Learning from scratch a confidence measure

Matteo Poggi
http://vision.disi.unibo.it/~mpoggi
Stefano Mattoccia
http://vision.disi.unibo.it/~smatt

University of Bologna
Department of Computer Science and Engineering (DISI)
Viale del Risorgimento 2
Bologna, Italy

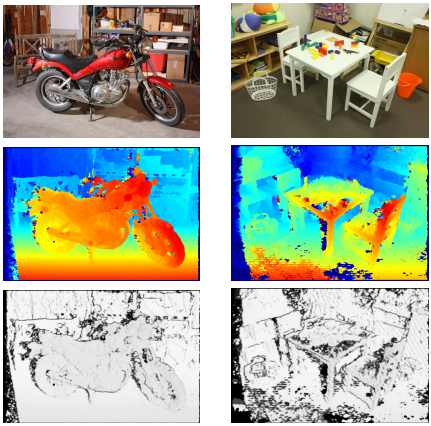

Figure 1: Reference image, disparity map and confidence computed by CCNN for *Motorcycle* and *Playtable* frames of the Middlebury 2014 training dataset.

In this paper, we propose a novel approach, referred to as Confidence Convolutional Neural Network (CCNN)[1], to predict the correctness of stereo matching by deploying a Convolutional Neural Network (CNN). In literature this is usually carried out by means of confidence measures [1] which encode the degree of reliability of the disparity assigned to each pixel by considering different cues: cost volume, reference image, disparity map and so on. Although some standalone measures are quite effective [1], recent works proved that combining a pool of them, within a machine learning framework, enables to significantly improve the overall effectiveness. In particular, Park & Yoon [2] represents state-of-the-art in this field, obtaining the best results according to the Area Under the Curve (AUC) evaluation protocol defined by Hu and Mordohai [1].

This paper proposes the first method that allows to obtain a confidence measure inferred from scratch by a CNN deploying as input cue only the disparity map computed by a stereo algorithm. This strategy makes our proposal suited even for out-of-the-box 3D sensors that typically do not provide the cues required by other methods.

Dataset/Alg.	Opt.	Park&Yoon	CCNN
KITTI/BM	0.137	0.179	**0.175**
KITTI/SGM	0.038	0.124	**0.099**
Middl./BM	0.093	0.114	**0.107**
Middl./SGM	0.042	0.093	**0.074**

Table 1: Average AUC on KITTI 2015 and Middlebury 2014 training datasets with BM and SGM algorithms. The lower, the better. Values closer to optimum are in bold.

For a fair comparison, we trained the proposed CCNN and Park & Yoon [2] on KITTI 2012 (more than 6 million samples), using the Block Matching stereo algorithm (BM). This provides more than 6 million samples for training. Then, we evaluated CCNN and Park & Yoon on KITTI 2015 training dataset processing the output of BM and Semi-Global Matching (SGM). We also cross-evaluated the two approaches, with BM and SGM stereo algorithms, on Middlebury 2014 training dataset. Table 1 reports average AUCs for CCNN and Park & Yoon, computed on KITTI 2015 and Middlebury 2014, for BM and SGM, in order to assess their effectiveness. Observing the table, we can notice that our proposal always outperforms state-of-the-art.

[1] The source code of CCNN and the trained network is available here: `http://vision.disi.unibo.it/~mpoggi`

[1] Xiaoyan Hu and Philippos Mordohai. A quantitative evaluation of confidence measures for stereo vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, pages 2121–2133, 2012.

[2] Min-Gyu Park and Kuk-Jin Yoon. Leveraging stereo matching with learning-based confidence measures. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.