A Stereo Matching Algorithm for Vehicle Binocular System

Fangyi Zhang, Gang Zhao(✉), Haiying Liu, and Wang Qin

College of Transportation, Shandong University of Science and Technology, Qingdao, China
479943377@qq.com

Abstract. In order to improve outdoor performance of vehicle binocular system, the stereo matching algorithm based on "3bit-Census Transformation & An Adaptive window aggregation based on edge truncation & Fast Parallax Calculation" was proposed. The stereo matching algorithm based on this framework improved the robustness, matching accuracy and efficiency of the calculation from different stages. The experimental results show that the algorithm proposed in this paper is better than the traditional algorithm and can meet the requirements of the vehicle binocular system.

Keywords: Binocular system · Stereo matching · Robustness · Matching accuracy · Calculation efficiency

1 Introduction

With the rapid development of computer and sensor technology, 3D reconstruction based on binocular system has been widely applied to intelligent vehicle positioning, obstacle measurement and other fields [1]. The results of stereo matching can directly affect the robustness, accuracy and efficiency of binocular system [2]. At present, the binocular stereo matching algorithm is divided into three categories: (1) global stereo matching; (2) semi-global stereo matching; (3) local stereo matching [3]. Global stereo matching and semi-global stereo matching have high accuracy with high complexity, so they are not suitable for intelligent vehicles. But conversely, although local stereo matching has relatively low accuracy, it has relatively low complexity as well, therefore, it is easy to be realized by vehicle binocular system [4].

Local stereo matching is mainly divided into four stages: (1) Cost calculation; (2) Cost aggregation; (3) Parallax calculation; (4) Parallax refinement and post-processing. Through the above stages, dense parallax images can be obtained from the left and right images collected by binocular system. The 3D model of driving environment can be reconstructed by combining the parallax image with the parameters of binocular camera. In order to strengthen the robustness, accuracy and efficiency of the vehicle binocular systems, a stereo matching framework based on "3bit-Census transformation & An adaptive window aggregation based on edge truncation & Fast parallax calculation" is proposed.

S. Liu and G. Yang (Eds.): ADHIP 2018, LNICST 279, pp. 514–523, 2019.
https://doi.org/10.1007/978-3-030-19086-6_57

2 3bit-Census Based on Average Confidence Interval

In the traditional Census transformation, the result is strongly dependent on the center pixel [5]. Once the center pixel is seriously interfered by the noise, the result of the Census transformation will change significantly, and the matching accuracy will be reduced. What's more, the relative gray values of pixels and center pixel are represented by 1-bit binary number, and the gray value space is divided into "2-gray value spaces (0 or 1)". This transformation method has poor ability to describe local features and poor distinction between different local features. In this paper, a 3bit-Census transformation based on the confidence interval is proposed. The steps of the calculation are as follows:

(1) The pixel value confidence interval and the pixel average value in the interval are calculated, which is used as the reference pixel of the 3bit-Census transformation.
(2) The traditional "2-gray value spaces" is divided into "8-gray value spaces" by using the logic combination of 3-bit binary numbers.

2.1 Pixel Average Value Calculation Based on Confidence Interval

In order to solve the problems that the transformation is highly dependent on the center pixel and the center pixel is susceptible to noise, the use of the pixel, which is seriously affected by the noise in the calculation should be reduced.

Assuming that the size of the transformation window is 'n * n', the pixel value $I_{(1,1)}, I_{(1,2)}, I_{(1,3)} \cdots I_{(\mathrm{n},n)}$ in the window is normal distribution from the average value $\overline{I}(o)$, and the standard deviation is σ. (i.e. $I \sim N(\overline{I}(o), \sigma^2)$)

Confidence interval is:

$$\left[\overline{I}(o) - \frac{\sigma}{n} Z_{\alpha/2}, \overline{I}(o) + \frac{\sigma}{n} Z_{\alpha/2}\right] \tag{1}$$

The reference pixel of the transformation based on confidence interval is:

$$I'(\mathrm{o}) = \frac{\sum_{i=1}^{m} I_i}{m}; I_i \in \left[\overline{I}(o) - \frac{\sigma}{n} Z_{\alpha/2}, \overline{I}(o) + \frac{\sigma}{n} Z_{\alpha/2}\right] \tag{2}$$

In this paper: $\alpha = 0.01, Z_{\alpha/2} = 2.68$

The confidence interval can effectively eliminate the pixel points that is seriously affected by the noise, and use the high quality pixel points in the confidence interval to calculate the average value as the transformation reference. The above method can effectively improve the robustness and the ability of describing local features.

2.2 3Bit-Census Transformation

In the process of 3bit-Census transformation, considering the amount of logic resources in the vehicle hardware platform, the "2-gray value spaces" of [0: $I'(o)$] and [$I'(o)$: 255] are reasonably extended to "8-gray value spaces". The method can be realized by

shift operation and adder, and it can reduce the amount of logic resource occupied by vehicle hardware platform effectively. The 3bit-Census transformation is shown in Eq. (3).

$$\beta(h,o) = \begin{cases} 111, & I(h) \geq 2I'(\mathrm{o}) \\ 110, & 2I'(\mathrm{o}) > I(h) \geq \frac{3}{2}I'(\mathrm{o}) \\ 101, & \frac{3}{2}I'(\mathrm{o}) > I(h) \geq \frac{5}{4}I'(\mathrm{o}) \\ 100, & \frac{5}{4}I'(\mathrm{o}) > I(h) \geq I'(\mathrm{o}) \\ 011, & I'(\mathrm{o}) > I(h) \geq \frac{3}{4}I'(\mathrm{o}) \\ 010, & \frac{3}{4}I'(\mathrm{o}) > I(h) \geq \frac{1}{2}I'(\mathrm{o}) \\ 001, & \frac{1}{2}I'(\mathrm{o}) > I(h) \geq \frac{1}{4}I'(\mathrm{o}) \\ 000, & I(h) < \frac{1}{4}I'(\mathrm{o}) \end{cases} \tag{3}$$

To verify the ability of the 3bit-Census transformation to describe local features, Fig. 1 shows two different windows: 'a' and 'b'. Equation (4) is the bit-string obtained by the Census transformation and the 3bit-Census transformation. Equation (5) is the hamming distance between the bit-strings.

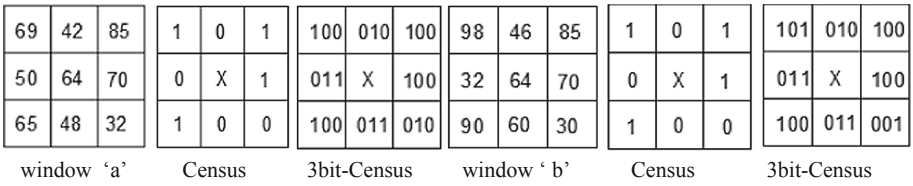


Fig. 1. The calculation process of bit string

$$\begin{cases} \mathrm{C(cen\_a)} = \{1,0,1,0,1,1,0,0\} \\ \mathrm{C(cen\_b)} = \{1,0,1,0,1,1,0,0\} \\ \mathrm{C(3bit\_cen\_a)} = \{1,0,0;0,1,0;1,0,0;0,1,1;\ 1,0,0;1,0,0;0,1,1;0,1,0\} \\ \mathrm{C(3bit\_cen\_b)} = \{1,0,1;0,1,0;1,0,0;0,1,0;\ 1,0,0;1,0,0;0,1,1;0,0,1\} \end{cases} \tag{4}$$

$$\begin{cases} \mathrm{Hamming[C(cen\_a), C(cen\_b)]} = 0 \\ \mathrm{Hamming[C(3\,bit\_cen\_a), C(3\,bit\_cen\_b)]} = 4 \end{cases} \tag{5}$$

The above calculation shows that the hamming distance between two bit-strings may be 0 after Census transformation of different windows, which cannot judge the difference between two windows, and then lead to mismatch. After 3bit-Census transformation, the hamming distance is 4, which can correctly judge the difference between different windows. It can be concluded that the stereo matching method based on 3bit-Census transformation can effectively improve the description ability of local features, the credibility of similarity measurement and the matching accuracy.

3 An Adaptive Window Aggregation Based on Edge Truncation

In the cost aggregation stage, the size and shape of the aggregation window plays a decisive role in the matching accuracy. If the aggregation window is too small, it cannot contain enough local information in the weak texture, so it is easy to cause mismatch [6]. If the aggregation window is too large, in the parallax discontinuous region or the edge of the object, the excessively large window would contain more pixel points with different depths, as a result, it is easy to cause mismatch too. In order to improve the cost aggregation level, the pixel value in the aggregation window should have similar depth, and the size and shape of the aggregation window should be adjusted adaptively according to the local features.

3.1 Sobel Edge Detection

Sobel operator is a first-order derivative edge detection operator [7]. It uses two convolution check images of 3 * 3 to process the convolution, and thus obtains the grayscale gradient: G_{x}, G_y and G, which is shown in Eqs. (6) (7) (8) (9).

$$G_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} * A \quad \text{and} \quad G_{\mathrm{y}} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * A \tag{6}$$

$$\begin{aligned} G_{\mathrm{x}} = {} & [f(x-1,y-1) + 2f(x-1,y) + f(x-1,y+1)] \\ & - [f(x+1,y+1) + 2f(x+1,y) + f(x+1,y-1)] \end{aligned} \tag{7}$$

$$\begin{aligned} G_{\mathrm{y}} = {} & [f(x-1,y+1) + 2f(x,y+1) + f(x+1,y+1)] \\ & - [f(x-1,y-1) + 2f(x,y-1) + f(x+1,y-1)] \end{aligned} \tag{8}$$

$$G = \sqrt{G_x^2 + G_y^2} \tag{9}$$

If $G > G_{\max}$, the point is one on the edge. Compared with other algorithms, Sobel algorithm has the advantages of simple calculation structure and easy implementation of vehicle hardware system. According to the above principle, the image collected by the vehicle binocular system are detected by Sobel operator. The detection results are shown in Fig. 2. The edge information of the object can ensure all the pixel value in the aggregation window having similar depth.

3.2 An Adaptive Window Based on Edge Truncation

In order to realize that the size and shape of the aggregation window being adjusted adaptively according to the local characteristics, an adaptive aggregation window is proposed in this paper. The stages to create the adaptive aggregation window are as follows:

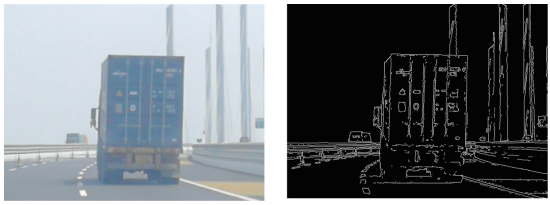
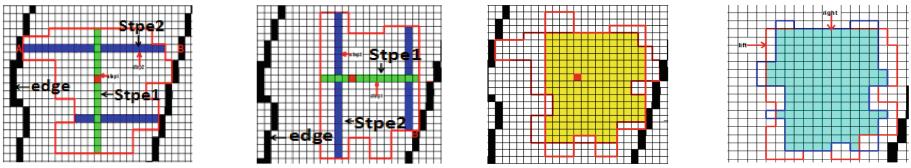


Fig. 2. Sobel edge detection

(1) Aggregating form column to row: (1) Taking the pixel point 'o' to be the starting point, the constraint condition and the edge truncation condition are used to determine the upper and lower endpoints in the vertical direction; (2) Starting from each pixel of the vertical pixel interval, the constraint conditions and the edge truncation condition determine the left and right endpoints. The aggregation result is shown in Fig. 3(a). Constraints and edge truncation conditions are shown in Eq. (10).



(a) column to row (b) row to column (c) single image (d) final window

Fig. 3. Aggregate window generation diagram

(2) According to the method of stage (1), the result of aggregation form row to column is shown in Fig. 3(b).
(3) Take the intersection of the above aggregation windows and determine the window for the single image, which is shown in Fig. 3(c).
(4) Taking the intersection of the aggregation windows in the left and right images to get the final aggregation window, which is shown in Fig. 3(d).

$$\begin{cases} 1.\, D_c(P,o) < \tau_1 \quad and \quad D_c(P, P-(0,1)) < \tau_1 \\ 2.\, D_s(P,o) < L_1 \\ 3.\, D_c(P,o) < \tau_2 \quad if \quad L_2 < D_s(P,o) < L_1 \\ 4.\, D(P,o) = |I(P) - I(o)| \\ 5.\, A < P < B \end{cases} \tag{10}$$

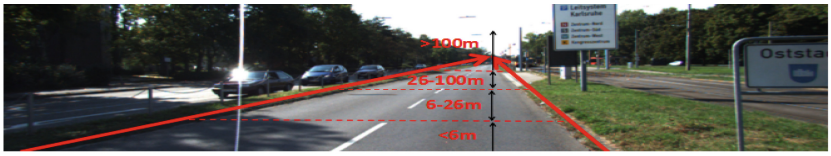
In the Eq. (10): 'o' is the central pixel; 'P' is the point to be aggregated; 'D_c' is the column difference; 'D_s' is the row difference; 'τ_1, τ_2, L_1, L_2' are the variable parameters, which are 30, 10, 15 and 8 in this paper. 'A' and 'B' are the pixel points on the edge. The matching costs is shown in Eq. (11):

$$L(o) = \sum_{i \in M} Hamming[C_{cen}(i), C_{cen}(i-d)] \tag{11}$$

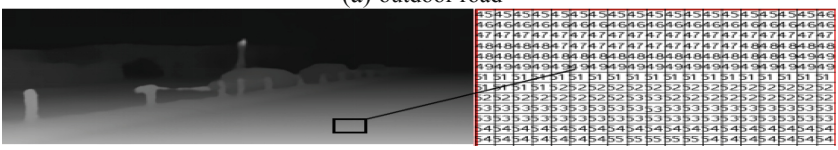
In Eq. (11): 'C_{cen}' is bit-string; 'L(o)' is cost aggregation; 'm' is the aggregation region shown in Fig. 3(d); 'i' is the pixel point in the aggregation window; 'd' is parallax.

4 Fast Parallax Calculation

The process of parallax calculation is based on the cost aggregation and determine the optimal matching point by calculating the minimum matching cost in the maximum parallax (d_{max}), which is shown in Eq. (12). Figure 4(a) shows the road image obtained by the vehicle binocular system, which has obvious depth variation. The depth of the object above the image is large and the parallax is small. The depth of the object below the image is small and the parallax is large. Figure 4(b) shows that parallax varies slightly in the row and more in the column. In other words, the depth change is mainly reflected in the column direction.



(a) outdoor road



(b) parallax variation image

Fig. 4. Outdoor road and parallax images

In the traditional parallax (d_{max}) calculation process, the parallax is fixed, which results in computational complexly and reduces the real-time performance. In order to solve the above problem, this paper proposed a method for calculating the parallax with variable maximum parallax: (1) Assuming that the image obtained at a certain time is the first frame, the initial 'd_{max}' is '150'. (2) The parallax ($d_{(x,\ y)\text{-}1}$) of the pixel points in the first image is calculated under the 'd_{max}'. (3) When the parallax of the second frame is calculated, the maximum parallax ($d_{(x,\ y)max\_2}$) of the corresponding point in the second image is shown in Eq. (13). The updating process is repeated every 5 frames.

$$J = Min\{L(j) | 0 \le j \le d_{\max}\} \tag{12}$$

$$d_{(x,y)max\_2} = 1.3 * d_{(x,y)\_1} \tag{13}$$

In Eqs. (12) (13): 'J' is the optimal matching point; L (j) is the matching cost in the range of maximum parallax.

5 Experiment

5.1 Experiment I: Robustness Analysis

In order to verify the robustness of the algorithm based on 3bit-Census, we use linear SAD algorithm and the nonlinear algorithm proposed in this paper to analyze the images collected by binocular system. The experiment used a pair of images from Middlebury [8] as processed images. There are obvious bicycle shadows in the image, and both images are 2988 * 2088 pixels. In this experiment, the maximum parallax of the two algorithms is '150'. The transformation window for 3bit-Census is 7 * 7. The image processing results are shown in Fig. 5.

(a)left image (b)right image (c) SAD algorithm (d)our nonlinear algorithm

Fig. 5. Comparison of results between linear SAD and our nonlinear algorithm

By analyzing the experiment, we can obtain the following conclusions:

(1) Based on the linear SAD algorithm, the background grayscale difference indicates that the background depth has a great change, which is not consistent with the experiment fact, on the contrary, based on the parallax image processed by the nonlinear algorithm, the background grayscale is uniform and the depth of the background is similar. By comparing with the original image, the experimental results are consistent with the experiment fact.
(2) Based on the parallax image processed by the linear SAD algorithm, the change of gray value of bicycle is poor, which indicates that the algorithm has low matching accuracy. The gray value based on our nonlinear algorithm is smooth, and the gray level of the rear wheel is consistent with the background gradually, which is consistent with the experimental fact.
(3) The image background based on linear SAD algorithm has obvious bicycle shape, which indicates that the algorithm is less robust to illumination, on the contrary, the parallax image based on our nonlinear algorithm is less affected by the shadow of the object, which shows that the algorithm is more robust to illumination.

5.2 Experiment II: Matching Accuracy Analysis

In order to analyze the matching accuracy of this algorithm, this paper uses Tsukubaer, Venuseddy, Teddy and Cones of Middlebury website to carry out off-line stereo matching experiment, and compares it with other high-quality algorithms. Off-line experiment requires 'visual studio 2012+ opencv2.4.8'. In the experiment, this paper will calculate the mismatch rate in three regions: 'nocc', 'all' and 'disc', respectively, and calculate the average mismatch rate of all images, which represents the accuracy of the algorithm. 'all' region includes half occlusion region but not image edge of object. Error threshold = 1.0. The experimental results are shown in Table 1. The results of the parallax image and the ground truth are compared as shown in Fig. 6.

Table 1. Comparison of matching accuracy between this algorithm and other algorithms

Algorithm	Tsukuba			Vensus			Teddy			Cones			Av Error
	nocc	all	disc	nocc	all	disc	nocc	all	disc	nocc	all	disc	
ADCENSUS [9]	1.07	1.48	5.7	0.09	0.25	1.15	4.10	6.22	10.9	2.42	7.25	6.95	3.97
SSCBP [10]	1.05	1.39	5.57	0.10	0.16	1.39	3.44	8.32	9.95	2.60	7.13	7.23	4.03
Our method	1.42	8.85	6.01	0.39	0.85	2.44	4.35	8.01	8.77	5.01	7.03	6.33	4.96
CrossLMF [11]	2.46	2.78	6.26	0.27	0.38	2.15	5.50	10.6	14.2	2.34	7.82	6.80	5.13
Sdds [12]	3.31	3.62	10.4	0.39	0.76	2.85	7.65	13.0	19.4	3.99	10.0	10.8	7.19
FastAggreg [13]	1.16	2.11	6.06	4.03	4.75	6.43	9.04	15.2	20.2	5.37	12.6	11.9	8.24
RTCsensus [14]	5.08	6.25	19.2	1.58	2.42	14.2	7.96	13.8	20.3	4.10	9.54	12.2	9.73
GC [15]	1.94	4.12	9.39	1.79	3.44	8.75	16.5	25.0	24.9	7.70	18.2	15.3	11.4

By analyzing the experiment, we can obtain the following conclusions:

(1) As shown in Table 1, the average error of the experiment results based on our algorithm is 4.96%, which is slightly higher than that of SSCBP and other high-quality matching algorithms. Compared with the algorithms such as CrossLMF, SDDS, FastAggreg, RTCensus and GC, the algorithm increases 0.17%, 2.23%, 3.28%, 4.77%, 6.44% respectively on average matching accuracy.

(2) As shown in Fig. 6, compared with the ground truth image, the image processed by our algorithm has a good overall effect. Compared with the SAD algorithm, the matching accuracy of this algorithm is obviously improved, the matching accuracy of the objects in the image is very high, and the edge of the object is smooth, so the algorithm proposed in this paper can meet the requirements of the vehicle binocular system better.

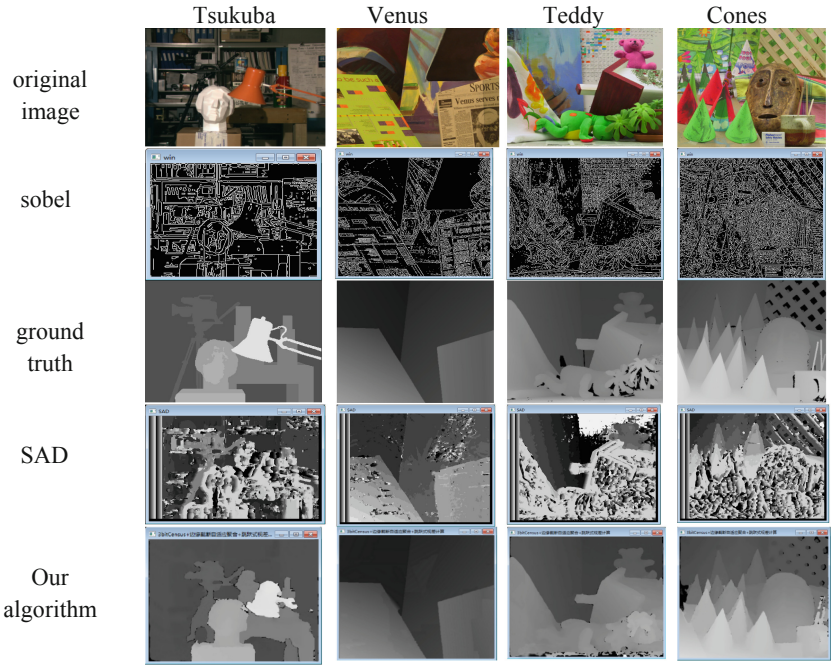


Fig. 6. Experimental parallax images and ground truth images

6 Conclusion

In this paper, based on the application background of vehicle binocular system, a stereo matching algorithm based on "3bit-Census transformation & An adaptive window aggregation based on edge truncation & Fast parallax calculation" was proposed. The algorithm has the following characteristics:

(1) The nonlinear 3bit-Census transformation has robustness to illumination and strong ability to describe local features, which lays a stable foundation for improving the stereo matching accuracy.
(2) Adaptive aggregation based on edge truncation realizes the transformation of adaptive aggregation window according to local features and edge truncation, which ensures that all pixels in the aggregation window have similar depth and improves the matching accuracy of the stereo matching.
(3) The fast parallax calculation method based on the variable maximum parallax avoids the calculation complexity, and enhances the matching efficiency and ensures the real-time performance, so the algorithm proposed in this paper can satisfy the vehicle binocular system working outdoors.

References

1. Tong, Z., Zhao, T., He, L., Wang, X.: Localization and driving speed detection for construction vehicles based on binocular vision. China Mech. Eng. **29**(4), 423–428 (2018)
2. Xiong, X., Hua, C., Fang, C., Chen, Y.: Research on free-form surface stereo matching method based on improved census transform. In: 35th Chinese Control Conference, pp. 4889–4893. IEEE, Chengdu (2016)
3. Lin, S., Yin, X., Tang, Y.: Research status and prospect of binocular stereo matching technology. Sci. Technol. Eng. **17**(30), 135–147 (2017)
4. Dehnavi, M., Eshghi, M.: FPGA based real-time on-road stereo vision system. J. Syst. Archit. **81**, 32–43 (2017)
5. Men, Y., Zhang, G., Men, C., Ma, N.: A stereo matching algorithm based on four-moded census and relative confidence plane fitting. Chin. J. Electron. **24**(4), 807–812 (2015)
6. Hirschmüller, H., Innocent, P., Garibaldi, J.: Real-time correlation-based stereo vision with reduced border errors. Int. J. Comput. Vision **47**(1–3), 229–246 (2002)
7. Kalra, A., Chhokar, R.: A hybrid approach using Sobel and canny operator for digital image edge detection. In: 2016 International Conference on Micro-Electronics and Telecommunication Engineering, pp. 305–310. IEEE, Ghaziabad (2017)
8. Middlebury Stereo Evaluation. http://vision.middlebury.edu. Accessed 9 June 2018
9. Mei, X., Sun, X., Zhou, M., Jiao, S., Wang, H., Zhang, X.: On building an accurate stereo matching system on graphics hardware. In: IEEE International Conference on Computer Vision Workshops, pp. 467–474. IEEE, Barcelona (2012)
10. Peng, Y., Li, G., Wang, R., Wang, W.: Stereo matching with space-constrained cost aggregation and segmentation-based disparity refinement. In: Three-Dimensional Image Processing, Measurement. International Society for Optics and Photonics, vol. 9393, pp. 939309-1-939309-11 (2015)
11. Do, M.N.: Cross-based local multipoint filtering. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 430–437. IEEE, Providence (2012)
12. Wang, Y., Dunn, E., Frahm, J.: Increasing the efficiency of local stereo by leveraging smoothness constraints. In: 2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission, pp. 246–253. IEEE, Zurich (2012)
13. Tombari, F., Mattoccia, S., Stefano, L., Addimanda, E.: Near real-time stereo based on effective cost aggregation. In: 19th International Conference on Pattern Recognition, pp. 1–4. IEEE, Tampa (2009)
14. Humenberger, M., Zinner, C., Weber, M., Kubinger, W., Vincze, M.: A fast stereo matching algorithm suitable for embedded real-time systems. Comput. Vis. Image Underst. **114**(11), 1180–1202 (2010)
15. Scharstein, D., Szeliski, R., Zabih, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In: Stereo and Multi-Baseline Vision, pp. 7–42. IEEE, Kauai (2002)