

IMPROVING DISPARITY MAP ESTIMATION FOR MULTI-VIEW NOISY IMAGES

Shiwei Zhou, Zhengyang Lou, Yu Hen Hu, Hongrui Jiang

University of Wisconsin, Madison, WI 53706, USA

ABSTRACT

A robust multi-view disparity estimation algorithm for noisy images is presented. The proposed algorithm constructs 3D focus image stacks (3DFIS) by projecting and stacking multi-view images and estimates a disparity map based on the 3DFIS. To make the algorithm robust to noise and occlusion, a texture-based view selection and patch size variation scheme based on texture map is proposed. Experiment results indicate that the proposed algorithm outperforms conventional stereo matching algorithms as well as previously reported multi-view disparity estimation algorithms under noisy conditions.

Index Terms— disparity map, multi-view images, 3D focus image stacks, noisy image

1. INTRODUCTION

Stereo matching for estimating disparity map using stereo images has been well-studied in computer vision [1-26]. Disparity map can also be estimated using multi-view images (more than two views) [27-30], and plays a critical role in multi-view image processing like denoising, super resolution, etc. However, most of existing algorithms are specifically tailored for noiseless images, and their performance deteriorates when the images are contaminated by noise, which further impedes future image processing procedures. In this paper, disparity estimation from multiple views under noise will be investigated.

Disparity estimation methods can be categorized into local and global types. Local methods [1-13] are relatively simple and efficient, which makes them capable of real-time processing. Typical local methods compute matching cost for each pixel within a support region using all possible disparities, and then performs a “winner-takes-all” (WTA) optimization [1]. The downside of local methods, however, is that their simple structures make them prone to producing inaccurate estimations in homogenous areas that have few textures, as well as in occluded regions that are invisible to partial of the cameras. Improvements of local methods focus on using more robust matching costs [2-7], adaptive support windows [8-10], efficient cost aggregation [10-13], etc. Global methods [14-23], on the other hand, demonstrate

superior performance in such areas where local methods fall short. They are often formulated in an energy-minimization framework that aims to minimize a global energy function consisting a data term and a number of regularization terms. A variety of algorithms have been proposed to solve the optimization problem, including simulated annealing [14, 15], belief propagation [16-18], graph cut [19-21], dynamic programming [22, 27], etc. Unfortunately, the high computational complexity of energy function minimization in global methods prevents them from being applied to real-time implementation. Meanwhile, machine learning techniques that assist matching cost computation have also been recently studied. Zbontar et al. [23] proposed to learn a similarity measure using a convolutional neural network (CNN). Multiple literature has been published to improve this CNN-based method, including hierarchical image segmentation [24], bilateral solver [25], and weakly supervised learning on unlabeled images [26]. However, the training stage of such methods requires high computational cost and hardware configurations such as GPU acceleration. Also, their performance on noisy images is unknown.

In this work, we adopt the simple framework of local methods for its efficiency, while implementing the 3D focus image stacks (3DFIS) to facilitate a more accurate disparity estimation in noisy condition. To overcome the issues caused by noise and occlusion, we further propose a view selection and patch size variation scheme based on texture map that is estimated from the 3DFIS.

2. DISPARITY ESTIMATION

We assume the multi-view images are acquired from a dense planar array of cameras. The cameras are placed at grid points $(s, t) \in \mathbf{Z}^2$ which is a set of 2D integer indices of the camera array. Without loss of generality, the center view, which is located at $(0, 0)$, is designated as the reference view. The images are assumed to have been calibrated and rectified so that their epipolar lines are parallel to the horizontal axis, and homography between images can be simplified to pure translation. In noisy conditions, the noise-corrupted images can be modeled as the sum of clean images and additive white Gaussian noise for each pixel (x, y) :

$$I_{s,t}(x, y) = I'_{s,t}(x, y) + n_{s,t}(x, y), \quad (1)$$

where $I'_{s,t}$ is the noiseless clean images, and $n_{s,t}$ is i.i.d. zero-

mean Gaussian noise with variance σ^2 .

2.1. 3D Focus Image Stacks

3D focus image stacks (3DFIS) was previously introduced by us for multi-view image denoising [30]. Assume there are K cameras in the camera array and each camera corresponds to a grid point (s, t) , i.e. there is a unique mapping from (s, t) to an integer k such that $1 \leq k \leq K$. Then each image $I_{s,t}(x, y)$ at the (s, t) view is shifted by (sd, td) and stacked upon each other to form a 3-dimensional matrix

$$F^d(x, y, k) = I_{s,t}(x - sd, y - td). \quad (2)$$

The 3D matrix $F^d(x, y, k)$ is called a *3D focus image stack (3DFIS)* with respect to disparity value d . It contains all the information from all cameras when the focal plane corresponds to disparity d . Ideally, if the (x, y) pixel at the target view $I_{0,0}(x, y)$ has the true disparity value d , then the entire column of $F^d(x, y, k)$, denoted by $F^d(x, y, :)$, should have the same pixel value, that is,

$$F^d(x, y, :) = I_{0,0}(x, y) \cdot \mathbf{1}_{K \times 1}, \quad (3)$$

where $\mathbf{1}_{K \times 1}$ is a vector consisting of all 1s.

2.2. Disparity Map Estimation using 3DFIS

With the 3DFIS constructed from multiple views, the disparity map can be estimated by exploring the consistency of images in each stack. Some previous works [29-31] generate *multi-focus images (MFI)* by simply averaging each 3DFIS as

$$I^d(x, y) = \frac{1}{K} \sum_{k=1}^K F^d(x, y, k), \quad (4)$$

and then compute matching cost using the sum of absolute difference (SAD) between each MFI and the reference view

$$C(x, y, d) = \frac{1}{n} \sum_{(i,j) \in N(x,y)} |I^d(i, j) - I_{0,0}(i, j)|, \quad (5)$$

where $N(x, y)$ is the square neighborhood of (x, y) , and n is the number of pixels in $N(x, y)$. The disparity value of (x, y) is then estimated as

$$\hat{d}(x, y) = \arg \min_d C(x, y, d). \quad (6)$$

The MFI-based method, while simple, often yields noisy, spurious disparity maps. In particular, occlusions due to sharp discontinuities on the disparity map are not considered in forming the MFI. In addition, flat or low-texture regions tend to produce ambiguous estimations due to matching errors. To address these issues, a robust matching cost, along with a texture-based view selection and patch size variation scheme is proposed.

2.2.1. Disparity Estimation with Robust Matching Cost

For each pixel (x, y) and the support window W centered at it, let us denote by \mathbf{v}_k^d a vector containing all pixel values

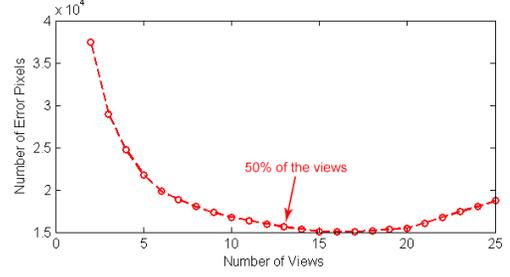


Fig. 1. Number of error pixels when using different number of views

within window W of k^{th} view in the focus image stack F^d . For convenience, let \mathbf{v}_1^d be the column corresponding to the reference view. We compute the vector difference between the k^{th} view ($k > 1$) and reference view ($k = 1$):

$$\tilde{\mathbf{v}}_k^d = \mathbf{v}_k^d - \mathbf{v}_1^d, \quad 2 \leq k \leq K. \quad (7)$$

Then sort the sequence $\{\|\tilde{\mathbf{v}}_k^d\|_1; 1 \leq k \leq K\}$ in increasing order such that $k \rightarrow k'$, $\|\tilde{\mathbf{v}}_{k'}^d\|_1 \leq \|\tilde{\mathbf{v}}_{k'+1}^d\|_1$. Here, $\|\tilde{\mathbf{v}}_k^d\|_1$ is the sum of absolute values of each element in $\tilde{\mathbf{v}}_k^d$. Define the matching cost as the mean value of h best $\|\tilde{\mathbf{v}}_{k'}^d\|_1$ ($1 < h \leq K$) as

$$C^*(x, y, d) = \frac{1}{n(h-1)} \sum_{k'=2}^h \|\tilde{\mathbf{v}}_{k'}^d\|_1, \quad (8)$$

where n is the number of pixels in patch vector \mathbf{v}_k^d . Finally, the disparity value for pixel (x, y) will be estimated as

$$\hat{d}^*(x, y) = \arg \min_d C^*(x, y, d). \quad (9)$$

It can be easily proved that the matching cost C^* in eq. (8) is greater than or equal to the C in eq. (5). In other words, patch mismatches tend to produce higher cost in eq. (8), and thus the proposed matching cost makes it easier to distinguish the true disparity from all other candidates, making the disparity estimation more robust to noise.

In eq. (8), the choice of h is critically important to the accuracy of the estimated disparity map. If a view is occluded by another object due to discontinuities in the disparity map, serious bleeding artifacts, as illustrated in Fig. 2(a)(b), may degrade the quality of the estimated disparity map. Previously, Kang et al. [32] proposed to use the best 50% of the frames (views) in computing the matching cost. However, as shown in Fig. 1, the number of views that yields best disparity estimation may need more than 50% of views. In this work, we propose a dynamic view selection heuristic based on the texture analysis of the reference view.

Apart from view selection, we also consider the patch size used in computing the matching score. Small patch sizes tend to reveal details but are vulnerable to noises in flat regions, and large patches will have the opposite behavior. Therefore, we propose a method to vary patch size selections based also on the texture analysis of the image.

2.2.2. Texture Map Estimation

Due to the corruption of noise, texture map estimated from the reference image contains large amount of noisy artifacts.

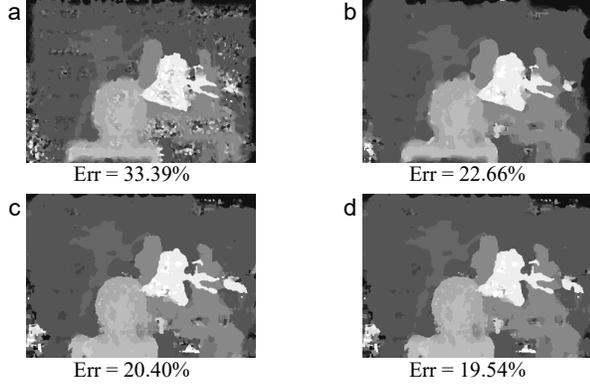


Fig. 2. Disparity maps using (a) fixed patch size (5×5) + all views; (b) variable patch size + all views; (c) variable patch size + 50% views; (d) variable patch size + view selection

In correspondence, we developed a novel multi-view strategy to estimate texture map such that the impact of noise can be reduced from redundant information of multiple views. Previously in Section 2.1, it was mentioned that the column vector $F^d(x, y, :)$ should have the same intensity value if the true disparity value at (x, y) is d , and vice versa. In other words, the variance of vector $F^d(x, y, :)$ is small only when the true disparity at pixel location (x, y) is d . However, this is only true for high-texture regions, where edges and pixel intensity variations are prevalent. In low-texture regions, due to the homogeneity of pixel values within the neighborhood, column vector $F^d(x, y, :)$ tends to contain similar values no matter whether the true disparity is d or not, which in turn makes the variance of $F^d(x, y, :)$ remains relatively small for all disparity values.

Given 3DFIS $F^d(x, y, k)$, the standard deviation $\sigma^d(x, y)$ is obtained for each pixel (x, y)

$$\sigma^d(x, y) = \sqrt{\frac{1}{K} \sum_{k=1}^K (F^d(x, y, k) - \bar{F}^d(x, y))^2}, \quad (10)$$

where $\bar{F}^d(x, y)$ is the mean value of vector $F^d(x, y, k)$. Then the strength of textures at (x, y) is defined as

$$\Sigma(x, y) = \frac{1}{d_{\max}} \sum_{d=1}^{d_{\max}} \sigma^d(x, y), \quad (11)$$

where d_{\max} is the maximum of candidate disparity values. To further reduce the impact of noise, we also apply a smoothing filter (e.g. Gaussian) to each σ^d . Fig. 3 shows an example of texture map of a multi-view dataset, where bright colors (large values) represent high textures, while low textures are identified as dark colors (small values). From the texture map, low and high texture regions can be identified to assist disparity estimation, which will be discussed next.

2.2.3. View Selection and Patch Size Variation

With the texture map, both patch sizes and number of views to be selected can be calculated accordingly. Assume the

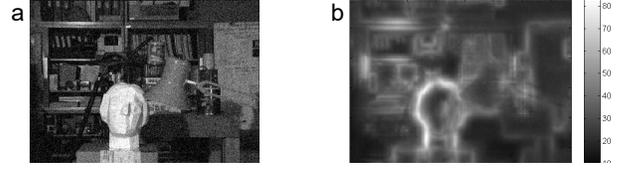


Fig. 3. (a) noisy image; (b) texture map

size of a patch may range from $L_{\min} \times L_{\min}$ to $L_{\max} \times L_{\max}$. In flat areas, the texture strength at pixel (x, y) , $\Sigma(x, y)$, decreases, and needs a larger patch size. When $\Sigma(x, y) \leq \Sigma_l$ (pre-determined lower threshold), the maximum patch size L_{\max} is used. Similarly, when $\Sigma(x, y) \geq \Sigma_u$ (pre-determined upper threshold), the minimum patch size L_{\min} is used. Then we define the patch size $L(x, y)$ as a linear function of texture strength $\Sigma(x, y)$

$$L(x, y) = \begin{cases} L_{\max} & , \Sigma(x, y) \leq \Sigma_l \\ \frac{L_{\max} - L_{\min}}{\Sigma_l - \Sigma_u} \cdot \Sigma(x, y) + \frac{\Sigma_l \cdot L_{\min} - \Sigma_u \cdot L_{\max}}{\Sigma_l - \Sigma_u} & , \Sigma_l < \Sigma(x, y) < \Sigma_u \\ L_{\min} & , \Sigma(x, y) \geq \Sigma_u \end{cases} \quad (12)$$

In practice, the values of Σ_l and Σ_u are related to the noise level and are determined empirically through experiments.

The number of views selected can also be related with the texture strength in a similar way. We define the number of selected views $V(x, y)$ as a function of patch size $L(x, y)$

$$V(x, y) = \begin{cases} 0.5K & , L(x, y) = L_{\min} \\ \frac{K}{2(L_{\max} - L_{\min})} \cdot L(x, y) + \frac{K \cdot L_{\max} - 2K \cdot L_{\min}}{2(L_{\max} - L_{\min})} & , L_{\min} < L(x, y) < L_{\max} \\ K & , L(x, y) = L_{\max} \end{cases} \quad (13)$$

where K is the total number of views. In other words, the number of views we selected is proportional to the patch size. Fig. 2 (a)-(d) illustrate the disparity maps using different strategies, ranging from the vanilla version from our previous works to the improved version proposed in this paper. As illustrated in Fig. 2(d), the robust view selection and patch size variation process can reduce the error of

Algorithm 1 Disparity Estimation

Input: 3DFIS $F^d, d = 1, \dots, d_{\max}$
Output: Disparity map D_{est}

- 1 **for** $d = 1 : d_{\max}$
- 2 **for** each pixel location (x, y)
- 3 Compute $\sigma^d(x, y)$ using eq. (10);
- 4 **end**
- 5 Apply Gaussian filter to σ^d ;
- 6 **end**
- 7 Estimate texture map Σ using eq. (11);
- 8 Estimate patch size L and number of views V using eq. (12) (13);
- 9 **for** $d = 1 : d_{\max}$
- 10 **for** each pixel location (x, y)
- 11 Compute $C^*(x, y, d)$ using eq. (8), with $h = V(x, y)$, and patch size $= L(x, y)$;
- 12 **end**
- 13 **end**
- 14 Compute estimated disparity map d_{est} using eq. (9);

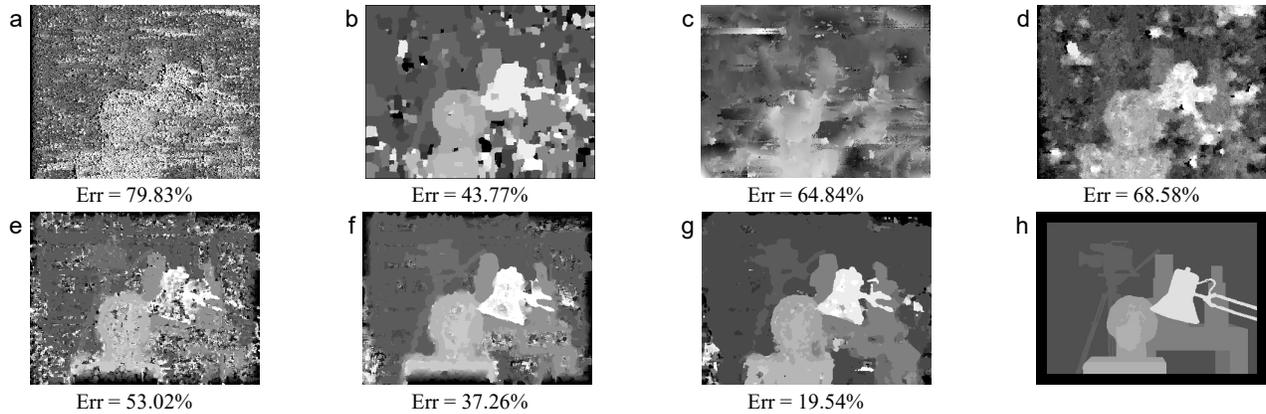


Fig. 4. Comparison of disparity map estimation at $\sigma = 20$: (a) Kolmogorov et al. [19]; (b) Klaus et al. [17]; (c) Tani ai et al. [21]; (d) Lee et al. [12]; (e) Miyata et al. [29]; (f) Zhou et al. [30]; (g) proposed; (h) ground truth

disparity estimation. An overview of the disparity estimation is illustrated in Algorithm 1.

3. EXPERIMENTS

We evaluate the proposed disparity estimation algorithm through experiments on several multi-view image datasets. Our experiments are conducted using the popular Middlebury Stereo Dataset [33]. For all datasets, white Gaussian noise with standard deviation $\sigma = 20$ is added. The evaluation criteria for measuring the quality of disparity map is the error percentage, which is defined as

$$Err(d_{est}) = \frac{1}{N} \sum_{i=1}^N (d_{est}(i) \neq d_{gt}(i)), \quad (14)$$

where d_{est} is the estimated disparity map, and d_{gt} is the ground truth. N is the total number of pixels in the image. As for the parameters, the maximum and minimum patch sizes are set to $L_{max} = 15$, $L_{min} = 5$. The upper and lower threshold Σ_u and Σ_l are defined as $\Sigma_u = 0.5\sigma + 19$, $\Sigma_l = 0.75\sigma + 5$, where σ is the noise standard deviation.

In Table 1, we compare the error percentage of different methods on various datasets when noise level $\sigma = 20$. Kolmogorov et al. [19] and Klaus et al. [17] are two of the conventional stereo matching algorithms that produce disparity maps with decent quality on noise-free images. Tani ai et al. [21] is currently the state-of-the-art algorithm on Middlebury evaluation (ver. 3). Lee et al. [12] is a recent local method that improves on cost aggregation. Miyata et al. [29] and our previous work [30] are multi-view algorithms that perform disparity estimation and denoising simultaneously. As can be observed, existing stereo matching algorithms lack robustness when images are contaminated by noise, though most of them produce excellent disparity map on noiseless images. In comparison, the proposed algorithm shows decent error rate in noisy conditions. Similar improvements can also be observed under higher noise levels with performance degradation for all methods, but will not be shown here due to page limit.

In Fig. 4, the proposed method is compared with other stereo and multi-view disparity estimation algorithms on “Tsukuba” dataset. Most stereo algorithms are severely underperforming due to noise interference. Our previous work shows improving robustness but suffers from bleeding artifacts near object boundaries caused by occlusions. The proposed method suppressed such artifacts thanks to the view selection procedure. Also, with the patch size variation, the flat regions are much smoother in our proposed disparity map.

4. CONCLUSION

In this paper, we have proposed a multi-view disparity map estimation algorithm that is robust to noise. By constructing the 3DFIS, with the proposed texture-based view selection and patch size variation scheme, our method is also able to handle the occlusion problem and ambiguity issue in low texture regions, while mitigating the degradation of estimation accuracy caused by noise. In the future, we would like to investigate different noise types other than Gaussian noise.

5. ACKNOWLEDGEMENT

This work was supported by the U.S. National Science Foundation through the Cyber-Physical System program under grant number CNS-1329481.

	Tsukuba	Barn	Cones	Venus
<i>Kolmogorov et al.</i>	79.38	65.66	72.89	73.66
<i>Klaus et al.</i>	43.77	36.89	40.03	50.17
<i>Tani ai et al.</i>	64.84	48.37	61.82	51.06
<i>Lee et al.</i>	68.58	53.75	60.12	68.93
<i>Miyata et al.</i>	53.02	52.68	61.33	60.70
<i>Zhou et al.</i>	37.26	43.33	49.52	54.06
<i>Proposed</i>	19.54	18.28	25.05	21.29

Table 1. Error percentage (%) comparison when $\sigma = 20$

6. REFERENCES

- [1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, no. 1, pp. 7–42, 2002.
- [2] S. Birchfield and C. Tomasi, "A pixel dissimilarity measure that is insensitive to image sampling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 4, pp. 401–406, 1998.
- [3] K. J. Yoon and I. S. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 650–656, Apr. 2006.
- [4] L. De-Maezta, A. Villanueva, and R. Cabeza, "Stereo matching using gradient similarity and locally adaptive support-weight," *Pattern Recognit. Letters*, vol. 32, no. 13, pp. 1643–1651, Oct. 2011.
- [5] Y. S. Heo, K. M. Lee, and S. U. Lee, "Robust stereo matching using adaptive normalized cross-correlation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 4, pp. 807–822, Apr. 2011.
- [6] J. Jiao, R. Wang, W. Wang, S. Dong, Z. Wang, and W. Gao, "Local stereo matching with improved matching cost and disparity refinement," *IEEE Multimedia*, vol. 21, no. 4, pp. 16–27, Oct./Dec. 2014.
- [7] Y. Zhan, Y. Gu, K. Huang, C. Zhang, and K. Hu, "Accurate image-guided stereo matching with efficient matching cost and disparity refinement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 9, pp. 1632–1645, Sep. 2016.
- [8] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: theory and experiment," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 9, pp. 920–932, Sep. 1994.
- [9] Y. Boykov, O. Veksler, and R. Zabih, "A variable window approach to early vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 12, pp. 1283–1294, Dec. 1998.
- [10] K. Zhang, J. Lu, and G. Lafruit, "Cross-based local stereo matching using orthogonal integral images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 7, pp. 1073–1079, Jul. 2009.
- [11] K. Zhang, Y. Fang, D. Min, L. Sun, S. Yang, and S. Yan, "Cross-scale cost aggregation for stereo matching," in *Proc. IEEE Int. Conf. ICCV*, pp. 1590–1597, 2014.
- [12] S. Lee, J. H. Lee, J. Lim, and I. H. Suh, "Robust stereo matching using adaptive random walk with restart algorithm," *Image and Vision Computing*, vol. 37, pp. 1–11, May. 2015.
- [13] L. Li, X. Yu, S. Zhang, X. Zhao, and L. Zhang, "3D cost aggregation with multiple minimum spanning trees for stereo matching," *Applied Optics*, vol. 56, no. 12, pp. 3411–3420, Apr. 2017.
- [14] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computer vision," *Journal of the American Statistical Association*, vol. 82, no. 397, pp. 76–89, 1987.
- [15] S. T. Barnard, "Stochastic stereo matching over scale," *Int. J. Comput. Vis.*, vol. 3, no. 1, pp. 17–32, 1989.
- [16] P. F. Felzenszwalb and D. R. Huttenlocher, "Efficient belief propagation for early vision," in *Proc. IEEE Conf. CVPR*, vol. 1, pp. 3017–3024, Jun. 2004.
- [17] A. Klaus, M. Sormann, and K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure," in *Proc. IEEE Int. Conf. Pattern Recognit.*, vol. 3, pp. 15–18, 2006.
- [18] M. Mozerov and J. van Weijer, "Accurate stereo matching by two step global optimization," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1153–1163, Mar. 2015.
- [19] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," in *Proc. IEEE Int. Conf. ICCV*, vol. 2, pp. 508–515, 2001.
- [20] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximation energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.
- [21] T. Taniyai, Y. Matsushita, Y. Sato, and T. Naemura, "Continuous stereo matching using local expansion moves," *arXiv preprint arXiv:1603.08328*, Mar. 2016.
- [22] S. Birchfield and C. Tomasi, "Depth discontinuity by pixel-to-pixel stereo," in *Proc. IEEE Int. Conf. ICCV*, pp. 1073–1080, Jan. 1998.
- [23] J. Zbontar and Y. LeCun, "Stereo matching by training a convolutional neural network to compare image patches," *J. Mach. Learn. Res.*, vol. 17, no. 1–32, Jan. 2016.
- [24] S. Drouyer, S. Beucher, M. Bilodeau, M. Moreaud, and L. Sorbier, "Sparse stereo disparity map densification using hierarchical image segmentation," in *International Symposium on Mathematical Morphology and Its Application to Signal and Image Processing*, pp. 172–184, May. 2017.
- [25] J. Barron and B. Poole, "The fast bilateral solver," in *Euro. Conf. Comput. Vision (ECCV)*, pp. 617–632, Oct. 2016.
- [26] S. Tulyakov, A. Ivanov, and F. Fleuret, "Weakly supervised learning of deep metrics for stereo reconstruction," in *Proc. IEEE Conf. ICCV*, pp. 1339–1348, Oct. 2017.
- [27] N. Anantrasirichai and C. N. Canagarajah, "Dynamic programming for multi-view disparity/depth estimation," in *Proc. IEEE Int. Conf. ICASSP*, vol. 2, pp. II-II, 2006.
- [28] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 606–619, Mar. 2014.
- [29] M. Miyata, K. Kodama, and T. Hamamoto, "Fast multiple-view denoising based on image reconstruction by plane sweeping," in *IEEE Conf. Visual Commun. Image Process.*, pp. 462–465, Dec. 2014.
- [30] S. Zhou, Y. H. Hu, and H. Jiang, "Multiple view image denoising using 3D focus image stacks," in *IEEE Global Conf. Signal Info. Process. (GlobalSIP)*, pp. 1052–1056, 2015.
- [31] K. Takahashi and T. Naemura, "Layered light-field rendering with focus measurement," *Signal Processing: Image Communication*, vol. 21, no. 6, pp. 519–530, 2006.
- [32] S. B. Kang, R. Szeliski, and J. Chai, "Handling occlusions in dense multi-view stereo," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, vol. 1, pp. I-103, Jun. 2001.
- [33] D. Scharstein and R. Szeliski, *Middlebury Stereo Evaluation [Online]*. Available: <http://vision.middlebury.edu/stereo/data/>