EFFICIENT BINOCULAR STEREO MATCHING BASED ON SAD AND IMPROVED CENSUS TRANSFORMATION

YUN ZHANG[1], WENXIANG CHEN[2], HAN LIU[3], JINHUA LIU[4], HUI DU[5]

[1]Institute of Zhejiang Radio and TV Technology, Communication University of Zhejiang, Hangzhou 310018, China
[2]School of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China
[3]School of Computer Science and Informatics, Cardiff University, Cardiff CF24 3AA, UK
[4]School of Electronics and Information, Communication University of Zhejiang, Hangzhou 310018, China
[5]School of New Media, Communication University of Zhejiang, Hangzhou 310018, China
E-MAIL: zhangyun@cuz.edu.cn, 731122137@qq.com, liuh48@cardiff.ac.uk, 20020186@cuz.edu.cn, duhui@cuz.edu.cn

Abstract:

Binocular stereo matching aims to obtain disparities from two very close views. Existing stereo matching methods may cause false matching when there are much image noise and disparity discontinuities. This paper proposes a novel binocular stereo matching algorithm based on SAD and improved Census transformation. We first perform improved Census transformation, and then get the matching costs by combining SAD and improved Census transformation. Finally we cluster the matching costs and calculate the disparities. To generate better disparities, we further propose the improved bilateral and selective filters to enhance the accuracy of disparities. Experimental results show that our binocular stereo matching can produce more accurate and complete disparities, and works well in complex scenes with irregular shapes and more objects, thus has wide applications in stereoscopic image processing.

Keywords:

stereo matching; disparities; SAD; Census transformation; bilateral and selective filters

1. Introduction

Nowadays, binocular vision has wide applications, such as human face recognition, object tracking and virtual reality (VR). Stereo matching, which aims to obtain the depth information from left and right image pairs shot by binocular cameras, is a key technology in binocular vision. Stereo matching has been studied for many years, and many excellent algorithms have been proposed. Most stereo matching algorithms are based on the similarity, epipolar, uniqueness, continuity and ordering constraints, and include the following steps: (1) matching cost computation; (2) matching cost aggregation; (3) disparity computation; (4) disparity refinement. The matching cost is decided by the differences of gray values of corresponding pixels in left and right images.

Traditional stereo matching methods, such as Sum of absolute difference (SAD) and sum of squared difference (SSD), which are designed for simple scenes, cannot process textureless images, and are sensitive to the light variations and noises. In contrast, normalized cross correlation (NCC) can better resist noises, and would not be affected by the light variations, but this method is computation-intensive. To cope with complex scenes and light variations, census transformation is proposed. Although this method is successful in texture-less images, it depends on the central pixel of the template, and thus the matching results may be degraded when the central pixel is affected by noises.

In this paper, we propose a novel method for more accurate and efficient stereo matching. In particular, we first perform census transformation in the left and right images; then we calculate the matching cost by combining SAD and the improved census transformation; finally, we cluster the matching cost and calculate the disparities. To further improve the quality and accuracy of disparities, we propose a post-processing method based on the improved bilateral filtering, which combines calculated disparities and original RGB images. In our method, the combination of disparity and color information can effectively solve the information loss problem in traditional bilateral filtering, and thus can obtain complete and accurate disparities. To fill in the holes after the bilateral filtering, we further propose a selective filtering approach, which is based on the

analysis of the filtered disparity histograms. Since the disparity calculation and filtering of each pixel are independent, we apply OpenMP for multi-thread parallel processing on CPU, which can save much time.

The remainder of this paper is organized as follows. Section 2 gives a brief summary of related work. In Section 3, we present the detailed algorithm of our stereo matching. Section 4 shows results and comparisons. Finally, we conclude this paper in Section 5.

2. Related work

Binocular stereo matching was first proposed by Robert [1], who applied computer vision approaches in 3D scenes. Following its proposal, stereo matching was studied further by an increasing number of scholars, leading to more advanced algorithms being proposed. In general, stereo matching can be classified into 2 categories: local stereo matching and global stereo matching. Local matching refers to the methods using different filters. Yoon et al [2] introduced a bilateral filter to stereo matching. With the adaptive weights, their method can be used to effectively improve the matching accuracy, but cannot ensure the efficiency. Hosni et al. [3] proposed a generic and fast cost-filtering framework for more efficient stereo matching. Yang [4, 5] proposed a non-local solution for matching cost aggregation and recursive bilateral filtering, which improve the matching accuracy and efficiency. Mei et al. [6] and Yao et al. [7] proposed a segment-tree based cost aggregation method for non-local stereo matching, which leads to advances in both disparity accuracy and processing speed. Zhang [8] proposed a cross-scale framework to improve the cost aggregation for accurate stereo matching. Cigla et al. [9] presented edge-aware recursive filters(REAF) for accurate and efficient stereo matching.

For the global stereo matching method, the disparities are calculated by minimizing a global energy function. Birchfield et al. [10] proposed an algorithm to detect depth discontinuities from stereo image pairs. Their method can handle large untextured regions and accelerate the dynamic programming. Hong et al. [11] proposed a new segmentation-based stereo matching using graph cuts [12], which is used to achieve the optimal solution by assigning disparity plane to each segment. MozerovW et al. [13] proposed to combine local cost-filtering and global energy minimization methods to improve the overall stereo matching by a two-step energy minimization algorithm using the MRF models. Their method can be used to effectively solve the stereo matching problem in occlusion regions. Zbontar et al. [14] applied the convolution neural network approach to predict the image patches matching, and used it to compute the stereo matching cost, which was further refined by cross-based cost aggregation and semi-global matching. Luo et al. [15] proposed a deep learning network to efficiently produce accurate results on GPU. Using the semi-global matching approach, Seki et al. [16] proposed a learning based penalties estimation method to predict accurate dense disparity map.

To evaluate the performance of stereo matching, Daniel et al. [17] proposed a systematic theoretic framework for stereo matching, and constructed the *Middlebury* testing platform, which has been widely used in stereo matching evaluation.

Although previous methods can efficiently produce accurate disparities in stereo matching, it not easy to implement them and the implementations may be failed for complex scenes. In addition, the learning based methods are not robust but depend on the training data. In this paper, we propose a robust and efficient algorithm based on SAD and improved census transformation. Our method is easy to implement and our results are very comparable to the ones obtained by using the state-of-the-arts methods on the public data sets.

3. Algorithm

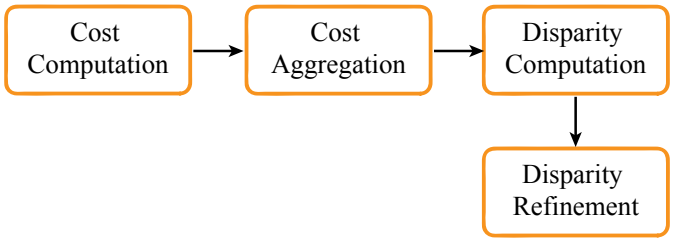

FIGURE 1. flowchart of stereo matching

As shown in Fig. 1, the binocular stereo matching framework consists of the 4 steps. In the cost computation step, the matching cost is decided according to the differences of gray corresponding pixels. Cost aggregation mainly refers to the filtering of the matching costs. In the disparity computation step, the disparity of each pixel is selected in a defined disparity range to minimize the matching costs. Disparity refinement aims to rectify the incorrect the disparities obtained by stereo matching.

In this paper, we propose to improve the accuracy and robustness of stereo matching by combining the SAD and improved census transformation. In particular, we first perform the census transformation with left and right images; then we calculate the matching costs based on SAD and improved census transformation; finally, we perform aggregation of the matching cost

and calculate the disparities. Details of the algorithm are provided in the following subsections.

3.1 Improved Census transformation

Traditional census transformation overly depends on the central pixel, and the incorrect matching rate is high on regions with discontinuous and single texture. To solve this problem, we further add the spatial information to express the disparities, and they are likely to share similar textures when the distance and the gray value between a central pixel and its neighbors are low. In this paper, we propose the weighted gray average of neighboring pixels for stereo matching, see Eq. (1).

$$\bar{I}_p = \frac{1}{w_p} \sum_{q 2 N_p} I_q s(p;q) c(I_p; I_q) \tag{1}$$

where

$$s(p;q) = exp(\frac{jjp \quad qjj^2}{2 \ s}); c(I_p; I_q) = exp(\frac{jjI_p \quad I_q jj^2}{2 \ c}); \tag{2}$$

p is the central pixel, and q is its neighboring pixels. $s(\ ); c(\ )$ are Gaussian functions, which determine the spatial and color differences between neighboring pixels. w_p is the normalized parameter. For central pixel p in regions with a single texture, I_p and $\bar{I}_p$ are similar, while for central pixel p in regions with discontinuous textures, I_p and $\bar{I}_p$ are very different. when the central pixel is destroyed by noises, we compare neighboring pixels(I_q) of I_p with $\bar{I}_p$ to reduce the impact of noise.

In our improved census transformation, each neighboring pixel is represented by 2 bits, which can better express the disparity variations. The equations are as follows.

$$(p;q) = \begin{cases} 11; & \bar{I}_p < I_q \quad and \quad (p) \quad "(p) \\ 10; & \bar{I}_p < I_q \quad and \quad (p) > "(p) \\ 01; & \bar{I}_p \quad I_q \quad and \quad (p) \quad "(p) \\ 00; & \bar{I}_p \quad I_q \quad and \quad (p) > "(p) \end{cases}; \tag{3}$$

where

$$(p) = jI_p \quad \bar{I}_p j; "(p) = \frac{\sum_{q 2 N_p} jI_q \quad \bar{I}_p j}{num(N_p)}; \tag{4}$$

$"(p)$ is the threshold for different textures, and $num(\ )$ is the number of neighboring pixels. Fig. 2 gives examples of our improved census transformation, and the sequences are used to measure the similarity of pixels from the left and right views.

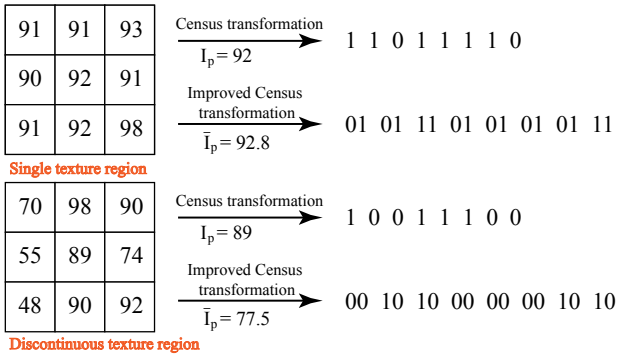

FIGURE 2. Improved Census transformation

3.2 Stereo matching based on SAD and Improved Census transformation

Traditional SAD method cannot be used to deal with images with weak textures, and might be disturbed by noises. Compared with SAD, census transformation can be used to solve the problem of weak textures, but may fail to process images with repetitive or similar textures. In this paper, we propose a novel stereo matching by combining SAD and improved census transformation, which can lead to effective reduction of incorrect matching in challenging cases, and the cost function is defined as follows.

$$C(p;d) = \ _{rCensus} C_{rCensus}(p;d) + \ _{SAD} C_{SAD}(p;d); \tag{5}$$

where

$$\begin{cases} C_{rCensus}(p;d) = Ham(T(p); T(p_d)) \\ C_{SAD}(p;d) = jI_l(p) \quad I_r(p_d)j \end{cases}; \tag{6}$$

$C_{rCensus}(\ )$ refers to the cost of improved census transformation, and $C_{SAD}(\ )$ is the cost based on the SAD method. p is a pixel in the left image, and p_d is the corresponding pixel in the right image with d as the disparity value. $Ham(\ )$ is the Hamming distance between the Census sequences of two pixels from the left and right views

3.3 Disparity Refinement

Disparities obtained in Section 3.2 may contain noises leading to loss of information, and cannot preserve the edges of objects in the scene, as shown in Fig. 3. In this section, we further refine the disparities to improve the accuracy and quality of stereo matching.

Improved bilateral filtering

We improve the bilateral filtering by combining the disparity and RGB information, definitions are shown as follows.

$$\widetilde{I}_p = \frac{1}{k_p} \sum_{q \in N_p} I_q f(p;q) g(I_p; I_q) h(I_p^c; I_q^c); \quad (7)$$

where

$$\begin{cases} \frac{1}{k_p} & = \sum_{q \in N_p} f(p;q) g(I_p; I_q) h(I_p^c; I_q^c) \\ f(p;q) & = exp(-\frac{\|p-q\|^2}{2\sigma_f^2}) \\ g(I_p; I_q) & = exp(-\frac{\|I_p - I_q\|^2}{2\sigma_g^2}) \\ h(I_p^c; I_q^c) & = exp(-\frac{\|I_p^c - I_q^c\|^2}{2\sigma_h^2}) \end{cases}; \quad (8)$$

I^c refers to the RGB image. $f(p;q), g(I_p; I_q), h(I_p^c; I_q^c)$ are the weights to measure the spacial, disparity and color similarity of neighboring pixels. The combination of RGB color and disparity information can better solve the information loss problem resulting from traditional bilateral filtering, and thus can obtain complete and accurate disparities.

Selective filtering

After the bilateral filtering, there are still some holes in the disparity map. To solve the 'hole' problem, we first draw a histogram for the disparities. Fig. 4 gives the histogram of a disparity image after bilateral filtering. We find that the number of pixels is relatively small for the gray value between 10 and 60, and pixels in this range are more likely to be hole regions. Based on the disparity histogram analysis, we propose the selective filtering approach, which only utilize valid neighboring pixels for bilateral filtering and thus can effectively complete the hole regions. The modified filter is shown as follows.

$$\widetilde{I}_p = \frac{1}{k_p} \sum_{q \in N_p; I_q \in T} I_q f(p;q) g(I_p; I_q) h(I_p^c; I_q^c); \quad (9)$$

where T is the range of valid disparities. Fig. 3 gives results of bilateral filtering and our selective filtering, which shows advantages of our method in completing hole regions. Fig. 3 shows the results of bilateral filtering, improved bilateral filtering and our method. Although the improved bilateral filtering is better than the initial bilateral filtering, there are still some holes. In comparison, our method can produce more accurate and complete disparities by combining bilateral and selective filtering.

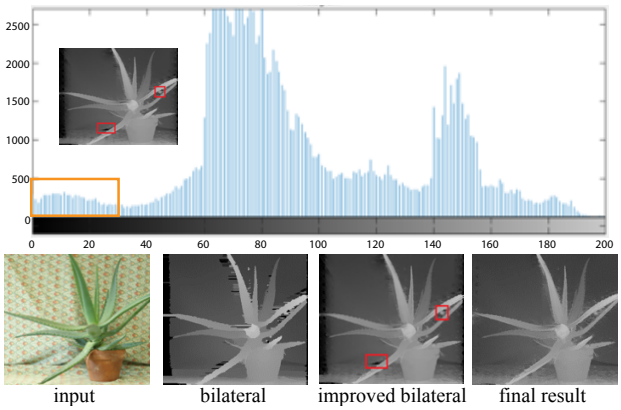

FIGURE 3. Stereo matching with selective filtering

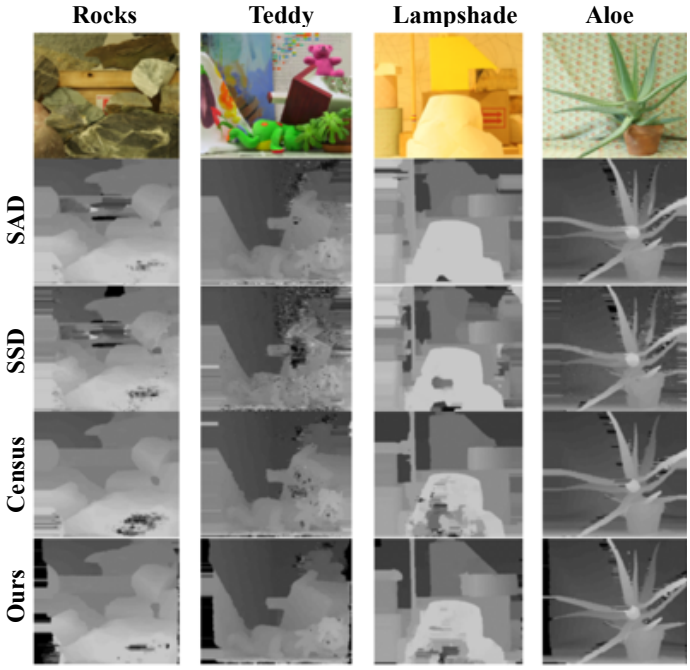

FIGURE 4. Comparisons of stereo matching

TABLE 1. Matching error rate

test examples	Different stereo matching methods			
	SAD	*SSD*	*Census*	*Ours*
Rocks1	9.80%	10.31%	7.44%	7.44%
Teddy	3.30%	3.56%	2.02%	7.44%
Lampshade1	0.68%	1.41%	0.37%	7.44%
Aloe	12.76%	12.82%	5.39%	7.44%

4 Results

We test our stereo matching algorithm using the public data set from Middlebury Computer vision page [18], which is

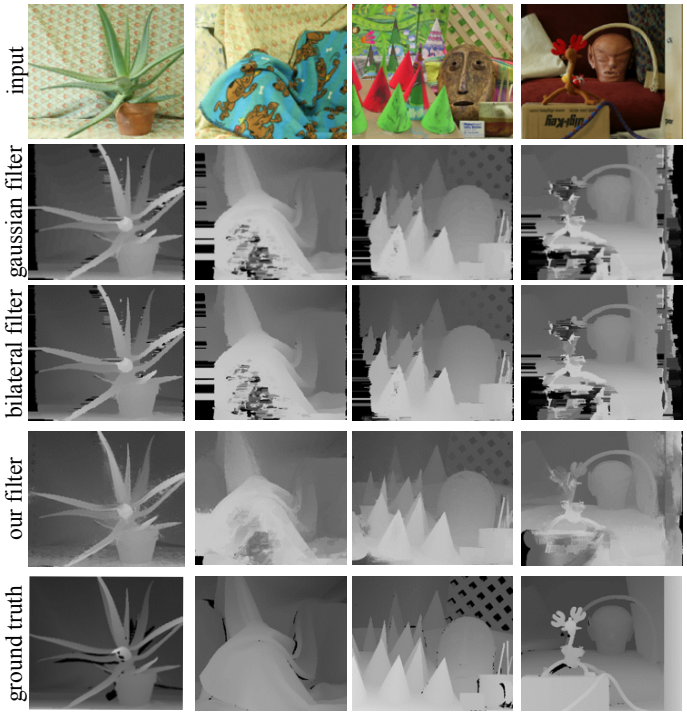

FIGURE 5. Comparisons of stereo matching after filtering

TABLE 2. Matching error rate after filtering

test examples	Comparison of different filters		
	gaussian	*bilateral*	*ours*
Aloe	9.80%	10.31%	7.44%
Cloth3	3.30%	3.56%	2.02%
Cones	0.68%	1.41%	0.37%
Reindeer	12.76%	12.82%	5.39%

widely used by previous stereo matching methods. Fig. 4 shows results of stereo matching by different methods. The first line shows the input RGB images, and the following lines show stereo matching results obtained by using SAD, SSD, Census and our method. Compared with other methods, the disparities obtained by using our method is more accurate, especially for regions with details. For regions with low textures and discontinuous disparities, our method is robust and can well preserve edges. Table 1 shows further quantitative comparisons, and our method has lower matching error rate than other methods.

Fig. 5 shows results of stereo matching after filtering by different filters. The first line is the input RGB images, and the following lines provide filtering results by the Gaussian filter, the bilateral filter and our filter. Compared with traditional filtering methods, our method can better preserve edges, complete disparity holes, and thus can obtain more accurate disparities, which are similar to the ground truth (See the last line of Fig. 5). Table 2 shows quantitative results of stereo matching after filtering. Compared with other methods, our method can better complete the disparity holes with much fewer matching errors, and works well in complex scenes with irregular shapes and more objects. We also tested our method in many examples of the data sets and complex scenes shot by ourselves, and the results are satisfactory. The main advantage of our method is that it shows the effectiveness and robustness of obtaining performance that is comparable to the one obtained using the state-of-the-art methods.

5 Conclusions

In this paper, we have proposed a novel algorithm for stereo matching based on SAD and improved census transformation. To reduce noises and holes in the disparities, we have further proposed improved bilateral and selective filters to refine the stereo matching results. Results and comparisons show that the disparities obtained by our method are more accurate, even in challenging cases, such as regions with low textures, discontinuous disparities and irregular shapes. Our stereo matching is efficient and easy to implement, and can be directly applied in stereo image/video processing and editing as a pre-processing step.

In the future, we will further study the stereo matching to adapt to more complex scenes and improve the matching accuracy. For real-time applications in stereo images/videos, we aim to accelerate the stereo matching through GPU optimization.

Acknowledgements

This work was supported by National Natural Science Foundation of China (61602402), Zhejiang Provincial Basic Public Welfare Research(LGG19F020001, 2017C3163, 2017C33167).

References

[1] L. G. Roberts, *Machine Perception of Three-Dimensional Solids*, ser. Outstanding Dissertations in the Computer Sciences. Garland Publishing, New York, 1963.

[2] 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.

[3] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 504–511, 2013.

[4] Q. Yang, "A non-local cost aggregation method for stereo matching," in *2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, June 16-21, 2012*, 2012, pp. 1402–1409.

[5] ——, "Recursive bilateral filtering," in *Computer Vision - ECCV 2012 - 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part I*, 2012, pp. 399–413.

[6] X. Mei, X. Sun, W. Dong, H. Wang, and X. Zhang, "Segment-tree based cost aggregation for stereo matching," in *2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, June 23-28, 2013*, 2013, pp. 313–320.

[7] P. Yao, H. Zhang, Y. Xue, M. Zhou, G. Xu, Z. Gao, and S. Chen, "Segment-tree based cost aggregation for stereo matching with enhanced segmentation advantage," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2017, New Orleans, LA, USA, March 5-9, 2017*, 2017, pp. 2027–2031.

[8] K. Zhang, Y. Fang, D. Min, L. Sun, S. Yang, and S. Yan, "Cross-scale cost aggregation for stereo matching," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 27, no. 5, pp. 965–976, 2017.

[9] C. Cigla, "Recursive edge-aware filters for stereo matching," in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2015, Boston, MA, USA, June 7-12, 2015*, 2015, pp. 27–34.

[10] S. Birchfield and C. Tomasi, "Depth discontinuities by pixel-to-pixel stereo," *International Journal of Computer Vision*, vol. 35, no. 3, pp. 269–293, 1999.

[11] L. Hong and G. Chen, "Segment-based stereo matching using graph cuts," in *2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004), with CD-ROM, 27 June - 2 July 2004, Washington, DC, USA*, 2004, pp. 74–81.

[12] Y. Boykov and M. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images," in *ICCV*, 2001, pp. 105–112.

[13] M. G. Mozerov and J. van de Weijer, "Accurate stereo matching by two-step energy minimization," *IEEE Trans. Image Processing*, vol. 24, no. 3, pp. 1153–1163, 2015.

[14] J. Zbontar and Y. LeCun, "Computing the stereo matching cost with a convolutional neural network," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, 2015, pp. 1592–1599.

[15] W. Luo, A. G. Schwing, and R. Urtasun, "Efficient deep learning for stereo matching," in *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, 2016, pp. 5695–5703.

[16] A. Seki and M. Pollefeys, "Sgm-nets: Semi-global matching with neural networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, 2017, pp. 6640–6649.

[17] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, no. 1-3, pp. 7–42, 2002.

[18] D. Scharstein, R. Szeliski, and H. Hirschmller, "Middlebury stereo vision page," http://vision.middlebury.edu/stereo/.