Survey of Disparity Map Algorithms Intended for Real Time Stereoscopic Depth Estimation

Erick Blankenberg and Scott Blankenberg

Motivation

Motivation for Investigating Stereo Depth Estimation

- Cost effective depth estimation is critical for economical autonomous vehicles
- LIDAR is expensive, uses active sensing, and produces sparse measurements which are augmented with video streams for object identification
- Stereo imaging systems in conjunction with cheap LIDAR is significantly less expensive, passive, and produce dense depth estimates but require a computationally expensive disparity estimation step

State of the Art Disparity Algorithms

- Traditional disparity estimation algorithms use area or feature based comparisons to find matching pixels and require significant sacrifices to achieve real-time speed such as local and sparse approaches
- Recent efforts from Uber Research and others in convolutional neural networks and differentiable structures bring unprecedented accuracy and real-time evaluation on CUDA enabled hardware. In our work we evaluate two such algorithms alongside a traditional disparity estimation algorithm.

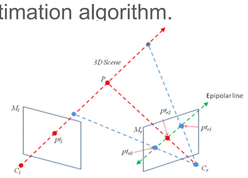


Figure 1: Image of Epipolar Line Structure

Related Work

Surveys

- Several exist for traditional algorithms.

Middlebury Database

- Databases of indoor scenes of increasing complexity
- API for testing and published list of ranked algorithms
- Benchmark used in several papers for comparison.

KITTI Vision Benchmark Suite

- Databases of outdoor scenes for scene-flow, stereo disparity, odometry, and object tracking.

Dataset

Middlebury Dataset 2001:

- 9 Images
- All are synthetic images made from intersecting textured planes.

Middlebury Dataset 2014:

- 25 Images
- Each is a complex indoor scene made from compositions of household objects.

Methods

- **Block Matching (Baseline method)**
  - Relatively simple and fast window based local matching method
  - Scans epipolar line and uses Winner Take All principle
  - Does not consider global context
    - Fails at discontinuities and plain regions
    - Susceptible to noise

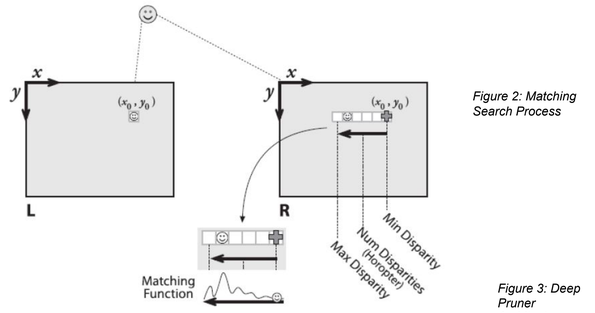


Figure 2: Matching Search Process

- **Deep Pruner**
  - Uses differentiable adaptation of PatchMatch algorithm to create a sparse representation of cost volume
  - Learns which range to prune for every pixel
  - Uses an image guided refinement module to improve performance

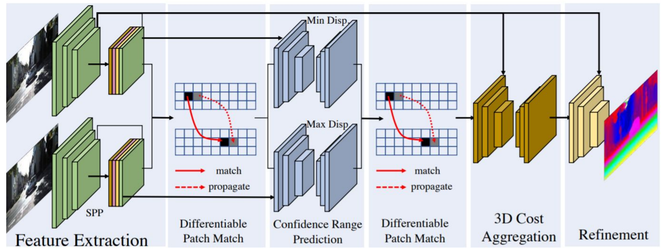


Figure 3: Deep Pruner Architecture

- **Pyramid Stereo Matching Network (PSM Net)**
  - Spatial pyramid pooling module for incorporating global image context
  - 3D convolutional neural network

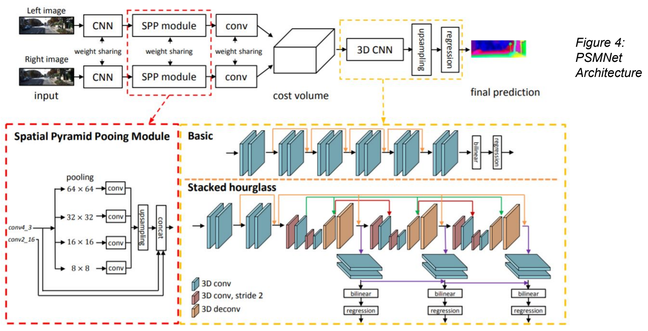
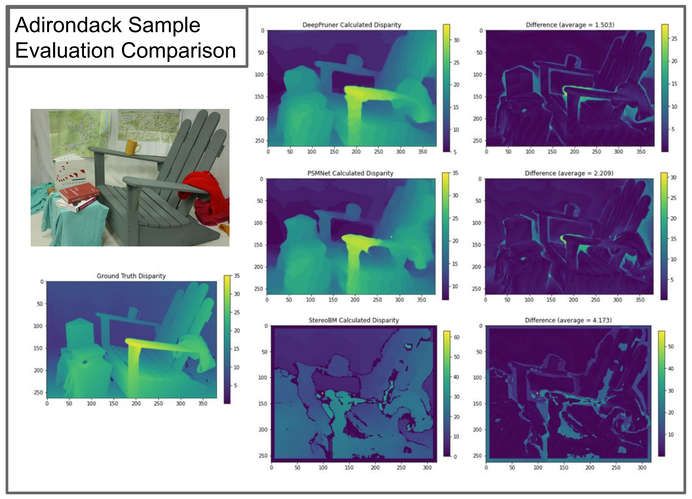


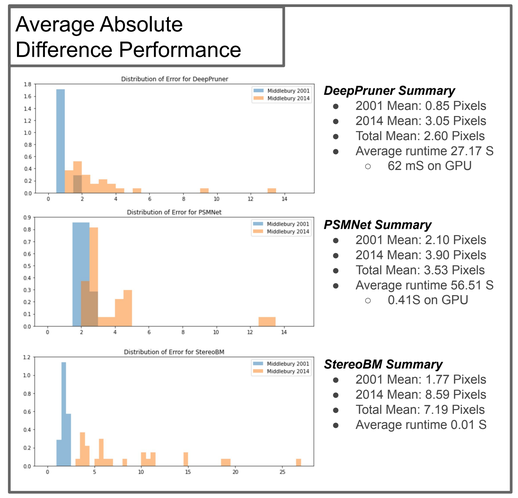
Figure 4: PSMNet Architecture

Experimental Results

Adirondack Sample Evaluation Comparison



Average Absolute Difference Performance



DeepPruner Summary

- 2001 Mean: 0.85 Pixels
- 2014 Mean: 3.05 Pixels
- Total Mean: 2.60 Pixels
- Average runtime 27.17 S
  - 62 mS on GPU

PSMNet Summary

- 2001 Mean: 2.10 Pixels
- 2014 Mean: 3.90 Pixels
- Total Mean: 3.53 Pixels
- Average runtime 56.51 S
  - 0.41S on GPU

StereoBM Summary

- 2001 Mean: 1.77 Pixels
- 2014 Mean: 8.59 Pixels
- Total Mean: 7.19 Pixels
- Average runtime 0.01 S

Conclusions

Quantitative

- Performance is good across the board on the 2001 dataset but complex scenes in the 2014 dataset are challenging
- State of the art deep-learning models are superior to the Stereo Block Matching algorithm provided in OpenCV.
- Runtimes are high but were run on a CPU (2.9 GHz Intel Core I5 in a 2015 Macbook)

Qualitative

- Deep-learning models tend to overpaint low disparity values on near objects over high disparity objects in the foreground creating a “halo” when the absolute difference is plotted. PSMNet does this more than DeepPruner.
- DeepPruner captures smooth gradients, PSMNet creates steps of flat patches over gradient sections.
- StereoBM generates flat estimates on gradients in all but a few 2001 images and oftentimes fails to produce an estimate (gives 0) for complex regions leading to high error.

Future Work

Database Scope:

- Extend survey to outdoor scenes from the KITTI database and use included occlusion maps.
- Include moderately complex 2006 Middlebury set.
- Include custom images.

More Algorithms

- Finish fuzzy logic disparity algorithm.
- Implement best scoring algorithm not based on deep-learning methods intended for real-time application from KITTI and Middlebury records.
- Implement deep-learning algorithms on GPUs to achieve speed comparable to claims of authors.

References

[1] Bodkin, Bryan Hale. "Real-Time Mobile Stereo Vision. " Master's Thesis, University of Tennessee, 2012.
[2] Gu, Tianyu. "Real time obstacle depth perception using stereo vision. " A thesis presented to the graduate school of the University of Florida for Master of Science in University of Florida, 2014.
[3] S. Duggal, S. Wang, W.-C. Ma, R. Hu, and R. Urtasun, "DeepPruner: Learning Efficient Stereo Matching via Differentiable PatchMatch," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Sep. 2019.
[4] J.-R. Chang and Y.-S. Chen, "Pyramid Stereo Matching Network," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Mar. 2018.