LGS: LOCAL GEOMETRICAL STRUCTURE-BASED INTEREST POINT MATCHING FOR WIDE-BASELINE IMAGERY IN URBAN AREAS

M. Chen[1, *], Q. Zhu[1], S. Yan[1], Y. Zhao[1]

[1] Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, 611756, China - minchen@home.swjtu.edu.cn; zhuq66@263.net; shyan@my.swjtu.edu.cn; ytzhao@my.swjtu.edu.cn

Commission I, WG I/8

KEY WORDS: Wide-baseline images, Urban area, Viewpoint change, Local geometrical structure, Interest point matching, Matching expansion

ABSTRACT:

Feature matching is a fundamental technical issue in many applications of photogrammetry and remote sensing. Although recently developed local feature detectors and descriptors have contributed to the advancement of point matching, challenges remain with regard to urban area images that are characterized by large discrepancies in viewing angles. In this paper, we define a concept of local geometrical structure (LGS) and propose a novel feature matching method by exploring the LGS of interest points to specifically address difficult situations in matching points on wide-baseline urban area images. In this study, we first detect interest points from images using a popular detector and compute the LGS of each interest point. Then, the interest points are classified into three categories on the basis of LGS. Thereafter, a hierarchical matching framework that is robust to image viewpoint change is proposed to compute correspondences, in which different feature region computation methods, description methods, and matching strategies are designed for various types of interest points according to their LGS properties. Finally, random sample consensus algorithm based on fundamental matrix is applied to eliminate outliers. The proposed method can generate similar feature descriptors for corresponding interest points under large viewpoint variation even in discontinuous areas that benefit from the LGS-based adaptive feature region construction. Experimental results demonstrate that the proposed method provides significant improvements in correct match number and matching precision compared with other traditional matching methods for urban area wide-baseline images.

1. INTRODUCTION

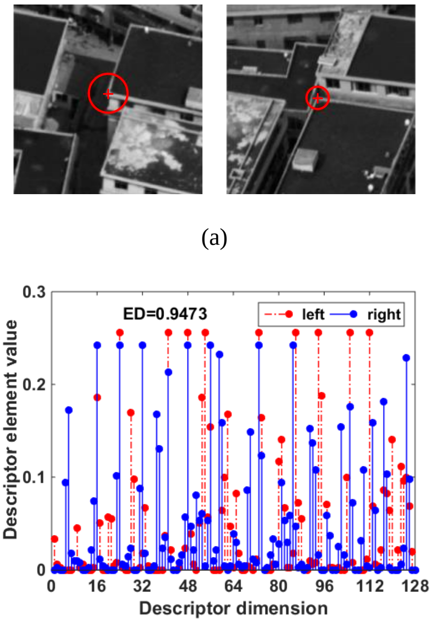
Image matching, which refers to the task of establishing correspondences between images with overlapping areas, is a fundamental issue in many photogrammetry and computer vision applications. In general, existing image matching methods can be grouped into two categories: area-based and feature-based methods (Gruen, 2012). Area-based methods often detect interest points on the source image and find the corresponding pixel in a search area on the target image. In the correspondence search stage, the intensity values in a rectangular window are simply adopted to describe the central pixel. Area-based methods can perform effectively in traditional aerial imagery because the small pitch and roll angles and relatively stable flight height of the aerial platform do not cause obvious geometrical differences between aerial images. Among area-based methods, image correlation is widely used because it is simple and easy to implement. Normalized cross-correlation was proposed to improve the robustness of cross-correlation to linear intensity variation (Ackermann, 1984; Lhuillier and Quan, 2002). The matching performance can be improved further by combining several matching strategies, e.g., least squares matching (Gruen and Akca, 2005). Although area-based methods can achieve sub-pixel level accuracy and even better, they are sensitive to image nonlinear intensity change and geometric deformation (Gruen, 2012), thus, these methods are unable to match wide-baseline urban area images with viewpoint change and parallax discontinuity.

Feature-based methods are more robust to image variation than area-based methods by constructing robust feature descriptors. Feature-based methods generally consist of three steps: feature detection, description, and matching. In the past several decades, numerous methods have been proposed for one of the steps or the entire procedure. Scale-invariant feature transform (SIFT) algorithm (Lowe, 2004) is one of the most popular methods due to its robustness to image rotation and scale change. With the success of the SIFT method, feature-based methods have been widely studied to address image geometric distortion and intensity changes. Some affine-invariant region detectors have been proposed to address image affine transformation (Matas et al., 2004; Mikolajczyk et al., 2005). The matching performance on images with viewpoint change was improved based on these affine-invariant features. However, fewer features can be detected than interest point detectors and the robustness to image viewpoint change is still limited. Other methods obtain better matching results by simulating image affine or projective space and performing feature matching in the simulated space (Morel and Yu, 2009; Yu et al., 2012). ASIFT is one of the best-known methods to deal with image viewpoint variation (Morel and Yu, 2009). It simulates the original image to cover the entire affine space in the beginning. Then, SIFT is adopted to detect and match features in the simulated affine space. ASIFT can find matches from the images even under significant viewpoint change. However, the high complexity limits the industrial application of ASIFT.

[*] Corresponding author

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-IV-2-W5-13-2019 | © Authors 2019. CC BY 4.0 License.

In the field of photogrammetry, high-precision position and orientation system (POS) data are typically used as auxiliary information to coarsely correct images and eliminate image geometric distortion caused by viewpoint change before feature matching. Then, a traditional feature matching method, such as SIFT, is adopted to find correspondences (Hu et al., 2015; Roth et al., 2017). Although this strategy helps improve the matching performance, the local geometric distortion is difficult to fit by the global transformation. The number of correct matches can be increased through sub-region-based matching (Sun et al., 2014; Ai et al., 2015) by dividing the input image into several sub-regions and matching features in the sub-regions. If high-precision POS data are unavailable in certain applications, then an initial matching can be conducted to obtain several matches to calculate a roughly geometric transformation between the images, and then images can be corrected and traditional feature matching method is adopted to generate correspondences; this method is called ISIFT in the present paper (Jiang and Jiang, 2017). The aforementioned methods can improve the matching performance through image correction, but several problems remain. First, image correction can only alleviate the geometric distortion of the plane scene to a certain extent. It is difficult to compute feature regions with similar image content between corresponding points that are located in areas with discontinuous parallax. Second, the methods using initial matching to correct images significantly depend on the performance of initial matching. Obtaining reliable initial matches between urban wide-baseline images is difficult using the existing feature matching methods (Figure 1).



(a)

(b)

Figure 1. An example of viewpoint change and parallax discontinuity between wide-baseline images in urban area. (a) shows a pair of oblique images. The two crosses denote a pair of correspondence. The red circle around each point denotes the feature region determined based on SIFT (Lowe, 2004). (b) shows the SIFT descriptors of the two points and the corresponding Euclidean distance (ED). The feature descriptors are dissimilar and difficult to match correctly.

Despite the progress in dealing with geometric distortions between images, the matching of wide-baseline images in urban area remains a problem. Feature matching for this type of images is becoming an unavoidable problem with the development of unmanned aerial vehicle, oblique photogrammetry, and mobile mapping technologies. Therefore, developing robust feature matching methods is crucial to ensure reliable correspondence in such images.

In this study, a new method is proposed to match interest points on wide-baseline images in urban areas on the basis of the following observations: 1) if the feature region computation is guided by the local geometrical structure (LGS) of each interest point, the computed feature regions will be robust to image viewpoint change; thus, the feature descriptors generated from these feature regions would have a greater likelihood of success in matching even in a discontinuous area; and 2) man-made objects are the main objects of urban area imagery and a large number of straight lines can be detected in addition to interest points. The main contributions of this study are threefold. First, we define a concept of LGS by exploring the geometric relationship between interest point and straight lines in the neighborhood of interest point. LGS distinguishes interest points as different categories (G1, G2, and G3) and guides the design of special matching methods for various types of interest points instead of using a common matching method for all types of interest points, thereby contributing to an improvement in the matching performance. Second, we propose two structure adaptive methods on the basis of LGS to construct viewpoint-invariant feature regions for interest points in G1 and G2, respectively. The proposed feature region construction methods help generate feature regions with consistent image content for corresponding points, thereby enabling the production of similar feature descriptors for corresponding interest points under serious geometric distortion and parallax discontinuity. Finally, we propose a matching expansion method in which affine invariants are introduced as geometric constraints and combined with feature descriptors similarity to increase the number of correct matches.

2. METHODOLOGY

This section details the proposed interest point matching method. We first define the concept of LGS and classify interest points into three groups. Then, an LGS-based structure adaptive feature (SAF) called LGS-SAF is developed, and the nearest neighbor distance ratio (NNDR) matching strategy (Lowe, 2004) is improved to match LGS-SAFs. Many initial matches are obtained in this stage and the epipolar geometry between images is estimated through random sample consensus (RANSAC) algorithm (Fischler and Bolles, 1981). As LGS-SAF can only be built for the interest points in group G1 (as discussed in Section 2.2), in the next two stages, we respectively propose an epipolar geometry-based method called EG-SAF to construct SAF and match the interest points in group G2 and the unmatched interest points in group G1, as well as a matching expansion method to match the interest points that have not been matched in the previous stages. The workflow of the proposed matching method is summarized in Figure 2.

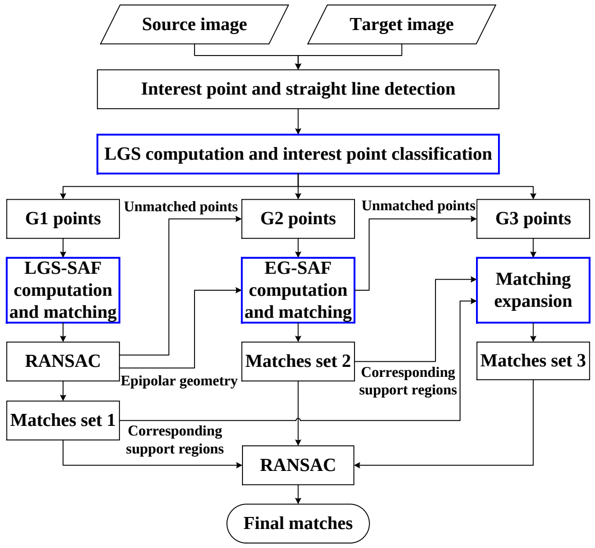

Source image
Target image
Interest point and straight line detection
LGS computation and interest point classification
G1 points
Unmatched points
G2 points
Unmatched points
G3 points
LGS-SAF computation and matching
EG-SAF computation and matching
Matching expansion
RANSAC
Epipolar geometry
Matches set 2
Corresponding support regions
Matches set 3
Matches set 1
Corresponding support regions
RANSAC
Final matches

Figure 2. Workflow of the proposed matching method. The blue boxed steps are the key steps of our workflow.

2.1 LGS definition and interest point classification

Man-made objects are the major type of objects on urban area images. Thus, numerous interest points and straight lines can be detected. Straight lines effectively describe the structure of objects. This paper proposes an LGS-based interest point matching method to address viewpoint change by exploring the geometrical relationship between interest points and straight lines. We define the concept of LGS for interest point as follows: for an interest point p_i, we first set a local window R_i that is centered on p_i and has a size of $m \times m$ pixels, where m is a user-defined parameter. The straight lines that have at least one point on the line dropping in R_i are selected. If parallel straight lines exist among these selected lines, only the longest one is retained. Thus, a straight line set $LS_i = \{l_1, l_2, \ldots, l_n\}$ is generated for interest point p_i, where n is the number of retained straight lines. For each straight line $l_j \in LS_i$, if one of its endpoints drops in R_i, then the direction from the endpoint closer to p_i to another endpoint is regarded as the direction of l_j, labeled as θ_j. If both endpoints are out of R_i, then l_j is split into two new straight lines from the projection point of point p_i on l_j and two opposite directions are respectively assigned to each of the two new straight lines. LS_i is updated with the split. Finally, corresponding to each straight line $l_j \in LS_i$, a vector starting from the interest point p_i with the direction θ_j and with the modulus $|l_j|$ is determined, where $|l_j|$ denotes the length of l_j. On the basis of all vectors, the LGS of interest point p_i is defined as

$$LGS(p_i) = \left\{ p_i, \left(|l_j|, \theta_j \right)_{j=1}^{N} \right\} \tag{1}$$

where N is the number of vectors, and $|l_j|$ and θ_j are the modulus and direction of the j-th vector.

On the basis of LGS, all interest points are classified into three groups (G1, G2, and G3). **G1:** If the LGS of an interest point has at least two vectors and salient point can be found in the directions of at least two vectors, then this interest point is classified into group G1. To find salient point in the direction of a vector $\mathbf{v}_j$, we define an impact zone (indicated by the dashed box in Figure 3) for $\mathbf{v}_j$. The parameter s is set to control the size of the impact zone. If a straight line exists with at least one point on the line dropping in the impact zone, and the intersection of the straight line and the vector is also in the impact zone, then this intersection is regarded as one salient point in the direction (Li and Yao, 2017). If more than one straight line satisfies the constraint, then more than one salient point is found in this direction. **G2:** If the LGS of an interest point has at least two vectors, but the number of vectors in which the salient point can be found is less than two, then this interest point is classified into group G2. **G3:** If the LGS of an interest point has less than two vectors, then this interest point is classified into group G3.

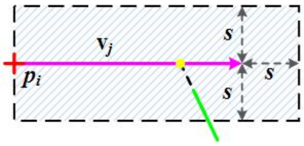


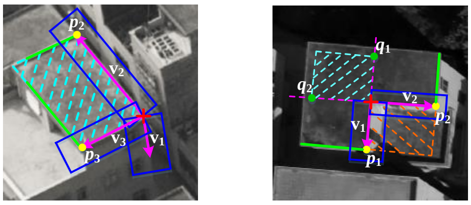
Figure 3. Salient point computation in the direction of one vector of LGS. The red cross denotes an interest point p_i. The purple arrow denotes a vector $\mathbf{v}_j$ of the LGS of p_i. The green line is a straight line dropping in the impact zone of $\mathbf{v}_j$. The yellow solid dot is a salient point in the direction of $\mathbf{v}_j$.

2.2 LGS-SAF computation and matching for interest points in group G1

The matching of interest points in group G1 is a key step in the proposed matching framework. It not only produces some point matches but also affects the following matching procedure. An LGS-SAF matching method is proposed to ensure the reliability of the matching of interest points in group G1.

2.2.1 Viewpoint robust LGS-SAF computation: For an interest point in group G1, salient points in the directions of at least two LGS vectors can be found according to the definition of G1. In the beginning, the vectors that have salient points in their directions are selected. If the number of selected vectors is greater than two, for every two selected vectors with a vectorial angle at 30°-150°, one salient point is selected from each vector and combined with the interest point to determine a parallelogram, i.e., the support region (image region normalized to feature region) of the interest point (Figure 4(a)). If the number of selected vectors is equal to two, then one salient point is selected from each vector and combined with the interest point to determine a support region. Meanwhile, two virtual points that are symmetrical to the salient points are selected and combined with the interest point to form another support region (Figure 4(b)). To improve the matching rate in the aforementioned support region determination, if more than one salient point exists in one direction, then every salient point is adopted to determine support region separately.

Figure 4 shows two examples of support region computation for interest points in G1. In Figure 4(a), the number of LGS vectors is equal to three. Among the three vectors, no salient point has been found in the direction of vector $\mathbf{v}_1$, and two salient points, (p_2 and p_3) have been found in the $\mathbf{v}_2$ and $\mathbf{v}_3$ directions, respectively. Therefore, salient points p_2 and p_3 are combined with the interest point to determine a support region (cyan dotted area). In Figure 4(b), there are only two LGS vectors. p_1 and p_2 are the salient points found in the $\mathbf{v}_1$ and $\mathbf{v}_2$ directions, respectively. These two salient points and the interest point determine a support region (orange dotted area). In addition, the symmetrical points of p_1 and p_2 are adopted as virtual salient points and combined with the interest point to form another support region (cyan dotted area).



(a) (b)

Figure 4. Support region computation for interest point in G1. (a) an example of more than two LGS vectors, and (b) an example of only two LGS vectors.

According to the support region computation method, some interest points have more than one support region. In this case, the interest point is regarded as several interest points and each one corresponds to a support region inspired by the main orientation assignment in the SIFT method (Lowe, 2004).

To compute feature descriptor conveniently, the irregular support region is normalized to a square feature region. The normalized feature region size is fixed as $T_r \times T_r$ for every interest point. In the normalization, the interest point is fixed at the lower left corner of the normalized feature region. Another vertex on the same diagonal line with the interest point in the support region corresponds to the upper right corner of the normalized feature region. A vector from the interest point to the diagonal point is formed respectively in the support region and normalized feature region. Then, the corresponding relationship of the two pairs of vertices on another diagonal line is determined according to the side of the vector where the vertex is located. A homography matrix between the support region and the normalized feature region is estimated through the four pairs of vertices. The normalization is performed based on the homography matrix.

This normalization method has three advantages. First, the fixed feature region size makes the feature region invariant to image scale change. Second, the method is rotation invariant by fixing the normalized location of the four vertices of the support region. Third, the normalization contributes to distinguish the interest points that share the same support region, e.g., the corners of the same building roof.

As SIFT descriptor, the normalized feature region is divided into $4 \times 4 = 16$ sub-regions. A histogram of gradient orientation (8 orientations) is computed in each sub-region and then accumulated to form a 128-dimensional descriptor. Finally, a normalization step is performed to improve the illumination robustness of the descriptor. Considering that image rotation has already been eliminated in the feature region normalization, we do not compute the main orientation for the feature and do not perform gradient orientation normalization in the descriptor construction.

2.2.2 Improved NNDR matching for LGS-SAF: Feature matching can be performed by using the existing matching methods, e.g., NNDR. For each source feature (feature on the source image), the NNDR method is to find the two target features (features on the target image) that have the smallest Euclidean distance with the source feature. If the ratio between the smallest distance over the second smallest distance is smaller than a threshold, then the target feature corresponding to the smallest distance is regarded as a match of the source feature. The NNDR method works effectively when combined with SIFT-like methods. However, for the LGS-SAFs proposed in this paper, the NNDR method usually fails because the two target features with the two smallest distance values may correspond to the same interest point because one interest point may generate more than one feature region and descriptor according to the proposed LGS-SAF computation method. In this situation, satisfying the ratio constraint in the NNDR method is difficult. Thus, the corresponding interest points cannot be matched (Figure 5). To overcome this problem, an improved NNDR (called I-NNDR) strategy is proposed to match LGS-SAFs as follows:

First, the Euclidean distance between all source feature descriptors and target feature descriptors are computed. For a source feature, the target feature with the smallest distance $d_{\min}$ is found. Then, a distance threshold $T_d = d_{\min} / T_{ratio}$ is computed, where T_{ratio} is the ratio threshold in the NNDR method. Finally, all target features with distance values smaller than T_d are selected. If all the selected features correspond to the same interest point, then this interest point is considered as the match point of the source feature.

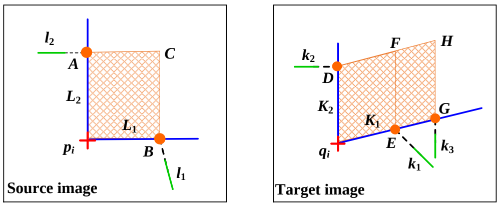


Figure 5. Comparison of NNDR and I-NNDR matching methods

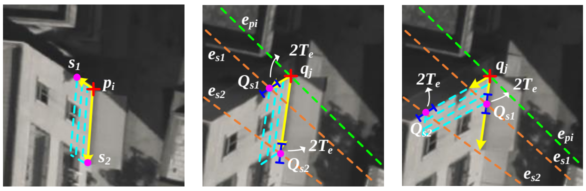
In Figure 5, p_i is an interest point on the source image. L_1 and L_2 are two straight lines forming the LGS of p_i. Straight line l_1 drops in the impact zone of L_1 and generates an intersection B. Straight line l_2 drops in the impact zone of L_2 and generates an intersection A. Thus, parallelogram $p_i ACB$ is a support region of interest point p_i. Interest point q_i on the target image is the corresponding point of p_i. Straight lines K_1 and K_2 are the corresponding lines of L_1 and L_2, respectively, which form the LGS of q_i. Straight lines k_1 and k_3 drop in the impact zone of K_1 and generate two intersections E and G. Specifically, k_1 is the corresponding line of l_1. Straight line k_2, the corresponding line of l_2, drops in the impact zone of K_2 and generates an intersection D. Thus, two parallelograms q_iDFE and q_iDHG are formed as support regions of q_i. Then, two feature descriptors are computed for q_i. If the textures in these two regions are similar (which is often the case with urban area images), the computed feature descriptors are highly similar. Then, the two feature descriptors may generate the two smallest distance values in the NNDR method. The distance ratio may be close to 1 and does not satisfy the ratio constraint in NNDR. Consequently, interest points p_i and q_i cannot be matched successfully. In comparison, they can be matched successfully by using the I-NNDR method because both of the two distance values that are smaller than T_d correspond to q_i.

After feature matching, the RANSAC algorithm is performed to eliminate outliers and estimate the fundamental matrix F between the source and target images.

2.3 EG-SAF computation and matching for interest points in G2

According to the definition of G2 and the construction of LGS-SAF, we cannot compute LGS-SAFs for interest points in G2. In this section, an epipolar geometry-based structure adaptive feature (called EG-SAF) matching method is proposed to deal with interest points in G2. The interest point set of G2 is updated before matching by adding the interest points in G1 that were not matched in the previous step. The EG-SAF is computed as follows (Figure 6):

First, for an interest point p_i on the source image, a support region is formed by every two LGS vectors. As shown in Figure 6(a), the endpoints s_1 and s_2 of the two branches are combined with p_i to determine a parallelogram as the support region of interest point p_i. Second, the epipolar lines e_{pi}, e_{s1}, and e_{s2} corresponding to points p_i, s_1, and s_2 are computed according to $e_{pi} = Fp_i$, $e_{s1} = Fs_1$, and $e_{s2} = Fs_2$, respectively, where F is the fundamental matrix between the source image and target image that was estimated in the previous step. Third, a candidate match set C_i^e of interest point p_i is found from the interest points in G2 on the target image based on a distance constraint between point and the epipolar line e_{pi}: if the distance between a point to e_{pi} is smaller than a threshold T_e, then this point is regarded as one candidate. Otherwise, it is eliminated as outlier. Then, for each candidate $q_j \in C_i^e$, every two of its LGS vectors are selected, and the intersections Q_{s1} and Q_{s2} of the LGS vectors and epipolar lines e_{s1} and e_{s2} are computed. Points Q_{s1}, Q_{s2} and the candidate point q_j are adopted to form a support region. As the corresponding relationship between the LGS vectors of p_i and the LGS vectors of q_j is unknown before matching, the two cases shown in Figures 6(b) and 6(c) should be considered to avoid mismatching caused by wrong corresponding relationship of LGS vectors. Thereafter, the feature region normalization and descriptor computation methods proposed in Section 2.2 are adopted to compute descriptor for EG-SAF. Finally, the I-NNDR strategy and RANSAC algorithm are conducted to find matches.



(a) (b) (c)

Figure 6. Computation of support region of EG-SAF. (a) support region of interest point on source image, (b) a case of support region of candidate point on target image, and (c) another case of support region of candidate point on target image.

Theoretically, the support region can be determined directly by points Q_{s1}, Q_{s2}, and q_j. However, in practice, errors in fundamental matrix and epipolar line estimation are inevitable, which makes the intersections Q_{s1} and Q_{s2} unreliable. To solve this problem, we find a salient point in the neighborhood of each intersection to replace the intersection and form a support region. The neighborhood is set along the direction of the LGS vector with width $2T_e$. The saliency of each pixel in the neighborhood is computed as Equation (2) and the pixel with the largest saliency value is regarded as the salient point.

$$S(g_k) = \left| I_{average}\left(G_k^l(N)\right) - I_{average}\left(G_k^r(N)\right) \right| \tag{2}$$

where $S(g_k)$ denotes the saliency of pixel g_k, $G_k^l(N)$ and $G_k^r(N)$ are the sets of N pixels on the left and right sides of pixel g_k, and $I_{average}(\ )$ is a function to compute the average of pixel grayscale values.

2.4 Matching expansion

In this section, we design a matching expansion to address the interest points in G3 and the unmatched interest points in the updated G2. In the beginning, a clustering step is performed to check whether an interest point drops in the support region of a matched interest point or not. If an interest point does not drop in the support region of any matched interest point, then this interest point is saved into a set S_{out}. Otherwise, it is saved into set S_{in}. For an interest point $X \in S_{in}$, its support region is determined on the basis of the support region of the matched interest point which it drops in. As shown in Figure 7, (p_i, q_i) is a pair of matched interest points. The blue parallelograms are the support regions of the two points. X is an interest point in set S_{in}. All interest points dropping in the support region of q_i are regarded as the candidate matches. Two salient points are determined and combined with the interest point to form a support region as the support region construction in LGS-SAF and EG-SAF. The salient points are determined according to the following rules: if line segment XA is shorter than line segment XB, then point B is regarded as the first salient point; otherwise, point A is regarded as the first salient point; if line segment XC is shorter than line segment XD, then point D is regarded as the second salient point; otherwise, point C is regarded as the second salient point. After the two salient points on the source image are determined, the salient points on the target image can be determined according to the corresponding relationship ($A \to E$, $B \to F$, $C \to G$, $D \to H$). Then, the support regions and feature descriptors of X and its candidate points can be computed by using the method proposed in Section 2.2.

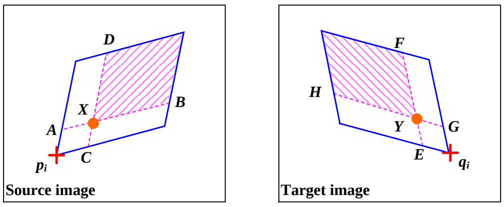


Figure 7. Matching expansion

Besides using the feature descriptor similarity as photometric constraint, affine invariants are also computed to be geometric constraints in our method. As the support region in LGS-SAF or EG-SAF is determined on the basis of a pair of close straight lines, all pixels in the same support region are approximately coplanar. Based on this assumption, the ratio $|XA|/|XB|$ is an affine invariant, where $|XA|$ and $|XB|$ denote the length of line segments XA and XB, respectively. Therefore, if candidate point Y is the correct match of point X, then two equations, $|XA|/|XB| = |YE|/|YF|$ and $|XC|/|XD| = |YG|/|YH|$, are obtained. Considering both the photometric and geometric constraints, we compute the similarity between the source feature and each candidate as Equation (3). If a candidate produces the highest similarity and the similarity value is larger than a threshold T_{sim}, then this candidate will be regarded as the match of the source feature.

$$Sim(X,Y) = \begin{cases} 0, & if \quad abs\left[\left(|XA|/|XB|\right)/\left(|YE|/|YF|\right)-1\right] > \tau \\ 0, & if \quad abs\left[\left(|XC|/|XD|\right)/\left(|YG|/|YH|\right)-1\right] > \tau \\ e^{-\|Desc_X - Desc_Y\|}, & otherwise \end{cases} \qquad (3)$$

where τ is an affine invariant threshold, $Desc_X$ denotes the feature descriptor of point X, and $Desc_Y$ denotes the feature descriptor of candidate Y.

The homography transformation between the source image and target image is estimated through the RANSAC algorithm based on all the matches obtained in the previous steps. The interest points in set S_{out} are matched as follows: First, a feature region with size $T_r \times T_r$ is determined for each interest point on the source image. Second, a parallelogram support region is determined and normalized to a $T_r \times T_r$ feature region for each interest point on the target image. The shape and size of all parallelogram support regions on the target image are the same, which are generated by mapping the $T_r \times T_r$ square with the homography transformation. Then, feature descriptors are computed for the interest points on the source and target images. Finally, feature matching is performed by using the NNDR strategy and epipolar constraint. After matching the interest points in S_{out}, a step of outlier elimination is conducted on the matches generated from all the previous matching procedures by using the RANSAC algorithm to produce the final matches.

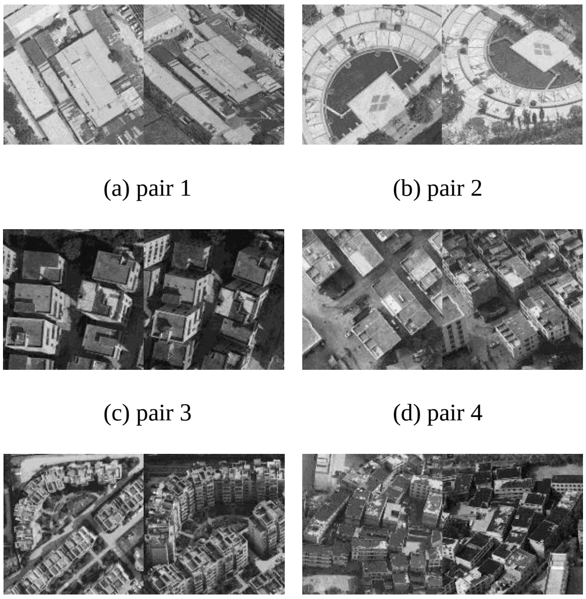
3. EXPERIMENTAL RESULTS AND ANALYSIS

3.1 Experimental datasets

In this study, six pairs of images in typical scenes (Figure 8) are selected to evaluate the robustness of the proposed method on various objects. The information of the experimental datasets, e.g., imagery system (IS), relative flight height (RFH), and ground sample distance (GSD), is provided in Table 1.

No.	IS	RFH	GSD	Image size (Unit: pixel)
1	IQ180	800 m	8 cm	1000×1000
2	IQ180	800 m	8 cm	1000×1000
3	Nikon D810	300 m	6 cm	1024×1024
4	SWDC-5	600 m	8 cm	1000×1000
5	SWDC-5	600 m	8 cm	1200×1200
6	SWDC-5	600 m	8 cm	1200×1200

Table 1. Details of experimental datasets



(a) pair 1

(b) pair 2

(c) pair 3

(d) pair 4

(e) pair 5

(f) pair 6

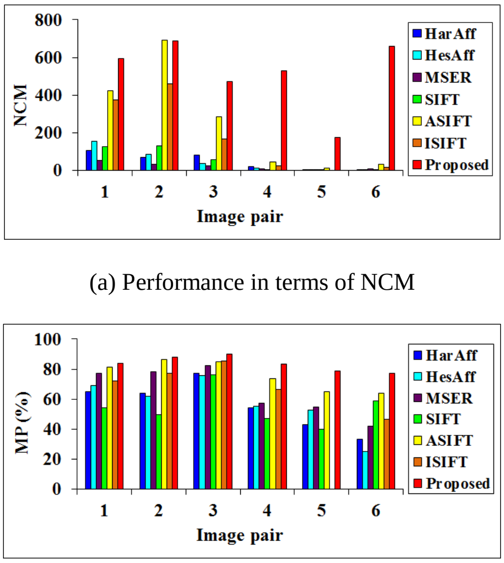
Figure 8. Experimental datasets

3.2 Matching performance evaluation

We compare our approach with six state-of-the-art feature matching methods: HarAff (Harris-Affine detector (Mikolajczyk and Schmid, 2004) combined with SIFT descriptor and NNDR matching strategy), HesAff (Hessian-Affine detector (Mikolajczyk and Schmid, 2004) combined with SIFT descriptor and NNDR matching strategy), MSER (MSER detector (Matas et al., 2004) combined with SIFT descriptor and NNDR matching strategy), SIFT (Lowe, 2004), ASIFT (Morel and Yu, 2009), and ISIFT method (Jiang and Jiang, 2017). The parameters of all comparative methods are set as recommended in the literature. In the proposed method, we adopt the Harris detector (Harris and Stephens, 1988) to detect interest points and the LSD algorithm (Von Gioi et al., 2010) to detect straight lines, and we fix the parameters $m \times m = 11 \times 11$, $s = 20$, $T_r \times T_r = 65 \times 65$, $T_e = 20$, $N = 5$, $\tau = 0.3$, and $T_{sim} = 0.65$ for all experiments.

3.2.1 Quantitative comparison: Two widely used indicators, number of correct matches (NCM) and matching precision (MP), are adopted to evaluate the performance of the proposed method. NCM is computed manually. MP is the percentage of correct matches out of all produced matches. Figure 9 shows the quantitative comparisons, where Figures 9(a) and 9(b) plot the NCM and MP values, respectively. Figure 9(a) shows that HarAff, HesAff, MSER, and SIFT can only obtain a small number of matches on all pairs, indicating that these four methods are sensitive to viewpoint change, especially in urban areas. Compared with the four methods, ASIFT and ISIFT improve the robustness to image viewpoint variation by designing different matching strategies. Among these strategies, ASIFT eliminates geometric distortion between images by simulating the image affine space and achieves improved performance. Particularly in image pair 2, the depth variation is not so significant that the affine space simulated by ASIFT can fit the local geometrical distortion. Thus, ASIFT achieves the best performance. However, on image pairs 4, 5, and 6 where the scene depth changes greatly, many local areas are not covered by the simulated affine space because the simulated affine space is discontinuous. Therefore, ASIFT obtains fewer matches on the three pairs of images. In addition, the feature region computation method in ASIFT encounters difficulty in generating similar feature regions for correspondence located in areas where the parallax is discontinuous. ISIFT is also based on the idea of image simulation and rough correction. However, ISIFT only simulates the entire image once. When the image scene depth changes greatly, the geometric transformation obtained by ISIFT can only correct parts of the image. Therefore, the performance improvement in ISIFT is limited. In addition, ISIFT relies on the performance of initial matching. As shown in Figure 9(a), in image pair 5, accurately estimating the geometric transformation between the images based on initial matches is difficult. Thus, ISIFT fails in the iteration.

Compared with the aforementioned method, the proposed method achieves the best performance in terms of NCM on all pairs of images except image pair 2. The main reason is that the proposed method can adaptively compute feature regions with consistent image content for correspondence according to the local geometrical structure of interest points. Whether the interest points are located in a planar or discontinuous area, the feature regions of corresponding points obtained by the proposed method have high similarity and are easily recognizable in the matching process. In addition, the matching expansion in the proposed method contributes toward obtaining additional matches. Figure 9(b) shows that the proposed method achieves the best performance in terms of MP.

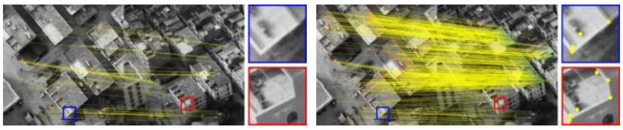


(a) Performance in terms of NCM

(b) Performance in terms of MP

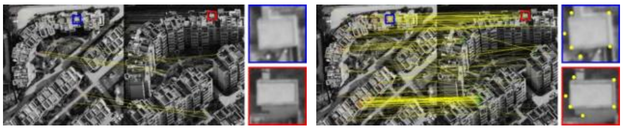
Figure 9. Quantitative comparisons

3.2.2 Qualitative comparison: In this subsection, we compare only the proposed method with ASIFT because the quantitative comparisons show that ASIFT achieves better performance than the other methods. Figures 10-12 present the matching results of ASIFT and the proposed method on image pairs 4, 5, and 6. Matches are linked with yellow lines.



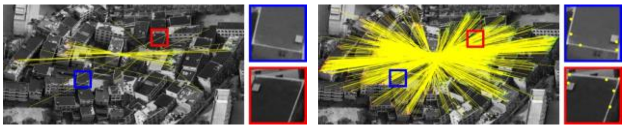
(a) ASIFT (b) Proposed

Figure 10. Matching results of ASIFT and the proposed method on image pair 4



(a) ASIFT (b) Proposed

Figure 11. Matching results of ASIFT and the proposed method on image pair 5



(a) ASIFT (b) Proposed

Figure 12. Matching results of ASIFT and the proposed method on image pair 6

In image pair 4 (Figure 10), most of the matches obtained by ASIFT are located on the ground. Almost all the interest points near the corners and edges of buildings have not been matched successfully. The reason is that the significant viewpoint and scene depth variation make the image content in the feature regions of corresponding interest points in the image area with discontinuous parallax dissimilar. Although ASIFT can alleviate the geometrical distortion to a certain extent by simulating affine space, the similarity of the image content in the regular feature regions centered on interest points is still low. In image pair 5 (Figure 11), ASIFT only obtained a small number of matches and almost all of the interest points on buildings failed to match. The reason is that the structure of buildings in image pair 5 is more complicated than that in image pair 4, and the discrete affine space simulated by ASIFT is difficult to correctly fit the geometrical distortion between corresponding local image areas. In image pair 6 (Figure 12), the intersection angle is more than 90 °, and ASIFT obtains only a small number of matches, which are mostly false.

Compared with ASIFT, the proposed method obtains better results on all three pairs of images. For example, in the enlarged sub-images in Figures 10-12, the proposed method can successfully produce several correct matches, whereas ASIFT fails in these areas. The reason is that the feature region, which is calculated adaptively according to the local structure of interest points in the proposed method, is robust to image viewpoint variation. This condition makes the feature descriptors of corresponding interest points highly similar. The quantitative and qualitative experiments demonstrate that the proposed method can solve the matching problem of wide-baseline images in urban areas.

4. CONCLUSION

In this study, a novel interest point matching method for wide-baseline images in urban areas is proposed. A concept of LGS is defined to guide the matching procedure. On the basis of LGS, interest points are classified into various categories and matched by designing suitable strategies instead of using a uniform matching strategy for all interest points. The experimental results demonstrate that the proposed method performs better than other state-of-the-art feature matching methods for wide-baseline images with significant viewpoint change in urban areas. However, the proposed LGS-based method highly depends on the image content, which indicates that the proposed method is effective for structured images. A possible future work is to improve the construction of LGS and make the proposed method perform well in structured and textured areas.

ACKNOWLEDGEMENTS

This research was supported by the National Natural Science Foundation of China (No. 41631174 and No. 41501492), the National Key Research and Development Program of China (No. 2016YFB0502603), and the Fundamental Research Funds for the Central Universities (No. 2682017CX084).

REFERENCES

Ai M., Hu Q., Li J., et al., 2015. A Robust Photogrammetric Processing Method of Low-Altitude UAV Images. *Remote Sens*., 7, pp. 2302-2333.

Ackermann F., 1984. Digital image correlation: performance and potential application in photogrammetry. *Photogramm. Rec*., 11(64), pp. 429-439.

Fischler M.A., and Bolles R.C., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commu. ACM*, 24(6), pp. 381–395.

Gruen A., 2012. Development and Status of Image Matching in Photogrammetry. *Photogramm. Rec*., 27(137), pp. 36-57.

Gruen, A., Akca, D., 2005. Least squares 3D surface and curve matching. *ISPRS J. Photogramm.*, 59(3), pp. 151-174.

Harris C., and Stephens M., 1988. A combined corner and edge detector. *In Proceedings of the 4th Alvey Vision Conference*. pp. 147-152.

Hu H., Zhu Q., Du Z., et al., 2015. Reliable spatial relationship constrained feature point matching of oblique aerial images. *Photogramm. Eng. Remote Sens*., 81(1), pp. 49-58.

Jiang S., and Jiang W., 2017. On-Board GNSS/IMU Assisted Feature Extraction and Matching for Oblique UAV Images. *Remote Sens*., 9(8), pp. 813.

Lhuillier M., and Quan L., 2002. Match propagation for image-based modeling and rendering. *IEEE T. Pattern Anal*., 24(8), pp. 1140-1146.

Li K., and Yao J., 2017. Line segment matching and reconstruction via exploiting coplanar cues. *ISPRS J. Photogramm.*, 125, pp. 33-49.

Lowe D G., 2004. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis*., 60(2), pp. 91-110.

Matas J., Chum O., Urban M., et al., 2004. Robust wide-baseline stereo from maximally stable extremal regions. *Image Vis. Comput*., 22(10), pp. 761-767.

Mikolajczyk K., and Schmid C., 2004. Scale and affine invariant interest point detectors. *Int. J. Comput. Vis*., 60(1), pp. 63-86.

Mikolajczyk K., Tuytelaars T., Schmid C., et al., 2005. A comparison of affine region detectors. *Int. J. Comput. Vis*., 65(1-2), pp. 43-72.

Morel J.M., and Yu G., 2009. ASIFT: A New Framework for Fully Affine Invariant Image Comparison. *SIAM J. Imaging Sci.*, 2(2), pp. 1-31.

Roth L., Kuhn A., Mayer H., 2017. Wide-baseline image matching with projective view synthesis and calibrated geometric verification. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 85(2), pp. 85-95.

Sun Y., Zhao L., Huang S., et al., 2014. L2-SIFT: SIFT feature extraction and matching for large images in large-scale aerial photogrammetry. *ISPRS J. Photogramm.*, 91, pp. 1-16.

Von Gioi R.G., Jakubowicz J., Morel J.M., et al., 2010. LSD: A fast line segment detector with a false detection control. *IEEE T. Pattern Anal*., 32(4), pp. 722-732.

Yu Y., Huang K., Chen W., et al., 2012. A novel algorithm for view and illumination invariant image matching. *IEEE T. Image Process*., 21(1), pp. 229-240.