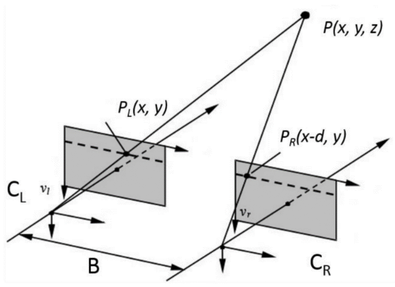
A PARALLEL CONVOLUTIONAL NEURAL NETWORKARCHITECTURE FOR STEREO VISION ESTIMATION

[1]Yao Chou, [1]Dah-Jye Lee, [2]Dong Zhang, [1]Karina Hill
Robotic Vision Lab,
[1]Department of Electrical and Computer Engineering
Brigham Young University, Utah, 84602, U.S.A.
[2]School of Electronics and Information Technology
Sun Yat-sen University, Guangzhou, Guangdong, China 510006

Abstract

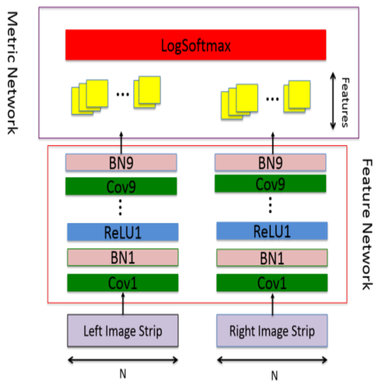
Extracting depth information from a stereo image pair is a commonly used method in 3-D computer vision. For robotics and unmanned vehicle applications that require real-time performance, speed is often more important than accuracy. In recent years, Convolutional Neural Networks (CNNs) have shown great success in many computer vision applications including classification, segmentation, object detection, edge detection, and stereo vision estimation. Existing network architectures for stereo vision estimation predict very little information during the forward pass and are only able to calculate the disparity for one pixel at a time. In this paper, we propose a parallel architecture to speed up disparity map computation by simultaneously processing all pixels on one horizontal line. We train and test our network on five Middlebury datasets. Our parallel architecture achieves at least a 10X speedup compared to existing networks. Its accuracy is also very competitive.

Disparity Calculation



$$z = \frac{fB}{d}$$

Parallel CNN Architecture

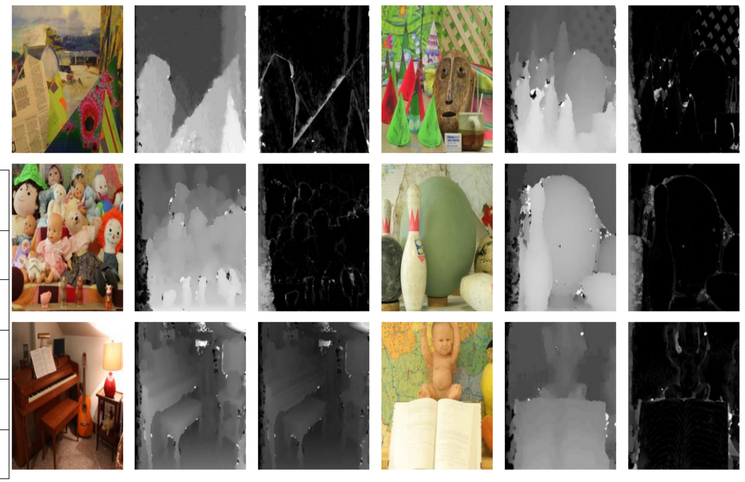


It is a joint learning network which consists of two sub-networks: a feature network and a metric network. The feature network maps two input images to feature representations, and the metric network calculates the disparities for all pixels.

Motion Classification and Depth Analysis Results

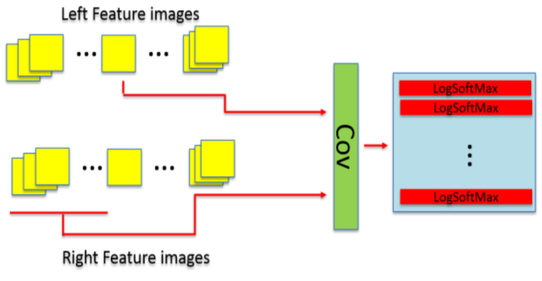
Our proposed parallel architecture generates an accurate disparity map in real-time (0.1 second using a graphics processing unit.

Middlebury	>2 pixels	>3 pixels	>4pixels	>5pixels
2001	84.87%	86.44%	87.11%	87.50%
2003	87.89%	90.00%	91.40%	92.26%
2005	82.83%	87.67%	90.22%	91.74%
2006	81.24%	82.24%	82.27%	82.83%
2014	73.04%	78.56%	83.10%	87.21%



Col 1 & 4: original left image, Col 2 & 5: stereo estimate, Col 3 & 6: disparity error.

Disparity Calculation Framework



We input two image strips into CNNs and simultaneously calculate the disparities of each pixel in the entire left strip, rather than matching a patch to the right strip. We use CNNs to extract features for both strips and then compute the matching cost between these two groups of features. Finally, we utilize a layer of parallel LogSoftMax functions to simultaneously estimate the disparities for all pixels in the strips. This parallel architecture significantly reduces the processing time and makes it suitable for real-time use.

Conclusion

In this paper, we propose a parallel CNN architecture for stereo vision estimation. Our results on the Middlebury datasets show that our network architecture produces accurate results. Unlike the existing CNN stereo vision architectures that either determine whether a pair of small image patches is a good match, or compute the disparity for one pixel in one forward pass time, our parallel architecture can estimate disparities for all pixels in the center row of an entire image strip in one single forward pass. Our architecture significantly speeds up the computation while provide very promising accuracy.