SCIENCE CHINA
Information Sciences

• LETTER •

January 2020, Vol. 63 119102:1–119102:3
https://doi.org/10.1007/s11432-019-9867-9

Effective two-view line segment reconstruction based on structure priors

Wei WANG[1*], Hainan CUI[2,3], Wei GAO[2,3] & Zhanyi HU[2,3]

[1]*School of Network Engineering, Zhoukou Normal University, Zhoukou* 466001*, China;*
[2]*National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing* 100190*, China;*
[3]*University of Chinese Academy of Sciences, Beijing* 100049*, China*

Received 25 January 2019/Accepted 1 April 2019/Published online 12 September 2019

Citation Wang W, Cui H N, Gao W, et al. Effective two-view line segment reconstruction based on structure priors. Sci China Inf Sci, 2020, 63(1): 119102, https://doi.org/10.1007/s11432-019-9867-9

Dear editor,
For the 3D reconstruction of the line segments detected in images, traditional line segment (LS) detection and matching methods [1, 2] usually encounter the following problems when noise and illumination variation are involved: (1) the detected LSs deviate from real scene structures; (2) only a minority of the detected LSs with better discriminating features can be correctly matched (see Figure 1(a)); (3) a certain amount of falsely matched LSs are generated. As a result, as shown in Figure 1(e) and (f), the 3D LSs reconstructed from the LS matches contain larger errors, and are also too sparse to describe the complete scene structures.

To address the above issues, this study presents an effective two-view LS reconstruction method based on structure priors (i.e., plane priors and angle priors).

Methodology. Our method is comprised of three main components: pre-processing, 3D LS inference, and 3D LS optimization.

(1) Pre-processing. Given two calibrated images I_1 and I_2, our method starts from identifying building regions using the method [3] (excluding those LSs in unrelated regions such as the sky and the ground), and detecting and matching the LSs in these building regions using the method [2].

Let $\mathcal{L}$ and $\overline{\mathcal{L}}$ denote the matched LSs and the non-matching LSs in I_1, respectively. In order to jointly infer the 3D LSs in $\overline{\mathcal{L}}$ and correct falsely matched LSs in $\mathcal{L}$, we construct an initial neighboring LS set $N(l)$ for each LS $l \in \mathcal{M}$ (where $\mathcal{M} = \mathcal{L} \cup \bar{\mathcal{L}}$) using the Delaunay triangulation method [2]. Note that $N(l)$ will be heuristically updated later. Moreover, the intersection angle prior A between two planes is set to $[0^0, 30^0, 45^0, 60^0, 90^0, 120^0, 135^0, 150^0]$.

(2) 3D LS inference using co-plane cues. 3D LS inference focuses on inferring more 3D LSs by exploiting potential scene planes, and the overall process is described as Algorithm 1.

Algorithm 1 3D LS inference using co-plane cues

Input: initial LS matches $\mathcal{L}$.
Output: inferred and corrected LS matches $\mathcal{L}^*$.
1: Generate seed planes S from $l \in \mathcal{M}$ and its neighboring LS set $N(l)$; add reliable LS matches to $\mathcal{L}^*$.
2: Explore dominant scene planes $\mathcal{H}$.
3: For each seed plane $P(l) \in S$,
  3.1: Generate its plane family $\mathcal{H}^* : \{h_i\}$ from the associated plane $h \in \mathcal{H}$;
  3.2: Assign the optimal plane h_i to unvisited LSs in $\{l, N(l)\}$;
  3.3: Add reliable LS matches to $\mathcal{L}^*$ and update $\mathcal{H}$;
  3.4: Add the LS set to S if any of its component LSs is assigned to the optimal plane;
  3.5: Update the neighbors of LS l.
4: Output $\mathcal{L}^*$.

• Generate seed planes. Given an LS $l \in \mathcal{M}$ and its neighboring LS set $N(l)$, for using co-plane cues, we construct the corresponding matched LS set $\mathcal{M}_l = \{l, N(l)\} \cap \mathcal{L}$, and triangulate each LS in $\mathcal{M}_l$ to the 3D LS. Then, we use all the 3D LSs to fit a local plane, and consider this plane as a

* Corresponding author (email: wangwei@zknu.cn)

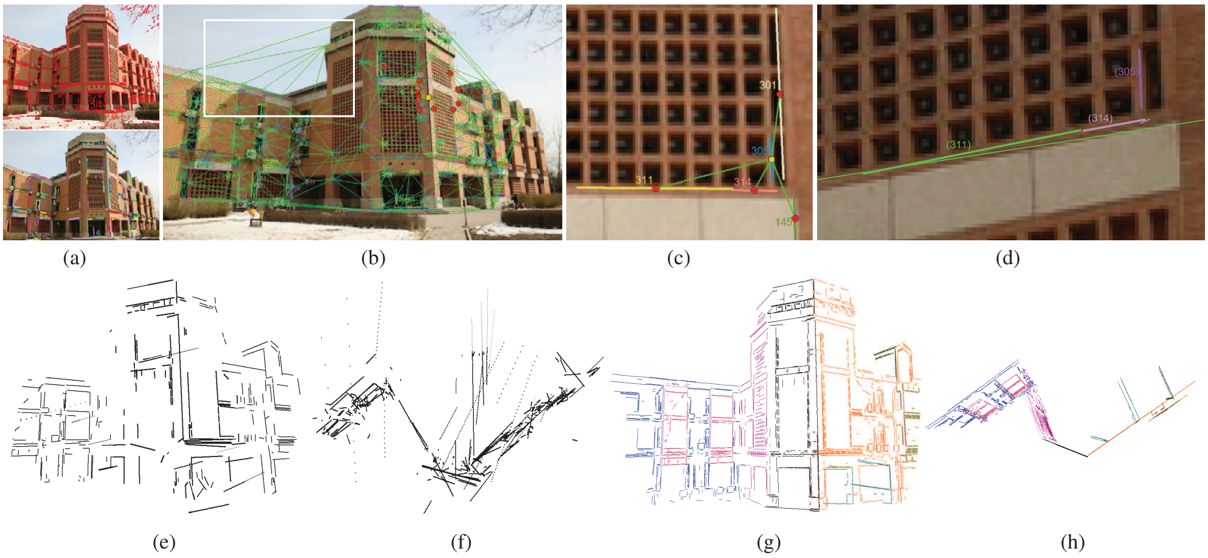

Figure 1 (Color online) 3D LS reconstruction. (a) Initial detected LSs (top) and fewer matched LSs (bottom) by the method of [2]; (b) the neighboring system constructed by the Delaunay triangulation method (the yellow square and red circles denote the current LS and its neighboring LSs, respectively); (c) initial non-matching LS #305 (yellow square dot) and its neighboring LSs (red circle dots) in I_1; (d) by finding co-planes, LS #(305) in I_2 is correctly matched with LS #305 in I_1, and initial falsely matched LSs (thin LSs) in I_2 (with #311 and #314 in I_1) are corrected as #(311) and #(314) (thick LSs); (e) 3D LSs reconstructed from initial LS matches; (f) the top-view of (e); (g) 3D LSs produced by our method (different colors denote the 3D LSs in different planes); (h) the top-view of (g).

seed plane $P(l)$ if the corresponding fitting error (i.e., the average distance between this plane and all the endpoints of these 3D LSs) is smaller than the threshold θ (set to 0.05).

• Explore dominant planes. For a seed plane $P(l)$, we further exploit the supporting 3D LSs corresponding to the LSs in $\mathcal{L}$ to detect dominant scene planes. To this end, for an LS $x \in \mathcal{L}$, we first take the corresponding 3D LS X as the support inlier of $P(l)$ if the average distance between $P(l)$ and the two endpoints of X is smaller than θ, and then update $P(l)$ by fitting all of its inliers. Finally, we select the updated plane $P(l)$ which is associated with at least K (set to 10) 3D LSs as the dominant scene plane, and add it to the set $\mathcal{H}$.

• Infer 3D LSs. For a dominant plane $h \in \mathcal{H}$ and its associated LS set $\{l, N(l)\}$, we first calculate the derived planes by rotating the plane h to each angle in A around the axes where the inlier 3D LSs lie, and construct the plane family $\mathcal{H}^* : \{h_i\}$. Then, according to $\mathcal{H}^*$, we attempt to infer or correct the 3D LS corresponding to each LS $x \in \{l, N(l)\}$ under the following two cases.

(i) LS $x \in \mathcal{L}$. In this case, we calculate the average distance between the two endpoints of its corresponding 3D LS X and each plane $h_i \in \mathcal{H}^*$, and select the plane h_i with the smallest distance as the candidate plane relevant to LS x. Then, we consider the candidate plane optimal to LS x if the smallest distance is smaller than θ or the following condition is met:

$$F(x, h_i(x)) < F(x, x'), \tag{1}$$

where $h_i(x)$ denotes the LS induced by the plane h_i in I_2, and $F(x, x')$ denotes the following DAISY feature [4] similarity between LSs x and x':

$$F(x, x') = \frac{1}{k} \sum_{i=1}^{k} s(F(x_i), F(x'_i)), \tag{2}$$

where x_i and x'_i denote the corresponding i-th pixel in LSs x and x', respectively, $F(x_i)$ and $F(x'_i)$ denote the corresponding DAISY features, and $s(\cdot)$ denotes the similarity measure.

In fact, LS $h_i(x)$ is more likely to match with LS x in both geometry and image feature. We thus update the initial LS match (x, x') by $(x, h_i(x))$.

(ii) LS $x \in \bar{\mathcal{L}}$. Here, we assign it the plane h_i^* if the following condition is met:

$$h_i^* = \arg \min_{h_i \in \mathcal{H}^*} \bar{F}(x, h_i(x)), \tag{3}$$

where $\bar{F}(x, h_i(x))$ denotes the set of the DAISY feature similarities (i.e., $F(x, h_i(x))$) that are smaller than the mean value of the DAISY feature similarities of all matched LSs.

• Update neighbors. For the above two cases, the 3D LS X can be obtained simply by back-projecting LS x from I_1 onto its assigned plane. Nevertheless, if LS x is not assigned to an appropriate plane, we consider that it may not belong to $P(l)$ and remove it from $\{l, N(l)\}$ (i.e., eliminating the relations between it and other LSs in $\{l, N(l)\}$). Conversely, for an isolated LS x removed from other LS sets neighboring to $\{l, N(l)\}$, it is added to $N(l)$ (i.e., constructing the relation between it and LS l) only if it meets the conditions of the two above cases with respect to the dominant planes associated with LS set $\{l, N(l)\}$; otherwise, it will be discarded after all seed planes are traversed.

• Update seed planes and dominant scene planes. In inferring 3D LSs, an LS set $\{l, N(l)\}$ will be taken as a seed plane if any of its component LSs obtains the optimal planes from other seed planes; moreover, the optimal plane assigned to an LS is simultaneously added to the set $\mathcal{H}$ to generate more dominant scene planes.

(3) Jointly optimizing 3D LSs and planes. After 3D LS inference, each LS in $\mathcal{M}$ can be assigned to a valid plane, and the reconstructed 3D LSs are basically reliable. To obtain the globally consistent 3D LSs, we further perform a cooperative optimization over these 3D LSs by minimizing the following energy function.

$$E(H) = \sum_{l \in \mathcal{M}} E(H_l) + \alpha \sum_{k \in \bar{N}(l)} E(H_l, H_k) + \beta \sum_{k,m \in \bar{N}(l)} E(A_{lkm}), \quad (4)$$

where H_l denotes the current plane assigned to LS $l \in \mathcal{M}$ and $\bar{N}(l)$ denotes the updated neighbors of LS l in inferring 3D LSs; $E(H_l)$, $E(H_l, H_k)$ and $E(A_{lkm})$ are the data term, plane regularization term and angle regularization term, respectively; α and β are the weights of the latter two terms (set to 0.6 and 0.4, respectively).

• Data term. $E(H_l)$ is used to measure the cost of assigning a plane H_l to the current LS l, and is defined as

$$E(H_l) = L(D(l, H_l)) + \kappa \cdot L(F(l, H_l(l))), \quad (5)$$

where $D(l, H_l)$ denotes the average distance between plane H_l and the two endpoints of LS L, $F(l, H_l(l))$ denotes the DAISY feature similarity between LS l and LS $H_l(l)$ induced by plane H_l in I_2, $L(\cdot)$ is the Logistic function, and κ is the weight of the image features (set to 0.5).

• Plane regularization. $E(H_l, H_k)$ encodes the structure prior that two neighboring LSs are likely to be coplanar in 3D space in a high possibility and is defined as

$$E(H_l, H_k) = \begin{cases} \mathrm{e}^{-d(L,K)} \cdot \mathrm{e}^{-F(l,k)}, & H_l \neq H_k, \\ 0, & \text{otherwise}, \end{cases} \quad (6)$$

where $d(L, K)$ denotes the average distance between 3D LS L and K, and $F(l, k)$ denotes the DAISY feature similarity defined as in (2).

• Angle regularization. $E(A_{lkm})$ encodes a higher-order structure prior in a simple manner, incorporates the intersection angles between planes, and is defined as:

$$E(A_{lkm}) = \begin{cases} \mathrm{e}^{-P_{lkm}}, & A(H_{lk}, H_{lm}) \notin A, \\ 0, & \text{otherwise}, \end{cases} \quad (7)$$

where H_{lk} denotes the plane determined by LSs l and k (similarly for H_{lm}), $A(H_{lk}, H_{lm})$ denotes the intersection angle between planes H_{lk} and H_{lm}, and $P_{lkm} = \min_{a \in A} |A(H_{lk}, H_{lm}) - a|$ denotes the minimum of the differences between $A(H_{lk}, H_{lm})$ and the angles in A.

As shown in Figure 1(g) and (h), through solving (4) using the method [5], the approximate optimal solution can be achieved, and the resulting 3D LSs can describe complete scene structures better.

Conclusion. Based on structure priors, our method first progressively exploits the potential planes corresponding to the matched and nonmatching LSs, and then globally regularizes the resulting planes and 3D LSs via the cooperative optimization method. Experimental results confirm that our method can effectively produce as many reliable 3D LSs as possible and their associated planes.

Acknowledgements This work was supported in part by National Key R&D Program of China (Grant No. 2016YFB0502002), National Natural Science Foundation of China (Grant Nos. 61872361, 61772444, 61421004, 61873264), National Science Foundation for Young Scientists of China (Grant No. 61703397), Natural Science Foundation of Henan Province (Grant No. 162300410347), and Key Scientific and Technological Project of Henan Province (Grant Nos. 162102310589, 192102210279).

References

1 von Gioi R G, Jakubowicz J, Morel J M, et al. LSD: a fast line segment detector with a false detection control. IEEE Trans Pattern Anal Mach Intell, 2010, 32: 722–732

2 Li K, Yao J. Line segment matching and reconstruction via exploiting coplanar cues. ISPRS J Photogramm Remote Sens, 2017, 125: 33–49

3 Zhao H S, Shi J P, Qi X J, et al. Pyramid scene parsing network. In: Proceedings of Computer Vision and Pattern Recognition, Honolulu, 2017. 6230–6239

4 Tola E, Lepetit V, Fua P. DAISY: an efficient dense descriptor applied to wide-baseline stereo. IEEE Trans Pattern Anal Mach Intell, 2010, 32: 815–830

5 Huang X F. Cooperative optimization for energy minimization: a case study of stereo matching. 2007. ArXiv:cs/0701057