Pseudo-LiDAR++: Accurate Depth for 3D Object Detection in Autonomous Driving

ICLR 2020

Summary

In this paper an extension to the recently proposed Pseudo-LiDAR framework is presented. A novel architecture for stereo depth estimation, called Stereo Depth Network (SDN), is proposed which overcomes inaccuracies in depth estimates at large distances by predicting depth directly instead of first computing a disparity map. Furthermore, a graph-based algorithm is devised that leverages sparse LiDAR measurements to correct the systematic bias of stereo depth estimators. The combination of these two components shows to outperform the results obtained using the Pseudo-LiDAR framework with the largest improvements reported on far-away objects.

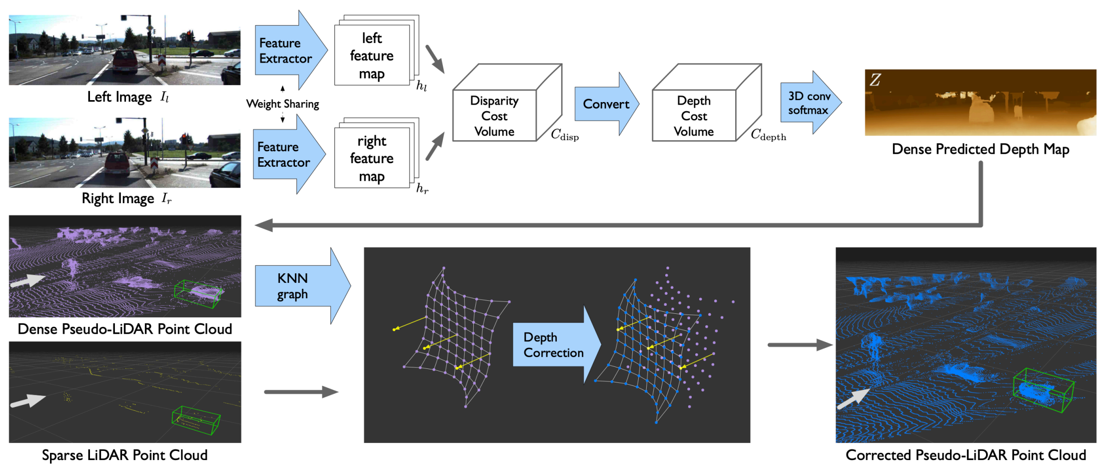

Figure 1: Illustration of the Pseudo-LiDAR++ pipeline. The Stereo Depth Network (SDN) for direct depth estimation is displayed in the upper half, the Graph-based Depth Correction (GDC) algorithm using sparse LiDAR in the lower half.

Main Contributions

- **Novel Depth Estimation Architecture** The authors present Stereo Depth Network (SDN) as a novel architecture for end-to-end trainable depth estimation without relying on disparity estimates. The approach is motivated by the quadratic relationship between errors in disparity and depth which results in degraded performance of conventional depth estimation networks for far-away objects. SDN is shown to outperform detection results obtained using previous state-of-the-art depth estimation models.
- **Depth Correction based on Sparse LiDAR** In order to correct to bias introduced by the discretization from depth to disparity, a graph-based approach is put forward that leverages accurate depth information obtained form a sparse LiDAR sensor to correct the depth estimate computed by the Stereo Depth Network. It is argued that a low number of LiDAR readings are sufficient to drastically increase the accuracy of the depth prediction.

Implementation Details

- **Depth Cost Volume** To be able to estimate depth directly, a replacement for the disparity cost volume used in conventional depth estimation models has to be formulated. The authors present the depth cost volume obtained using bilinear interpolation and the depth-to-disparity transform. All convolutional operations are performed on the depth cost volume to avoid the effect of disparity errors being more severe at larger distances.

- **Sparse LiDAR Generation** A sparse 4-beam LiDAR is simulated given the original KITTI LiDAR measurements. The points are ordered according to their elevation angle and only points corresponding to the resolution of a sparse LiDAR are selected.

- **Graph-based Depth Correction (GDC)** Depth correction is performed using the predicted depth map and the sparse LiDAR measurements. A K-nearest-neighbor graph is constructed and the edge weights are computed such that the depth of any points can be estimated from the depth of its neighbors. The correction step is then performed by changing the depth of those pixels for which accurate LiDAR measurements exist and then updating the depth of all other points in a way that the weights computed in the previous step can still be used to reconstruct the depth of any point. Thus, a bias in depth estimate can be removed while maintaining local consistency.

Evaluation

- **Ablation Study on SDN and GDC** Ablation experiments are performed to test the isolated contribution of SDN and GDC. It is shown that both components significantly enhance the performance compared to previous arts.

- **KITTI 3D and BEV Object Detection** A large number of experiments are presented including stereo baseline models, Pseudo-LiDAR models as well as models using full LiDAR measurements. It can be seen that the combination of SDN and GDC leads to a considerable performance increase compared to the methods using Pseudo-LiDAR, especially for IoU of 0.7.

		IoU = 0.5			IoU = 0.7		
Detection algorithm	Input	Easy	Moderate	Hard	Easy	Moderate	Hard
3DOP [4]	S	55.0 / 46.0	41.3 / 34.6	34.6 / 30.1	12.6 / 6.6	9.5 / 5.1	7.6 / 4.1
MLF-STEREO [42]	S	-	53.7 / 47.4	-	-	19.5 / 9.8	-
S-RCNN [21]	S	87.1 / 85.8	74.1 / 66.3	58.9 / 57.2	68.5 / 54.1	48.3 / 36.7	41.5 / 31.1
PL: AVOD [36]	S	89.0 / 88.5	77.5 / 76.4	68.7 / 61.2	74.9 / 61.9	56.8 / 45.3	49.0 / 39.0
PL: PIXOR*	S	89.0 / -	75.2 / -	67.3 / -	73.9 / -	54.0 / -	46.9 / -
PL: P-RCNN	S	88.4 / 88.0	76.6 / 73.7	69.0 / 67.8	73.4 / 62.3	56.0 / 44.9	52.7 / 41.6
PL++: AVOD	S	89.4 / 89.0	79.0 / 77.8	70.1 / 69.1	77.0 / 63.2	63.7 / 46.8	56.0 / 39.8
PL++: PIXOR*	S	89.9 / -	78.4 / -	74.7 / -	79.7 / -	61.1 / -	54.5 / -
PL++: P-RCNN	S	89.8 / 89.7	83.8 / 78.6	77.5 / 75.1	82.0 / 67.9	64.0 / 50.1	57.3 / 45.3
PL++: AVOD	L# + S	90.2 / 90.1	87.7 / 86.9	79.8 / 79.2	86.8 / 70.7	76.6 / 56.2	68.7 / 53.4
PL++: PIXOR*	L# + S	95.1 / -	85.1 / -	78.3 / -	84.0 / -	71.0 / -	65.2 / -
PL++: P-RCNN	L# + S	90.3 / 90.3	87.7 / 86.9	84.6 / 84.2	88.2 / 75.1	76.9 / 63.8	73.4 / 57.4
AVOD [16]	L + M	90.5 / 90.5	89.4 / 89.2	88.5 / 88.2	89.4 / 82.8	86.5 / 73.5	79.3 / 67.1
PIXOR* [47, 22]	L + M	94.2 / -	86.7 / -	86.1 / -	85.2 / -	81.2 / -	76.1 / -
P-RCNN [35]	L	96.3 / 96.1	88.6 / 88.5	88.6 / 88.5	87.8 / 81.7	86.0 / 74.4	85.8 / 74.5

Figure 2: Experimental results on the KITTI Benchmark. The same object detectors are compared for Pseudo-LiDAR, Pseudo-LiDAR++ with and without depth correction as well as the full LiDAR models.

References

This summary is solely based on my understanding of the original paper. All images used here are taken from the original paper as well. The paper can be found under the following link:
`https://arxiv.org/pdf/1906.06310.pdf`