Patch Based Confidence Prediction for Dense Disparity Map

Akihito Seki[1,2]
akihito.seki@toshiba.co.jp

Marc Pollefeys[2]
marc.pollefeys@inf.ethz.ch

[1]Corporate R&D Center
Toshiba Corporation, Japan

[2]Department of Computer Science
ETH Zürich, Switzerland

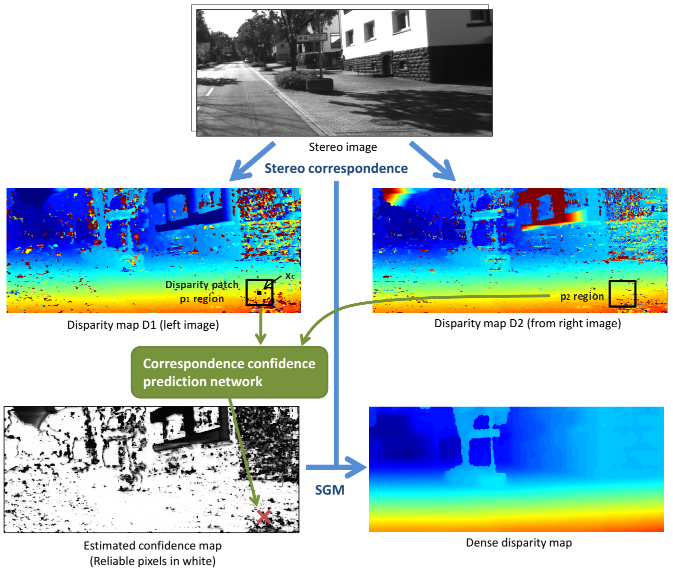

Figure 1: Overview of our method.

Confidence of stereo correspondences is useful information to improve quality of the disparity maps. Many features to predict the confidence have been proposed. Learning based confidence measure[6] combines these features and is able to outperform their individual usage. These features and classifiers are carefully designed, however beneficial information might be undescribed or their representation might be too redundant.

As shown in Fig.1, we propose a novel confidence prediction method to overcome the problem. We design a disparity patch which takes into account the ideas of conventional confidence features. The patches are used as inputs of a Convolutional Neural Network(CNN) so that the discriminative features and classifier are simultaneously trained. In order to handle trade-off between accuracy and computation time, we propose three types of network structures and their input patches. Moreover, the confidence is incorporated into Semi Global Matching (SGM) [4] to acquire dense disparity map. SGM is widely used for dense disparity estimation due to its high accuracy while keeping low computation cost. In the following, we will briefly explain both methods and experimental results.

Confidence estimation with a CNN: We leverage the disparity patch and introduce the knowledge of the conventional features. The patch consists in a two channels. 1st channel is coming from an idea that neighboring pixels on a disparity map D_1 which have consistent disparities are more likely to be correct matching[7]. In 2nd channel, a disparity D_2 from another image is considered such that the matches from left to right image should be consistent with those from right to left[1]. We employ a shallow CNN for the sake of reducing potential computation cost of the network, however, the network is still slow computation because the output of the network for each pixel has to be computed from scratch. We also propose speed-up networks by modifying preprocessing of the patches and network structure.

Confidence fusion with SGM: SGM has two parameters in order to control discontinuities of disparity map. We assume the discontinuities are likely to have the large magnitude of the image gradient using the same assumption as the original SGM, but not all large gradient pixels correspond to them. We consider the pixels with high confidence should be trusted and are able to be discontinuities easily. Hence, penalties at the high confidence pixel are designed to be decreased.

Figure 2 and Table 1 show evaluation results based on sparsification curve and its area under curve (AUC) value. Better confidence prediction methods have AUC values that are closer to the optimal curve: It means the method removes incorrect correspondence pixels while keeping the correct ones. Our methods outperform state of the art method[6] on both accuracy and computation time.

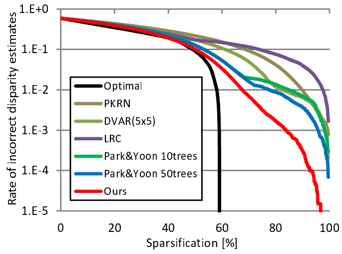

Figure 2: Sparsification plots on 123rd frame on KITTI 2012.

Method		AUC [$\times 10^{-2}$]	Time [sec.]
Optimal		3.95	-
Ours		**4.20**	28.5
	fast	4.50	**0.3**
	hybrid	4.48	0.5
Park& Yoon[6]	50 trees	4.70	2.2
	10 trees	5.15	0.4

Table 1: Comparison of overall AUC value and computation time.

Rank	Method	Error	Runtime[sec.]
1	Ours with MC-CNN-acrt	**2.36%**	68*
2	Displets v2[3]	2.37%	265
3	VDS(anonymous)	2.42%	68*
4	MC-CNN-acrt[8]	2.43%	67*

Table 2: Out-Noc error on KITTI 2012 testing dataset by May 1st 2016. "*" at Runtime means GPU computation.

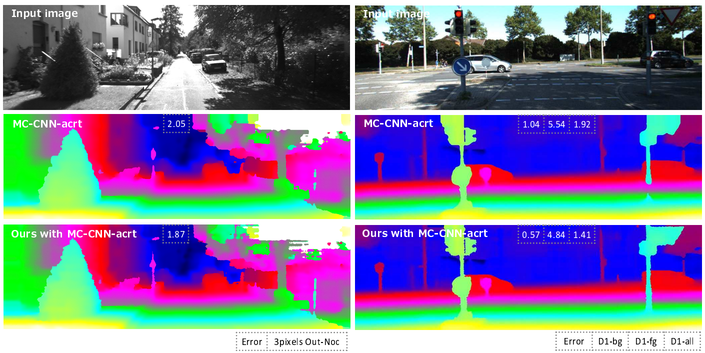

Figure 3: Example results of original MC-CNN-acrt and our fusion method with MC-CNN-acrt on KITTI 2012(left) and 2015(right).

Table 2 and Figure 3(left) show the accuracy of dense disparity map on KITTI 2012[2] testing dataset . We got the best accuracy when MC-CNN-acrt[8] was employed as a similarity measure. On KITTI 2015[5], we got the second best without the need for a strong foreground shape prior[3]. For details, please refer to the main paper.

[1] P. Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. *MVA*, 6:35–49, 1993.

[2] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the KITTI vision benchmark suite. In *Proc. CVPR*, 2012.

[3] F. Guney and A. Geiger. Displets: Resolving stereo ambiguities using object knowledge. In *Proc. CVPR*, 2015.

[4] H. Hirschmuller. Stereo processing by semiglobal matching and mutual information. *Trans. on PAMI*, 30(2):328–341, 2008.

[5] M. Menze and A. Geiger. Object scene flow for autonomous vehicles. In *Proc. CVPR*, 2015.

[6] M.G. Park and K.J. Yoon. Leveraging stereo matching with learning-based confidence measures. In *Proc. CVPR*, 2015.

[7] A. Spyropoulos, N. Komodakis, and P. Mordohai. Learning to detect ground control points for improving the accuracy of stereo matching. In *Proc. CVPR*, 2014.

[8] J. Zbontar and Y. LeCun. Computing the stereo matching cost with a convolutional neural network. In *Proc. CVPR*, 2015.