Supplementary material for:
P\textsuperscript{2}Net: Patch-match and Plane-regularization for Unsupervised Indoor Depth Estimation

Zehao Yu\textsuperscript{1,2}, Lei Jin\textsuperscript{1,2}, and Shenghua Gao\textsuperscript{1,3}

\textsuperscript{1} ShanghaiTech University
\textsuperscript{2} DGene Inc
\textsuperscript{3} Shanghai Engineering Research Center of Intelligent Vision and Imaging
\{yuzh,jinlei,gaoshh\}@shanghaitech.edu.cn
https://github.com/svip-lab/Indoor-SfM Learner

1 Surface normal visualization

We provide more visualizations of surface normal prediction on the ScanNet [1] dataset. In our implementation, we directly fit the surface normal from ground truth depth annotation. Black pixels indicate invalid regions where no ground truth depths are provided. Compared to MovingIndoor [3], our surface normal estimation better preserves the boundary of the planar regions, thanks to our superpixel constraint.

2 Point clouds visualization

We further provide some point clouds visualization on NYUv2 [2] and ScanNet [1] dataset in Figure 2.

3 The effect of different patterns.

We compare the effect of different patterns in our Patch-match module. We experiment with different $N$s and report the result in Table 1. Setting $N$ to 3 gives best results.
Table 1: Comparison between different patterns in our Patch-match module.

<table>
<thead>
<tr>
<th>N</th>
<th>rms</th>
<th>rel</th>
<th>$\delta &lt; 1.25$</th>
<th>$\delta &lt; 1.25^2$</th>
<th>$\delta &lt; 1.25^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.629</td>
<td>0.173</td>
<td>0.746</td>
<td>0.939</td>
<td>0.984</td>
</tr>
<tr>
<td>2</td>
<td>0.618</td>
<td>0.170</td>
<td>0.748</td>
<td>0.937</td>
<td>0.984</td>
</tr>
<tr>
<td>3</td>
<td>0.612</td>
<td>0.166</td>
<td><strong>0.758</strong></td>
<td><strong>0.945</strong></td>
<td><strong>0.985</strong></td>
</tr>
<tr>
<td>4</td>
<td>0.634</td>
<td>0.173</td>
<td>0.741</td>
<td>0.938</td>
<td>0.984</td>
</tr>
</tbody>
</table>
Fig. 1: Visualization of surface normal results on the ScanNet [1] dataset. From left to right: input RGB, MovingIndoor [3], our results and surface normal fitted from ground truth depth. Black pixels in ground truth indicate invalid regions where no depth ground truth are provided.
Fig. 2: Point cloud visualization. From left to right: input RGB from NYUv2, point cloud in 3D, RGB from ScanNet, point cloud in 3D.
References