Using CNNs to Estimate Depth from Stereo Imagery

Tyler Jordan, Skanda Shridhar, Jayant Thatte
Department of Electrical Engineering, Stanford University

Motivation

• 3D TV / Free Viewpoint TV
• Virtual Reality / Head-mounted displays
• Augmented Reality
• Computer Vision
• Autonomous Vehicles

Convolutional Neural Network

Convolutional Neural Networks are interconnected layers of artificial neurons (perceptrons) that are trained to create a model for image classification. Each layer corresponds to a set of filters which are applied to the output of the previous layer ultimately resulting in a classification label. Our CNN is trained to calculate the similarity of pixels in the stereo imagery at various disparities.

Experimental Results

Major objects in the scenes like the road, signs, and cars are accurate in the disparity maps. The right and left edges are not as clean as the center of the image due to the lack of redundant data. The CNN approach performs far better than the naive plane-sweep approach.

References