M^3VSNet: Unsupervised Multi-metric Multi-view Stereo Network

Baichuan Huang\(^{1,3}\), Hongwei Yi\(^{2}\), Can Huang\(^{3}\), Yijia He\(^{3}\), Jingbin Liu\(^{1}\), Xiao Liu\(^{3}\)

1. Wuhan University 2. Peking University 3. Megvii Technology Limited

huangbaichuan@whu.edu.cn
https://whubaichuan.github.io/
Introduction

Multi-view stereo (MVS)

Images

Feature

Sparse

SFM

Dense

MVS

Applications

Traditional method

Supervised learning method

Heavy workload on 3D training ground-truth data

Unsupervised method

MVSnet
1. Determine the final depth map based on the constraints of photometric consistency.
2. Use geometric consistency to measure whether the depth between multiple views is consistent or not.
A novel unsupervised multi-metric MVS network for dense point cloud reconstruction without any supervision.

In non-ideal environments, the same color can’t be guaranteed in multi-view images. We extract semantic feature from VGG network.

We combine the pixel-level and feature level unsupervised loss function.

Code is available: https://github.com/whubaichuan/M3VSNet
## Result

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean Distance(mm)</th>
<th>Acc.</th>
<th>Comp.</th>
<th>overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Furu</td>
<td>0.612</td>
<td>0.939</td>
<td>0.775</td>
<td></td>
</tr>
<tr>
<td>Tola</td>
<td>0.343</td>
<td>1.190</td>
<td>0.766</td>
<td></td>
</tr>
<tr>
<td>Colmap</td>
<td>0.400</td>
<td>0.664</td>
<td>0.532</td>
<td></td>
</tr>
<tr>
<td>SurfaceNet</td>
<td>0.450</td>
<td>1.043</td>
<td>0.746</td>
<td></td>
</tr>
<tr>
<td>MVSNet(D=192)</td>
<td>0.444</td>
<td>0.741</td>
<td>0.592</td>
<td></td>
</tr>
<tr>
<td>Unsup_MVS</td>
<td>0.881</td>
<td>1.073</td>
<td>0.977</td>
<td></td>
</tr>
<tr>
<td>MVS2</td>
<td>0.760</td>
<td>0.515</td>
<td>0.637</td>
<td></td>
</tr>
<tr>
<td>M3VSNet(D=192)</td>
<td>0.636</td>
<td>0.531</td>
<td>0.583</td>
<td></td>
</tr>
</tbody>
</table>

*lower is better*

### DTU dataset

- (a) Ground Truth
- (b) MVSNet
- (c) M³VSNet w/o Feature-wise Loss
- (d) M³VSNet

### Tanks & Temples dataset

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean</th>
<th>Family</th>
<th>Francis</th>
<th>Horse</th>
<th>Light house</th>
<th>M60</th>
<th>Panther</th>
<th>Play-around</th>
<th>Train</th>
</tr>
</thead>
<tbody>
<tr>
<td>M3VSNet</td>
<td>37.67</td>
<td>47.74</td>
<td>24.38</td>
<td>18.74</td>
<td>44.42</td>
<td>43.45</td>
<td>44.95</td>
<td>47.39</td>
<td>30.31</td>
</tr>
<tr>
<td>MVS2</td>
<td>37.21</td>
<td>47.74</td>
<td>21.55</td>
<td>19.50</td>
<td>44.54</td>
<td>44.86</td>
<td>46.32</td>
<td>43.48</td>
<td>29.72</td>
</tr>
<tr>
<td>SurfaceNet</td>
<td>26.30</td>
<td>36.97</td>
<td>11.01</td>
<td>7.25</td>
<td>37.55</td>
<td>26.94</td>
<td>26.18</td>
<td>42.98</td>
<td>21.53</td>
</tr>
</tbody>
</table>

*more efficient*
Thanks for your attention!

Baichuan Huang
huangbaichuan@whu.edu.cn
https://whubaichuan.github.io/
Wuhan University, China