Spatio-Temporal Depth Estimation in Real-Time

HyunJun Jung
(Supervisor : Benjamin Busam)
Spatio-Temporal Depth Estimation in Real-Time

Background and Motivation
I want depth

Well... Network can predict depth with an image!
I want better depth

Hmmm... more views gives better result!

Train with two views [1]

Train with multi views [2]

[1] Unsupervised Monocular Depth Estimation with Left-Right Consistency Godard et al. CVPR 2017
Spatio-Temporal Depth Estimation in Real-Time
Spatio-Temporal Depth Estimation in Real-Time

Is... Based on \textbf{Active Stereo Net} and \textbf{Multi View Stereo Net}

Let’s talk about them first!
Multi View Stereo Net

- It uses Multiview information to get very accurate depth map
- Variance metric for cost volume makes it possible to use arbitrary number of views

MVSNet: Depth Inference for Unstructured Multi-view Stereo, Yao et al. ECCV 2018
Active Stereo Net

- It upsamples without upconvolution (fast)
- Network trains in unsupervised manner (no depth GT required)

ActiveStereoNet: End-to-End Self-Supervised Learning for Active Stereo Systems
Zhang et al. ECCV 2018
Spatio-Temporal Depth Estimation in Real-Time

Is.. Based on **Active Stereo Net** and **Multi View Stereo Net**

**MVSNet Part :**
- Using Multiview input image to predict coarse level depth

**ASN Part :**
- Bilinear interpolation & guided refinement will upsample the depth *(much faster than upconvolution)*
- Self supervised training scheme will be applied at last.

Let’s predict good depth with Multiview, in real time
Spatio-Temporal Depth Estimation in Real-Time

- Extract feature & Diff. H. Warp
  - $L_{t-1}$
  - $L_t$
  - $R_t$
  - $L_{t+1}$
- Variance metric
  - Cost volume
- Soft argmin
  - $d$ (1/8 res depth)
- Full res depth
- Spatial invalidation (ex. Reference check)
- Temporal invalidation (ex. Optical flow)
Spatio-Temporal Depth Estimation in Real-Time

Coarse depth prediction:
1. Read multi-view
2. Get each view
3. Feature extraction
4. 1/8 res feature map
5. Warp to first view
6. [view_idx : 0]
7. Coarse depth
8. Soft argmin
9. Reduce dimension
10. Calculate variance
11. {weight = n_views}
12. Warped each view

Depth refinement:
13. Downsample to ¼ res
14. Bilinear upsampling to ¼ res
15. Image guided refinement
16. Refined ¼ res depth
17. Downsample to ½ res
18. Bilinear upsampling to ½ res
19. Image guided refinement
20. Refined ½ res depth
21. Bilinear upsampling to ½ res
22. Image guided refinement
23. Refined full res depth

HyunJun Jung - Spatio-Temporal Depth Estimation in Real-Time
Final Goal

Train multi view monocular camera to predict depth.

Input
- Realtime multiview monocular view
- R & sT from monocular SLAM (s = unknown scale)

Output
- Realtime depth map
- Scaling factor for Translation

Possible application
- Whenever we are limited to use monocular camera but need tracking (ex, endoscopy camera)
Spatio-Temporal Depth Estimation in Real-Time

Plans, Timeline
Plan for Implementation

Stereo Camera Phase

• Supervised training with exact extrinsic (R, t) from simulation
• Supervised training with estimated extrinsic from SLAM
• Self-supervised training with estimated extrinsic

Mono Camera Phase

• Supervised training with extrinsic (R, scaled t).
  *(scaling factor will be also predicted)*
• Self-supervised version with same condition.
## Timeline for Project

<table>
<thead>
<tr>
<th>Name</th>
<th>Begin date</th>
<th>End date</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMSD</td>
<td>10/15/18</td>
<td>2/8/19</td>
</tr>
<tr>
<td>Kick off presentation</td>
<td>11/5/18</td>
<td>11/5/18</td>
</tr>
<tr>
<td>Short sprint for CVPR</td>
<td>10/15/18</td>
<td>11/23/18</td>
</tr>
<tr>
<td>Preparing for training</td>
<td>10/15/18</td>
<td>11/9/18</td>
</tr>
<tr>
<td>Training Network</td>
<td>11/10/18</td>
<td>11/15/18</td>
</tr>
<tr>
<td>Preparing supplemental data</td>
<td>11/16/18</td>
<td>11/23/18</td>
</tr>
<tr>
<td>Preparing for demo</td>
<td>11/19/18</td>
<td>2/1/19</td>
</tr>
<tr>
<td>Acquire more dataset</td>
<td>11/19/18</td>
<td>12/7/18</td>
</tr>
<tr>
<td>Train for demo</td>
<td>12/17/18</td>
<td>1/18/19</td>
</tr>
<tr>
<td>Code for demo</td>
<td>1/7/19</td>
<td>1/25/19</td>
</tr>
<tr>
<td>Test for demo</td>
<td>1/21/19</td>
<td>2/1/19</td>
</tr>
<tr>
<td>Intermediate presentation</td>
<td>12/10/18</td>
<td>12/14/18</td>
</tr>
<tr>
<td>Prepare</td>
<td>12/10/18</td>
<td>12/13/19</td>
</tr>
<tr>
<td>D-day</td>
<td>12/13/18</td>
<td>12/14/18</td>
</tr>
<tr>
<td>Final presentation</td>
<td>1/28/19</td>
<td>2/8/19</td>
</tr>
<tr>
<td>Prepare</td>
<td>1/28/19</td>
<td>2/4/19</td>
</tr>
<tr>
<td>D-day</td>
<td>2/4/19</td>
<td>2/8/19</td>
</tr>
</tbody>
</table>

- **Short sprint till 11/23**
- **Preparing for Demo till Feb**

HyunJun Jung - Spatio-Temporal Depth Estimation in Real-Time
Spatio-Temporal Depth Estimation in Real-Time

Q & A