Carnegie Mellon University
Mechanical Engineering

LiDAR Guided Boundary Bleeding Refinement for Depth Map

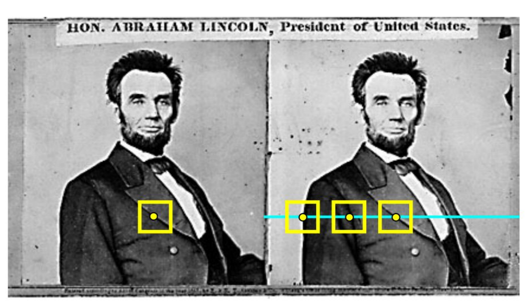
Zi Li, Fei Lu

{zili, flu2}@andrew.cmu.edu

Introduction

High resolution depth map has widespread applications in SLAM, segmentation, detection, planning, etc. However, producing a good quality depth map is still challenging.

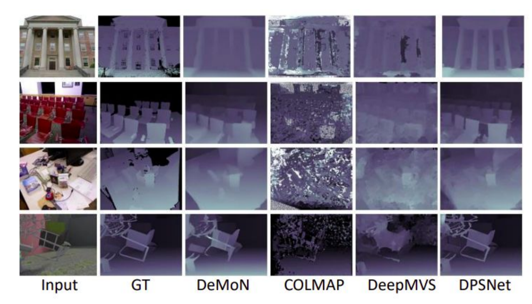
Stereo Block Matching



Pros: Easy to implement and parallel programming.
Cons: Subjected to repeated patterns, featureless areas, specularities issues.

Related work:
Semi-global Block Matching, StereoBM, Efficient Large-Scale Stereo Matching , etc.

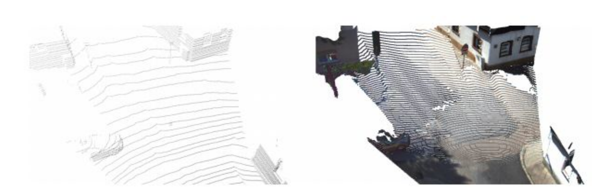
Deep Learning Depth Estimation



Pros: Smooth map, less noise, high overall accuracy.
Cons: Hard to generalize to unseen data, boundary bleeding, long tail issues.etc.

Related work:
PSMNet, DeMoN, DeepMVS, DPSNet.etc

Pointcloud Upsampling



Pros: Inference on real measured data.
Cons: Lidar projection covers very sparse areas, a good algorithm is required to make efficient upsampling.

Related work:
Delaunay triangulation, edge-preserving filters, etc.

How can we take advantages of methods above using multi-modal data?
A depth map could also be used directly for 3D object detection like point cloud and generates much denser pseudo-points!

Reference

[1] Cristiano Premebida, Luis Garrote, Alireza Asvadi, A. Pedro Ribeiro, and Urbano Nunes. High-resolution lidar-based depth mapping using bilateral filter. CoRR, abs/1606.05614, 2016.

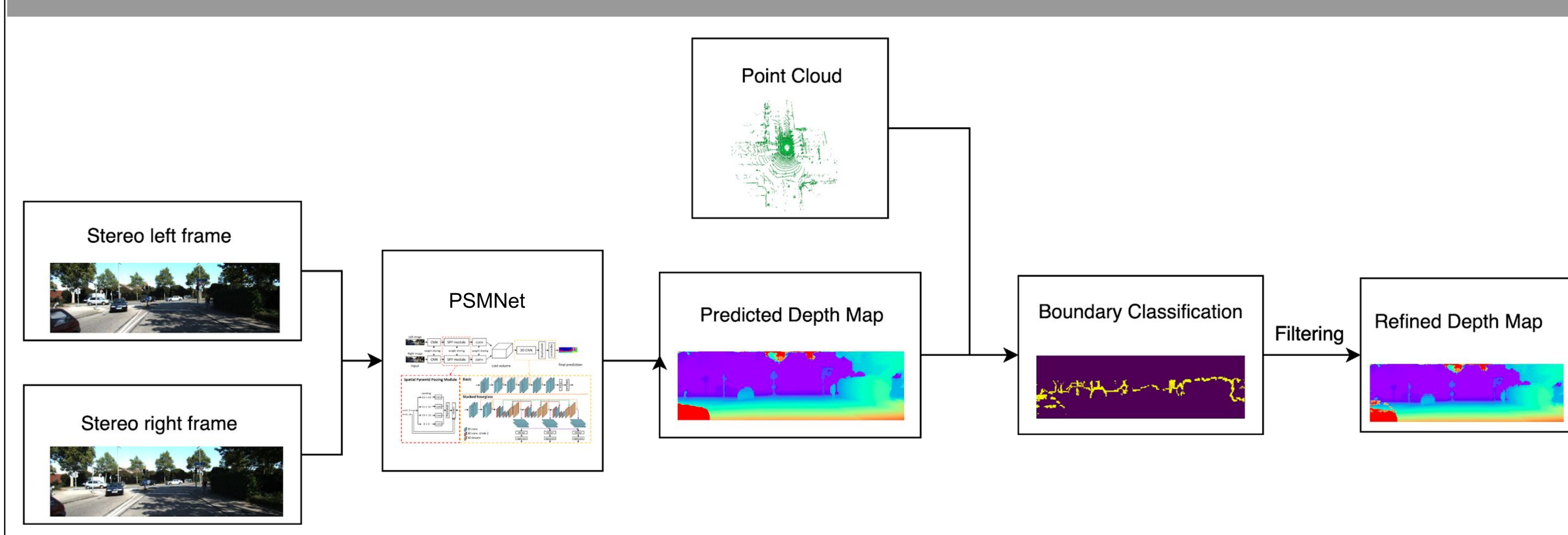
[2] Yan Wang, Wei-Lun Chao, Divyansh Garg, Bharath Hariharan, Mark Campbell, and Kilian Q. Weinberger. Pseudo-lidar from visual depth estimation: Bridging the gap in 3d object detection for autonomous driving. CoRR, abs/1812.07179, 2018.

[3] X. Weng and K. Kitani. Monocular 3D Object Detection with Pseudo-LiDAR Point Cloud. arXiv:1903.09847, 2019.

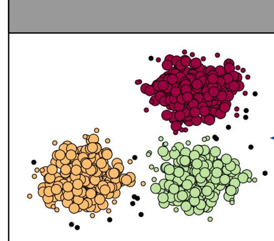
[4] Charles Ruizhongtai Qi, Wei Liu, Chenxia Wu, Hao Su, and Leonidas J. Guibas. Frustum pointnets for 3d object detection from RGB-D data. CoRR, abs/1711.08488, 2017.

Depth Map Refinement

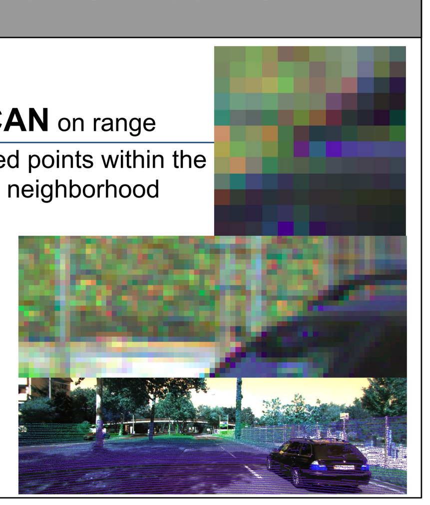
Pipeline



Boundary Classification



DBSCAN on range
on projected points within the 13×13 neighborhood



Dispersion Measurement

$$DF = \left|\frac{r_k - r_{k+1}}{r_k + r_{k+1}}\right|, \quad k = 1, \cdots, nR,$$

Boundary pixels have more than 1 cluster

Bilateral Filtering

A **weighted range estimation** is computed from neighborhood

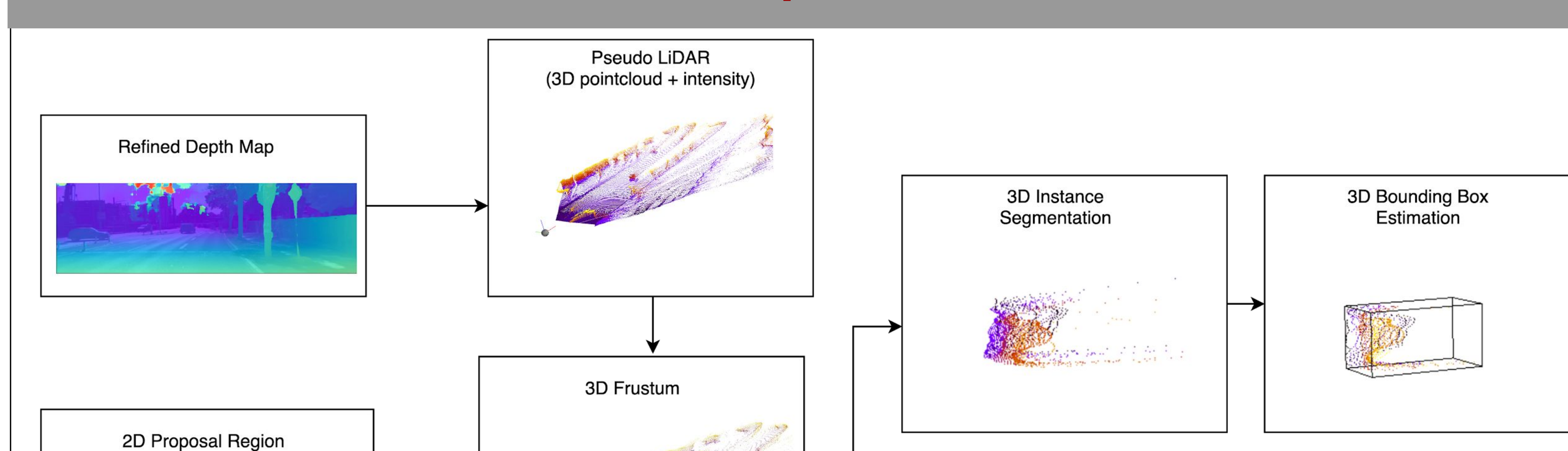
$$r_0^* = \frac{1}{W}\sum_{\mathbf{x}_i \in R} G_{\sigma_s}(||\mathbf{x}_0 - \mathbf{x}_i||) G_{\sigma_r}(|r_0 - r_i|) r_i$$

$$G_{\sigma_s} = \frac{1}{1 + (||\mathbf{x}_0 - \mathbf{x}_i||)}, \qquad G_{\sigma_r} = \frac{1}{1 + (|r_0 - r_i|)}.$$

$$r_0 = \min(r_i), \forall r_i \in R$$

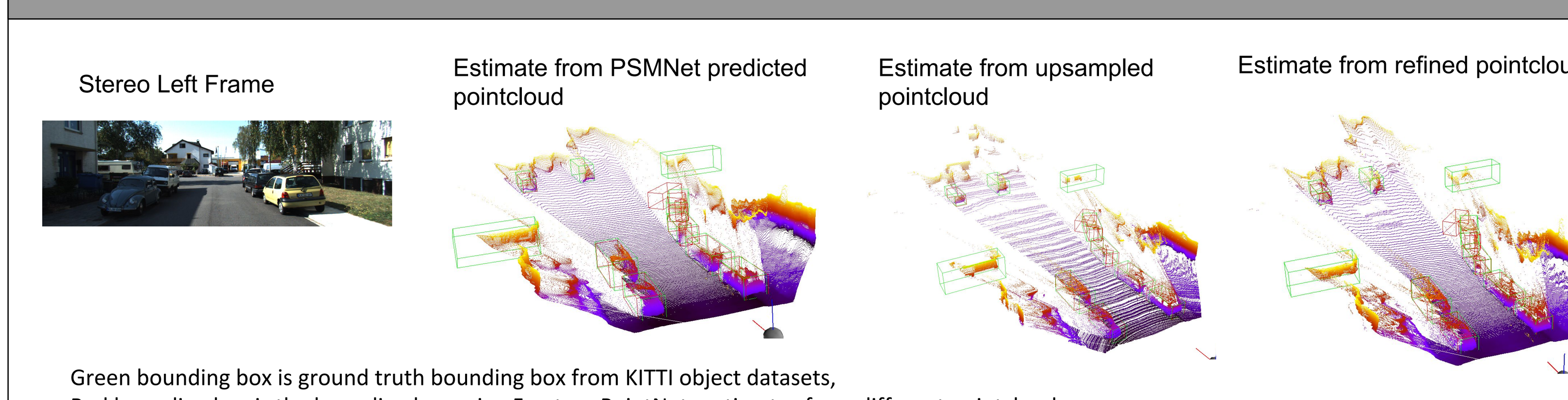
Pseudo LiDAR 3D Detection

Pipeline



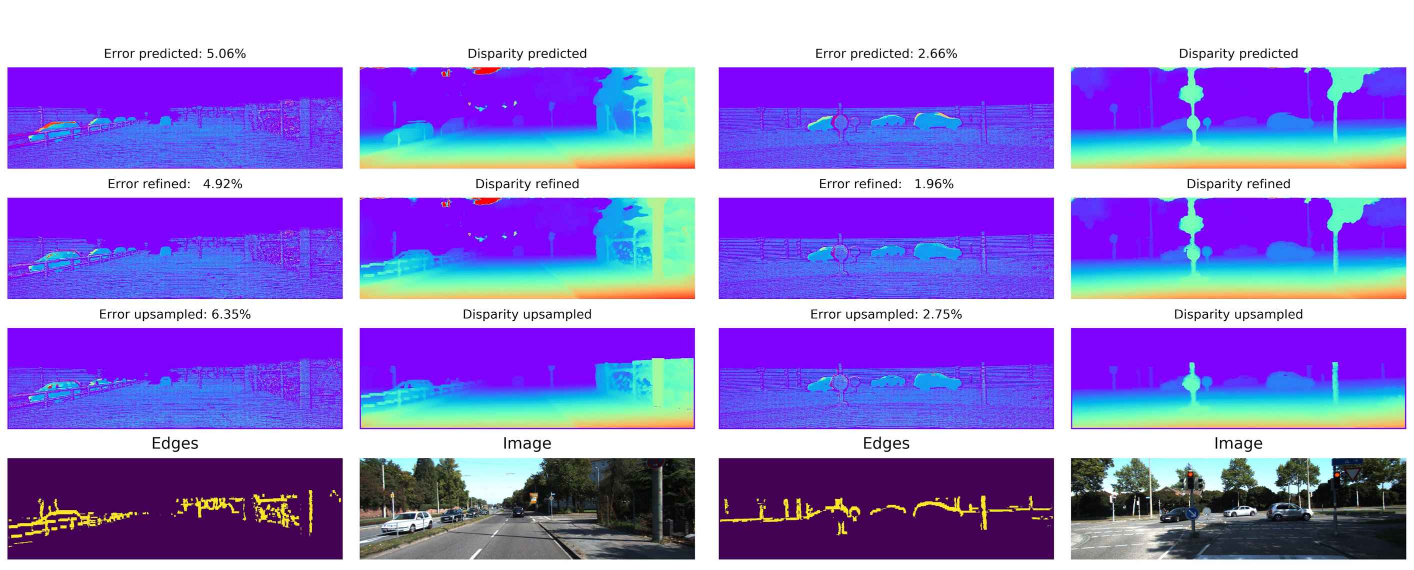
3D Bounding Box Estimation with Pre-trained Frustum PointNets

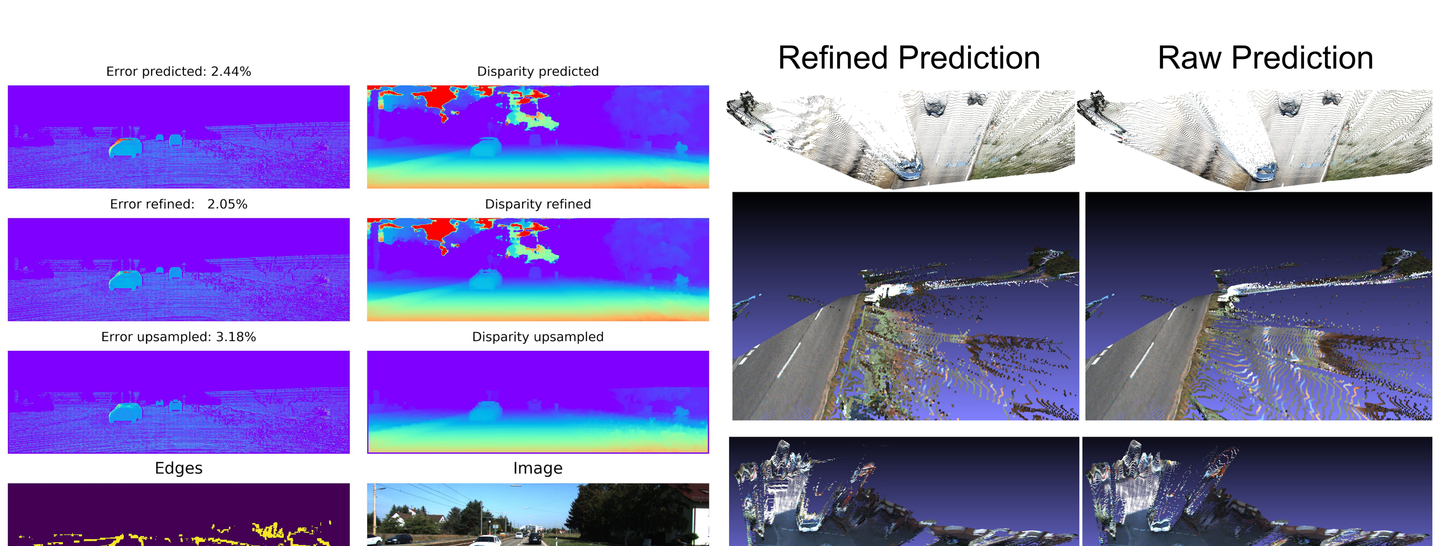
3D Bounding Box Comparison



Green bounding box is ground truth bounding box from KITTI object datasets,
Red bounding box is the bounding box using Frustum PointNets estimates from different pointcloud.

Results and Analysis





Depth Evaluation

Method	D1-fg	D1-bg	D1-all	Sig-Err pixels
Raw Prediction	11.61	2.13	4.01	1163.52
Refined Prediction	9.93	2.39	4.13	830.50
LiDAR Upsampled	11.87	4.23	6.24	750.93

Table 1: The benchmark is evaluated using a one-to-one match style, instead of the KITTI official evaluation tool which requires interpolation to maintain the same density. Sig-Err pixels here means the average number of pixels which exist error greater than 15. Blue is the baseline provided by [2].

3D Objection Detection Evaluation

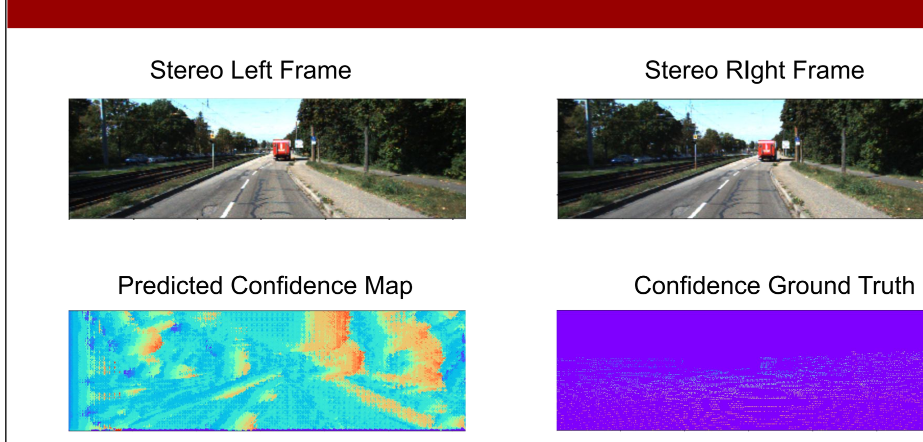
	Easy		Medium		Hard	
	BEV	3D	BEV	3D	BEV	3D
Raw Prediction	74.45	56.20	65.96	48.82	57.04	42.62
Refined Prediction	75.52	73.67	66.72	64.84	57.57	54.22

Table 2: The benchmark is evaluated on car category only with IoU=0.5. Notice that we used a pre-trained model provided by [2] on a proportion of the whole KITTI detection testing data. Blue is the baseline provided by [2].

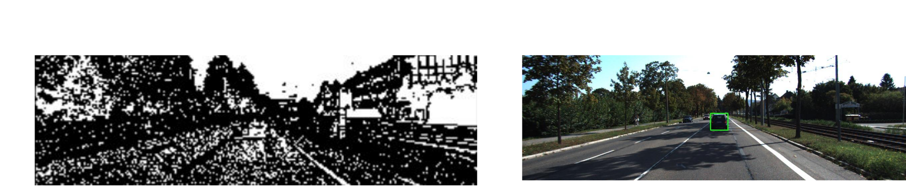
Conclusion:

- After bilateral filter refinement, the depth map presents much lower error variance and improved foreground mean accuracy.
- Edge-preserving upsampling effectively reduced the long tail at object boundaries.
- Refined pseudo-LiDAR presents a better 3D object detection result than raw PSMNET predicted pseudo points on the same detection model.

Experimentation



Confidence map: it can be used as weighting map to provide reasonable fusion with multi-modal data. We tried to trained the same PSMNet from scratch, using smooth L1 loss, minimizing the L1 distance with sparse gt confidence.But after 10+ hours' training and parameter tuning, the network was still not able to produce precise confidence map.



Reducing exhaustive search: Using canny filters, or 2D mask rcnn to give a potential edge appearing areas, and do DBSCAN only on those areas.

ICP registration: Find the transform between predicted pseudo points and real point cloud, and apply RanSAC to rule out those unmatched data.