Multiple Homography Estimation via Stereo Line Matching for Textureless Indoor Scenes

Sihui WANG
School of Information and Communication Engineering
Beijing University of Posts and Telecommunications
Beijing, China
scott_won@sina.cn

Xuesong ZHANG*
School of Computer Science
Beijing University of Posts and Telecommunications
Beijing, China
xuesong_zhang@bupt.edu.cn

***Abstract*—For robot navigation applications, indoor scenes usually contain texture-less planar surfaces, which renders traditional point-based matching approaches deficient. As an emerging alternative, line-based methods have attracted growing interest. In our paper, we develop a unified approach for line detection and stereo matching that can produce appealing results in un-textured scenes. In addition, we propose a co-planar line classification approach, based on which we are able to enforce the epipolar line constraints to multiple homography compatibility validation. Experimental results demonstrate the efficiency of our method as well as its improved capability on multi-homography estimation over traditional methods.**

Keywords-line detection; stereo matching; coplanarity; multiple homography; epipolar line constraint

I. INTRODUCTION

In the vision-based simultaneous localization and mapping (SLAM) community, low-level image features, such as the local points, are widely used. Despite that point features are well-studied, easy to track and convenient to handle [1], it has been shown that the point-based methods such as SIFT rely on properties like textures or local structures [2] and will not produce satisfactory results in indoor environment where textured-objects become less frequent [3]. Compared with the point-based methods, line-based methods are more robust and can provide much more structural information. In addition, line features are abundant in the man-made environment [4]. As a result, line-based methods have emerged as a desirable alternative to point-based methods.

However, several difficulties are posed in the field of line-based feature detection and matching, namely the fragmentation of a single line, the inaccuracy of the locations of the endpoints, the less distinctive appearance of line segments, the lack of geometric constraint such as the epipolar constraint in point matching [5], etc. To address these problems, several approaches have been proposed. Guoxuan Zhang, Jin Han Lee, Jongwoo Lim and Hong Suh's papers [1-6] managed to ameliorate the issues of fragmentation and less distinctive appearance of line segments, while Lilian Zhang and Reinhard Koch's papers [7] introduced a line matching algorithm that can produce fast matching results in the indoor scenes. They both contribute to the improvement on the line detection and matching performance using different techniques. In Section 2, we will review the related works and integrate the state-of-the-art algorithmic blocks into one framework that can produce fast and optimal line matching results for texture-less indoor scenes.

Reconstructing the 3D scene geometry is an important problem in computer vision that can be applied in robot auto-navigation and augmented reality. Traditional structure from motion (SFM) is a point-base method that is rendered obsolete in texture-less indoor scenes [8]. As we switch from points to lines, line-based plane extraction algorithm should be adopted to cluster lines into coplanar bundles. In Section 3, we will illustrate our method for 3D reconstruction and coplanar classification.

The problem of planar region extraction and the problem of homography estimation is somehow intertwined. According to Hyunwoo Kim and Sukhan Lee [9], each corresponding line pairs between two views can generate a plane hypothesis and an underlying planar homography. By evaluating each line's compatibility with the planar homography, we can cluster the candidate lines into coplanar bundles. Conversely, in this paper we will demonstrate that by exploiting the coplanar relations and the epipolar line constraint, mutually compatible homographies that reflect the true scene structure can be obtained. Zygmunt L. Szpak, Wojciech Chojnacki and Anton van den Hengel [10] have stated that from a practical perspective, failing to enforce consistency constraints on multiple homographies will lead to inconsistent estimates of epipolar geometry between two views. In our paper, we observed that from a practical perspective, enforcing the epipolar line constraint will lead to an improvement in multiple homography estimation. A discussion on this topic will be presented in Section 4.

II. LINE DETECTION AND STEREO MATCHING

A. Algorithmic Building Blocks

Line detection algorithms that are based on the calculation of grayscale gradient share a common pipeline. Despite the

* Partially supported by National Natural Science Foundation of China (61871055,61701036), and partially by the Fundamental Research Funds for the Central Universities (2018RC54).

similar framework, the performance of line detection algorithms varies greatly, depending on the implementation details. In the indoor texture-less scenes, low-contrast edges might be of significant structural implications and the mis-detection of them might lead to a flawed reconstruction of the scene geometry. Meanwhile, the fragmented line segments as the input will cause the line matching algorithm to be computationally expensive and inexact. In a word, for the purpose of line detection in the indoor texture-less scene, the algorithm should be sensitive enough to detect the less distinctive segments and the algorithm should also be noise-resistant so that the extracted edges won't be highly fragmented.

Jin Han Lee, Sehyung Lee, Guoxuan Zhang, Jongwoo Lim and Il Hong Suh [6] proposed a Canny-based line extractor and integrated it with the MSLD [11] descriptor for line detection and matching. In our experiment, it is shown that this line detection algorithm is capable of detecting weak textures as well as generating non-fragmented line segments.

However, line matching based on MSLD descriptor is not optimal in the indoor texture-less scenes. Lilian Zhang and Reinhard Koch [7] proposed the LBD descriptor for line matching. According to them, LBD descriptor is more efficient to compute and it is faster to generate the matching results than the state-of-the-art methods.

Yet there are a couple of drawbacks of the originally proposed LBD-based approach. First, in the originally proposed LBD-based approach, LBD descriptor is incorporated with EDLine detector [12] and the latter is proved to be inefficient to detect low-contrast edges in our experiments. Second, in the originally proposed LBD-based approach, efforts are made to overcome the scale changes and the global rotation, which won't produce any improvement in our cases given that we apply the algorithm to calibrated, stereo cameras.

B. *Our Approach: A Hybrid Algorithm*

Based on the preceding analysis, we proposed our approach that is built on three strategies.

The first strategy is to use Canny edge detector and to merge the edges into segments by geometric coherence. Canny edge detector is sensitive enough to extract the weak textures, while the merging algorithm can filter out the noise-like edges and retain the linear features with a minimum length. This treatment can obtain the low-contrast line segments that might have significant structural implications in the indoor texture-less scenes. Meanwhile, due to the screening out of the noise-like edges and the merging of linear features into line segments, the number of resulting candidate lines won't be large, which is vital for the performance of LBD-based line matching approach.

The second strategy is to compute the LBD descriptors for the extracted line segments. In our cases, the calculation in multiple scale space is unnecessary. Hence, we can remove the down-sampling part of the original approach and perform the computation of the descriptors only on the original image, which helps to reduce the total computational time.

The third strategy is to adopt the graph matching method using spectral technique. In the originally proposed approach, a global rotational angle is estimated first to overcome the rotation problem. Then, LBD descriptors in all scale spaces are checked to calculate the descriptor distance between two lines. In our implementation, the process of global rotational angle estimation is removed. The process of descriptor distance calculation is simplified in that all we need to do is to calculate the descriptor distance only in the original images.

As the result, our algorithm shows equal line matching performance with the original LBD-based approach and the total computational time is reduced. If we evaluate the process of line detection and line matching as a whole, our algorithm will out-perform both Canny edge detector + MSLD descriptor algorithm and the EDLine + LBD descriptor algorithm in our application scenarios.

In Figure 1, the performance of LBD and our algorithm are compared. It is shown that our algorithm can detect low-contrast segments and produce superior results with less fragmentation.

III. Coplanar Line Classification for Efficient Homography Proposal

Planar homographies and Manhattan hypothesis can be exploited to obtain the coplanar line bundles [8][9]. In our paper, we demonstrate that we can reconstruct the 3D-line first. Then, the coplanar line classification can be obtained by the application of geometric constraints if the reconstruction error is not too large. In our experiments, line detection and stereo matching typically consume about 0.60s and all other processes (3D reconstruction, co-planarity classification and multiple homography estimation) typically cost about 0.13~0.20s. Figure 2 shows the result produced by our co-planarity classification algorithm. It is shown that our algorithm has low computational complexity and can obtain

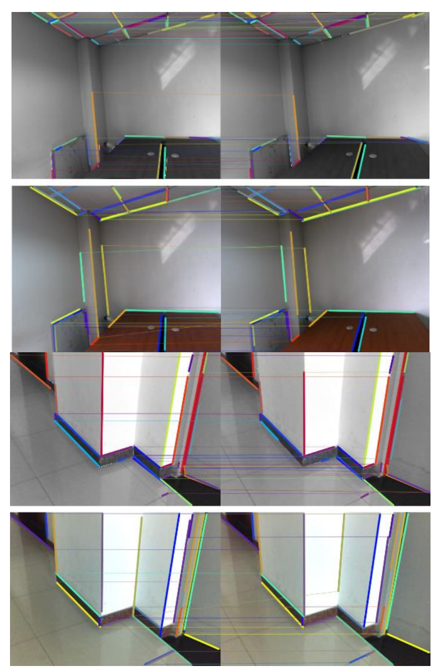

Figure 1. Line detection and stereo matching performance of LBD (top) and our method (bottom)

satisfactory results provided that the reconstruction error is not too large.

In 3D geometry, line pairs are divided into three categories: the intersecting lines, the parallel lines and the skew lines. Skew lines have nothing to do with co-planarity, while the co-planarity of intersecting lines and parallel lines should be dealt with carefully so that our classification conform with the common sense of the real world.

There are three steps in our approach.

First step: d**etermining line pairs relationship**. In this step, distances and angles between the lines are calculated and the practical thresholds are set to determine which line pair is parallel, intersecting or skew. In addition, we maintain a matrix named CHECK whose elements are set to 'false' initially. We use this matrix to mark which line pair is processed while which line pair is not processed.

S**econd step: processing intersecting line pairs**.

We loop through the line pairs. As skew line pairs have nothing to do with the coplanar calculation, we set the corresponding elements in the matrix CHECK to true, indicating that these line pairs have been processed.

We will leave the parallel line pairs to the next step, so we just overlook these line pairs in this stage.

Any intersecting line pair is considered as a coplanar hypothesis. We call the line pair that first generates the coplanar hypothesis the seed line pair. Starting from the seed line pair, we expand the hypothetical plane by merging lines that satisfy the coplanar constraint into the hypothetical plane. If the hypothetical plane can't expand any more, we call the set of lines we obtain the *closure* of the coplanar hypothesis. Any element in the matrix CHECK that corresponds to a line pair in the closure should be set to true to indicate that the line pair has been processed.

Sometimes the seed line pair will generate a sub-optimal closure. For example, the seed line pair l_1, l_2 generates a closure of the coplanar hypothesis $\{l_1, l_2, l_3\}$ while the seed line pair l_2, l_3 generates a closure of the coplanar hypothesis $\{l_2, l_3, l_4, l_5\}$. In this situation, we should merge the closures. Generally speaking, if A, B are two coplanar hypotheses and l_1, l_2 are two non-collinear lines satisfying the condition $l_1, l_2 \in A \cap B$, then we should substitute the coplanar hypothesis A ∪ B for A and B.

After obtaining the optimal coplanar hypothesis, we should filter out the *pseudo-planes*, that is, the coplanar line pairs that don't actually form a solid plane. From geometric intuition, it is obvious that if AB, CD are two intersecting 3D-lines, EF is a 3D-line that crosses AB or CD, and EF is not coplanar to ABCD or ABDC, then we say that AB and CD form a pseudo-plane and remove the coplanar hypotheses that containing AB and CD. Hence, we obtain the coplanar hypotheses generated by the intersecting line pairs.

Third step: processing the remaining parallel line pairs.

A number of parallel line pairs are marked 'true' in the matrix CHECK to indicate that these parallel line pairs have been processed. Parallel line pairs of this kind are mostly feature lines fit into the same coplanar hypothesis. Our aim in this stage is to identify the parallel structural line pairs and generate new coplanar hypotheses. The operations here are similar to the operations performed in step two. First, loop through the matrix CHECK again to find out the unprocessed parallel line pairs. Second, verify that the line pairs don't form pseudo-planes. Third, generate the coplanar hypothesis and calculate the maximum closure. Hence, we obtain the coplanar hypotheses generated by the parallel line pairs.

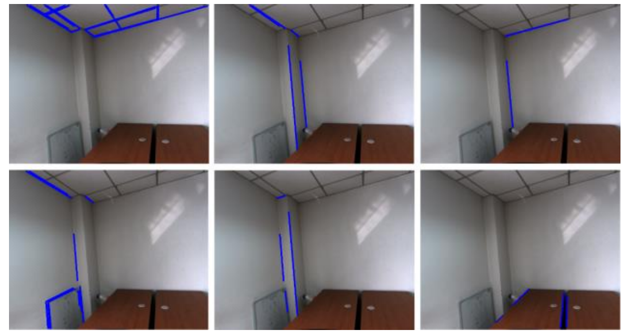

Figure 2. Coplanar lines extracted by our co-planarity classification algorithm

Put the coplanar hypotheses from step two and the coplanar hypotheses from step three together, we obtain the coplanar classification of the lines.

IV. Epipolar Line Constrained Multiple Homography Estimation

A. *Epipolar Line Constraint*

According to [10], the planar homography for stereo cameras can be represented by:

$$H_i = I_3 + \frac{1}{d_i} K t n_i^T K^{-1} \tag{1}$$

where n_i is the unitary outward normal vector of plane i, and d_i is the distance from plane i to the origin of the world coordinate system. The intrinsic parameter matrix is $K_1 = K_2 = K$, and the motion of cameras is $R_1 = R_2 = I_3$, $t_1 = 0$ and $t_2 = t$.

Therefore, we obtain the epipolar line constraints between the two views:

$$\begin{cases} x_2 = \left(1 + \frac{t}{d_i} n_1\right) x_1 + \frac{t}{d_i} n_2 y_1 + \frac{f t n_3 - x_0 t n_1 - y_0 t n_2}{d_i} & (2) \\ y_2 = y_1 & (3) \end{cases}$$

where f is the focal length of the camera, $f_x = f_y = f$ and $K = \begin{pmatrix} f & 0 & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{pmatrix}$.

B. *Consistency Constraints of Multiple Homographies*

According to [10], for any two given homographies H_i, H_j, the discriminant of the cubic equation $p_{H_i H_j}(\lambda) = 0$ should be zero:

$$\Delta\left(p_{H_i H_j}\right) = 0, \forall i \neq j \tag{4}$$

These constraints are called the multiplicity constraints.

In addition, we have:

$$\det\left[\left(H_j^{-1} H_i\right)^n, \left(H_l^{-1} H_k\right)^m\right] = 0, \quad \forall i \neq j, \forall k \neq l, \forall m, n \in N^+ \tag{5}$$

where [*,*] is the Lie bracket operator and [S,T]=ST-TS.

These constraints are called the singularity constraints.

In fact, it is easy to mathematically verify that:

If two monocular cameras share the same intrinsic parameter matrix $K_1 = K_2 = K$ and the motion of the cameras is $t_1 = 0$, $t_2 = (t, 0, 0)^T$, $R_1 = I_3$, $R_2 = \begin{pmatrix} cos\theta_z & sin\theta_z & 0 \\ -sin\theta_z & cos\theta_z & 0 \\ 0 & 0 & 1 \end{pmatrix}$, then the singularity constraints will be automatically satisfied by enforcing the epipolar line constraints.

C. Multiple Homography Estimation

Zygmunt L. Szpak, Wojciech Chojnacki and Anton van den Hengel [13] [14] developed an algorithm for multiple homography estimation using the latent variable parametrization. Their algorithm is point-based and requires the iteration process of the Levenberg-Marquardt method. In our paper, we demonstrate that the structural information of lines and planes can be exploited to simplify the estimation of multiple homographies.

Suppose that $L = \{l_i | i = 1,2,3, \dots, n\}$ is a set of coplanar lines in the left view and $R = \{r_i | i = 1,2,3, \dots, n\}$ is a set of coplanar lines in the right view. Suppose that (l_i, r_i) is the corresponding line pair. In principle, 4 pairs of non-collinear corresponding points can define a homography matrix. However, in practice we use the two endpoints and the middle point of each $l_i \in L$ and the corresponding points (which are obtained by epipolar constraints) in the right view as the corresponding point pairs to calculate the homography matrix. Our intention behind this approach is to avoid the degenerative cases and to obtain a relatively good estimation of homography matrix.

After determining the corresponding points, it is general practice to use RANSAC to calculate the homography matrix. However, we argue here that the outliers which have large re-projective errors are mostly excluded in the process of line matching and coplanar estimation, and the epipolar line constraint alone can guarantee the correctness of the homography estimation and further simplify the calculation.

Suppose that $x(x_1, y_1)$ is a point in the left view and $x'(x_2, y_2)$ is the corresponding point in the right view. As we obtain x' by epipolar line constraint, it is assured that $y_1 = y_2$. Following the epipolar line constraints, it can be deduced that:

$$x_2 = h_{11}x_1 + h_{12}y_1 + h_{13}, \tag{6}$$

$$H = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{7}$$

We represent the points in the left view by $P_1(x_{11}, y_1)$,…, $P_n(x_{n1}, y_n)$ and we represent the corresponding points in the right view by $Q_1(x_{12}, y_1)$,…, $Q_n(x_{n2}, y_n)$. Then it is easy to verify the following equations based on the epipolar line constraint:

$$Ah=b \tag{8}$$

where

$$A = \begin{pmatrix} x_{11} & y_1 & 1 \\ x_{21} & y_2 & 1 \\ \dots & \dots & \dots \\ x_{n1} & y_n & 1 \end{pmatrix}, h = \begin{pmatrix} h_{11} \\ h_{12} \\ h_{13} \end{pmatrix}, b = \begin{pmatrix} x_{12} \\ x_{22} \\ \dots \\ x_{n2} \end{pmatrix}$$

Hence, the problem of multiple homography estimation in question reduces to the problem of the least square minimization of linear equations. In this paper, we adopt the SVD method to calculate the homography matrix. Suppose that $A = U\begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^T$ by SVD and $U = (U_n, \bar{U})$ in which U_n is the first n columns of U, then we have $h = (h_{11}, h_{12}, h_{13})^T = V\Sigma^{-1}U_n^T b$. Hence, the homography matrix can be recovered by (7).

Compared with RANSAC, the homography matrix generated by our method is more consistent according to the consistency constraints of multiple homographies. The discussion on this topic will be presented in the Section 5.

V. Experimental Results

A. Experiments on Synthetic Data

We tested the performance of our method and the RANSAC method on several planar scenes with noise. Planar scenes were generated first, then location deviations of corresponding points were added as the noise. The location deviation d is a random variable that is uniformly distributed in $[0, 20M]$ where M is the noise level of the scene. The result shows that as the noise level increases, the incompatibility of the multiple homographies estimated by RANSAC method also increases. Nevertheless, the multiple homographies estimated by our method remains compatible with each other. The results of our experiment are shown in Figure 3.

B. Experiments on Real Data

We use four pairs of stereo pictures to evaluate the performance of our method and the RANSAC method on real data. As Figure 4 shows, the coplanar lines are labeled manually beforehand.

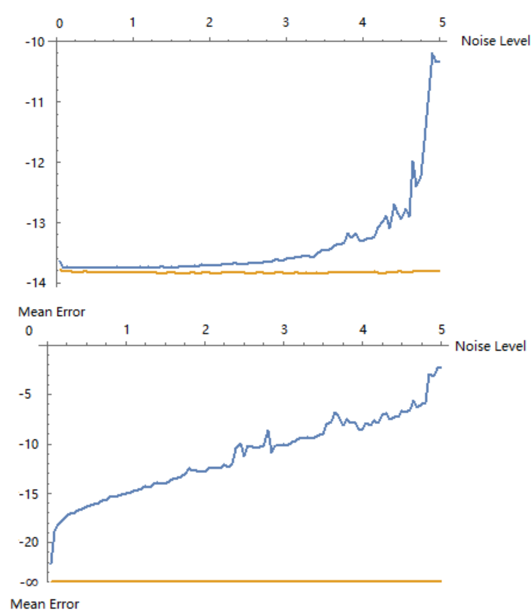

Figure 3. Logarithm of mean error of multiplicity constraints (top) and logarithm of mean error of singularity constraints (bottom) on synthetic data. The blue curves are obtained by the RANSAC method, while the orange curves are obtained by the epipolar line constraints.

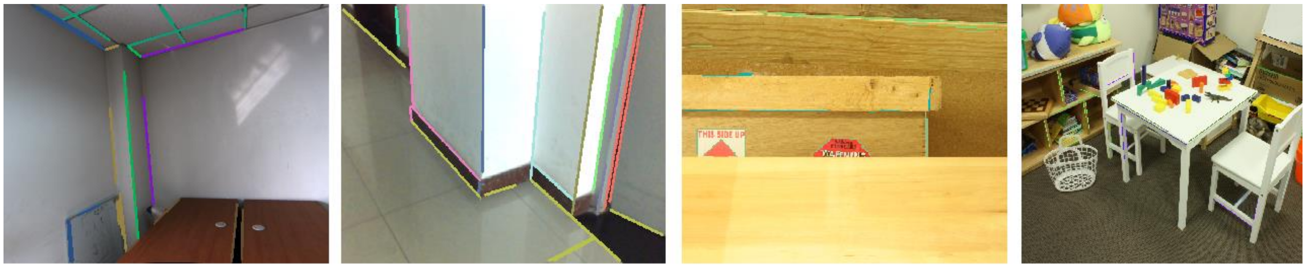

Figure 4. Images used to evaluate the compatibility of multiple homographies. Coplanar lines are labeled before-hand.

The mean multiplicity constraint value is defined by $\frac{1}{N_1}\sum_{i\neq j}\left|\Delta\left(p_{H_iH_j}\right)\right|$ where N_1 is the total number of pairs of H_i, H_j satisfying $i \neq j$. The mean singularity constraint value is defined by $\frac{1}{N_2}\sum_{i\neq j,k\neq l}\left|\det\left[\left(H_j^{-1}H_i\right)^n, \left(H_l^{-1}H_k\right)^m\right]\right|$ where N_2 is the total number of tetrads $\{H_i, H_j, H_k, H_l\}$ satisfying $i \neq j$ and $k \neq l$. The mean value of both constraints are calculated and the results are shown in table 1. In practical perspective, it seems that our method produces slightly better result of multiple homography estimation if compatibility is measured by the multiplicity constraints. If compatibility is measured by the singularity constraints, then the multiple homography estimation produced by the RANSAC method will get significantly worse when large re-projective error occurs. Meanwhile, the multiple homographies that are produced by our method will always satisfy (5).

VI. CONCLUSION

In the field of line detection, the combination of Canny edge detector and the segments merging algorithm based on geometric coherence provides a cure for the issue of fragmentation and the issue of poor recognition of low-contrast segments. Combining Canny-based line detector and LBD-based line matching algorithm together, we are able to produce optimal stereo matching results in the indoor texture-less scenes.

The advantage of co-planarity classification algorithm based on geometric constraints is its low computational complexity. The process of stereo matching and co-planarity classification help to build high-level image features which simplify the calculation of planar homography. By enforcing the epipolar line constraints, the multiplicity constraints are reduced and the singularity constraints are totally eliminated, indicating that the compatibility of multiple homographies is improved after enforcing the epipolar line constraints.

TABLE I. COMPATIBILITY EVALUATION ON REAL DATA

Data	**Multiplicity constraints**		**Singularity constraints**	
	RANSAC	***Epipolar***	***RANSAC***	***Epipolar***
Image1	1.59×10^{-9}	1.49×10^{-14}	1.10×10^{-4}	0
Image2	2.29×10^{-6}	1.25×10^{-14}	5.97	0
Image3	12744	5.27×10^{-16}	4.34×10^{14}	0
Image4	13.61	1.98×10^{-14}	N.A.	0

REFERENCES

[1] Guoxuan Zhang, Jin Han Lee, Jongwoo Lim and Il Hong Suh, "Building a 3-D line-based map using stereo SLAM", IEEE Transactions on Robotics, 31(6):1-14, January 2016.

[2] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors", PAMI 27(2005), pp. 1615-1630.

[3] L. Wang, U. Neumann and S. You, "Wide-baseline image matching using line signatures", ICCV, 2009, pp. 1311-1318.

[4] Juan López, Roi Santos, Xosé R. Fdez-Vidal and Xosé M. Pardo, "Two-view line matching algorithm based on context and appearance in low-textured images", Pattern Recognition, Volume 48, Issue 7, July 2015, pp. 2164-2184.

[5] Bin Fan, Fuchao Wu and Zhanyi Hu, "Robust line matching through line-point invariants", Pattern Recognition, Volume 45, Issue 2, February 2012, pp. 794-805.

[6] Jin Han Lee, Sehyung Lee, Guoxuan Zhang, Jongwoo Lim, Wan Kyun Chumg and Il Hong Suh, "Outdoor place recognition in urban environments using straight lines", Proceedings – IEEE International Conference on Robotics and Automation, 2014. 5550-5557.

[7] Lilian Zhang and Reinhard Koch, "An efficient and robust line segment matching approach based on LBD descriptor and pairwise geometric consistency", Journal of Communication and Image Representation, Volume 24, Issue 7, October 2013, pp. 794-805.

[8] Chelhwon Kim and Roberto Manduchi, "Planar structures from line correspondences in a Manhattan world", Asian Conference on Computer Vision, 2014, pp. 509-524.

[9] Hyunwoo Kim and Sukhan Lee, "Multiple planar region extraction based on the coplanar line pairs", Proceedings – IEEE International Conference on Robotics and Automation, pp. 2059-2064.

[10] Zygmunt Szpak, Wojciech Chojnacki and Anton van den Hengel, "Robust multiple homography estimation: an ill-solved problem", CVPR 2015, pp. 2132-2141.

[11] Zhiheng Wang, Fuchao Wu and Zhanyi Hu, "MSLD: A robust descriptor for line matching", Pattern Recognition, Volume 42, Issue 5, May 2009, pp. 941-953.

[12] Cuneyt Akinlar and Cihan Topal, "Edlines: a real-time line segment detector with a false detection control", Pattern Recognition Letters, Volume 32, Issue 13, 1 October 2011, pp. 1633-1642.

[13] Wojciech Chojnacki, Zygmunt Szpak, Michael Brooks and Anton van den Hengel, "Multiple homography estimation with full consistency constraints", Proceedings – 2010 Digital Image Computing: Techniques and Applications, DICTA 2010.

[14] Zygmunt Szpak, Wojciech Chojnacki, Anders Eriksson and Anton van den Hengel, "Sampson distance based joint estimation of multiple homographies with uncalibrated cameras", Computer Vision and Image Understanding, Volume 125, August 2014, pp. 200-213.

AUTHORS' BACKGROUND

1. This form helps us to understand your paper better, which will not be published.

First Author's Name: Sihui WANG	
Position (Prof., Assoc. Prof., Asst. prof., Dr., Mr., Ms.)	**Mr.**
Research Field	Computer Vision
Personal Webpage	
Second Author's Name: Xuesong ZHANG	
Position (Prof., Assoc. Prof., Asst. prof., Dr., Mr., Ms.)	Assoc. Prof.
Research Field	Computer Vision, Computational Photography
Personal Webpage	
Third Author's Name:	
Position (Prof., Assoc. Prof., Asst. prof., Dr., Mr., Ms.)	
Research Field	
Personal Webpage	