

## Research Article

# Stereo Matching Algorithm Based on HSV Color Space and Improved Census Transform

Chen Lv <sup>1</sup>, Jiahan Li,<sup>1</sup> Qiqi Kou <sup>2</sup>, Huandong Zhuang,<sup>1</sup> and Shoufeng Tang<sup>1</sup>

<sup>1</sup>School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China

<sup>2</sup>School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China

Correspondence should be addressed to Qiqi Kou; kouqiqi@cumt.edu.cn

Received 22 August 2020; Revised 23 April 2021; Accepted 7 July 2021; Published 20 July 2021

Academic Editor: Yong Aaron Tan

Copyright © 2021 Chen Lv et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aiming at the problem that stereo matching accuracy is easily affected by noise and amplitude distortion, a stereo matching algorithm based on HSV color space and improved census transform is proposed. In the cost calculation stage, the color image is first converted from RGB space to HSV space; moreover, the hue channel is used as the matching primitive to establish the hue absolute difference (HAD) cost calculation function, which reduces the amount of calculation and enhances the robustness of matching. Then, to solve the problem of the traditional census transform overrelying on the central pixel and to improve the noise resistance of the algorithm, an improved census method based on neighborhood weighting is also proposed. Finally, the HAD cost and the improved census cost are nonlinearly fused as the initial cost. In the aggregation stage, an outlier elimination method based on confidence interval is proposed. By calculating the confidence interval of the aggregation window, this paper eliminates the cost value that is not in the confidence interval and subsequently filters as well as aggregates the remaining costs to further reduce the noise interference and improve the matching accuracy. Experiments show that the proposed method can not only effectively suppress the influence of noise, but also achieve a more robust matching effect in scenes with changing exposure and lighting conditions.

## 1. Introduction

Computer vision studies how to let computers obtain high-level and abstract information from images and videos. Feature extraction is one of the key steps. Features generally include image grayscale features, color features, and texture features [1]. Research directions of computer vision include target tracking, superresolution reconstruction [2], and machine learning [3, 4]. In recent years, the stereo matching technology has gradually become a research hotspot in the field of computer vision and has been widely used in unmanned driving, three-dimensional reconstruction, virtual reality, target tracking, and other fields [5, 6]. In stereo matching, researchers find the corresponding points in two or more images in the same scene and calculate the parallax to restore the scene depth. Scharstein and Szeliski [7] divided the stereo matching algorithm into global stereo matching and local stereo matching. The global stereo matching

algorithm [8–10] usually uses the minimized energy function instead of cost aggregation to select the best parallax value, which can obtain a high-precision disparity map, but the complicated calculation leads to the limitations in practical applications. The local stereo matching algorithm uses the local information of pixels to construct a supported window, calculates the cost of all pixels in the window and aggregates to replace the cost value of a single pixel, and finally uses the Winner Take All (WTA) [11] algorithm to obtain the disparity map. Since the local stereo matching algorithm has fast computing speed and low hardware requirements, its accuracy will be more and more higher with the continuous development of research, and its application scenarios are getting much more wider.

Similarity measures of most stereo matching algorithms are based on pixel brightness or gray level information. In actual industrial applications, the binocular camera is affected by the external environment and internal

photosensitive components. There may be many problems in the two images it collects, such as noise interference, light distortion, or inconsistent exposure, which make the effect of this type of algorithm greatly reduced. To solve the above problems, Zabih and Woodfill [12] proposed the census transform for stereo matching, which can effectively suppress the influence of amplitude distortion. Its principle is to map the size relationship between the neighboring pixels and the center pixel into a binary string and calculate the similarity by comparing the Hamming distance. However, the census algorithm overly relies on the central pixel, resulting in an unsatisfactory result in a noisy environment. To overcome that shortcoming, many scholars have proposed many methods on the basis of the census method. Zhu et al. [13] proposed a three-state census with noise tolerance to enhance the robustness of the algorithm under noisy environments. Chen et al. [14] calculated the uniformity of the star-shaped neighborhood of the center pixel and replaced the gray value of the center pixel with the mean value of the minimum uniformity. Chang et al. [15] proposed the MCT algorithm to perform census transform on 6 fixed pixels. The above algorithms suppress the amplitude distortion and the noise interference to a certain extent. However, the census-based method only relies on the size relationship between pixels as the basis for similarity judgment in the cost calculation process, so it loses pixel gray information and distance information, which affects the matching accuracy.

Based on the above analysis, this paper proposes an antinoise matching method based on HSV color space and improved census transform. In the cost calculation stage, the original image is converted from RGB color space to HSV color space, whose hue channel is used as a matching primitive to establish the HAD cost calculation function and suppress the impact of amplitude distortions. Then, the traditional census transform is improved. The weighted value of the window is used to compare with the gray value of the center pixel. When the difference between them is too large, the weighted value is used to replace the gray value of the center pixel, thereby reducing noise interferences. Finally, the cost values of HAD and the improved census method are mapped to  $[0, 1]$  and merged as the initial cost. In the cost aggregation stage, a confidence interval-based outlier elimination method is proposed. The cost values which are not in the confidence interval will be eliminated; in addition, the remaining costs will be filtered and aggregated, so as to further reduce the influence of noise and improve the matching accuracy in nonideal environments.

## 2. Algorithm Description

The proposed method consists of cost calculation, cost aggregation, choice of disparity, etc. These parts are introduced as follows.

### 2.1. Cost Calculation

**2.1.1. HAD Cost.** The traditional RGB color space is composed of three brightness-related color components: R, B,

and G. Once the exposure or lighting conditions changed, the values of the three components will change significantly. At this time, it is difficult to use the RGB color space to make similarity judgments. The improved cost calculation method based on census usually loses color information, which affects the matching accuracy. To solve the above problems, a cost calculation method of HAD is proposed in this section. This method first converts the image from the RGB color space to the HSV color space. The H channel is the hue containing rich color information. S channel and V channel are the saturation and the brightness, respectively. When exposure or lighting conditions changed, the value of the H channel is relatively stable, while the S channel and V channel are easily affected. Therefore, when calculating the cost, the value of the H channel is used as the matching primitive, and subsequently, the values of S channel and V channel are discarded. In this way, the amount of calculations can be reduced, and the robustness of the algorithm in the case of amplitude distortions can be enhanced. This paper uses the truncation threshold  $\tau_H$  to prevent the abnormality of a single pixel from having too much impact on the overall cost. The HAD cost calculation function is expressed as

$$C_H(i, d) = \min[|I_H(i) - I'_H(i_d)|, \tau_H], \quad (1)$$

where  $i$  and  $i_d$  represent the corresponding pixels in the reference and the target graphs when the disparity is  $d$ ,  $C_H(i, d)$  is the cost of these two pixels in the HSV color space, which is used to reflect the degree of difference in hue. The larger the value, the more the difference.  $I_H(i)$  and  $I'_H(i_d)$ , respectively, represent the values of these two pixels in the H channel, and  $\tau_H$  represents the hue truncation threshold. Figure 1 shows the test image of Aloe under different exposure and lighting conditions. The exposures of Figures 1(a) and 1(b) are different. Figures 1(a) and 1(c) have different light conditions. The exposure and lighting conditions of Figures 1(b) and 1(c) are different. Furthermore, the center of the red circle in the three pictures is the same pixel point. This paper counts the value of this point in each channel of RGB and HSV and the gray value. To facilitate comparison, these values are mapped to the interval of  $[0, 255]$ , and the results are shown in Table 1.

It can be seen that, due to the inconsistency of exposure and lighting conditions, the values of the three channels of the same pixel in the RGB color space have changed significantly. The same is true for the gray value of the same pixel. What is more, in the HSV color space, the values of the S channel and the V channel are also affected by the exposure and illumination. However, the value of the H channel of the same pixel has a small change, especially in the two images Figures 1(a) and 1(b) with different exposures and the difference is only 2, which is much smaller than that of other channels. In order to show the degree of change of each channel value more intuitively, this paper calculated the standard deviations of the R, G, B, gray, and H channels. The results are shown in Table 2.

In Table 2, the standard deviation of the pixel in the center of the circle of the three images Figures 1(a)–1(c) in the H channel is 4.78, which is far less than that of the R, G,

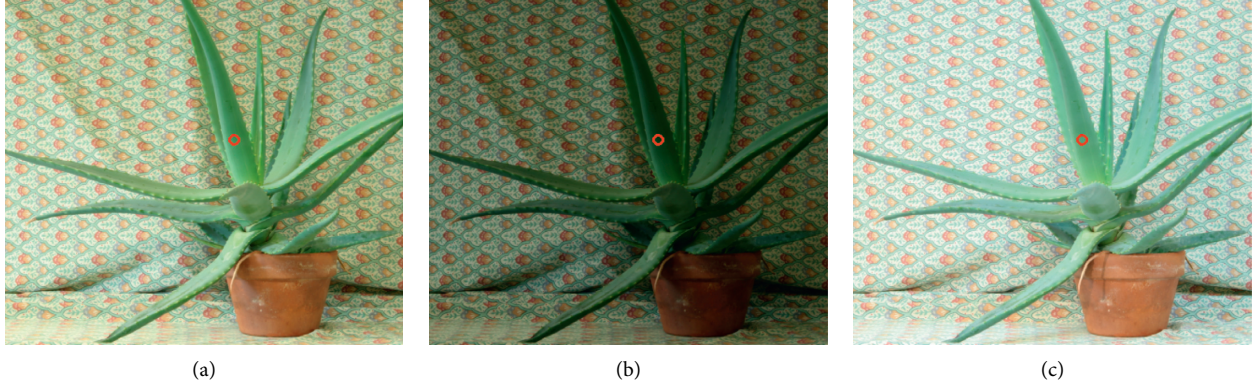


FIGURE 1: Aloe under different exposure and lighting conditions.

TABLE 1: Values of different channels.

Image	Channels						
	R	G	B	H	S	V	Gray
Figure 1(a)	96	142	89	79	95	142	122
Figure 1(b)	31	54	28	81	123	54	44
Figure 1(c)	115	161	121	90	73	161	143

TABLE 2: Standard deviations of different channels.

$\sigma_R$	$\sigma_G$	$\sigma_B$	$\sigma_{\text{gray}}$	$\sigma_H$
35.97	46.61	38.58	20.46	4.78

B, and gray channels. It shows that using the image hue information as a matching primitive can effectively reduce the impact of exposure and light changes.

**2.1.2. Improved Census Method.** The traditional census algorithm can suppress the impact of amplitude distortions to a certain extent, but it relies too much on the center pixel. Once the center pixel value is disturbed by noise, the census encoding will change significantly, leading to mismatches. A typical method is to take the mean of neighborhood pixels instead of the center pixel for census transform [15]. However, this method has two obvious shortcomings. One is that it ignores the particularity of the center pixel, and the other is that it ignores the distance relationship between the center pixel and the neighboring pixels. Aiming at the above shortcomings, an improved census method is proposed. According to the distance relationship between neighborhood pixels and central pixels, the gray values of pixels in the rectangular window are weighted and summed to obtain a new value:

$$I_\omega(x, y) = \sum_{u=-m}^m \sum_{v=-n}^n I(x+u, y+v) \omega(u, v). \quad (2)$$

In formula (2),  $I_\omega(x, y)$  is the weighted result centered on  $(x, y)$ ,  $(x+u, y+v)$  is the neighborhood coordinate of  $(x, y)$ ,  $(u, v)$  can be considered as the position of the window where the current point is located, and the weighted window can be regarded as a rectangular region with the size of  $(2m+1) \times (2n+1)$ , and  $\omega(u, v)$  is a weighting function and can be expressed as

$$\omega(u, v) = \eta e^{-(u^2+v^2)}. \quad (3)$$

Among them, the parameter  $\eta$  is used for normalization to ensure that the sum of the weights is 1, which is expressed as

$$\eta = \left( \sum_{u=-m}^m \sum_{v=-n}^n e^{-(u^2+v^2)} \right)^{-1}. \quad (4)$$

Using formula (2) for weighting, this section takes into account the uniqueness of the center pixel, while considering the distance relationship between the neighboring pixels and the center pixel. The closer to the center pixel, the greater its weight and the more contribution to the weighting result. Figure 2 shows the weight distribution map with a size of  $3 \times 3$ .

This paper compares  $I_\omega(x, y)$  and  $I(x, y)$  and then sets the error threshold  $T_{\text{mid}}$  to determine the value of the center pixel as follows:

$$I_{\text{mid}}(x, y) = \begin{cases} I(x, y), & |I(x, y) - I_\omega(x, y)| \leq T_{\text{mid}}, \\ I_\omega(x, y), & |I(x, y) - I_\omega(x, y)| > T_{\text{mid}}. \end{cases} \quad (5)$$

It can be seen from formula (5) that when the value difference between  $I_\omega(x, y)$  and  $I(x, y)$  is greater than  $T_{\text{mid}}$ ,  $I_\omega(x, y)$  is used instead of  $I(x, y)$ ; otherwise,  $I(x, y)$  is still used as the central pixel for census encoding.

Finally, this paper performs census transform on the image to get the census code as follows:

$$\text{Str}(x, y) = \bigotimes_{u=-w}^w \bigotimes_{v=-h}^h \xi(I_{\text{mid}}(x, y), I(x+u, y+v)), (x+u, y+v) \neq (x, y). \quad (6)$$

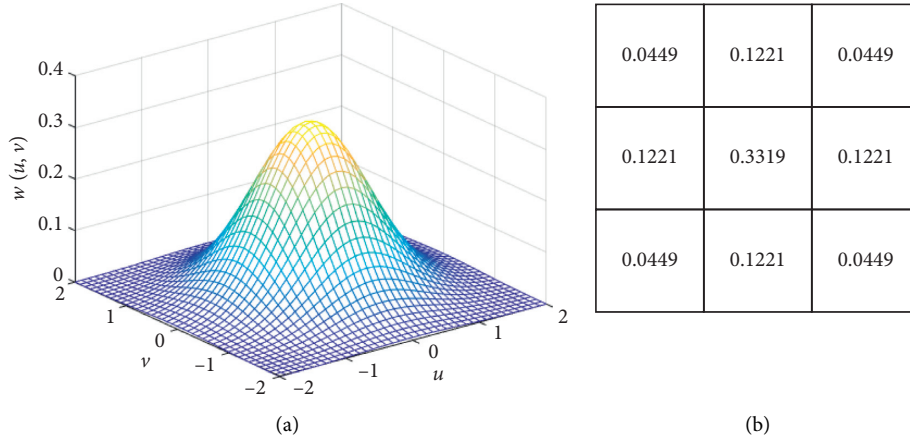


FIGURE 2: The support window weight distribution map with a size of  $3 \times 3$ . (a) The 3D weight distribution map and (b) the 2D coordinate weight distribution map.

where  $\text{Str}(x, y)$  is the census code corresponding to the center pixel,  $(x, y)$  is the center pixel coordinate,  $(x + u, y + v)$  is the neighborhood pixel coordinate, and  $I_{\text{mid}}(x, y), I(x + u, y + v)$  are their gray values, respectively. The neighborhood selection range is a rectangular area with a width of  $[-w, w]$  and a height of  $[-h, h]$ .  $\oplus$  represents the bitwise connector.  $\xi(\cdot)$  is the mapping function of the census transform, which is expressed as

$$\xi(I_{\text{mid}}(x, y), I(x + u, y + v)) = \begin{cases} 0, & I_{\text{mid}}(x, y) \leq I(x + u, y + v), \\ 1, & I_{\text{mid}}(x, y) > I(x + u, y + v). \end{cases} \quad (7)$$

Figure 3 shows the comparison charts of the census transform. Under the nonnoise condition, the traditional census code is 00011000, as shown in Figure 3(a). When the gray value of the center pixel is disturbed by noise from 50 to 70, the transformation process of the traditional census method and the improved method in this paper are shown in Figures 3(b) and 3(c), respectively. Because the traditional census transform overly relies on the central pixel, the generated code is quite different from the original code. There are 6 bits in Figure 3(b), which are different from the original code. However, in Figure 3(c), the code obtained by the weighted summation method is only one bit different from the original code, which indicates that the proposed method in this section can effectively suppress the influence of noise and improve the matching robustness.

**2.1.3. Cost Fusion.** In Sections 2.1.1 and 2.1.2, the cost calculation methods of HAD and improved census are, respectively, proposed. Compared with the original methods, these two methods can effectively reduce the amplitude distortion and noise influence. This section fuses the two methods to achieve a more robust initial cost.

The range of pixels of the H channel is  $[0^\circ, 360^\circ]$ . However, the value of the Hamming distance in the census transform is related to the window size. This paper uses the  $9 \times 9$  transform window, and the values' range is  $[0, 80]$ . It can be illustrated that the units and the ranges of the two

matching primitives—HAD and census—are different, so the simple linear superposition cannot be used for fusion. In addition, nonlinear mapping can suppress the influence of outliers. Therefore, this section uses a nonlinear function to map the cost values of HAD and improved census to the interval of  $[0, 1]$ , respectively. The mapping function expression is

$$\rho(x) = 1 - e^{-(x/\lambda)}. \quad (8)$$

In formula (8), the influence on the result will decrease smoothly as  $x$  increases. When  $x$  is larger than a certain value,  $\rho(x)$  tends to stabilize and finally converges to 1. In addition, parameter  $\lambda$  is used to control the rate of convergence and suppress outliers. The corresponding results of its different values are shown in Figure 4. The mapping results of HAD and improved census are combined to get the final cost  $C_{H-\text{CEN}}(i, d)$  as follows:

$$C_{H-\text{CEN}}(i, d) = 2 - e^{-(C_H(i, d)/\lambda_H)} - e^{-(C_{\text{CEN}}(i, d)/\lambda_{\text{CEN}})}, \quad (9)$$

where  $C_H(i, d)$  and  $C_{\text{CEN}}(i, d)$  are the HAD cost and the improved census cost, respectively, and  $\lambda_H$  and  $\lambda_{\text{CEN}}$  are the control parameters of these two costs.

**2.2. Cost Aggregation of Outlier Eliminations.** The cost aggregation can improve the pixel identification and get the correct disparity more easily. However, when multiple pixels in the aggregation window are interfered by noise and cause abnormal cost value, its aggregation result will also be affected. In order to solve the above problems, a cost aggregation method for removing outliers is proposed in this section. In the cost aggregation stage, the outlier value is eliminated according to the confidence interval of the pixels in the window, so that a reliable aggregation area can be obtained. The proposed specific method is as follows.

First, this method calculates the means and the standard deviations of all cost values in the rectangular aggregation area centered on pixel  $i$ . The formulas for the means and the standard deviations are

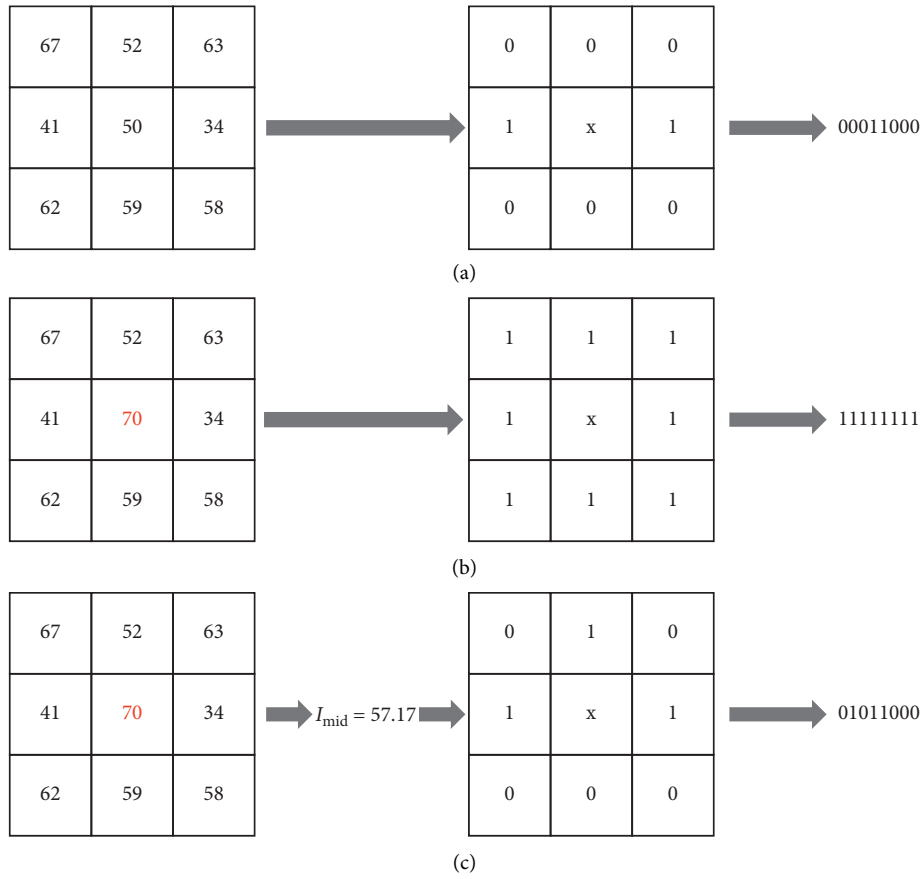


FIGURE 3: Comparison charts of census transform. (a) Traditional census method under noise-free environment; (b) traditional census method under noisy environment; (c) method proposed in this section under noisy environment.

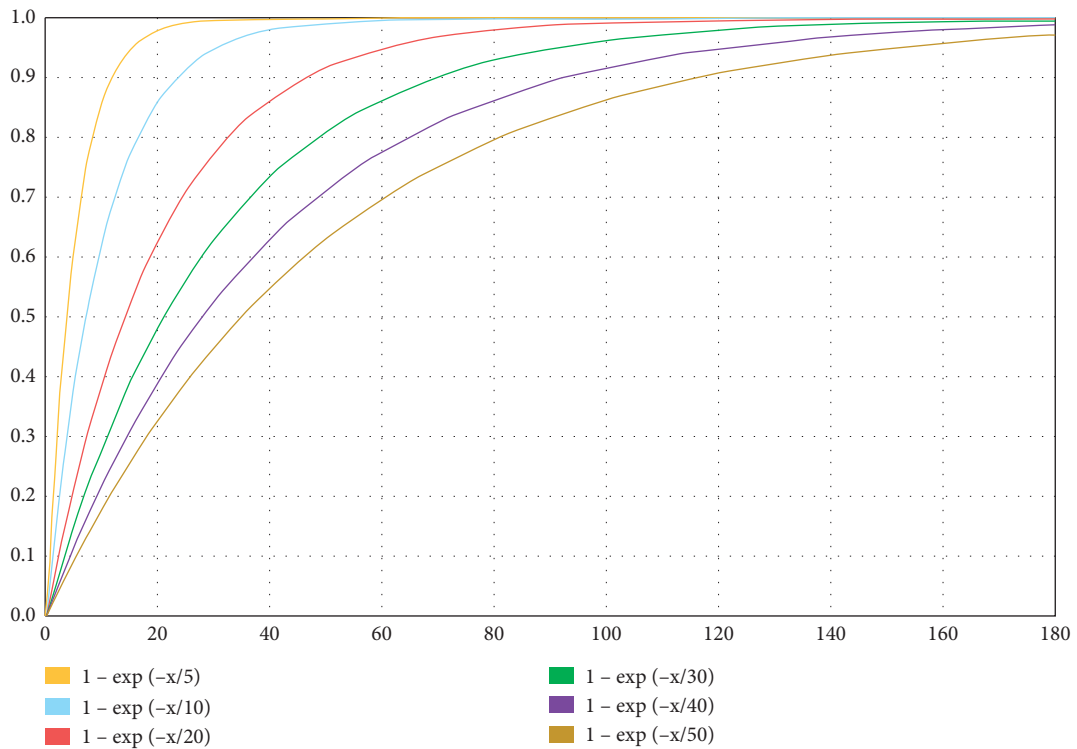


FIGURE 4: Different values of  $\lambda$  correspond to different results.

$$\begin{aligned}\bar{C}_{H-CEN}(i, d) &= \frac{1}{n} \sum_{j \in N(i)} C_{H-CEN}(j, d), \\ \sigma(i, d) &= \sqrt{\frac{1}{n} \sum_{j \in N(i)} (C_{H-CEN}(j, d) - \bar{C}_{H-CEN}(i, d))^2},\end{aligned}\quad (10)$$

where  $j$  is the pixel in the aggregation window  $N(i)$  and  $n$  is the number of pixels.  $\bar{C}_{H-CEN}(i, d)$  and  $\sigma(i, d)$ , respectively, represent the mean and the standard deviation of all cost values in  $N(i)$  when the disparity is  $d$ .

Second, it calculates the confidence interval of the aggregation area based on the mean  $\bar{C}_{H-CEN}(i, d)$  and the standard deviation  $\sigma(i, d)$ , which is expressed as

$$\left[ \bar{C}_{H-CEN}(i, d) - \frac{\sigma(i, d)}{\sqrt{n}} z_{\alpha/2}, \bar{C}_{H-CEN}(i, d) + \frac{\sigma(i, d)}{\sqrt{n}} z_{\alpha/2} \right], \quad (11)$$

where  $\alpha$  represents the significance level, when the confidence level is 95%, the value of  $\alpha$  is 0.05, and  $z_{\alpha/2}$  is the critical value. Furthermore, according to the normal distribution table,  $z_{0.025} = 1.96$ .

The specific process of removing outliers is shown in Figure 5. The inputs are left and right images of Gaussian noise with a standard deviation of 10. In the images, a rectangular aggregation area of size  $9 \times 9$  is selected. And after the cost calculation, the cost value is calculated according to the above method to obtain  $\bar{C}_{H-CEN}(i, d) = 38.25$  and  $\sigma(i, d) = 10.62$ . The confidence interval is [35.94, 40.56]. The outlier value is shown in red in Figure 5. After removing the outliers, a reliable aggregation area is obtained.

Finally, the cost values of all pixels in the aggregation area  $N(i)$  are traversed, and then the cost values are marked as outliers and removed, which are not in the interval shown in formula (11). The area after the outliers are removed is denoted as  $N'(i)$ , the number of pixels contained in  $N'(i)$  is denoted as  $n'$ , and  $N'(i)$  is aggregated to obtain the cost as follows:

$$C_{agg}(i, d) = \frac{1}{n'} \sum_{j \in N'(i)} C_{H-CEN}(j, d) K(i, j), \quad (12)$$

where  $C_{agg}$  is the aggregation cost of point  $i$  when the disparity is  $d$ . Moreover,  $K(i, j)$  is the similarity core, which is used to measure the similarity between pixels  $i$  and  $j$ . The higher the similarity, the smaller the value is. What is more, the guided filtering is selected as the similar kernel function.

According to the aggregation result, the WTA algorithm is used for the disparity selection and finally the disparity map generation.

### 3. Experiment and Analysis

The experiment is based on the VS2013 platform, combined with the OpenCV3.2 open source vision library and compiled with C++ language. The computer hardware environment is as follows. CPU: Intel (R) Core i7-4700MQ, main

frequency: 2.4 GHz, memory: 8 GB, and software environment: Windows 8.1 x64 system. The test set comes from the KITTI2015 and Middlebury stereo matching test platforms. According to the Middlebury evaluation standard, the error threshold is set to 1, which means that once the difference between the disparity map obtained by the algorithm and the real disparity is greater than 1, the point of the disparity map will be considered as a mismatch. In order to verify the effectiveness of the proposed method more objectively, none of the experimental results have been subjected to any disparity refinement processing. In addition, the experimental parameters are shown in Table 3.

**3.1. Noise Immunity Test.** To verify the performance of the proposed method in a noisy environment, three cost calculation methods are selected for comparing with it, namely, the improved census method (MCT) [15], the combined census and gradient method (CT + Grad) [16], and the AD combined with gradient method (AD + Grad) [17]. The experimental dataset consists of 4 groups of standard test images provided by Middlebury. Herein, salt and pepper noise with a density of 1%, 2%, 5%, and 10% and gaussian noise with a standard deviation of 5, 10, 15, and 20 are added to the reference images. To ensure the consistency of the experiment, this thesis uses the same aggregation method. In addition, the average mismatch rates of the four methods in the nonoccluded area are compared in Table 4.

According to the experimental results, in the noise-free environment and the salt and pepper noise environment, the method in this paper and the AD + Grad method can achieve better matching results. Moreover, the mismatch rates of the five experiments are all lower than the MCT and CT + Grad methods. Among them, the AD + Grad method performs better than the proposed method in the case of no noise. However, under the condition of 1%, 2%, 5%, and 10% salt and pepper noise, the proposed method is superior to the AD + Grad method. Moreover, in the Gaussian noise environment, the mismatch rate of the AD + Grad method has increased significantly, which is much higher than the other three methods. What is more, the mismatch rate of the method in this paper remains the lowest among the four methods. Peculiarly under Gaussian noise with a standard deviation of 20, the mismatch rate of the proposed algorithm is 2.04% less than the second-ranked method named CT + Grad. With the increase in noise, the mismatch rate of the proposed method has always been maintained at a low level, and the gap with the other three methods has gradually widened. Therefore, it is obvious that the cost calculation method based on the HSV color space and improved census proposed in this paper can effectively suppress noise interference.

In order to further verify the effect of the improved census method on noise suppression, the KITTI2015 dataset is selected for testing. KITTI includes image pairs acquired in real road scenes. In practical applications, noise will inevitably be mixed due to the influence of the hardware used to collect images and the signal transmission. In order to simulate noise interference, the standard test chart in the



FIGURE 5: Outlier elimination process.

TABLE 3: Parameters for the proposed method.

Parameters	$m$	$n$	$\tau_H$	$T_{mid}$	$\lambda_H$	$\lambda_{CEN}$	$n$	$z_{\alpha/2}$
Values	1	1	9	8	21	35	80	1.96

TABLE 4: Average error matching rates of different cost computation methods in the noise-free environment and the noise environment.

Algorithm	Noiseless	Salt and pepper noise				Gaussian noise			
		1%	2%	5%	10%	5	10	15	20
MCT	5.01	5.98	6.53	7.80	10.08	7.63	10.33	13.32	15.96
CT + Grad	4.59	5.30	5.74	6.80	8.47	7.23	9.88	13.01	15.16
AD + Grad	<b>4.04</b>	4.75	5.20	6.12	7.36	11.27	27.06	47.69	63.49
Proposed	4.40	<b>4.63</b>	<b>4.75</b>	<b>5.12</b>	<b>5.87</b>	<b>6.32</b>	<b>8.80</b>	<b>10.88</b>	<b>13.12</b>

KITTI test set is selected. In addition, salt and pepper noise with a noise density of 10% and Gaussian noise with a standard deviation of 10 in the reference picture is added for comparison experiments. The experiment was carried out in the same color space and the same aggregation method was used. The experimental results are shown in Figure 6. Figures 6(a) and 6(b) are the test images with added salt and pepper noise and Gaussian noise, respectively. Figures 6(c) and 6(e) show the algorithm mismatch rates under the salt and pepper noise before and after the improvement of census, which can be observed to decrease by 0.54%. Moreover, Figures 6(d) and 6(f) show the mismatch rates of

the algorithm under the Gaussian noise before and after the improvement of census, respectively. The improved algorithm is 9.83% lower than before. Therefore, experiments show that the improved cost calculation method of census can effectively suppress noise interference.

To further verify the universality of the proposed outlier elimination method based on confidence interval in cost aggregation, pepper and salt noise with density of 5% was added into four groups of standard test images and three different similar kernel functions were selected as the cost aggregation methods, namely, box filter [18], guided filter [19], and minimum spanning tree (MST) [20]. The

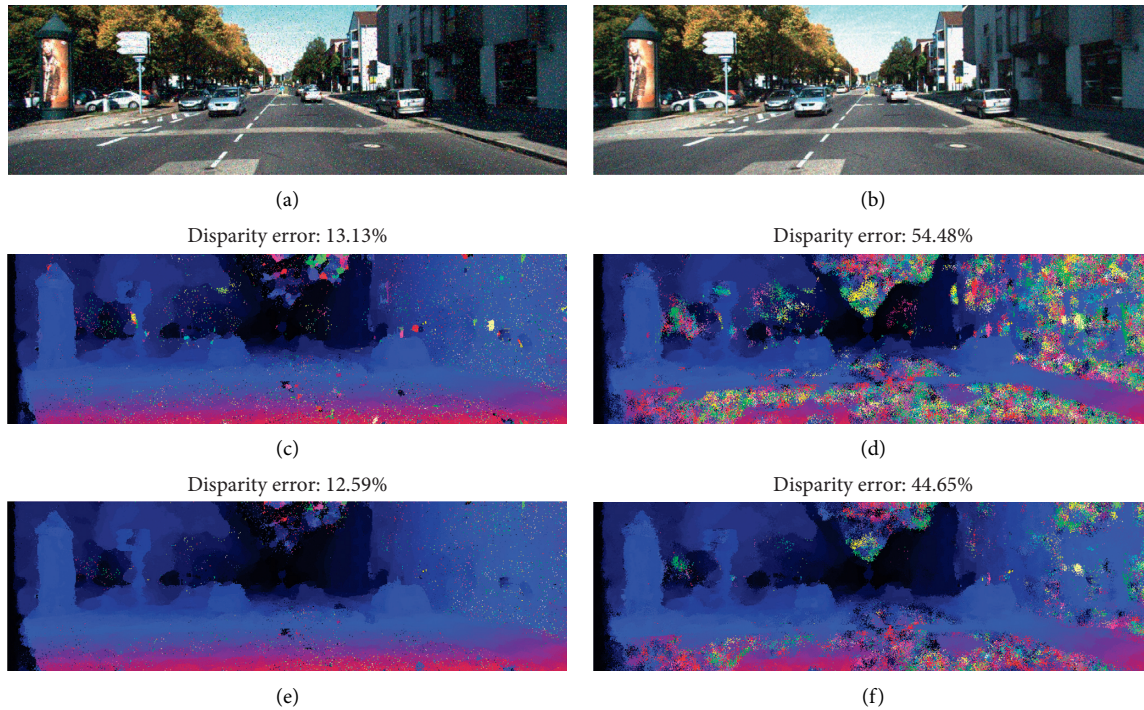


FIGURE 6: Performance verification of the improved census method in a noisy environment.

experiments count the mismatch rates of the original methods and of the methods added in the outlier elimination step in the nonocclusion area. In addition, the H-CEN cost calculation method proposed in Section 2.1 is used in all experiments. The experimental results are shown in Table 5.

By analyzing the above table, a conclusion could be drawn that there are great differences between experimental results of different aggregation methods; also, the mismatching rates of the three aggregation methods, namely, box filtering, guided filtering, and minimum spanning tree, decrease successively. By comparing the effect before and after the improvement of the same method, it can be found that the mismatch rate of the improved aggregation method in this paper is lower than that of the original method. Furthermore, in order to show the effect more intuitively, the histogram corresponding to the experimental results is shown in Figure 7.

From Figure 7, it can be observed that, under the salt and pepper noise with the density of 5%, the proposed outlier elimination method based on the confidence interval can effectively reduce the mismatching rate of the above three aggregation methods, of which the box filtering effect is the most effective. In addition, the effect of the minimum spanning tree is the weakest. The main reasons are as follows. The box filter itself has poor noise resistance, and the generated disparity map of the box filter has a high mismatch rate. Moreover, the outlier elimination method can effectively eliminate the abnormal value and improve the matching accuracy. However, the minimum spanning tree is proposed based on the overall idea, which can suppress the influence of noise to a certain extent and its mismatching rate is already at a low

level. Therefore, the effect of outlier elimination is limited.

Figure 8 presents the disparity maps generated by different aggregation methods of the test image Tsukuba before and after adopting the outlier elimination method based on the confidence interval.

Based on the above images, the quality of the disparity maps generated by the guided filtering and the minimum spanning tree is significantly higher than that of the box filtering. Comparing the two images, Figures 8(e) and 8(f), the area above the lamp in Figure 8(e) is disturbed by noise, and the generated disparity map has white noise, while noise in the same area in Figure 8(f) is significantly reduced. By comparing the matching effect of the table lamp position in Figures 8(c)–8(h), it can be found that the disparity map used the outlier elimination method proposed in this paper has a clearer outline of the table lamp and the table lamp bracket is more complete, which fully demonstrates that the improved method can effectively suppress noise.

**3.2. Exposure and Lighting Experiments.** Exposure and lighting experiments are performed on the 7 sets of test images of Aloe, Art, Baby1, Baby2, Bowling 1, Cloth1, and Dolls in the Middlebury dataset. In addition, their disparity search ranges are [0, 71], [0, 75], [0, 46], [0, 52], [0, 77], [0, 58], and [0, 74], respectively. The scaling factor of disparities are all 3. There are several methods participating in the comparative experiment: SAD method [21], census and gradient combined method under guided filtering (CG-GF) [16], AD and gradient combined method under guided filtering (AG-GF) [17], census and gradient combined



TABLE 5: Error matching rates before and after the improvement.

Stereo pairs	Original method			Improved method		
	Box filter	Guided filter	MST	Box filter	Guided filter	MST
Tsukuba	10.60	5.40	2.44	7.95	2.90	2.23
Venus	7.85	4.49	3.15	5.40	3.14	3.13
Teddy	13.44	11.38	7.15	11.70	9.80	7.04
Cones	6.18	5.93	4.15	5.65	4.63	4.13
Average	9.52	6.80	4.23	7.68	5.12	4.13

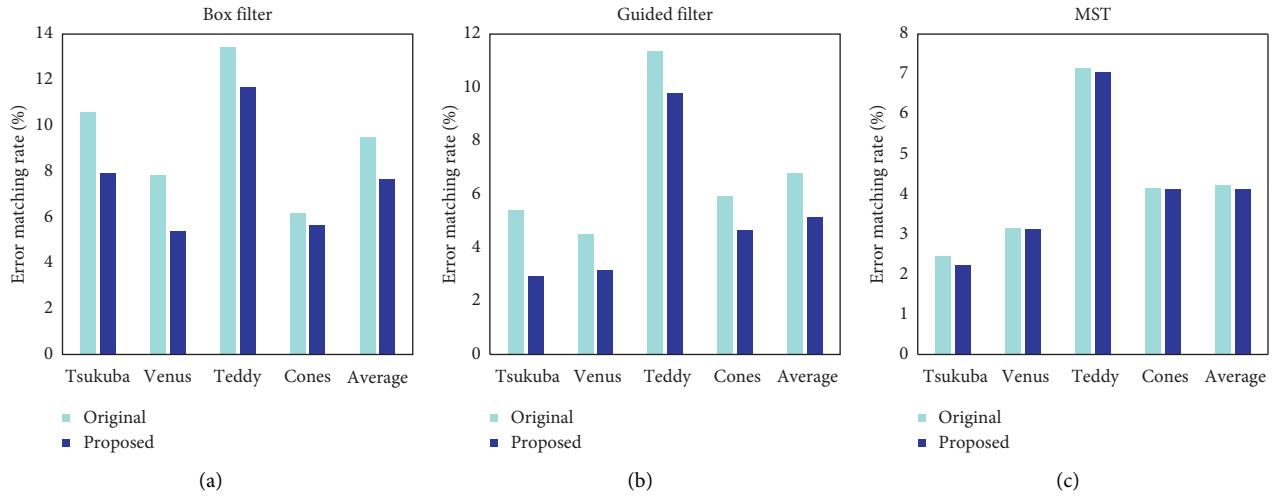


FIGURE 7: Histograms of experimental results of different cost aggregation methods. (a) Box filter; (b) guided filter; (c) minimum spanning tree.

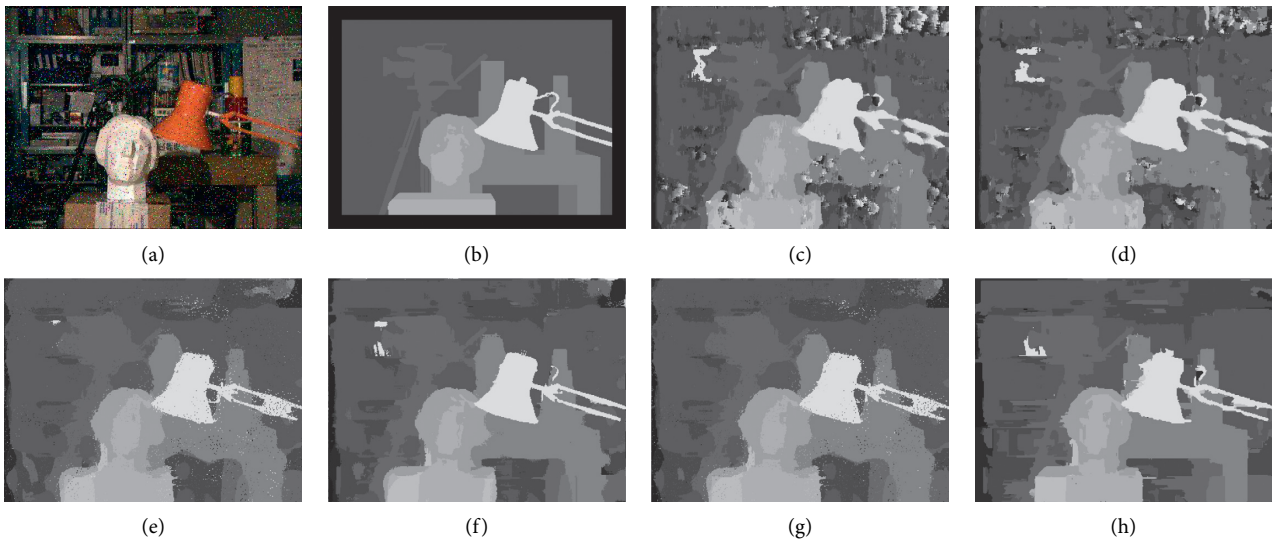


FIGURE 8: Disparity maps generated by different aggregation methods. (a) 5% salt and pepper noise image; (b) ground truth map; (c) box filter; (d) improved box filter; (e) guided filter; (f) improved guided filter; (g) MST; (h) improved MST.

method under MST (CG-MST), and AD and gradient combined method under MST (AG-MST) [20]. At the same time, in order to verify the effect of the HSV color space in the cost calculation method of HAD proposed in this paper, an ablation control group using RGB color space under the framework of this paper is added.

Figures 9–14 show a part of experimental results among them. Figures 9–11 are disparity maps generated by different algorithms under inconsistent exposure conditions. Figures 12–14 are the disparity maps generated by different algorithms under inconsistent lighting conditions, respectively. In these figures, the above five

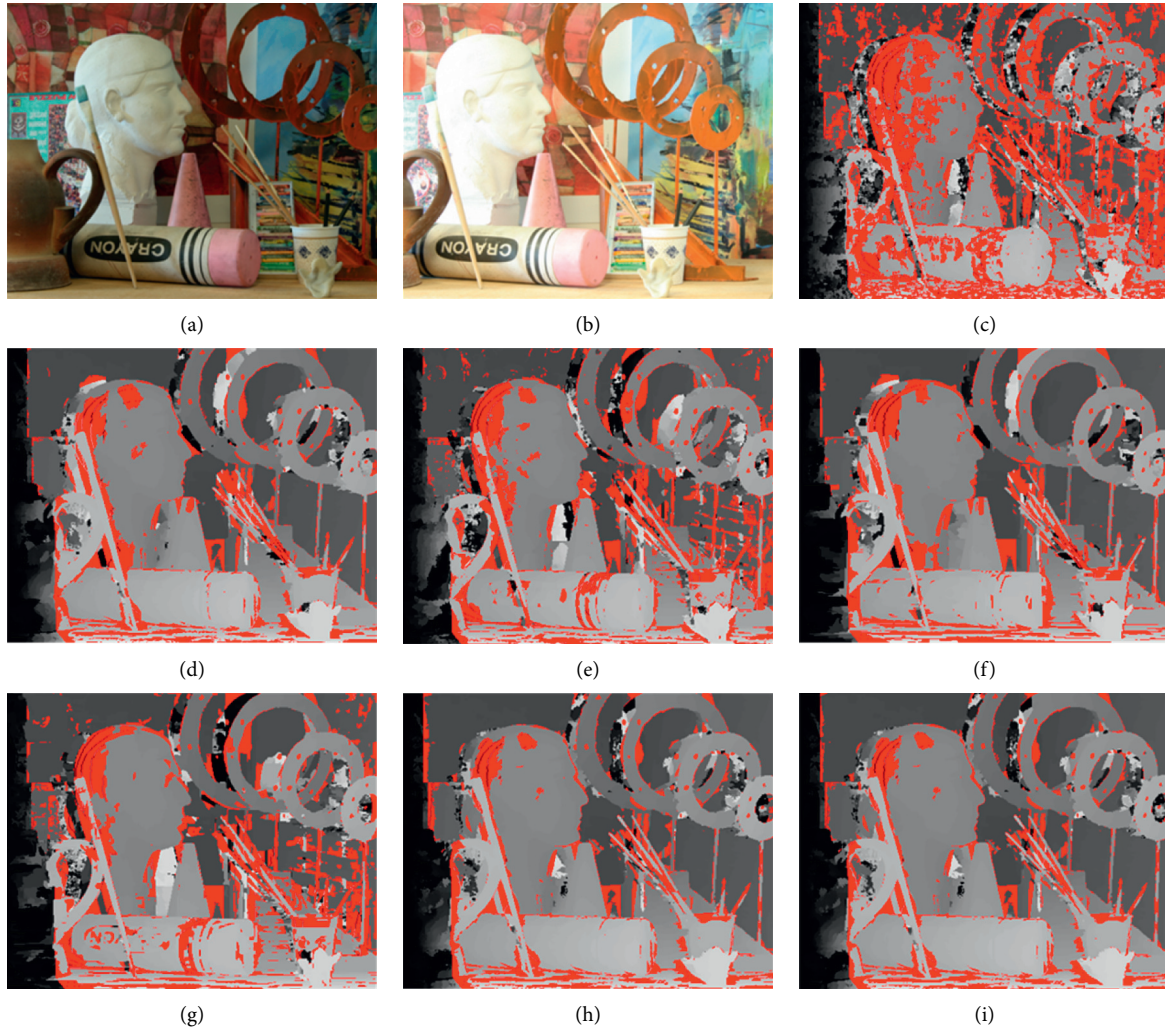


FIGURE 9: Art's disparity maps generated by different algorithms when exposure conditions are not consistent. (a) Left image; (b) right image; (c) SAD; (d) CG-GF; (e) AG-GF; (f) CG-MST; (g) AG-MST; (h) proposed RGB; (i) proposed HSV.

methods correspond to Figures 9(c)–9(g), respectively. Furthermore, Figure 9(h) is the experimental result using RGB color space under the framework of this article and Figure 9(i) is the experimental result using HSV color space. In addition, the red pixels in the figures are mismatched points in the nonoccluded area.

From the above figures, it can be observed that the SAD method performs poorly when the exposure and lighting conditions are changed, and there are a large amount of mismatch points in the images. Based on the experimental results of the previous section, despite that the methods based on AD and gradient combination (AG-GF and AG-MST) have a high matching accuracy in the ideal environment, their matching precisions drop sharply, and the matching results are not good when faced with the amplitude distortion. Besides, the methods based on the MST aggregation (CG-MST, AG-MST) perform better under salt and pepper noise, but there are also many mismatches in the experiment of this section. However, the method based on RGB space proposed in this paper has fewer mismatch points in several comparison algorithms, and the method

based on HSV space proposed in this thesis not only performed outstandingly in the experiment of the previous section, but also achieved a better matching effect when the exposure and the illumination were changed. It has the least number of mismatch points in red. Tables 6 and 7 show the mismatch rates of Aloe, Art, Baby1, Baby2, Bowling1, Cloth1, and Dolls in the nonoccluded area under different exposure and lighting conditions.

When the exposure conditions change, the matching accuracy of the CG-GF method, proposed RGB (Pro-RGB), and proposed HSV (Pro-HSV) proposed in this paper are significantly better than that of other methods. Among them, the CG-GF method ranks first for the test images of Aloe, Baby1, Bowling1, and Cloth1. What is more, for Dolls two sets of test images, the performance of the proposed RGB is the best. The proposed HSV method ranked first in the Art test chart. Moreover, it also has the lowest average mismatch rate for all test images, indicating that the overall effect is better than that of the CG-GF method. In addition, only in the Dolls test chart, the proposed HSV has a slightly lower accuracy than the proposed RGB. The results of the

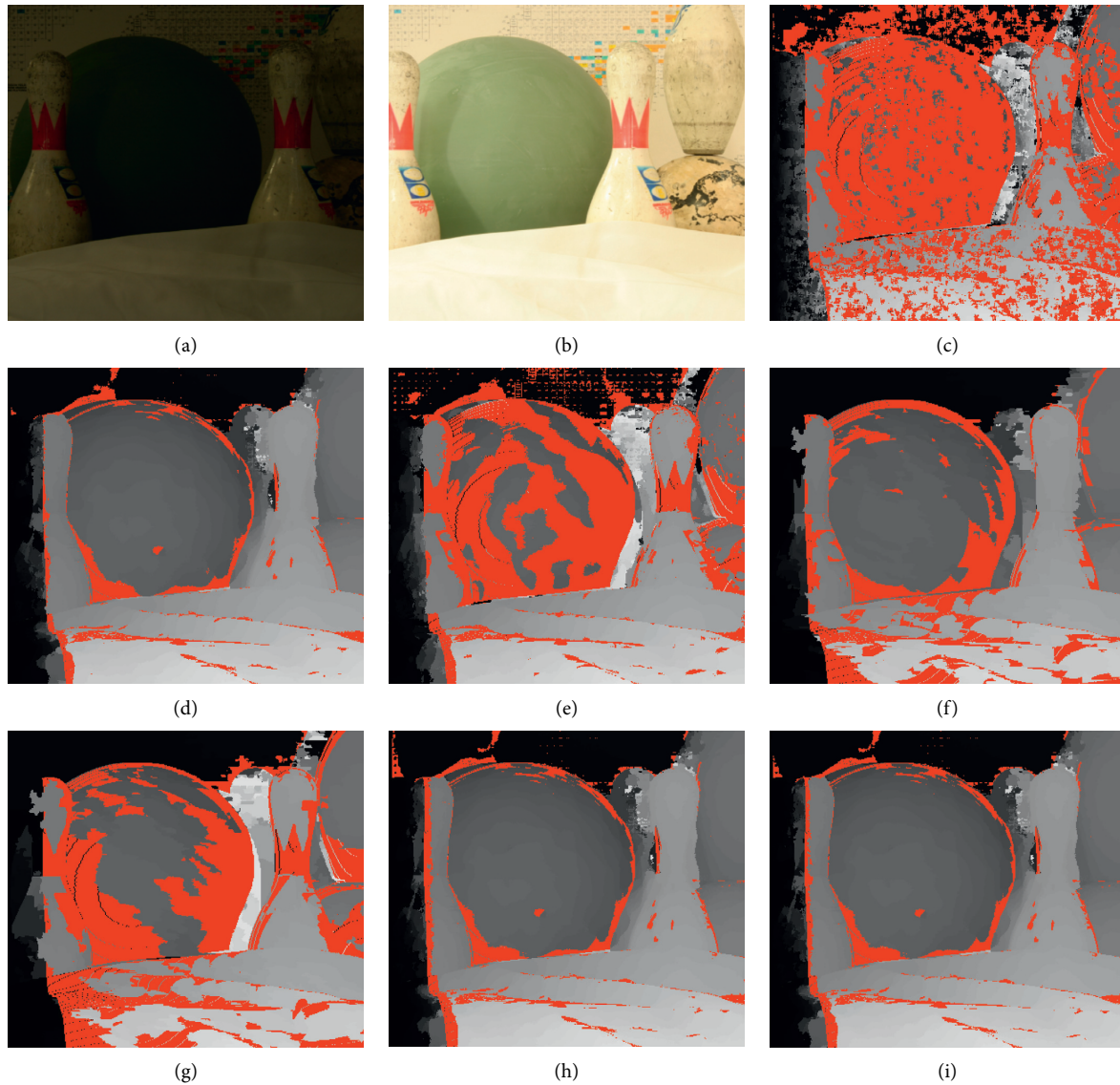


FIGURE 10: Bowling1's disparity maps generated by different algorithms when exposure conditions are not consistent. (a) Left image; (b) right image; (c) SAD; (d) CG-GF; (e) AG-GF; (f) CG-MST; (g) AG-MST; (h) proposed RGB; (i) proposed HSV.

remaining groups of test charts are better than the proposed RGB, which verifies the superiority of the algorithm using HSV color space. When the lighting conditions change, the proposed HSV ranks first in 5 groups of the 6 groups of test images, which has an absolute advantage. Especially in the Baby2 figures, its accuracy is 8.18% ahead of AG-GF, which has a good performance. In addition, from the point of view of the average mismatch rate, even the excellent method called CG-GF, its average mismatch rate is still 5.69% higher than the proposed method, which illustrates that the matching effect of the proposed method is far better than the other 5 methods in the condition of the changing lighting scenes.

Experimental results have proved that the proposed method can reduce the impact caused by inconsistent exposure and lighting conditions and improve the matching accuracy under the amplitude distortion environment.

### 3.3. Experiments on Middlebury Stereo Evaluation-Version 3.

In order to verify the performance of the proposed algorithm, this paper selects Motorcycle, Playroom, Playtable, and Vintage in Middlebury Stereo Evaluation-Version 3 to conduct experimental comparisons. The comparison results are shown in Figure 15. Figures 15(i)–15(l) show the performance of the algorithm before improvement. Figures 15(m)–15(p) show the performance of the algorithm after improvement. It can be seen that the algorithm in this paper suppresses the error rate of pixel matching, which is better than the algorithm before improvement. This article uses PSNR as the evaluation index for quantitative analysis, and the PSNR after the improvement is shown in Table 8. The PSNR of the proposed algorithm in Motorcycle, Playroom, and Playtable has been improved. Among them, the algorithm has improved by 1.35 dB on Playtable. It was only slightly weaker on Vintage than the algorithm before the improvement, which reflects the good effect of the algorithm.

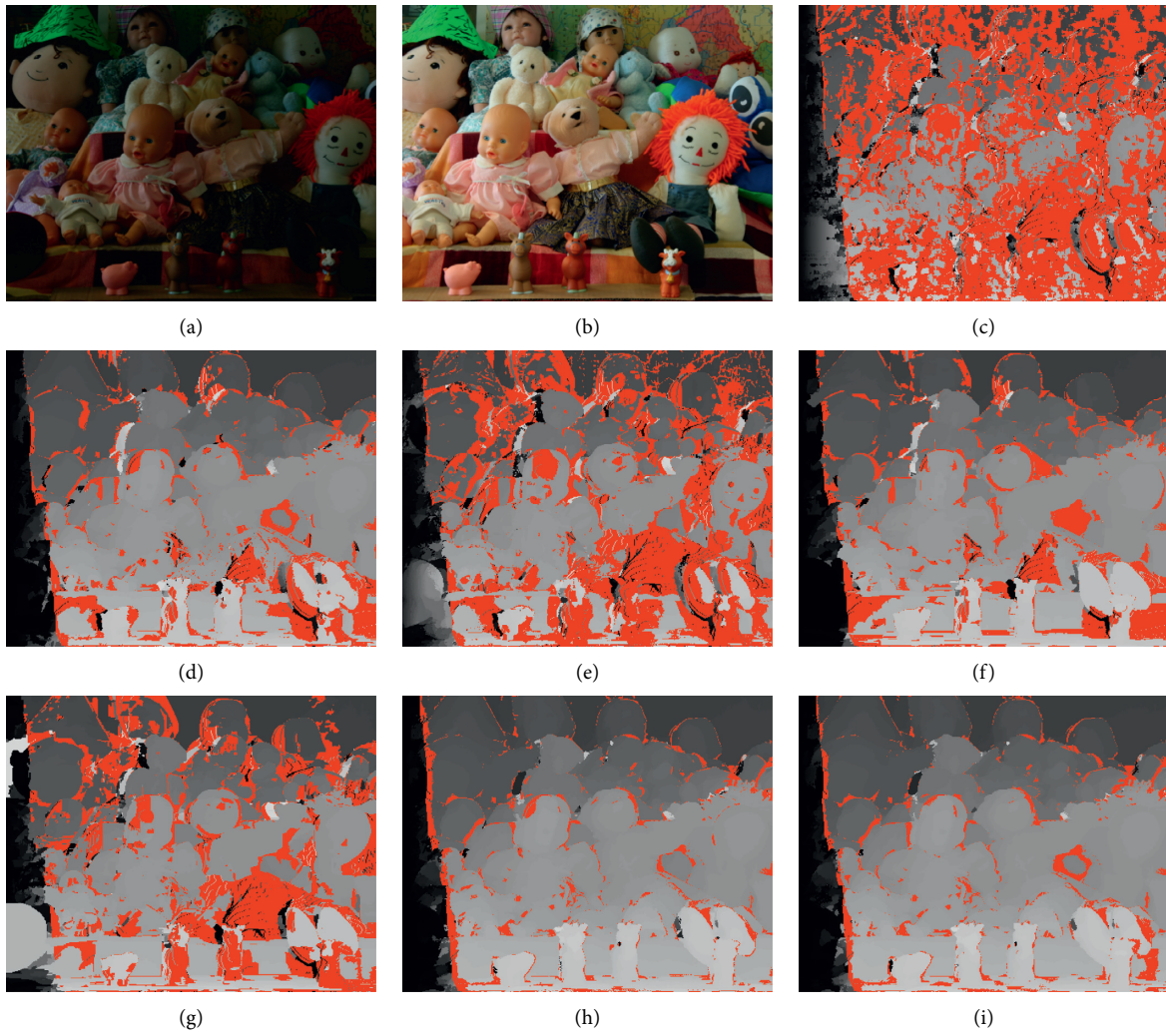


FIGURE 11: Dolls' disparity maps generated by different algorithms when exposure conditions are not consistent. (a) Left image; (b) right image; (c) SAD; (d) CG-GF; (e) AG-GF; (f) CG-MST; (g) AG-MST; (h) proposed RGB; (i) proposed HSV.

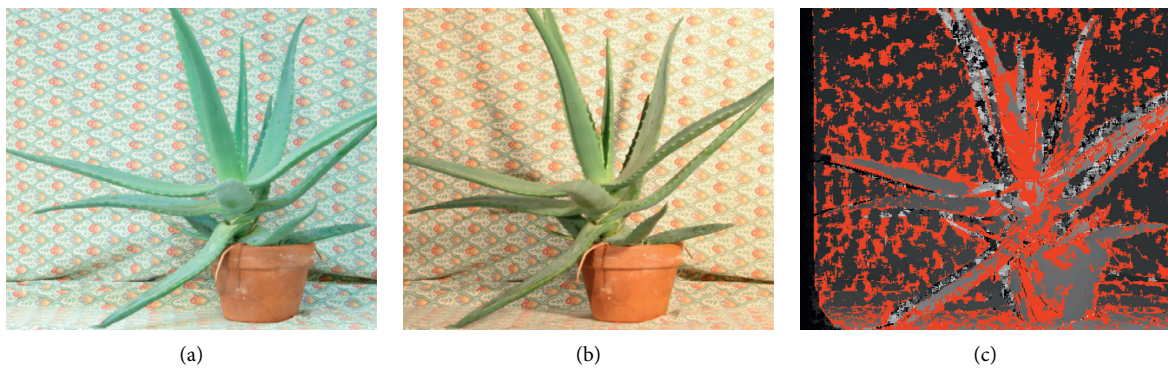


FIGURE 12: Continued.

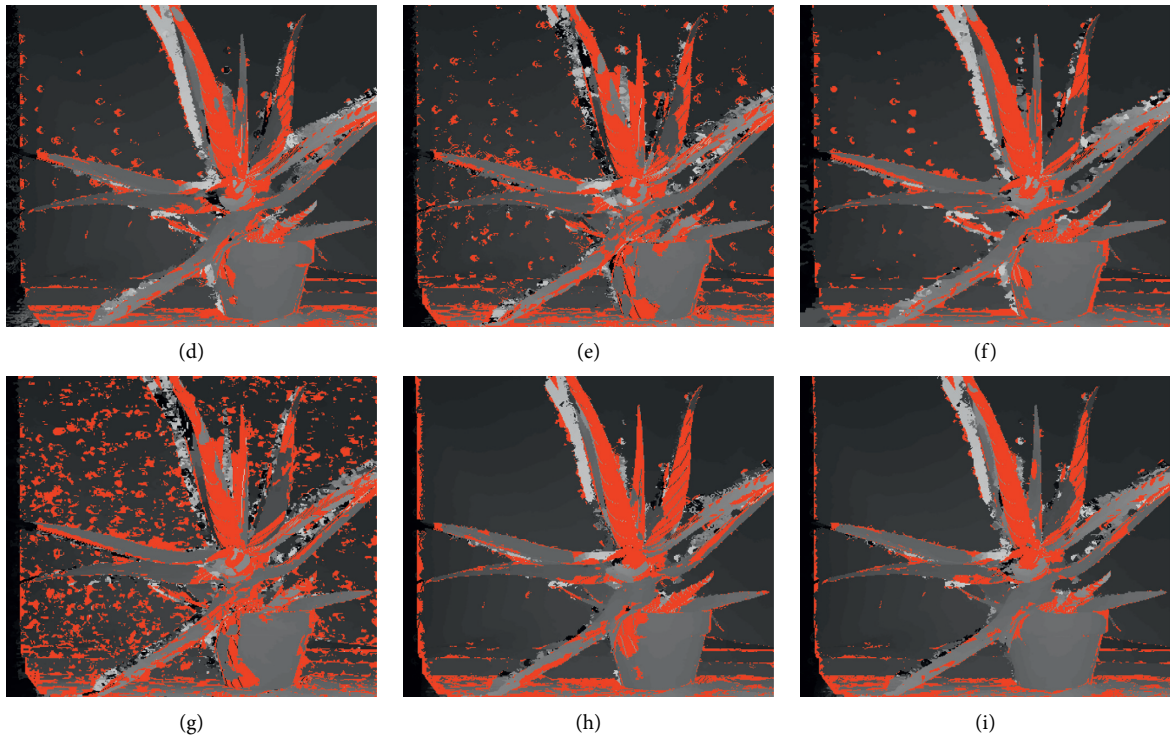


FIGURE 12: Aloe's disparity maps generated by different algorithms when illumination conditions are not consistent. (a) Left image; (b) right image; (c) SAD; (d) CG-GF; (e) AG-GF; (f) CG-MST; (g) AG-MST; (h) proposed RGB; (i) proposed HSV.

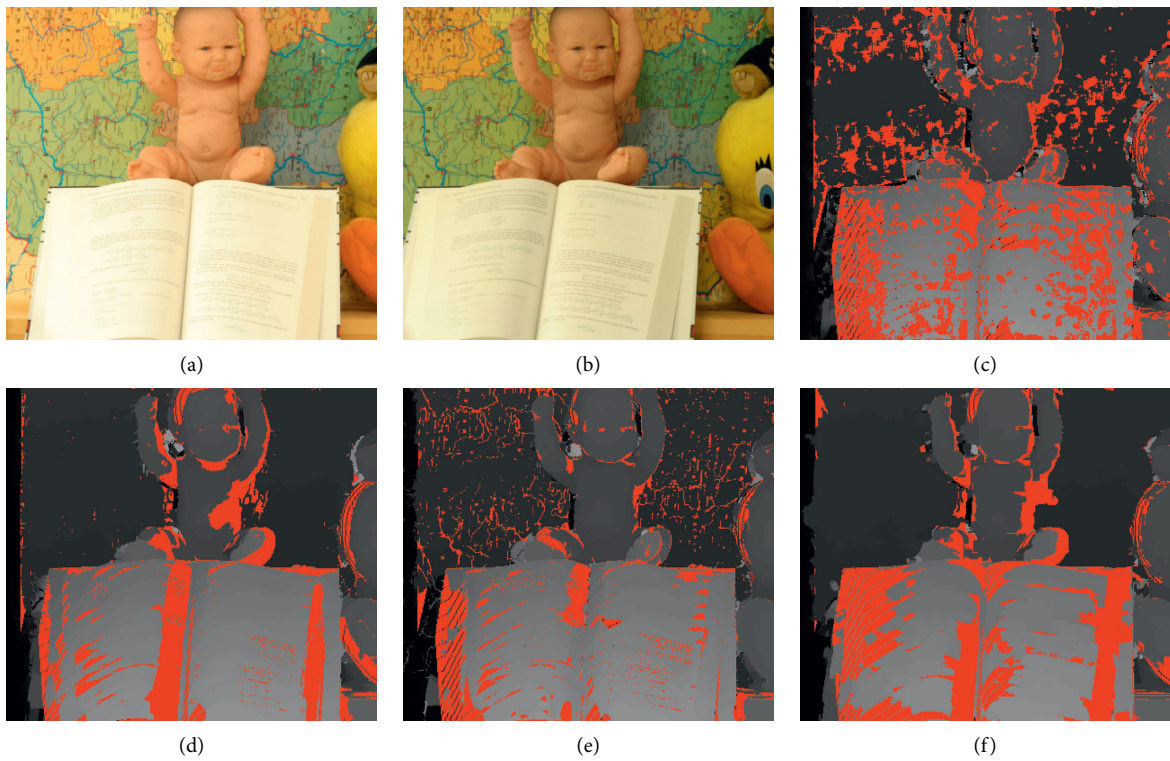


FIGURE 13: Continued.

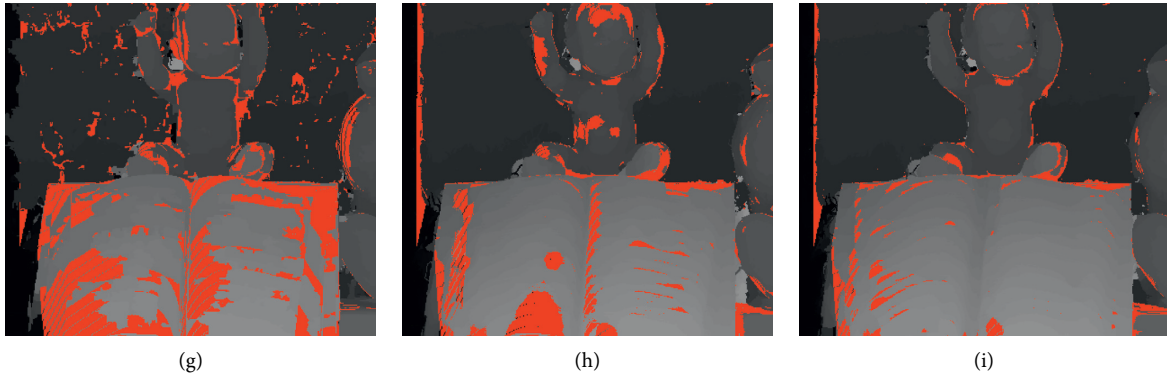


FIGURE 13: Baby2's disparity maps generated by different algorithms when illumination conditions are not consistent. (a) Left image; (b) right image; (c) SAD; (d) CG-GF; (e) AG-GF; (f) CG-MST; (g) AG-MST; (h) proposed RGB; (i) proposed HSV.

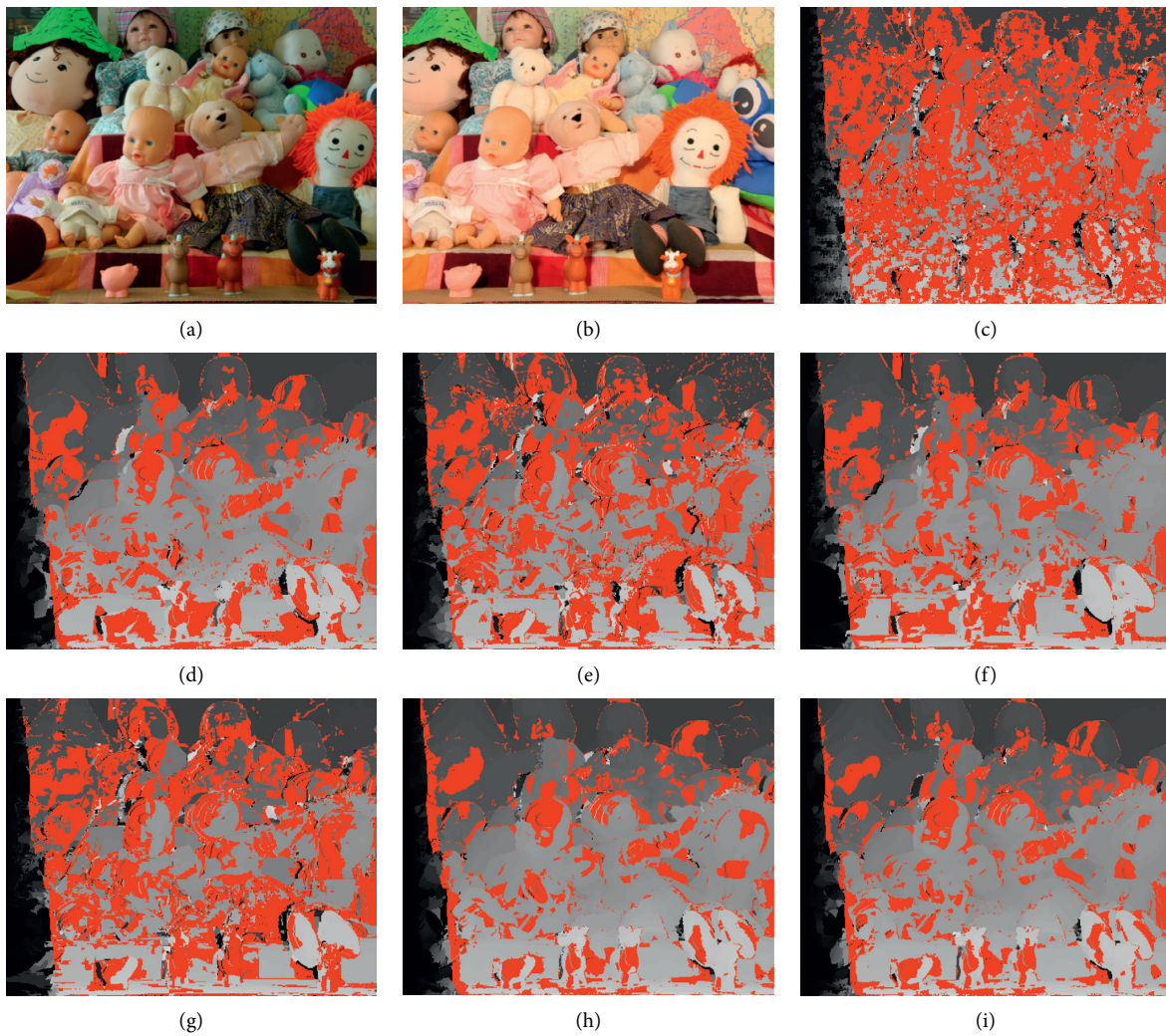


FIGURE 14: Dolls' disparity maps generated by different algorithms when illumination conditions are not consistent. (a) Left image; (b) right image; (c) SAD; (d) CG-GF; (e) AG-GF; (f) CG-MST; (g) AG-MST; (h) proposed RGB; (i) proposed HSV.

TABLE 6: Error matching rates in nonoccluded regions of different methods under different exposures.

Algorithm	Aloe	Art	Baby1	Bowling1	Cloth1	Dolls	Average
SAD	29.78	41.87	20.17	62.01	31.34	63.64	41.47
CG-GF	<b>6.92</b>	15.51	<b>2.58</b>	<b>7.70</b>	<b>1.54</b>	16.67	8.49
AG-GF	13.05	22.97	10.16	37.23	7.48	41.30	22.03
CG-MST	7.00	17.55	8.32	19.48	1.28	19.95	12.26
AG-MST	22.92	24.06	10.20	36.29	9.47	31.24	22.36
Pro-RGB	7.98	14.63	4.18	8.84	2.52	<b>8.14</b>	7.41
Pro-HSV	7.08	<b>14.07</b>	3.39	8.22	1.83	8.90	<b>7.25</b>

TABLE 7: Error matching rates in nonoccluded regions of different methods under different illumination conditions.

Algorithm	Aloe	Art	Baby2	Bowling1	Cloth1	Dolls	Average
SAD	39.05	58.65	27.95	35.58	13.24	66.71	40.20
CG-GF	15.52	36.82	16.37	20.42	2.23	31.41	20.46
AG-GF	19.98	39.33	13.29	15.66	3.65	46.94	23.14
CG-MST	15.09	36.30	21.01	22.05	<b>1.57</b>	32.96	21.50
AG-MST	28.05	38.10	19.58	22.52	8.98	48.62	27.64
Pro-RGB	17.94	37.69	10.8	15.11	2.62	<b>25.77</b>	18.32
Pro-HSV	<b>13.49</b>	<b>31.10</b>	<b>5.11</b>	<b>10.76</b>	1.95	26.20	<b>14.77</b>



FIGURE 15: Performance verification of the improved census method.

TABLE 8: PSNR for the proposed method

Numbers of Figure 15	15(i)	15(m)	15(j)	15(n)	15(k)	15(o)	15(l)	15(p)
PSNR (dB)	10.79	10.85	8.72	9.06	10.64	11.99	3.23	3.22

## 4. Conclusions

This paper proposes a new stereo matching algorithm based on HSV color space and improved census transform, which effectively suppresses the effects of ambient noise and exposure lighting inconsistencies in the stereo matching. First, the RGB image is converted into HSV color space by establishing the HAD cost calculation function, and it reduces the impact of amplitude distortions. Then, to address the problem of traditional census transform overrelying on central pixels, an improved census method for neighborhood weighting is proposed. Last but not the least, the HAD cost and improved census cost are mapped to the  $[0, 1]$  and further merged. In the stage of cost aggregation, an outlier culling method based on confidence interval is proposed to further reduce the effect of noise and improve the matching accuracy in nonideal situations. Experimental results on noise and amplitude distortion environment demonstrate that the proposed method can not only effectively suppress the influence of noise but also maintain a low mismatch rate under scenes with changing exposure and lighting conditions. In the follow-up research, the matching accuracy of the proposed algorithm will be further optimized in the depth discontinuous area.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

## Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities under Grant no. 2020QN49.

## References

- [1] Q. Kou, D. Cheng, L. Chen, and Y. Zhuang, "Principal curvatures based local binary pattern for rotation invariant texture classification," *OPTIK*, vol. 193, Article ID 162999, 2019.
- [2] L. Chen, Q. Kou, D. Cheng, and J. Yao, "Content-guided deep residual network for single image super-resolution," *OPTIK*, vol. 202, Article ID 163678, 2020.
- [3] P. Kaur, B. Singh Khehra, and A. P. Singh Pharwaha, "Deep transfer learning based multiway feature pyramid network for object detection in images," *Mathematical Problems in Engineering*, vol. 2021, Article ID 5565561, 13 pages, 2021.
- [4] Z. S. Chafi and H. Afrakhte, "Short-term load forecasting using neural network and particle swarm optimization (PSO) algorithm," *Mathematical Problems in Engineering*, vol. 2021, Article ID 5598267, 10 pages, 2021.
- [5] Z. Zanzan, W. Hui, W. Qiong, and L. Zhiguang, "3D shape reconstruction based on digital holography of stereo matching technology," *Laser & Optoelectronics Progress*, vol. 51, no. 11, Article ID 110901, 2014.
- [6] A. Howard, "Real-time stereo visual odometry for autonomous ground vehicles," in *Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3946–3952, IEEE, Nice, France, September 2008.
- [7] 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.
- [8] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," in *Proceedings of the 8th IEEE International Conference on Computer Vision, ICCV 2001*, pp. 508–515, Vancouver, Canada, July 2001.
- [9] J. X. Xu, Q. Li, Y. Liu et al., "Stereo matching algorithm based on color weights and tree dynamic programming," *Acta Optica Sinica*, vol. 37, no. 12, Article ID 1215007, 2017.
- [10] O. Veksler, "Stereo correspondence by dynamic programming on a tree," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 384–390, IEEE, San Diego, CA, USA, June 2005.
- [11] D. Zhang, "The algorithm of WTA stereo matching based on the human detection method," *Electronic Test*, vol. 5, pp. 66–68, 2013.
- [12] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," *European Conference on Computer Vision*, Springer, Berlin, Germany, pp. 151–158, 1994.
- [13] S. P. Zhu, L. N. Yan, and L. Zheng, "Stereo matching algorithm based on improved census transform and dynamic programming," *Journal of the Optical*, vol. 36, no. 4, pp. 216–224, 2016.
- [14] S. T. Chen, Y. L. Wu, and Y. Y. Zhang, "S-census transform algorithm with variable cost," *Computer Engineering and Design*, vol. 39, no. 2, pp. 414–419, 2018.
- [15] N. Y.-C. Chang, T.-H. Tsai, B.-H. Hsu, Y.-C. Chen, and T.-S. Chang, "Algorithm and architecture of disparity estimation with mini-census adaptive support weight," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 6, pp. 792–805, 2010.
- [16] R. Zhao, W. Ouyang, and X. Wang, "Cross-scale cost aggregation for stereo matching," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 590–1597, Columbus, OH, USA, June 2014.
- [17] A. Hosni, M. Bleyer, and M. Gelautz, "Secrets of adaptive support weight techniques for local stereo matching," *Computer Vision and Image Understanding*, vol. 117, no. 6, pp. 620–632, 2013.
- [18] M. J. Hannah, *Computer Matching of Areas in Stereo Images*, Stanford University, Stanford, CA, USA, 1974.
- [19] K. He, J. Sun, and X. Tang, "Guided image filtering," *European Conference on Computer Vision*, Springer, Berlin, Germany, pp. 1–14, 2010.



- [20] Q. Yang, "A non-local cost aggregation method for stereo matching," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1402–1409, IEEE, Providence, RI, USA, June 2012.
- [21] T. Kanade, H. Kano, S. Kimura, A. Yoshida, and K. Oda, "Development of a video-rate stereo machine," in *Proceedings 1995 IEEE/RSJ International Conference on Intelligent Robots and Systems. Human Robot Interaction and Cooperative Robots*, vol. 3, pp. 95–100, IEEE, Pittsburgh, PA, USA, August 1995.