SCIENCE CHINA
Information Sciences

• MOOP •

July 2023, Vol. 66 174101:1–174101:3
https://doi.org/10.1007/s11432-022-3631-6

ASPPR: active single-image piecewise planar 3D reconstruction based on geometric priors

Wei WANG[1,2], Qiulei DONG[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 17 April 2022/Revised 4 August 2022/Accepted 9 October 2022/Published online 3 April 2023

Citation Wang W, Dong Q L, Hu Z Y. ASPPR: active single-image piecewise planar 3D reconstruction based on geometric priors. Sci China Inf Sci, 2023, 66(7): 174101, https://doi.org/10.1007/s11432-022-3631-6

Single-image piecewise planar reconstruction aims at recovering piecewise planar structures of a building from an image. Considering the intrinsic spatial ambiguity in a single image, it is an enormous challenge for existing geometry and learning-based methods [1, 2] to obtain reliable results on complex building structures. To address the problem, motivated by the successful use of interactive strategies in other visual tasks, we propose an active single-image piecewise planar reconstruction (ASPPR) method that integrates automatic plane inference and interactive plane refinement using geometric priors.

Methodology. In an image, each polygonal region typically corresponds to a building plane, and its component edge is frequently associated with a double plane (specific angles or occlusions occur between two planes) or single plane (a plane is adjacent to a non-plane or self-occlusions occur in a building) structures. Accordingly, the following three patch-based convolutional neural networks (CNNs) are constructed to detect the geometric priors corresponding to single or double structures where H_1 and/or H_2 denote the associated planes.

(1) A-GP: Angles (i.e., 0°, 45°, 90°, and 135°) and occlusions (including occluding structures with double and single planes) are detected between planes H_1 and H_2.

(2) C-GP: Corners (i.e., convex and concave) constructed by planes H_1 and H_2 are detected.

(3) O-GP: Orientations (i.e., frontal, right-facing, and left-facing) of planes H_1 and/or H_2 with respect to the camera are detected.

The A-GP, C-GP, and O-GP CNNs are trained using an improved ResNet-34 with the ResNet2 module [3] and the training samples are collected from public data sets with ground-truth 3D points. More specifically, in each image, line segments are first detected using existing line segment detectors, and a square patch centered at each line segment is then extracted. Then, the samples are collected according to the fitted planes (H_1 and/or H_2) using the 3D points corresponding to two parts on both sides of each line segment (e.g., if the planes H_1 and H_2 intersect each other, the intersection angle is then calculated to construct a sample for the former four classes of the A-GP CNN; otherwise, a sample is collected for the fifth class of the A-GP CNN).

Based on geometric priors, the proposed ASPPR method is designed by integrating four components: interactive region partitioning, automatic plane inference, interactive plane modification, and automatic plane optimization.

(1) Interactive region partitioning. To reliably reconstruct each component plane of a building, the corresponding polygonal region is first annotated by continuously clicking on its potential vertices. In practice, the initial edges of the annotated polygonal region frequently deviate from real 3D line structures (i.e., the boundaries of the plane) owing to inaccurate clicks. It is therefore necessary to automatically correct these edges to construct an accurate polygonal region. More specifically, for an initial edge x of the annotated polygonal region, its slope is first updated using the following three constraints:

(i) For constraints constructed using an automatically detected line segment, the slope of edge x is replaced with that of an automatically detected line segment y when the following condition is met:

$$P(x,y) = \left(\bar{d}(x,y) < k_1\right) \wedge \left(\bar{s}(x,y) < k_2\right) \wedge \left(\bar{l}(y) > k_3\right), \tag{1}$$

where $\bar{d}(x,y)$ and $\bar{s}(x,y)$ denote the average distance and the difference in slope angle between edge x and line segment y, respectively; $\bar{l}(y)$ denotes the length of line segment y; k_1, k_2 and k_3 are the corresponding thresholds.

(ii) For constraints constructed by two parallel edges, when edges x and y belong to the same polygonal region, the slope of edge x is replaced with that of edge y when $\bar{s}(x,y) < k_2$ and $P(A,y) > P(A,x)$, where $P(A,x)$ (the same holds for $P(A,y)$) is defined as

$$P(A,x) = P(A|x) \cdot P(x), \tag{2}$$

$P(A|x)$ is the output probability of the A-GP CNN, $P(x)$ is defined as

$$P(x) = \sum\nolimits_{i \in Z} \delta\left(d(i,x) < k_1\right)/\bar{l}(x), \tag{3}$$

* Corresponding author (email: qldong@nlpr.ia.ac.cn)

Z denotes the points detected using existing edge detection methods, and $\delta(\cdot)$ is equal to 1 when the input condition is true, and is 0 otherwise.

(iii) For constraints constructed with geometric consistency, when edges x and y belong to different polygonal regions, the slope of edge x is replaced with that of edge y when $\bar{d}(x,y) < k_1$ is met and edge y has been corrected.

After updating the slopes of all initial edges of the annotated polygonal region, the lines that the initial edges lie on are computed based on their mid-points and new slopes. Furthermore, as shown in Figure 1(b), the resulting lines intersect each other to construct a new polygonal region and replace the annotated polygonal region. Furthermore, for each component edge of a polygonal region, the A-GP, C-GP, and O-GP CNNs are used to detect the corresponding geometric priors associated with double and single plane structures. To this end, three patches with different sizes (64×64, 128×128 and 256×256) centered at the current edge are extracted to feed into each of the three CNNs, respectively, and the geometric prior with the highest of the three output probabilities is taken as the optimal result.

(2) Automatic plane inference. For a polygonal region produced through interactive region partitioning, each of its edges is associated with double and single plane structures, and the planes can be automatically inferred using the constraints constructed using the related geometric priors. Without loss of generality, with ASPPR, the vertical edges are used to conduct the plane inference and other edges are taken as auxiliary conditions. More specifically, the vertical edge l with reliable geometric priors and its associated planes (called H_l^L and/or H_l^R) is first selected to construct the constraints for inferring the planes associated with other vertical edges. To this end, the following criterion is defined to measure the overall reliability of the geometric priors of a vertical edge:

$$M(l) = (1-\kappa)\cdot R(N_l) + \kappa \cdot A(N_l, \theta_l), \tag{4}$$

where $R(N_l)$ and $A(N_l, \theta_l)$ denote the orientation and angle consistencies, respectively, and κ is the weight parameter.

The orientation consistency $R(N_l)$ is defined as

$$R(N_l) = \exp\left(-\max\left(\left\langle N_l^L, \bar{N}_l^L\right\rangle, \left\langle N_l^R, \bar{N}_l^R\right\rangle\right)\right), \tag{5}$$

where $\langle x, y\rangle$ denotes the angle between orientations x and y, and N_l^L and $\bar{N}_l^L$ (the orientations of planes H_l^L) are obtained by computing the direction perpendicular to two vanishing directions produced using the edges of the current polygonal region and the O-GP CNN, respectively (the same holds for the orientations N_l^R and $\bar{N}_l^R$ of planes H_l^R).

The angle consistency $A(N_l, \theta_l)$ is defined as

$$A(N_l, \theta_l) = \begin{cases} \exp(-|\langle N_l^L, N_l^R\rangle - \theta_l|), & \Lambda_{LR} = \Lambda_l, \\ 0, & \text{otherwise}, \end{cases} \tag{6}$$

where θ_l and Λ_l denote the learned angle and corner between planes H_l^L and H_l^R, respectively, and Λ_{LR} denotes the corner computed using the orientations N_l^L and N_l^R.

According to the defined $M(l)$, the geometric priors of a vertical edge l are considered reliable when the corresponding $M(l)$ value is larger than a pre-defined threshold θ. Furthermore, the vertical edge l with the largest $M(l)$ value (hereafter called a seed vertical edge) is selected to compute planes H_l^L and H_l^R according to the geometric priors where the space height is set according to the prior knowledge of a typical range of building heights.

Then, starting from the seed vertical edge l, the planes corresponding to its left and right neighboring vertical edges (denoted by vertical edges l_- and l_+) are inferred using their associated geometric priors. Furthermore, this process will be repeated by taking vertical edges l_- and l_+ as seed vertical edges until each polygonal region is assigned to a plane. As indicated in Algorithm 1 and Figure 1(c), we take vertical edge l_- as an example to describe the process of the plane inference (the same holds for vertical edges l_+).

Algorithm 1 Geometric prior-based plane inference

Input: Planes associated with seed vertical edge l.
Output: Planes associated with vertical edges on the left side of vertical edge l.
Initialization: Plane set $\mathcal{M}$ = null;
1: For vertical edge l_-:
2: $H_{l_-}^R = H_l^L$ and $H_{l_-}^L$ = null;
3: IF θ_{l_-} is reliable:
4: IF $N_{l_-}^L$ is reliable: Compute $H_{l_-}^L$, update Λ_{l_-} using θ_{l_-} and $N_{l_-}^L$, and go to step 8;
5: IF Λ_{l_-} is reliable: Compute $H_{l_-}^L$, update $N_{l_-}^L$ using θ_{l_-} and Λ_{l_-}, and go to step 8;
6: ELSE
7: IF $\Lambda_{l_-}^{LR} = \Lambda_{l_-}$: Update θ_{l_-} using $\langle N_{l_-}^L, N_{l_-}^R\rangle$ and go to step 3;
8: IF $H_{l_-}^L \neq$ null and the occlusion is invalid for double plane structures:
9: Save $H_{l_-}^L$ and $H_{l_-}^R$ to $\mathcal{M}$;
10: IF $M(l) > \theta$: Set vertical edge l_- to the current seed vertical edge l and go to step 1;
11: ELSE
12: Save $H_{l_-}^R$ to $\mathcal{M}$;
13: Select a new seed vertical edge as the current seed vertical edge l and go to step 1;
14: Output plane set $\mathcal{M}$.

(3) Interactive plane modification. To interactively correct false planes produced through automatic plane inference with minimal user cost, as shown in Figures 1(d)–(f), three types of interactive manners based on click correlations (called IPM-1, IPM-2, and IPM-3) are defined as follows.

(i) IPM-1 indicates increasing (or decreasing) the depth of a plane by conducting different types of clicks (e.g., left or right clicks) at the center of the current polygonal region.

(ii) IPM-2 indicates the setting of two planes to the same plane by conducting different types of clicks at the centers of the polygonal regions corresponding to the current and changed planes, respectively.

(iii) IPM-3 indicates a clockwise (or anticlockwise) change in the orientation of a plane with a pre-specified interval by conducting different types of clicks at the centers of the polygonal regions corresponding to the current and reference planes, respectively.

(4) Automatic plane optimization. A global plane optimization method that incorporates image cues and geometric priors is utilized to improve the reconstruction accuracy of piecewise planar structures. More specifically, letting $\mathcal{R}$ denote the set of polygonal regions, to assign the optimal plane to each polygonal region, the plane is optimized by minimizing the following cost function:

$$E(\mathcal{H}) = \sum_{r\in\mathcal{R}} \bar{E}(\mathcal{H}_r) + \alpha \cdot \sum_{s\in N(r)} \widetilde{E}(\mathcal{H}_{r,s}) + \beta \cdot \sum_{s\in N(r)} \hat{E}(\mathcal{H}_{r,s}), \tag{7}$$

where $\mathcal{H}_r$ and $\mathcal{H}_{r,s}$ denote the current plane assigned to polygonal region r and two planes corresponding to two neighboring polygonal regions r and s, respectively, $N(r)$

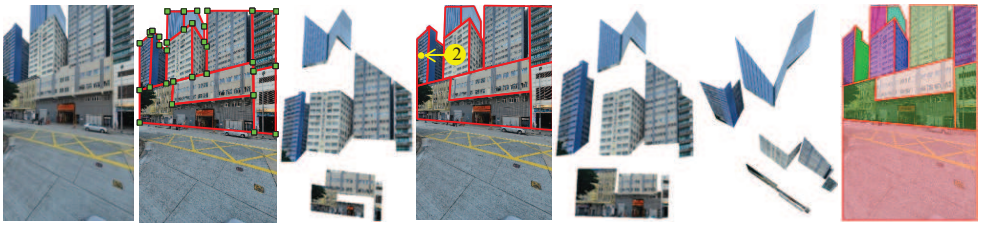

Figure 1 (Color online) Example of the proposed ASPPR method. (a) Input image; (b) interactively partitioned polygonal regions using clicks (green squares); (c) automatically inferred planes; (d) interactively corrected planes using click correlations (IPM-2); (e) and (f) two views of the corrected planes; (g) globally optimized planes (different colors denote different planes).

is the set of polygonal regions neighboring to polygonal region r, and α and β are the weight parameters.

The unary potential encodes the depth and orientation prior of the current plane $\mathcal{H}_r$ assigned to polygonal region r, and is defined as

$$E\left(\mathcal{H}_r\right)=(1-\rho)\cdot V\left(n_r,\bar{n}_r\right)+\rho\cdot D\left(\mathcal{H}_r\right), \tag{8}$$

where n_r and $\bar{n}_r$ denote the orientation of plane $\mathcal{H}_r$ and the orientation perpendicular to two vanishing directions produced using the edges of polygonal region r, respectively, and ρ is the weight used to balance the orientation prior $V\left(n_r,\bar{n}_r\right)$ and depth prior $D\left(\mathcal{H}_r\right)$, in which $V\left(x,y\right)$ denotes the difference between orientations x and y, and $D\left(\mathcal{H}_r\right)$ is computed according to the following three cases.

Case I: Polygonal region r is adjacent to the ground, and $D\left(\mathcal{H}_r\right)$ is defined as

$$D\left(\mathcal{H}_r\right)=T\left(\bar{d}\left(l^r_{\text{bottom}},l^{\mathcal{H}_r}_{\text{depth}}\right)\Big/H_{PR}\right), \tag{9}$$

where l^r_{bottom} is the bottom edge of polygonal region r; $l^{\mathcal{H}_r}_{\text{depth}}$ is the line segment generated according to the depth of plane $\mathcal{H}_r$; H_{PR} is the maximum height of the polygonal regions in $\mathcal{R}$; and $T\left(\cdot\right)$ is the tanh function.

Case II: Polygonal region r is adjacent to a polygonal region s that is corrected to a reliable plane $\mathcal{H}_s$ through interactive plane modification, and $D\left(\mathcal{H}_r\right)$ is defined as

$$D\left(\mathcal{H}_r\right)=T\left(\bar{d}\left(L_r,L_s\right)/\bar{D}_s\right), \tag{10}$$

where L_r and L_s denote the 3D line segments produced by back-projecting the shared edge to planes $\mathcal{H}_r$ and $\mathcal{H}_s$, respectively, and $\bar{D}_s$ denotes the average depth value of the points on the 3D line segment L_s.

Case III: Polygonal region r is not adjacent to the ground or the polygonal region described in Case II, and $D\left(\mathcal{H}_r\right)$ is computed using the method in Case II when one neighbor of polygonal region r is assigned a reliable plane measured through (4).

The first pairwise potential encodes the angle regularization between two neighboring planes and is defined as

$$\widetilde{E}\left(\mathcal{H}_{r,s}\right)=\begin{cases}T\left(\left|\langle\mathcal{H}_r,\mathcal{H}_s\rangle-A_m\right|\right), & \langle\mathcal{H}_r,\mathcal{H}_s\rangle\notin A,\\ 0, & \text{otherwise},\end{cases} \tag{11}$$

where A and $A_m\in A$ denote pre-defined angles and the angle having the largest difference with angle $\langle\mathcal{H}_r|\mathcal{H}_s\rangle$, respectively.

The second pairwise potential encodes the appearance regularization between two neighboring polygonal regions, and is defined as

$$\hat{E}\left(\mathcal{H}_{r,s}\right)=\begin{cases}\text{S\_GP}\left(\mathcal{H}_r,\mathcal{H}_s\right), & \mathcal{H}_r\neq\mathcal{H}_s,\\ 0, & \text{otherwise},\end{cases} \tag{12}$$

where $\text{S\_GP}\left(\mathcal{H}_r,\mathcal{H}_s\right)$ denotes the output probability of the S-GP CNN where the two patches on both sides of the shared edge between polygonal regions r and s are extracted and taken as inputs.

To reliably solve (7), the candidate planes used are produced by augmenting the plane obtained through automatic plane inference and interactive plane modification in an automatic manner based on the principle of IPM-1 and IPM-3 (i.e., rotating a plane at the pre-defined angles to produce new planes). Finally, starting from the polygonal regions described in Cases I and II, the cooperative optimization method proposed in [4] is applied to solve (7). As shown in Figure 1(g), the optimized planes are reliable.

Conclusion. The study presents an active single-image piecewise planar reconstruction method based on geometric priors. With this method, the geometric priors learned using CNNs are applied to partition the building regions into polygonal regions in a click-based interactive manner, and automatically infer the planes corresponding to the resulting polygonal regions in a progressive manner. Furthermore, under the guidance of an interactive plane refinement based on the click correlations, the inferred planes are globally optimized under a unified framework incorporating image cues and geometric priors. Extensive experiments (see Appendixes A–C) demonstrate that the proposed method can produce more satisfactory results.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61991423, U1805264), Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB32050100), and Beijing Municipal Science and Technology Project (Grant No. Z211100011021004).

Supporting information Videos and Appendixes A–C. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Liu X, Zhao Y, Zhu S C. Single-view 3D scene reconstruction and parsing by attribute grammar. IEEE Trans Pattern Anal Mach Intell, 2017, 40: 710–725
2 Qian Y, Furukawa Y. Learning pairwise inter-plane relations for piecewise planar reconstruction. In: Proceedings of European Conference on Computer Vision, 2020. 330–345
3 Gao S H, Cheng M M, Zhao K, et al. Res2Net: a new multi-scale backbone architecture. IEEE Trans Pattern Anal Mach Intell, 2019, 43: 652–662
4 Huang X. Cooperative optimization for energy minimization: a case study of stereo matching. 2007. ArXiv:cs/0701057