

Neighborhood Correlation and Window Adaptive Stereo Matching Algorithm

Yibo Du ^{1,2,3} and Kebin Jia ^{1,2,3,*}

¹ Faculty of Information Technology, Beijing University of Technology, Beijing, 100124

² Beijing Laboratory of Advanced Information Networks, Beijing, 100124

³ Beijing Key Laboratory of Computational Intelligence and Intelligent System,
Beijing University of Technology, Beijing, 100124

* kebinj@bjut.edu.cn

Received August 2019; revised October 2019

ABSTRACT. *In view of the poor noise sensitivity of traditional stereo matching algorithms and the problem of mismatching in non-smooth areas, a neighborhood-based stereo matching algorithm is proposed. Relevance and window adaptive stereo matching algorithm. Firstly, Uses weighted average sum of neighborhood cross windows to assign central pixels. Secondly, by setting adaptive threshold, according to the correlation between neighborhood pixels and central pixels in the window, it carries out secondary cost. In the process of cost aggregation, adaptive windows with changing color thresholds are used, and noise screening is carried out in this process. In the disparity refinement stage, the disparity map is further optimized by combining left-right consistency detection and regional voting. Finally, a quadratic linear interpolation operation is performed on the image. In this paper, we use the standard stereo image of Middlebury test platform to carry out experiments. The experimental results show that the proposed method can effectively reduce the image's sensitivity to noise, and the mismatch rate is lower than that of many matching algorithms.*

Keywords: Stereo matching; Parallax image; Neighborhood correlation; Cost aggregation

1. Introduction. Binocular vision technology is one of the hotspots of many scholars' research at present. It uses binocular camera [1] to shoot the scene, it can acquire the three-dimensional information of the scene.

In 1960s, Marr first combined visual computing theory with binocular stereo matching technology [2]. Local and global stereo matching methods are the mainstream algorithms nowadays. The matching algorithm of area blocks applied by Wang et al. [3]. It used fixed window and calculated the matching degree. Roy et al. applied a global matching algorithm based on graph optimization theory [4]. Tin et al. constructed and optimized the global energy function of the segmented image [5]. Wang et al. optimized the disparity map by calculating the cost values of eight directions using the semi-global matching algorithm [6]. Dinh proposes a robust matching cost function, which can adaptively construct arbitrary shape regions [7]. The global method mentioned above can improve the accuracy of disparity, but it is difficult to obtain the global energy function and refine in practice.

It is worth pointing out that Zabih et al. first applied Census to stereo matching [8]. This method can ensure high matching accuracy, but it relies too much on the central pixel. In order to solve these problems, many scholars have proposed different algorithms

based on Census method. Fan [9] added noise tolerance in the initial cost calculation stage to obtain stable cost. Zhu et al. introduced noise tolerance in Census transformation [10]. Men et al. combined the dynamic programming method in Census transformation [11]. Beak et al. combined Census cost value with the gradient change value of surrounding pixels [12]. Guo et al. adopted semi-global matching algorithm based on Census [13]. Chai et al. combines Census algorithm with traditional SAD algorithm to improve the matching accuracy [14]. However, the above methods don't consider the influence of noise, and the improvement of accuracy is limited. For the cost aggregation, Zhang et al. proposed an aggregation method based on cross-window [15]. Mei et al. analyzed the insufficiency of Zhang's work [16], and improve the matching accuracy by re-judging the arm length of cross window. But this method does not take into account the influence of noise.

Based on the analysis of above research, this paper adopts a algorithm based on neighbor-information constraints. In cost calculation, central pixel is assigned by means of window weighted mean. In aggregation stage, cost are aggregated by adaptive window. And noise elimination strategy is introduced in the process of construction. Finally, disparity is optimized by combining left-right consistency detection and regional voting.

The structure of this paper is as follows: Section 2 focus on the Census algorithm, including cost calculation, aggregation and refinement. Section 3 is the experimental results to show the effectiveness of our method.

2. Algorithmic Description.

2.1. Principle of traditional Census algorithm. Census algorithm is a region-based matching algorithm. Its main principle is to determine the size of the surrounding and central pixels in the support window. If it is larger than the central pixel, it will be marked as 1, otherwise marked as 0. Specific formulas are as follows:

$$\ell(p, p') = \begin{cases} 0, & \text{if } (p < p') \\ 1, & \text{if } (p > p') \end{cases} \quad \text{Cens}(p) = \bigotimes_{p' \in W(u,v)} \ell(p, p') \quad (1)$$

P represents the central pixel of the window, p' represents the surrounding pixel. $W(u, v)$ represents the supporting window. By comparing the neighboring pixels with the central pixel one by one, a series of 0, 1 binary numbers are obtained. By calculating the Hamming distance of the bit strings, the similarity of the two regions can be judged. The schematic diagram of the Census transform is shown in Fig. 1.

However, Census has its limitation. In the non-smooth region, pixels will change greatly. Comparing the size of the central pixel with other pixel will lead to mismatching. As shown in Fig. 2, although the neighboring pixels have changed a lot, the Hamming distances are same in two different regions.



FIGURE 1. Census Transform schematic diagram

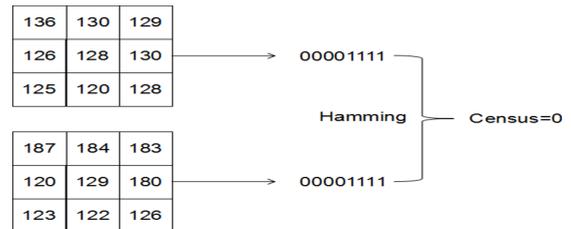


FIGURE 2. Mismatching transformation

2.2. The Algorithm in This Paper. Our methods flow chart is shown in Fig. 3

A cross-weighted template is first established to weighted average sum the central pixels in the window. As shown in Figure 4, the cross-weighted template can reduce the impact of boundary mutation on the central pixel by weighting. Re-assigning the mean value to the central pixel can reduce the dependence on the central pixel.

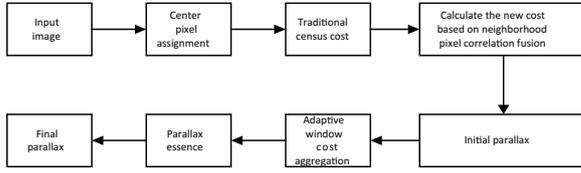


FIGURE 3. The flow chart of the algorithm in this paper

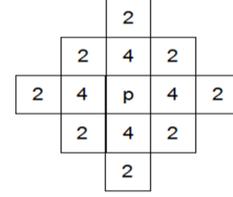


FIGURE 4. Cross-weighted template

Then, a double constraint algorithm based on neighborhood information is proposed in this paper. On the basis of Census cost calculation, a new cost is merged. The specific methods are as follows:

1) Calculate the absolute value of difference between neighborhood and center pixels in support window. 2) Compared with the appropriate threshold. If it is greater than threshold, the marker is 0. Otherwise, 1. 3) By encoding all neighborhood pixels, a binary bit string can be obtained. 4) Calculate the new cost for the two searched regions. 5) Combine the traditional census cost with the new cost. The specific formula is as follows:

$$C(p, d) = (1 - \partial)f_c(p, d) + \partial f_g(p, d) \quad (0 < \partial < 1)$$

$$f_c(p, d) = 1 - \exp\left[-\frac{C_{census}(p, d)}{\lambda_c}\right] \quad f_g(p, d) = 1 - \exp\left[-\frac{C_g(p, d)}{\lambda_g}\right] \quad (2)$$

$C(p, d)$ denotes the computational cost after fusion, C_{census} denotes the Census cost, and C_g denotes the new cost value. f_c and f_g map the calculated cost value to the $[0, 1]$ interval by exponential form. λ is the empirical parameter. ∂ is a constant whose value is between 0 and 1. In the non-smooth region, the parameters are larger, and C_g will have a greater impact on the matching performance. On the contrary, C_{census} will assign a higher weight. The size of ∂ is related to the threshold set in step 2. In this paper, an adaptive threshold method is used to set a larger threshold for the distance from the central pixel in the support window. If the number of pixels larger than the threshold is more than $1/3$ of the total number, the region is considered to be a non-smooth area.

6) The local area is restrained by the Census and the new cost. Then the initial disparity is calculated. 7) Cost aggregation: In this paper, three-constraint adaptive window construction method is used to sum and average all the cost values inside the window. The constraints of window construction are as follows:

$$D_c(p_1, p) < \tau_1, D_c(p_1, p_1 + (1, 0)) < \tau_1 \quad D_s(p_1, p) < L_1$$

$$D_c(p_1, p) < \tau_2, \text{if } L_2 < D_s(p_1, p) < L_1 \quad (\tau_2 < \tau_1, L_2 < L_1) \quad (3)$$

D_c, D_s represents the difference of color and distance between the pixel P1 and the central pixel P, and τ_1, τ_2, L_1, L_2 represents the different color and distance thresholds respectively. Figure 5 is an example.

As shown in the left image, extension construction is carried out in four directions from P. The first constraint is to stop extension if the absolute value of color difference between p and p1 or between p2 and p1 is greater than the threshold τ_1 . The second constraint is to stop extension if the distance between p and p1 is greater than the threshold L_1 . The

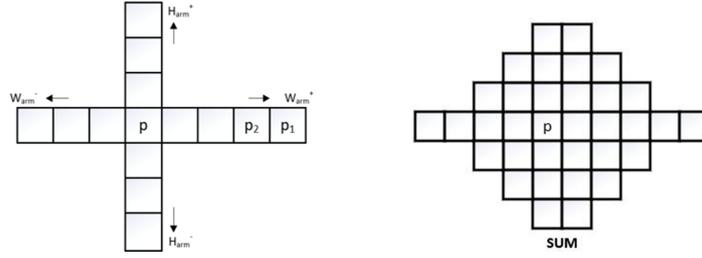


FIGURE 5. Window Construction Sample Diagram

third constraint : when the first constraint condition is satisfied after a long extension of p , the color threshold is lowered to τ_2 and the decision is continued. In this paper, an adaptive threshold which varies according to the distance is proposed in view of the fact that the operation of setting the fixed threshold is prone to mismatch. The extended distance is expanding and the color threshold should be decreasing. Based on this, an adaptive color threshold is established as follows:

$$\tau_3 = \tau_1 - \frac{l(p_1, p)}{L_1} \tau_1 + \beta_1 \tau_1 \quad \tau_4 = \tau_2 - \frac{l(p_1, p) - L_2}{L_1 - L_2} \tau_2 + \beta_2 \tau_2 \quad (4)$$

τ_3 and τ_4 are the newly defined thresholds, $l(p_1, p)$ represents the distance between the p_1 and p , β_1 and β_2 are empirical parameters. According to the above formula, then (3) is changed to:

$$\begin{aligned} D_c(p_1, p) < \tau_3, D_c(p_1, p_1 + (1, 0)) < \tau_1 \quad D_s(p_1, p) < L_1 \\ D_c(p_1, p) < \tau_4, \text{ if } L_2 < D_s(p_1, p) < L_1 \quad (\tau_2 < \tau_1, \tau_4 < \tau_3, L_2 < L_1) \end{aligned} \quad (5)$$

8) Adding noise rejection strategy. For each pixel, a small cross window is constructed, such as the gray-white region of Figure 6. The maximum, minimum and median of the pixels in the gray area are counted, and whether $p_{\min} < p_{\text{mid}} < p_{\max}$ is satisfied or not is judged. If not, the cross window is extended to one pixel unit in each direction, then the above condition is re-counted and judged. If satisfied, the target pixel p_{xy} is judged: $p_{\min} < p_{xy} < p_{\max}$. If this condition is not satisfied, the point is regarded as a noise point and the value p_{xy} is replaced by P_{mid} . The extension stops when any direction of the cross window reaches the boundary. And in each extension process, only four pixels will be added at most: black area.

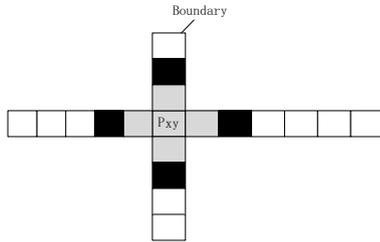


TABLE 1. Value of each parameter

∂	τ_1	τ_2	L_1	L_2	β
0.2 0.8	20	6	34	17	0.2

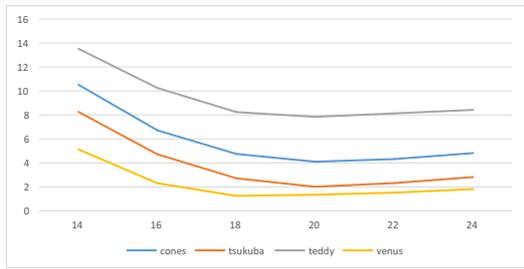
FIGURE 6. Denoising schematic diagram

9) Disparity refinement. First, left and right consistency detection is used to distinguish occlusion points from mismatched points in images. For mismatched points, the method of region voting is adopted. Finally, a quadratic linear interpolation operation is performed on the image, and put this operation in the cost aggregation stage: for a certain pixel p , its true parallax d_p is obtained by the following formula:

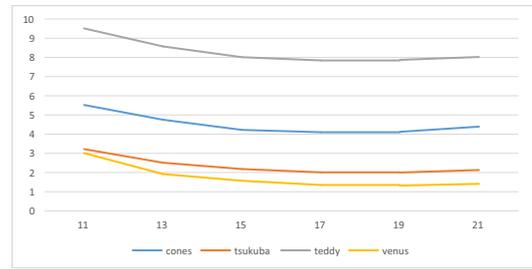
$$d_p = d - \frac{C(p, d_+) - C(p, d_-)}{2(C(p, d_+) + C(p, d_-) - 2C(p, d))} (d_+ = d + 1 \quad d_- = d - 1)$$

3. Experimental Results and Performance Analysis. The standard images on Middlebury platform are selected for experiment. The experimental environment is: PC, Intel (R) Core i5-3470 processor. The operating system is 64-bit Win7 system. The parameters used are shown in Table 1.

Firstly, the rationality of the selected parameters is verified. As shown in Fig. 7 (a), when the distance threshold remains unchanged, the rationality of the color threshold is determined for four images, and the average mismatch rate between the test and the real is taken as criterion. The vertical axis is the mismatch rate, while the horizontal axis is the different values of the color threshold when L2 is 17. Fig. 7(b) indicates the influence of the distance parameter L2 on the mismatch rate when the color threshold is 20. We can see that when $\tau_1 < 20$, the mismatch rate will increase a lot, and when the color threshold is greater than 20, it will change slowly. Moreover, when $L > 17$ is used, the change of mismatch rate is very small.



(a) Mismatch Rate under Different Color Threshold



(b) Mismatch Rate under Different Distance Thresholds

FIGURE 7. Images with different Gaussian noise

In the initial cost calculation stage, the proposed are experimented, and the rectangular area is selected as support window to calculate the cost. The comparison results are shown in Figure 8. The horizontal axis is a rectangular window of M size for experiment, and the vertical axis is the average mismatch rate.

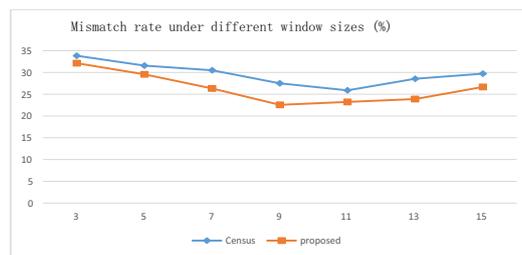


FIGURE 8. Average Mismatch Rate under Different Window Sizes

Verify the sensitivity to noise: Two standard images are added with Gauss noise with different standard deviations. The sensitivity to noise is judged according to the change degree of PSNR. As shown in Table 2, PSNR of the Census is much faster than proposed when noise is increasing.

By adding 2%, 5%, 10% and 15% salt and pepper noise to the above two images, the average mismatch rate is judged. The comparison results are shown in Table 3. With the increase of concentration, more mismatch points will be generated, but the mismatch rate is still lower than that of Mei’s adaptive window aggregation algorithm. This proves that the proposed aggregation algorithm has better robustness to noise.

TABLE 2. Psnr value under different gauss noise processing (cones, teddy)

	cones				teddy			
Standard	0.5	2	5	8	0.5	2	5	8
Census	6.22	5.10	4.22	4.06	6.88	5.52	4.98	4.88
Proposed	5.97	4.94	4.42	4.39	6.35	5.73	5.37	5.33

TABLE 3. Average mismatch rate of non-occluded areas under different salt and pepper noise concentrations %

Noise	none	2%	5%	10%	15%
Ref. [16]	4.54	6.67	11.73	26.88	48.63
Proposed	3.82	5.02	8.65	17.33	38.46

In order to verify the innovation of this algorithm. The initial cost are calculated and aggregated. The comparison results are shown in Fig. 9. First column: standard image. Second: real image. Third: aggregated image of Census. Fourth: mismatched image after comparing with a real image. Fifth: aggregated image of proposed method. Last: mismatch image. We can see that our method can effectively reduce the mismatching rate.

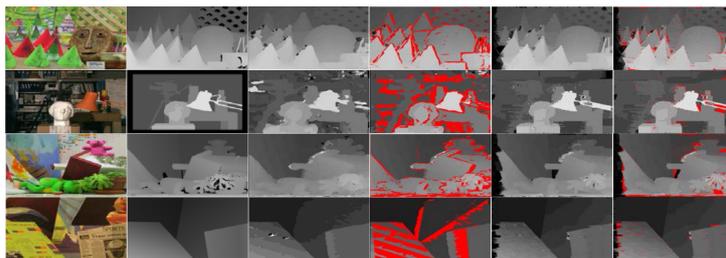


FIGURE 9. Contrast and Error of Result Charts

As shown in Figure 10, the top is the disparity map processed by Census, and the bottom obtained by proposed. We can observe the “peak”, “lamp”, “table” reflects the structure of the object is better maintained.

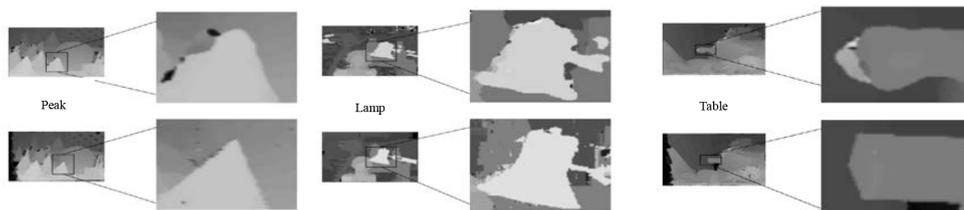


FIGURE 10. Contrast of Local Structural Diagrams

The proposed algorithm also compared with other improved algorithms in non-occluded areas. As shown in Table 4.

From table, we can see that proposed algorithm can greatly improve matching accuracy. In this paper, optimization methods are added on the basis of aggregation. Fig. 11 shows the contrast effects of four types of images before and after optimization. The first column is the original image, second: real image, third: before optimization, fourth: optimized. To verify the reliability of the proposed algorithm, the comparison experiments between

the proposed algorithm and other algorithms are carried out. The comparison data of mismatch rate are shown in Table 5.

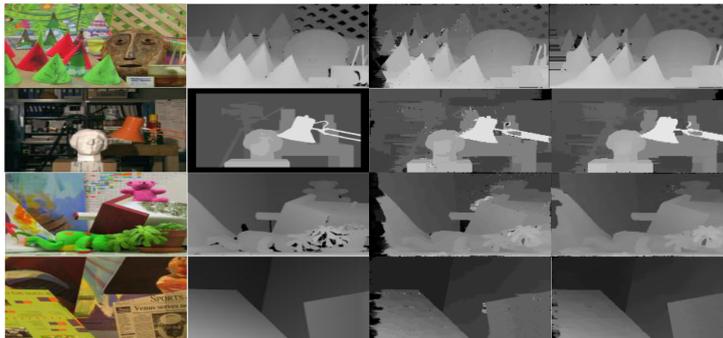


FIGURE 11. Contrast chart before and after optimization

TABLE 4. Comparison of mismatch rates of different algorithms in non-occluded areas %

Algorithms	cones	tsukuba	teddy	venus	avg%
Census	21.89	27.41	23.43	27.78	25.13
Proposed	4.10	2.01	7.84	1.35	3.82
Ref.[9]	4.04	2.53	7.57	1.60	3.93
SG-census [13]	12.92	4.80	8.05	1.91	6.92
Mp-census [17]	6.78	4.50	11.32	3.55	6.53

TABLE 5. Comparison table of mismatch rates of different %

Algorithms	cones		tsukuba		teddy		venus		avg%
	all	nocc	all	nocc	all	nocc	all	nocc	
Proposed	8.85	3.46	1.26	0.92	10.72	6.09	0.86	0.34	4.06
Ref. [9]	9.75	3.29	3.23	2.52	12.35	5.70	1.05	0.23	4.74
Ref. [11]	9.26	3.79	4.15	3.62	11.6	5.68	1.87	1.08	5.13
Seg-CT [12]	13.77	6.15	5.40	4.57	11.44	6.27	1.93	1.19	6.34
SAD + CT [14]	7.65	4.09	1.98	1.29	12.60	6.02	0.74	0.53	4.36
GRD [18]	16.13	4.45	3.59	2.72	17.56	7.45	4.12	1.68	7.21
MCADSR [19]	11.1	3.51	4.15	3.62	14.7	7.57	0.87	0.48	5.75
GradAdaptWt [20]	7.67	2.61	2.63	2.26	13.10	8.00	6.99	1.39	5.58
AdaptAggrDP [21]	13.2	5.53	3.50	1.57	14.3	6.79	2.69	1.53	6.14

The average mismatch rate of proposed is 4.06%. The overall matching performance of proposed is better than that of other algorithms.

4. Conclusions. Aiming at the problem that Census algorithm has strong noise sensitivity and is prone to mismatching in non-smooth areas of images, a algorithm based on neighborhood information constraints is proposed. Firstly, the neighborhood weighted average assignment of the central pixel is carried out. Then the original cost is fused with the newly calculated cost. The initial cost is aggregated by adaptive window based on color change, and noise elimination strategy is introduced in process. In optimization stage, disparity map is processed by combining left-right consistency detection and regional voting.

Acknowledgement. This paper is supported by the Project for the National Natural Science Foundation of China under Grants No. 61672064, the Beijing Natural Science Foundation under Grant No. 4172001, and the Beijing Laboratory of Advanced Information Network under Grant No. PXM2019_014204_500029.

REFERENCES

- [1] Jin Y, Lee M, Enhancing binocular depth estimation based on proactive perception and action cyclic learning for an autonomous developmental robot, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 1, pp. 169-180, 2018.
- [2] Marr D, Poggio T, Cooperative computation of stereo disparity, *Science*, vol. 194, No. 4262, pp. 283-287, 1976.
- [3] Wang F, Jia K, Feng J, The real-time depth map obtainment based on stereo matching, *Euro-china Conference on Intelligent Data Analysis and Applications*, pp. 138-144, 2016.
- [4] Roy S, Cox I J, A maximum-flow formulation of the n-camera stereo correspondence problem, *Proceedings of IEEE International Conference on Computer Vision*, pp. 492-49, 1998.
- [5] T. T. San, N. War, Stereo matching algorithm by hill-climbing segmentation, *2017 IEEE 6th Global Conference on Consumer Electronics (GCCE), IEEE*, 2017.
- [6] Ende W, Yalong Z, Liangyu P, et al. Stereo matching algorithm based on the combination of matching costs, *2017 IEEE 7th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER). IEEE*, 2017.
- [7] Dinh V Q, Nguyen V D, Jeon J W, Robust matching cost function for stereo correspondence using matching by tone mapping and adaptive orthogonal integral image, *IEEE Trans Image Process*, vol. 24, no. 12, pp. 5416-5431, 2015.
- [8] Zabih R, Woodfill J, Non-parametric local transforms for computing visual correspondence, *European Conference on Computer Vision*, vol. 801, pp. 151-158, 1994.
- [9] Fan H R, Yang F, Pan X. An improved Census transform and gradient fusion stereo matching algorithm, *Journal of Optics*, pp. 267-277, 2018.
- [10] Zhu Shiping, Yan Lina and Li Zheng, Stereo matching algorithm based on improved Census transform and dynamic programming, *Journal of Optics*, vol. 36, no. 4, 2016.
- [11] Yubo M, Guoyin Z, Chaoguang M, et al, A stereo matching algorithm based on four-moded census and relative confidence plane fitting, *Chinese Journal of Electronics*, vol. 24, no. 4, pp. 807-812, 2015.
- [12] Beak E T, Ho Y S, Cost aggregation with guided image filter and superpixel for stereo matching, *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA). IEEE*, 2016.
- [13] Guo S, Xu P, Zheng Y, Semi-global matching based disparity estimate using fast Census transform, *International Congress on Image & Signal Processing. IEEE*, 2017.
- [14] Chai Y, Cao X J. Stereo matching algorithm based on joint matching cost and adaptive window, *2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, pp. 442-446, 2018.
- [15] Zhang K, L, Lafruit G, Cross-based local stereo matching using orthogonal integral images, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 7, pp. 1073-1079, 2009.
- [16] Mei X. Sun X, Zhou M, et, al, On building an accurate stereo matching system on graphics hardware, *Proceedings of IEEE International Conference on Computer Vision Workshops*, pp. 467-474, 2011.
- [17] Kwak J C, Park T R, Koo Y S, et al, Implementation of Improved Census Transform Stereo Matching on a Multicore Processor, *Multimedia and Ubiquitous Engineering*, pp. 989-995, 2013.
- [18] Zhang K, Fang Y, Min D, et al, Cross-scale cost aggregation for stereo matching, *Computer Vision & Pattern Recognition. IEEE*, 2014.
- [19] Shan Y, Hao Y, Wang W, et al, Hardware acceleration for an accurate stereo vision system using mini-census adaptive support region, *ACM Transactions on Embedded Computing Systems*, vol. 13, no. 4, pp. 1-24, 2014.
- [20] De-Maeztu L, Villanueva A, Cabeza R, Stereo matching using gradient similarity and locally adaptive support-weight, vol. 32, no. 13, pp. 1643-1651, 2011.
- [21] Wang L, Yang R, Gong M, Real-time stereo using approximated joint bilateral filtering and dynamic programming, *Journal of Real-Time Image Processing*, vol. 9, no. 3, pp. 447-461, 2014.