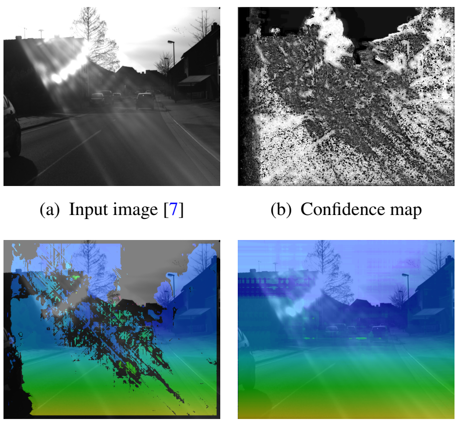
Leveraging Stereo Matching with Learning-based Confidence Measures

Min-Gyu Park and Kuk-Jin Yoon
Computer Vision Laboratory, School of Information and Communications, Gwangju Institute of Science and Technology, South Korea

Stereo matching has long been an important topic in computer vision, and its difficulties are thoroughly examined in the literature. Based on those researches, several papers address the feasibility of detecting mismatched pixels [6] not only to improve the quality of disparity maps [1] but also to leverage mid-level scene representation [8]. Moreover, the problem of detecting mismatched pixels becomes more important as the degree of ill-conditioning increases because the current solutions usually fail to find correct answers as described in Fig. 1(c). In this paper, we consider how far a learning-based confidence prediction approach can leverage stereo matching for the practical use in general outdoor environments.



(a) Input image [7] (b) Confidence map

(c) Semi-global Matching [4] (d) Proposed

Figure 1: Stereo matching results in a challenging environment. (b) shows the predicted confidence map. (c) and (d) show estimated disparity maps overlaid on the input image.

First, we analyze the characteristics of various confidence measures in the regression forest framework in order to select most effective confidence measures among various measures. Then, we train the regression forest once again to predict the confidence of a pixel by using selected measures. In the testing step, we predict the confidence of a pixel using an ensemble of tree outputs. Afterwards, we manipulate the predicted confidence value as

$$\hat{Q}(\mathbf{p}) = Q(\mathbf{p})\overline{Q} + (1 - Q(\mathbf{p}))\underline{Q}, \tag{1}$$

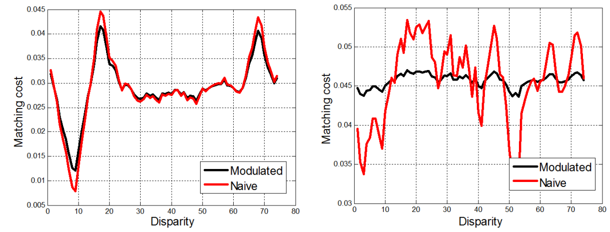
where $\underline{Q}$ and $\overline{Q}$ indicate lower and upper bounds for confidence values, and $Q(\mathbf{p})$ represents the predicted confidence value for a pixel $\mathbf{p}$. Here, the upper and lower bounds are computed in the training step to take the quality of trained forests or classifiers into account.

Second, we incorporate confidence information into stereo matching algorithms by employing it in order to modulate the initial matching cost,

$$\hat{C}(\mathbf{p}, d) = \hat{Q}(\mathbf{p})C(\mathbf{p}, d) + (1 - \hat{Q}(\mathbf{p})) \sum_{k \in \mathcal{D}} \frac{C(\mathbf{p}, k)}{|\mathcal{D}|}, \tag{2}$$

where $C(\mathbf{p}, d)$ is the per-pixel matching cost of $\mathbf{p}$ for a disparity value d and $\mathcal{D}$ is the set of possible disparity values. The latter term is the mean of the matching costs multiplied by $(1 - \hat{Q}(\mathbf{p}))$. As the probability of correctness increases, the modulated cost increasingly depends on the original matching costs. Otherwise, the modulated costs becomes a mean value. Figure 2 shows the result of the matching cost modulation, in which matching costs of confident pixels appear similar to the initial matching costs whereas unconfident pixels are flattened inversely proportional to the confidence value. Therefore, disparity values of unreliable pixels can be easily dominated by the neighboring pixels.

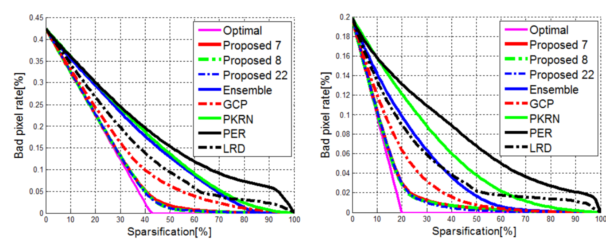
This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.



(a) (b)

Figure 2: Modulated matching costs. Matching costs of unreliable pixels in (b), which are not likely to give correct solutions, are flattened depending on the predicted confidence values whereas confident pixels (a) have similar costs to its original matching costs.

For the experiment, we evaluated error detection performance in terms of the sparsification curve and the improvement of stereo matching algorithms in terms of the bad pixel rate. Two sparsification curves are drawn in Fig. 3 that confirm the superior performance of the proposed method in detecting unreliable pixels. We adopted the proposed cost modulation scheme to semi-global matching [4] and fast cost volume filtering [5] algorithms. The proposed method reduced bad pixel rates 1.22% for the KITTI dataset [2] and 0.67% for the Middlebury dataset [9] in average. Furthermore, we observed a significant improvement for challenging outdoor datasets [7]. Because, the challenging dataset contains a large number of unreliable pixels that violate underlying assumptions of binocular stereo matching. It is worth noting again, the detection of unreliable pixels is necessary for these kinds of input images. We suggest an effective way to improve existing algorithms in these environments through the matching cost modulation scheme. A detailed explanation is given in the paper and the supplementary material.



(a) Sparsification on KITTI Frame 30 (b) Sparsification on KITTI Frame 50

Figure 3: Comparison of sparsification curves for various confidence measures including learning based approaches [1, 3] and individual confidence measures [3, 6].

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

[2] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. Vision meets robotics: The kitti dataset. *International Journal of Robotics Research (IJRR)*, 2013.

[3] R. Haeusler, R. Nair, and D. Kondermann. Ensemble learning for confidence measures in stereo vision. In *CVPR*, pages 305–312, 2013.

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

[5] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, and M. Gelautz. Fast cost-volume filtering for visual correspondence and beyond. *PAMI*, 35(2):504 – 511, 2013.

[6] X. Hu and P. Mordohai. A quantitative evaluation of confidence measures for stereo vision. *PAMI*, 34(11):2121–2133, 2012.

[7] S. Meister, B. Jähne, and D. Kondermann. Outdoor stereo camera system for the generation of real-world benchmark data sets. *Optical Engineering*, 51(02): 021107–1–021107–6, 2012.

[8] D. Pfeiffer, S. Gehrig, and N. Schneider. Exploiting the power of stereo confidences. In *CVPR*, pages 297–304, 2013.

[9] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47:7–42, 2002.