CONVOLUTIONAL COST AGGREGATION FOR ROBUST STEREO MATCHING

Somi Jeong *Seungryong Kim* *Bumsub Ham* *Kwanghoon Sohn*

School of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea
E-mail: khsohn@yonsei.ac.kr

ABSTRACT

Although convolutional neural network (CNN)-based stereo matching methods have become increasingly popular thanks to their robustness, they primarily have been focused on the matching cost computation. By leveraging CNNs, we present a novel method for matching cost aggregation to boost the stereo matching performance. Our insight is to learn the convolution kernel within CNN architecture for cost aggregation in a fully convolutional manner. Tailored to cost aggregation problem, our method differs from hand-crafted methods in terms of its convolutional aggregation through optimally learned CNNs. First, the matching cost is aggregated with cost volume unary network, and then optimized with explicit disparity boundary, estimated through disparity boundary pairwise network, within a global energy minimization. Experiments demonstrate that our method outperforms conventional hand-crafted aggregation methods

Index Terms— stereo matching, convolutional neural networks, cost aggregation, global energy minimization

1. INTRODUCTION

Stereo matching has been one of the most important and fundamental tasks for numerous computer vision applications, such as 3-D scene reconstruction and intermediate view generation [1, 2].

Many stereo matching methods primarily aim to estimate a spatially smooth but discontinuity preserved disparity map for a pair of stereo images. To achieve this goal, many approaches utilize a Markov Random Field (MRF)-based energy function [3, 4, 5], where the disparities are determined in a unary term and the spatially smooth and discontinuity preserved property is ensured in a pairwise term. They reliably estimate the disparity by minimizing a global energy function, but they have limitations on computational complexity. Other approaches, called local approaches, estimate the disparity by measuring correlation of color intensities within a local window. They achieve this goal by applying an edge-aware filtering (EAF) based cost aggregation [6, 7, 8, 9, 10]. Generally, they are much faster to obtain disparity than energy-based global approaches. However, they have challenges in defining reliable cost aggregation function and selecting the optimal window size and shape.

Generally, cost aggregation step is to regularize matching costs from neighboring pixels with an explicit kernel function [1]. Box-filtering [1], which used an average filter within fixed support window, is a simple cost aggregation method. To provide more robust aggregation performance, Yoon and Kweon [6] used bilateral weight using color similarity and geometric proximity. They assumed that there exists a high correlation between boundaries of color image and disparity map. Several cost aggregation methods have adopted robust EAF methods, such as guided filter [9, 11] and domain transform [12, 10] to smooth the matching cost while preserving the disparity boundaries efficiently. Instead of using the fixed window, cross based aggregation method [8] used an adaptive window, and aggregated matching cost within window only. The performances of these methods largely depend on the hand-crafted kernel constructed using color image, and it is based on the assumption that color and disparity boundaries are coherent. The inherent discrepancy between them, however, may degrade their aggregation performance. To overcome above limitation, few methods proposed to use the disparity boundary and the color boundary together [13]. However, it was also based on the hand-crafted kernel, so its performance was still limited.

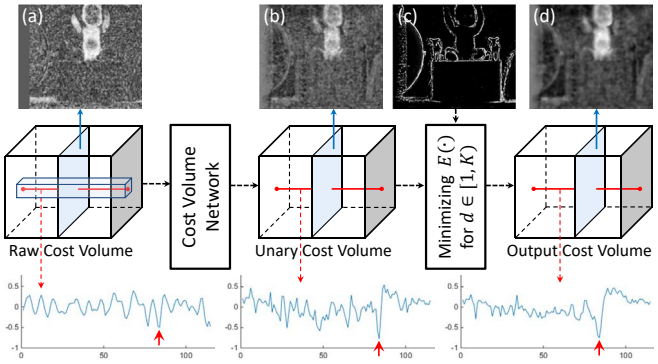


Fig. 1. Framework of our proposed cost aggregation. From an initial raw cost volume in (a), we aggregate the matching costs through sucessive convolutions within CNNs to produce aggregated unary cost volume in (b). Moreover, using estimated pairwise disparity boundaries in (c), we also aggregate the unary cost volume by minimizing a global energy function to estimate final output cost volume in (d).

Recently, various computer vision tasks have been reformulated using convolutional neural networks (CNNs) due to their robustness, such as image classification [14, 15], object detection [16, 17], and image segmentation [18], and they show the magnificently improved results. CNNs also have contributed to improve the stereo matching performance by adjusting it to the matching cost computation step. Some methods proposed to learn features via networks to classify two input patches matching or not, and computed the matching costs using these features [19, 20, 21]. However, due to the inherent challenges of stereo matching tasks, despite the CNN-based matching costs, it is hard to estimate accurate disparities only using the matching costs. In order to boost the performance, they applied conventional hand-crafted aggregated methods such as cross-based cost aggregation [8] and box-filtering [1], but they still show the limited performance.

To overcome this limitation, we propose a novel cost aggregation method by leveraging CNNs, illustrated in Fig. 1. Our insight is to learn the kernel function for the cost aggregation in a fully convolutional manner. To this end, we use CNNs architecture to aggregate the raw matching cost, where the parameters are learned directly

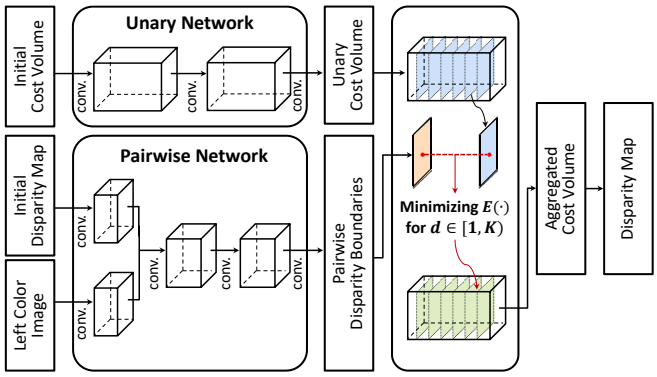


Fig. 2. The network architecture of our proposed cost aggregation.

from a ground-truth disparity map. To boost the aggregation performance, we use additional CNNs to predict the tentative disparity boundary. The outputs of those networks are then combined by minimizing a global energy function on each cost slice. Experimental results show that our proposed method outperforms conventional cost aggregation methods on the Middlebury benchmark [22].

2. PROBLEM FORMULATION AND CHALLENGES

Given a rectified pair of stereo images I, I', stereo matching aims at estimating disparity D for each pixel $p = [p_{\mathbf{x}}, p_{\mathbf{y}}]^T$. Matching costs are first measured for pixel p across disparity candidates $d = \{1, ..., K\}$, where K is the maximum disparity range, such that $C(p, d) = S(I(p), I'(p - [d, 0]^T))$, where $S(\cdot, \cdot)$ is the cost function for measuring the dissimilarity, e.g., census transform [23] or MC-CNN [19]. Note that matching costs are defined in 3-D space to build 3-D cost volume.

To eliminate the effects of outliers and produce reliable disparity maps, raw matching costs are aggregated from the neighboring pixels [1]. Based on the assumption of discontinuity consistency between color image and disparity map, most existing methods utilize color image as a guidance to aggregate the matching costs [9, 10, 24]. The aggregated matching cost C' is obtained as follows:

$$C'(p, d) = \sum\nolimits_{q \in \mathcal{N}_p} w_I(p, q) C(q, d), \tag{1}$$

where $\mathcal{N}_p$ is a local aggregation window centered at pixel p, and $w_I(p, q)$ is a normalized edge-aware weight within $\mathcal{N}_p$. By finding a minimum of $C'(p, d)$ across the disparity candidates d, the final disparity D can be obtained by $D(p) = \operatorname{argmin}_d C'(p, d)$.

Conventionally, most existing cost aggregation methods have been focusing on how to optimally design the edge-aware weights in a hand-crafted manner [1]. However, they have inherent limitations to provide an optimal performance. Firstly, since they are formulated with hand-crafted features, their performance depends on the matching cost functions and the parameter settings such as truncation value or window size [25]. To obtain the optimal aggregation performance, there is no alternative but to change the applied cost function or tune the parameter settings. As a result, it is hard to determine the optimal parameter settings that provide the consistently reliable performance for all cost functions. Secondly, the edge-aware weight derived from the only color image cannot define the disparity boundaries optimally because of an inherent color-disparity discrepancy [13]. Finally, 2-D cost aggregation window cannot consider a correlation on matching costs across disparity search spaces, which can contribute to boost the cost aggregation performance.

3. PROPOSED METHOD

3.1. Overview

By leveraging CNNs, our objective is to design a novel cost aggregation method that reformulates the cost aggregation step in learning framework. Our key algorithmic idea is that optimal kernels for cost aggregation can be learned as convolution kernels of CNNs. To this end, we formulate two sub-networks as shown in Fig. 2, unary and pairwise networks. We employ the unary network, which input is an initial cost volume. It aggregates the matching costs through successive convolution layers, and estimates the aggregated cost volume, called unary cost volume. It inherently aggregates matching costs not only within the local spatial window but also across the disparity search spaces efficiently and effectively.

In the pairwise network, to boost the aggregation performance, the tentative disparity boundaries are estimated using both color image and initial disparity map. Using estimated disparity boundaries, the unary cost volume is aggregated on each cost slice by minimizing global energy function. Unlike conventional hand-crafted weights [9, 10], our disparity boundaries provide reliable performances.

3.2. Cost Volume Unary Network

In the cost volume unary network, the raw matching costs are aggregated with successive convolutions. In this section, we clarify the concept of our cost aggregation approach by using a single convolution, and it is easily extended into multiple convolution cases. With convolution property, an aggregated cost value at pixel p can be estimated by summing the matching cost values on 3-D local neighborhood $\mathcal{M}_p$ weighted by convolution kernels $\mathbf{W}_u$ as

$$\begin{aligned} C'(p, k) &= \mathbf{W}_u^k * C(p, d) \\ &= \sum\nolimits_{(q,d) \in \mathcal{M}_p} \mathbf{W}_u^k(q, d) C(p - q, d), \end{aligned} \tag{2}$$

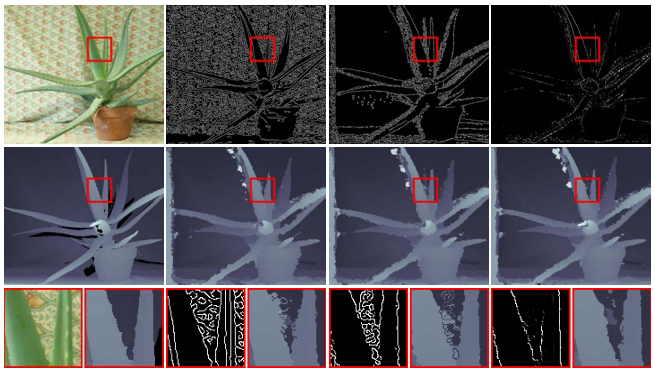
where (q, d) is defined as pixels within 3-D local aggregation neighborhood $\mathcal{M}_p$, and $\mathbf{W}_u^k$ is the k^{th} convolution kernel.

Compared to conventional cost aggregation kernels [9, 10], convolutional kernels $\mathbf{W}_u$ in CNNs have two benefits. Firstly, they are learned with a large number of ground-truth disparity maps, and thus it provides highly robust performance. With the optimally learned convolution kernels, our approach can provide consistently reliable aggregation performance regardless of the cost functions and the parameter settings. Secondly, they aggregate the matching costs within not only spatial local neighborhood but also disparity search spaces. When determining the disparity among the matching costs for all search ranges, the matching costs can provide helpful cues to estimate the confidence of disparity [26]. Thus, by aggregating the matching costs within 3-D local neighborhood, the disparity can be estimated more optimally and effectively.

The unary network has 3 convolutional layers, consisting of 5×5 sized 256, 256 and K convolutional kernels followed by batch normalization and ReLU activation function. To preserve the spatial resolution, any pooling operator is not used. Since the disparity candidates $d = \{1, ..., K\}$ are defined in the discrete domain, we treat it as a K-class classification problem, and each disparity can be treated as one class. Inspired by fully convolutional networks [18], we train our network to classify K-disparities using softmax loss in order to predict pixel-wise disparity with ground-truth disparity.

3.3. Disparity Boundary Pairwise Network

Even though the cost volume unary network itself works well for aggregating the matching costs, disparity boundary can boost the



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

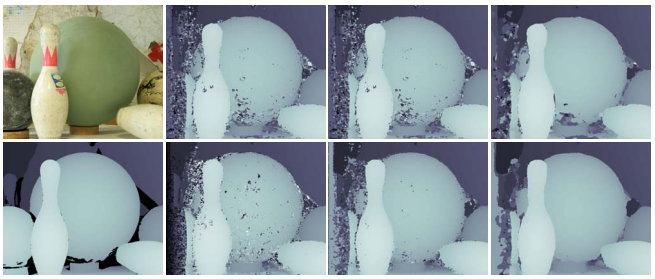
Fig. 3. Comparison of cost aggregations using different boundaries. Top: (a) left color image and boundaries of (b) color image, (c) initial disparity map, and (d) proposed method. Bottom: (a) ground-truth disparity map and resultant disparity maps of (b),(c) and (d).

aggregation performance and localization ability around disparity boundaries. To this end, we employ the disparity boundary pairwise network to predict the tentative disparity boundary.

There are some boundary detection methods using CNNs [27, 28]. The color image is used as input of CNNs and they show outstanding boundary detection performance than others using hand-engineered method. However, even using CNNs, estimating the disparity boundary is not easy task. The color image boundary can be clearly defined, but there is inherent discrepancy between color and disparity map. It is unsuitable to estimate the disparity boundary optimally just using the color image. Moreover, we attempt to use initial disparity map as input of CNNs. It is also not successful attempt to estimate the exact disparity boundary because of its outliers.

We propose the joint usage of color image I and initial disparity map $D'(p) = \text{argmin}_d C(p,d)$ as inputs in pairwise network. It can contribute the disparity boundary estimation in a synergic manner combined the color image and the initial disparity map. Note that although there were a few attempts [13, 29] to estimate the disparity boundaries using these two inputs, their performances were still limited due to their hand-crafted formulation. Since these two inputs are mutually beneficial, our pairwise network can predict highly robust disparity boundaries such that $B(p) = \mathbf{W}_b * \{I(p), D'(p)\}$, where $\mathbf{W}_b$ represents the parameters of pairwise network. As exemplified in Fig. 3, while the color and the initial disparity boundary images contain irrelevant boundaries on complex texture or outlier regions, our disparity boundary provides a high consistency to the ground-truth disparity map, and as a result, it boosts the disparity estimation performance.

This network consists of 5 convolutional layers, and convolutional filter size is 5×5 followed by batch normalization and ReLU activation functions. To fuse multiple inputs, we formulate this network as three sub-sets of convolutional layers for initial disparity feature extraction, color feature extraction, and fusion. For extracting the features from two inputs, first two sets of convolutions have 64 convolutional kernels. Then, these two kind of features are concatenated. The fusion layer consists of three convolutional layers, consisting of 128, 128 and 2 convolutional kernels. We formulate the disparity boundary estimation as a binary classification problem for boundary or non-boundary class. We define ground-truth disparity boundary map obtained using Canny edge detection on the ground-truth disparity map [30]. To achieve sharper boundaries, a non-maximum suppression scheme is further applied.



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

Fig. 4. Comparison of cost aggregations on initial cost volumes using (top) census transform [23] and (bottom) MC-CNN [19]: (a) left color image and ground-truth disparity map, (b) initial disparity maps, and refined disparity maps using (c) only unary cost volume network and (d) both unary and pairwise network combined with global energy function minimization.

3.4. Global Inference

To boost the performance, the unary cost volume C' and the tentative disparity boundary map B should be integrated in a synergistic manner. We fuse them using the global energy minimization defined on each cost slice. We formulate the global energy function based on weighted least squares (WLS) optimization framework. With omitting k for simplicity, the aggregated matching cost C'' can be obtained by minimizing the following energy function:

$$E(C'') = \sum\nolimits_p (C''(p) - C'(p))^2 + \lambda \sum\nolimits_p \sum\nolimits_{q \in \mathcal{N}_p^4} w_B(p,q)(C''(p) - C''(q))^2, \quad (3)$$

where λ is a regularization parameter, $\mathcal{N}_p^4$ is a 4-neighborhood, and w_B is a weight defined with the disparity boundary map B as

$$w_B(p,q) = \exp(-\|B(p) - B(q)\|^2 / \sigma_B), \quad (4)$$

where σ_B is a Gaussian bandwidth parameter. Our energy function consists of unary and pairwise terms, balanced by the regularization parameter. The unary term lets aggregated matching cost harmonize well with the unary matching costs, and the pairwise term smooths aggregated matching costs using the disparity boundary as the guidance.

By minimizing the energy function on each matching cost slice, the aggregated matching costs finally can be obtained such that

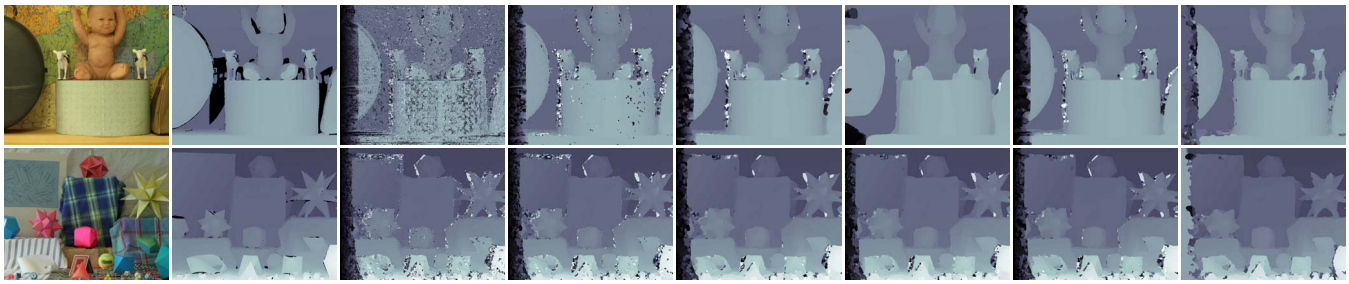
$$\mathbf{C}'' = (\mathbf{I} + \lambda(\mathbf{L} - \mathbf{W}_B))^{-1}\mathbf{C}' \quad (5)$$

where $\mathbf{C}''$, $\mathbf{C}'$ and $\mathbf{W}_B$ are matrix forms defined for all pixels p of C'', C' and w_B, respectively. $\mathbf{L}$ is a Laplacian matrix for $\mathbf{W}_B$. With recent fast solver [31], such minimization can be very efficiently performed. Since the matching costs are aggregated through all image domain, the effects of outliers within local neighborhood can be definitely reduced. The final disparity map can be achieved such that $D(p) = \text{argmin}_d C''(p,d)$.

4. EXPERIMENTAL RESULTS

4.1. Experimental Settings

We implemented the proposed networks using VLFeat MatConvNet toolbox [32]. The filter weights of each layer were initialized by Gaussian distribution with zero mean and a standard deviation of 0.001. We normalized the inputs by subtracting the mean and dividing by the standard deviation. We set the learning rate as 0.001.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 5. Comparison of qualitative evaluations on the Middlebury benchmark [22] for cost aggregations on initial cost volumes using (top) census transform [23] and (bottom) MC-CNN [19]: (a) left color images, (b) ground-truth disparity maps, (c) initial disparity maps, refined disparity maps using (d) BF [1], (e) GF [11], (f) DT [12], (g) CBCA [8], and (h) our method.

To build the initial cost volume, census transform [23] and MC-CNN [19] were utilized. For the census transform [23], we cropped 9×9 image patch to represent each image position as bit vector, sized 81. For the MC-CNN [19], we adapted the fast architecture. Especially, we used the cost volume without any post-processing techniques refered in [19]. We directly used the code provided by the authors.

With this initial cost volume, we evaluated our method compared with conventional cost aggregation methods, such as box filter (BF) [1], guided filter (GF) [11], domain transform (DT) [12], and cross-based aggregation (CBCA) [8]. We utilized the built-in codes in MATLAB for BF [1] and GF [11] and utilized the code from the authors for DT [12]. We implemented the code for CBCA [8]. For a fair comparison, no post-processing techniques were employed.

For quantitative evaluations, we evaluated the error rate by measuring the percentage of bad matching pixels, whose the absolute disparity error is greater than 2 pixel, in two subsets: the pixels in the non-occluded regions (non-occ) and in the whole region (all).

4.2. Datasets

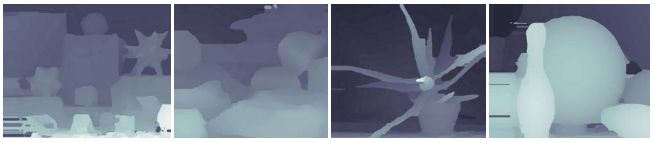
We used the image pairs of Middlebury stereo dataset [22], specifically half resolutions of the year 2005 and 2006, which have 6 and 21 rectified pairs of stereo image respectively with 700×550 resolution and 115 maximum disparity. We excluded 5 stereo pairs, *i.e. Lamp1, Lamp2, Midd1, Midd2 and Plastic* because they contain large textureless regions and could degenerate overall performance of our learning. Among 22 images, we divided them into 16 pairs in training set and 6 pairs in test set. Even though the training images are only 16 pairs, our networks were formulated as the pixel-by-pixel classification problem, so each pixel can be seen as one training candidate, *i.e.* over 5 million pairs are used for training our network, which is similar with [19, 21].

4.3. Results

We first evaluated the component-wise analysis of two sub-networks in our method, *i.e.*, unary network and pairwise network, as in Fig. 4. We then evaluated quantitative comparisons of conventional cost aggregation methods in Fig. 5 and Table 1. As expected, conventional hand-crafted aggregation methods such as BF [1], GF [11], DT [12], and CBCA [8] show limited performances for both two cases using census transform [23] and MC-CNN [19] for the initial matching costs. Compared to these methods, our aggregation method using only the unary cost volume network has already shown outstanding performances thanks to learned optimal kernels through CNNs. Moreover, minimizing our global energy function with the estimated disparity boundaries further boosted the aggregation performance.

Table 1. Comparion of qualitative evaluations on the Middlebury benchmark [22] of proposed method and conventional methods.

Methods	Census [23]		MC-CNN [19]	
	non-occ	all	non-occ	all
Initial	24.492	31.247	8.858	17.038
BF [1]	9.279	17.581	7.442	15.680
GF [11]	7.870	15.804	6.522	14.455
DT [12]	12.247	16.292	6.368	14.291
CBCA [8]	7.680	15.345	6.206	14.316
Ours-Unary	**5.359**	**14.356**	**5.036**	**13.102**
Ours	**4.558**	**13.012**	**4.398**	**11.820**



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

Fig. 6. Visualization of refined disparity maps of our method through post-processing as in [19] on initial cost volumes using (a),(b) census transform [23] and (c),(d) MC-CNN [19].

Fig. 6 illustrates estimated disparity maps of our method through post-processing similar to [19], including semiglobal matching [33], interpolation, subpixel enhancement and refinement. It proves that our method shows the reliable performance for the cost aggregation.

5. CONCLUSION

We presented an efficient cost aggregation method by leveraging CNNs to obtain an accurate disparity map. Based on the insight that optimal kernel function for cost aggregation can be learned as convolutions in CNNs, we formulated two sub-networks, unary cost volume network and pairwise disparity boundary network, and trained these network in a fully convolutional manner. To collaborate these outputs efficiently, we employed the minimization of global energy function on each cost slice. We evaluated our proposed method on Middlebury benchmark [22], which demonstrates that our method definitely outperforms other conventional cost aggregation methods.

6. ACKNOWLEDGMENTS

This work was supported by Institute for Information and communications Technology Promotion(IITP) grant funded by the Korea government(MSIP)(No.2016-0-00197).

7. REFERENCES

[1] D. Scharstein and R. Szeliski, "A taxanomy and evaluation of dense two-frame stereo correspondence algorithms," *IJCV*, vol. 47, no. 1-3, pp. 7–42, 2002.

[2] B. Ham, D. Min, C. Oh, M. N. Do, and K. Sohn, "Probability-based rendering for view synthesis," *IEEE Trans. IP*, vol. 23, no. 2, pp. 870–884, 2014.

[3] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A comparative study of energy minimization methods for markov random fields with smoothness-based priors," *IEEE Trans. PAMI*, vol. 30, no. 6, pp. 1068–1080, 2008.

[4] K. Yamaguchi, T. Hazan, D. McAllester, and R. Urtasun, "Continuous markov random fields for robust stereo estimation," *In ECCV*, 2012.

[5] S. Kim, B. Ham, B. Kim, and K. Sohn, "Mahalanobis distance cross-correlation for illumination-invariant stereo matching," *CSVT*, vol. 24, no. 11, pp. 1844–1859, 2014.

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

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

[8] K. Zhang, J. Lu, and G. Lafruit, "Cross-based local stereo matching using orthogonal integral images," *CSVT*, vol. 19, no. 7, pp. 1073–1079, 2009.

[9] A. Honsi, C. Rhemann, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," *IEEE Trans. PAMI*, vol. 35, no. 2, pp. 504–511, 2013.

[10] C.C. Pham and J.W. Jeon, "Domain transformation-based efficient cost aggregation for local stereo matching," *CSVT*, vol. 23, no. 7, pp. 1119–1130, 2013.

[11] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. PAMI*, vol. 35, no. 6, pp. 1397–1409, 2013.

[12] E. Gastal and M. Oliveira, "Domain transform for edge-aware image and video processing," *In ACM SIGGRAGH*, 2011.

[13] D. Chen, M. Ardabilian, and L. Chen, "Depth edge based trilateral filter method for stereo matching," *In ICIP*, 2015.

[14] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet classification with deep convolutional neural networks," *In NIPS*, pp. 1097–1105, 2012.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *In CVPR*, 2016.

[16] R. Girshick, J. Donahue, T. Darrell, and Malikk J., "Rich feature hierarchies for accurate object detection and semantic segmentation," *In CVPR*, 2014.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *In ECCV*, 2014.

[18] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *In CVPR*, 2015.

[19] J. Zbontar and Y. LeCun, "Stereo matching by training a convolutional neural network to compare image patches," *JMLR*, vol. 17, no. 1-32, pp. 2, 2016.

[20] Z. Chen, X. Sun, and L. Wang, "A deep visual correspondence embedding model for stereo matching costs," *In ICCV*, 2015.

[21] W. Luo, A.G. Schwing, and R. Urtasun, "Efficient deep learning for stereo matching," *In CVPR*, 2016.

[22] D. Scharstein and C. Pal, "Learning conditional random fields for stereo," *In CVPR*, 2007.

[23] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," *In ECCV*, 1994.

[24] J. Lu, H. Yang, D. Min, and M. N. Do, "Patchmatch filter: Efficient edge-aware filtering meets randomized search for fast correspondence field estimation," *In CVPR*, 2013.

[25] M. Gong, R. Yang, L. Wang, and M. Gong, "A performance study on different cost aggregation approaches used in real-time stereo matching," *IJCV*, vol. 75, no. 2, pp. 283–296, 2007.

[26] C.L. Zitnick and T. Kanade, "A cooperative algorithm for stereo matching and occlusion detection," *IEEE Trans. PAMI*, vol. 22, no. 7, pp. 675–684, 2000.

[27] Gedas Bertasius, Jianbo Shi, and Lorenzo Torresani, "Deepedge: A multi-scale bifurcated deep network for top-down contour detection," *In CVPR*, 2015.

[28] Wei Shen, Xinggang Wang, Yan Wang, Xiang Bai, and Zhijiang Zhang, "Deepcontour: A deep convolutional feature learned by positive-sharing loss for contour detection," *In CVPR*, 2015.

[29] K.R. Kim and C.S. Kim, "Adaptive smoothness constraints for efficient stereo matching using texture and edge information," *In ICIP*, 2016.

[30] J. Canny, "A computational approach to edge detection," *IEEE Trans. PAMI*, , no. 6, pp. 679–698, 1986.

[31] 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. IP*, vol. 23, no. 12, pp. 5638–5653, 2014.

[32] V. Andrea and K. Lenc, "Matconvnet: Convolutional neural networks for matlab," http://www.vlfeat.org/matconvnet/.

[33] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. PAMI*, vol. 30, no. 2, pp. 328–341, 2008.