

# Optical Engineering

OpticalEngineering.SPIEDigitalLibrary.org

## Improved census transform for noise robust stereo matching

Jongchul Lee  
Daeyoon Jun  
Changkyoung Eem  
Hyunki Hong

# Improved census transform for noise robust stereo matching

Jongchul Lee,<sup>a</sup> Daeyoon Jun,<sup>b</sup> Changkyoung Eem,<sup>c</sup> and Hyunki Hong<sup>d,\*</sup>

<sup>a</sup>GSAIM, Chung-Ang University, Department of Imaging Science and Arts, 221 Huksuk-dong, Dongjak-ku, Seoul 156-756, Republic of Korea

<sup>b</sup>Chung-Ang University, School of Integrative Engineering, 221 Huksuk-dong, Dongjak-ku, Seoul 156-756, Republic of Korea

<sup>c</sup>Chung-Ang University, College of ICT Engineering, 221 Huksuk-dong, Dongjak-ku, Seoul 156-756, Republic of Korea

<sup>d</sup>Chung-Ang University, School of Integrative Engineering, 221 Huksuk-dong, Dongjak-ku, Seoul 156-756, Republic of Korea

**Abstract.** Census transform (CT), a stereo matching algorithm, has a strong advantage in radial distortion and brightness changes. However, CT is noise-sensitive because it compares the brightness of a single central pixel based on the brightness values of neighborhood pixels within a matching window. Star-census transform, which compares the brightness of pixels separated by a certain distance along a symmetrical pattern within the matching window, is presented. The proposed method can select the distance between the pixels for comparison and comparison patterns. The experiment results show that the proposed method yields a better performance than the previous CT methods. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.OE.55.6.063107](https://doi.org/10.1117/1.OE.55.6.063107)]

Keywords: stereo matching; 3-D depth estimation; census transform; noise robustness.

Paper 160461 received Mar. 28, 2016; accepted for publication May 31, 2016; published online Jun. 16, 2016.

## 1 Introduction

Stereo vision establishes the correspondence between both views as stereo matching and calculates the dense disparity and three-dimensional (3-D) depth information. The stereo matching method is largely divided into two parts: local and global. Also, the stereo matching algorithm consists of several steps, including initial matching cost calculation, cost aggregation, cost calculation, and refinement.<sup>1</sup>

The local method applies a matching window to find the correspondence between the reference image and the target image. It is more efficient than the global method because it searches only the designated area, whereas the global method explores the entire image area including the neighborhood area. The global method defines an energy model by using various conditions, such as uniqueness and continuity, and determines the matching information by minimizing the energy function of an entire image. The global method can obtain more exact difference values than the local method because it processes a search of the entire image repeatedly. However, the implementation of the global method is complex and is not suitable for real-time processing due to the amount of computational complexity. Representative examples of the global method include belief propagation,<sup>2,3</sup> graph-cut,<sup>4</sup> and dynamic programming.<sup>5</sup>

For noise robustness in stereo matching, we improve the original census transform (CT), which is one of the most widely used local methods to calculate the initial matching cost. Hirsh Müller and Scharstein<sup>6</sup> compared the performance of various stereo matching algorithms including CT, based on the image change due to different camera exposures and light conditions. CT compares the relative brightness between two pixels and converts the comparison result into a bit-string. The local methods, such as sum of absolute or squared difference and normalized cross correlation,

compare the values of all pixels in the matching window. CT obtains a robust result toward brightness change and radiometric distortion because it determines only the relative high and low level of brightness between two pixels.<sup>7</sup> However, CT is noise-sensitive because it compares the relative difference of brightness of a pixel with neighborhood pixels based on one single central pixel. In other words, the probability of false matching greatly increases when the central pixel is affected by noise or other conditions.<sup>8,9</sup>

Several suggestions have been proposed, such as the mini-census transform (MCT)<sup>8</sup> and generalized-census transform (GCT),<sup>9</sup> to improve the performance of the existing CT. CT compares all the pixels in the neighborhood based on the central pixels within the matching window; thus, if the size of the matching window increases, the computational complexity also increases. MCT compares only six pixels of the neighborhood, which are selected empirically, with a central pixel.<sup>8</sup> This method has shown good performance with less computational complexity than the existing CT. Furthermore, GCT was proposed for robust matching performance toward Gaussian noise.<sup>9</sup> This method applies pixels separated by a certain distance within the matching window, not the central pixel. MCT has a drawback whereby it is affected by noise due to the characteristic of comparing the brightness values of neighborhood pixels based on a single central pixel. On the other hand, GCT is robust to noise because it compares the neighborhood pixels to each other and the pixels are symmetrically set within a matching mask.

This paper introduces star-census transform (SCT), which compares the pixels separated by a certain distance symmetrically. The purpose of the proposed method is to initiate the sampling of neighborhood pixels in a symmetrical pattern excluding the central pixels of the matching window. It then compares the previous and current sampling points consecutively along a scan pattern. Compared with GCT, which compares the brightness of sampling points separated by a certain distance, the proposed method is more robust to

\*Address all correspondence to: Hyunki Hong, E-mail: [honghk@cau.ac.kr](mailto:honghk@cau.ac.kr)

noise due to its comparison of brightness values between the sampling points. The sampling points and the distance (edge length) between these points can be diversely selected, depending on the correlation with the central pixel of the matching window, the degree of noise within the image, and the computational complexity requirement. The proposed SCT can be utilized in areas such as real-time stereo matching, as well as feature points matching and tracking.

## 2 Census Transform

CT compares the brightness values between the pixels.<sup>6,7</sup> Equations (1) and (2) show the CT equation and comparison process between brightness values. It describes the bit-string  $C(p)$  as a value compared by operator  $\otimes$  subjected to the brightness value  $I_p$  of central pixel  $p$  and the brightness value  $I_q$  of neighborhood pixel  $q$  in a matching window  $W$ . Here, function  $\xi(\cdot)$  returns 1 if the brightness value of the neighborhood pixel is higher than the counterpart of the central pixel, and 0 if the brightness value of the neighborhood pixel is lower than the counterpart of the central pixel by comparing the brightness values between pixels.  $C(p)$  is the encoding bit-string and consists of 0 and 1. It refers to a relative brightness distribution with neighborhood pixels on the basis of the central pixel of the window.

The bit-string  $C_l(p)$  on the left reference image is computed, and then the bit-strings  $C_r(p+d)$  on the right target image are calculated within the search range of the maximum disparity of  $d$ . If the matching window is  $m \times m$ , the bit-string  $m \times m - 1$  is obtained. Also, as indicated in Eq. (3), the hamming distance value, which is implemented with exclusive logical OR arithmetic operation toward bit-strings of  $C_l(p)$  and  $C_r(p+d)$ , is calculated. Then, CT determines an initial difference value of pixel  $p$ , which is equal to the lowest cost value of the difference value  $d_{\text{int}}(p)$ .

$$C(p) = \otimes_{q \in W} \xi(I_p, I_q), \quad (1)$$

$$\xi(I_p, I_q) = \begin{cases} 1, & \text{if } I_p < I_q, \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

$$d_{\text{int}}(p) = \underset{d}{\operatorname{argmin}} \sum \operatorname{Hamming}[C_l(p), C_r(p+d)]. \quad (3)$$

MCT, which was proposed to reduce the computation complexity of the existing CT, compares only some parts of the pixels in the window based on the central pixel. In other words, on the  $5 \times 5$  sized matching window in Fig. 1(a), MCT only performs the comparison arithmetic

operation of the six dark-colored pixels.<sup>8</sup> GCT, in Fig. 1(b), is the CT method, which compares the neighborhood pixels with each other in a symmetrical pattern on the basis of the central pixel.<sup>9</sup> GCT is more robust to noise than the other methods because it compares the brightness values of various neighborhood pixels in the direction of the arrow, not the central pixel.

## 3 Proposed Method

### 3.1 Star-Census Transform

The existing CT compares the brightness values of the neighborhood pixels on the basis of the central pixel within the matching window; thus, if noise occurs, the false matching ratio greatly increases. The proposed SCT is a method used to compare the neighborhood pixels by symmetrical patterns in a consecutive manner, rather than comparing them with the central pixel. Here, it is crucial to select the sampling pixels, which are the subjects for comparison in the window.

Fire and Archibald<sup>9</sup> calculated the average degree of correlation between neighborhood pixels on the basis of the central pixel on Middlebury benchmark images: Tsukuba, Venus, Teddy, Cones, and so on. The result demonstrated that the highest correlation for the neighborhood pixels is shown for the distance of  $\sim 2$  pixels from the central pixel. Thus, checking the pixel area located 2 pixels from the central pixel of the matching window is an effective procedure for reducing the false matching. Based on this property, Fire proposed GCT, which compares the brightness values of neighborhood pixels in the symmetrical direction, without a central pixel.

To reduce the noise that can arise from a particular location such as a central pixel, the proposed SCT samples the pixels separated by a certain distance within the matching window and compares the brightness values of the corresponding points. Also, the scan pattern of each sampling point should be designed symmetrically to obtain the same results even from a rotated image. Here, we employ the Chebyshev distance on a spatial space ( $x$  and  $y$  coordinates), where the distance between two pixels is the greatest of their differences along any coordinate dimension.

All the sample points for comparison in the matching window are connected, and the last sampling point is compared to the initial sampling point. In other words, the relative brightness heights of all sample points are analyzed along the scan pattern. The proposed method compares two neighborhood pixels consecutively on the basis of a random sample point in the matching window, so the matching accuracy is higher than that of the existing CT. Also, the initial sample point is compared to the last sample point. Thus, the scan pattern is connected as one line. This means that the distribution of the relative level of brightness values of all pixels can be measured. For example, a  $5 \times 5$  sized matching window can be compared with the sample points of the distance from 1 to 4, and the maximum distance increases depending on the size of the matching window. The proposed method can produce a variety of patterns depending on the size of the matching window, the compared distance between pixels, and the position of sampling points. Figures 2(a) and 2(b) show the SCT pattern with an edge length (comparison distance) of 2 in a  $3 \times 3$  matching window and an edge length (comparison distance) of 3 in a  $5 \times 5$  matching window:

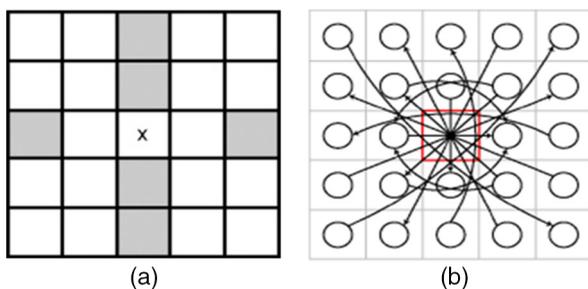
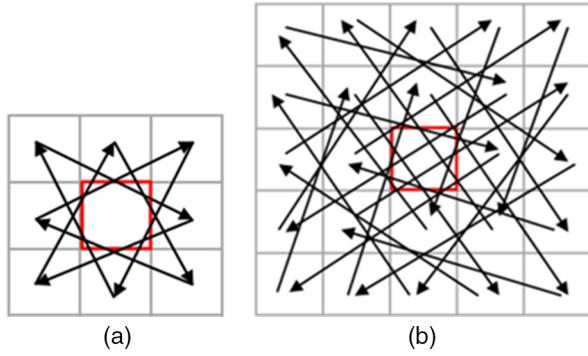


Fig. 1 (a) MCT<sup>8</sup> and (b) GCT.<sup>9</sup>



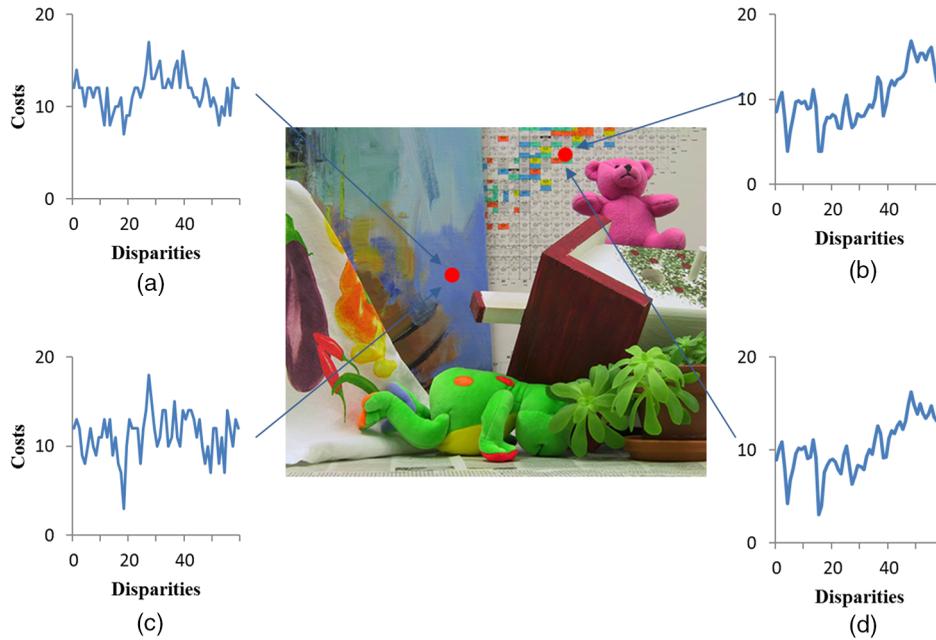
**Fig. 2** Basic pattern of SCT: matching windows of (a)  $3 \times 3$  (edge distance 2) and (b)  $5 \times 5$  (edge distance 3).

$$\xi(I_q, I_{q'}) = \begin{cases} 1, & \text{if } I_q - I_{q'} > 0 \\ 0, & \text{otherwise} \end{cases}, \quad (4)$$

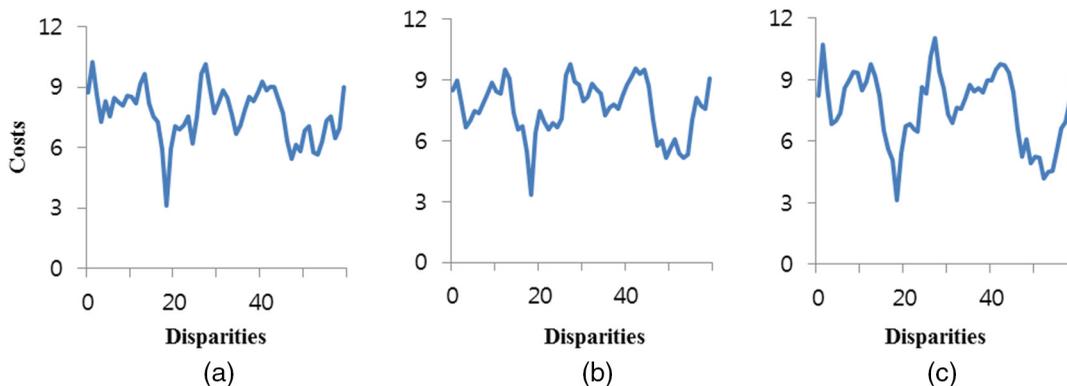
$$C(p) = \bigotimes_{q \in W} \xi(I_q, I_{q'}). \quad (5)$$

The proposed SCT can be defined by Eqs. (4) and (5). In Eq. (4), if the difference between brightness value  $I_q$  of neighborhood pixel  $q$  (excluding the central pixel  $p$  of the matching window) and the brightness value of  $I_{q'}$  of the neighborhood pixel  $q'$ , which is separated by a certain distance) is greater than 0, then it returns 1. Otherwise, the remaining pixels return 0. By using operator  $\bigotimes$  in Eq. (5), the comparison result of brightness values of pixels in the matching window of size  $W$  is converted into a bit-string.

The brightness values of the sampling points in a consecutive manner depending on the scan patterns in matching window are compared. The number of sampling points in which brightness has been compared is doubled in the same sized matching window, so it has a more robust performance for stereo matching. Figure 3 describes the average matching cost distribution obtained by CT and SCT, from two random areas (indicated in the red circle) of the Teddy image, which is the Middlebury standard image. These two areas are front-parallel planes, which face the camera and toward each other. The ground truth difference values of these two random areas are 18 and 15. Figures 3(a) and 3(b) show the distribution of



**Fig. 3** Comparison of matching costs of [(a) and (b)] CT, and [(c) and (d)] SCT.



**Fig. 4** Comparison of matching costs with edge length of (a) 2, (b) 3, and (c) 4.

**Table 1** PKRN values of obtained disparity maps.

	Tsukuba	Venus	Teddy	Cones	Avg.
CT	1.160	1.152	1.127	1.144	1.146
SCT (24-2-1)	1.164	1.172	1.143	1.170	1.162
SCT (24-3-1)	1.158	1.163	1.136	1.161	1.154

**Table 2** PKRN values of inlier disparity maps.

	Tsukuba	Venus	Teddy	Cones	Avg.
CT	1.199	1.197	1.165	1.191	1.188
SCT (24-2-1)	1.204	1.213	1.182	1.217	1.204
SCT (24-3-1)	1.195	1.203	1.173	1.205	1.194

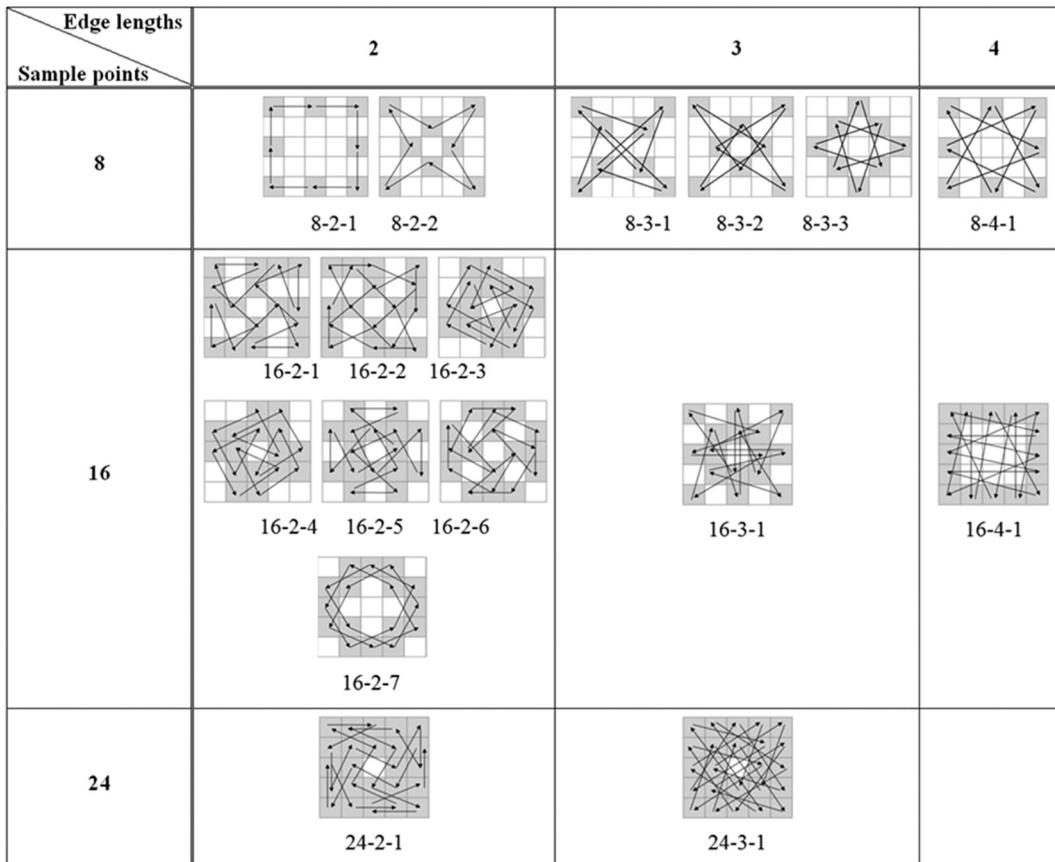
matching cost from CT, and Figs. 3(c) and 3(d) show those from SCT. Figure 3(b) shows two minimum errors of the minimum cost values; it is therefore difficult to obtain exact disparity information. However, in Figs. 3(c) and 3(d), the minimum matching cost values are clearly obtained at 18 and 15 disparity levels, respectively.

Figure 4 shows the matching cost distribution according to the edge length (the distance between the sample points)

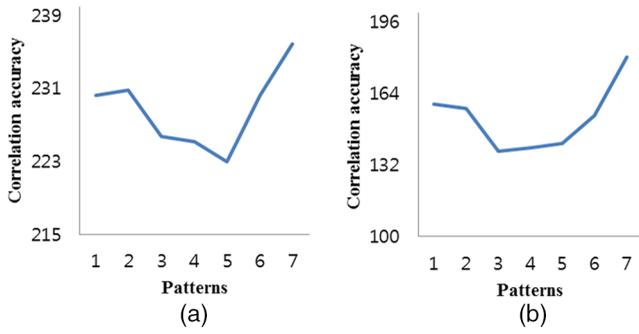
of SCT. In the case of the 24 sampling points in the  $5 \times 5$  matching window, there is no comparison pattern at the edge length of 4: pixels are sampled redundantly. We therefore sampled 16 points in the  $5 \times 5$  matching window in Fig. 4. The average matching cost in the marked areas of Fig. 3 (Teddy image) is then computed. In other words, Figs. 4(a)–4(c) indicate the cost error distribution with 16 sampling points in the matching window when the edge lengths are set to 2, 3, and 4, respectively. Compared to Fig. 4(c), the case of Fig. 4(a) is relatively easier to distinguish the location of the minimum matching cost. In this case, the edge length between sample points in the matching window is set as 2; we can thus obtain more reliable distribution of matching cost than in other cases. This result is consistent with the analysis result of the previous study.<sup>9</sup>

In the three cases shown in Fig. 4, peak-ratio naive (PKRN) is employed to measure the method of obtaining the reliable disparity value of the minimum matching cost obtained from the Winner Takes All method.<sup>10,11</sup> PKRN determines the reliability of the final disparity value as Eq. (6), by calculating the proportion of the disparity value of the minimum matching cost ( $C_1$ ) and the disparity value of the second minimum matching cost ( $C_2$ ). Generally, if the difference between  $C_1$  and  $C_2$  increases (if PKRN value becomes larger), it is determined that the reliable disparity value is obtained:

$$PKRN = \frac{C_2 + \epsilon}{C_1 + \epsilon}, \tag{6}$$



**Fig. 5** SCT scan pattern according to sampling points and edge lengths.



**Fig. 6** Comparison of summation of on average correlation values of seven different patterns with (a) no noise and (b) Gaussian noise.

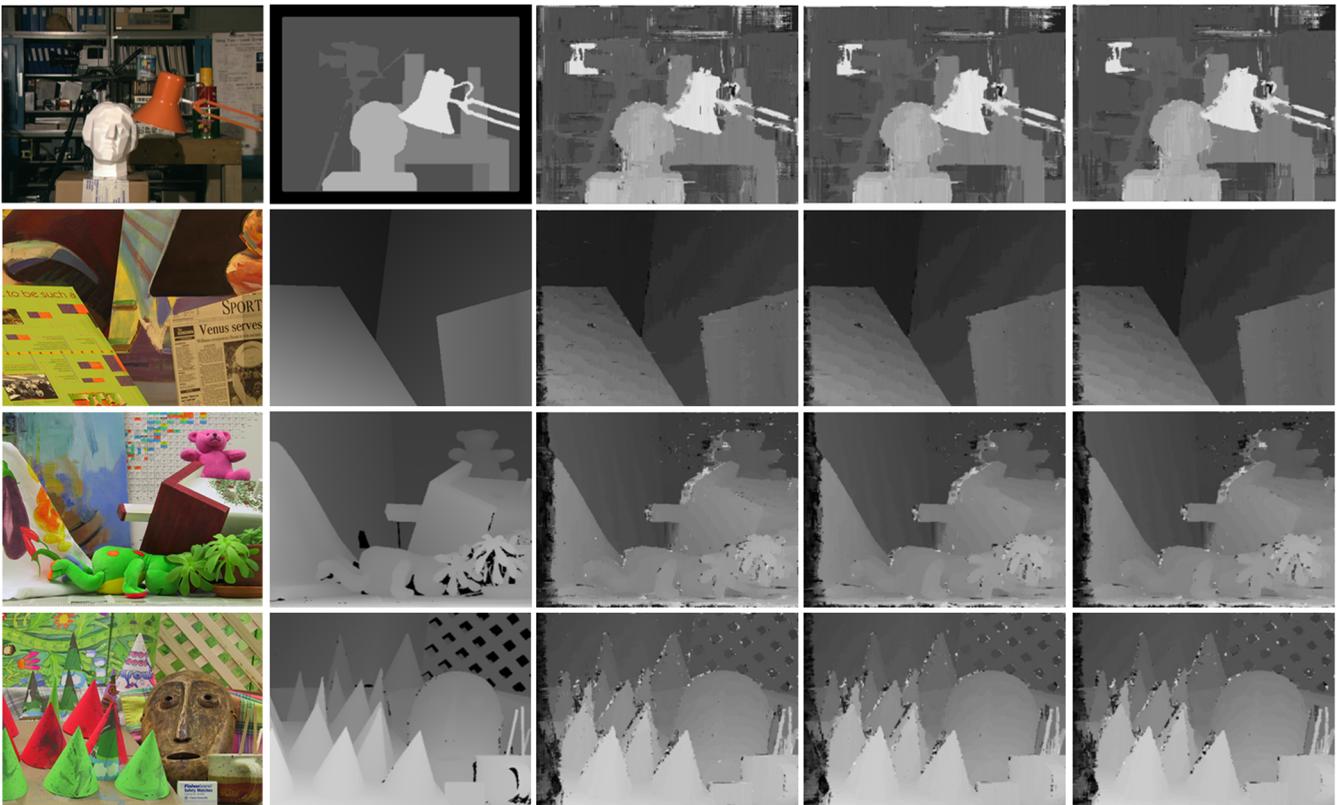
where  $\epsilon$  is a value for the case when the minimum matching cost ( $C_1$ ) is 0 ( $\epsilon = 10$ ). In Fig. 4, PKRN is used just to determine the reliability of the final disparity value according to three edge length parameters.  $\epsilon$  in Eq. (6) has no influence on the matching performance.

Table 1 describes the PKRN reliability results on the disparity value of Middlebury standard images (Tsukuba, Venus, Teddy, and Cones). When 24 points are sampled in a  $5 \times 5$  matching window, PKRN results by the SCT method are greater than those by the CT method. Here, “24-2-1” refers to the first of scan patterns with 24 sample points and the edge length of 2. Table 2 shows the PKRN results from inlier areas by a left-right consistency check. From Tables 1 and 2, the proposed SCT method with the edge length of 2 can obtain better reliable disparity information than the previous CT method.

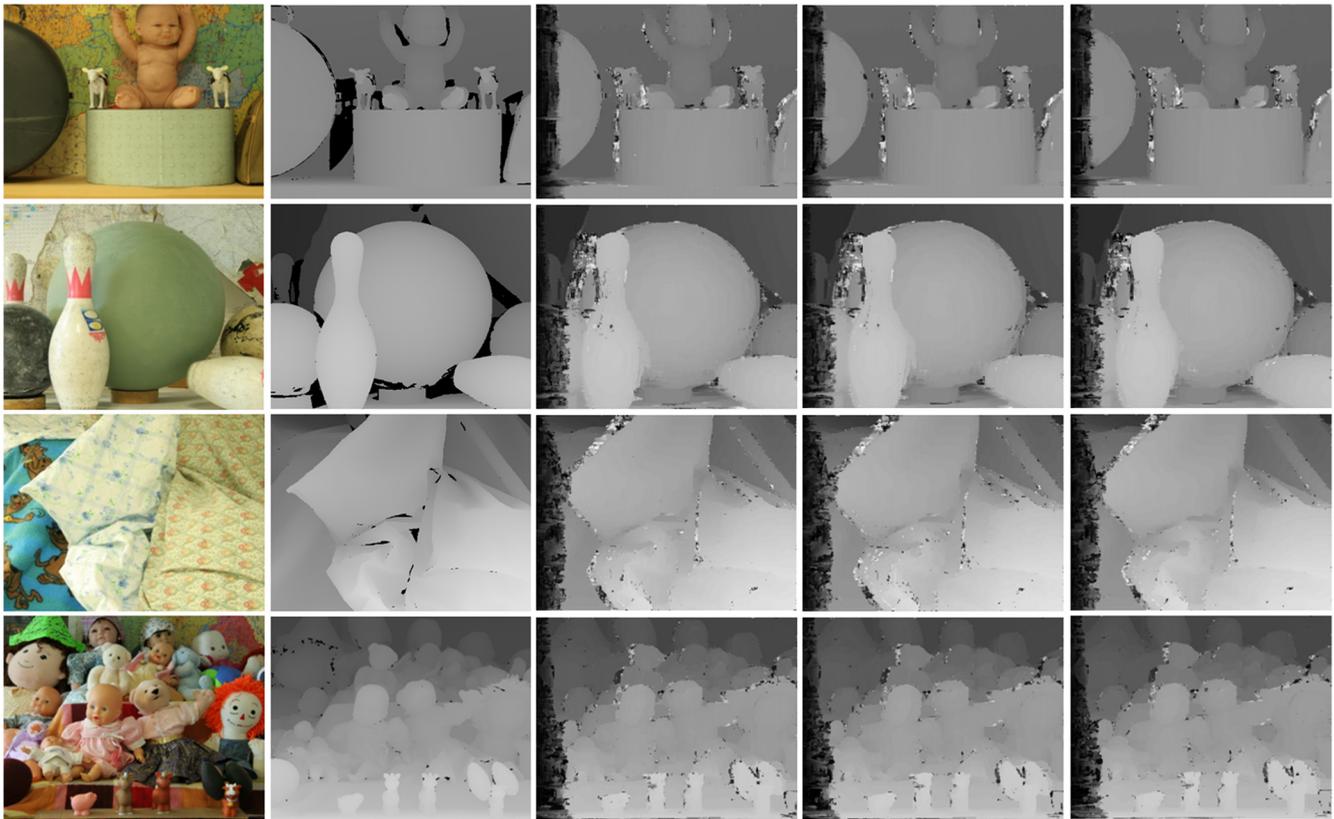
### 3.2 Selection of Sample Points

The proposed SCT can perform a comparison operation up to 24 times in the  $5 \times 5$  matching window. When the matching window size increases, the number of bit-strings produced from the comparison operation increases. If the number of sampling points is reduced based on the scan pattern, the number of comparison operations also decreases. The edge length (comparison distance) in the  $5 \times 5$  matching window is 1 to 4. To ensure the accuracy and robustness against noise, the proposed method considers the patterns with edge lengths of 2 to 4, excluding the distance of 1. Figure 5 shows the possible patterns on the  $5 \times 5$  matching window according to the edge lengths and the number of sample points (shown as dark color). For example, if eight points are sampled, three patterns can be generated: (1) two scan patterns with the edge length of 2, (2) three scan patterns with the edge length of 3, and (3) one scan pattern with the edge length of 4. In Fig. 5, “8-2-1” refers to the first scan pattern with eight sample points and an edge length of 2. The initial sample point and the last sample point are connected, and the sample points are consecutively compared along their symmetrical pattern.

Even if the edge lengths are same (if two pixels are separated by the same distance) in the matching window, we can choose several patterns, as shown in Fig. 5. When pixels in the matching mask are sampled, each point in the neighboring pixels has been correlated differently with a central pixel. Therefore, it is crucial to determine which pixel should be placed in a scan pattern in the area-based stereo matching window. Previous studies<sup>9</sup> suggested the average correlation



**Fig. 7** Tsukuba, Venus, Teddy, and Cones images, ground truth, and disparity map obtained from proposed SCT: sample points of 8 (8-2-2), 16 (16-2-2), and 24 (24-2-1) (from left to right).



**Fig. 8** Baby 3, Bowling 2, Cloth 2, and Dolls images, ground truth, and disparity map obtained from proposed SCT: 8 sample points (8-2-2), 16 (16-2-2), and 24 (24-2-1) (from left to right).

**Table 3** Comparison of false matching ratio results (nonocclusion regions).

Methods	Tsukuba	Venus	Teddy	Cones	Avg.	Baby3	Bowling2	Cloth2	Dolls	Avg.	
CT (full $5 \times 5$ )	8.20	1.78	8.15	4.96	5.77	3.56	6.64	3.78	6.76	5.18	
MCT	7.56	1.47	8.26	5.38	5.66	4.06	7.29	5.32	9.59	6.56	
GCT (12 edges)	6.20	1.65	7.72	6.04	5.40	4.76	9.09	5.66	10.42	7.48	
GCT (16 edges)	5.49	1.72	7.84	5.95	5.25	4.61	8.39	5.04	9.47	6.87	
SCT 8 points	8-2-2	<b>7.92</b>	<b>1.28</b>	<b>8.35</b>	<b>4.77</b>	<b>5.58</b>	<b>4.43</b>	<b>7.71</b>	<b>3.93</b>	<b>7.50</b>	<b>5.89</b>
	8-3-1	8.78	2.06	9.59	6.44	6.71	5.74	9.73	5.53	10.71	7.92
	8-4-1	7.05	2.24	11.02	7.22	6.88	6.81	11.15	6.40	12.17	9.13
SCT 16 points	16-2-1	6.49	0.93	8.13	4.49	5.01	4.36	7.48	4.01	7.47	5.83
	16-2-2	<b>6.29</b>	<b>0.92</b>	<b>7.93</b>	<b>4.31</b>	<b>4.86</b>	<b>4.18</b>	<b>7.37</b>	<b>3.81</b>	<b>7.15</b>	<b>5.62</b>
	16-2-3	8.16	1.55	8.16	5.01	5.72	4.66	8.00	4.48	8.28	6.35
	16-2-4	8.32	1.34	8.46	5.16	5.82	4.70	7.87	4.48	8.33	6.34
	16-2-5	7.78	1.18	8.59	5.21	5.69	4.97	7.83	4.65	8.89	6.58
	16-2-6	6.13	1.11	7.99	4.40	4.90	4.18	7.15	3.93	7.26	5.63
	16-2-7	7.84	1.23	8.24	4.68	5.49	4.53	7.44	3.92	7.36	5.81
SCT 24 points	16-3-1	7.10	1.29	7.56	4.97	5.23	3.90	6.97	4.44	8.48	5.94
	16-4-1	6.39	1.82	9.02	6.09	5.83	4.85	9.29	6.04	11.18	7.84
	24-2-1	<b>6.52</b>	<b>0.96</b>	<b>7.79</b>	<b>4.52</b>	<b>4.94</b>	<b>3.93</b>	<b>6.27</b>	<b>3.63</b>	<b>6.60</b>	<b>5.10</b>
	24-3-1	7.95	1.73	9.00	5.84	6.13	5.32	9.06	5.15	9.68	7.30

relationship between the central pixel and the neighborhood pixels on Middlebury standard images (Tsukuba, Venus, Teddy, and Cones). Based on this, it is possible to select the sample points with a high correlation relationship within the matching window. Figure 6 shows the average correlation calculated from each sample point along the seven different patterns (with 16 sample points and edge length of 2). The experiment results indicate that four patterns (16-2-1, 16-2-2, 16-2-6, and 16-2-7) show a high correlation, even if the Gaussian noise ( $\sigma = 5.12$ ) is added. By employing appropriate scan patterns from the correlation distribution of an input image, we can improve the overall performance of stereo matching.

## 4 Experimental Results

### 4.1 Benchmark Results Analysis

The computer used in the experiment is Intel(R) Core(TM) i7-3770 CPU 3.40 GHz, Nvidia Geforce GTX 760. The Middlebury benchmark datasets<sup>12</sup> for the performance test of stereo matching are employed. In the stereo matching framework, we compute an initial matching cost and apply cross-based aggregation<sup>13,14</sup> of the initial costs in the support region. Then, a final disparity map is obtained using a median filter (Figs. 7 and 8). This experiment framework focuses on comparing the performance of the previous CT, MCT, and GCT methods with the proposed SCT method.

If advanced cost aggregation and optimization processes are included, we can sufficiently obtain better stereo matching performance.

Table 3 shows the false matching ratio of final disparity to ground truth in the nonocclusion regions by using the proposed SCT pattern (Fig. 5). A percentage of bad matching pixels is computed with the absolute difference between the computed disparity map and the ground truth disparity map. Here, a threshold value that means a disparity error tolerance is set to 1.

In Table 3, matching performance depends on both the scan pattern and the image local properties (intensity distribution) to some degrees. That means, there is no optimal scan pattern to guarantee the best matching performance in any input image. When SCT patterns with 8, 16, and 24 sample points and an edge length of 2 are employed, the most accurate matching performance is obtained. By considering the experimental results (Table 3), we can determine an appropriate scan pattern suitable for input images. Also, when comparing eight sampling points in the proposed pattern, the proposed SCT method showed better performance than the existing CT and MCT on the four benchmark images (Tsukuba, Venus, Teddy, and Cones). By using 16-2-2 patterns with 16 sample points on four benchmark images, the best performance with a 4.86% error percentage is obtained. In other reference images (baby 3, bowling 2, cloth 2, and dolls), the previous CT method achieved better stereo

**Table 4** False matching ratio in Gaussian noise (nonocclusion regions).

Methods		30 dB	25 dB	20 dB	15 dB	10 dB	Avg.
CT (full 5 × 5)		9.55	13.12	30.28	43.09	73.41	33.89
MCT		11.34	15.82	35.89	47.98	74.58	37.12
GCT (12 edges)		8.46	11.62	23.12	32.00	63.46	27.73
GCT (16 edges)		7.92	10.16	20.46	28.39	60.21	25.42
SCT 8 points	8-2-2	8.00	10.47	22.25	32.36	63.56	27.32
	8-3-1	9.35	11.30	22.34	30.94	61.20	27.02
	8-4-1	9.62	11.76	22.66	30.68	60.08	26.96
SCT 16 points	16-2-1	7.07	9.51	20.89	29.86	60.81	25.62
	16-2-2	<b>6.83</b>	<b>9.06</b>	<b>19.74</b>	<b>28.63</b>	<b>59.67</b>	<b>24.78</b>
	16-2-3	8.32	10.83	23.59	34.23	65.59	28.51
	16-2-4	8.22	10.99	23.78	33.50	65.58	28.41
	16-2-5	8.43	11.45	24.62	34.97	66.11	29.11
	16-2-6	7.04	9.23	20.52	30.15	62.41	25.87
	16-2-7	7.50	9.85	20.93	30.68	62.65	26.32
	16-3-1	<b>7.39</b>	<b>9.13</b>	<b>19.40</b>	<b>27.58</b>	<b>59.60</b>	<b>24.62</b>
	16-4-1	8.26	10.46	20.87	28.73	59.66	25.59
	SCT 24 points	24-2-1	<b>6.48</b>	<b>8.67</b>	<b>19.34</b>	28.79	61.99
24-3-1		8.30	10.12	20.53	28.79	59.86	25.52

**Table 5** False matching ratio in impulse noise (nonocclusion regions).

Methods		2%	5%	10%	20%	Avg.
CT (full 5 × 5)		8.65	23.06	64.94	88.59	46.31
MCT		9.15	19.33	48.94	83.07	40.12
GCT (12 edges)		7.77	13.99	34.96	75.37	33.02
GCT (16 edges)		7.71	13.56	31.87	<b>72.75</b>	31.47
SCT 8 points	8-2-2	7.32	12.21	33.20	76.32	32.26
	8-3-1	9.22	15.15	34.95	74.83	33.53
	8-4-1	9.76	16.29	35.98	74.29	34.08
SCT 16 Points	16-2-1	6.28	10.23	28.90	73.87	29.82
	16-2-2	<b>5.97</b>	<b>9.64</b>	<b>27.01</b>	<b>72.93</b>	<b>28.88</b>
	16-2-3	7.78	13.49	35.88	77.68	33.70
	16-2-4	7.86	13.39	36.18	77.61	33.76
	16-2-5	7.82	13.81	37.17	77.95	34.18
	16-2-6	6.33	11.14	30.44	75.22	30.78
	16-2-7	7.14	11.78	32.20	75.73	31.71
	16-3-1	7.17	12.37	30.78	72.85	30.79
	16-4-1	8.43	14.91	34.68	73.97	32.99
	SCT 24 points	24-2-1	<b>5.90</b>	10.28	29.35	74.89
24-3-1		8.19	13.48	31.66	73.12	31.61

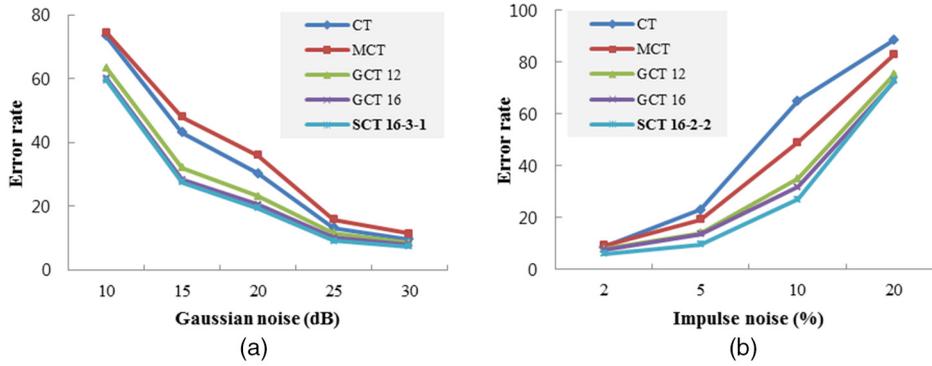


Fig. 9 Comparison of false matching ratios in (a) Gaussian and (b) impulse noise.

matching than MCT and GCT. In the proposed method, we obtained the best performance (5.10%) by using the 24-2-1 scan pattern with 24 sample points and an edge length of 2. In Table 3, best performance values according to the number of sample points (8, 16 and 24) are indicated in bold font.

### 4.2 Stereo Matching Performance in Noise

To examine the noise robustness of the proposed method, Gaussian noise and impulse noise are applied to the Tsukuba, Venus, Teddy, and Cones images. Gaussian noise with a signal-to-noise ratio (SNR) of 10, 15, 20, 25, and 30 dB and an impulse noise with the pixel-to-noise ratio of 2, 5, 10, and 20% are applied, respectively.

Table 4 shows the average false matching ratio results according to the amount of Gaussian noise (dB). If Gaussian noise exists, the existing CT and MCT methods obtain unreliable disparity maps overall, regardless of the amount of noise. Since Gaussian noise affects every pixel of the image evenly, the proposed method obtains reliable disparity results to some degree. When 16 sampling points are considered, the 16-3-1 scan pattern showed the best performance.

When Gaussian noise with a higher SNR (30, 25, and 20 dB) is added, the 24-2-1 scan pattern obtained the best

performance. In addition, if Gaussian noise with a lower SNR (15 and 10 dB) is added, the 16-3-1 scan pattern achieves the best performance. In other words, for the case of relatively small noise, a more reliable disparity result by a scan pattern with more sampling points is obtained. In the image degraded much by Gaussian noise, we should choose patterns, which are suitable to identify the brightness distribution of the image.<sup>9</sup> In the total average false matching ratio results, the best performance is obtained by using a scan pattern with 16 sampling points (16-3-1 and 16-2-2). In Table 4, best performance values according to the number of sample points (16 and 24) are indicated in bold font.

Table 5 shows the average false matching ratio in Benchmark images (Tsukuba, Venus, Teddy, and Cones) with impulse noise. When an impulse noise of 2%, 5%, 10% and 20% is added, the average false matching ratio of GCT with an edge length of 16 is 31.47%, and that of the proposed SCT with a 16-2-2 pattern is 28.88%. In the case where impulse noise was applied in relatively small amounts (2%, 5%, and 10%), SCT showed much better performance than the other methods. When impulsive noise is significantly increased (20%), the performance of GCT was relatively better than that of the other methods. However,

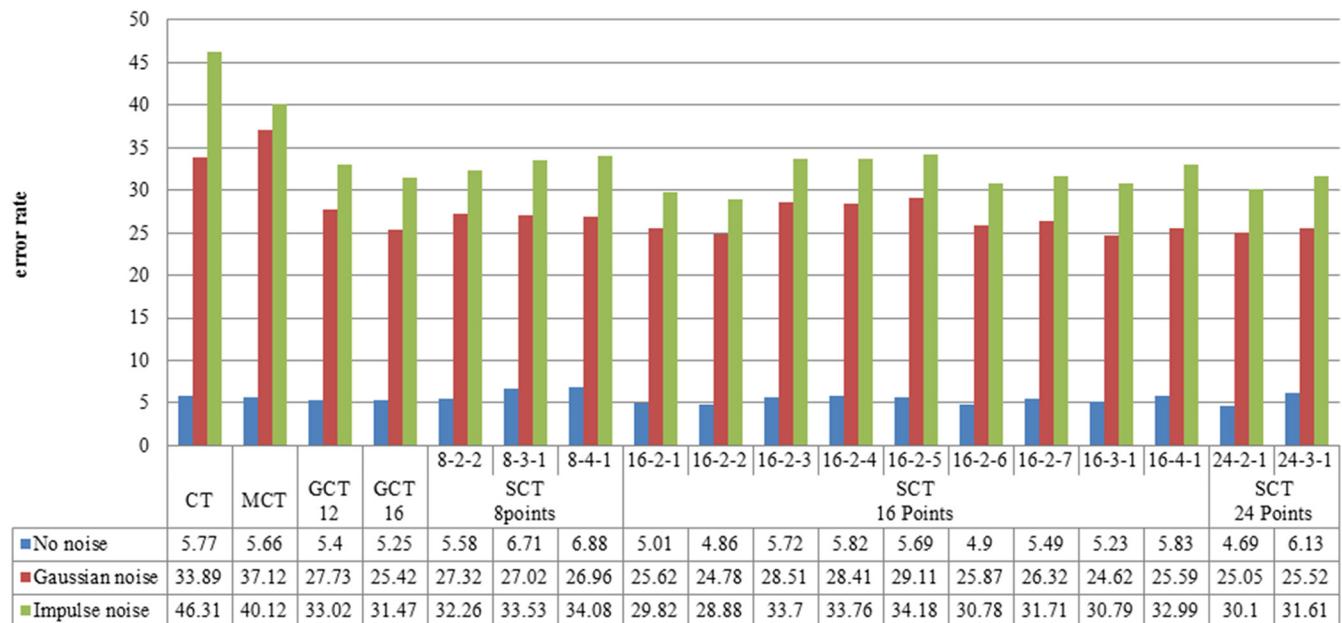


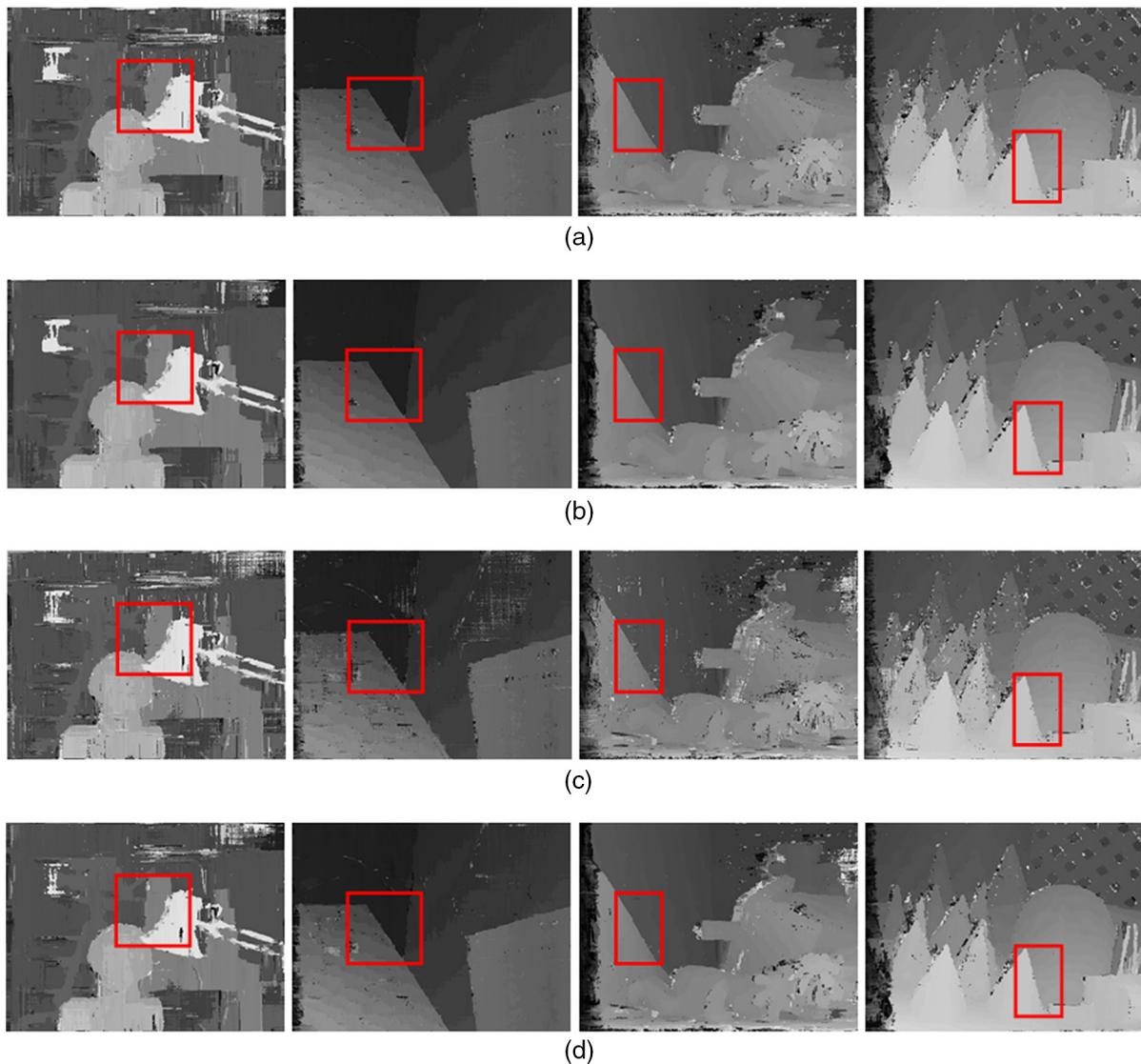
Fig. 10 Average false matching ratios in Gaussian noise, impulse noise, and no noise.

since the input stereo views are considerably degraded, it is difficult to obtain reliable disparity results. The error ratio by GCT is 72.75%. In Table 5, best performance values according to both the number of sample points (16 and 24) and the degree of impulse noise are indicated in bold font. Table 5 shows the best performance (24.78%) in average is obtained by using a 16-2-2 scan pattern.

Figures 9 and 10 show the average false matching results (Tables 4 and 5) by MCT, GCT, and proposed SCT when Gaussian noise and impulse noise are applied. The performance of CT and MCT is significantly affected by Gaussian noise and impulse noise. On the contrary, the proposed SCT shows relatively reliable stereo matching performance. In conclusion, in the case of stereo view with no noise (Benchmark images), the best performance is obtained by using SCT with an edge length of 2. If Gaussian noise is applied, we obtain the best reliable disparity map by using SCT with an edge length of 3. These results are consistent with the correlation distribution of the center pixel and neighborhood in the matching window.<sup>9</sup>

In Tables 4 and 5, we have evaluated matching performance both in Gaussian noise and in impulse noise with several SNR conditions. From the evaluation results (Tables 4 and 5), we determine the scan pattern with 16 sample points in a  $5 \times 5$  matching window to cope with practical situation, where the noise is generated in image acquisition.

These results are consistent with the correlation distribution of the center pixel and neighborhood in the matching window.<sup>9</sup> In case a pixel belongs to a different surface other than the surface on which the kernel pixel is, the encoding result may be affected by local brightness distribution. Figure 11 shows the disparity results in no noise and Gaussian noise. In Fig. 11, the disparity value distribution at surface discontinuities is indicated in the red rectangle. Some foreground regions become thicker (about one pixel) at surface discontinuities than ground truth depths. However, Table 6 shows that SCT obtains more reliable initial disparity results even at surface discontinuities than the existing CT. Here, the first scan pattern with 24 sample points and an



**Fig. 11** Disparity maps by [(a) and (c)] existing CT [(b) and (d)] and by SCT in no noise and Gaussian noise (Table 6).

**Table 6** False matching ratio of initial disparity maps.

		No noise		Gaussian noise (30 dB)		Impulse noise (5%)	
		CT	SCT	CT	SCT	CT	SCT
Tsukuba	Non-occ	40.0	31.9	46.2	39.1	53.8	49.1
	All	41.1	33.2	47.1	40.2	54.5	50.0
	Disc	42.1	36.1	46.9	40.7	55.7	51.9
Venus	Non-occ	41.3	29.8	52.9	43.1	61.8	55.2
	All	42.2	31.0	53.7	44.0	62.4	56.0
	Disc	45.0	37.3	52.5	44.5	62.7	58.8
Teddy	Non-occ	50.3	38.8	61.2	51.1	72.9	69.9
	All	55.4	45.1	65.1	56.1	75.6	72.9
	Disc	56.6	50.3	64.3	57.4	75.5	73.3
Cones	Non-occ	39.9	26.5	48.1	34.2	66.8	62.6
	All	46.7	34.8	53.9	41.6	70.4	66.7
	Disc	49.2	39.8	55.6	45.4	70.7	68.4
Avg.		45.8	36.2	54.0	44.8	65.2	61.2

edge length of 2 (24-2-1) is employed. The performances are evaluated in the nonoccluded region “non-occ,” all (including half-occluded) regions “all,” and regions near depth discontinuities “disc” region, respectively.

In conclusion, though matching results at surface discontinuities may be affected by coplanarity of the central point of the search window and the sample points, overall performance by SCT is much more reliable than the existing CT.

## 5 Conclusion

This paper presents an improved CT method with a star-like scan pattern. The brightness values of the sampling points separated by a certain distance within the stereo matching window are compared in a symmetrical manner. The drawback of the existing CT is that the computation complexity increases: as the matching window size increases, the number of conversed bit-strings also increases. In the proposed SCT, we can choose an appropriate scan pattern in accordance with the processing speed, correlation distribution of the subject image, and types of noise. From the experiment results in Gaussian noise and impulse noise, the proposed SCT achieved relatively more reliable matching performance, even when using smaller sample points than those of the existing methods. The proposed method is also useful in other areas, such as feature matching and tracking.

## Acknowledgments

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (Grant No. 2013R1A1A2008953).

## References

1. D. Scharstein and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” *Int. J. Comput. Vision* **47**(1), 7–42 (2002).
2. A. Klaus, M. Sormann, and K. Karner, “Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure,” in *Proc. of Int. Conf. on Pattern Recognition*, pp. 15–18 (2006).
3. Q. Yang et al., “Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling,” *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(3), 492–504 (2008).
4. J. Worby and W. J. MacLean, “Establishing visual correspondence from multi-resolution graph cuts for stereo-motion,” in *Proc. of Computer Vision and Pattern Recognition*, pp. 313–320 (2007).
5. M. Gong and Y. Yang “Real-time stereo matching using orthogonal reliability-based dynamic programming,” *IEEE Trans. Image Process.* **16**(3), 879–884 (2007).
6. H. Hirschmüller and D. Scharstein, “Evaluation of stereo matching costs on images with radiometric differences,” *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(9), 1582–1599 (2009).
7. R. Zabih and J. Woodfill, “Non-parametric local transforms for computing visual correspondence,” in *Proc. of European Conf. on Computer Vision*, pp. 151–158 (1994).
8. N. Chang et al., “Algorithm and architecture of disparity estimation with mini-census adaptive support weight,” *IEEE Trans. Circuits Syst. Video Technol.* **20**(6), 792–805 (2010).
9. W. Fire and J. Archibald, “Improved census transforms for resource-optimized stereo vision,” *IEEE Trans. Circuits Syst. Video Technol.* **23**(1), 60–73 (2013).
10. X. Hu and P. Mordohai, “Evaluation of stereo confidence indoors and outdoors,” in *Proc. of Computer Vision and Pattern Recognition*, pp. 1466–1473 (2010).
11. D. Pfeiffer, S. Gehrig, and N. Schneider, “Exploiting the power of stereo confidences,” in *Proc. of Computer Vision and Pattern Recognition*, pp. 297–304 (2013).
12. D. Scharstein, R. Szeliski, and H. Hirschmüller, “Middlebury stereo vision,” <http://vision.middlebury.edu/stereo/> (16 December 2015).
13. K. Zhang, J. Lu, and G. Lafruit, “Cross-based local stereo matching using orthogonal integral images,” *IEEE Trans. Circuits Syst. Video Technol.* **19**(7), 1073–1079 (2009).
14. X. Mei et al., “On building an accurate stereo matching system on graphics hardware,” in *Proc. of IEEE Int. Conf. on Computer Vision Workshops on GPUs for Computer Vision*, pp. 467–474 (2011).

**Jongchul Lee** received his BS degree in computer engineering from Academic Credit Bank System, Seoul, Republic of Korea, in 2007. He is currently pursuing a MS degree in the Department of Imaging Science and Arts, Graduate School of Advanced Imaging Science, Multimedia and Film (GSAIM) at Chung-Ang University, Seoul. His research interests include stereo vision, computer vision, and augmented reality.

**Daeyoon Jun** received his BS degree in electronic software engineering from Hansei University, Gunpo, Republic of Korea, in 2015. He is currently pursuing a MS degree in the School of Integrative Engineering, Chung-Ang University, Seoul, Republic of Korea. His research interests include stereo vision and computer vision.

**Changkyoung Eem** received his BS, MS, and PhD degrees in electronic engineering from Hanyang University, Republic of Korea, in 1990, 1992, and 1999, respectively. From 1995 to 2000, he worked at DACOM R&D Center. In 2000, he founded a network software company, IFeelNet Co., and worked for Bzweb Technologies as CTO from 2006 to 2009 in the United States. Since 2014, he has been an industry university cooperation professor at Chung-Ang University, Republic of Korea. His research interests include stereo vision, computer vision, and augmented reality.

**Hyunki Hong** received his BS, MS, and PhD degrees in electronic engineering from Chung-Ang University, Republic of Korea, in 1993, 1995, and 1998, respectively. From 1998 to 1999, he worked as a researcher in the Automatic Control Research Center, Seoul National University, Republic of Korea. From 2000 to 2014, he was a professor in GSAIM at Chung-Ang University. Since 2014, he has been a professor in the School of Integrative Engineering, Chung-Ang University. His research interests include stereo vision, computer vision, and augmented reality.