

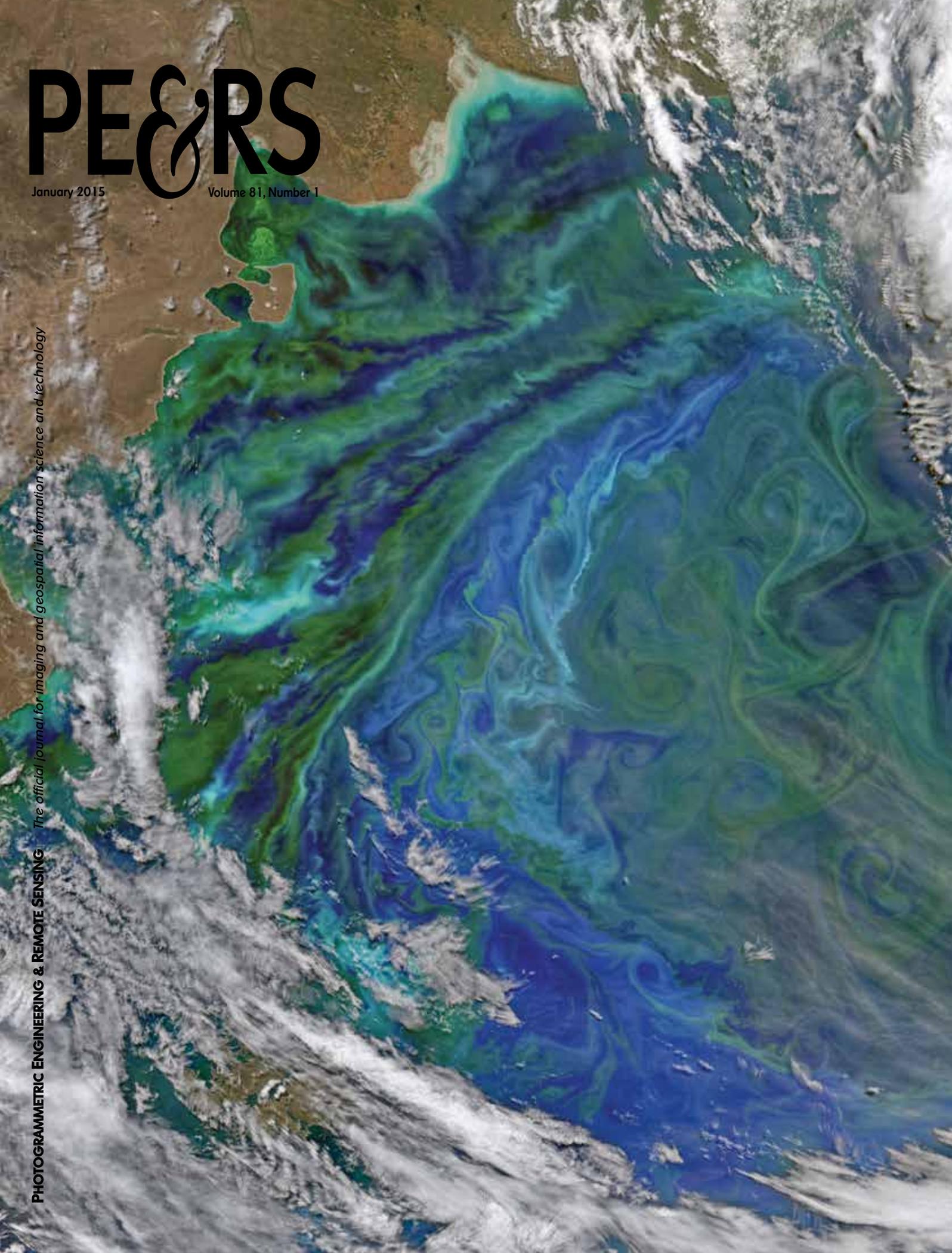
# PE&RS

January 2015

Volume 81, Number 1

*The official journal for imaging and geospatial information science and technology*

**PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING**



### COLUMNS

Letter from the Executive Director	4
Grids and Datums— <i>Republic of The Marshall Islands</i>	13
Mapping Matters	15
Book Review— <i>Elements of photogrammetry with Application in GIS, Fourth Edition</i>	19

### ANNOUNCEMENTS

ASPRS 2015 Annual Conference	1
Signatures—SAC Blog	21
January GeoByte—Relationship of and Transformations Between most Common Reference Frames Used in the U.S.	48
Call for Papers	68

### DEPARTMENTS

Certification	18
Region News	20
ASPRS News	22
Classifieds	22
New Members	23
Industry News	24
Calendar	48
Forthcoming Articles	48
Who's Who in ASPRS	79
Sustaining Members	80
Instructions for Authors	82
Membership Application	84

### HIGHLIGHT ARTICLE

**5 Geospatial Technologies Lead the Living to the Dead**  
*Nancy K. O'Hare, Brandon P. Adams, Marguerite Madden, and Thomas R. Jordan*

### PROFESSIONAL INSIGHT—AN INTERVIEW

**11 UNM ASPRS Student Chapter**



**11** UNM | *Geography & Environmental Studies*

### PEER-REVIEWED ARTICLES

**25 Correction of Distortions in YG-12 High-Resolution Panchromatic Images**  
*Yonghua Jiang, Guo Zhang, Deren Li, Xinming Tang, Wenchao Huang, and Litao Li*

A method by which the positioning accuracy of YG-12 high-resolution panchromatic images using GCPs is significantly improved by eliminating inaccuracies caused by interior distortions and serious time-synchronization errors.

**37 Optimal Land Cover Mapping and Change Analysis in Northeastern Oregon Using Landsat Imagery**

*Michael Campbell, Russell G. Congalton, Joel Hartter, and Mark Ducey*

compare a number of well-established techniques with some new methods using a two-county area in northeastern Oregon as a case study.

**49 Reliable Spatial Relationship Constrained Feature Point Matching of Oblique Aerial Images**

*Han Hu, Qing Zhu, Zhiqiang Du, Yeting Zhang, and Yulin Ding*

A reliable feature point matching method for oblique aerial images using the spatial relationships of the point correspondences to remove outliers.

**59 Sub-Pixel-Scale Land Cover Map Updating by Integrating Change Detection and Sub-Pixel Mapping**

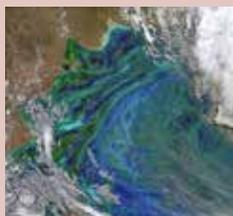
*Xiaodong Li, Yun Du, and Feng Ling*

A land cover map updating method that involves the use of a current coarse-resolution remotely sensed image and a previous fine-resolution land cover map to update fine-resolution land cover maps.

**69 Mapping Wetlands and Phragmites Using Publicly Available Remotely Sensed Images**

*Yichun Xie, Anbing Zhang, and William Welsh*

Standard procedures for integrating NAIP (National Agriculture Imagery Program) and Landsat images with multiple processes of ground truthing, image classification and validation.



Late spring and summer weather brings blooms of color to the Atlantic Ocean off of South America. The Patagonian Shelf Break is a biologically rich patch of ocean where airborne dust from the land, iron-rich currents from the south, and upwelling currents

from the depths provide a bounty of nutrients for the grass of the sea—phytoplankton. In turn, those floating sunlight harvesters become food for some of the richest fisheries in the world.

The Visible Infrared Imaging Radiometer Suite (VIIRS) on Suomi NPP captured this false-color view of phytoplankton-rich waters off of Argentina on December 2, 2014. Scientists in NASA's Ocean Color Group used three wavelengths (671, 551, and 443 nanometers) of light to highlight different plankton communities in the water. Bands of color not only reveal the location of plankton, but also the dynamic eddies and currents that carry them.

The aquamarine stripes and swirls are likely coccolithophores, a type of phytoplankton with microscopic calcite shells that can give water a chalky color. The various shades of green are probably a mix of diatoms, dinoflagellates, and other species.

Blooms occur off of Patagonia because warmer, saltier waters from the subtropics meet the colder, fresher waters flowing up from the south. These currents collide along what oceanographers call a shelf-break front, a turbulent area of vertical and horizontal mixing on the edge of the continental shelf.

NASA images by Norman Kuring, NASA's Ocean Color Group. For more information, visit <http://earthobservatory.nasa.gov/IOTD/view.php?id=84870>

**APPLICATIONS PAPER**

# Reliable Spatial Relationship Constrained Feature Point Matching of Oblique Aerial Images

Han Hu, Qing Zhu, Zhiqiang Du, Yeting Zhang, Yulin Ding

## Abstract

*This paper proposes a reliable feature point matching method for oblique images using various spatial relationships and geometrical information for the problems resulted by the large view point changes, the image deformations, blurring, and other factors. Three spatial constraints are incorporated to filter possible outliers, including a cyclic angular ordering constraint, a local position constraint, and a neighborhood conserving constraint. Other ancillary geometric information, which includes the initial exterior orientation parameters that are obtained from the platform parameters and a rough DEM, are used to transform the oblique images geometrically and reduce the perspective deformations. Experiment results revealed that the proposed method is superior to the standard SIFT regarding both precision and correct matches using images obtained by the SWDC-5 system.*

## Introduction

Beginning in 2000, oblique aerial camera systems garnered attention from the photogrammetry community due to their ability to capture the facades of buildings and their ability to be briefly interpreted (Petrie, 2009). Many penta-view camera systems that feature four 45° oblique cameras and one nadir camera, including Pictometry (Gerke and Kerle, 2011), MIDAS (Madani, 2012), and SWDC-5 that are used in this paper, have collected numerous datasets. However, traditional photogrammetry techniques and software are designed primarily for nadir images and are difficult to adapt for oblique aerial images. New challenges have been posed to photogrammetry practitioners to integrate all of the images to extract more compact and accurate information, especially for oblique views (Jacobsen, 2009; Nyaruhuma *et al.*, 2012; Fritsch and Rothermel, 2013).

Exterior orientation (EO) parameters for the images are necessary prior to 3D reconstruction (Gerke, 2009), texture mapping (Wang *et al.*, 2008), position measurement (Xiong *et*

*al.*, 2014), and other mapping applications. Because all of the cameras are installed on a rigid platform and held stationary after manufacturing, the bore-sight angles and translations (called platform parameters) between the nadir view and the four oblique views are fixed in the ideal condition (Wiedemann and Moré, 2012) and are calibrated using retro-reflect coded targets in the calibration field (Fraser, 1997). In theory, only the EO parameters for the nadir images need to be estimated in the bundle adjustment. However, due to reinstallation, limitations in mechanical manufacturing, and possible asynchronous exposures among different cameras, the platform parameters are not stable and can only be considered as fixed in a single flight (Jacobsen, 2009). In this situation, a combined bundle adjustment with both the nadir and oblique images in the same block are necessary, which requires sufficient tie points to sew the block together.

It turns out that feature matching between nadir images and oblique images is astonishingly difficult because of the obvious difference in their appearances, which consists of occlusions, perspective deformations, light conditions, and blur that are caused by the wide baseline and large tilting angles (Yao and Cham, 2007; Yang *et al.*, 2012). Additionally, in production practice, an existing solution with traditional software is to process images for each camera separately and then manually select enough inter-camera tie points to assemble different blocks of images together. However, this solution is not only time consuming, but it is also prone to inter-camera inconsistencies due to the lack of accurate tie points.

To eliminate the drudgery of manual selection and to improve the quality of bundle adjustment, we propose to amend the standard feature matching process by injecting additional spatial relationships of feature points into the process to increase the reliability, rather than only using the appearance information of the images. Specifically, we propose a cyclic angular order constraint, a local position constraint and a neighborhood conserving constraint. Furthermore, the initial geometric information of the five cameras that is obtained from the global positioning system (GPS) and inertial measurement unit (IMU) onboard the aircraft, calibrated platform parameters, and a rough DEM are used to geometrically transform the images, in order to reduce the perspective deformations. In our previous work (Zhu *et al.*, 2007), a filter strategy using information content is proposed to improve the repeatability of the interest points and the reliability of the matches, which is also based on the appearance information. However, when the images are essentially dissimilar in appearance,

---

Han Hu, Zhiqiang Du, and Yeting Zhang are with the State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University, P.O. Box C310, 129 Luoyu Road, Wuhan, Hubei, 430079, P.R. China (huan@whu.edu.cn).

Qing Zhu is with the National-local Joint Engineering Laboratory of Spatial Information Technology for High-speed Railway Running Safety, Southwest Jiaotong University, P.R. China; and the State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University, P.R. China.

Yulin Ding is with the State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University, P.R. China, and the Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong.

---

Photogrammetric Engineering & Remote Sensing  
Vol. 81, No. 1, January 2015, pp. 49–58.  
0099-1112/15/811–49

© 2014 American Society for Photogrammetry  
and Remote Sensing  
doi: 10.14358/PERS.81.1.49

the spatial relationships are more important cues that are responsible for reliable feature correspondence. Furthermore, the spatial relationship constraint has been confirmed to be a powerful method for wide baseline-dense matching in our previous work.

In the next section, feature matching methods based on appearance and spatial relationships are briefly introduced. Then, we present two core innovative steps of the proposed methods followed by information for the configurations of the oblique camera system and the datasets obtained. The performance of the proposed methods is subsequently evaluated leading to concluding remarks.

## Related Works

In an early investigation of image matching, three fundamental properties of the correct correspondences were set up (Ullman, 1979): *Similarity*: the correspondences must be similar in appearance among the matching images. *Proximity*: the correspondences must have tenable spatial relationships. For example, in the situation of repeated patterns, images of textureless, spatial relationships provide clues for establishing correspondences or removing outliers among identical features. *Exclusion*: the final feature matches should be established one-on-one and can simply be enforced by cross check, which is a two-step procedure of forward and backward matching.

In the photogrammetry and computer vision communities, feature matching methods have been prosperous; most of these methods resorted to appearance information alone to identify correspondences. To the best of our knowledge, the earliest feature matching method dates to the Moravec (1981) corner detector and normalized cross correlation (NCC). Since then, the robust pyramid matching strategy (Wang, 1990) has been the software standard, which uses intersecting points to detect corners (Moravec, 1981; Förstner and Gülch, 1987; Harris and Stephens, 1988) and NCC to match the features from coarse to fine. In applications where accuracy is important, the least square matching (LSM) strategy is also mandatory for locating the matches in a sub-pixel position (Gruen, 1985). The sophisticated paradigm that uses pyramid matching and LSM is widely adopted in both aerial and close range applications. Liang and Heipke (1996) integrate an automatic relative orientation method with a modified Moravec detector and a coarse-to-fine matching strategy into their method. Several hundreds of well distributed correspondences are found, and the results indicate that internal accuracies reached less than 0.3 pixels, even in scanned digital images. Lerma *et al.* (2013) compare different matching strategies for close range applications, in which artificial retro-reflect targets are not used. They confirm that NCC and LSM result in better accuracies for images with near parallel bore-sight directions.

The methods that match features with NCC can be considered to use the appearance information in a fixed square window around the corners. Consequently, NCC is essentially sensitive to scale and rotation changes (Lowe, 2004). Lindeberg (1993) proposes to detect blob-like features that are stable in the scale space, and the same author also provides a practical solution to build the scale space that is the Laplacian of Gaussian (Lindeberg, 1998), which has triggered the development of numerous invariant features. Based on the scale space theory, Lowe (2004) presents the milestone work of scale invariant feature transform (SIFT), which detects features in the scale space using the difference of Gaussian and describes the image blob with the histograms of gradients. Mikolajczyk and Schmid (2005) compare various features against different situations and claim that SIFT presents the best performances, except for the case of large affine deformations.

Some approaches attempt to restrict the features to be affine-invariant (Matas *et al.*, 2004; Mikolajczyk and Schmid, 2004); however, these approaches can cause either decreases in the amount of detected features or performance losses in the cases of small affine deformations (Lowe, 2004). Attempts to surpass the performance of SIFT have focused on descriptor dimensions (Ke and Sukthankar, 2004), speed (Bay *et al.*, 2008) and affinity (Morel and Yu, 2009).

The methods described above use appearance information either in a square window or in a salient blob. The extracted feature descriptors are matched using NCC and Euclidean distances, respectively. When the appearances of the images are fundamentally dissimilar, their performances will predominantly decrease and cause a huge amount of false correspondences. Spatial relation constraints are herein adopted to guide the matching process and remove outliers. Zhang (1988) invented a feature matching method called a “bridging mode” which is still widely used in the VirtuoZo software. The bridging mode assumes that a single feature point is not capable of describing a feature; thus, three points that form an image segment are adopted: two protruding points with a minimal intensity gradient and one point with a maximal gradient. The matching window is formed by the relation in the image segment. Zhang *et al.* (1991) extend the bridging mode to support search in two directions to enrich the matching information and use dynamic programming to increase the global consistency. Although the matching window is adapted by the spatial relationships, the matching criterion is still determined by the NCC; therefore, the issues of the NCC still exist for this method.

Unlike the low-level feature descriptors, which have similarity measurements that can be simply defined by Euclidean distances, the high-level spatial relations are complex to quantify. Existing methods that combine blob-like feature descriptors and spatial relations are usually formulated as a graph matching problem (Caetano *et al.*, 2009). Li *et al.* (2005) use a Bayesian formulation, which is modeled by a Markov random field (MRF). The likelihood term enforces the similarities of the feature appearance/descriptors, and the *a priori* term encodes the cyclic order constraint of the Delaunay triangulations. The model is casted to the corresponding factor graph and solved with the max-product algorithm. Torresani *et al.* (2008) formulate the feature matching problem as a unified energy minimization task that encodes additional information, including feature appearance, geometric compatibility, spatial coherence, and is solved using dual decomposition. The authors noticed that their method was equivalent to the graph matching problem that embodies pairwise constraints in edges and similarity measurements in the vertices of graphs. Liu *et al.* (2012) propose a similar method and consider the optimization problem as finding two matched graphs with minimum non-rigid transformation errors. However, these graph matching approaches, which combine appearance and structure information in a unified framework, are NP-hard because all of the combinations of the binary labels that denote true/false matches must be exhausted to find the global solution (Li *et al.*, 2005; Caetano *et al.*, 2009). The space complexity of the algorithm is generally  $O(n^4)$  (Caetano *et al.*, 2009; Torki and Elgammal, 2010), and the time complexity is  $O(n^3)$  (Liu *et al.*, 2012), where  $n$  is the number of features; this approach is often not appropriate for aerial images. Even 800 correspondences will take almost 1,200 seconds to solve the problem (Liu *et al.*, 2012), compared to nearly real time with the standard RANSAC approach. In our practical experience, the number of correspondences of a single stereo pair result in a magnitude of  $10^3$  to  $10^4$ .

## Methodology

### Algorithm Principle

As shown in Figure 1. We add two steps, which are detailed in the following Section, into the standard feature matching procedure. First, we use geometric transformation to relieve the perspective deformation between the oblique image and nadir images. Second, three spatial relation constraints are adopted to remove the remaining outlier by the random sample consensus (RANSAC) approach using epipolar information. Because epipolar constraint is only imposed on a line, there are still possibilities for unfiltered false correspondences when the image format is large. In this situation, additional constraints to filter off remaining outliers are necessary.

### Preprocessing

Before detection and matching of the feature points, we exploit the initial EO parameters to geometrically transform the images in order to relieve the perspective deformations as shown in Figure 1. Assuming the terrain is exactly a plane, a one-to-one correspondence can be established between the nadir and the oblique images. After defining a plane  $\pi \rightarrow z - h_{flight} = 0$  to roughly approximate the terrain in the coordinate system of vertical camera, a homography matrix  $H$  is estimated between the image plane of the vertical and oblique images. The homography matrix  $H$  is calculated in Equation 1 (Hartley and Zisserman, 2004) after giving calibrated camera matrix  $K_V$  and  $K_O$  for the vertical and oblique cameras, respectively:

$$H = K_O(R + \mathbf{t}\mathbf{n}_d^T)K_V^{-1} \quad (1)$$

where  $\mathbf{n}_d = \mathbf{n}/d$ ,  $\mathbf{n}$  is the normal vector of plane  $\pi$  in the coordinate system of the nadir image,  $d$  is the distance between its center and  $\pi$ , and  $[R, \mathbf{t}]$  is the relative rotation and translation between the nadir and oblique images. Using the homography matrix  $H$ , the image coordinate of the oblique camera  $\mathbf{m}_O$  can be mapped to the vertical view by  $\mathbf{m}_V = H\mathbf{m}_O$ . Using a specific resampling method, such as bilinear interpolation or bicubic interpolation, the pixels of the oblique image can be remapped to match the vertical view. Therefore, the perspective deformation in the oblique image is relieved.

To compensate for the differences of approximately  $180^\circ$  in the yaw angle between strips, a special method that treats all three rotation angles of the nadir images as zeros is adopted. Therefore, similar geometric transformations also apply to all of the nadir images. To enclose all of the transformed pixels, the rectified images must be enlarged and shifted according to the transformed boundaries. In this way, we can assume that the scale, rotation, and other perspective deformations have been eliminated.

The features are then detected, described and matched on the rectified images as shown in Figure 1; however, the point coordinates are reverse transformed to the original image coordinates because the geometric transformation defined by the homography matrix is invertible. Then, the follow-up outlier removal steps are performed because some constraints are only justifiable in the original image, such as the epipolar constraint. Because no invariant ability should be imposed on the features, we chose to use the FAST corner detector (Rosten *et al.*, 2010) as suggested by Jazayeri and Fraser (2010) and a BRIEF descriptor (Calonder *et al.*, 2012) for efficiency consideration. The binary descriptors are then matched in hamming distance using approximate nearest neighbor search techniques (Muja and Lowe, 2012), then reverse transformed to the original image coordinates. We also adopt the RANSAC approach to initially remove outliers; however, dozens of false correspondences out of the  $10^4$  matches still exist, even if the RANSAC threshold is set to 1 pixel, which will sometimes prevent the bundle adjustment from converging.

### Spatial Relationship Constrained Outlier Removal

In this study we impose three more constraints on the initial correspondences based on spatial relations and classify them as inliers or outliers after quantifying the constraints, including the cyclic angular order constraint, local position consistency constraint, and neighborhood conserving constraint. The first constraint assume that any non-rigid transform of the correspondences will not change the cyclic angular order between them; the second assumes that the positions of the correspondences in a local neighborhood should not abruptly change; and the last constraint that the

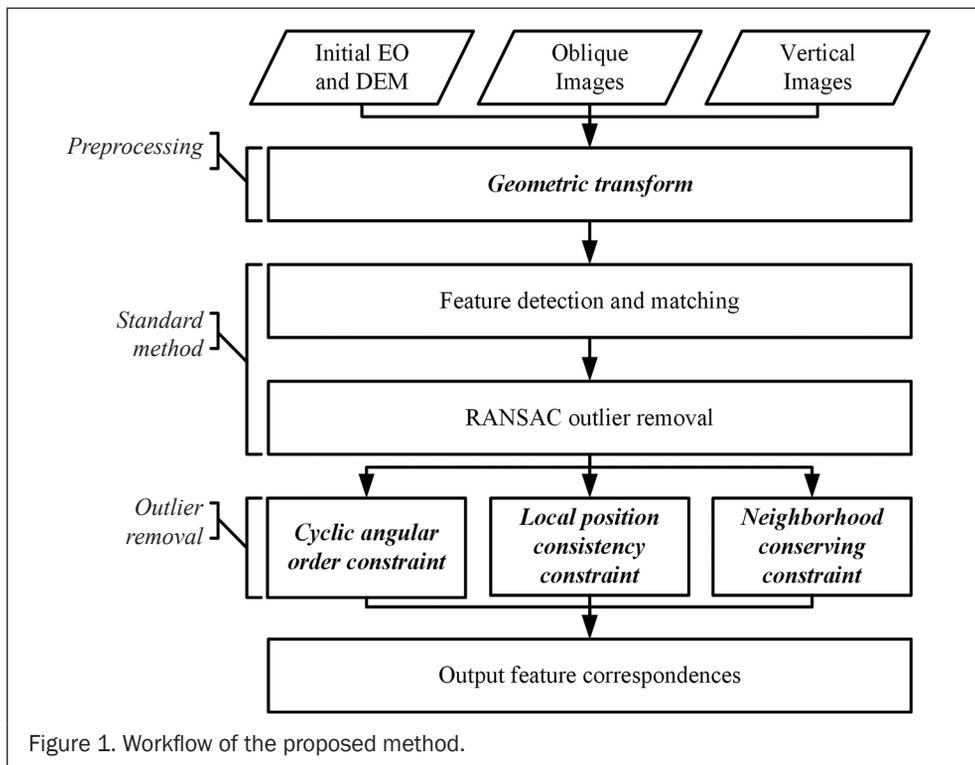


Figure 1. Workflow of the proposed method.

neighborhood relationship should be conserved in correspondences. It should be noted that all the three constraints are not processed sequentially, but in parallel with the refined correspondences after RANSAC outlier removal as shown in Figure 1. After executed in parallel, we removed the union set of the outliers detected in each constraint.

#### Cyclic Angular Order Constraint

In this study, the angular order of a feature point  $i$  in the reference image is denoted as  $S_i$ , where  $i$  is the point number, and that in the matching image is denoted as  $S'_i$ .  $S$  is formatted as the sequence of point numbers of the  $K$  nearest neighbors (denoted as  $N(i)$ ;  $k = 6$  is adopted in this study), which is ordered clockwise. For example, as shown in Figure 2a the angular order for point No. 96 is  $S_{96} = \{103, 98, 94, 95, 97, 104\}$  and that in the matching image (Figure 2b) is  $S'_{96} = \{97, 104, 103, 98, 95, 94\}$ . Similarly, for the false correspondence point No. 94, the angular orders are  $S_{94} = \{97, 104, 103, 95, 96, 98\}$  and  $S'_{94} = \{104, 103, 97, 96, 95, 98\}$ .

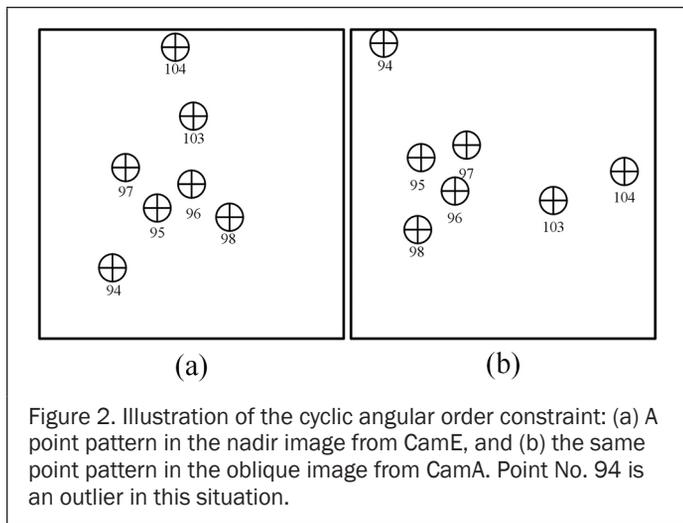


Figure 2. Illustration of the cyclic angular order constraint: (a) A point pattern in the nadir image from CamE, and (b) the same point pattern in the oblique image from CamA. Point No. 94 is an outlier in this situation.

We can consider the sequences of  $S$  and  $S'$  as two strings, and their differences can be measured by the edit distance. The edit distance quantifies the dissimilarity by counting the minimum number of operations to transform one string to another; the operations include insertion, deletion, and substitution of a single character. To compensate for the rotation effects, the cyclic edit distance (denoted as  $CED(S, S')$ ) is adopted (Klein, 1998; Liu *et al.*, 2012). For example, as discussed above,  $CED(S_{96}, S'_{96}) = 2$  because it involves one deletion and one assertion of No. 94; similarly,  $CED(S_{94}, S'_{94}) = 4$ .

Under the assumption that a non-rigid transformation will not change the cyclic angular order of a point around its neighbors, we have also found that the cyclic angular order constraint is sometimes indeterminate. Thus, we relax the constraint to the following statement:

**Cyclic Angular Order Constraint** - For each correspondence  $i$ , if  $CED(S_i, S'_i) \geq 4$ , then label correspondence  $i$  an outlier.

The constraint is established based on the following reasons: (a) When no outlier exists, the  $CED$  will be zero under the assumption that a non-rigid transformation will not change the cyclic angular order; (b) For a correct correspondence, if only one outlier exists among its KNN and the outlier changes the angular order, then  $CED(S_i, S'_i)$  will always be two after careful investigations; and (c) For a false correspondence, if the outlier changes the angular order, then  $CED(S_i, S'_i)$  will always be larger than four. Although the constraint will

sometimes cause a false alarm (i.e., label the correct correspondence an outlier) and neglect outliers, these cases are rare or can be remedied by other constraints. For example, there are two special cases: (a) more than two outliers exist around a correct correspondence, and (b) the outlier does not change the angular order. The first case is rather rare because the correspondences have already been filtered by RANSAC, and we only chose small neighbors. The latter case may be remedied by other steps, which is the reason that we use multiple spatial constraints.

#### Local Position Consistency Constraint

Using the initial correspondences, we can estimate a perspective transform function  $T(\cdot)$  to approximately align the reference and matching image such that a point  $\mathbf{p}(x, y)$  in the reference image can be mapped to the point  $\mathbf{p}'(x', y')$  in the matching image as  $\mathbf{p}' \approx T(\mathbf{p})$ . For example, Sun *et al.* (2014) use a 2D translation to simulate a global transformation, and the affine transformation is adopted by Han *et al.* (2012). In this paper, the affine model is calculated as:

$$T(\mathbf{p}) = A\mathbf{p} + \mathbf{b} \quad (2)$$

where  $A$  is a  $2 \times 2$  matrix and  $\mathbf{b}$  is a  $2 \times 1$  translation vector. However, the non-rigid deformation caused by variations in elevation and perspective transform cannot be simplified by the affine model. Absolute discrepancies exist between the points in the matching image and  $T(\mathbf{p})$  as denoted by the following  $2 \times 1$  residual vector of  $r(\mathbf{p})$ :

$$r(\mathbf{p}) = \mathbf{p}' - T(\mathbf{p}). \quad (3)$$

The local position consistency constraint is based on the assumption that after the affine transformation, the residuals should be consistent across a local area. More explicitly, for a point  $\mathbf{p}_i$  and its KNN  $\{\mathbf{p}_k \mid k \in N(i)\}$ , we calculate the average residuals of the neighbors  $\mu_r = \sum_k r(\mathbf{p}_k) / K$ , and the mean and standard deviation of their lengths as  $\mu_{\|r\|} = \sum_k \|r(\mathbf{p}_k)\| / K$  and  $\sigma_{\|r\|} = \sqrt{\sum_k (\|r(\mathbf{p}_k)\| - \mu_{\|r\|})^2 / K}$ , respectively. It should be noted that, in most cases,  $\|\mu_r\| \neq \mu_{\|r\|}$ . The local position consistency constraint imposes the following statement on the residual length and direction:

**Local Position Consistency Constraint:** The residual for a point  $\mathbf{p}_i$  in the reference image should satisfy the following constraint, otherwise label it as an outlier.

$$\begin{cases} r(\mathbf{p}_i) \bullet \mu_r > 0 \\ \|r(\mathbf{p}_i)\| \in [\mu_{\|r\|} - 3\sigma_{\|r\|}, \mu_{\|r\|} + 3\sigma_{\|r\|}] \end{cases} \quad (4)$$

where the first restricts the residual to be approximately in the same direction of its neighbors and the second impose constrain on the size of the residuals.

#### Neighborhood Conserving Constraint

The neighborhood conserving constraint is based on the assumption that the neighbors of a point in the reference image are also the neighbors in the matching image. Therefore, we compute both the KNN of  $\mathbf{p}_i$  and  $\mathbf{p}'_i$  as  $N(i)$  and  $N'(i)$ , respectively. Ideally, the two ID sets should coincide with each other. The similarity of two sequence vectors of the points can simply be measured by their intersection, which is denoted as  $I(\mathbf{p}_i, \mathbf{p}'_i)$ . Similar to the second constraint, the statistical average  $\mu_I$  and standard deviation  $\sigma_I$  are obtained, and the third constraint is formally stated as follows:

**Neighborhood Conserving Constraint:** The intersection number for the KNN of the two points in a correspondence,  $I(\mathbf{p}_i, \mathbf{p}'_i)$ , should satisfy the following:

$$I(\mathbf{p}_i, \mathbf{p}'_i) > \mu_i - 3\sigma_i. \quad (5)$$

#### Algorithm Complexities

As described above, previous matching algorithms that adopted spatial constraints often resort to the graph matching method, thus resulted in prohibitively expensive time and space complexities at the magnitude of  $O(n^3)$ , where  $n$  is the number of correspondences or even higher (Torki and Elgammal, 2010), which was unable to extend to aerial images applications. It can be noted that for all the three constraints, KNN search for all the correspondences are adopted and its time complexity is  $O(n \log n)$  in average. For the first constraint, we need to calculate the cyclic edit distance for each points at the time complexity of approximate  $O(Kn)$ , where  $K$  is the selected number of nearest neighbors. For the second constraint, we need to calculate the statistical information of each cliques at the complexity of  $O(Kn)$ . And for the last one, only global statistical information is needed with time complexity of  $O(n)$ . Considering that  $K$  is a relative small number, the total time complexity is only at  $O(n \log n)$  level. Furthermore, it is obviously that the space complexity is  $O(n)$ . So the low algorithm complexities have guaranteed the extendibility of our method.

## Experimental Evaluations

### Configurations and Dataset

In this study, the oblique imagery system, SWDC-5, adopts the “Maltese Cross” configuration (Petrie, 2009). This type of configuration consists of five cameras: a single camera (*CamE*)

pointing at nadir and four oblique cameras (*CamA*, *CamB*, *CamC*, and *CamD*). One pair of oblique cameras (*CamA* and *CamC*) point in opposite directions across the strips, whereas the other pair of cameras (*CamB* and *CamD*) view the same strip. Using a cross configuration system, the intuitive merits of having oblique views from four different directions are clearly emphasized by the ability to capture all of the possible facades of the buildings as shown in Figure 3.

The entire dataset consists of 27 strips that are collected during several flights by LEADOR SPATIAL over Jinyang, which is new development in the mountainous city of Guiyang. In fact, the flight conditions over the area were not good due to air movement and cloudy skies. The complete block contains 1,711 images for each camera at a strip overlap of 60 percent and side overlap of 50 percent, from which only some continuous strips of the fifth flight were chosen here. The average elevation of the coverage area is approximately 1,300 m, and the expected relative flight height is 600 m. The principal distances of the camera are approximately 50 mm for the nadir images and 80 mm for the oblique images. With this configuration, the ground sample distance (GSD) is approximately 0.08 m. In this study, the image distortion and principle point are pre-rectified to zero and to the image center, respectively. Additionally, a bundle adjustment of all of the nadir images was previously conducted using the integrated sensor orientation (Ip *et al.*, 2007). The initial EO information was obtained from the GPS/IMU devices and several manually selected ground control points to eliminate the datum shift. The mean square error of unit weight  $\sigma_0$  is 0.3 pixels, and we treat triangulated 3D points from nadir images as control points in the experiments because no ground control points were available. Furthermore, the theoretical positioning error of the nadir images are at centimeters level. The platform parameters are calibrated in a calibration field as shown in Table 1. The initial EO parameters for *CamA* to *CamD* are estimated

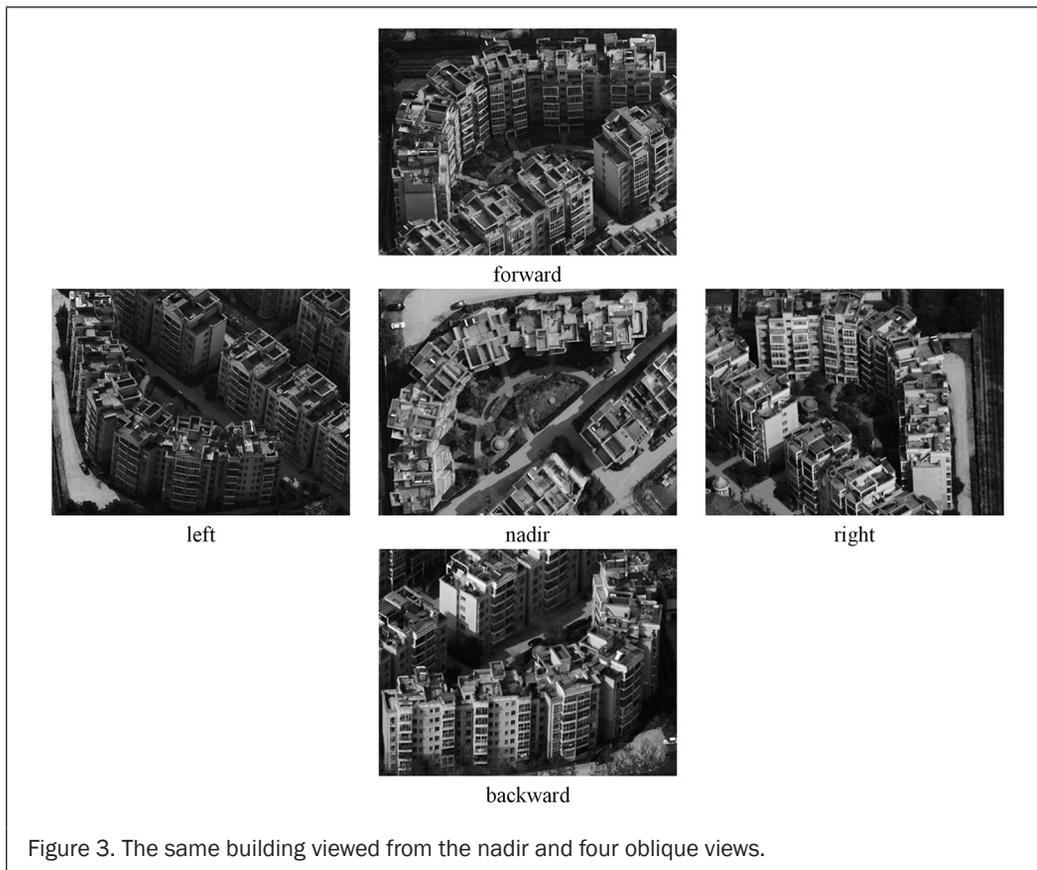


Figure 3. The same building viewed from the nadir and four oblique views.

with the platform parameters and the adjusted EO for *CamE*. Due to reinstallation and other miscellaneous reasons, the initial EO parameters may cause meters-level 3D positioning errors. However, the initial accuracies for the oblique EO parameters are enough, because they only contribute to the geometrical transformations.

TABLE 1. PLATFORM PARAMETERS OF THE CAMERA SYSTEM; THE EULERIAN ANGLES ARE IN THE ORDER OF  $\omega \rightarrow \varphi \rightarrow \kappa$ . CAMERA IS POINTING AT NADIR AND CAMERA TO CAMD ARE THE FOUR OBLIQUE CAMERAS

	Translation $P_{OE}$			Rotation $R_{OE}$		
	$X(m)$	$Y(m)$	$Z(m)$	$\varphi(^{\circ})$	$\omega(^{\circ})$	$\kappa(^{\circ})$
<i>CamA-CamE</i>	0.109	0.003	0.04	-0.744	44.824	90.704
<i>CamB-CamE</i>	0.019	0.13	0.056	-44.550	-0.171	180.187
<i>CamC-CamE</i>	-0.111	-0.014	0.037	0.705	-44.911	-89.346
<i>CamD-CamE</i>	-0.002	-0.138	0.046	45.459	0.412	0.248

To qualitatively and quantitatively evaluate the performance of the proposed methods, two types of experiments are conducted on two small subsets of images from the fifth flight, which include a dense built-up area and a rural area with vegetation and bare soil, as shown in Figure 4. First, we visually and statistically compare the match results with the standard SIFT methods using the two datasets. Next, we apply bundle adjustment to a block of 150 images and evaluate the quality of the bundle adjustment to quantify the performance of the correspondences.

### Performance Evaluation of the Matching Results

In Plate 1, the match results of the proposed method and the standard SIFT method are displayed. For both methods, the ratio match (Lowe, 2004) is adopted, which detects the two nearest neighbors and only reserves those correspondences that exhibit a distance ratio between the two neighbors of less than 0.75. Furthermore, the cross check described previously is also adopted to ensure the one-on-one matching property. The successive RANSAC approach using a fundamental matrix follows the ratio match with a threshold of 1 pixel. To present a better understanding of the matching results, the key points and the images of the oblique views are geometrically transformed without a shift such that the correspondences are approximately parallel to each other. Otherwise, the correspondences will intersect and distort the meaning of the visual comparisons. The results of the proposed method are shown in Plate 1a and 1c. For the others, the matching results are obtained with the SIFT on the original images. The first column in each row provides the intuitional distributions of the correspondences, and the second column provides the matching results. The number of correspondences and the distribution are superior to those using SIFT, especially in the built-up areas where SIFT encountered enormous outliers after the RANSAC approach and most matches are centralized at few areas (Plate 1b). Furthermore, even in seriously occluded areas, such as the built-up areas in Plate 1a, we are also able to detect enough well-distributed tie points on building roofs when the shape of the roofs are relatively regular and simple.

To present a more comprehensive understanding of the performance of the proposed method, the matching results

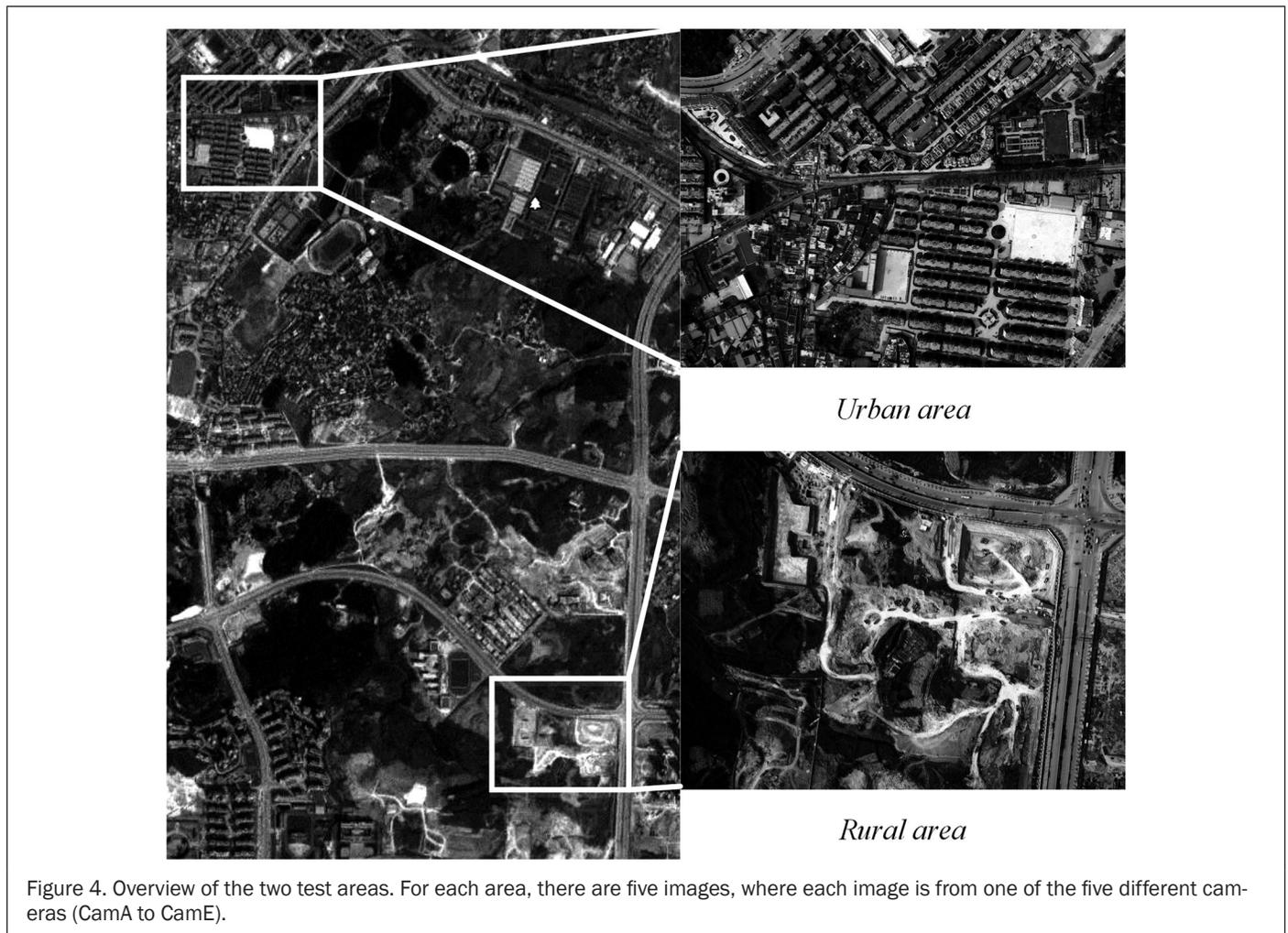


Figure 4. Overview of the two test areas. For each area, there are five images, where each image is from one of the five different cameras (CamA to CamE).

for the two areas are listed in Table 2. The number of final matches, the outliers detected by the spatial relationships constraint and the false alarms are presented. The false alarms are obtained by manual inspecting the detected outliers; specifically, they are the number of correct correspondences that are labeled as outliers. The false alarms rate is very low, partly because of the sparseness of the remaining outliers after previous outlier handling steps, such as cross check, ratio match and RANSAC. Another characteristic observed is that the performances for the images of the urban areas are slightly inferior to those of the rural areas. This is because occlusions are more severe in the previous scenario.

Turning into compare with other methods or software, we also use SIFT, VisualSFM (Wu, 2011) and Photoscan to test matching results with the same images. In the experiments, the SiftGPU (Wu, 2007) is adopted for the SIFT implementation, which is also used in VisualSFM for feature detection. It can be noted that whereas the performance of SIFT decreases dramatically in the urban areas, the proposed method performs considerably well; the geometric transformation relieves the perspective deformation, especially for regions that are parallel to the horizontal ground (i.e., the building roof and planar ground). In fact, in the urban area, more

correspondences are detected on the building roofs than on the ground, which are more inclined to be occluded and textureless. Furthermore, in the case of matching between images of different oblique cameras, SIFT will totally fail after RANSAC outlier removal and VisualSFM also will not perform well, as shown in the last two rows of Table 2a and 2b. The translational tilting angles (Morel and Yu, 2009) between the images are approximately 90°, which makes it almost impossible for the partially affine invariant SIFT descriptor to establish correct correspondences. However, the proposed method is also capable of handling these cases. Although Photoscan is able to obtain enough tie points, its performance is still inferior to the proposed method and after careful examinations, we discover that Photoscan will produce enormous false matches. In fact, when images is significantly different, large amount of outliers are expected if not handled appropriately.

Furthermore, one of the major considerations on the performance for practical applications is the runtime speed. As described above, some previous works on feature points matching with spatial relationship constraints are impracticable in applications of aerial images due to prohibitively high space and time complexities (Liu *et al.*, 2012). In order to provide some perspectives into the algorithm details, we

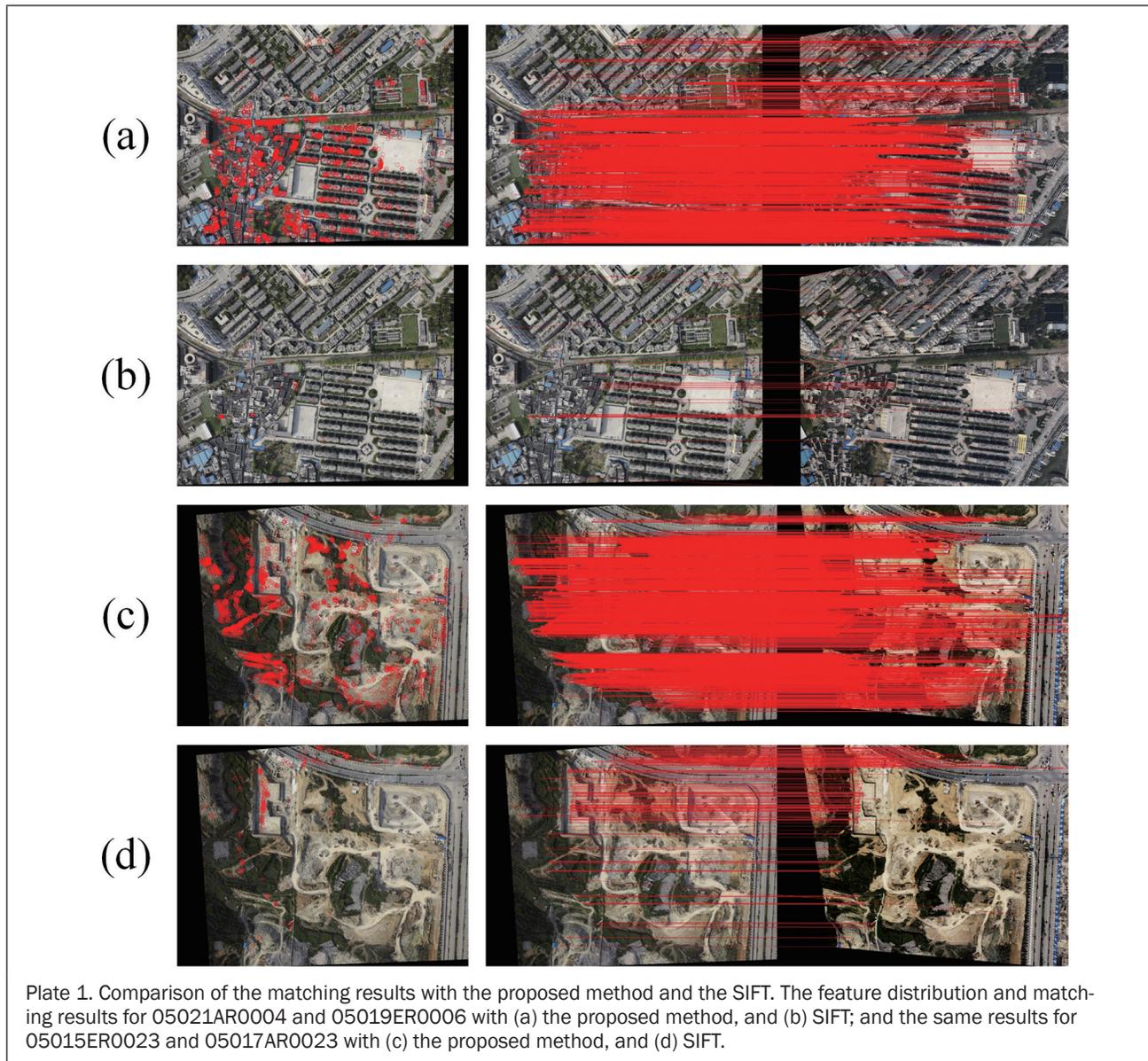


TABLE 2. MATCHING RESULTS FOR THE TWO TEST AREAS; THE 'E' DENOTES THE NADIR IMAGES AND A TO D DENOTE OBLIQUE IMAGES

a) Tests in the rural area, which features vegetation, bare earth and roads.

Pair	Proposed Method			SIFT	VisualSFM	Photoscan
	#Matches	#Outliers	#False Alarm	#Matches	#Matches	#Matches
E-A	3595	31	0	215	413	1369
E-B	9017	32	0	1730	182	2913
E-C	6498	32	0	2277	248	2145
E-D	7026	23	0	2315	294	2434
A-C	1253	21	0	N/A	N/A	78
B-D	1181	6	0	N/A	N/A	612

b) Tests in the built-up area, which features dense and tall buildings.

Pair	Proposed Method			SIFT	VisualSFM	Photoscan
	#Matches	#Outliers	#False Alarm	#Matches	#Matches	#Matches
E-A	3398	31	0	36	61	940
E-B	3943	41	0	138	62	934
E-C	4461	56	1	43	36	1730
E-D	4688	41	0	49	34	1096
A-C	726	17	0	N/A	N/A	286
B-D	2596	36	0	N/A	N/A	231

TABLE 3. PERFORMANCE RESULTS FOR THE RUNTIME SPEED OF THE PROPOSED METHOD

Pair	Feature detection (ms)	KNN search (ms)	RANSAC filter (ms)	Spatial filter (ms)	Miscellaneous (ms)	Total (ms)
E-A	8847	8290	9	26	102	17274
E-B	10607	10329	618	26	99	21679
E-C	10034	9038	30	28	107	19237
E-D	10814	9587	380	32	13	20826

TABLE 4. ADJUSTED RESULTS OF THE PLATFORM PARAMETERS FOR THE OBLIQUE IMAGES; THE BOLD VALUES DENOTE THE INTERNAL ACCURACIES OF THE PLATFORM PARAMETERS

	Translation $P_{OE}$			Rotation $R_{OE}$		
	$X(m)$	$Y(m)$	$Z(m)$	$\varphi(^{\circ})$	$\omega(^{\circ})$	$\kappa(^{\circ})$
CamA-CamE	0.107	0.003	0.038	-0.770	44.820	90.720
	<b>3.20E-03</b>	<b>3.29E-03</b>	<b>3.26E-03</b>	<b>3.07E-04</b>	<b>2.30E-04</b>	<b>4.31E-04</b>
CamB-CamE	0.018	0.118	0.051	-44.583	-0.155	180.223
	<b>2.93E-03</b>	<b>2.27E-03</b>	<b>2.87E-03</b>	<b>1.83E-04</b>	<b>1.78E-04</b>	<b>2.12E-04</b>
CamC-CamE	-0.094	-0.012	0.024	0.751	-44.974	-89.306
	<b>2.81E-03</b>	<b>3.14E-03</b>	<b>2.96E-03</b>	<b>2.91E-04</b>	<b>2.19E-04</b>	<b>2.90E-04</b>
CamD-CamE	-0.001	-0.116	0.039	45.519	0.374	0.295
	<b>2.98E-03</b>	<b>2.43E-03</b>	<b>2.98E-03</b>	<b>1.79E-04</b>	<b>1.73E-04</b>	<b>2.50E-04</b>

demonstrate the runtime speed with respect to different modules of the method. As summarized in Table 3, although we have carefully chosen the feature detection method and diligently optimized the KNN search, which is specially tuned for speed, the bottleneck still lies in feature detection and KNN search for the match candidates. The first two steps operate on all the feature points, whose amount is at the level of  $10^5$ , however the RANSAC filter and spatial filter proposed in this paper operate on the matching candidates succeeded to previous steps. This helps to explain the performance. Due to the low time complexities of the spatial filter as described above, the runtime of the three constraints are nearly negligible when compared to the hotspots in the procedures.

#### Performance Evaluation of Bundle Adjustment

To further evaluate the accuracies of the matching results, we select 150 images from seven strips (30 images for each camera) and conduct a bundle adjustment on the block of images using tie points that are generated and filtered by the

proposed method. In the bundle adjustment, the EO parameters for the nadir images (CamE) are held fixed because these images have already been oriented with ground control points. Because all of the strips are from the same flight, we considered that the platform parameters at all of the exposure positions are the same and unknown in the adjustment (Jacobsen, 2009). Among all of the images, we created 3,390 pair-wise images matches because of the dense overlap ratio for the oblique camera system. After connecting all of the matches, we select hundreds of points that are evenly distributed within each image. In the selection, we prefer the points that connect more images, and the maximum tied points even linked 31 images. Overall, there are 8,339 object points that formulate 47,522 image tie points for the bundle adjustment.

The mean square error of the unit weight is  $\sigma_0 = 0.46$  pixels after the bundle adjustment, and the adjusted platform parameters together with their internal accuracies are listed in Table 4. The internal accuracies for the adjusted platform parameters are quite good because of the well-distributed block

TABLE 5. EVALUATION RESULTS WITH THE GROUND CONTROL POINTS: (A) THE PROJECTION ERROR OF THE OBLIQUE IMAGES, AND (B) THE 3D POSITION ERRORS TRIANGULATED BY POINTS ON THE OBLIQUE IMAGES

(a) Projection errors of the ground control points				
	<i>RMSE x (pixels)</i>	<i>RMSE y (pixels)</i>	<i>RMSE xy (pixels)</i>	<i>Max. RMSE (pixels)</i>
Before adjustment	7.45	11.50	13.70	25.78
After adjustment	0.60	0.81	1.00	4.18
(b) 3D position errors of the ground control points				
	<i>RMSE XY (m)</i>	<i>RMSE Z (m)</i>	<i>RMSE (m)</i>	<i>Max. RMSE (m)</i>
Before adjustment	0.603	0.942	1.114	3.504
	<b><math>8.04 \times 10^{-4}</math></b>	<b><math>1.26 \times 10^{-3}</math></b>	<b><math>1.49 \times 10^{-3}</math></b>	<b><math>4.6 \times 10^{-3}</math></b>
After adjustment	0.072	0.100	0.123	0.479
	<b><math>9.6 \times 10^{-4}</math></b>	<b><math>1.33 \times 10^{-4}</math></b>	<b><math>1.64 \times 10^{-4}</math></b>	<b><math>6.38 \times 10^{-4}</math></b>

network with multiple tie points among all of the images. The platform parameters changed only slightly compared to those before the adjustment as shown in Table 1.

To evaluate the consistencies in the block, we select another 3,000 sets of tie points from the remaining correspondences after removing the points used for the bundle adjustment. Each set must connect at least two nadir images and two oblique images. Thus, we treat the 3D points triangulated by the nadir images as ground control points. The projection errors and the 3D position errors of the oblique images are listed in Table 5, including the root mean square errors (RMSE) of the image coordinates and object space. The projection errors decreased nearly an order of magnitude, even with the small changes in the platform parameters. The position errors are reduced to approximately 1.5 GSD. When taking the flight height into, the relative accuracies (RMSE divided by relative flight height) are shown as the bolded rows in Table 5b, which is sufficient for subsequent processing considering the precision of the triangulated control points and the small baseline-height ratio of the medium format oblique images (Colomina and Molina, 2014).

## Conclusions

We determined that translational shifts and rotational skewing exists in the platform parameters for the SWDC-5, even if the parameters are expected to be fixed after offline calibration, which will cause tens of pixels misalignments on the oblique images and more than one-meter positioning errors. Due to occlusions and perspective deformations, tie point matching for the images of oblique camera systems are extraordinarily difficult. To surmount this problem, we exploited the initial geometric information to rectify the images and reduce the perspective deformations. Furthermore, spatial relationship constraints are incorporated into the feature matching procedures and serve as important clues to remove outliers, which still exist after using the common approach to handling outliers because of the consequences of the essential image dissimilarities in appearances. Experimental evaluations and comparisons reveal that the proposed method outperforms the standard SIFT in both the numbers and distributions of the correspondences. After a bundle adjustment with the connected points that are generated by the proposed method, the translation and skewing of the platform parameters are remedied. Furthermore, the internal inconsistencies of the block are reduced to a satisfactory level. An obvious trend in processing oblique images is to exploit the measurement potentials (Haala, 2013), and our future works will focus on developing efficient and suitable method for the dense image matching of oblique images.

## Acknowledgments

This study was supported by the National Basic Research Program of China (973 Program, No. 2010CB731801) and the National High Technology Research and Development Program of China (863 Program, No. 2012AA121305).

## References

- 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.
- Caetano, T.S., J.J. McAuley, L. Cheng, Q.V. Le, and A.J. Smola, 2009. Learning graph matching, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(6):1048–1058.
- Calonder, M., V. Lepetit, M. Ozuysal, T. Trzcinski, C. Strecha, and P. Fua, 2012. BRIEF: Computing a local binary descriptor very fast, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(7):1281–1298.
- Colomina, I., and P. Molina, 2014. Unmanned aerial systems for photogrammetry and remote sensing: A review, *ISPRS Journal of Photogrammetry and Remote Sensing*, 92:79–97.
- Förstner, W., and E. Gülch, 1987. A fast operator for detection and precise location of distinct points, corners and centres of circular features, *ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, 02-04 June, Interlaken, Switzerland, pp. 281–305.
- Fraser, C.S., 1997. Innovations in automation for vision metrology systems, *The Photogrammetric Record*, 15(90):901–911.
- Fritsch, D., and M. Rothermel, 2013. Oblique image data processing - Potential, experience and recommendation, *Proceedings of Photogrammetric Week 2013*, 09-13 September, Stuttgart, Germany, pp. 73–88.
- Gerke, M., 2009. Dense matching in high resolution oblique airborne images, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38(3/W4):77–82.
- Gerke, M., and N. Kerle, 2011. Automatic structural seismic damage assessment with airborne oblique Pictometry imagery, *Photogrammetric Engineering & Remote Sensing*, 77(9):885–898.
- 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.
- Haala, N., 2013. The landscape of dense image matching algorithms, *Proceedings of Photogrammetric Week 2013*, 09-13 September, Stuttgart, Germany, pp. 271–284.
- Han, Y.K., Y.G. Byun, J.W. Choi, D.Y. Han, and Y. Kim, 2012. Automatic registration of high-resolution images using local properties of features, *Photogrammetric Engineering and Remote Sensing*, 78(3):211–221.
- 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.
- Hartley, R., and A. Zisserman, 2004. *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, UK, 672 p.

- Ip, A., N. El-Sheimy, and M. Mostafa, 2007. Performance analysis of integrated sensor orientation, *Photogrammetric Engineering & Remote Sensing*, 73(1):1–9.
- Jacobsen, K., 2009. Geometry of vertical and oblique image combinations, *Remote Sensing for a Changing Europe: Proceedings of the 28<sup>th</sup> Symposium of the European Association of Remote Sensing Laboratories, Istanbul, Turkey*, 02-05 June 2008, pp. 16.
- Jazayeri, I., and C.S. Fraser, 2010. Interest operators for feature-based matching in close range photogrammetry, *The Photogrammetric Record*, 25(129):24–41.
- Ke, Y., and R. Sukthankar, 2004. PCA-SIFT: A more distinctive representation for local image descriptors, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2004)*, 27 June-02 July, Washington, D.C., pp. 506–513.
- Klein, P.N., 1998. Computing the edit-distance between unrooted ordered trees, *Proceedings of the 6<sup>th</sup> Annual European Symposium on Algorithms (ESA98)*, 24-26 August, Venice, Italy, pp. 91–102.
- 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.
- Li, Y., Y. Tsin, Y. Genc, and T. Kanade, T., 2005. Object detection using 2D spatial ordering constraints, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, 20-26 June, San Diego, California, pp. 711–718.
- Liang, T., and C. Heipke, 1996. Automatic relative orientation of aerial images, *Photogrammetric Engineering & Remote Sensing*, 62(1):47–55.
- 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.
- Liu, Z., J. An, and Y. Jing, 2012. A simple and robust feature point matching algorithm based on restricted spatial order constraints for aerial image registration, *IEEE Transactions on Geoscience and Remote Sensing*, 50(2):514–527.
- Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, 60(2):91–110.
- Madani, M., 2012. Accuracy potential and applications of MIDAS aerial oblique camera system, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 39(PART B1):127–132.
- 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, 2004. Scale and affine invariant interest point detectors, *International Journal of Computer Vision*, 60(1):63–86.
- Mikolajczyk, K., and C. Schmid, 2005. A performance evaluation of local descriptors, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630.
- Moravec, H.P., 1981. Rover visual obstacle avoidance, *Proceedings of the 7<sup>th</sup> International Joint Conference on Artificial Intelligence (IJCAI 1981)*, 24-28 August, Vancouver, Canada, pp. 785–790.
- Morel, J., and G. Yu, 2009. ASIFT: A new framework for fully affine invariant image comparison, *SIAM Journal on Imaging Sciences*, 2(2):438–469.
- Muja, M., and D.G. Lowe, 2012. Fast matching of binary features, *Proceedings of the Ninth Conference on Computer and Robot Vision (CRV2012)*, 27-30 May, Toronto, Canada, pp. 404–410.
- Nyaruhuma, A.P., M. Gerke, G. Vosselman, and E.G. Mtaló, 2012. Verification of 2D building outlines using oblique airborne images, *ISPRS Journal of Photogrammetry and Remote Sensing*, 71:62–75.
- Petrie, G., 2009. Systematic oblique aerial photography using multiple digital cameras, *Photogrammetric Engineering & Remote Sensing*, 75(2):102–107.
- 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.
- Sun, Y., L. Zhao, S. Huang, L. Yan, and G. Dissanayake, 2014. L2-SIFT: SIFT feature extraction and matching for large images in large-scale aerial photogrammetry, *ISPRS Journal of Photogrammetry and Remote Sensing*, 91:1–16.
- Torki, M., and A. Elgammal, 2010. One-shot multi-set non-rigid feature-spatial matching, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010)*, 13-18 June, San Francisco, California, pp. 3058–3065.
- Torresani, L., V. Kolmogorov, and C. Rother, 2008. Feature correspondence via graph matching: Models and global optimization, *Proceedings of the European Conference on Computer Vision (ECCV 2008)*, 12-18 October, Marseille, France, pp. 596–609.
- Ullman, S., 1979. *The Interpretation of Visual Motion*, The MIT Press, Cambridge, Massachusetts, 229 p.
- Wang, M., H. Bai, and F. Hu, 2008. Automatic texture acquisition for 3D model using oblique aerial images, *International Conference on Intelligent Networks and Intelligent Systems (ICINIS'08)*, 01-03 November, Wuhan, China, pp. 495–498.
- Wang, Z., 1990. *Principle of Photogrammetry: With Remote Sensing*, Press of Wuhan Technical University of Surveying and Mapping, Wuhan, 455 p.
- Wiedemann, A., and J. Moré, 2012. Orientation strategies for aerial oblique images, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 39(Part B1):185–189.
- Wu, C., 2007. SiftGPU: A GPU implementation of Scale Invariant Feature Transform (SIFT), URL: <http://cs.unc.edu/~ccwu/siftgpu> (last date accessed: 18 November 2014).
- Wu, C., 2011. VisualSFM: A Visual Structure from Motion System, URL: <http://ccwu.me/vsfm> (last date accessed: 18 November 2014).
- Xiong, X., Y. Zhang, J. Zhu, and M. Zheng, 2014. Camera pose determination and 3-D measurement from monocular oblique images with horizontal right angle constraints, *IEEE Geoscience and Remote Sensing Letters*, 11(11):1976–1980.
- Yang, H., S. Zhang, and Y. Wang, 2012. Robust and precise registration of oblique images based on scale-invariant feature transformation algorithm, *IEEE Geoscience and Remote Sensing Letters*, 9(4):783–787.
- Yao, J., and W. Cham, 2007. Robust multi-view feature matching from multiple unordered views, *Pattern Recognition*, 40(11):3081–3099.
- Zhang, Z., 1988. A new approach of epipolar-line image matching - Bridging Mode, *Journal of Wuhan Technical University of Surveying and Mapping*, 13(4):19–27.
- Zhang, Z., J. Zhang, and X. Wu, 1991. Developing of bridging mode and global image matching, *Journal of Wuhan Technical University of Surveying and Mapping*, 16(3):1–11.
- 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.

(Received 18 April 2014; accepted 04 August 2014; final version 24 August 2014)