3D LINE SEGMENT RECONSTRUCTION ON PIECEWISE PLANAR SCENES

Kai Li *Jian Yao*[†]

School of Remote Sensing and Information Engineering, Wuhan University, P.R. China

[†] jian.yao@whu.edu.cn

ABSTRACT

This paper presents a new 3D Line Segment (LS) reconstruction method for piecewise planar scenes. Given a set of LS matches identified from two images, the algorithm first clusters them into groups where LS matches in each group are related by a same planar homography induced by a same space plane, which is subsequently estimated from the putatively reconstructed 3D LSs obtained by triangulating each pair of LS correspondences in the group. The final reconstructed 3D LSs are obtained by projecting the image LSs into the estimated space plane. To reduce the incidence of the false grouping of LS matches, we transfer the problem of LS match grouping into a multi-label optimization problem and solve it by graph cuts. The advantages of our algorithm over others are that it can generate satisfactory results with much fewer images and can recover the space planes where the reconstructed 3D LSs lie, which is beneficial for upper level applications.

Index Terms— 3D Line Segment Reconstruction, Piecewise Planar Scene Reconstruction, Graph Cuts.

1. INTRODUCTION

Feature point based 3D reconstruction methods [1, 2, 3, 4] might fail in man-made environments with a low amount of distinctive interest points, e.g. urban and indoor scenes. In these environments, exploiting LSs extracted from images for 3D reconstruction can be an alternative way because LSs are often abundant in man-made scenes. Another benefit for exploiting LSs for 3D reconstruction is that we can obtain the wireframe of a scene with very few images and generate a recognizable model for the scene. This is because man-made objects (e.g. buildings) can often be outlined by a set of LSs. For example, to reconstruct the scene in Fig 1(d), our proposed 3D LS reconstruction method can generate the 3D model shown in Fig 1(e) using only two stereo images. It is easy to recognize the scene from this 3D model, but hardly possible to achieve this from the extremely sparse point clouds obtained by point based 3D reconstruction methods.

Despite of the above benefits for exploiting LSs for 3D reconstruction, it is often hard to reliably reconstruct 3D LSs because of the unstableness and low locating accuracy of the extracted LSs. For example, to reconstruct the LSs in the scene shown in Fig 1(a), which can approximately be regarded to lie on a single space plane, when we used traditional way by triangulating the LS correspondences identified from two stereo images, we obtained the 3D LSs shown in Fig 1(b). As we can see, many of the 3D LSs are reconstructed mistakenly. To solve this problem, some methods [5, 6, 7] resort to exploit multiple (three or more) images photographing a same scene to decrease the ambiguities and get robust reconstruction results, which involves establishing LS correspondences between multiple images or some sophisticated hypothesizing-and-testing procedures.

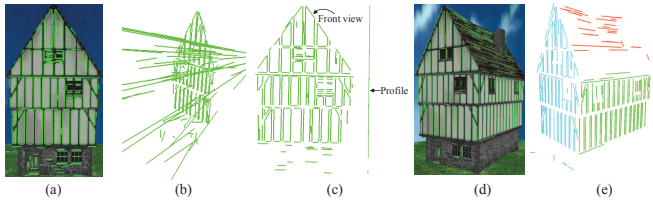


Fig. 1. An example of a problem in 3D line segment reconstruction and the results obtained by our proposed solution.

In this paper, we present a new solution to this problem, which requires only two images. We first cluster LS matches identified from two images into groups that LS matches in each group are related by a same planar homogrpahy inducing by a single space plane, which is then estimated from the putatively reconstructed 3D LSs obtained by triangulating LS correspondences in this group. Once the space plane for each LS match group has been recovered, the final 3D LSs can be obtained by projecting the LSs in one image to the space plane. To reliably cluster LS match, we transfer the problem of LS match clustering into a multi-label optimization problem and solve it by graph cuts [8]. Fig 1(c) shows our refined 3D LS reconstruction results for this scene. As we can see, the problem existing in Fig 1(b) is solved and the space plane for the scene is correctly recovered. For scenes comprised of multiple planes, as shown in Fig 1(d), our method successfully reconstructs the 3D LSs, shown in Fig 1(e), using only two stereo images and recovers the three main planes in the scene, with the correct annotations of the reconstructed 3D LSs w.r.t. the space planes they belong to. The contributions of this paper hence include the two aspects. First, we propose a new solution for solving the ambiguities in 3D LS reconstruction by LS match clustering and space plane estimation. Second, we propose to solve the problem of LS match grouping by solving a multi-label optimization problem using graph cuts.

2. RELATED WORKS

We divide the existing LS reconstruction methods into two categories according to whether a method requires LS matching before the reconstruction procedure. Many methods in this category focus on the exploitation of different mathematical representation for a 3D line to establish the projective relationship between a 2D line and its 3D correspondence, which is the foundation for 3D LS reconstruction and camera calibration based on lines. The projective relationship between a 2D and 3D lines is not as explicitly as that the for points [9]. A series of representations for lines in 3D space have been proposed. They are plücker coordinates [10, 11, 12], pair of points [13, 14, 15, 16, 9], pair of planes[9], a unitary direction vector and a point on the line [17], the intersections with two orthogonal planes [18], and a more recent one, Cayley representation [19]. Under these representations, the authors proposed different methods

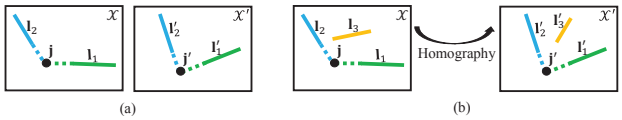


Fig. 2. An illustration of the principle ideas of the LS matching method presented in [25].

for reconstructing 3D lines and/or estimating the camera parameters. Some methods have been proposed to specifically reconstruct 3D LSs on certain scenes, like scenes meeting Manhattan world assumption [20, 21], piecewise planar scenes [22] and poorly textured scenes [23]. The prior of the scenes helps decrease the uncertainties for reconstructing 3D LSs.

Some recent algorithms in this area attempt to free the reconstruction procedure from the heavy dependence on LS matching procedure because it is hard to get reliable LS correspondences in some kind of scenes, such as poorly-textured indoor environments and scenes containing wiry structures (e.g. power pylons [6]). The majority of these methods adopt the strategy to first generate a set of 3D hypotheses for each extracted LSs, either through sampling the depth of the endpoints of 3D LSs to camera centers [7], or triangulating some putative LS correspondences after enforcing some soft constraints on the extract LSs [6, 5], and then validate the hypotheses by projecting them back to images. In [24], a novel algorithm is proposed to obtain 3D LSs with an unknown global scale from a single image capturing a Manhattan world scene. It is possible to do so because LSs on this special scene can only distribute in three orthogonal directions. This fact tremendously deceases the degrees of freedom for all reconstructed LSs.

Our method belongs to the first category and the reconstruction procedure is based on our LS matching method presented in [25]. We focus only on obtaining 3D LSs, as well as recovering the main planes of piecewise planar scenes. The camera parameters are obtained by some external camera calibration methods, or by our LS matching method [25]. The advantages of our algorithm over others are that it can generate satisfactory results with much fewer images and can recover the space planes where the reconstructed 3D LSs lie, which is beneficial for upper level applications, such as scene understanding [26, 27], building facade segmentation [22, 28, 29, 30].

3. ALGORITHM

3.1. Line Segment Matching

The proposed LS reconstruction method is based on our LS matching method presented in [25], which provides some crucial input for the LS reconstruction, we briefly summarize it here to help understand this paper.

Refer to Figure 2(a), suggest I_1 and $\mathbf{l}_2$ are two LSs from the first image $\mathcal{X}$, and $\mathbf{l}'_1$ and $\mathbf{l}'_2$ from the second image $\mathcal{X}'$. The goal of LS matching is to establish the corresponding relationship between $\mathbf{l}_1$ and $\mathbf{l}'_1$, and $\mathbf{l}_2$ and $\mathbf{l}'_2$. For this aim, the proposed method intersects $\mathbf{l}_1$ and $\mathbf{l}_2$, and $\mathbf{l}'_1$ and $\mathbf{l}'_2$, generating junction points $\mathbf{j}$ and $\mathbf{j}'$, respectively. At the places of the two formed junctions, the method constructs two structures called Line-Junction-Line (LJL), consisting a junction point and the two LSs forming it. To match the two LJLs, ($\mathbf{l}_1$, $\mathbf{j}$, $\mathbf{l}_2$) and ($\mathbf{l}'_1$, $\mathbf{j}'$, $\mathbf{l}'_2$), the method describes them with a robust descriptor, followed by some effective matching strategies. Once the two LJLs are successfully matches, two LS matches, ($\mathbf{l}_1$, $\mathbf{l}'_1$) and ($\mathbf{l}_2$, $\mathbf{l}'_2$), and one point match ($\mathbf{j}$, $\mathbf{j}'$) can be obtained. The point matches obtained from all LJL matches found in the two images can be used to estimate the fundamental matrix $\mathbf{F}$ between them. The above strategy can only match LSs which lie near with other LSs because only spatially adjacent LSs are intersected and used to constructed LJLs. For those LSs which lie far away from others and are not used to construct LJLs, they are matched by the local planar homography $\mathbf{H}$ computed from a neighboring LJL match and the estimated $\mathbf{F}$. Please refer to [25] to know how to achieve this. For example, in Figure 2(b), to match $\mathbf{l}_3$ and $\mathbf{l}'_3$, the proposed method finds their adjacent LJL match, ($\mathbf{l}_1$, $\mathbf{j}$, $\mathbf{l}_2$) and ($\mathbf{l}'_1$, $\mathbf{j}'$, $\mathbf{l}'_2$), and estimate a local planar homography using the LJL match and $\mathbf{F}$. Since the local homography can establish point-to-point correspondence, it can be used to map the endpoints of $\mathbf{l}_3$ to $\mathcal{X}'$, generating an estimated correspondence for $\mathbf{l}_3$, $\mathbf{l}^h_3$. If ($\mathbf{l}_3$, $\mathbf{l}'_3$) is a correct match, $\mathbf{l}^h_3$ must be very near to $\mathbf{l}'_3$. In this way, whether ($\mathbf{l}_3$, $\mathbf{l}'_3$) is a correct match can be determined.

LS matches and the local homographies are used as input for the subsequent 3D LS reconstruction procedure. The camera parameters for each image can also be recovered from the point matches obtained along with LJL matches, if they are not available.

3.2. Two-view Based 3D Line Segment Reconstruction

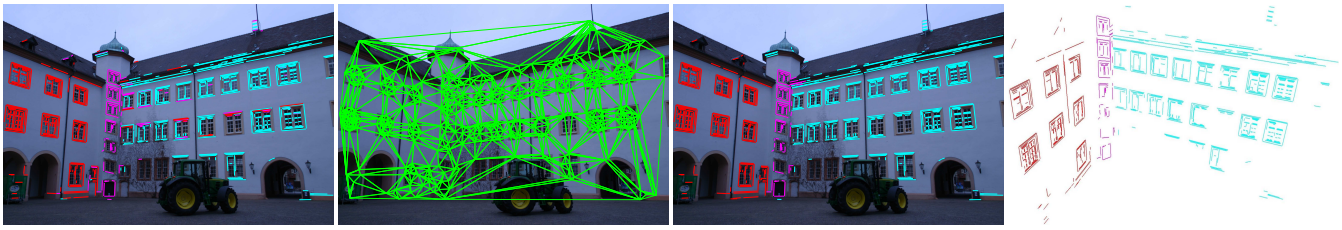
For two images, $\mathcal{X}$ and $\mathcal{X}'$, after matching the extracted LSs [15] from them using the above-presented method, we obtain a LS match set, $\mathcal{M} = \{(\mathbf{l}_v, \mathbf{l}'_v)\}_{v=1}^{V}$, a homography matrix set, $\mathcal{H} = \{\mathbf{H}_i\}_{i=1}^{I}$, and the projection matrices for the two images, $\mathbf{P}$ and $\mathbf{P}'$. The 3D LS set corresponds to LS matches in $\mathcal{M}$ is $\mathcal{L} = \{\mathbf{L}_v\}_{v=1}^{V}$, which is obtained by triangulating the two LS correspondences in each match.

Line Segment Match Grouping The projections of space LSs from one space plane into two images would be related by a single homogrpahy. Based on this fact, we cluster the obtained LS matches in $\mathcal{M}$ using the homography matrices in $\mathcal{H}$. For a LS match $(\mathbf{l}_v, \mathbf{l}'_v) \in \mathcal{M}$, we find the homography $\mathbf{H}_i \in \mathcal{H}$ which minimizes the distance a pair of LSs according to a homography:

$$d = \mathbf{l}'^{\top}\mathbf{H}\mathbf{e}_1 + \mathbf{l}'^{\top}\mathbf{H}\mathbf{e}_2 + \mathbf{l}^{\top}\mathbf{H}^{-1}\mathbf{e}'_1 + \mathbf{l}^{\top}\mathbf{H}^{-1}\mathbf{e}'_2 \qquad (1)$$

where $\mathbf{e}_{i=1,2}$ and $\mathbf{e}'_{j=1,2}$ denote the two endpoints of $\mathbf{l}$ and $\mathbf{l}'$, respectively. After this, some homography matrices in $\mathcal{H}$ are assigned with some LS matches from $\mathcal{M}$, forming a LS match group set, $\mathcal{S} = \{\mathcal{G}\}_{n=1}^{N}$, where $\mathcal{G}_n$ denotes the n-th LS match group whose elements are from $\mathcal{M}$. Since LS matches in each group are clustered by a same homography matrix induced by one space plane, their corresponding 3D LSs in $\mathcal{L}$ are supposed to lie on the space plane. We can estimate the space plane from the 3D LSs corresponding to matches in the group. But before that, we merge some the groups in $\mathcal{S}$ to ensure line segment matches induced by 3D LSs coming from a same space plane are clustered into only one group. For two LS match groups, $\mathcal{G}_i$ and $\mathcal{G}_j$, suppose they are formed based on the homographies, $\mathbf{H}_i$ and $\mathbf{H}_j$, respectively. If LS matches in $\mathcal{G}_i$ are consistent with $\mathbf{H}_j$, and vice versa, we merge the two groups into one. Here, a group of LS matches are *consistent* with a homography means the average of their distances according to the homography (defined in Eq. 1) is a small value (2 pixels in this paper). After this, we obtain a new LS group set, $\mathcal{S}$.

Space Plane Estimation For a LS match set $\mathcal{G}_i \in \mathcal{S}$, suppose the 3D LSs corresponding to all LS matches in $\mathcal{G}_i$ form a set $\mathcal{L}_s$, which is subset of $\mathcal{L}$. All 3D LSs in $\mathcal{L}_s$ are supposed to lie on the single space plane. We estimate it from the endpoints of the 3D LSs in $\mathcal{L}_s$ using RANSAC. Once the space plane has been estimated, we can recompute the homography relating the LS matches in $\mathcal{G}_i$ and accept it only when the majority of LS matches $\mathcal{G}_i$ are consistent with it. After processing all elements in $\mathcal{S}$, we obtain a space plane set $\mathcal{P} = \{\mathbf{P}_m\}_{m=1}^{M}$, and the updated LS match set $\mathcal{S} = \{\mathcal{G}\}_{m=1}^{M}$, whose elements correspond to elements of $\mathcal{P}$.



(a) (b) (c) (d)

Fig. 3. Some intermediate results generated by the proposed method based on two images.

Graph Cuts Based Line Segment Match Regrouping We found that it often brings in mistakes when we group LS matches only based on their distances according to estimated homographies that a LS match which is supposed to be assigned into one group but is clustered to another group mistakenly. These kinds of mistakes occur when there are several similar space planes in the scene and the estimated homographies are not so accurate, especially when the extracted LSs are not so precisely located. For instance, Fig 3(a) shows an example of the LS match grouping result based on only the distances of LS matches according to the estimated homographies. We draw in different colors the matched LSs in one of the two images to differentiate the groups to which they belong. The LSs drawn in the same color should appear on the same scene plane if they are correctly grouped. But, as we can see, a considerable number of matched LSs are mistakenly clustered. The spatial smoothness assumption that coplanar space LSs should be spatially adjacent with each other needs to be exploited for more reliably LS match grouping.

We propose to solve this problem by solving a multi-label optimization problem through minimizing the following energy function:

$$E = \sum_{p} D_p(l_p) + \sum_{p,q} V_{p,q}(l_p, l_q), \tag{2}$$

where the data term D_p measures the cost of p being assigned with the label l_p, and the smoothness term, $V_{p,q}$ encourages a piecewise smooth labeling (i.e. regularizes the solution) by assigning a cost whenever neighboring objects p and q being assigned with labels l_p and l_q, respectively. As to our problem, the data term D_p is the cost of a LS match $p = (\mathbf{l}_p, \mathbf{l}'_p)$ being labeled to belong to a certain group l_p, in which the homography relating the matches is assumed to be $\mathbf{H}_{l_p}$. Then, D_p can be calculated from Eq. 1. The smooth term $V_{p,q}$ measures the cost of two neighboring LS matches p and q are labeled to belong to groups l_p and l_q, respectively. We simply set it as a constant equaling 4 in pixels for all the labels. In this way, all terms in the objective function are defined. We resort to graph cuts to minimize the object function. We need to construct an adjacent graph to enable graph cuts to be exploited to solve our problem. Inspired by [31, 32], which constructed Delauney triangles for feature points, we constructed Delauney triangles using the midpoints of matched LSs in one image to define the adjacent relationship between LS matches, as shown in Fig 3(b). The weights for each edge in the adjacent graph is assigned by the Gaussian function according to the distance between the two vertices to encourage vertices with smaller distances being assigned with a same label in a higher possibility. The regrouping result corresponding the solution of the minimum of Eq.2 are shown in Fig 3(c). Comparing Fig 3(a) and (c), we can observe that almost all mistakes are corrected.

Once the falsely grouped LS matches are regrouped into correct groups, the final 3D LS corresponding to each match can be obtained simply by projecting one of the two LSs to its corresponding space plane. To remove the falsely reconstructed 3D LSs brought by few falsely grouped matches still existing after enforcing the smooth constraint. We intersect adjacent 3D planes and remove the 3D LSs that are beyond the intersection. After this, we obtain the final 3D LSs shown in Fig 3(d). As we can see, the three main planes in this scene are correctly recovered and all 3D LSs are well reconstructed and correctly annotated w.r.t. the space planes on which they lie.

3.3. Multi-view Based 3D Line Segment Reconstruction

If more than two images are available, it is easy to extend the above two-view based 3D LS reconstruction method to deal with multiple views. We just need combine the results obtained from each pair of images. In details, we first reconstruct a set of space planes, $\mathcal{P}_1$, and the corresponding set of 3D LSs, $\mathcal{L}_1$ lying on these planes using the first two images. The two sets are used to initialize the overall space plane set, $\mathcal{P}_o$, and the overall 3D LS set, $\mathcal{L}_o$, for the whole scene. The subsequent images are used to refine the two initial overall sets. For each added image, it will be used to reconstruct 3D LSs with its previous one, generating a new space plane set, $\mathcal{P}_i$ and a new 3D LS set, $\mathcal{L}_i$. For a space plane $\mathbf{P}_i^j \in \mathcal{P}_i$ and its corresponding 3D LS set is $\mathcal{L}_i^j$, if it is consistent with a space plane $\mathbf{P}_o^m \in \mathcal{P}_o$, whose corresponding 3D LS set is $\mathcal{L}_o^m$, we merge $\mathbf{P}_i^j$ and $\mathbf{P}_o^m$ into a new space plane estimating from all 3D LSs in $\mathcal{L}_i^j$ and $\mathcal{L}_o^m$. 3D LSs in $\mathcal{L}_i^j$ and $\mathcal{L}_o^m$ are then projected into the new space plane. Otherwise, we regard $\mathbf{P}_i^j$ as a new plane and update $\mathcal{P}_o$ and $\mathcal{L}_o$ as: $\mathcal{P}_o = \mathcal{P}_o \cup \{\mathbf{P}_i^j\}$ and $\mathcal{L}_o = \mathcal{L}_o \cup \mathcal{L}_i^j$. After processing all images, there would exist a considerable amount of duplications in $\mathcal{L}_o$ because a same 3D LS can be visible in more than two views and be reconstructed in multiple times. We remove the duplications by keeping only one 3D LS for a group of ones which are spatially adjacent and have very similar directions.

4. EXPERIMENTS

We report in this section the experimental results of our method on a synthesis image dataset and a real image dataset [1].

Synthesis images The synthesis image dataset [7] has 80×3=240 images photographing around a CAD model from the upper, middle and bottom viewpoints. Each round consists of 80 images separated by a constant angle interval, 4.5°. An example image in the dataset is shown in Fig 4(a). We employed for experiments only the 80 images for the middle round because we found in our initial experiments that the reconstruction result generated by our method based on the 80 images is negligibly different from that based on all 240 images, but the running time dropped significantly. Our reconstruction result is shown in Fig 4(b). We can observe from this figure that the main planes in this scene are correctly recovered, and LSs on scene are precisely represented in the reconstructed model, with correct annotations of the planes to which they belong. We overlapped our result

[1] More experimental results and the generated 3D models are available on http://kailigo.github.io/projects/LineReconstruction.

(a) (b) (c) (d) (e)

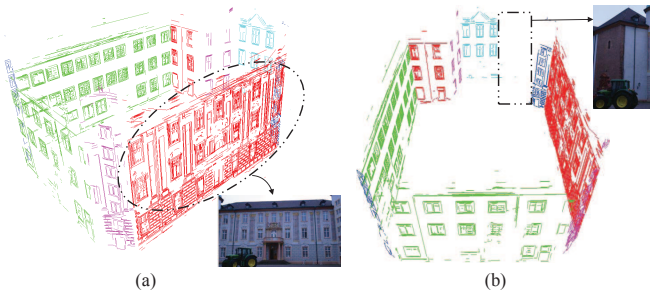
Fig. 4. The 3D line segment reconstruction results of the proposed method on a synthesis image dataset.

	$\rho = 1.0$				$\rho = 0.6$			
	Ours	[7]	[6]	[5]	Ours	[7]	[6]	[5]
ME	0.083	0.162	0.065	NR	0.076	0.137	0.044	0.029
RMSE	0.135	0.291	0.196	NR	0.111	0.189	0.080	0.046

Table 1. The comparative Mean Error (ME) and Root Mean Square Error (RMSE) of the reconstruction results obtained by our method and several other ones on a synthesis dataset. NR denotes that the corresponding measure data is Not Reported in the authors' paper(s).

with the group truth CAD model to qualitatively evaluate the reconstruction accuracy, as shown in Fig 4(c). As we can see, the vast majority of the reconstructed LSs (in black) cling to or closely approach the ground truth model, indicating that they are precisely reconstructed. To quantitatively evaluate the reconstruction accuracy, following [7, 6, 5], we calculated the Hausdorff distances between densely sampled points along the 3D LSs and the ground truth CAD model and computed the Mean Error (ME) and Root Mean Square Error (RMSE). We compare our results with the other methods, as that presented in Table 1. Since we could not get the implements of the compared methods, we can only compare our measure data with those reported in the authors' papers. Table 1 shows that when we set the cutoff threshold $\rho = 1.0$, as that applied in [6], the ME of our result is much better than [7], and only slightly inferior to [6]. But for the RMSE, our result is the best one. When we set $\rho = 0.6$ as that used in [5], our result is better than [7], but worse than both [6] and [5]. However, one fact needs to be noticed is that both the ME and RMSE for our result increase slightly as we increased ρ from 0.6 to 1.0, while they increase more significant for these two methods. It is a reasonable inference that if we increase ρ again, our method will be the best one among all these methods because our method has strong constraints for avoiding the emergence of gross outliers, but the others do not. Our reconstructed 3D LSs are obtained by projecting image LSs to the estimated space plane. If the space planes are precisely estimated, they will provide strong constraints on the locations of the 3D LSs. Once LS matches are correctly grouped, it is unlikely to have gross outliers. In fact, we found that when ρ is greater than 1.5, both ME and RMSE for our result remain unchanged and fix at 0.099 and 0.192, respectively. This indicates that all the sampled 3D points on the reconstructed 3D LSs have errors less than this threshold, and gross outliers do not exist.

The above experiment shows that our method can generate comparable results with others using one third of the images they used. In fact, our method can generate satisfactory result with even less images. Fig 4(d) show the 3D LS reconstruction result of the proposed method on 27 images obtained by taking one from every three images among the 80 images we used, and Fig 4(e) shows the corresponding overlapping result between the reconstructed model and the ground truth model. We compare this model, referred as $\mathcal{M}_{27}$, with that generated from 80 images, referred as $\mathcal{M}_{80}$. Qualitatively speaking, there is no significant different between $\mathcal{M}_{27}$ and $\mathcal{M}_{80}$, except some missing LSs on the roof and bottom of the captured house; the walls of the house are identically and completely reconstructed in both models. The great overlap between $\mathcal{M}_{27}$ and the ground truth model shown in Fig 4(e) indicates that no gross outliers exist in this model, just the same as $\mathcal{M}_{80}$. Quantitatively speaking, the ME and RMSE for $\mathcal{M}_{27}$ are 0.068 and 0.098, respectively, when the cutoff threshold $\rho = 0.6$, and are 0.071 and 0.108, respectively, when $\rho = 1.0$. In both cases, the two measures for M_{27} are slightly better than that for M_{80}. Moreover, same as M_{80}, both ME and RMSE for M_{27} increase slightly with a greater ρ.



(a) (b)

Fig. 5. The 3D line segment reconstruction result of the proposed method on a real image dataset.

Real images We then conducted experiment on a real image dataset [33], which contains 30 images photographing a scene consisting of several houses. The reconstruction result is shown in Fig 5. As we can see, the 3D LSs lying on the main planes of scene are well reconstructed. The details of scenes are precisely presented in the reconstructed model, see the bricks and windows of the selected dashed elliptical region shown in Fig 5(a). However, our method failed to reconstruct 3D LSs on the two main planes of this scene shown in the selected rectangle region in Fig 5(b). This is because only several LSs are extracted on these two planes and a even fewer LS matches are obtained. Our method is unable to reliably estimate a space plane when LS matches induced by 3D LSs on the plane are too few, and hence incapable to obtain the 3D LSs on it. The same reason also explains the missing on our model the representations of scene LSs on the roofs of the houses, and also the false reconstruction of the 3D LSs lying on the bottom of the house for the synthesis image dataset, as shown in Fig 4(b). Incorporating point matches into space plane estimation seems to be a solution for this problem, as did in [22]. Solving this problem will be our immediate future plan.

Running time The proposed LS reconstruction method is easy to implement, and can be very efficient because it involves only some simple geometric operations of the 2D and 3D LSs. The algorithm is currently implemented based on MATLAB. The unrefined codes took 32.2s on the synthesis dataset using 80 images, and 52.2s on the real image dataset.

5. CONCLUSIONS

In this paper, we have presented a new algorithm for reconstructing 3D LSs on piecewise planar scenes. We propose a new solution for solving the uncertainties in 3D LS reconstruction by estimating space planes from clustered LS matches and projecting image LSs to the space planes. To limit the incidence of the false clustering of LS matches, we propose to solve the problem of LS match clustering by solving a multi-label optimization problem using graph cut. Experiments on both synthesis and real image datasets verify the effectiveness of the proposed method and it is superiority to others for its advantages on obtaining satisfactory results using much fewer images and on reconstructing precise 3D LSs while recovering the space where they lie.

6. REFERENCES

[1] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski. Building Rome in a day. Communications of the ACM, Vol. 54, No. 10, Pages 105-112, October 2011.

[2] C. Wu. Towards linear-time incremental structure from motion. In *3DV*, 2011.

[3] Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 32(8):1362-1376, 2010.

[4] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: exploring photo collections in 3D. *ACM Transactions on Graphics*, 25(3): 835-846, 2006.

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

[6] M. Hofer, A. Wendel, and H. Bischof. Incremental line-based 3D reconstruction using geometric constraints. In *BMVC*, 2013.

[7] A. Jain, C. Kurz, T. Thormahlen, and H. P. Seidel. Exploiting global connectivity constraints for reconstruction of 3D line segments from images. In *CVPR*, 2010.

[8] Y. Boykov, O. Veksler, and R. Zabih. Efficient approximate energy minimization via graph cuts. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 36(12), 1222-1239, 2001.

[9] R. Hartley and A. Zisserman. Multiple view geometry in computer vision. Cambridge university press, 2003.

[10] D. Matinec and T. Pajdla. Line reconstruction from many perspective images by factorization. In *CVPR*, 2003.

[11] B. Přibyl, P. Zemčík, and M. Čadík Camera pose estimation from lines using plücker coordinates. In *BMVC* 2015.

[12] A. Bartoli and P. Sturm. Structure-from-motion using lines: Representation, triangulation, and bundle adjustment. *Computer Vision and Image Understanding*, 100(3):416-441, 2005.

[13] P. Smith, I. D. Reid, and A. J. Davison. Real-time monocular SLAM with straight lines. In *BMVC* 2006.

[14] 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.

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

[16] A. F. Habib, M. Morgan, and Y. R. Lee Bundle adjustment with self-calibration using straight Lines. *The Photogrammetric Record*, 17(100), 635-650, 2002.

[17] C.J. Taylor and D.J. Kriegman Structure and motion from line segments in multiple images. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 17(11), 10211032, 1995.

[18] M. E. Spetsakis and J. Y. Aloimonos. Structure from motion using line correspondences. *International Journal of Computer Vision*, 4(3), 171-183, 1990.

[19] L. Zhang and R. Koch Structure and motion from line correspondences: representation, projection, initialization and sparse bundle adjustment. *Journal of Visual Communication and Image Representation*, 25(5), 904-915.

[20] G. Schindler, P. Krishnamurthy, and F. Dellaert. Line-based structure from motion for urban environments. In *3DPVT*, 2006.

[21] C. Kim and R. Manduchi Planar Structures from Line Correspondences in a Manhattan World. In *ACCV*, 2014.

[22] S. N. Sinha, D. Steedly, and R. Szeliski. Piecewise planar stereo for image-based rendering. In *ICCV*, 2009.

[23] H. Bay, A. Ess, A. Neubeck, and L. Van Gool. 3D from line segments in two poorly-textured, uncalibrated images. In *3DPVT*, 2006.

[24] S. Ramalingam and M. Brand. Lifting 3D Manhattan lines from a single image. In *ICCV*, 2013.

[25] K. Li, J. Yao, X. Lu, L. Li, and Z. Zhang. Hierarchical line matching based on line-junction-line structure descriptor and local homography estimation. doi:10.1016/j.neucom.2015.07.137, 2016.

[26] D. Hoiem, A. Efros, and M. Hebert. Closing the loop in scene interpretation. In Computer Vision and Pattern Recognition. In *CVPR*, 2008.

[27] J. Pan. Coherent Scene Understanding with 3D Geometric Reasoning. *Ph.D. Thesis, Carnegie Mellon University*, 2014.

[28] O. Teboul, L. Simon, P. Koutsourakis, and N. Paragios. Segmentation of building facades using procedural shape priors. In *CVPR*, 2010.

[29] J. A. Delmerico, P. David, and J. J. Corso. Building facade detection, segmentation, and parameter estimation for mobile robot stereo vision. *Image and Vision Computing*, 31(11), 841-852, 2013.

[30] J. Lee, Y. Lu, and D. Song. Planar building facade segmentation and mapping using appearance and geometric constraints. In *IROS*, 2014.

[31] A. Delong, A. Osokin, H. N. Isack, and Y. Boykov. Fast approximate energy minimization with label costs. *International Journal of Computer Vision*, 96(1), 1-27, 2012.

[32] T. T. Pham, T. J. Chin, J. Yu, and D. Suter. The random cluster model for robust geometric fitting. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 36(8), 1658-1671, 2014.

[33] C. Strecha, W. V Hansen, L. V. Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *CVPR*, 2008.