DEEP STEREO CONFIDENCE PREDICTION FOR DEPTH ESTIMATION

Sunok Kim[†] *Dongbo Min*[‡] *Bumsub Ham*[†] *Seungryong Kim*[†] *Kwanghoon Sohn*[†]

[†]School of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea
[‡]Department of Computer Science and Engineering, Chungnam National University, Daejeon, Korea
E-mail: khsohn@yonsei.ac.kr

ABSTRACT

We present a novel method that predicts a confidence to improve the accuracy of an estimated depth map in stereo matching. In contrast to existing learning based approaches relying on hand-crafted confidence features, we cast this problem into a convolutional neural network, learned using both a matching cost volume and its associated disparity map. As the size of the matching cost volume varies depending on a search range of stereo image pairs, we propose to use a top-K matching probability volume layer so that an input size for convolutional layers remains unchanged. Experimental results demonstrate that the proposed method outperforms the state-of-the-art confidence estimation approaches on various benchmarks.

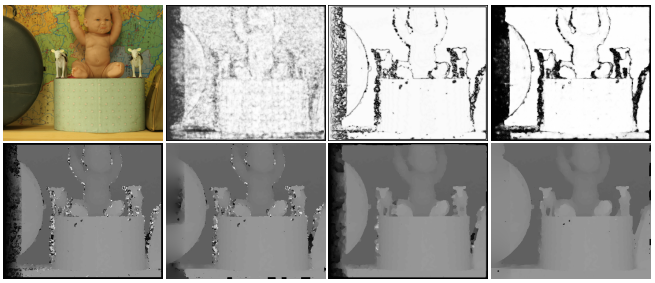
Index Terms— confidence prediction, stereo matching, depth refinement, convolutional neural networks, matching probability

1. INTRODUCTION

As one of the most important topics in computer vision, stereo correspondence has been actively studied over the last few decades. The disparity map (or depth map) obtained using stereo correspondence is widely used in many applications such as 3D reconstruction [1], object detection [2], and driver assistance system [3].

Though numerous approaches have been proposed to provide highly accurate disparity maps [4, 5, 6, 7, 8], even state-of-the-art methods still show a limited performance due to inherent difficulties for that task, such as occlusion, saturation, specularity, and textureless region [6]. To overcome these limitations, most stereo matching methods count on a disparity refinement step, where they first find mismatched pixels in an estimated disparity map and then refine the disparity map with reliable pixels [9, 10, 11, 12, 13, 14, 15]. In this step, predicting the confidence of an estimated disparity is one of the most important issues. Conventionally, the left-right consistency check or peak ratio [11] have been commonly used as an input feature for estimating the confidence map. However, the single confidence feature cannot reliably estimate the confidence across various scenes [10].

Recently, learning based methods using multiple confidence features have been developed [12, 13, 14, 15], showing a substantial accuracy gain over existing simple approaches. They first define a set of confidence features and then train a classifier, e.g., random forest [16], to predict a confidence map. However, all of these techniques utilize hand-crafted confidence features followed by simple classifiers. Since the performance of hand-crafted confidence features is sensitive to matching cost functions or datasets, it is difficult to select optimal confidence features that consistently guarantee high performance [15]. To overcome these problems, one method attempts to learn both confidence feature and classifier simultaneously through convolutional neural networks (CNNs) [17]. It, however, encodes confidence features using left and right disparity maps only, and possibly miss useful cues that can be learned from the matching cost volume computed from stereo matching methods.



(a) (b) (c) (d)

Fig. 1. Importance of joint learning using the matching cost volume and disparity map: (a) a left color image and initial disparity map estimated by MC-CNN [3], confidence maps and refined disparity maps estimated using (b) matching cost volume only, (c) disparity map only, and (d) both matching cost volume and disparity map. By using matching cost volume and disparity map simultaneously, our method provides a highly reliable performance.

In this paper, we introduce a novel CNNs architecture, where both matching cost volume and its corresponding disparity map are used to encode more discriminative confidence features. We design a joint network architecture that uses two different types of data as inputs. Specifically, we formulate three sub-networks including matching cost feature extractor, disparity feature extractor, and fusion network. To deal with a varying size of matching cost volume accoding to stereo pairs, we propose a top-K matching probability layer, where the matching cost volume is normalized and projected into a fixed-length of matching cost volume space. Fig. 1 shows the outstanding performance of the proposed confidence estimation that uses both the matching cost volume and disparity map. Experimental results further demonstrate that our method outperforms existing hand-crafted confidence measures [15] as well as CNNs-based confidence estimation method [17] on various benchmarks.

2. PROPOSED METHOD

2.1. Problem Statement and Model Architecture

Let us define stereo image pairs $\left(I_i^L, I_i^R\right)$ for pixel $i = [i_{\mathrm{x}}, i_{\mathrm{y}}]^T$. The objective of stereo matching is to estimate a disparity for each pixel i. Since any stereo matching methods cannot provide a fully reliable disparity map, we aim at estimating the confidence of the estimated disparity map in a learning framework and refining them. Using the existing stereo matching approaches, we build matching

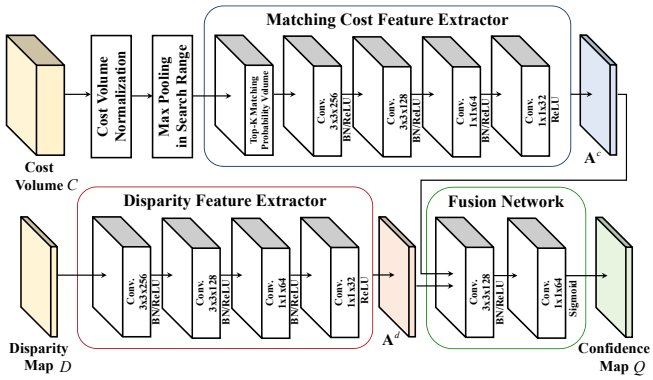


Fig. 2. The network architecture of proposed learning framework.

cost volume $C_{i,d}$ across disparity candidates $d \in \{1, ..., s\}$, where s is a disparity search range, and estimate its associated disparity $D_i = \text{argmin}_d C_{i,d}$. A ground truth confidence Q_i^* is defined as 1 when the absolute difference between the estimated disparity D_i and the ground truth disparity map D_i^* is less than a fixed value ϵ, and otherwise 0. By learning the relationship between matching cost volume $C_{i,d}$ with its associated disparity D_i and a ground truth confidence Q_i^*, the confidence of D_i can be then estimated such that $Q_i \in [0, 1]$.

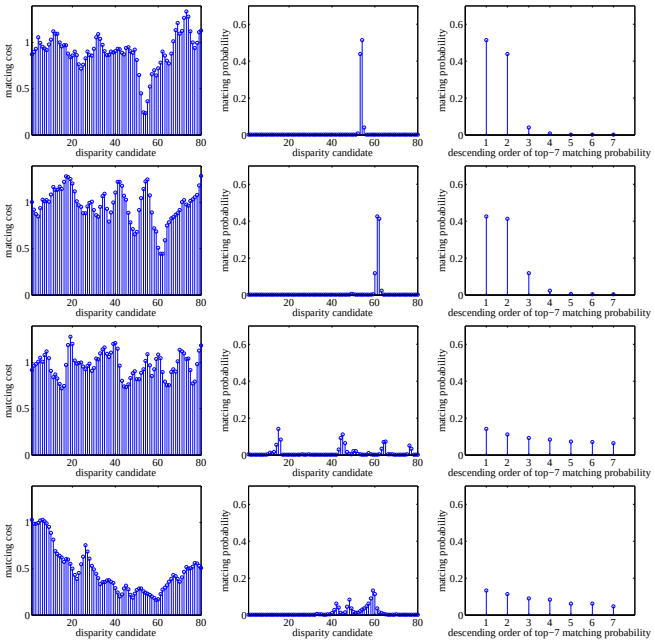
To this end, we design a network architecture as in Fig. 2 that estimates the confidence by exploiting matching cost volume and its associated disparity map as inputs. The overall network consists of three sub-networks: a matching cost feature extractor, a disparity feature extractor, and a fusion network. The matching cost and disparity feature extractor networks are modeled by feed-forward processes such that $\mathbf{A}^c = \mathcal{F}_{\mathbf{W}^c}(C)$ and $\mathbf{A}^d = \mathcal{F}_{\mathbf{W}^d}(D)$ respectively, where $\mathbf{W}^c$ and $\mathbf{W}^d$ are each network parameter, and $\mathbf{A}^c$ and $\mathbf{A}^d$ are intermediate features. Note that the size of matching cost volume varies depending on the search range of stereo image pairs. To deal with the search range variation of matching cost volume, we propose a top-K matching probability volume layer which enables the search range-invariant convolutions while improving the feature extraction performance. We adopt the top-K matching probability volume P instead of directly using the matching cost volume C. The two activations of the matching cost and disparity feature extractor networks are concatenated and used as an input of the fusion network to predict the per-pixel confidence map $Q = \mathcal{F}_{\mathbf{W}^f}(\mathbf{A}^c, \mathbf{A}^d)$ where $\mathbf{W}^f$ is the fusion network parameter.

2.2. Learning Confidence Estimator via CNNs

Our CNN architecture design is inspired by two intuitions that: 1) useful cues for confidence prediction can be extracted from both matching cost and an initial disparity map, and 2) top-K matching probability volume can solve the search range variation problem of the matching cost volume while improving the discriminability power.

2.2.1. Network design

At the heart of our strategy to boost the performance is the fusion strategy of two inputs consisting of matching cost volume and initial disparity map. Due to their heterogeneous attributes, a direct concatenation of two raw inputs does not provide an optimal performance. Alternatively, we may simply fuse intermediate outputs obtained from individual sub-networks for two raw inputs, but the prediction output cannot be fused optimally since the contribution of each input might vary for each pixel. Therefore, inspired by [18], we



(a) (b) (c)

Fig. 3. Effectiveness of top-K matching probability volume: (a) raw matching cost of MC-CNN [3] for KITTI dataset, (b) matching probability volume using Eq. (1), and (c) descending order of top-7 matching probability using Eq. (2). First and second row represent reliable pixels, and third and fourth row represent unreliable pixels.

build the CNN architecture with one fusion network $\mathcal{F}_{\mathbf{W}^f}(\mathbf{A}^c, \mathbf{A}^d)$ as well as two individual sub-networks $\mathcal{F}_{\mathbf{W}^c}(C)$ and $\mathcal{F}_{\mathbf{W}^d}(D)$. We also consider the search range variation of the matching cost volume C by adding top-K matching probability layer, which will be described in the Sec. 2.2.2.

$\mathcal{F}_{\mathbf{W}^c}(C)$ and $\mathcal{F}_{\mathbf{W}^d}(D)$ consist of 4 convolutional layers, followed by batch normalization (BN) and rectified linear units (ReLUs) except for the last convolutional layer. In $\mathcal{F}_{\mathbf{W}^f}(\mathbf{A}^c, \mathbf{A}^d)$, the inputs are concatenated intermediate activations, $\mathbf{A}^c$ and $\mathbf{A}^d$. The fusion network consists of 2 convolutional layers, followed by BN and ReLUs. The sigmoid function is used for the last activation to train the binary classifier. With the output of fusion network $Q = \mathcal{F}_{\mathbf{W}^f}(\mathbf{A}^c, \mathbf{A}^d)$, we learn all the network parameters by minimizing the cross entropy loss:

$$\mathcal{L} = -\sum_i [Q_i^* \log Q_i + (1 - Q_i^*) \log(1 - Q_i)]. \quad (1)$$

2.2.2. Top-K matching probability layer

The matching cost volume C, computed using any existing stereo matching methods [8, 3], can be used as an input for the matching cost feature extractor network $\mathcal{F}_{\mathbf{W}^c}(C)$. However, the range of matching cost values may vary depending on the stereo matching setup, e.g., raw matching cost computation, cost aggregation, and optimization steps. We address this issue by normalizing the matching cost volume $C_{i,d}$ at a pixel i and a disparity d as follows:

$$P_{i,d} = \frac{\exp(-C_{i,d}/\sigma_F)}{\sum_l \exp(-C_{i,l}/\sigma_F)}, \quad (2)$$

where $P_{i,d}$ represents a matching probability. σ_F is a parameter to adjust a flatness of the matching cost and $l \in \{1, ..., s\}$ where s is a

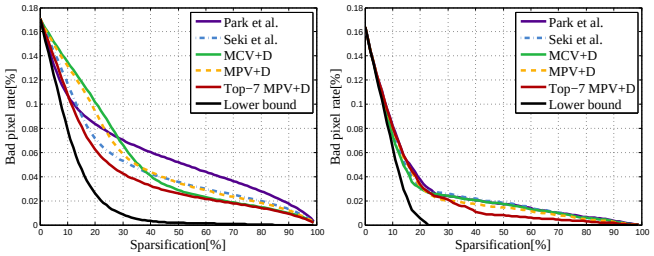


Fig. 4. Average sparsification curve on (a) KITTI (training : MID 2005) and (b) MID (training : KITTI 2012).

search range. Compared to softmax loss, our normalization scheme can adjust the flatness of matching cost volumes. As exemplified in Fig. 3(a), the absolute scale of the matching cost $C_{i,d}$ varies, and this problem can be alleviated in the matching probability volume $P_{i,d}$ as in Fig. 3(b).

In Fig. 3(b), we can also find that most matching cost probability values are close to 0, which do not help to convey useful cues. Such redundant parts rather distract the performance of confidence prediction. More seriously, the size of matching cost volume can vary according to the disparity search range s of stereo image pairs. Thus, we also propose to use top-K matching probability layer, where the matching probability P of Eq. (2) is projected into a fixed length input P^K, where $P^K_{i,k}$ for $k = \{1, 2, ..., K\}$ represents the k-th maximum values among $P_{i,d}$ for $d = \{1, ..., s\}$. Note that this layer does not have learnable parameters, but it is differentiable, thus enabling the network to learn regardless of search range of the matching cost volume. In Fig. 3(c), the top-K matching probability contains enough information to classify reliable and unreliable pixels.

2.3. Depth Refinement with GCPs-Based Propagation

The predicted confidence map can be used to refine a disparity map through ground control points (GCPs) based propagation scheme. We first set pixels that have higher confidence values than the threshold δ to GCPs. With the set of GCPs, we obtain a final disparity map $\widehat{D}$ by minimizing the following objective function defined on the Markov Random Field (MRF) framework [19]:

$$\sum_i (h_i(\widehat{D}_i - D_i)^2 + \lambda \sum_{j \in \mathcal{N}_i} \omega_{i,j}(\widehat{D}_i - \widehat{D}_j)^2), \quad (3)$$

where h_i is a binary mask to mark the position of GCPs, *i.e.*, it is 1 for GCPs and 0 otherwise, and the weight $\omega_{i,j}$ between pixels i and j is defined as follows:

$$\omega_{i,j} = \frac{h_i k_{i,j} \exp(-\|D_i - D_j\|^2/\sigma_D)}{\sum_{j \in \mathcal{N}_i} h_i k_{i,j} \exp(-\|D_i - D_j\|^2/\sigma_D)}, \quad (4)$$

where $k_{i,j}$ is the bilateral affinity between the color values of the pixels i and j, and their spatial locations. σ_D is the standard deviation of the Gaussian function, and $\mathcal{N}_i$ represents a local 4-neighborhood for pixel i.

By minimizing Eq. (4) with respect to $\widehat{D}$, the final disparity map $\widehat{D}$ can be estimated such that

$$(\mathbf{H} + \lambda \mathbf{L})\,\widehat{\mathbf{D}} = \mathbf{H}\mathbf{D} \quad (5)$$

where $\mathbf{D}$ and $\widehat{\mathbf{D}}$ are the vector forms of the estimated initial disparity map D and output disparity map $\widehat{D}$ and $\mathbf{L}$ is the sparse Laplacian matrix having the element $-\omega_{i,j}$ for $i \neq j$ and $\sum_{j \in \mathcal{N}_i} \omega_{i,j}$ otherwise.

Table 1. The average AUC values$\times$100 for KITTI [20], MID [21], and MPI [22] dataset. The AUC value of ground truth confidence is measured as 'lower bound'. The result with the lowest AUC value in each experiment is highlighted.

Training	KITTI 2012			MID 2005		
Testing	KIT.	MID	MPI	KIT.	MID	MPI
Park *et al.* [15]	2.59	4.27	2.43	4.04	4.11	2.82
Seki *et al.* [17]	2.54	4.18	2.26	2.77	4.13	2.56
MCV+D	2.59	4.02	-	3.14	4.25	-
MPV+D	2.56	4.00	-	3.37	3.98	-
Top-7 MPV+D	**2.45**	**3.96**	**2.09**	**2.69**	**3.94**	**2.09**
Lower bound	1.67	3.33	1.55	1.67	3.33	1.55

Table 2. The average BMP for KITTI [20], MID [21], and MPI [22] dataset. The BMP of refined disparity map using ground truth confidence map is measured as 'lower bound'. The result with the lowest BMP in each experiment is highlighted.

Training	KITTI 2012			MID 2005		
Testing	KIT.	MID	MPI	KIT.	MID	MPI
Initial disparity	18.09	26.31	23.43	18.09	26.31	23.43
Park *et al.* [15]	12.67	20.91	16.43	13.01	21.66	21.78
Seki *et al.* [17]	11.08	18.82	15.42	11.23	20.91	19.52
MCV+D	12.40	24.06	-	12.73	18.82	-
MPV+D	11.25	18.06	-	12.56	15.66	-
Top-7 MPV+D	**10.22**	**15.53**	**12.52**	**10.58**	**12.19**	**12.83**
Lower bound	9.07	6.53	5.59	9.07	6.53	5.59

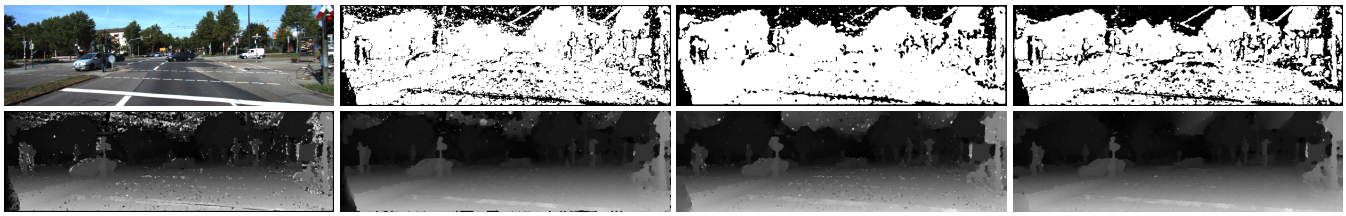
3. EXPERIMENTAL RESULTS

In order to evaluate the performance of the proposed method, we compared our method with the state-of-the-art methods including hand-crafted confidence measure, Park *et al.* [15], and CNNs-based confidence measure, Seki *et al.* [17], on various dataset [21, 22, 20]. We trained the network parameters using 8 frames in KITTI 2012 [23] as used in [17] and 5 images in Middlebury 2005 (MID 2005) [21] and tested on 21 images in Middlebury 2006 (MID), 23 images in MPI-Sintel (MPI), and 200 frames in KITTI 2015 (KITTI). The matching cost volume was computed with the MC-CNN [3] (KITTI 2015 fast network provided by [24]). The disparity search range for KITTI [20] and MID [21] is 80 and for MPI [22], it varies for each image. The threshold ϵ for the ground-truth confidence map is 3 for KITTI and MPI and 1 for MID following [14, 15]. For constructing the matching probability volume, we set σ_F as 0.05 for KITTI and 0.15 for MID. For GCPs-based propagation, we set σ_D and δ to 10 and 0.7 using cross-validation, respectively. We empirically set K to 7. For training CNNs, we used the MatConvNet library [25] and set the number of training patches to 150K. The learned parameters are fixed for all the experiments.

Furthermore, we evaluated component-wise contributions of the proposed method by comparing the network trained with matching *cost* volume and disparity map (MCV+D) and with matching *probability* volume and disparity map (MPV+D). Note that the proposed method is the network trained with top-7 matching probability volume and disparity map (Top-7 MPV+D).

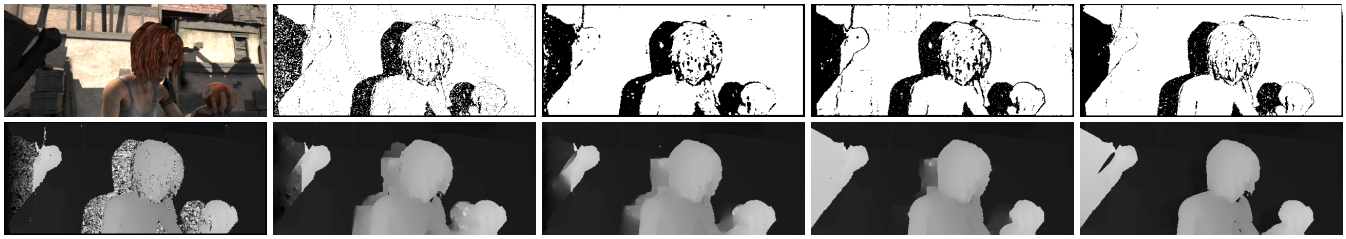
3.1. Confidence Measure Analysis

To evaluate the robustness of the confidence estimation, we used the sparsification curve and its area under curve (AUC) as used in



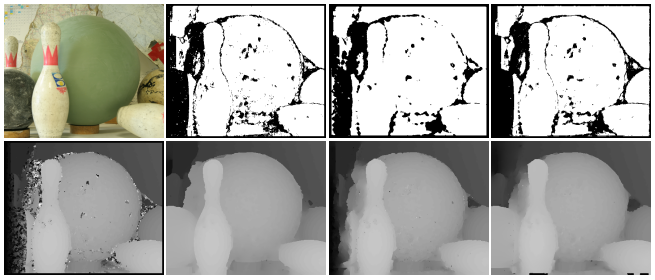
(a) (b) (c) (d)

Fig. 5. Qualitative evaluation on the KITTI dataset [20]: (a) input left color image and initial disparity map using MC-CNN [3], confidence maps and refined disparity maps estimated by (b) Park *et al.* [15], (c) Seki *et al.* [17], and (d) the proposed method (Top-7 MPV+D).



(a) (b) (c) (d) (e)

Fig. 6. Qualitative evaluation on the MPI dataset [22]: (a) input left color image and initial disparity map using MC-CNN [3], confidence maps and refined disparity maps estimated by (b) Park *et al.* [15], (c) Seki *et al.* [17], (d) the proposed method (Top-7 MPV+D), and (e) the ground truth confidence map.



(a) (b) (c) (d)

Fig. 7. Qualitative evaluation on the MID dataset [21]: (a) input left color image and initial disparity map using MC-CNN [3], confidence maps and refined disparity maps estimated by (b) Park *et al.* [15], (c) Seki *et al.* [17], and (d) the proposed method (Top-7 MPV+D).

[14, 15, 26]. The sparsification curve draws a bad pixel ratio while successively removing the pixels in descending order of confidence value in the disparity map. The lower AUC value means the higher accuracy of the confidence map. The lower bound of sparsification curve is obtained with a ground truth confidence map Q^*. Average sparsification curves for KITTI and MID datasets are shown in Fig. 4. The results show that the proposed confidence estimator (Top-7 MPV+D) exhibits a better performance than conventional methods [15, 17] and the trained network using raw matching cost volume (MCV+D) and matching probability volume (MPV+D). The average AUC values for KITTI, MID, and MPI are summarized in Table 1. It shows that the proposed method (Top-7 MPV+D) always achieves the lowest AUC values, demonstrating its outstanding performance in predicting mismatched pixels of the disparity map.

3.2. Stereo Matching Analysis

To verify the effectiveness of the confidence measures, we refined the disparity map through the GCPs-based propagation using the confidence map estimated by several approaches. To evaluate the quantitative performance, we measured an average bad matching percentage (BMP) [21] for the MID [21], MPI [22], and KITTI [20] datasets. Table 2 shows the BMP at MID [21], MPI [22], and KITTI [20] datasets. For MID, we computed the BMP at non-occluded pixels only. Since the KITTI benchmark provides a sparse ground truth disparity map, we evaluated the BMP only at sparse pixels with the ground truth disparity values. The results of extensive experiments show that the proposed method achieves the lowest BMP. Qualitative evalutaions for the KITTI [20], MPI [22], and MID [21] datasets are provided in Figs. 5-7, respectively. As expected, the refined disparity map with our confidence map shows better quality.

4. CONCLUSION

In this study, we have presented a learning framework for estimating stereo matching confidence by using both matching cost volume and initial disparity map in CNNs. It is assumed that the optimal confidence features can be learned from the matching probability volume together with the initial disparity map. With the depth refinement method using the proposed confidence estimation method, we obtained an accurate and robust disparity map for public datasets as well as for challenging outdoor environments. Though the confidence estimation is based on the CNN architecture, the depth refinement step still relies on the hand-crafted approach. As future work, we will study a learning-based approach that refines a depth map in a deep convolutional neural network framework.

5. ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (NRF-2016R1A2A2A05921659).

6. REFERENCES

[1] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, and P. Sayd, "Real time localization and 3d reconstruction," *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 363–370, Jun. 2006.

[2] S. Gupta, R. Girshick, P. Arbelaez, and J. Malik, "Learning rich features from rgb-d images for object detection and segmentation," *in Proc. Eur. Conf. Comput. Vis.*, pp. 345–360, Sep. 2014.

[3] J. Zbontar and Y. Lecun, "Computing the stereo matching cost with a convolutional neural network," *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1592–1599, Jun. 2015.

[4] K. Yoon and I. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 650–656, 2006.

[5] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, 2008.

[6] D. Min and K. Sohn, "Cost aggregation and occlusion handling with wls in stereo matching," *IEEE Trans. Image Process.*, vol. 17, no. 8, pp. 1431–1442, 2008.

[7] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 3017–3024, Jun. 2011.

[8] S. Kim, B. Ham, B. Kim, and K. Sohn, "Mahalanobis distance cross-correlation for illumination invariant stereo matching," *IEEE Trans. Circ. Syst. Vid. Techn.*, vol. 24, no. 11, pp. 1844–1859, 2014.

[9] Q. Yang, L. Wang, R. Yang, H. Stewenius, and D. Nister, "Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 3, pp. 492–504, 2009.

[10] P. Mordohai, "The self-aware matching measure for stereo," *in Proc. IEEE Int. Conf. Comput. Vis.*, pp. 1841–1848, Sep. 2009.

[11] X. Hu and P. Mordohai, "A quantitative evaluation of confdience measures for stereo vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2121–2133, 2012.

[12] C. Varekamp, K. Hinnen, and W. Simons, "Detection and correction of disparity estimation errors via supervised learning," *in Proc. IEEE Int. Conf. 3D Imaging*, pp. 1–7, Dec. 2013.

[13] R. Haeusler, R. Nair, and D. Kondermann, "Ensemble learning for confidence measrues in stereo vision," *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 305–312, Jun. 2013.

[14] A. Spyropoulos, N. Komodakis, and P. Mordohai, "Learning to detect ground control points for improving the accuracy of stereo matching," *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1621–1628, Jun. 2014.

[15] M. Park and K. Yoon, "Leveraging stereo matching with learning-based confidence measures," *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 101–109, Jun. 2015.

[16] L. B. Statistics and L. Breiman, "Random forests," *Mach. Learn.*, vol. 63, no. 4, pp. 5–32, 2001.

[17] A. Seki and M. Pollefeys, "Patch based confidence prediction for dense disparity map," *in Proc. Brit. Mach. Vis. Conf.*, vol. 10, Sep. 2016.

[18] Y. Li, J. B. Huang, N. Ahuja, and M. H. Yang, "Deep joint image filtering," *in Proc. Eur. Conf. Comput. Vis.*, pp. 154–169, Oct. 2016.

[19] D. Min, S. Choi, J. Lu, B. Ham, K. Sohn, and M. N. Do, "Fast global image smoothing based on weighted least squares," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5638–5653, 2014.

[20] M. Menze and A. Geiger, "Object scene flow for autonomous vehicles," *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 3061–3070, Jun. 2015.

[21] [Online] Avaliable: http://vision.middlebury.edu/stereo/.

[22] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A naturalistic open source movie for optical flow evaluation," *in Proc. Eur. Conf. Comput. Vis.*, pp. 611–625, Oct. 2012.

[23] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," [Online] http://www.cvlibs.net/datasets/kitti/.

[24] J. Zbontar and Y. Lecun, "Stereo matching by training a convolutional neural network to compare image patches," *J. Machi. Learn. Research*, vol. 17, pp. 1–32, 2016.

[25] A. Vedaldi and K. Lenc, "Matconvnet: Convolutional neural networks for matlab," *in Proc. ACM Int. Conf. Multi.*, pp. 689–692, Oct. 2015.

[26] A. Spyropoulos and P. Mordohai, "Correctness prediction, accuracy improvement and generalization of stereo matching using supervised learning," *Int. J. Comput. Vis.*, vol. 118, no. 3, pp. 300–318, 2015.