Joint Point and Line Segment Matching on Wide-Baseline Stereo Images

Kai Li[1] Jian Yao[†,1,2] Menghan Xia[1] Li Li[1]
[1]School of Remote Sensing and Information Engineering, Wuhan University, P.R. China
[2]State Key Laboratory of Geo-information Engineering, P.R. China
[†]jian.yao@whu.edu.cn

Abstract

This paper presents an method that matches points and line segments jointly on wide-baseline stereo images. In both two images to be matched, line segments are extracted and those spatially adjacent ones are intersected to generate V-junctions. To match V-junctions from the two images, we extract for each of them an affine and scale invariant local region and describe it with SIFT. The putative V-junction matches obtained from evaluating their description vectors are refined subsequently by the epipolar line constraint and topological distribution constraint among neighbor V-junctions. Since once a pair of V-junctions are matched, the two pairs of line segments forming them are matched accordingly. A part of line segments from the two images are therefore matched along with V-junction matches. To get more line segment matches, we further match those left unmatched line segments by the local homographies estimated from their adjacent V-junction matches. Experiments verify the robustness of the proposed method and its superiority to both some famous point and line segment matching methods on wide-baseline stereo images. In addition, we also show the proposed method can make it easier for 3D line segment reconstruction.

1. Introduction

Image matching is a vital procedure for many high-level computer vision problems, such as 3D reconstruction, structure from motion, object recognition, etc. A common pipeline for image matching commences image feature extraction, like feature points, line segments, edges, followed by feature matching. An image matching method often targets to match only one type of image feature. If we want to use several features combinatorially for higher level applications, like 3D reconstruction, several methods have to be applied on the same images in order, which is less efficient. Several line segment matching methods [1, 2, 3] exploit point matching results to help match line segment. These methods first get point matches using the existing methods and then match the extracted line segments by utilizing the obtained point matches. A common problem for these methods is that the line segment matching results rely heavily on the obtained point matches.

Unlike those image matching methods which match points or line segments separately, or those use point matching results for line segment matching, this paper presents a new algorithm that matches points and line segments jointly through matching V-junctions generated by intersecting adjacent line segments. This can be achieved because once a pair of V-junctions from two images are matched, the corresponding relationship between the two pairs of line segments forming them are established accordingly. With this idea, a foremost problem to be solved is how to match V-junctions from two images. We propose to extract from each V-junction a scale and affine invariant region and describe it with SIFT. Through evaluating the description vectors of V-junctions from two images, we get some putative V-junction matches, and some line segment matches. Besides, we propose an effective strategy to refining the obtained putative V-junction matches by exploiting the epipolar line constraint and topological distribution constraint among neighbor V-junctions. Another crucial problem to be solved is how to match line segments that are spatially separated with others and are not used to form V-junctions, and can not therefore be matched along with V-junctions. The solution we propose is to match them by estimating local homographies from their neighboring matched V-junctions. Experiments show that our method is more robust than some famous point matching methods and can produce more line segment matches than the state-of-the-art line segment matching methods with higher accuracy in most cases. In addition, we also show our result can facilitate to reconstruct 3D line segments.

2. Related Works

We present in this section first some point matching methods and then some line segment matching methods related to our method.

2.1. Point Matching Methods

Point-based image matching has been widely investigated and numerous methods have been proposed. The widely acknowledged methods are those first extract invariant local regions and then describe these regions with some kind of descriptors. The most famous local region detectors are maximally stable external regions (MSER) [4], edge-based regions (EBR) [5], scale-invariant feature transform (SIFT) [6], and a more recent one presented in [7], etc. In [8], some of the most famous local region detectors were compared. The same authors also gave a summarize and evaluation of local region description methods in [9]. A more recent and comprehensive survey about local region extraction and description methods is in [10].

2.2. Line Segment Matching Methods

Line segment matching methods in existing literatures can generally be classified into two categories: methods that match line segments individually and those in groups. Some methods matching line segments as individuals exploit the photometric information associated with individual line segments, such as intensity [11, 12], gradient [13, 14, 15], and color [16] in the local regions around line segments. All these methods underlie the assumption that there are considerable overlaps between corresponding line segments, which leads to the failure of these methods when corresponding line segments share insufficient corresponding parts. Other methods matching line segments as individuals leverage point matches for line segment matching [17, 1, 2]. These methods first find point matches using the existing point matching methods, and then exploit invariants between coplanar points and line(s) under certain image transformations to evaluate line segments from two images. Line segments which meet the invariants are regarded to be in correspondence. A common disadvantage of these methods is that they depend heavily on the point matching results and once insufficient point matches were found before, these methods will generate inferior results. Methods matching line segments in groups are more complex, but more constraints are available for disambiguation. In [18], the stability of the relative positions of the endpoints of a group of line segments in a local region under various image transformations is exploited. This method is robust in some challenging situations. However, the dependence on the approximately corresponding relationship between the endpoints of line segment correspondences leads to the tendency of this method to produce false matches when substantial disparity exists in the locations of the endpoints.

A more common way to match line segments in groups is to match them in pairs. Our method, and the two methods [3, 19] from which our method derives, adopt this way. Compared with the two methods, our method makes improvements in these aspects. When compared with [19], first, a more robust and efficient (without the time-consuming procedure of sampling the scales of local regions as that did in [19]) way is proposed to match V-junctions of line segments from two images. We extract from each junction an affine and scale invariant region and describe it with SIFT. Second, an effectively iterative scheme is proposed to refine the obtained putative V-junction matches. Third, we propose to match individual line segments which can not be matched along with V-junctions by using the local homographies estimated from their neighboring matched V-junctions. Both the V-junction refinement procedure and the individual line segment matching procedure are absent in [19]. Our method is therefore more robust and can get more line segment matches. When compared with [3], on the one hand, rather than relying on the point matches obtained by some external point matching methods to provide the global and local constraints for disambiguation, our method is self-sustaining. It generates point matches itself to help match line segments. On the other hand, though both our method and [3] use local homographies for matching line segments, the major difference lies in the way of estimating the local homographies (with also some minor differences on applying them). In [3], a local homography is estimated from at least 4 pairs of assumed coplanar point correspondences obtained by SIFT, while in our method, a local homography is estimated only from a pair of V-junction correspondences and the fundamental matrix. Both of them are obtained by our method in previous stages and can therefore always be guaranteed.

The remaining parts of this paper are organized as follows. Section 3 presents the ways we generate, describe and matching V-junctions. The strategies of matching those left individual line segments and 3D line segment reconstruction are introduced in Section 4 and Section 5, respectively. The experimental results are shown in Section 6 and the conclusions are drawn in Section 7.

3. V-Junction Matching

We present the ways of generating, describing and matching V-junctions orderly in this section.

3.1. V-Junction Generation

Only the intersecting junctions of 2D line segments whose 3D correspondences are coplanar in the scene are stable with camera motions and can find their 2D correspondences in other images. Therefore, we need to determine the 2D line segments whose 3D correspondences are coplanar in 3D space to generate junctions. However, it is hardly possible to determine the 3D coplanarity of 2D line segments only from a image without the projective information of the camera. But adjacent line segments possess a higher probability to be coplanar in 3D space due to the spatial proximity. So, it is an alternative way to intersect

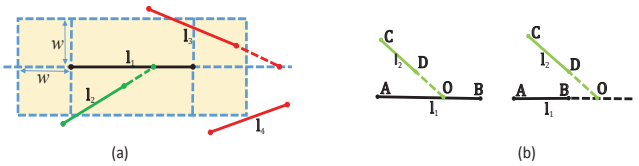

Figure 1: Constructing V-junction: (a) Finding 2D line segments whose 3D correspondences are possibly coplanar in 3D space; (b) Two situations where the two line segments are used to construct V-junctions.

neighboring line segments to get reliable junctions. We use a similar method as that presented in [19] to generate junctions. Refer to Figure 1(a), for a line segments $\mathbf{l}_1$, we define the rectangle (filled in yellow in the figure), $\mathcal{R}$, as its affect region, which centers at the midpoint of $\mathbf{l}_1$ and has the width of $|\mathbf{l}_1|+2w$ and the height of $2w$, where $|\mathbf{l}_1|$ denotes the length of $\mathbf{l}_1$ and w is a user-defined parameter. Any line segment satisfying the following two conditions is assumed to be coplanar with $\mathbf{l}_1$ in 3D space. First, at least one of the two endpoints drops in $\mathcal{R}$. Second, its intersection with $\mathbf{l}_1$ also drops in $\mathcal{R}$. Under these two conditions, in Figure 1(a), only $\mathbf{l}_2$ is accepted to be coplanar with $\mathbf{l}_1$ in 3D space.

There exist two situations where two line segments assume to be coplanar in 3D space, as shown in Figure 1(b). In the left case where the intersection lies on one of the two line segments (not on their extensions), two V-junctions, $\widehat{\mathbf{AOC}}$ and $\widehat{\mathbf{BOC}}$ are constructed. In the right case where the intersection lies on the extensions of the two line segments, only one V-junction, $\widehat{\mathbf{AOC}}$ are constructed.

3.2. V-Junction Description

To match the constructed V-junctions from two images, we first exact from each V-junction a scale and affine-invariant region by exploiting the relationship between the V-junction and the two line segments forming it and other neighbor line segments, and then describe the region with SIFT descriptor.

For a V-junction, we detect from each of the two line segments forming it a stable point which has the largest intensity change among all points along the line segment. Our idea derives from the edge-based region (EBR) detector proposed in [5], which exploits features of edges to detect invariant regions. We exploit features of line segments to extract invariant regions for V-junctions. We incorporate the relationship between neighbor line segments to make the extracted invariant regions less sensitive to noise and location imprecision of line segments.

Suppose $\mathbf{l}$ is one of the two line segments forming a V-junction, $\mathcal{V}$. If it was used to construct V-junctions besides $\mathcal{V}$, as the line segment $\overline{\mathbf{AB}}$ shown in Figure 2(a), the junction points are selected as the candidates of the stable point. This is reasonable because line segments lie on region borders, and if $\mathbf{l}$ meets a line segment, it likely reaches a region

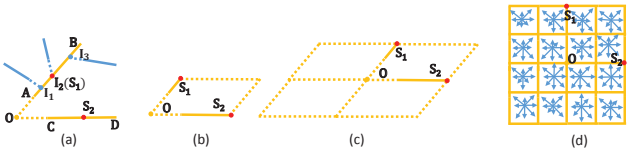

Figure 2: An illustration of the scale and affine invariant local region extraction and description procedures of our proposed method: (a) Finding stable points $\mathbf{S}_1$ and $\mathbf{S}_2$ for line segments $\overline{\mathbf{AB}}$ and $\overline{\mathbf{CD}}$ in the V-junction $\widehat{\mathbf{BOD}}$; (b) The extracted local region; (c) The expanded local region; (d) Describing the normalized local region with SIFT.

Figure 3: An example of finding scale and affine invariant local regions on two stereo images [22] with severe viewpoint change. The parallelograms drawn in different colors are the extracted local regions. Only a subset of all parallelograms are shown in the two images for better interpretation.

border. The junction of $\mathbf{l}$ and the line segment it meets is very possibly to be a stable point. On the other hand, if $\mathbf{l}$ was not used to construct any V-junction besides $\mathcal{V}$, as the line segment $\overline{\mathbf{CD}}$ shown in Figure 2(a), all points on $\mathbf{l}$ are regarded as candidates stable points. In this way, we collect a set of candidate stable point $\mathcal{C} = \{\mathbf{S}_i\}_{i=1}^{N}$, where N is the number of candidates. The best candidate is the one which has the most abrupt intensity change along the line segment. For a candidate $\mathbf{S}_i$, we collect 5 points on both its sides along $\mathbf{l}$ and gather two sets of the pixel intensities, $\mathcal{I}_1(\mathbf{S}_i)$ and $\mathcal{I}_2(\mathbf{S}_i)$. The intensity change of $\mathbf{S}_i$ is calculated as: $I_d(\mathbf{S}_i) = |median(\mathcal{I}_1(\mathbf{S}_i)) - median(\mathcal{I}_2(\mathbf{S}_i))|$, where $median(\cdot)$ means the median value of the elements in a set. The $\mathbf{S}_i$ with the maximal value of $I_d(\mathbf{S}_i)$ is selected as the final stable point $\mathbf{S}$. Median value of a set of intensities is used to compute the intensity change of a candidate point because this can reduce the influence of noise among the sets.

After finding a stable point for both line segments forming a V-junction as $\mathbf{S}_1$ for the line segment $\overline{\mathbf{AB}}$ and $\mathbf{S}_2$ for the line segment $\overline{\mathbf{CD}}$ shown in Figure 2, the parallelogram determined by these two stable points and the junction is identified as the invariant region for the V-junction as shown in Figure 2(b). Figure 3 shows the extracted invariant regions by our method on two images with different sizes and

great viewpoint change. In this extreme case, some (nearly) identical regions are extracted in the two images.

Since most signal variations exist near line segments, we expand the extracted parallelogram around the junction point into a larger one, as shown in Figure 2(c). Next, we normalize the expanded region into a square through affine transformation to make it affine-invariant. Finally, we describe the square with SIFT. The size of the square is suggested to be 41×41 in [8]. But we found in our case it produced better matching results when the size of the square is set as 21×21.

3.3. V-Junction Matching

To match V-junction from two images, the general way is to evaluate the Eucliean distances between their description vectors. But since the two line segments forming a V-junction locate in a local region, the crossing angle of them should vary at a small range under most image transformations. Let $(\mathcal{V}, \mathcal{V}')$ be a pair of V-junctions to be matched and (θ, θ') be the crossing angles of the two pairs of line segments. If $(\mathcal{V}, \mathcal{V}')$ is a correct match, the difference between θ and θ' should be less than a small threshold ϵ_1 (set as 30° in this paper), *i.e.*, $|\theta - \theta'| < \epsilon_1$. This simple constraint can help discard many false candidates before evaluating their description vectors and thus contributes to better matching results.

There inevitably exist false ones among the putative V-junction matches obtained after the above procedure. We refine them by the following two ways. First, we estimate the fundamental matrix for the two images using the putative V-junction matches and keep only the inliers. This epipolar line constraint can filter out those false matches that lie near the corresponding epipolar lines. We use another effective way to further refine the obained V-junction matches by exploiting the stability of the topological distribution of a group of V-junctions in a local region.

Refer to Figure 4, for a V-junctions $\mathcal{V}_c$, the two line segments forming it and their reverse extensions form a coordinate-like structure. Its neighbors distribute in the four quadrants. This topological distribution is relatively stable with image transformations. *i.e.*, after some kinds of image transformations, while $\mathcal{V}_c$ is transformed into $\mathcal{V}'_c$, its neighbors should change consistently. To apply this constraint for refining V-junction matches, for the candidate V-junction match, $(\mathcal{V}_c, \mathcal{V}'_c)$, we collect the K ($K = 10$ used in this paper) nearest matched V-junctions as $\tilde{\mathcal{N}} = \{\mathcal{V}_i\}_{i=1}^K$ and $\tilde{\mathcal{N}}' = \{\mathcal{V}'_j\}_{j=1}^K$ for $\mathcal{V}_c$ and $\mathcal{V}'_c$, respectively. If $(\mathcal{V}_c, \mathcal{V}'_c)$ is a correct match, the following two conditions must be satisfied. The first one is that there should exist a sufficient large proportion (0.5 used in this paper) of correspondences in $\tilde{\mathcal{N}}$ and $\tilde{\mathcal{N}}'$. Second, the matches in $\tilde{\mathcal{N}}$ and $\tilde{\mathcal{N}}'$ that have its two V-junctions lying in the same quadrants of the coordinates formed by $\mathcal{V}_c$ and $\mathcal{V}'_c$ should account for a great ratio of the total matches; the ratio is set as 0.8 in this paper.

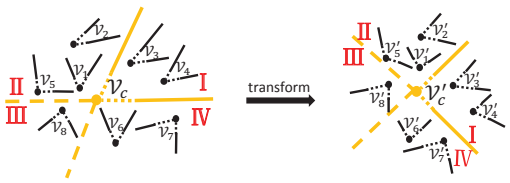

Figure 4: An illustration of the topological distribution of a V-junction $\mathcal{V}_c$ and its neighbor V-junctions before and after image transformations.

With the guidance of the epipolar line geometry between the two images and topological distribution constraint among neighboring V-junctions, the V-junctions from the two images can be matched exhausted by alternatively adding new matches and deleting false ones until no more match can be added.

4. Individual Line Segment Matching

Line segments that have not been matched along with V-junctions will be further matched as individuals. They are first grouped according to those matched V-junctions, and then matched in corresponding groups based on the local homographies estimated from the pairs of V-junction correspondences in corresponding groups.

4.1. Local Homography Estimation

We have assumed the two line segments forming a V-junction are coplanar in 3D space. Therefore, a V-junction match generates two coplanar line segment matches, which can be used to estimate a local homography with the combination of the estimated fundamental matrix.

A planar homography $\mathbf{H}$ is determined by eight degrees of freedom, necessitating 8 independent constraints to find a unique solution. However, when the fundamental matrix $\mathbf{F}$ between two images is known, then $\mathbf{H}^\top\mathbf{F}$ is skew-symmetric [23], *i.e.*,

$$\mathbf{H}^\top\mathbf{F} + \mathbf{F}^\top\mathbf{H} = 0. \tag{1}$$

The above equation gives five independent constraints on $\mathbf{H}$, and the other three are required to fully describe a homography. One line match provides two independent constraints [24], resulting in that the system is over-constrained since two coplanar line matches exist in our case.

The homography induced by a 3D plane $\boldsymbol{\pi}$ can be represented as

$$\mathbf{H} = \mathbf{A} - \mathbf{e}'\mathbf{v}^\top, \tag{2}$$

where the 3D plane is represented by $\boldsymbol{\pi} = (\mathbf{v}^\top, 1)$ in the projective reconstruction with camera matrices $\mathbf{C} = [\mathbf{I}|\mathbf{0}]$ and $\mathbf{C}' = [\mathbf{A}|\mathbf{e}']$. The homography maps a point in one 2D plane to another 2D plane. For a line segment match

$(\mathbf{l}, \mathbf{l}')$, suppose $\mathbf{x}$ is an endpoint of $\mathbf{l}$, the homography maps it to its corresponding point $\mathbf{x}'$ as: $\mathbf{x}' = \mathbf{H}\mathbf{x}$. Since $\mathbf{l}$ and $\mathbf{l}'$ correspond with each other, $\mathbf{x}'$ must be a point lying on $\mathbf{l}'$, that is ${\mathbf{l}'}^\top \mathbf{x}' = 0$. Therefore, we obtain

$$ {\mathbf{l}'}^\top (\mathbf{A} - \mathbf{e}'\mathbf{v}^\top)\mathbf{x} = 0. \tag{3} $$

Arranging the above equations, we finally get

$$ \mathbf{x}^\top \mathbf{v} = \frac{\mathbf{x}^\top \mathbf{A}^\top \mathbf{l}'}{\mathbf{e}'^\top \mathbf{l}'}, \tag{4} $$

which is linear in $\mathbf{v}$. Each endpoint of a line segment in a line match provides an equation, and two line segment matches totally provide four constraint equations. A least-square solution of $\mathbf{v}$ can be obtained from the four equations. The local homograpy $\mathbf{H}$ is then computed from Eq. (2).

4.2. Individual Line Segment Matching

Let $\mathcal{M} = \{(\mathcal{V}_m, \mathcal{V}'_m)\}_{m=0}^{T}$ be the set of T V-junction matches from two images, where $(\mathcal{V}_m, \mathcal{V}'_m)$ denotes the m-th V-junction match found before. Let $\mathcal{K} = \{\mathbf{l}_i\}_{i=1}^{M}$ and $\mathcal{K}' = \{\mathbf{l}'_j\}_{j=1}^{N}$ be the two groups of individual line segments, which have not been matched before, from the two images, respectively. For each individual line segment $\mathbf{l}_i \in \mathcal{K}$ or $\mathbf{l}'_j \in \mathcal{K}'$, we find four of its nearest matched V-junctions and assign it into the corresponding 4 groups. After that, any matched V-junction collects zero to multiple individual line segment(s). Individual line segments collected by corresponding V-junctions from two images are assessed and matched separately.

Suppose $\mathbf{l}$ and $\mathbf{l}'$ are a pair of individual line segments to be evaluated and they are collected by the matched V-junctions, $\mathcal{V}$ and $\mathcal{V}'$, respectively. We first check whether the direction difference of them is consistent with the direction differences of the two pairs of matched line segments brought by $\mathcal{V}$ and $\mathcal{V}'$. The directions of adjacent line segments should change similarly under image transformations. Let σ_1 be the mean value of the direction differences of the two pairs of line correspondences brought by $\mathcal{V}$ and $\mathcal{V}'$ and σ_2 be the direction difference of $\mathbf{l}$ and $\mathbf{l}'$. If $|\sigma_2 - \sigma_1| < \epsilon_2$, where ϵ_2 is a user-defined threshold set as 20° in this paper, we accept ($\mathbf{l}$, $\mathbf{l}'$) temporarily and take it for further evaluation. Next, we test ($\mathbf{l}$, $\mathbf{l}'$) again by using the brightness constraint [16], which requires the brighter side of two corresponding line segments should be the same. The brighter side of a line segment refers to the side where the average intensity of pixels in a small profile along the line segment is greater than the other side.

If $(\mathbf{l}, \mathbf{l}')$ satisfies the above constraints, we then evaluate it by the local homography $\mathbf{H}$ estimated from $\mathcal{V}$ and $\mathcal{V}'$. We map $\mathbf{l}$ and $\mathbf{l}'$ by $\mathbf{H}$, generating their correspondences $\mathbf{l}_h$ for $\mathbf{l}$ and $\mathbf{l}'_h$ for $\mathbf{l}'$, respectively. The average of the four distances, including the perpendicular distances of two endpoints of $\mathbf{l}'_h$ to $\mathbf{l}$ and the perpendicular distances of the two endpoints of $\mathbf{l}_h$ to $\mathbf{l}'$, is defined as the mapping error of ($\mathbf{l}$, $\mathbf{l}'$), which is used to measure the similarity between $\mathbf{l}$ and $\mathbf{l}'$. After that, there may exist cases that one line segment in one image is matched with several line segments in the other image. We select the pair with the minimal mapping error as the correct match and reject others.

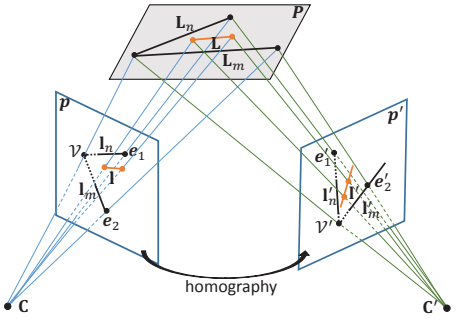

Figure 5: An illustration of reconstructing 3D line segment by reconstructing the corresponding endpoints of line segments through exploiting the local homography.

5. 3D Line Segment Reconstruction

Traditional way to get 3D line segments is to triangulate corresponding line segments from multiple views [21, 20, 26], which is often complex and inefficient. We present here a simple but effective line segment reconstruction method by taking advantage of the result obtained by our image matching method. Refer to Figure 5, $\mathcal{V}$ and $\mathcal{V}'$ are two V-junctions in correspondence identified by our method. They are formed by the pairs of line segments ($\mathbf{l}_m$, $\mathbf{l}_n$) and ($\mathbf{l}'_m$, $\mathbf{l}'_n$), respectively. ($\mathbf{l}$, $\mathbf{l}'$) is a pair of corresponding individual line segments matched by using the strategies presented in Section 4 when they are grouped into $\mathcal{V}$ and $\mathcal{V}'$, respectively. Since we can estimate the local homography from $\mathcal{V}$ and $\mathcal{V}'$ and this local homography can establish point-to-point correspondence, we can transfer reconstructing 3D line segments into reconstructing their endpoints. As shown in Figure 5, we find in the second plane $\mathbf{p}'$ the correspondence of each endpoint of line segments in the first plane $\mathbf{p}$, and triangulate corresponding endpoints to get 3D line segments in scene plane $\mathbf{P}$.

6. Experimental Results

We present in this section first how we fixed the parameters of the method and then the performance evaluation of our local region detector, followed by the point and line segment matching results. At last, we give an example 3D line segment reconstruction result obtained by our method.

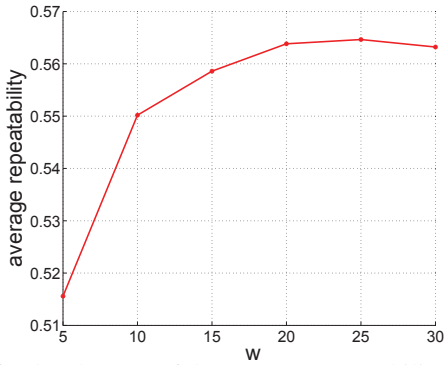

Figure 6: The changes of the average repeatability with different values of the parameter w.

6.1. Parameters Setting

Our proposed method has some parameters, but only the parameter w is crucial which determines how adjacently two line segments lie that they are used to intersect to form a junction. The remaining parameters are used to strengthen some constraints and the fluctuations of their values make no great differences on the matching results. We fixed them after some initial experiments, and their defaulted values were presented in the preceding text. Here, we present only how we fixed w.

A bigger w will result in more V-junctions, and more resultant V-junction matches. However, excessive V-junctions, especially when many of them can not find their correspondences in another group of V-junctions, hamper the matching since more interferences are involved. Besides, both more computation time and memory are required to match them. We adopted the method introduced in [8] by calculating the repeatability of the extracted local regions for V-junctions from different images to select a proper value for w. We tested on the famous datasets, *graffiti*, *leuven*, *boat*, *bikes* and *ubc* [8], in which the images are related by global homographies. We sampled w from 5 to 30 at the step of 5 and calculated the average repeatability for all image pairs with each w. The change of the average repeatability along with w is shown in Figure 6. We can observe from this figure that the repeatability curve increases when w is less than 20, and is stable until w is bigger than 25, where the curve begins to drop. Thus, both 20 and 25 are proper values for w. To obtain less junctions and reduce the computation time, $w = 20$ was selected in this paper.

6.2. Evaluation of the Local Region Detector

One of the main contributions of this paper is the local region detector for V-junctions. We evaluated it using the same strategy presented in [8] by calculating the repeatability of the extracted local regions in different images. We adapted our detector into the evaluation framework proposed in [8] and compared the repeatability obtained by our detector with those embedded detectors in the framework.

	bikes	*boat*	*leuven*	*graffiti*	*ubc*
Ours	3	4	4	2	6
Hessian-affine	1	1	2	5	1
Harris-Affine	4	3	3	6	2
MSER	5	2	1	1	3
IBR	6	6	5	3	5
EBR	2	5	6	4	4
Salient	7	7	7	7	7

Table 1: The relative ranks of our detector compared to other detectors with respect to the repeatability of the extracted local regions on some datasets established in [8].

	(a)	(b)	(c)	(d)	(e)	(f)	(g)
SIFT	9	6	132	60	2	65	6
Ours	200	321	585	206	80	169	137

Table 2: The numbers of total point matches obtained by our method and SIFT on the 7 image pairs shown in Figure 8.

Table 1 shows the relative ranks our detector compared to other detectors on some of the established datasets. Note that we got the relative ranks shown in this table when we set the overlap error parameter as 20%. The relative ranks may change slightly with different overlap error parameter. In this table, 1 represents the highest repeatability while 7 represents the lowest repeatability. As we can see from this table, though our detector is not the generally best detector, it does have some advantages over some detectors. Besides, our detector is specially designed to get V-junction matches. When matching V-junctions, apart from evaluating their description vectors, the important crossing angle constraint (see Section 3.3) can be applied to help disambiguation. Our local region detector, when combined with this additional constraint, can be very discriminative for matching V-junctions, as shown in the subsequent section.

6.3. Point Matching Results

We used the seven image pairs, (a)∼(g) shown in Figure 8, to evaluate our method for point matching. All these image pairs were collected from publicly available datasets [18, 27, 8, 15] and are characterized by some extreme image transformations, namely, viewpoint, scale and rotation changes, non-uniform light change; or poorly-textured scene (image pair (e)). We tested on these extreme image pairs to show the robustness of our methods. For comparison, we also show the corresponding results obtained by SIFT[1]. The visualized results are shown in the first two columns of Figure 8 and the corresponding statistical results are shown in Table 2. We can obverse that our method has an overwhelming advantage on the number of the obtained total matches on all these image pairs. Though we did not count the numbers of correct matches among the total matches, which is tedious and error-prone, the much more

[1]The implementation is from http://www.cs.ubc.ca/ lowe/keypoints/

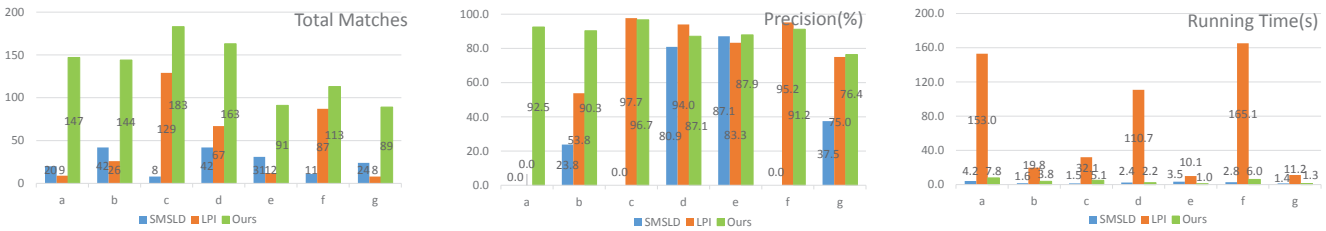

Figure 7: Line segment matching results obtained by our method, LPI [2] and SMSLD [15] on the 7 image pairs shown in Figure 8. We evaluated the three methods by comparing the three measures: the total number of matches, the precision of the obtained matches and the running time.

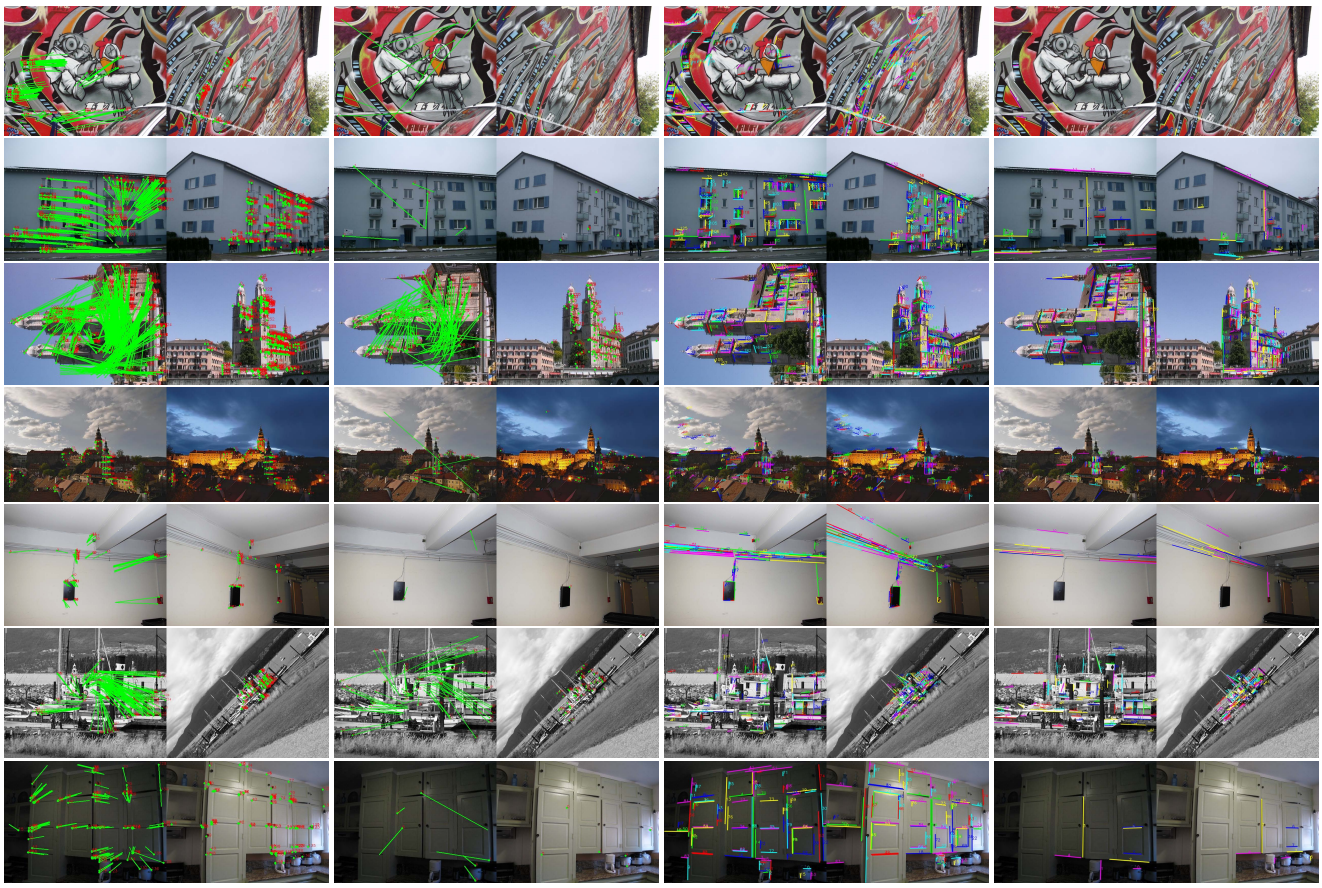

Figure 8: The visualized point matching and line segment matching results of our method and other methods on 7 representative image pairs. These images pairs will be referred as (a)∼(g) in order for later use. The first two columns visualize the point matching results obtained by our method and SIFT, respectively. For better accessing the obtained point matching results, for each image pair, we drawn the optical flow of the matched points in the first image and label two corresponding points from the two images with the same number. The corresponding statistical results are listed in Table 2. The last two columns visualize the line segment matching results obtained by our method and LPI [1], respectively. Two corresponding line segments from two images are drawn in the same color and are labeled with the same number at the middle. The corresponding statistical results are listed in Figure 7. Please zoom in for better interpretation of the results.

consistent optical flows shown in Figure 8 of the matched points in the first images for our method than that of SIFT indicates the accuracies of our results are much higher than SIFT. Thus, compared with SIFT, our method outperforms both in the number of the total correct matches and the accuracy. This proves the robustness and effectiveness of our proposed local region detector, matching strategy, and match-refinement strategies.

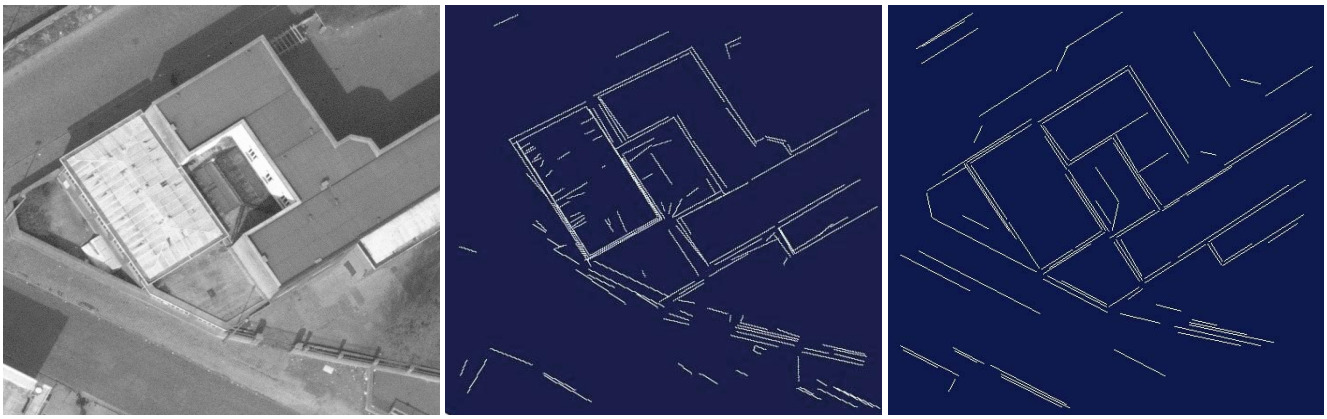

Figure 9: 3D line segment reconstruction results obtained by our method and the method presented in [26]. (Left): One of the used images. (Middle): 3D line segment reconstruction result obtained by our method using the first two images of the total six images. (Right): 3D line segment reconstruction result obtained by the method presented in [26] using all six images.

6.4. Line Segment Matching Results

We also compared our method with some state-of-the-art of line segment matching methods. Two methods with the implementations publicly available were used for the comparison. They are Line-Point-Invariant (LPI) [2] and SMSLD [15]. To make the comparison convincing, we followed the two methods and used LSD [25] for line segment extraction. The comparative results are shown in Figure 7 and the visualized the results obtained by our method and LPI are shown on the last two columns in Figure 8. We can observe from Figure 7 that our method has an overwhelming advantage on the number of total matches over the other two methods, and achieves the highest accuracies for most of the image pairs. It is remarkable that on image pairs (a), (c) and (f), the accuracies of the results obtained by SMSLD are 0, which means all its obtained matches are wrong. This situation also occurs for LPI on image pair (a) that all the obtained nine matches are wrong. This is because, as mentioned before, LPI relies on the point matches obtained by SIFT, but SIFT could not get enough correct point matches on this image pair. The same reason also explains why LPI obtained line segment matches with low accuracies in image pair (b), (e) and (g), where SIFT get inferior results, as shown in the first two columns of Figure 8 and Table 2. As to running time, SMSLD is slightly better than our method, and much better than LPI.

6.5. Line Segment Reconstruction Results

Line segment reconstruction is not the main focus of this paper. We present here only a sample result we have obtained to show the feasibility of our proposed method presented in Section 5. We will keep working on this area and report our algorithm in the future publications. The sample result is shown in Figure 9. There are six images for the scene, and we used only the first two to generate our result. As we can see, the outline of the scene is well reconstructed by our method and our result is just slightly inferior to that obtained by [26] itself, which were obtained from all the six images. It is sure that a better result can be obtained when more images are employed. We will extend our method to multiple views in the near future.

7. Conclusions

We have introduced in this paper a new image matching method that can get point matches and line segment matches jointly on wide-baseline stereo images by exploiting V-junctions of adjacent line segments. To match V-junctions from two images, we propose to extract from each of them an affine and scale invariant local region and describe it with SIFT, followed by an effective matching strategy. For those line segments that can not be matched along with V-junctions, we propose to match them by the local homographies estimated from their adjacent V-junction matches. Comparisons with both the state-of-the-art point and line segment matching methods verify the robustness and superiority of the proposed method. In addition, we also show our method can facilitate 3D line segment reconstruction.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Project No. 41571436), the Hubei Province Science and Technology Support Program, China (Project No. 2015BAA027), State Key Laboratory of Geo-information Engineering (Project No. SKLGIE2014-M-3-6), the Jiangsu Province Science and Technology Support Program, China (Project No. BE2014866), and Key Laboratory for National Geographic Census and Monitoring, National Administration of Surveying, Mapping and Geoinformation.

References

[1] B. Fan, F. Wu, and Z. Hu. Line matching leveraged by point correspondences. In *CVPR*, 2010. 1, 2, 7

[2] B. Fan, F. Wu, and Z. Hu. Robust line matching through line-point invariants. *Pattern Recognition*, 45(2):794–805, 2012. 1, 2, 7, 8

[3] M. Al-Shahri and A. Yilmaz. Line matching in wide-baseline stereo: a top-down approach. *IEEE Transactions on Image Processing*, 23(9):4199-4210, 2014. 1, 2

[4] J. Matas, O. Chum, M. Urban and T. Pajdla Robust wide-baseline stereo from maximally stable extremal regions. *Image and vision computing*, 22(10), 761-767, 2004. 2

[5] T. Tuytelaars and L. Van Gool. Matching widely separated views based on affine invariant regions. *International Journal of Computer Vision*, 59(1):61-85, 2004. 2, 3

[6] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004. 2

[7] H. Deng, W. Zhang, E. Mortensen, T. Dietterich, and L. Shapiro Principal curvature-based region detector for object recognition. In *CVPR*, 2007. 2

[8] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, and L. Van Gool. A comparison of affine region detectors. *International Journal of Computer Vision (IJCV)*, 65(1-2), 43-72. 2, 4, 6

[9] K. Mikolajczyk and C. Schmid . A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 27(10):1615–1630, 2005. 2

[10] J. Bernal, F. Vilarino, and J. Snchez Feature detectors and feature descriptors: Where we are now. *Universitat Autonoma de Barcelona, Barcelona*, 2010. 2

[11] C. Schmid and A. Zisserman. Automatic line matching across views. In *CVPR*, 1997. 2

[12] C. Baillard, C. Schmid, A. Zisserman, and A. Fitzgibbon. Automatic line matching and 3D reconstruction of buildings from multiple views. In *ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery*, 1999. 2

[13] Z. Wang, F. Wu, and Z. Hu. MSLD: A robust descriptor for line matching. *Pattern Recognition*, 42(5):941–953, 2009. 2

[14] L. Zhang and R. Koch. An efficient and robust line segment matching approach based on LBD descriptor and pairwise geometric consistency. *Journal of Visual Communication and Image Representation*, 24(7):794-805, 2013. 2

[15] B. Verhagen, R. Timofte, and L. Van Gool. Scale-invariant line descriptors for wide baseline matching. In *WACV*, 2014. 2, 6, 7, 8

[16] H. Bay, V. Ferrari, and L. Van Gool. Wide-baseline stereo matching with line segments. In *CVPR*, 2005. 2, 5

[17] M. I. Lourakis, S. T. Halkidis, and S. C. Orphanoudakis. Matching disparate views of planar surfaces using projective invariants. *Image and Vision Computing*, 18(9):673–683, 2000. 2

[18] L. Wang, U. Neumann, and S. You. Wide-baseline image matching using line signatures. In *ICCV*, 2009. 2, 6

[19] H. Kim, S. Lee, and Y. Lee. Wide-baseline stereo matching based on the line intersection context for real-time workspace modeling. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, 31(2):421–435, 2014. 2, 3

[20] M. Hofer, M. Donoser, H. Bischof. Semi-global 3D line modeling for incremental structure-from-motion. In *BMVC*, 2014. 5

[21] M. Hofer, M. Maurer, and H. Bischof. Improving sparse 3D models for man-made environments using line-based 3D reconstruction. In *3DV*, 2014. 5

[22] D. Mishkin, M. Perdoch, and J. Matas. Two-view matching with view synthesis revisited. In *IVCNZ*, 2013. 3

[23] Q. -T. Luong and T. Viéville. Canonical representations for the geometries of multiple projective views. *Computer Vision and Image Understanding*, 64(2):193–229, 1996. 4

[24] A. Anubhav, C. V. Jawahar, and P. J. Narayanan. A survey of planar homography estimation techniques. *Centre for Visual Information Technology, Tech. Rep. IIIT/TR/2005/12*, 2005. 4

[25] R. G. Von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall. LSD: A fast line segment detector with a false detection control. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 32(4):722–732, 2012. 8

[26] T. Werner and A. Zisserman: New techniques for automated architectural reconstruction from photographs. In *ECCV*, 2002. 5, 8

[27] H. Shao, T. Svoboda, T. Tuytelaars, and T. Van Gool. H-PAT indexing for fast object/scene recognition based on local appearance. In *IVR*, 2003. 6