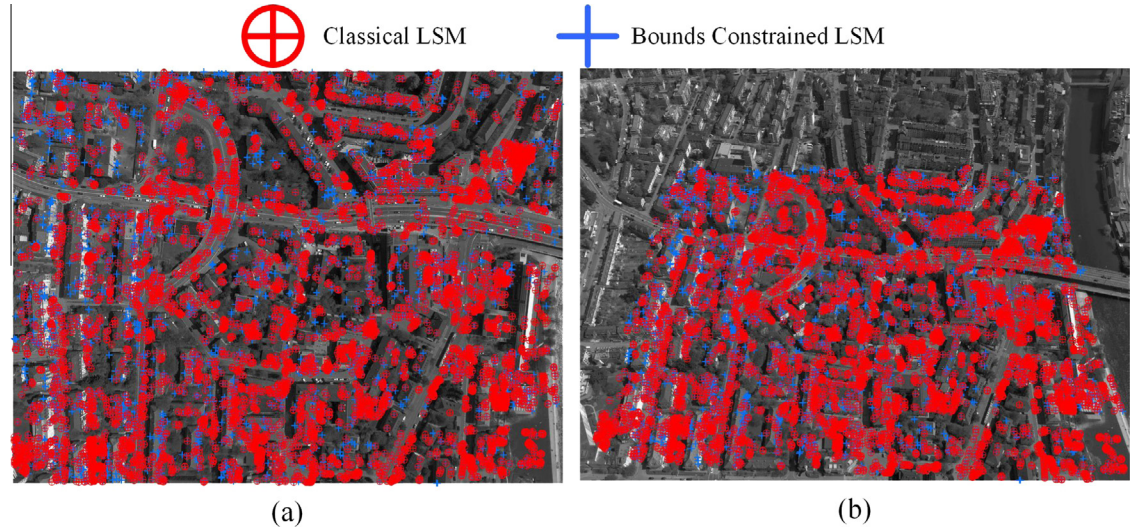

Fig. 8. Overview of the matching results after outlier removal with classical LSM and the proposed method in the residential area of Zurich. (a) The nadir view and (b) an oblique view are shown. To show the improvements compared to the classical LSM, the red circle are overlaid over the results of the proposed method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with the maximum NCC in a small search window around a feature point (Lerma et al., 2013). In fact, the NCC is equivalent to the SSD if the affine transformation and illumination effect are not taken into consideration. Therefore, in the ideal situation, a sub-pixel will be located around the integral pixel in the region of $[-0.5, 0.5]$ pixels. Regarding other factors, the range should also be not very large. Therefore, for the unknown parameters a_{13} and a_{23}, which account for the sub-pixel translations, the boundaries can also be gauged before optimization, as long as they are pre-located using the NCC. Furthermore, the contrast and brightness should also reside in a reasonable region, as shown in Fig. 3. It can be noted that a contrast value that varies in the range of $[-0.5, 2]$ and a brightness value in the range of $[-50, 50]$ will be sufficient to address the illumination differences.

As discussed above, all eight unknown parameters have strong physical significances and should be constrained to reasonable ranges. Therefore, we propose adding lower and upper boundaries to the unknown parameters in Eq. (4) as the follows,

$$\begin{aligned} &\min_x \tfrac{1}{2}\sum_{r,c}[I_1(r,c) - k_1 I_2(r',c') - k_2]^2 \\ &s.t. 1-\delta_1 < a_{11} < 1+\delta_1, 1-\delta_1 < a_{22} < 1+\delta_1 \\ &\quad -\delta_2 < a_{12} < \delta_2, -\delta_2 < a_{21} < \delta_2 \\ &\quad -\delta_3 < a_{13} < \delta_3, -\delta_3 < a_{23} < \delta_3 \\ &\quad \delta_4 < k_1 < 1/\delta_4, -\delta_5 < k_2 < \delta_5 \end{aligned} \tag{5}$$

where δ_1 and δ_2 control the degrees of the affine deformations and 0.2 will provide ample space for the potential deformations in real applications of oblique images; δ_3 controls the shifts from the initial corresponding positions, and 3–5 pixels are assigned to compensate for various unknown factors; δ_4 and δ_5 control the illumination changes, and as shown in Fig. 3, values of δ_4 = 0.5 and δ_5 = 50 are used.

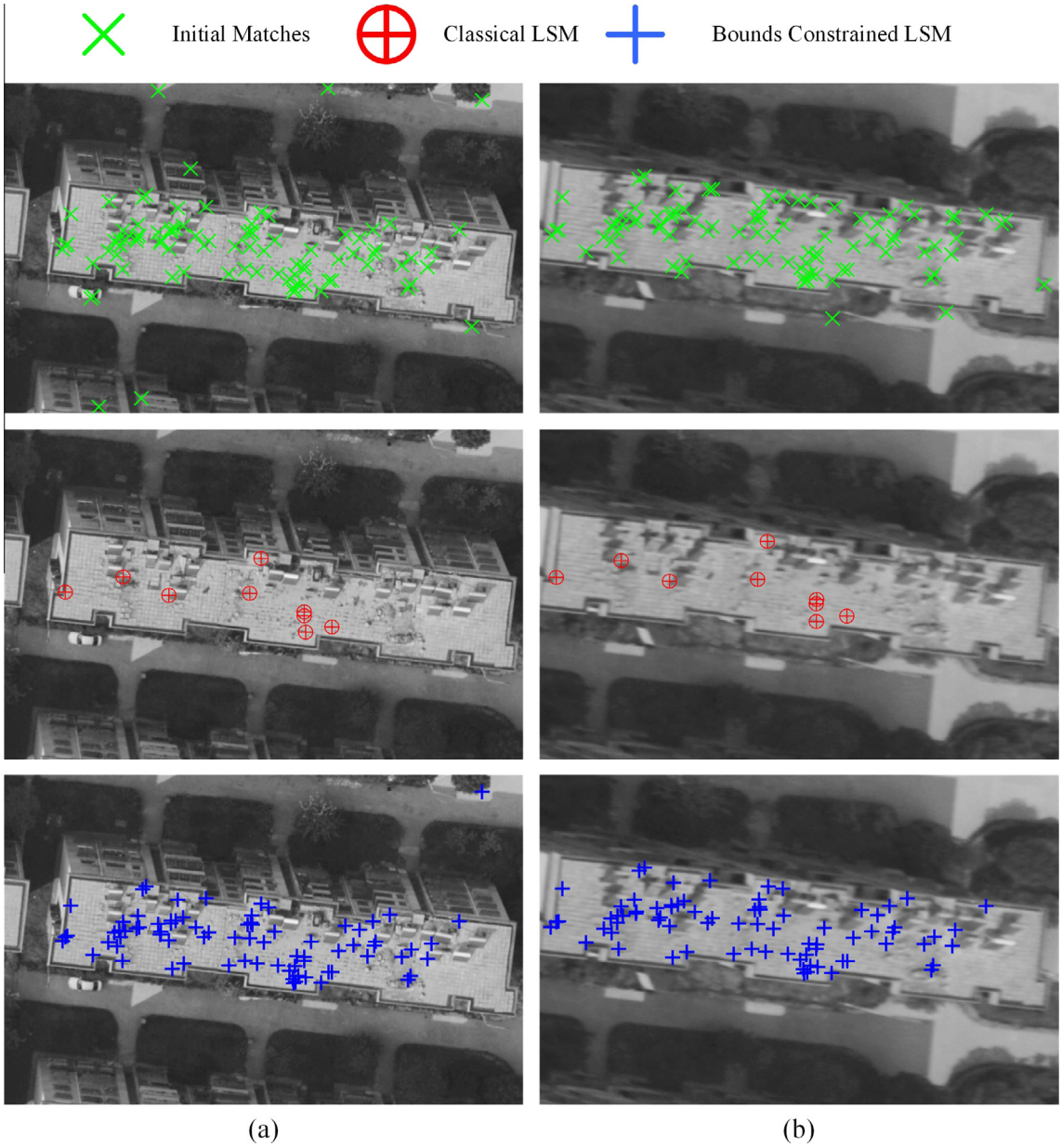

Fig. 9. Comparison of matching results on a building roof in the urban area of Jinyang. (a) Matching points overlaid on the rectified nadir image and (b) matching results overlaid on the rectified oblique image. The three rows show the results of initial matches using classical LSM and the proposed method.

3.3. Robust loss function

The optimization problem defined in Eq. (4) can be substituted in the following form,

$$\begin{aligned} &\min_{x} \sum_{r,c} \rho(c(\boldsymbol{x})) \\ &c(\boldsymbol{x}) = I_1(r,c) - k_1 I_2(r',c') - k_2 \\ &\rho(s) = s^2/2 \end{aligned} \tag{6}$$

where $c(\boldsymbol{x})$:$R^n \rightarrow R$, is the cost function and $\rho(s)$ is the square loss function. The cost function actually determines the penalty for one pixel, which is, in LSM, the pixel difference. The loss function determines how much loss will be imposed as the penalty for aggregating into the final minimization problem, and for the least-squares method, this is the square of the difference.

After defining LSM by the form in Eq. (6), it is intuitive to change the cost and loss functions to a more robust form. For example, we could change the cost function using the gradient $G(r, c)$ for the two images rather than the original grayscale images as $c(\boldsymbol{x}) = G_1(r, c) - k_1 G_2(r', c') - k_2$, as used by Campbell and Wu (2008), which has been proven to be equivalent to classical LSM. As mentioned previously, another attempt has been made, similar to adopting the NCC as the cost function, in the work of Zheltov and Sibiryakov (1997). However, in the end, the authors also declaimed equivalency to the SSD approach in classical LSM. Therefore, in this work, we do not attempt to amend the cost function.

In fact, when using the square differences as the loss function, it is implicitly assumed that the noise modeled in Eq. (1), which is added to each observation, is multivariate Gaussian distributed (Szeliski, 2011). This assumption is appropriate if the noise modeled in each pixel is only the result of tiny random errors, such

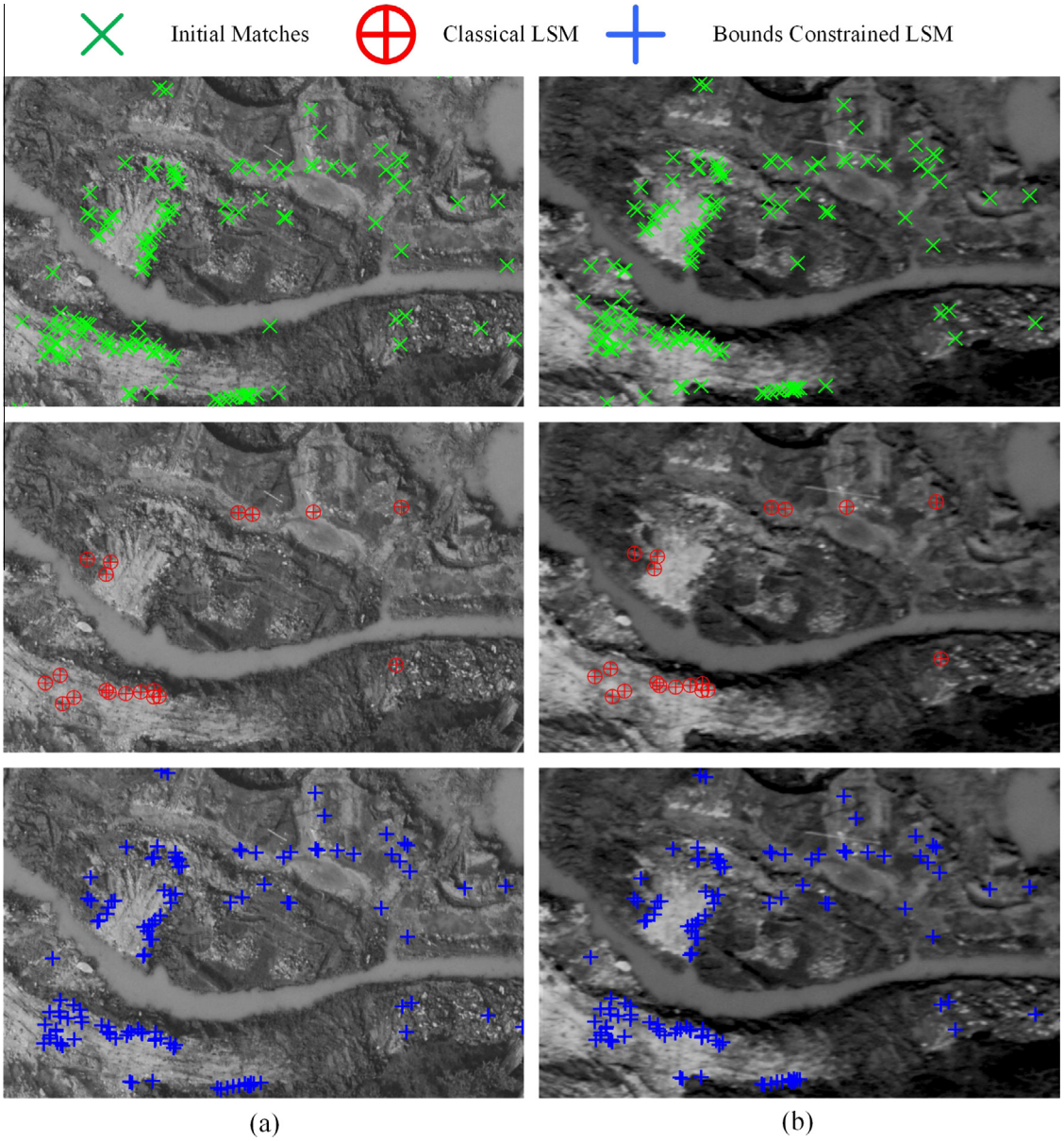

Fig. 10. Comparison of matching results on bare earth in the rural area of Jinyang. (a) Matching points overlaid on the rectified nadir image and (b) matching results overlaid on the rectified oblique image. The three rows show results of initial matches using classical LSM and the proposed method.

as in the scenario of nadir images, where the image differences are not quite significant. However, in the case of oblique images, the imaging quality may be significantly different; the noise in Eq. (1) may not follow the Gaussian distribution due to gross errors or outliers. To account for this situation, the robust loss function can be adopted. For example, an early attempt to solve this problem consisted of determining the truncated square loss as $\rho(x) = \min(x^2, V)$ (Szeliski, 2011), which simply extrudes the square loss to the maximal threshold V. Another example consisted of representing the loss with the L_1 norm as $\rho(x) = |x|$, which is also termed the total variation. To handle the problem of potential outliers in the LSM window, we adopted the Huber loss (Hastie et al., 2009) function as shown below,

$$\rho(s) = \begin{cases} s^2/2, & |s| \leqslant a \\ a|s| - a^2/2, & |s| > a \end{cases} \tag{7}$$

where a is an input parameter that reflects the outlier threshold. Fig. 4 compares the square loss and the Huber loss. It can be noted that, in the region of small cost, s, Huber loss is identical to the square loss, and thus retains the merits of a least-squares solver. In the region of large cost, which is probably caused by outliers, the losses only increase linearly with the cost rather than quadratically with the square of the loss. The outlier region is determined by the parameter a, and in this work, $a = 20$ is the value that is chosen. Therefore, the influences of outliers are suppressed.

After describing both the bound constraints and the robust loss function, we could formally reformulate the LSM problem as the following,

Bounds Constrained Least-squares Matching: Given two images, $I_1(r, c)$ and $I_2(r, c)$, in a small window, LSM is used to find the optimal solution, $\boldsymbol{x} = [a_{11}, a_{12}, a_{13}, a_{21}, a_{22}, a_{23}, k_1, k_2]^T$, to the following bound constrained non-linear optimization problem,

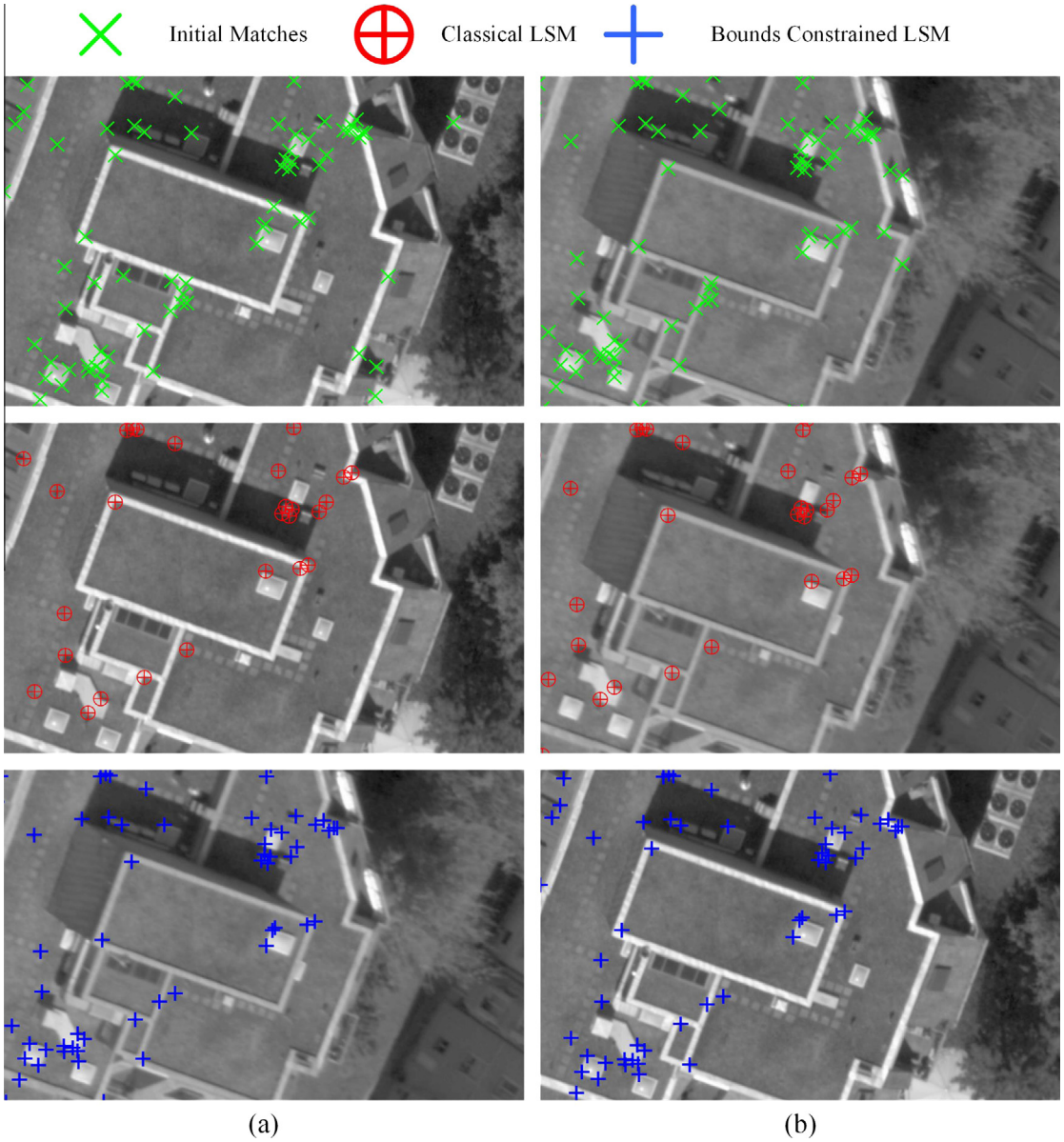

Fig. 11. Comparison of matching results on a roof top in Zurich. (a) Matching points overlaid on the rectified nadir image and (b) matching results overlaid on the rectified oblique image. The three rows show the results of initial matches using classical LSM and the proposed method.

$$
\begin{aligned}
&\min_x f(\boldsymbol{x}) = \sum_{r,c} \rho(c(\boldsymbol{x})) \\
&c(\boldsymbol{x}) = I_1(r,c) - k_1 I_2(r',c') - k_2 \\
&\rho(s) = \begin{cases} s^2/2, & |s| \leqslant a \\ a|s| - a^2/2, & |s| > a \end{cases} \\
&s.t.\ 1-\delta_1 < a_{11} < 1+\delta_1, 1-\delta_1 < a_{22} < 1+\delta_1 \\
&\quad -\delta_2 < a_{12} < \delta_2, -\delta_2 < a_{21} < \delta_2 \\
&\quad -\delta_3 < a_{13} < \delta_3, -\delta_3 < a_{23} < \delta_3 \\
&\quad \delta_4 < k_1 < 1/\delta_4, -\delta_5 < k_2 < \delta_5
\end{aligned} \tag{8}
$$

3.4. Solver for bound constrained nonlinear optimization

Unlike the least-squares problem defined in Eq. (3) and the general purpose optimization problem in Eq. (4), which can be solved by the traditional first order gradient descent method or the second order Newton method (Boyd and Vandenberghe, 2004), solving the bounds constrained non-linear optimization problem in Eq. (8) is a non-trivial task. Fortunately, for this problem, several publicly available solvers exist (Neumaier, 2014; Mittelmann, 2015), such as the bounds constrained version of limited memory BFGS algorithm (L-BFGS-B), active set algorithm with conjugate gradient approach (ASA-CG) and Levenberg–Marquardt (LM) methods. The L-BFGS-B is a quasi-Newton method and thus only requires the first order partial derivatives to simulate the second order Hessian matrix. It was originally proposed by Byrd et al. (1995) and was recently amended and optimized by Morales and Nocedal (2011). The latter formulation, which is a gradient projection method, was proposed by Hager and Zhang (2006), who claimed that it was more stable than the L-BFGS-B. The last method is a trust region approach and can be extended to the bounded constrained version. We have tested all of the above solvers and have chosen the bounds constrained routine of the LM package using openblas (Zhang, 2015) as the underlying linear solver because it provides a good balance between accuracy and runtime.

Table 3
Comparison of the convergent performances for the classical solver and the bounds-constrained solvers. #all represents the average iteration for all candidates, #converged represents the average iteration for the converged matches and runtime represents the time costs of all the candidates. The maximum iteration is set to 30.

Pair	Classical solver			Bound constrained solver		
	#all	#converged	runtime (ms)	#all	#converged	runtime (ms)
(a) Tests in the urban area of Jinyang, which features dense buildings						
E–A	11.308	3.403	2995	2.924	2.924	2163
E–B	11.823	4.714	2955	2.896	2.896	2206
E–C	9.982	3.649	2676	2.897	2.897	1982
E–D	12.029	4.181	3685	2.935	2.935	2516
A–C	6.661	2.920	945	2.911	2.911	989
B–D	8.919	3.550	1702	2.928	2.928	1489
E–E	14.109	5.670	15,059	2.981	2.981	8799
(b) Tests in the rural area of Jinyang, which features bare earth, vegetation and roads						
E–A	14.202	4.066	4997	2.961	2.961	2827
E–B	16.130	4.524	3855	2.960	2.960	2052
E–C	15.304	5.383	4315	2.951	2.951	2858
E–D	14.049	5.178	4573	2.954	2.954	2559
A–C	8.838	3.784	749	2.926	2.926	694
B–D	8.930	4.003	461	2.945	2.945	374
E–E	17.258	6.997	25032	2.990	2.990	12,427
(c) Tests in the residential area of Zurich, which features complex combinations						
E–A	9.042	4.518	1926	2.940	2.940	1665
E–B	10.337	4.916	1791	2.952	2.952	1395
E–C	8.095	4.051	1259	2.935	2.935	1130
E–D	9.548	4.378	1598	2.948	2.948	1289
A–C	5.428	2.615	664	2.927	2.927	736
B–D	6.568	2.899	655	2.930	2.930	589
E–E	9.152	5.794	5127	2.972	2.972	4045

To incorporate the L-BFGS-B and ASA-CG into LSM, we only have to feed the subroutines that calculate the loss value $f(\boldsymbol{x})$ and the gradient $df(\boldsymbol{x})/d\boldsymbol{x}$ into the solver, and we do not need to optimize details inside the solver. The loss value is defined in Eq. (8) and the Jacobian are calculated as the following,

$$\frac{df(\boldsymbol{x})}{d\boldsymbol{x}} = \sum_{r,c} \frac{d\rho(c(\boldsymbol{x}))}{dc(\boldsymbol{x})} \frac{dc(\boldsymbol{x})}{d\boldsymbol{x}} \tag{9}$$

where in this case $\frac{d\rho(c(\boldsymbol{x}))}{dc(\boldsymbol{x})} = c(\boldsymbol{x})$; if $|c(\boldsymbol{x})| < a$ and $c(\boldsymbol{x}) > a$, a is used; and $-a$ is used if $c(\boldsymbol{x}) < a$. For the latter part, $\frac{dc(\boldsymbol{x})}{d\boldsymbol{x}}$, the Jacobian is the same as that discussed in Gruen (1985). For the LM method, because it only accepts functions that are expressed as the sum of square loss, to incorporate the Huber loss, we have to modify the optimization function in Eq. (8) as $f(\boldsymbol{x}) = \sum_{r,c}\left(\sqrt{\rho(c(\boldsymbol{x}))}\right)^2$ and provide the cost and Jacobian for each pixel as described above.

Another *ad hoc* amendment for the solvers is the termination criterion. In general, three termination criteria are adopted in most of the optimization solvers, including the projected gradient tolerance, the function tolerance and the parameter tolerance. The first criterion is used in most of the solvers discussed above (Byrd et al., 1995; Hager and Zhang, 2006) and is in the increments of the parameter after it is projected to the axis defined by the last value $\boldsymbol{x}_k$. The second criterion stops the optimization procedure if the difference of the cost functions, as defined in Eq. (9), between the iterations is less than a certain proportion of $f(\boldsymbol{x})_k$. Similarly, we can define the last criterion, parameter tolerance. However, we have found that the thresholds for these criteria are obscure because they only account for numerical significances, and the default values set in these solvers require too many iterations or do not converge for the proposed LSM optimization in Eq. (8). Therefore, we proposed another termination criterion based on the shape of the affine transformation of the square matching window, which is works cooperatively. We applied the affine transformation in Eq. (2) to the four corners, (0, 0), (0, w−1), (w−1, w−1), (w−1, 0), where w is the width of the LSM window, and recorded the coordinates between the iterations. If the maximal movement of the corners is smaller than a threshold, then the optimization is stopped and considered a convergence. However, for aerial applications, the potential accuracy for LSM is approximately 0.1–0.5 pixels (Gruen, 2012), and the threshold can be assigned accordingly.

4. Experimental results and analyses

4.1. Test data descriptions

To evaluate the performances of the proposed method, we conducted experiments on two test areas with different oblique aerial camera systems as shown in Fig. 5. The first dataset covers a newly developed city area in China, Jinyang, with a domestic oblique camera system – SWDC-5. The SWDC-5 is a typical penta-view camera system using the "Maltese Cross" configuration, as detailed in previous studies (Petrie, 2009; Lemmens, 2014b), which includes a nadir camera and four oblique cameras of 45 degrees. The flight lines were designed to be from north to south. We denote the nadir view as *CamE*, the cross-flight views as *CamA* and *CamC* for east and west directions, the along-flight views as *CamB* and *CamD* for south and north directions, respectively. The GSDs were 8 cm for the nadir images and 6–20 cm for the oblique images. The designed GSD was obtained by controlling the principle distances for the nadir and oblique cameras, which were approximately 50 mm and 80 mm. To test the performances and efficiencies of the innovative LSM method in real world applications, we selected two test areas, which include a built-up area and a rural area. In each area, there were 5 images from each camera covering the same region, as shown in Fig. 5a and b.

The other dataset covers the Zurich area, which were collected by a penta-view Leica RCD30 in the EuroSDR/ISPRS benchmark project (Cavegn et al., 2014). There are two major differences between the SWDC-5 and RCD30: (a) the tilting angle of the RCD30 is approximately 35 degrees rather than 45 degrees for SWDC-5 and (b) the principal distances for both the nadir and obli-

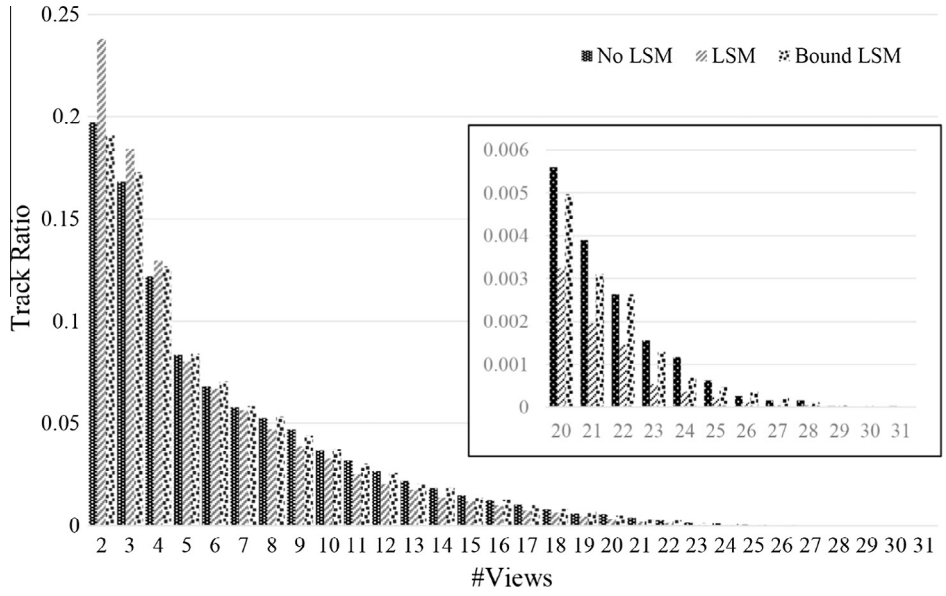

Fig. 12. Statistics of the tracks with three different strategies. "Track Ratio" denotes the number of views divided by the number of total tracks and "#views" denotes the number of covered views for each track. The enlarged figure shows results with "#views" larger than 20. The average numbers of views are 6.248, 5.585 and 6.197 for the strategies with no LSM, traditional LSM and bounds constrained LSM, respectively.

que images of the RCD30 are the same (53 mm). We also selected 5 images covering the same area as that shown in Fig. 5c. In order to summarize the differences, Table 1 compares some of the major features of these two datasets, including tilting angles, focal lengths, usages of forward motion compensation technique, GSDs and the image sizes.

4.2. Matching results

The matching workflow is as the following. First, the images are all pre-rectified to relieve the perspective deformations before the matching procedure, as described in our previous paper (Hu et al., 2015). Then, we detected the FAST corners (Rosten et al., 2010) that served as the interest points, as suggested by Jazayeri and Fraser (2010), and described them using BRIEF (Binary Robust Independent Elementary Features) descriptors (Calonder et al., 2012) for efficiency consideration. Then, the best matches were searched forward and backward using FLANN (Fast Library for Approximate Nearest Neighbors) (Muja and Lowe, 2009), and we only retained the matches that passed the cross check. Before LSM, we located the correspondences with the maximum NCC in a small search window, as suggested by Lerma et al. (2013), to determine the initial integer pixel locations. Furthermore, we only reserved the correspondences that created an NCC larger than a certain threshold (0.8 was used). In addition, the points were refined by LSM to locate the correspondences of sub-pixel accuracy or they were removed if LSM was not convergent. We removed potential outliers with the RANSAC approach using pair-wise epipolar geometry (Hartley and Zisserman, 2004) and spatial outlier filters (Hu et al., 2015).

Table 2 shows performance comparison of classical LSM and the proposed bound constrained version. In this comparison, we performed experiments using several scenarios, including nadir–oblique (row 1–4), oblique–oblique (row 5 and 6) and nadir–nadir (row 7) configurations. The number of matches after the cross check of the nearest match using the descriptors (#Cross check), the number that exceeded the NCC threshold (#NCC), the number that were convergent in LSM (#Convergences) and the number after outlier removal (#Filtered) are shown in Table 2. For the Jinyang dataset, as shown in parts (a) and (b) in Table 2, it can be noted that we can reach convergent LSM matches for all of the matching candidates after NCC screening, compared to only 20–50% using classical solvers in the oblique cases. After outlier detection, the improvement of matches is more than two-fold in these cases. In addition, the superior performance of the proposed method is consistent in all the cases, including scenarios between nadir and oblique cameras, where the translation tilt (Morel and Yu, 2009) angle is approximately 45°, and scenarios between the oblique cameras, where the translation tilt angle is approximately 90°. This finding has confirmed that the proposed method is much more robust and stable than classical LSM solvers.

For each match pair, we also recorded the average NCC value after convergent LSM matches, denoted as Avg. NCC in Table 2. An interesting finding is that the average NCC of classical LSM is consistently higher than that of the proposed solver, although both methods showed higher NCC values after sub-pixel localization. A reasonable explanation is that the classical solver can only reach convergence in cases where the appearances of the two matching area are almost identical, and the proposed solver can also handle cases with significant blurriness, illumination differences and, likely, noise. Another piece of evidence is that in the case of the nadir-nadir scenario, the bounds-constrained solver can also produce NCC values as well as the traditional solver. In this case, the classical solver can also achieve a convergence rate of approximately 90%.

Similar results were observed for the test area in the Zurich, as shown in the part (c) of Table 2; however, there were some slight differences that should be mentioned due to the configuration of the camera system. We noticed that classical LSM produced better results in this test than the test using the SWDC-5 system in Jinyang. The convergent ratios for the nadir-oblique matching scenarios were greater than 50% for almost all the cases. Furthermore, the average NCC value was significantly higher both before and after LSM. This can be explicitly explained by the better image qualities due to smaller tilting angles and the adoption of forward motion compensation technique, which will alleviate the blur effect.

To further compare the behavior of classical LSM and the proposed method, we also exported the matches and overlaid the correspondences on the images. Figs. 6–8 show the overall matching

Table 4
Comparison of the BA results of the three strategies. For each row, we record the mean square error of unit weight and Root Mean Square Errors (RMSE).

Strategy	σ_0 (pixels)	RMSE X (m)	RMSE Y (m)	RMSE Z (m)	RMSE XYZ (m)
(a) Results in the urban area					
No LSM	0.64	0.10	0.10	0.15	0.21
LSM	0.48	0.07	0.07	0.11	0.15
Bound LSM	0.48	0.07	0.06	0.12	0.15
(b) Results in the rural area					
No LSM	0.84	0.14	0.15	0.22	0.30
LSM	0.65	0.09	0.10	0.15	0.20
Bound LSM	0.65	0.10	0.10	0.16	0.21

results with the two methods. For the Jinyang dataset, in Figs. 6 and 7, it can be noted that although classical LSM could also produce relatively evenly distributed matches, we can obtained a better distribution, especially in the boundary areas. This occurred because, in these areas, the image deformations were more severe. However, the improvement in the distribution of matches was less significant in the Zurich area, as shown in Fig. 8. This is because the configuration is different; the footprint of the nadir view was completely covered by the oblique view, and the deformation problem might have been less severe due to the smaller tilting angles.

Furthermore, when we zoomed into a small local area, as shown in Figs. 9–11, we found that the proposed LSM method recalls more matches than classical LSM. It can be noted that oblique images in the right column will generally show more blurriness than nadir images as the reason of larger tilting angles and larger GSD. Due to this blurriness, the classical method can achieve only a very few correct matches. As detailed in Table 2, we can obtain 100% convergences, and after outlier filtering, a large amount of correct matches still remain. The additional correspondences are quite important for generating multiview tie points, which are essential for reliable bundle adjustment (BA).

4.3. Runtime analysis

To further evaluate the performances of the bounds-constrained solver for LSM, we demonstrated the convergence rate in the two test areas. Regarding the convergence rate, we were interested in the average iterations for matching pairs. Additionally, the runtime performance is also an important consideration for practical applications. The performances for the two test areas are detailed in Table 3, in which part (a) describes results for the urban areas in Jinyang, part (b) for the rural areas in Jinyang and part (c) for the Zurich area. Three metrics were recorded for the classical and the proposed methods, which included the average iterations for the converged LSM matches, the average iterations for all of the correspondences after NCC screening and the runtime performances. The runtime performances indicate the overall time consumed in a parallel execution time with 12×2.67 GHz CPU cores.

First, for the iteration statistics, it can be noted that the number of iterations of the classical solver is only a slightly inferior (3–5 iterations) to its counterpart (approximately 3 iterations); however, for the total average iterations, the performance is much worse (larger than 10 iterations). This occurs because, although classical solver may handle some simpler cases well enough, for challenging scenes, it may not converge and thus reach the maximum number of iterations, which is set to 30 as the default value in many photogrammetry software programs. Furthermore, we have set an early termination step, when the corners exceed twice the LSM window. Second, the runtime may be theoretically proportional to the average number of iterations; however, in our experience, the runtime is only slightly better than that of the classical solver. This is due to the underlying line search steps in the solvers, which may adaptively detect the step size and make the iterations slower but more efficient. Considering the above analyses, the proposed solver will produce better results as well as slight efficiency improvements.

4.4. Bundle adjustment analysis

In order to compare the influences of the LSM on the succeeded BA of oblique images, we have conducted bundle adjustment in two test areas of Jinyang, including an urban and a rural areas. The urban and rural areas have 140 and 135 images, respectively, where the EO parameters for the nadir images have already been oriented with integrated sensor orientation and thus kept fixed in this experiment. All the matching steps are the same as described in Section 4.2, except for the LSM stage: three different strategies are tested, including without LSM (nor NCC thresholding), with traditional LSM and with the bound constrained version. Furthermore, in order to conduct bundle adjustment, pairwise matches are connected into tracks using the connected components method (Agarwal et al., 2011), where one track contains multiple tie points of the same object point. And for each image, the tie points are gridded, in each grid only one track is selected and we prefer tracks covering more views, because these points tend to be more salient and robust.

The number of views of the tracks are recorded and normalized with the total number of tracks, because for different strategies the total tracks selected are not the same. Fig. 12 compares the three different strategies, and it can be noted that classical LSM has higher ratio for tracks with lower overlaps (from 2 to 4) and consistently smaller ratio for higher overlaps. And this can also be reflected by the average number of overlapped views, which are 6.248, 5.585 and 6.197 for the strategies with no LSM, traditional LSM and bounds constrained LSM, respectively. Another interesting finding is that, after NCC thresholding, the average overlapping number is still comparable to the strategy without NCC and LSM. NCC is still a powerful matching criterion for oblique images as long as the images are globally rectified to remove geometrical deformations.

For the evaluation of BA results, the EO parameters of the nadir images have already been adjusted and only the EO parameters for the oblique images are estimated. The theoretical accuracies for the EO parameters of the nadir images are at the level of a few centimeters and thousandths of a degree for position and rotation angles, respectively. Because no checkpoints are available, we use 4-folded tie points covering 2 nadir images and 2 oblique images, in which triangulated object points using nadir images serve as check points. The mean square error of unit weight (σ_0) obtained by the BA and the Root Mean Square Error (RMSE) of the check points are summarized in Table 4. It can be noted that the BA results without LSM are inferior to the two counterparts using LSM with regard to both σ_0 and RMSE of check points, at least for features without subpixel localization. Although it's more convenient and fast to match images with feature descriptor only, this

experiment demonstrates that the additional integer pixel localization with NCC and subpixel refinement with LSM is worthful. In fact, the additional matching stages using NCC and LSM are also suggested by Lerma et al. (2013) for close range photogrammetry, who have reported about twofold improvement in those cases compared to about 20–50% for our experiments with aerial oblique images. However, it should be also noted that there are almost no differences on the BA qualities between traditional LSM and the bounds constrained LSM. This is because that the bounds constrained LSM is not intended to improve the matching accuracy of LSM when convergent, but improve the robustness. The difference may appear when the network of tie points is weak enough to influence the BA, e.g. when the matches are too sparse to form a strong tie point connection.

5. Conclusions

This paper proposed a convergence assured LSM method by explicitly assigning lower and upper boundaries of the parameters with physical significance in the optimization. In this manner, the search space is reduced, and the unknown parameters will not stray from the true value. Furthermore, robust estimation is also used to handle possible outliers in the gray values of pixels to mitigate the influence of noise. As described above, we redefined LSM in the form of a general purpose bounds-constrained optimization and solved it using publicly available solvers. The proposed method is able to guarantee convergence of LSM when the initial matches have an appropriate NCC value and located in an integer pixel with local maximum NCC response. A slight improvement in runtime can be expected due to the faster convergence. Experimental evaluations have proven the effectiveness of the proposed method on two datasets with penta-view oblique configurations that involves different tilting angles and principal distance configurations. Another interesting finding of this paper is that decreasing the tilting angle of the oblique views will relieve the challenges in feature point matching, however at the cost of worse visibility of building facades. And the tilting angle should be carefully designed in cooperation with the intended application and the flight design. Furthermore, because the convergence of LSM is directly related with image quality, some techniques to improve the image quality, such as the forward motion compensation enabled in several oblique camera systems to remove the blur effect, may also improve the convergence of LSM.

This paper deals with manned aerial oblique images with systematic image acquisition, however, a rising interest in the photogrammetric community is the processing of oblique UAV images and terrestrial images, which also feature oblique views, but are unordered and contain large amount of images; these types of data require specific methods to determine the image clusters and efficient matching selection algorithms (Schindler et al., 2015). Furthermore, this paper only addresses the feature matching problem, and in order to exploit oblique images for 3D reconstruction dense image matching (DIM) methods are required. Because DIM and feature matching are complementary with each other, the combination of different matching strategies may produce more desirable results with better preserved sharp edges and less noises (Zhang, 2005; Remondino et al., 2014).

Acknowledgements

This study was supported by the National Natural Science Foundation of China (Nos. 41471320 and 41501421) and we would also like to extend our thanks to the ISPRS/EuroSDR Benchmark on Dense Image Matching (Cavegn et al., 2014) for providing the test datasets in Zurich.

References

Ackermann, F., 1984. Digital image correlation: performance and potential application in photogrammetry. Photogrammetric Rec. 11 (64), 429–439.
Agarwal, S., Furukawa, Y., Snavely, N., Simon, I., Curless, B., Seitz, S.M., Szeliski, R., 2011. Building Rome in a day. Commun. ACM 54 (10), 105–112.
Althof, R.J., Wind, M.G., Dobbins, J.T., 1997. A rapid and automatic image registration algorithm with subpixel accuracy. IEEE Trans. Med. Imaging 16 (3), 308–316.
Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., 2008. Speeded-up robust features (SURF). Comput. Vis. Image Underst. 110 (3), 346–359.
Birchfield, S., Tomasi, C., 1998. A pixel dissimilarity measure that is insensitive to image sampling. IEEE Trans. Pattern Anal. Mach. Intelligence 20 (4), 401–406.
Björck, Å., 1996. Nonlinear least square problem. In: Björck, Å. (Ed.), Numerical Methods for Least Squares Problems. Society for Industrial and Applied Mathematics, Philadelphia, USA, pp. 339–358.
Boyd, S.P., Vandenberghe, L., 2004. Convex Optimization, first ed. Cambridge University Press, Cambridge, p. 727.
Byrd, R.H., Lu, P., Nocedal, J., Zhu, C., 1995. A limited memory algorithm for bound constrained optimization. SIAM J. Sci. Comput. 16 (5), 1190–1208.
Calonder, M., Lepetit, V., Ozuysal, M., Trzcinski, T., Strecha, C., Fua, P., 2012. BRIEF: computing a local binary descriptor very fast. IEEE Trans. Pattern Anal. Mach. Intelligence 34 (7), 1281–1298.
Campbell, N.A., Wu, X., 2008. Gradient cross correlation for sub-pixel matching. Int. Arch. Photogrammetry, Remote Sensing Spatial Inf. Sci. 37 (Part B7), 1065–1070.
Cavegn, S., Haala, N., Nebiker, S., Rothermel, M., Tutzauer, P., 2014. Benchmarking high density image matching for oblique airborne imagery. Int. Arch. Photogrammetry, Remote Sensing Spatial Inf. Sci. XL-3, 45–52.
Chen, Q., Defrise, M., Deconinck, F., 1994. Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition. IEEE Trans. Pattern Anal. Mach. Intelligence 16 (12), 1156–1168.
Debella-Gilo, M., Kääb, A., 2011. Sub-pixel precision image matching for measuring surface displacements on mass movements using normalized cross-correlation. Remote Sensing Environ. 115 (1), 130–142.
Foroosh, H., Zerubia, J.B., Berthod, M., 2002. Extension of phase correlation to subpixel registration. IEEE Trans. Image Process. 11 (3), 188–200.
Förstner, W., 1982. On the geometric precision of digital correlation. Int. Arch. Photogrammetry, Remote Sensing Spatial Inf. Sci. 24 (3), 176–189.
Fraser, C., 1997. Innovations in automation for vision metrology systems. Photogrammetric Rec. 15 (90), 901–911.
Gruen, A., 1985. Adaptive least squares correlation: a powerful image matching technique. S. Afr. J. Photogrammetry, Remote Sensing Cartography 14 (3), 175–187.
Gruen, A., 2012. Development and status of image matching in photogrammetry. Photogrammetric Rec. 27 (137), 36–57.
Gruen, A., Baltsavias, E., 1988. Geometrically constrained multiphoto matching. Photogrammetric Eng. Remote Sensing 54 (5), 633–641.
Hager, W.W., Zhang, H., 2006. A new active set algorithm for box constrained optimization. SIAM J. Optimization 17 (2), 526–557.
Harris, C., Stephens, M., 1988. A combined corner and edge detector. In: Proc. Alvey Vision Conference, Manchester, UK, 31 August – 2 September, pp. 147–152.
Hartley, R., Zisserman, A., 2004. Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge, UK, p. 672.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning, Second ed. Springer, Berlin, Germany, p. 763.
Hu, H., Zhu, Q., Du, Z., Zhang, Y., Ding, Y., 2015. Reliable spatial relationship constrained feature point matching of oblique aerial images. Photogrammetric Eng. Remote Sensing 81 (1), 49–58.
Ip, A., El-Sheimy, N., Mostafa, M., 2007. Performance analysis of integrated sensor orientation. Photogrammetric Eng. Remote Sensing 73 (1), 1–9.
Jazayeri, I., Fraser, C.S., 2010. Interest operators for feature-based matching in close range photogrammetry. Photogrammetric Rec. 25 (129), 24–41.
Lemmens, M., 2014a. Oblique imagery: the standard for mapping. GIM Int. 28 (3), 14–17.
Lemmens, M., 2014b. Digital oblique aerial cameras (1). GIM Int. 28 (4), 20–25.
Lerma, J.L., Navarro, S., Cabrelles, M., Seguí, A.E., Hernández, D., 2013. Automatic orientation and 3D modelling from markerless rock art imagery. ISPRS J. Photogrammetry Remote Sensing 76, 64–75.
Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60 (2), 91–110.
Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L., 2005. A comparison of affine region detectors. Int. J. Comput. Vision 65 (1–2), 43–72.
Mittelmann, H.D., 2015. Decision Tree for Optimization Software, <http://plato.asu.edu/guide.html> (Accessed 30 June, 2015).
Morales, J.L., Nocedal, J., 2011. Remark on "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound constrained optimization". ACM Trans. Math. Software (TOMS) 38 (1), 7:1–7:4.
Moré, J.J., Thuente, D.J., 1994. Line search algorithms with guaranteed sufficient decrease. ACM Trans. Math. Software (TOMS) 20 (3), 286–307.
Morel, J., Yu, G., 2009. ASIFT: a new framework for fully affine invariant image comparison. SIAM J. Imaging Sci. 2 (2), 438–469.
Morgan, G.L.K., Liu, J.G., Yan, H., 2010. Precise subpixel disparity measurement from very narrow baseline stereo. IEEE Trans. Geosci. Remote Sensing 48 (9), 3424–3433.

Mostafa, M.M.R., Schwarz, K., 2001. Digital image georeferencing from a multiple camera system by GPS/INS. ISPRS J. Photogrammetry Remote Sensing 56 (1), 1–12.

Muja, M., Lowe, D.G., 2009. Fast approximate nearest neighbors with automatic algorithm configuration. In: Proc. International Conference on Computer Vision Theory and Application (VISSAPP2009). INSTICC Press, Lisboa, Portugal, pp. 331–340, 5–8 February.

Neumaier, A., 2014. Local Optimization Software. <http://www.mat.univie.ac.at/~neum/glopt/ software_l.html#uncon> (Accessed 11 September, 2014).

Nex, F., Gerke, M., Remondino, F., Przybilla, H.J., Bäumker, M., Zurhorst, A., 2015. ISPRS Benchmark for multi-platform photogrammetry. ISPRS Ann. Photogrammetry, Remote Sensing Spatial Inf. Sci. II-3/W4, 135–142.

Petrie, G., 2009. Systematic oblique aerial photography using multiple digital cameras. Photogrammetric Eng. Remote Sensing 75 (2), 102–107.

Remondino, F., 2006. Detectors and descriptors for photogrammetric applications. Int. Arch. Photogrammetry, Remote Sensing Spatial Inf. Sci. XXXVI (Part3), 49–54.

Remondino, F., Spera, M.G., Nocerino, E., Menna, F., Nex, F., 2014. State of the art in high density image matching. Photogrammetric Rec. 29 (146), 144–166.

Rosten, E., Porter, R., Drummond, T., 2010. Faster and better: a machine learning approach to corner detection. IEEE Trans. Pattern Anal. Mach. Intelligence 32 (1), 105–119.

Rupnik, E., Nex, F., Remondino, F., 2014. Oblique multi-camera systems – orientation and dense matching issues. Int. Arch. Photogrammetry, Remote Sensing Spatial Inf. Sci. 38 (W1), 107–114.

Rupnik, E., Nex, F., Toschi, I., Remondino, F., 2015. Aerial multi-camera systems: accuracy and block triangulation issues. ISPRS J. Photogrammetry Remote Sensing 101, 233–246.

Schindler, K., Hartmann, W., Havlena, M., 2015. Recent developments in large-scale tie-point search. In: Proc. 55th Photogrammetric Week (PHOWO2015), Stuttgart, Germany, September 7–11, pp. 175–182.

Shortis, M.R., Clarke, T.A., Robson, S., 1995. Practical testing of the precision and accuracy of target image centering algorithms. In: Proc. Photonics East'95, International Society for Optics and Photonics, pp. 65–76.

Shortis, M.R., Clarke, T.A., Short, T., 1994. Comparison of some techniques for the subpixel location of discrete target images. In: Proc. Photonics for Industrial Applications, International Society for Optics and Photonics, pp. 239–250.

Stone, H.S., Orchard, M.T., Chang, E., Martucci, S.A., 2001. A fast direct Fourier-based algorithm for subpixel registration of images. IEEE Trans. Geosci. Remote Sensing 39 (10), 2235–2243.

Szeliski, R., 2011. Computer Vision: Algorithms and Applications, first ed. Springer, London, p. 812.

Szeliski, R., Scharstein, D., 2004. Sampling the disparity space image. IEEE Trans. Pattern Anal. Mach. Intelligence 26 (3), 419–425.

Tian, Q., Huhns, M.N., 1986. Algorithms for subpixel registration. Comput. Vision, Graphics, Image Process. 35 (2), 220–233.

Wang, Z., 1990. Principle of Photogrammetry: with Remote Sensing. Press of Wuhan Technical University of Surveying and Mapping, Wuhan, p. 455.

Wiedemann, A., Moré, J., 2012. Orientation strategies for aerial oblique images. Int. Arch. Photogrammetry, Remote Sensing Spatial Inf. Sci. 39 (Part B1), 185–189.

Wong, K.W., Lew, M., Wiley, A.G., 1988. 3D metric vision for engineering construction. Int. Arch. Photogrammetry Remote Sensing 27 (B5), 647–656.

Xiao, J., Gerke, M., Vosselman, G., 2012. Building extraction from oblique airborne imagery based on robust façade detection. ISPRS J. Photogrammetry Remote Sensing 68, 56–68.

Xiong, X., Zhang, Y., Zhu, J., Zheng, M., 2014. Camera pose determination and 3-D measurement from monocular oblique images with horizontal right angle constraints. IEEE Geosci. Remote Sensing Lett. 11 (11), 1976–1980.

Xu, L., Jia, J., Kang, S.B., 2012. Improving sub-pixel correspondence through upsampling. Comput. Vision Image Understanding 116 (2), 250–261.

Yu, Z., Gang, C., Che, R., Liu, C., Wei, T., 2002. Bilinear interpolation centroid algorithm used for circular optical target location. In: Proc. Second International Conference on Image and Graphics, International Society for Optics and Photonics, pp. 333–339.

Zhang, L., 2005. Automatic digital surfece model (DSM) generation from linear array images. PhD. Thesis, ETH, Zürich, Switzerland, 199 p.

Zhang, X., 2015. OpenBLAS: An optimized BLAS library, <http://www.openblas.net/> (Accessed 16th January, 2016).

Zhang, Y., Shen, X., 2013. Direct georeferencing of airborne LiDAR data in national coordinates. ISPRS J. Photogrammetry Remote Sensing 84, 43–51.

Zheltov, S., Sibiryakov, A., 1997. Adaptive subpixel cross-correlation in a point correspondence problem. In: Proc. 4th Conference on Optical 3-D Measurement Techniques, Zurich, Switzerland, 29 September – 2 October, pp. 86–95.

Zhu, Q., Wu, B., Wan, N., 2007. A sub-pixel location method for interest points by means of the Harris interest strength. Photogrammetric Rec. 22 (120), 321–335.