CBMV: A Coalesced Bidirectional Matching Volume for Disparity Estimation
Konstantinos Batsos, Changjiang Cai, Philippos Mordohai

Motivation

- Purely data-driven learning-based methods have been successful in stereo matching by replacing hand-crafted rules with data-driven predictions. However, these methods tend to overspecialize in the training domain.
- Conventional matching functions enforce desirable invariances which could have been learned from data.
- With CBMV we aim to generate a robust matching volume able to compete with purely data-driven methods, but also to generalize much better in different domains.
- We integrate conventional matching methods in a supervised learning framework to achieve robustness.
- We show competitive results with the fast MC-CNN architecture and an improved capability to generalize on Middlebury 2014, KITTI 2012 and 2015, and ETH3D benchmarks using the same model.

Hypothesis: This is due to the learning method not being directly exposed to image appearance.

Method

- Four matchers are coalesced with bidirectional confidence features to create CBMV.
- We use a random forest classifier to predict matching likelihood.
- The final disparity maps are extracted from the matching volume using the optimization and filtering pipeline of MC-CNN.

Features

- The cost computed by each of the four different matchers $C_{	ext{min}}(x,y,d) = \text{CE/N/S/U}(S(x,y,d))$.
- For each matching function we compute two features, Ratio and Likelihood in a bidirectional manner.

\[
\begin{align*}
R(x,y,d) &= \frac{C_{\text{min}}(x,y,d)}{\min_{x,y} C_{\text{min}}(x,y,d)} \\
L(x,y,d) &= \sum \left( \frac{C_{\text{min}}(x,y,d) - C_{\text{min}}(x,y,d)}{\min_{x,y} C_{\text{min}}(x,y,d)} \right)^{\frac{1}{2}}
\end{align*}
\]

Bidirectional Matching

- For a given element of the matching volume, the ratio and likelihood features are computed along the yellow and red lines corresponding to the right and left epipolar lines respectively.
- Black dots denote a few intersections of left and right epipolar lines on the matching volume. Each intersection is a matching hypothesis linking a pixel in the left image with a pixel in the right image.

Training

CBMV is trained on the training and additional sets of the Middlebury 2014 dataset for ALL benchmark submissions.

Results

<table>
<thead>
<tr>
<th></th>
<th>bad-2.0</th>
<th>bad-2.0</th>
<th>argenmr</th>
<th>rsn-emr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Middlebury 2014 Dev set</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC-CNN sad</td>
<td>8.08%</td>
<td>19.1%</td>
<td>17.9%</td>
<td>55.0%</td>
</tr>
<tr>
<td>CBMV(sad)</td>
<td>11.1%</td>
<td>20.5%</td>
<td>14.4%</td>
<td>46.9%</td>
</tr>
<tr>
<td>MC-CNN sad</td>
<td>9.47%</td>
<td>20.6%</td>
<td>10.3%</td>
<td>57.7%</td>
</tr>
</tbody>
</table>

Generalization

- For our robust vision challenge submission, we replaced the post processing and optimization pipeline with the local expansion algorithm [T. Taniai et al. Continuous 3D label stereo matching using local expansion moves]. Our results show that CBMV does not underfit or overfit any of the benchmarks.

Robust Vision Challenge

For our robust vision challenge submission, we replaced the post processing and optimization pipeline with the local expansion algorithm [T. Taniai et al. Continuous 3D label stereo matching using local expansion moves]. Our results show that CBMV does not underfit or overfit any of the benchmarks.

<table>
<thead>
<tr>
<th>dataset</th>
<th>MCBV</th>
<th>MCBV</th>
<th>CBMV</th>
<th>CBMV</th>
</tr>
</thead>
<tbody>
<tr>
<td>metric</td>
<td>bad-2.0</td>
<td>bad-2.0</td>
<td>all</td>
<td>all</td>
</tr>
<tr>
<td>Middlebury 2014</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBMV_ROB</td>
<td>7.60%</td>
<td>13.3%</td>
<td>4.56%</td>
<td>5.06%</td>
</tr>
<tr>
<td>CBMV</td>
<td>11.1%</td>
<td>20.5%</td>
<td>4.14%</td>
<td>5.37%</td>
</tr>
</tbody>
</table>
| Acknowledgements

This research has been supported by the National Science Foundation under awards IIS 1522729 and 1637761. We are grateful to Enrique Dunn for constructive discussions.