Review for GA-Net: Guided Aggregation Net for End-to-end Stereo Matching

Can Koz

February 5, 2020

1 Summary

In stereo matching tasks, the main goal is to match two images by finding disparity between the two images. Optimal disparity estimation depends on the cost function and in this paper two novel neural net layers are proposed to calculate the cost. Instead of widely used 3D conv layers this approach uses deep guided aggregation networks and outperforms the state of the art approaches. This approach also requires less computational/memory complexity on both Scene Flow Dataset and KITTI benchmarks.

2 Strengths

Authors propose two novel layers to increase performance while decreasing the computational cost for stereo matching task. State of the art methods uses many 3D convolutional layers for stereo reconstruction. In the paper, two novel cost aggregation layers are defined. The first one, semi global guided aggregation (SGA) layer, is a differential approximation of semi-global matching (SGM) which aggregates the matching cost in different directions. The second one, local guided aggregation layer (LGA) recover the loss of details caused by the downsampling and upsampling layers. With these 2 GA layers and additional two 3D convolutinal layers, performance is increased with respect to state of the art GC-Net which has 19 3D convolutional layers. This novel approach also has 1/100 complexity as that of 3D convolutional layers.

3 Weaknesses

Similar to other stereo matching problems, large textureless regions and occlusions remain unsolved. There is significant improvement in computational complexity, but it can be improved further.

4 Evaluation

The proposed method is evaluated on Scene Flow and KITTI datasets. Authors tried different neural network architectures and different number of GA layers. They compared the result with other architectures, such as 3D Convolutional layers and SGM layers. Their results shows 0.14% improvement on KITTI dataset and 0.9% improvement on Scene Flow dataset.

5 Final Comments and Future Work

In comparison with previous work done in stereo matching, GA-Net shows significant improvoment with less computational complexity. 3D convolutions can be replaced and better solutions for large textureless regions can be developed in future work.