The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

Wide-Baseline Image Matching Based on Coplanar Line Intersections

Hyunwoo Kim and Sukhan Lee

Abstract— **This paper presents a novel method of wide-baseline image matching based on the intersection context of coplanar line pairs especially designed for dealing with poorly textured and/or non-planar structured scenes. The line matching in widely separated views is challenging because of large perspective distortion and the violation of the planarity assumption in local regions. To overcome the large perspective distortion, the local regions are normalized into the canonical frames by rectifying coplanar line pairs to be orthogonal. Also, the 3D interpretation of the intersection context of the coplanar line pairs helps to match the non-planar local regions by adjusting the region of interest of the canonical frame according to the different types of 3D non-planar structures. Compared to previous approaches, the proposed method offers efficient yet robust wide-baseline line matching performance under unreliable detection of end-points of line segments and poor line topologies or junction structures. Comparison studies and experimental results demonstrate the accuracy of the proposed method for various real world scenes.**

I. Introduction

Establishing feature correspondences between images under different illuminations, viewpoints, and backgrounds is an important and fundamental problem in computer vision and image processing, and recently it has been getting attention due to its tremendous needs in diverse applications such as image search, scene modeling, visual recognition, and augmented reality.

Many of the image matching methods to date have been proposed under the proposition that interest points can be detected and matched based on the invariant properties associated with photometrics. To deal with widely separated views, local affine/similarity invariant features have been actively developed and broadly employed in many vision applications during the last decades, including in maximally stable extremal regions (MSER) [1], edge based regions (EBR) [2], and scale-invariant feature transform (SIFT) [3]. The methods and performances of those kinds of local affine invariant features were reviewed and compared in [4]. Furthermore, Rothganger *et al.* [5] represented 3D objects in terms of local affine invariant features. However, those approaches are effective only for richly structured scenes with sufficient textural information that allows them to be used for the extraction and matching of interest points. Other than the perspective distortion, in widely separated views, the local regions of non-planar structures, such as 3D junctions/corners and 3D boundaries/edges, can not be approximated by local planar regions because they include multiple planes and are easily occluded by other planes [6]. Therefore, the local features need to be matched in non-planar structured scenes considering these 3D occlusions.

In real-world situations, it is often the case that scenes may contain poorly textured objects, obscuring images of interest [7]. In this case, line features can be good alternative image features to interest points because man-made objects are often configured with several well-defined geometric shapes that offer distinct 3D lines and edges [8], [9]. In applications such as scene modeling and 3D object recognition, the detection and matching of line features are required, regardless of those of interest points. While line features are regarded as robust to environmental variations for detection and localization in 2D image planes, they are difficult to match because of the lack of photometric invariance to be used for measuring similarity.

In the early years, the line matching approaches were developed based on the photometric properties by adapting the interest point-based approaches to line segments. Schmid and Zisserman [10] automatically matched line segments by exploiting the intensity neighborhood of the line segments guided by the epipolar constraints between different camera views in order to provide point to point correspondences along the line segments. Werner and Zisserman [11] improved the previous algorithm with resolving the resulting ambiguity by a search to register the photometric neighborhood. Bay *et al.* [12] obtained line segment correspondences by comparing the histograms of the neighboring color profiles in both views, and a topological filter was used for refinement matching. Those approaches presume that accurate camera geometry should be estimated or extra topological relation analysis should be performed prior to use.

To overcome the poor discriminating power of the photometric properties of line segments, the geometrical properties have been investigated as an alternative. In [12], a topological filter [13] was consecutively followed after histogram based matching in order to identify correct line matches while removing mismatches. Wang *et al.* [14] proposed a robust line matching method for affine distortion and 3D viewpoint changes, in which line segments are clustered

This work is supported in part by the Intelligent Robotics Development Program, one of the Frontier R&D Programs funded by the Ministry of Knowledge Economy (F0005000-2009-31), in part by the KORUS Tech Program administered by the KOTEF (Korea Industrial Technology Foundation) with the fund provided by the Ministry of Knowledge Economy, in part by the MEST (Ministry of Education, Science and Technology), Korea, under the WCU (World Class University) Program supervised by the KOSEF (Korea Science and Engineering Foundation) (R31-2008-000-10062-0), and in part by Basic Science Research Program through the National Research Foundation of Korea(NRF) by the MEST (2010-0004359).

Hyunwoo Kim is with the department of new media of Korean German Institute of Technology, Seoul, Korea. hwkim@kgit.ac.kr

Sukhan Lee is with the school of information and communication engineering of Sungkyunkwan University, Suwon, Korea. (Corresponding author, Email: lsh@ece.skku.ac.kr)

into local groups, or line signatures, according to spatial proximity. Those approaches have been proved successful for line matching for widely separated views; however, they are affected inaccuracies in the end points of detected line segments.

Another group of researchers has been utilized junction features for line matching. Vincent and Laganière [15] matched junctions by estimating the local perspective distortion between the neighborhoods of junctions, then estimated a fundamental matrix based on a constrained minimization assuming crude camera pose estimates. Bay *et al.* [16] identified polyhedral junctions resulting from the intersections of the line segments, then segmented the images into planar polygons using an algorithm based on a Binary Space Partitioning tree. Micusík *et al.* [17] detected and matched rectilinear structures based on vanishing point detection for widely separated views. However, junction-based and rectilinear-structure-based approaches can only be applied to well-structured scenes, in which lines and junctions are robustly extracted and/or vanishing points are stably estimated.

To address the challenge of line matching, we rely on the observation that many parts of man-made objects are constructed of local planar patches. According to this observation, the set of local planar patches can be considered as a good candidate of geometric primitives for the scene modeling of man-made objects. We adapt an image feature based on coplanar line intersections, called the "Line Intersection Context Feature (LICF)," which was recently introduced in [18]. The scenes are modeled by local planar regions using LICFs, but the scenes do not necessarily need to be piecewise planar.

In [18], the normalized cross correlation (NCC) is used as a descriptor for matching LICFs among different views. Although NCC is found to be very effective for narrow-baseline stereo matching, it is not adequate for matching LICFs in widely separated views because the local region patches of LICFs suffer large perspective distortions.

In this paper, we propose a novel robust region descriptor of the LICF feature, which is adequate for wide-baseline image matching by compensating for perspective distortions and handling non-planar 3D effects. The novelties of the proposed method are as follows. (1) First, we propose a robust region descriptor of LICFs in order to compensate for large perspective distortion between widely-separated views. (2) Second, the 3D interpretation of the intersection context in non-planar structure scenes facilitates the ability to robustly match non-planar structures such as 3D boundaries/edges and 3D junctions/corners and the patches on 3D planar patches. (3) Last, no a-priori knowledge such as camera geometry or 3D scene modeling is given for line matching. Moveover, when camera geometry is provided, the method can be simplified to be faster and more robust.

In Section II, the image feature LICF is reviewed and a robust region descriptor for wide-baseline stereos is proposed. In Section III, the multi-local feature matching technique is described. Section IV presents the comparison studies and experimental results in real world scenes including 3D line reconstruction. Finally, Section V concludes with a discussion and ideas for future works.

II. Feature Extraction and Description

A. Review on Line Intersection Context Feature (LICF)

In this section, the LICF is shortly reviewed (refer to [18] for details). Given line segments extracted in an image, intersecting lines are paired when both lines are closely located, i.e., the end point of line segments are located from their intersection within a certain distance by a proximity rule. The intersecting line pairs include both coplanar and non-coplanar in 3D. The discrimination is based on the fact that, when an intersection of an intersecting line pair exists in 3D, a match of the intersection can be found in a second view. Therefore, coplanar line pairs are determined by finding intersecting line pairs that has correspondence between different views in the local intersection areas.

The LICF contains geometric information, as well as photometric information. The former is the positional information of the intersection computed from a line pair, and the latter is the region information of the local image patch centered at the intersection position. Note that the LICF includes a region descriptor and a geometric primitive.

The LICFs are represented by the intersection positions and the corresponding region patches in an image, as follows:

$$\mathscr{F} \equiv \{\mathbf{x}_k, P(\mathbf{x}_k), \mathscr{L}_{pair,k}\} = \{\mathbf{x}_k, P(\mathbf{x}_k), \mathbf{l}_{\pi(k)_1}, \mathbf{l}_{\pi(k)_2}\}, \quad (1)$$

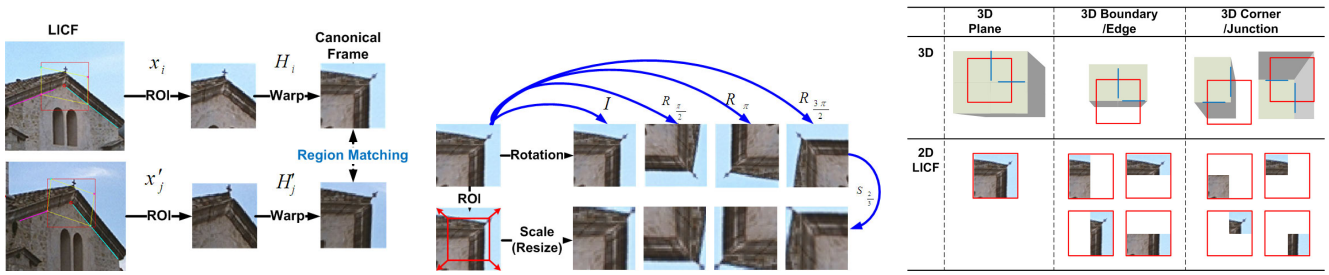
where $k = 1, \cdots, \#$ of Π. $\mathbf{x}_k$ denotes the intersection positions of the corresponding intersecting line pairs $\mathscr{L}_{pair,k} (= \{\mathbf{l}_{\pi(k)_1}, \mathbf{l}_{\pi(k)_2}\})$, and $P(\mathbf{x}_k)$ denotes the region patch centered at the intersection position $\mathbf{x}_k$. For convenience, an LICF ($\mathscr{F}_k$) sometimes refers to only the position ($\mathbf{x}_k$) instead of the set including the position and the neighboring region patch ($P(\mathbf{x}_k)$).

B. Robust Feature Description

Affine/similarity invariant region descriptors such as MSER, EBR, and SIFT can be considered good candidates for matching the intersection context in widely separated views. These descriptors transform a local region into an affine covariant region, through which the affine distortion of the local region is compensated. However, those region descriptors cannot be directly applied to LICF matching because the intersection context does not contain sufficient photometric information to normalize the local region using the analysis of its covariance matrix compared to those of other interest-point-based local regions.

To cope with large perspective distortion in local region matching, a novel affine/projective invariant region descriptor for LICFs needs to be developed. Instead of the covariance matrix analysis of local regions, the coplanar line pair of LICF is directly used for the normalization of the local region.

As shown in Figure 1(a), the angle between a coplanar line pair significantly changes with viewpoint. To match LICFs between different views, the perspective distortion



(a) Canonical frames (b) Rotation and scale variations of a canonical frame (c) 2D LICFs w.r.t. 3D non-planar structures

Fig. 1. Robust region descriptors of LICFs and the 3D interpretation.

of the local region of the intersection context need to be compensated. This is accomplished by rectifying a coplanar line pair into a special configuration in which the coplanar line pair is orthogonal. We call the normalized image region the "Canonical Frame" by adapting the term used in the context of the affine covariant regions [1]. The rectification process is achieved by estimating 2D homography H_k from a region patch $P(\mathbf{x}_k)$ to a canonical frame C_k, and the transformation is represented by

$$C_k \equiv C(\mathbf{x}_k) = P(H_k \mathbf{x}_k). \tag{2}$$

The homographies can be computed using the four intersecting points of the coplanar line pair and the bounding box of the local region to their corresponding points in the canonical frame. The size of the bounding box is equal to that of the canonical frame, which is related to the region matching process and can be designed by a user by considering the trade-off between the speed and accuracy and the degree of plane locality of a given scene. Note that although the true variation is anisotropic in all image directions, the isotropic scale variation can be considered a good approximation in practice.

1) Scale variation and rotational ambiguity: Next, the region descriptor encodes the scale variation by adjusting the region of interest (ROI) in the canonical frame. Without explicit smoothing in scale space, scale variation can be achieved by adjusting the size of the bounding box in the intersection context. The ROI in the intersection context is determined by the predefined scale steps, and the ROI is resized into the size of the canonical frame using bilinear interpolation. Given the scale s, the scale matrix is computed by $S_s = \begin{pmatrix} s & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 1 \end{pmatrix}$. The transform resizes the canonical frame and a bilinear interpolation is followed to normalize the size into that of the canonical frame. The scale-variant canonical frame with a scale factor s is represented by

$$S_k^s \equiv P(S_s H_k \mathbf{x}_k). \tag{3}$$

In addition, the canonical frame has rotational ambiguity in four classes, so their corresponding rotational matrices are applied into the canonical frames. The 2D rotational ambiguities are represented by R_r, where $r = \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$. The extended feature set of an LICF $\mathbf{x}$ with a rotation class r and scale variations s is represented by

$$RS_k^{s,r} \equiv P_k(R_r S_s H_k \mathbf{x}_k), \tag{4}$$

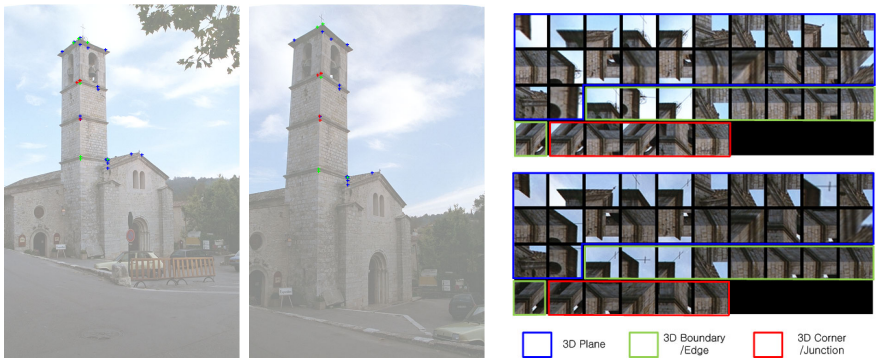
where $r = \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$ and $s = \{1, \frac{2}{3}\}$. Additional scale change can be implemented by adding more scale factors into the cost of the computational complexity. The procedure is depicted in Figure 1(b).

2) Extension to 3D Non-planar Structures: Many local region descriptors assume that the scene is locally planar; however, the assumption is violated for non-planar structures such as 3D junctions/corners and 3D boundaries/edges [6], [19]. The more widely the scenes are separated, the more the local planarity assumption is violated in those non-planar structures. To deal with those 3D non-planar structures, the 3D interpretations of LICFs are investigated.

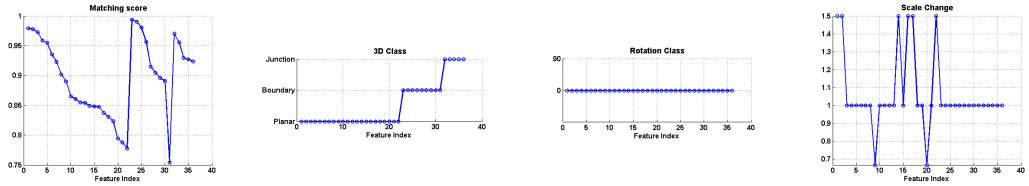
The 3D non-planar structures of local image regions can be classified into three categories: 3D planar patches, 3D boundaries/edges, and 3D junctions/corners, as illustrated in Figure 1(c). Other 3D structures can be approximately categorized into the aforementioned three classes. First, the 3D planar patch meets the local planarity, so the projective invariant region descriptor of the intersection context is sufficient. Second, for the 3D boundaries, one member of the line pair is located on the boundary between an object and the background and the other is located inside the object. In this case, only the object side of the intersection context need to be used for matching. Last, for the 3D junctions, the coplanar line pair envelopes the junction or the corner of the object, so the image region corresponding to the object, i.e., the quarter of the canonical frame, is used for matching.

According to those 3D interpretations of the LICFs, the canonical frame, i.e., the scale-variant canonical frame $RS_k^{s,r}$, is extended into the nine local features: one type for the canonical frame, four types for the 3D boundaries/edges, and anther four types for the 3D junctions/corners, as presented in Figure 1(c). Nine local features are constructed by applying window operations with different ROIs within the canonical frame. The windows are represented by W_{11}, $\{W_{21}, W_{22}, W_{23}, W_{24}\}$, and $\{W_{41}, W_{42}, W_{43}, W_{44}\}$ for the planar case, the 3D boundaries, and the 3D junctions, respectively. At last, we have the final robust region descriptor represented by

$$T_k^{s,r,w} \equiv W_w RS_k^{s,r,w} \equiv W_w P_k(R_r S_s H_k \mathbf{x}_k), \tag{5}$$



(a) 3D classified feature positions of non-planar structures (b) 3D classified canonical frames



(c) Matching score (NCC) (d) 3D class estimation (e) Rotation (class) estimation (f) Scale estimation

Fig. 2. LICF examples and the analysis of 3D non-planar structures.

where $r = \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$, $s = \{1, \frac{2}{3}\}$, and $w = \{11; 21, 22, 23, 24; 41, 42, 43, 44\}$. Note that the descriptor is extracted by image warping using the transform $R_r S_s H_k$ and the following windowing operation W_w only selects the ROI of the transformed image.

III. Feature Matching

A. Multi-local Feature Matching

The LICFs in wide-baseline settings are the set of local features containing multiple rotation classes, multiple scale steps, and multiple 3D types, which have four classes, two (or more) steps, and nine ROIs, respectively. Matching all of the local features between different views not only requires too much computational complexity, but also it does not guarantee correct matches. For effective and efficient LICF matching, a multi-local feature matching scheme based on NCC is proposed.

For either the multiple scale steps or the multiple rotation classes, the class with the highest matching score is selected as a matching candidate. First, the scale changes monotonically, so we can pick the optimal scale value by choosing the scale class with the highest NCC score. Given multiple scale steps (1 and $2/3$), the scale steps between different views cover more scale differences ($\{2/3, 1, 3/2\}$). In the current implementation, the scale step covers the scale range from 0.5 to 2.0 when the matching power of NCC is considered [20]. Also, for selection from the multiple rotation classes, clearly there is one and only solution to resolve the rotation ambiguity. The selection of the class with the highest NCC score makes sense.

Conversely, for the multiple 3D types (3D planar patches, 3D boundaries, and 3D junctions), an LICF tends to match in more than one of the 3D types due to the fact that 3D type features share the same canonical frame with different ROIs, as designed, and the 3D effects gradually appear according to the degree of viewpoint change in the local area. Therefore, each type of 3D non-planar structure needs to be considered as a different local feature and should be matched independently, then combined into one type of local feature in the final step. Within each type, the matching feature with the highest NCC is considered as a matching candidate, and multiply matched positions are combined by choosing the 3D type with the largest ROI regions (i.e., 3D planar patch, 3D boundary, and 3D junction, in that order). To reduce computational complexity, the feature type of 3D planar patch is first matched, then the matching LICFs are removed. Then, the feature type of 3D boundary is matched using the remaining features, and the same process is performed for the feature type of the 3D junction. For multi-local matching of LICFs using NCC, the same process used with interest points is adopted [21]. Also, after computing NCC scores from both images, the most strongly correlated matches in both images are selected.

Finally, a RANSAC-based refinement [18] is used to find correct LICF matches while simultaneously estimating the fundamental matrix from the matching candidates using the multi-local feature matching. The matching LICFs between two views, I and I', are represented by

$$\mathscr{M}_{\mathbf{x},\mathbf{x}'} = \{\mathbf{x}_k, \mathbf{x}'_k; \mathscr{L}_{pair,k}, \mathscr{L}'_{pair,k}\}, \tag{6}$$

where $k = 1, \ldots, K$, $\{\mathbf{x}_k, \mathscr{L}_{pair,k}\} \in \mathscr{F}$, $\{\mathbf{x}'_k, \mathscr{L}'_{pair,k}\} \in \mathscr{F}'$, and K denotes the number of matching LICFs. The fitting error, given a fundamental matrix F, is defined by the symmetric transfer error:

$$\mathscr{E}_{\mathbf{x},\mathbf{x}'} = \frac{1}{K} \sum_{i=1}^{K} d(\mathbf{x}'_i, F\mathbf{x}_i)^2 + d(\mathbf{x}_i, F^T\mathbf{x}'_i)^2 \tag{7}$$

where $d(\mathbf{x},\mathbf{y}) = (\mathbf{y}^T\mathbf{x})/\sqrt{\mathbf{y}_1^2+\mathbf{y}_2^2}$, the distance of point $\mathbf{x}$ from line $\mathbf{y}$.

Figure 2 demonstrates matching LICFs and the matching details. In Figure 2(a), matching LICFs are presented from a widely-separated image pair. They are drawn in different colors according to the 3D type, and the local planar patches are shown in detail in Figure 2(b). The matching score, the estimated 3D type, the rotation class, and the estimated scale are plotted in Figures 2(c), 2(d), 2(e), and 2(f), respectively. The estimated classifications are reasonable. In the scene, the rotation class is one because there is no serious 2D rotation, and the scale changes are mostly due to perspective distortion in the local areas. The 3D type classification looks reasonable, but it does not explicitly separate the 3D type classes. One feature can be assigned to several classes depending on the degree of 3D effects in the local region patch. For example, a 3D junction far away from the camera can be approximated using a local planar patch instead of 3D junction type.

1) Line Segment Matching: Given a matching LICF $\mathscr{M}_{\mathbf{x},\mathbf{x}'}$, matching between individual line segments $\{\mathbf{l}_{\pi(k)_1},\mathbf{l}_{\pi(k)_2}\}$ and $\{\mathbf{l}'_{\pi'(k)_1}$ and $\mathbf{l}'_{\pi'(k)_2}\}$ must be resolved from the matching pairs $\{\mathscr{L}_{pair,k},\mathscr{L}'_{pair,k}\}$. Using the estimated epipolar geometry from the matching stage, the angles between the line segments of the coplanar line pair and the epipolar line are compared in order to pair them into the line segment with a smaller angle difference. After the individual matching of the coplanar line pair, the multiple matches are discarded. The final matching lines are represented by

$$\mathscr{M}_{\mathbf{l},\mathbf{l}'} = \{\mathbf{l}_{\psi(k)},\mathbf{l}'_{\psi'(k)}\}, \psi(k) \subset \Pi, \psi'(k) \subset \Pi', \qquad (8)$$

where $k = 1,\cdots,\#$ of Ψ and Ψ'. Ψ and Ψ' denote the index set of matching line segments from the first and second views, respectively, and Ψ and Ψ' have the same number of elements after one-to-one matching.

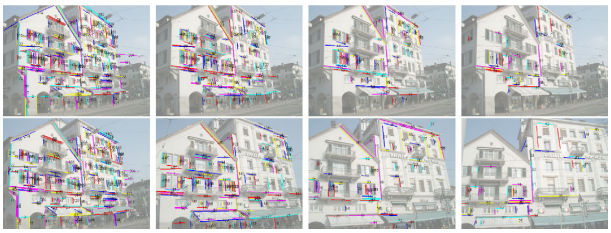
IV. Experimental Results

To evaluate the performance of the proposed method, it is compared with state-of-the-art technologies: the SIFT method [3], [22] for general purpose matching and the line-signature-based matching (LS) method of [14] for wide-baseline line matching. In the figures, the SIFT method and the LS method are referred to as "SIFT" and "LS," respectively. The proposed method provides two separate matching outputs: matching LICFs and matching lines. The matching LICFs are not only intermediate matches for the final individual line matching, but also they provide interest-point-like matches, so that we can compare the matches with matching SIFT features. Then, the matching lines, the final output of the proposed method, are measured and evaluated. The result of matching LICFs and that of matching lines are referred to as "LICF" and "LICF+Line," respectively.

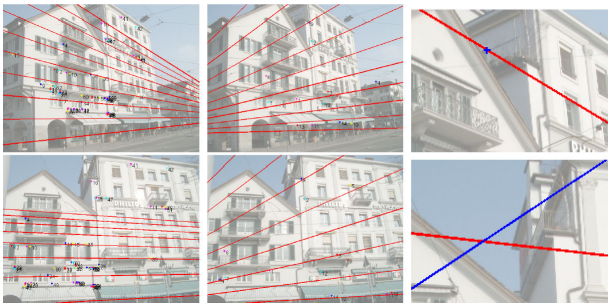
For quantitative analysis, the number of matching features after RANSAC refinement are used, while the conventional matching evaluation methods count correct matches manually. The difference is not that large, i.e., no more than 10% in our experiments, thanks to the RANSAC-based refinement. In the plot of the numbers of matching features, the absolute value explains how many features can be correctly matched. In addition, the relative value with respect to the reference view, or the graph slope, describes the degradation in the quality of matching in comparison with the perfect match because the matching results between the reference view and itself can be considered as a perfect matching case. In addition, the symmetric transfer error is also used to evaluate the matching accuracies of LICF and SIFT features [23], [18]. Exceptionally large values of the symmetric transfer error mean that the method failed at feature matching. The initial minimum consensus sample set is remained as the solution, so the sample set is considered as a inlier and its error has extremely large value.

A. Widely Separated Views

For comparison purposes, image pairs from richly textured outdoor scenes, in which both line features and interest point features can be detected and matched in the same scene, are collected through the public data sets [24], [25], [4].



(a) Matching lines between the reference view and Views 2-5



(b) Detailed comparison study between View 1 & 5

Fig. 5. Matching results of the sequence "Zubud." In Figure 5(b), matching LICFs and matching SIFT features are displayed in order, overlaying the epipolar constraints, in the first two columns. In the right most column, the blue point on the top is transferred into the epipolar lines on the bottom. The epipolar lines from the SIFT and the proposed method are shown in blue and red, respectively.

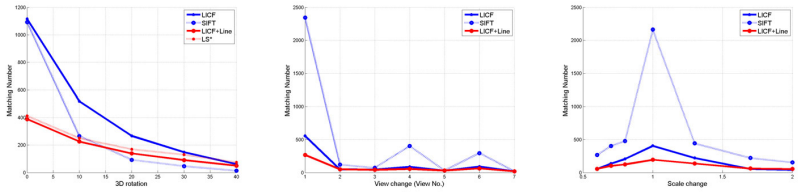
1) 3D rotation: Figure 3(a) presents an image sequence, "Zubud," with different 3D rotational variations of the image plane approximately ranging from 10° to 40°. The experimental result demonstrates that the matching number decreases and the symmetric transfer error increases when the 3D rotation angle increases. Quantitatively, the matching number and the symmetric transfer error are compared in the first columns of Figures 4(a) and 4(b), respectively. Qualitatively, the matching lines are shown in Figure 5(a).

(a) The sequence "Zubud". Views 1-5, in order. The reference view is View 1.

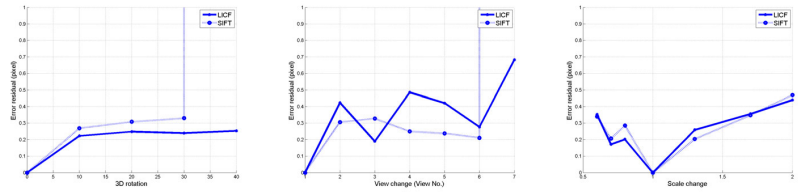
(b) The sequence "Kampa". Views 1-7, in order. The reference view is View 1.

(c) The sequence "INRIA". Views 1-7, in order. The reference view is View 3.

Fig. 3. Test image sequences for widely separated views. The reference views are boxed in red. The sequence "Zubud" is taken from scene #157 in the Zubud data set.



(a) The number of matching features



(b) The symmetric transfer errors of the estimated fundamental matrix

Fig. 4. Quantitative results. (Left) The sequence "ZuBud." (Middle) The sequence "Kampa." (Right) The sequence "INRIA."

The comparison results show that the proposed method is comparable to the LS method and superior to the SIFT with respect to 3D rotational variation in terms of matching number and symmetric transfer error. (Note that, although the direct comparison between the LS method by [14] is not clear because their line matching result is not shown in their paper and the line detection module also affects the results, the comparison may be informative and reasonable.) In addition, the estimated epipolar geometries after refinement are compared between the SIFT method and the proposed method. While the proposed method give the correct estimation of the fundamental matrix, the SIFT method fails to match features, resulting in a large symmetric transfer error. The point transfer of the position selected from the building wall shows the difference more clearly.

2) Perspective distortion: Figure 3(b) is an image sequence "Kampa," with perspective distortions among different views. The experimental results show that the proposed method works well for those kinds of large perspective distortions, and the results are comparable to those of the SIFT method. The matching number and the symmetric transfer error are compared in the middle columns of Figures 4(a) and 4(b), respectively.

When the numbers of matching features (LICFs and SIFTs) and the symmetric transfer errors are compared, the symmetric transfer error between the reference view and View 3 shows that the SIFT method is more accurate than the proposed method, but the result between the reference view and View 7 show that the proposed method works well while the SIFT method fails in terms of the number of matching features and the symmetric transfer error.

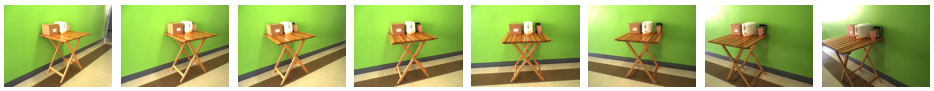
3) Scale change: Figure 3(c) is an image sequence "INRIA" with scale and 2D rotational variations. The scale variation ranges from approximately 0.6 to 2.0 times. The matching number and the symmetric transfer error are compared in the last columns of Figures 4(a) and 4(b), respectively.

B. Poorly Textured Indoor Scenes

Furthermore, the proposed method is applied to poorly textured scenes with widely separated views. These scenes are more challenging than the scenes in Section IV-A because

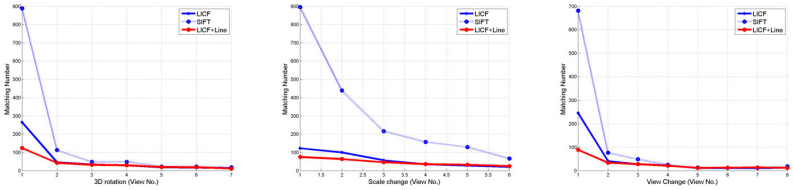
(a) The sequence "Dinning_Table." Views 1-7, in order. The reference view is View 1.

(b) The sequence "Kitchen." Views 1-6, in order. The reference view is View 1.

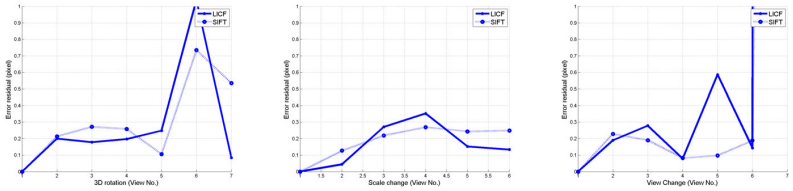


(c) The sequence "Tea_Table." Views 1-8, in order. The reference view is View 1.

Fig. 6. Test image sequences of poorly texture scenes.



(a) The number of matching features



(b) The symmetric transfer error of the estimated fundamental matrix

Fig. 7. Quantitative results. (Left) The sequence "Dinning_Table." (Middle) The sequence "Kitchen." (Right) The sequence "Tea_Table."

they not only lack textural information, but also the line features are hard to detect when the normal direction of surface embedding of the lines is close to that of the epipolar plane.

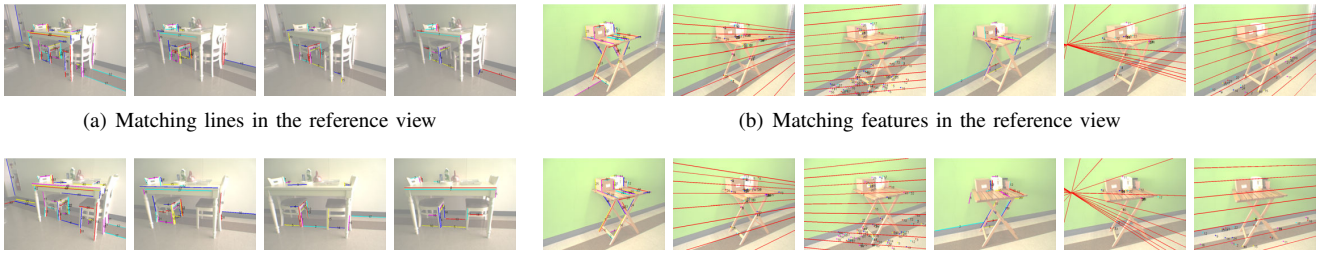
Figure 6 shows test image sequences taken from poorly textured indoor scenes. The scenes "Dinning_Table" and "Tea_Table," shown in Figures 6(a) and 6(c), respectively, are captured by rotating cameras around tables and chairs, so they include large perspective distortions due to 3D pose differences between the camera and the objects. The scene "Kitchen," shown in Figure 6(b), is taken by a translating/approaching camera, so it includes large scale changes in the sequence.

1) 3D Rotation: The experimental result for the scene "Dinning_Table," shown in the left column of Figure 7, is similar to that of the textured outdoor scene "Zubud," but the matching number decreases faster than that of the textured scene when the rotation angle increases. This is because the discrimination powers of the features in poorly textured scenes are weaker than those in textured scenes. Moreover, when the long edge of the dinning table is aligned with the horizontal line of the image space, the LICFs belonging to the dinning table are aligned in same epipolar line. Due to this kind of matching ambiguity in the scene "Dining_Table," the matching results in Views 5-7 are very unstable in terms of matching number and symmetric transfer error, as shown in the left column of Figure 7.

However, although the matching results of line segments are not perfectly correct, they are good enough to explain the scene structure of the chairs, floor lines, and dinning tables, as shown in Figures 8(a) and 8(c).

2) 3D Translation: For the scene "Kitchen," the experimental results show that both the SIFT method and the proposed method work well because this kind of translating camera motion results in only scale changes, and the scene is approximately modeled by several piecewise planar patches. The plots of the number of matching features and the symmetric transfer error, shown in the middle column of Figure 7(b), demonstrate the stabilities and accuracies of both methods.

3) Repetitiveness and Symmetry: The scene "Tea_Table" is very challenging because the tea table and the boxes on the table have repetitive and/or symmetric structures and patterns. The number of matching features is very small compared to those of other previous sequences, and the symmetric transfer error is very unstable, as shown in the



(a) Matching lines in the reference view

(b) Matching features in the reference view

(c) Matching lines in Views 3, 4, 5, and 6, in order

(d) Matching features in Views 2 & 4, in order

Fig. 8. Matching results between the reference view and some sample views. (Left) The scene "Dinning_Table." (Right) The scene "Tea_Table." In Figures 8(b) and 8(d), (The first three rows) Matching lines, matching LICFs, and matching SIFT features, in order, between the reference view and View 2. (The last three rows) Between the reference view and View 4. The figures are best viewed in color and with PDF magnification.

right column of Figure 7.

For instance, Figures 8(b) and 8(d) shows that both the SIFT method and the proposed method result in failures between the reference view and View 7. The camera geometry between the reference view and View 2 corresponds to a narrow baseline, and the matching result is quite correct. However, since the reference view and View 4 are widely separated and poorly textured with regard to repetitiveness and symmetry in structure, the matching results are very unstable, showing a case in which the proposed method does not provide the best solution.

V. Conclusion and Future Works

In this paper, a wide-baseline line matching algorithm was introduced to overcome large perspective distortion and non-planar local structure using a non-planar robust feature descriptor based on the intersection context of coplanar line pairs and its 3D interpretation. The experimental results showed that the performance is comparable and complimentary to those of state-of-the-art matching methods, such as the SIFT method and the LS method.

Future works will include the handling of matching ambiguity due to the symmetry of structure, the repeatability of texture, and the improvement of matching accuracy by combining both the SIFT method and the proposed method.

References

[1] J. Matas, O. Chum, U. Martin, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *Proceedings of British Machine Vision Conference*, vol. 1, London, 2002, pp. 384–393.

[2] T. Tuytelaars and L. Van Gool, "Matching widely separated views based on affine invariant regions," *Int. J. Comput. Vision*, vol. 59, no. 1, pp. 61–85, 2004.

[3] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[4] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *Int. J. Comput. Vision*, vol. 65, no. 1-2, pp. 43–72, 2005.

[5] F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce, "3d object modeling and recognition using affine-invariant patches and multi-view spatial constraints," in *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2003, pp. II: 272–277.

[6] A. Vedaldi and S. Soatto, "Features for recognition: Viewpoint invariance for non-planar scenes," in *Proceedings of the International Conference on Computer Vision (ICCV)*, vol. 2, 2005, pp. 1474–1481.

[7] E. Kim, G. Medioni, and S. Lee, "Planar patch based 3d environment modeling with stereo camera," in *RO-MAN07*, Jeju island, Korea, August 26-29 2008, pp. 516–521.

[8] C. Baillard, C. Schmid, A. Zisserman, A. Fitzgibbon, and O. O. England, "Automatic line matching and 3d reconstruction of buildings from multiple views," in *In ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery, IAPRS Vol.32, Part 3-2W5*, 1999, pp. 69–80.

[9] B. Micusík and J. Kosecka, "Piecewise planar city 3d modeling from street view panoramic sequences," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Location: Miami, FL, USA, 20-25 June 2009, pp. 2906 – 2912.

[10] C. Schmid and A. Zisserman, "Automatic line matching across views," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, vol. 0, p. 666, 1997.

[11] T. Werner and A. Zisserman, "New techniques for automated architectural reconstruction from photographs," in *ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part II*. London, UK: Springer-Verlag, 2002, pp. 541–555.

[12] H. Bay, V. Ferrari, and L. Van Gool, "Wide-baseline stereo matching with line segments," in *CVPR*, 2005, pp. I: 329–336.

[13] V. Ferrari, T. Tuytelaars, and L. V. Gool, "Wide-baseline muliple-view correspondences," in *CVPR*, June 2003, pp. I: 718–728.

[14] L. Wang, U. Neumann, and S. You, "Wide-baseline image matching using line signatures," in *Proceedings of the International Conference on Computer Vision (ICCV)*, 2009.

[15] E. Vincent and R. Laganière, "Junction matching and fundamental matrix recovery in widely separated views," in *Proc. British Machine Vision Conf.*, London, UK, September 2004, pp. 77–86.

[16] H. Bay, A. Ess, A. Neubeck, and L. V. Gool, "3d from line segments in two poorly-textured, uncalibrated images," in *Proceedings of the Third International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)*, June 2006.

[17] B. Micusík, H. Wildenauer, and J. Kosecka, "Detection and matching of rectilinear structures," in *CVPR*, 2008.

[18] H. Kim and S. Lee, "A novel line matching method based on intersection context," in *IEEE International Conference on Robotics and Automation (ICRA)*, Anchorage, AK, May 3-8 2010.

[19] P. Moreels and P. Perona, "Evaluation of features detectors and descriptors based on 3d objects," *Int. J. Comput. Vision*, vol. 73, no. 3, pp. 263–284, 2007.

[20] F. Zhao, Q. Huang, and W. Gao, "Image matching by multiscale oriented corner correlation," in *ACCV (1)*, 2006, pp. 928–937.

[21] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006.

[22] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," http://www.vlfeat.org/, 2008.

[23] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.

[24] T. Werner, "Lmatch: Matlab toolbox for matching line segments accross multiple calibrated images," http://cmp.felk.cvut.cz/ werner/software/lmatch/, 2007.

[25] H. Shao, T. Svoboda, and L. V. Gool, "Zubud-zurich buildings database for image based recognition," Swiss Federal Institute of Technology, Tech. Rep. Technical report No. 260, 2003.