IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 8, 2020, accepted June 17, 2020, date of publication June 22, 2020, date of current version July 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3003919

Quaternary Census Transform Based on the Human Visual System for Stereo Matching

SEOWON JI[1], SEUNG-WOOK KIM[1], (Member, IEEE), DONGPAN LIM[2], SEUNG-WON JUNG[1], (Senior Member, IEEE), AND SUNG-JEA KO[1], (Fellow, IEEE)

[1]School of Electrical Engineering, Korea University, Seoul 136-713, South Korea
[2]Samsung Electronics Company Ltd., Hwaseong 18448, South Korea

Corresponding author: Sung-Jea Ko (sjko@korea.ac.kr)

This work was supported in part by the S. LSI Division, Samsung Electronics Company Ltd., Hwaseong, South Korea.

ABSTRACT The census transform is a non-parametric local transform that is widely used in stereo matching. This transform encodes the structural information of a local patch into a binary code stream representing the relative intensity ordering of the pixels within the patch. Despite its high performance in stereo matching, the census transform often generates identical binary code streams for two different patches because it simply thresholds the pixels within the patch at the center pixel intensity. To overcome this problem, we introduce a quaternary census transform that encodes the local structural information into a quaternary code stream by employing both the relative intensity ordering and the minimum visibility threshold of the human eye known as the just-noticeable difference. Moreover, because the human eye activates different areas of the retina based on brightness, the patch size for the proposed quaternary census transform adaptively varies depending on the luminance of each pixel. Experimental results on well-known Middlebury stereo datasets prove that the proposed transform outperforms the other census transform-based methods in terms of the accuracy of stereo matching.

INDEX TERMS Census transform, depth estimation, disparity map, human visual system, similarity cost calculation, stereo image processing, stereo matching.

I. INTRODUCTION

Stereo matching is one of the most extensively studied topics in computer vision [1]–[3]. In general, stereo matching is composed of four steps: similarity cost calculation, cost aggregation, disparity selection, and disparity refinement [4], [5]. Among these steps, the similarity cost calculation is the most important, because the performance of the other three steps heavily depends on the similarity cost.

To achieve high-performance of stereo matching, various similarity cost calculation methods have been proposed such as sum of absolute difference [6], relative gradient [7], normalized cross-correlation [8], Mahalanobis distance cross-correlation [9], and census transform [10]. Among these methods, the census transform has been popularly used for stereo matching because it recorded the highest performance [11]. The census transform [10] summarizes the local image structure of a square patch as a binary code stream that represents the relative intensity ordering of the pixels in the patch by thresholding the pixels within the patch using the center pixel intensity. However, the census transform solely depends on the relative intensities of pixels rather than the intensity values themselves. This often gives rise to matching ambiguity that generates an identical binary code stream for flat and textured patches.

To alleviate this drawback, assorted variants of the census transform have been proposed. Mei *et al.* combined the census transform with the conventional pixel-wise absolute difference (AD-Census) to take the advantages of both [12]. Shi *et al.* applied the census transform in the relative gradient [7] domain instead of the conventional spatial domain (RG-Census) [13]. Chang *et al.* proposed a modified version of the census transform called a trinary cross color census transform (TCC-Census) [14]. The TCC-Census extends the census transform to a three-level transform by additionally encoding the pixel intensities in a certain margin around the center pixel intensity and utilizes the cross-square shaped patch. Lee *et al.* presented a three-moded

The associate editor coordinating the review of this manuscript and approving it for publication was Yue Zhang.

census transform (3M-Census) that not only extends the census transform to the three-level transform as in the case of TCC-Census, but also combines it with other similarity cost calculation methods such as color and gradient differences [15]. These methods, the TCC-Census [14] and 3M-Census [15], apply the additional margin to alleviate the matching ambiguity caused by the census transform. However, the TCC-Census and 3M-Census result in another type of matching ambiguity in flat regions of the image because they discard the structural information of those regions. Recently, a four-moded census transform (4M-Census) [16], [17] that extends the census transform to a four-level transform by additionally exploiting mean values of patch has been proposed. However, the 4M-Census gets only a minor performance gain of stereo matching.

Taken together, the existing variants of the census transform can be classified into two approaches: applying the census transform with another method or modifying the census transform itself. The focus of the first approach is to find a method that can supplement the limitation of the census transform. However, this approach can be counterproductive. For example, the pixel-wise absolute difference in the AD-Census may undermine the strength of the census transform such as its invariance to monotonic global gray-level shift. The second approach, on the other hand, is more versatile than the first approach because, if necessary, it can be easily extended to the first approach.

In this study, we introduce a quaternary census transform (QCT) that adopts two properties of the human visual system to solve the limitations of the conventional census transform and other census transform-based methods. First, the proposed transform utilizes the just-noticeable difference (JND), which is the minimum visible threshold of the human eye, along with the conventional relative intensity ordering to generate a quaternary code stream that summarizes more detailed structural information. Second, based on the characteristic of the human retina where the area of the activated light receptors varies according to brightness, the proposed transform exploits a variable-sized patch (VSP) whose size changes depending on the luminance of each pixel. Experimental results on the Middlebury stereo datasets [18]–[21] demonstrate that the proposed similarity measure outperforms the conventional measures in terms of the accuracy of stereo matching.

The remainder of this paper is organized as follows. The census transform and its variations are presented in Section II. In Section III, we present the QCT and describe the detail of the proposed transform. Experimental results are presented in Section IV to validate the effectiveness of the proposed transform. In Section V, we conclude the study.

II. RELATED WORKS

A. CENSUS TRANSFORM

The census transform [10] encodes the local structural information of a pixel $\mathbf{u}$ in an input image I into a binary code stream that represents the relative intensity ordering of the pixels within the patch as follows:

$$C(\mathbf{u}) = \otimes_{\mathbf{v}\in W} f\left(I_{\mathbf{u}}, I_{\mathbf{v}}\right), \tag{1}$$

where the symbol $\otimes$ denotes concatenation, W represents the set of pixels in a fixed-size square patch around $\mathbf{u}$, and $I_{\mathbf{v}}$ represents the intensity at pixel $\mathbf{v}$. In (1), the binary encoding function (BEF) f is defined as follows:

$$f\left(I_{\mathbf{u}}, I_{\mathbf{v}}\right) = \begin{cases} 1, & \text{if } I_{\mathbf{u}} < I_{\mathbf{v}}, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Then, the similarity between the pixels $\mathbf{u}_l$ and $\mathbf{u}_r$ in the left and right images, respectively, is computed as follows:

$$S\left(\mathbf{u}_l, \mathbf{u}_r\right) = 1 - \frac{H\left(C\left(\mathbf{u}_l\right), C\left(\mathbf{u}_r\right)\right)}{N(W)}, \tag{3}$$

where $N(\cdot)$ is the number of pixels inside the input patch except for the center pixel, and $H(\cdot,\cdot)$ returns the Hamming distance between the input binary code streams which stands for the number of entries at which the corresponding codes are different.

B. TRINARY CROSS COLOR CENSUS TRANSFORM (TCC-CENSUS)

The TCC-Census [14] utilizes the trinary encoding function (TEF) f_T to encode the pixels within a fixed-sized cross-square patch (CSP) into a binary code stream as follows:

$$C_T(\mathbf{u}) = \otimes_{\mathbf{v}\in W_T} f_T\left(I_{\mathbf{u}}, I_{\mathbf{v}}\right), \tag{4}$$

where W_T refers to the set of pixels in the fixed-sized CSP and the TEF is defined as

$$f_T\left(I_{\mathbf{u}}, I_{\mathbf{v}}\right) = \begin{cases} 10, & \text{if } I_{\mathbf{u}} + \alpha < I_{\mathbf{v}}, \\ 01, & \text{if } I_{\mathbf{v}} < I_{\mathbf{u}} - \alpha, \\ 00, & \text{otherwise.} \end{cases} \tag{5}$$

In (5), α indicates the margin proportional to the intensity of the center pixel, which is computed as

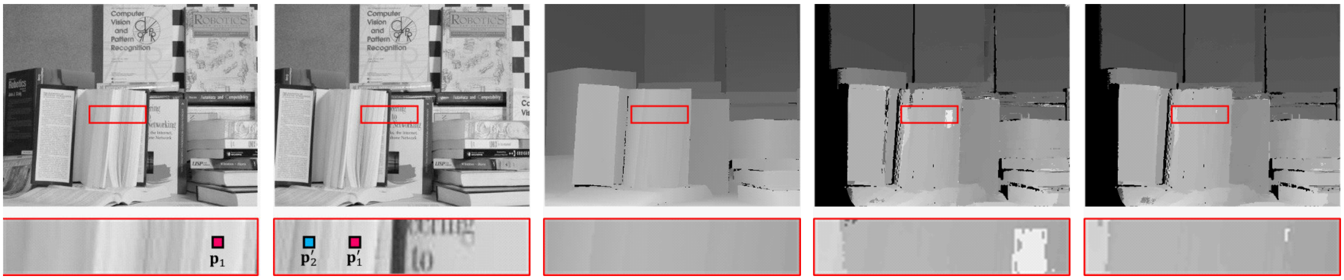
$$\alpha = \left[\frac{I_u}{\beta}\right], \tag{6}$$

where $[\cdot]$ denotes the nearest integer function and β is a user-defined constant. As in the case of the census transform, the similarity between $\mathbf{u}_l$ and $\mathbf{u}_r$ is computed using the Hamming distance in (3).

C. THREE-MODED CENSUS TRANSFORM (3M-CENSUS)

The 3M-Census [15] computes the similarity for stereo matching by using the distance of intensity and gradient as well as the Hamming distance. To calculate the similarity cost, the 3M-Census integrates the three distances as

$$\begin{aligned} S_{\mathrm{3MCT}}\left(\mathbf{u}_l, \mathbf{u}_r\right) = 3 &- \exp\left(-\frac{H\left(C_T\left(\mathbf{u}_l\right), C_T\left(\mathbf{u}_r\right)\right)}{\gamma_H}\right) \\ &- \exp\left(-\frac{\Delta I_{\mathbf{u}_l,\mathbf{u}_r}}{\gamma_I}\right) \\ &- \exp\left(-\frac{\Delta G_{\mathbf{u}_l,\mathbf{u}_r}}{\gamma_G}\right), \end{aligned} \tag{7}$$



(a) (b) (c) (d) (e)

FIGURE 1. **Input stereo image pair and resultant disparity maps of "Books" from the Middlebury stereo datasets. The magnified parts of the images are shown in the bottom row. (a) Pixel $\mathbf{p}_1$ in the left image; (b) $\mathbf{p}'_1$ in the right image that corresponds to $\mathbf{p}_1$, and $\mathbf{p}'_2$ in the right image that dose not corresponds to $\mathbf{p}_1$; (c) ground truth disparity map; (d) resultant disparity map obtained using the census transform; (e) resultant disparity map obtained using the proposed method.**

(a)

226	226	228	222	223	226	180	184	214
225	$\mathbf{p}_1$ (227)	229	222	$\mathbf{p}'_1$ (223)	227	180	$\mathbf{p}'_2$ (185)	214
226	228	228	222	222	227	179	187	215

(b)

0	0	1	0	0	1	0	0	1
0	$\mathbf{p}_1$	1	0	$\mathbf{p}'_1$	1	0	$\mathbf{p}'_2$	1
0	1	1	0	√0	1	0	1	1

(c)

1	1	2	1	1	2	1	1	√3
1	$\mathbf{p}_1$	2	1	$\mathbf{p}'_1$	2	1	$\mathbf{p}'_2$	√3
1	2	2	1	√1	2	1	2	√3

FIGURE 2. **3 × 3 patches with center pixels $\mathbf{p}_1$, $\mathbf{p}'_1$, and $\mathbf{p}'_2$ and their encoded patches. For the encoded patches with center pixels $\mathbf{p}'_1$ and $\mathbf{p}'_2$, neighboring pixels that have a Hamming distance value of "1" as compared to that of $\mathbf{p}_1$ are highlighted with a check symbol and red number. (a) 3 × 3 patches with a center pixel of $\mathbf{p}_1$, $\mathbf{p}'_1$, and $\mathbf{p}'_2$; (b) encoded patches of (a) using the BEF; (c) encoded patches of (a) using the proposed QEF.**

where $\Delta I_{\mathbf{u}_l,\mathbf{u}_r}$ and $\Delta G_{\mathbf{u}_l,\mathbf{u}_r}$ represent the intensity and gradient distance between $\mathbf{u}_l$ and $\mathbf{u}_r$, respectively, and γ_H, γ_I, and γ_G are empirical parameters.

D. LOCAL BINARY PATTERNS AND ITS VARIANTS

The local binary pattern (LBP) [22] is widely employed in the field of texture classification, whereas the census transform is successfully adopted for stereo matching. The LBP encodes binary code stream, which represents the local structural pattern same as the census transform. In stereo matching, binary code streams are used with the Hamming distance to calculate the similarity between two pixels in the horizontal scanline. However, in texture classification, all binary code streams of an image are gathered to generate the histogram, which is used as the image descriptor. In similar way to other variants of the census transform, the LBP has been modified to achieve the better performance. Similar to the TCC-Census and 3M-Census that use multi-level encoding function, local ternary pattern (LTP) [23] and elongated quinary pattern (EQP) [24] were proposed. There were also attempts to change the shape or size of the patch [25], [26]. Similar to the RG-Census, there were several LBP-based methods that perform the domain transformation before encoding the binary code stream [26]–[28]. However, owing to the difference between stereo matching and texture classification, these methods [26]–[28] adopt the domain transformation to make the image descriptor rotation-invariant, which is not desired for stereo matching.

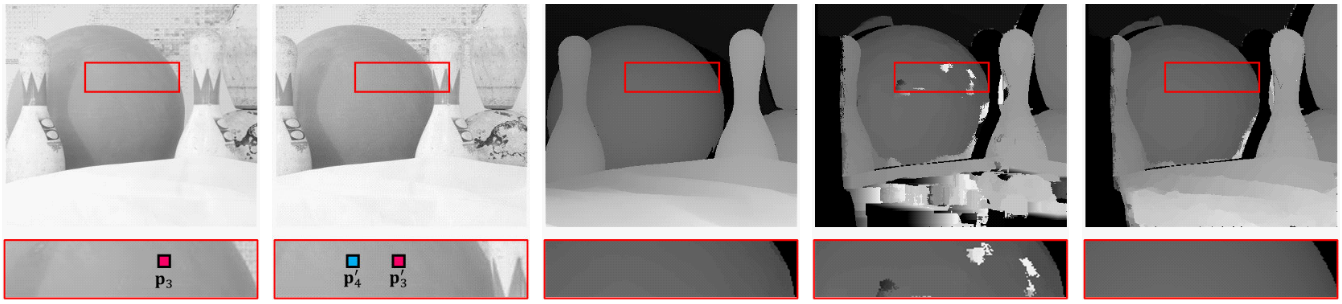
III. PROPOSED METHOD

Precise similarity calculation is necessary for robust stereo matching. To overcome the limitations of the conventional census transform-based methods discussed in Section II, both accurate encoding function and support of adaptive patch sizes are required. In this section, we present a QCT that adopts a JND-based quaternary encoding function (QEF) and VSP that automatically changes its size depending on the luminance of each pixel.

A. JND-BASED QCT

The census transform produces a binary code stream by using the BEF. Because the BEF employs relative intensity ordering and simply thresholds the neighboring pixels using the center pixel intensity, the BEF generates the identical code "1" for a pixel having a higher intensity than the center pixel irrespective of the magnitude of the intensity difference.

In stereo matching, the census transform often fails to distinguish a flat patch consisting of pixels having intensities similar to the center pixel from a textured patch consisting of pixels having intensities noticeably different from the center pixel. Figs. 1 and 2 illustrate an example in which the census transform fails to distinguish the pixel in the flat region from the pixel in the textured region in "Books" from the Middlebury datasets [18]–[21]. For a pixel $\mathbf{p}_1$ in the left image of Fig. 1(a), pixels $\mathbf{p}'_1$ and $\mathbf{p}'_2$ in the right image of Fig. 1(b) are corresponding pixel and not-corresponding pixel to $\mathbf{p}_1$, respectively. Fig. 2(a) shows 3 × 3 patches at the center pixels $\mathbf{p}_1$, $\mathbf{p}'_1$, and $\mathbf{p}'_2$. As shown in Fig. 2(a), $\mathbf{p}'_1$ belongs to the same flat area as $\mathbf{p}_1$, where the center and neighboring pixels have similar intensities. On the other hand, $\mathbf{p}'_2$ is located in a textured area around a vertical edge, where the center pixel and the three right most pixels show noticeable intensity differences. Using the BEF,



(a) (b) (c) (d) (e)

FIGURE 3. Input stereo image pair and resultant disparity maps of "Bowling1" from the Middlebury stereo datasets. The magnified parts of the images are shown in the bottom row. (a) Pixel $\mathbf{p}_3$ in the left image; (b) $\mathbf{p}'_3$ in the right image that corresponds to $\mathbf{p}_3$, and $\mathbf{p}'_4$ in the right image that dose not corresponds to $\mathbf{p}_3$; (c) ground truth disparity map; (d) resultant disparity map obtained using the TCC-Census; (e) resultant disparity map obtained using the proposed method.

162	163	163	161	162	162	179	179	174
161	$\mathbf{p}_3$ (162)	163	161	$\mathbf{p}'_3$ (161)	163	180	$\mathbf{p}'_4$ (178)	176
162	163	162	160	163	162	180	175	181

(a)

00	00	00	00	00	00	00	00	00
00	$\mathbf{p}_3$	00	00	$\mathbf{p}'_3$	00	00	$\mathbf{p}'_4$	00
00	00	00	00	00	00	00	00	00

(b)

1	2	2	1	2	2	√2	2	√1
1	$\mathbf{p}_3$	2	1	$\mathbf{p}'_3$	2	√2	$\mathbf{p}'_4$	√1
1	2	1	1	2	√2	√2	√1	√2

(c)

FIGURE 4. 3×3 patches with center pixels $\mathbf{p}_3$, $\mathbf{p}'_3$, and $\mathbf{p}'_4$ and their encoded patches. For the encoded patches with center pixels $\mathbf{p}'_3$ and $\mathbf{p}'_4$, neighboring pixels that have a Hamming distance value of "1" as compared to that of $\mathbf{p}_3$ are highlighted with a check symbol and red number. (a) 3×3 patches with a center pixel of $\mathbf{p}_3$, $\mathbf{p}'_3$, and $\mathbf{p}'_4$; (b) encoded patches of (a) using the TEF; (c) encoded patches of (a) using the proposed QEF.

these three patches are encoded to binary patches as shown in Fig. 2(b). Thus we obtain the similarities between $\mathbf{p}_1$ and $\mathbf{p}'_1$ as $S\left(\mathbf{p}_1, \mathbf{p}'_1\right) = 0.875$ and between $\mathbf{p}_1$ and $\mathbf{p}'_2$ as $S\left(\mathbf{p}_1, \mathbf{p}'_2\right) = 1$. As a result, $\mathbf{p}_1$ is mismatched to $\mathbf{p}'_2$, and the disparity map obtained using the census transform as shown in Fig. 1(d) has the disparity error compared to the ground truth disparity map in Fig. 1(c).

Although the TCC-Census does not have the aforementioned mismatching problem of the census transform, it faces difficulty in discriminating between different flat patches. This is because the TCC-Census adopts the TEF that encodes pixels having intensities similar to the center pixel into a single code and discards the fine structural information of flat patches. Figs. 3 and 4 demonstrate the failure case of the TCC-Census in "Bowling1" from the Middlebury datasets. Figs. 3(a) and (b) show pixel $\mathbf{p}_3$ in the left image and pixels $\mathbf{p}'_3$ and $\mathbf{p}'_4$ in the right image, respectively. These three pixels belong to the flat patches as shown in Fig. 4(a), and their encoded patches obtained by using the TEF are identical as in Fig. 4(b). Since $S\left(\mathbf{p}_3, \mathbf{p}'_3\right) = S\left(\mathbf{p}_3, \mathbf{p}'_4\right) = 1$, $\mathbf{p}_3$ can be mismatched to $\mathbf{p}'_4$. As a result, the disparity map obtained using the TCC-Census in Fig. 3(d) has the disparity error as compared to the ground truth disparity map in Fig. 3(c).

To encode a more detailed structure of the patch and alleviate the aforementioned problems, we propose an intuitive and simple encoding function that stores distinct local structural information by allocating code values based on the visibility of the intensity difference. Motivated by the conventional methods [14], [15], [23], [24], we define our multi-level encoding function as follow:

$$f'_Q\left(I_\mathbf{u}, I_\mathbf{v}\right) = \begin{cases} 3, & \text{if } I_\mathbf{u} + x < I_\mathbf{v}, \\ 2, & \text{if } I_\mathbf{u} < I_\mathbf{v} \leq I_\mathbf{u} + x, \\ 1, & \text{if } I_\mathbf{u} - x < I_\mathbf{v} \leq I_\mathbf{u}, \\ 0, & \text{otherwise,} \end{cases} \tag{8}$$

where x is the minimum visibility threshold. To correctly define the minimum visibility threshold x, we exploit the property of the human visual system known as JND. The JND is the minimum visibility threshold of the human eye that is inversely proportional to brightness [29], [30]. The ratio of the JND to brightness B is modeled as follows:

$$\frac{\mathrm{JND}}{B} = k, \tag{9}$$

where k represents the Weber fraction. To apply the JND to the image, B can be replaced by the background luminance obtained by applying Gaussian filtering to the image [31]. The JND of pixel $\mathbf{u}$, denoted as $\mathrm{JND}_\mathbf{u}$, can be written as

$$\mathrm{JND}_\mathbf{u} = k \cdot I^B_\mathbf{u}, \tag{10}$$

where $I^B_\mathbf{u}$ is the intensity of the background luminance at pixel $\mathbf{u}$. Using (8) and (10), we define the JND-based QEF f_Q as follows:

$$f_Q\left(I^B_\mathbf{u}, I_\mathbf{v}\right) = \begin{cases} 3, & \text{if } I^B_\mathbf{u} + \mathrm{JND}_\mathbf{u} < I_\mathbf{v}, \\ 2, & \text{if } I^B_\mathbf{u} < I_\mathbf{v} \leq I^B_\mathbf{u} + \mathrm{JND}_\mathbf{u}, \\ 1, & \text{if } I^B_\mathbf{u} - \mathrm{JND}_\mathbf{u} < I_\mathbf{v} \leq I^B_\mathbf{u}, \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

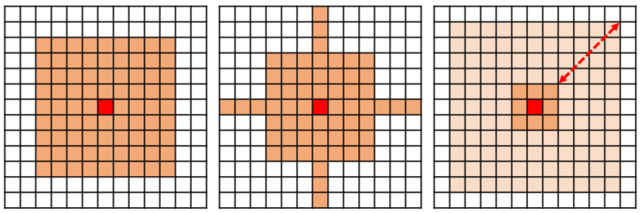


FIGURE 5. Two patches of the conventional methods and proposed VSP. (a) Square patch with a radius of 4 (Census transform); (b) cross-square patch (TCC-Census); (c) VSP with a radius ranging from 1 to 5.

Fig. 2(c) illustrates that the proposed QEF successfully alleviates the problem of the BEF. Unlike the BEF that produces indistinguishable encoded patches, the proposed QEF generates encoded patches which distinguish $\mathbf{p}_2'$ from $\mathbf{p}_1$ by employing both the JND and relative intensity ordering. By using the QEF, the similarity values are obtained as $S\left(\mathbf{p}_1, \mathbf{p}_1'\right) = 0.875$ and $S\left(\mathbf{p}_1, \mathbf{p}_2'\right) = 0.625$, and thus $\mathbf{p}_1$ can be correctly matched to $\mathbf{p}_1{}'$. As a result, the disparity map obtained using the proposed method no longer suffers from the disparity error in that region as shown in Fig. 1(e). The QEF also alleviates the limitation of the TEF producing identical encoded patches for different flat patches. With the QEF, the similiarity values are obtained as $S\left(\mathbf{p}_3, \mathbf{p}_3'\right) = 0.875$ and $S\left(\mathbf{p}_3, \mathbf{p}_4'\right) = 0.125$, as shown in Fig. 4(c). By preserving the fine structural information of the flat region, the QEF generates more similar encoded patches for the corresponding pixel pair, $\mathbf{p}_3$ and $\mathbf{p}_3'$. Thus, the proposed method successfully produces the disparity map, as shown in Fig. 3(e).

B. VARIABLE-SIZED PATCH (VSP)

To compute an accurate similarity cost using census transform-based methods, a patch with a proper size is as crucial as a precise encoding function. However, only a few studies have been conducted to find proper patch sizes for these methods. The census transform [10], AD-Census [12], and RG-Census [13] simply apply a square patch with a fixed size as shown in Fig. 5(a). TCC-Census [14] and 3M-Census [15] also utilize a fixed-sized patch with a different shape, namely a cross-square shaped patch, as shown in Fig. 5(b).

In order to find an appropriately patch size for similarity cost computation, we exploit another property of the human visual system. In the field of texture classification, there has already been successful examples that observe micro and macro patterns simultaneously by mimicking human retina characteristics [25]. According to [32], [33], the human retina contains two types of light receptors: rods and cones. Rods are located throughout the retina and function at low brightness, whereas cones are densely concentrated at the center of the retina and function at high brightness. Therefore, as the brightness level decreases, the periphery of the retina becomes more sensitive than the central region of the retina because of the distribution of rods and cones.

Based on this property of the retina, we carried out an experiment, wherein under varying background luminance conditions, the impact of patch size on the accuracy of stereo matching was investigated. The stereo matching that employs the proposed encoding function using various sized patches with a radius ranging from 1 to 5 was applied to five stereo image pairs from the Middlebury stereo datasets [18]–[21] ("Baby1," "Flowerpots," "Wood1," "Teddy," and "Tsukuba"). Every pixel of these images was classified as one of the four groups according to the intensity of its background luminance. Then, the stereo matching accuracy was computed in terms of the average percentage of bad pixels (APBP) for each group, which is defined as follows:

$$\mathrm{APBP}(\%) = \frac{100}{N(G)} \sum_{\mathbf{p}\in G} I\left(\left|D_{\mathbf{p}} - d_{\mathbf{p}}\right| > 1\right), \tag{12}$$

where G represents one of the four groups previously mentioned, and $N(G)$ returns the number of pixels in each group. In addition, $I(\cdot)$ stands for the indicator function, and $D_{\mathbf{p}}$ and $d_{\mathbf{p}}$ represent the ground truth disparity value and resultant disparity value at pixel $\mathbf{p}$, respectively.

Table 1 indicates that when the background luminance is low, the stereo matching using a larger patch tends to record a higher accuracy. Based on this observation, the proposed transform exploits a VSP whose patch size is inversely proportional to the intensity of the background luminance. A radius of the VSP at pixel $\mathbf{u}$ is defined as follows:

$$r_{\mathbf{u}} = \left\lceil r_{\max} \cdot \exp\left(-\frac{I_{\mathbf{u}}^B}{c}\right)\right\rceil, \tag{13}$$

where $\lceil\cdot\rceil$ denotes the ceiling function, $r_{\max}$ is the maximum radius of the VSP, and c is a constant to adjust the sensitivity of the background luminance. Therefore, the proposed quaternary code stream of $\mathbf{u}$ is obtained as follows:

$$C_Q(\mathbf{u}) = \otimes_{\mathbf{v}\in W_Q} f_Q\left(I_{\mathbf{u}}^B, I_{\mathbf{v}}\right), \tag{14}$$

where W_Q is the set of pixels in the VSP. The quaternary code stream for a pixel in the target image is obtained in the same manner. Finally, the similarity between pixels $\mathbf{u}_l$ and $\mathbf{u}_r$ in the left and right images, respectively, is calculated as:

$$S_Q(\mathbf{u}_l, \mathbf{u}_r) = 1 - \frac{H\left(C_Q(\mathbf{u}_l), C_Q(\mathbf{u}_r)\right)}{N\left(W_Q\right)}, \tag{15}$$

where $N(\cdot)$ is the number of pixels inside the input patch except for the center pixel, and $H(\cdot,\cdot)$ returns the Hamming distance between the input quaternary code streams.

IV. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed transform, 23 stereo image pairs and their corresponding ground-truth disparity maps available in the Middlebury stereo datasets [18]–[21], [34] were used. Each image was converted into a *lab* color space, and l channel was utilized as input data. The proposed QCT was compared with the popular

TABLE 1. APBP (%) on Four Background Luminance Groups with Different Patch Radii.

Data	Background luminance groups	Patch 1 (3x3)	Patch 2 (5x5)	Patch 3 (7x7)	Patch 4 (9x9)	Patch 5 (11x11)
Baby1	Low (0∼25%)	6.588	5.118	4.782	4.682	**4.382**
	Mid-Low (25∼50%)	2.985	2.244	2.038	**1.860**	1.969
	Mid-High (50∼75%)	5.687	5.393	**5.293**	5.449	5.697
	High (75∼100%)	4.336	**4.742**	4.956	4.364	4.778
Flowerpots	Low (0∼25%)	7.654	7.836	7.976	**7.561**	8.007
	Mid-Low (25∼50%)	8.606	7.998	8.031	**7.955**	7.996
	Mid-High (50∼75%)	4.122	3.681	**3.629**	3.889	4.514
	High (75∼100%)	**6.770**	7.207	7.005	7.274	7.778
Wood1	Low (0∼25%)	2.592	2.528	2.391	2.293	**2.198**
	Mid-Low (25∼50%)	16.468	12.714	9.652	**7.245**	7.849
	Mid-High (50∼75%)	11.149	**10.859**	11.497	11.824	12.076
	High (75∼100%)	**0.138**	0.275	0.276	0.276	0.276
Teddy	Low (0∼25%)	8.838	7.925	7.617	7.492	**7.425**
	Mid-Low (25∼50%)	5.584	4.767	**4.627**	4.826	5.134
	Mid-High (50∼75%)	6.373	**5.805**	5.820	5.947	6.053
	High (75∼100%)	13.766	**13.535**	13.979	14.679	16.046
Tsukuba	Low (0∼25%)	3.174	2.851	2.826	**2.676**	2.686
	Mid-Low (25∼50%)	4.027	3.098	**3.096**	3.219	3.277
	Mid-High (50∼75%)	6.022	**5.226**	6.244	6.297	6.306
	High (75∼100%)	**1.080**	1.131	1.234	1.320	1.852

The best and second-best results are boldfaced and underlined, respectively.

TABLE 2. APBP (%) on the 23 Stereo Image Pairs and the Average APBP (%).

Data	non-occlusion						all					
	Census transform	TCC Census	4M Census	$\mathrm{QCT}_{\mathrm{SP}}$ [1]	$\mathrm{QCT}_{\mathrm{CSP}}$ [2]	QCT [3]	Census transform	TCC Census	4M Census	$\mathrm{QCT}_{\mathrm{SP}}$ [1]	$\mathrm{QCT}_{\mathrm{CSP}}$ [2]	QCT [3]
Art	13.48	12.55	24.22	12.65	**11.45**	11.71	30.84	30.52	39.44	30.50	**29.50**	29.73
Baby1	4.33	4.37	8.44	4.04	3.85	**3.82**	11.97	11.84	15.61	11.38	11.12	**11.10**
Baby2	4.95	7.94	8.67	4.48	4.77	**4.31**	12.68	15.05	16.56	12.17	12.35	**11.93**
Books	13.79	13.24	16.76	13.11	12.70	**12.17**	23.55	22.97	25.49	22.96	22.58	**22.09**
Bowling1	9.15	9.60	14.60	8.54	**8.10**	8.17	24.44	24.76	28.96	23.63	**23.21**	23.25
Cloth1	2.76	2.30	6.35	2.18	2.17	**2.13**	12.36	11.88	14.88	11.58	11.62	**11.57**
Cloth2	5.91	5.48	12.17	5.57	**5.42**	5.44	18.84	18.46	24.04	18.57	**18.38**	18.39
Cloth3	3.63	3.18	7.42	3.00	**2.91**	2.91	13.73	13.18	16.41	12.99	12.92	**12.90**
Cloth4	2.62	2.49	6.63	2.45	**2.32**	2.36	17.38	17.26	20.03	17.01	**16.80**	16.84
Dolls	8.24	6.97	15.69	7.00	6.82	**6.82**	19.81	18.55	26.08	18.53	**18.35**	18.43
Flowerpots	12.31	12.31	16.26	11.51	10.89	**10.89**	32.05	31.28	35.42	31.42	30.99	**30.98**
Midd1	42.43	46.95	46.34	42.18	43.23	**41.42**	47.94	51.90	51.13	47.67	48.68	**47.04**
Moebius	11.10	11.65	17.05	10.19	**9.70**	9.84	21.62	22.00	26.93	20.94	**20.37**	20.64
Plastic	36.56	51.66	38.19	35.38	38.14	**35.30**	42.86	56.70	44.76	42.37	44.62	**42.18**
Reindeer	12.39	14.23	21.77	10.49	10.54	**10.20**	26.73	28.09	34.51	25.37	25.28	**25.17**
Rocks1	6.49	6.11	9.99	5.59	5.55	**5.46**	17.67	17.28	19.94	16.56	16.53	**16.46**
Rocks2	4.07	3.51	7.28	3.52	**3.45**	3.47	17.33	16.76	19.42	16.76	**16.70**	16.72
Wood1	8.60	9.77	15.04	8.49	7.91	**7.71**	20.43	21.27	26.30	21.02	20.36	**20.26**
Wood2	5.06	8.07	7.95	4.85	4.49	**4.47**	17.02	19.62	19.41	17.21	**16.74**	16.78
Cones	5.33	4.35	9.46	4.52	4.29	**4.25**	17.75	16.86	20.90	17.03	**16.78**	16.80
Teddy	9.03	9.85	15.23	8.58	8.44	**8.31**	19.63	20.43	24.85	19.36	19.20	**19.12**
Tsukuba	4.35	5.96	4.38	3.77	4.61	**3.41**	23.55	24.85	23.52	23.28	23.86	**22.87**
Venus	2.02	2.17	2.87	1.98	**1.79**	1.84	5.11	5.18	6.24	5.00	**4.77**	4.90
Average	9.94	11.07	14.47	9.31	9.28	**8.97**	21.53	22.46	25.25	21.01	20.94	**20.70**

The best and second-best results are boldfaced and underlined, respectively.

[1] $\mathrm{QCT}_{\mathrm{SP}}$ – QEF with the SP with a radius of 4

[2] $\mathrm{QCT}_{\mathrm{CSP}}$ – QEF with the CSP

[3] QCT – QEF with the VSP

census transform [10], AD-Census [12], RG-Census [13], TCC-Census [14], and 4M-Census [17] proposed recently.

Because the proposed transform is composed of the QEF and the VSP, we conducted experiments on three versions of the proposed transform to validate the effectiveness of each. The first version, $\mathrm{QCT}_{\mathrm{SP}}$, exploits the proposed QEF with a fixed-sized square patch (SP) applied to the census transform, and the second version, $\mathrm{QCT}_{\mathrm{CSP}}$, utilizes the proposed QEF with a fixed-sized cross-square patch (CSP) applied to the TCC-Census. Finally, the last version, QCT, exploits the QEF with the VSP. In case of the AD-Census and RG-Census that use the original census transform with other methods, we adopt parameter settings in their works. For the census transform and 4M-Census, the radius of the SP was set to 4, which records the highest accuracy with the census transform in the experimental data as shown in Fig. 6. In the case of the TCC-Census, the parameter for the TEF, β, was set to 50, and the radius of the CSP was set to 5. For the proposed QCT, r_{max} and c were empirically set to 7 and 70 as shown in Fig. 7. The Weber fraction, k, was set to 0.14 according

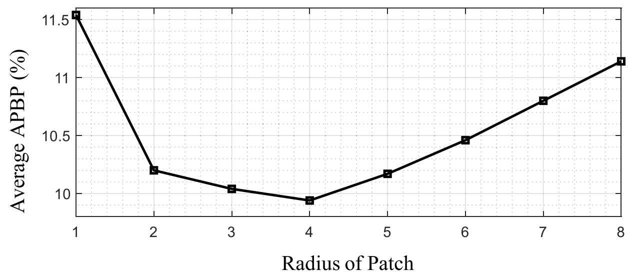


FIGURE 6. Average APBP of the census transform with different patch radii on the 23 stereo image pairs. The radius varies from 1 to 8.

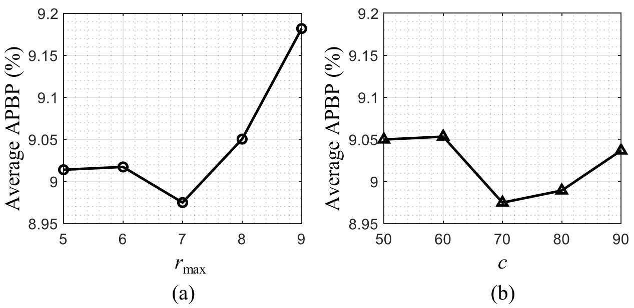


FIGURE 7. Average APBP of the proposed QCT with different parameters. (a) Average APBP of the QCT with different r_{max}; (b) average APBP of the QCT with different c.

to [35]. Finally, the background luminance was obtained by applying Gaussian filtering with a standard deviation of 1.5.

The accuracy of stereo matching was evaluated in terms of the APBP in two types of regions: the non-occlusion region and all regions. Note that the similarity aggregation [6] was performed, and the disparity refinement was excluded from all the experiments for fair comparison. Tables 2 and 5 list the APBP of the stereo matching methods on various image pairs and the average APBP of all image pairs, where the best and second-best results are boldfaced and underlined, respectively. In Table 2, results of the aforementioned three versions of QCT, the census transform and conventional methods that modify the encoding function or patch shape of the census transform are listed. Note that all regions include the occlusion area where the disparity value could not be estimated. Thus, the APBP in all regions is much higher than that in the non-occlusion region.

A. BEF VS. QEF

To validate the effectiveness of the QEF as compared to the conventional BEF, the QCT_{SP} using the same sized SP was compared to the conventional census transform that utilizes the BEF. Because the QEF stores local structural information in the patch more precisely than the BEF does, the QCT_{SP} showed a 6.36% improvement in average APBP as compared to the census transform. In particular, the QCT_{SP} recorded a considerable improvement in accuracy as compared to the census transform in "Cloth1," "Cloth3," "Dolls," "Reindeer," and "Cones" as listed in the left two columns of Table 3. Fig. 8 demonstrates the experimental

TABLE 3. Accuracy Improvement Ratio of the Proposed QEF (Top Five Data).

QEF compared to the BEF		QEF compared to the TEF	
Data	Improvement ratio (%)	Data	Improvement ratio (%)
Cloth1	20.91	Baby2	39.93
Cloth3	17.51	Plastic	26.18
Dolls	15.02	Reindeer	25.95
Reindeer	15.37	Wood2	44.42
Cones	15.18	Tsukuba	22.67

TABLE 4. Accuracy Improvement Ratio of the Proposed VSP (Top Five Data).

VSP compared to the SP		VSP compared to the CSP	
Data	Improvement ratio (%)	Data	Improvement ratio (%)
Art	7.47	Baby2	9.59
Books	7.17	Books	4.20
Wood1	9.13	Midd1	4.19
Wood2	7.94	Plastic	7.45
Tsukuba	9.55	Tsukuba	25.93

data and resultant disparity maps of "Cones," where the BEF failed to distinguish between flat and textured areas. As shown in Figs. 8(d) and (e), the resultant disparity map of the census transform suffered from a disparity error in the enlarged region where the flat and textured areas appeared repeatedly. QCT_{SP}, on the other hand, successfully resolved the problem of the BEF and produced a resultant disparity map with reduced disparity errors in that region as shown in Figs. 8(f) and (g).

B. TEF VS. QEF

The QCT_{CSP} that applies the QEF with the CSP was compared to the TCC-Census to verify that the QEF has an advantage over the conventional TEF. Unlike the TEF that generates an identical encoded patch for different flat patches, the proposed QEF produces distinct encoded patches. Therefore, the QCT_{CSP} showed an average performance improvement of 16.16% as compared to the TCC-Census. As listed in the right two columns of Table 3, in the "Baby2," "Plastic," "Reindeer," "Wood2," and "Tsukuba" which contain broad flat areas, the QCT_{CSP} showed an appreciable performance improvement as compared to the TCC-Census. Fig. 9 illustrates the experimental data and resultant disparity maps of "Wood2." At the bottom of each image, the area with low texture in the input left image is displayed. As shown in Figs. 9(d) and (e), the resultant disparity map of the TCC-Census suffered from an enormous disparity error in flat areas because the TEF could not distinguish flat areas as previously mentioned. However, Figs. 9(f) and (g) demonstrate that the QCT_{CSP} significantly reduced the disparity error.

C. SP & CSP VS. VSP

We compared the proposed VSP with the conventional SP and CSP to validate the effectiveness of the VSP. To accomplish this, the QCT was compared to QCT_{SP} and QCT_{CSP}. Compared to QCT_{SP} and QCT_{CSP}, the QCT showed an average performance improvement of 3.58% and 3.33%, respectively. Notably, because the VSP was used instead of SP,

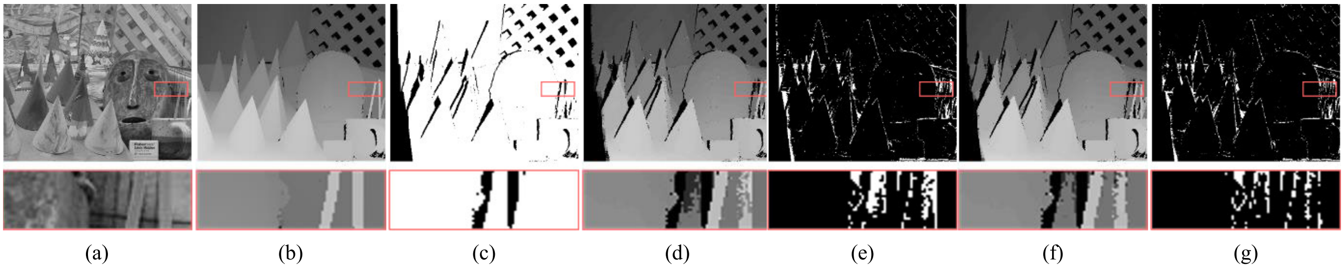


FIGURE 8. Experimental data and results of "Cones" from the Middlebury stereo datasets. The magnified parts of the images are shown in the bottom row. (a) Input left image; (b) ground truth disparity map; (c) occlusion map; (d) disparity map obtained by the census transform; (e) error map of (d) in the non-occlusion area; (f) disparity map obtained by QCT_{SP}; (g) error map of (f) in the non-occlusion area.

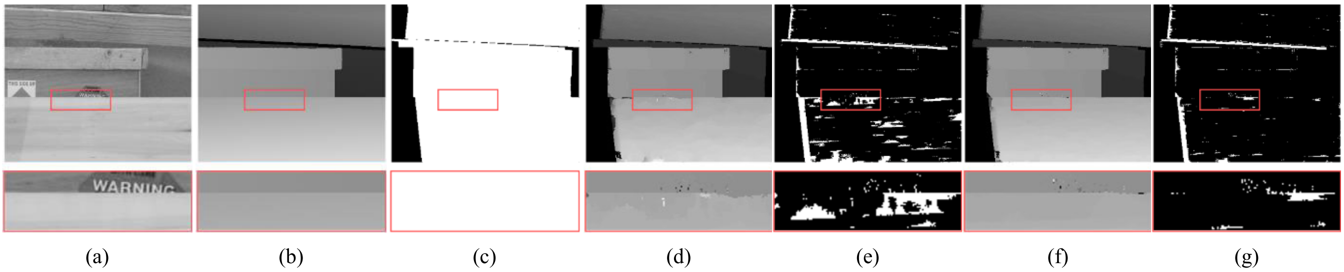


FIGURE 9. Experimental data and results of "Wood2" from the Middlebury stereo datasets. The magnified parts of the images with low texture area are shown in the bottom row. (a) Input left image; (b) ground truth disparity map; (c) occlusion map; (d) disparity map obtained by the TCC-Census; (e) error map of (d) in the non-occlusion area; (f) disparity map obtained by QCT_{CSP}; (g) error map of (f) in the non-occlusion area.

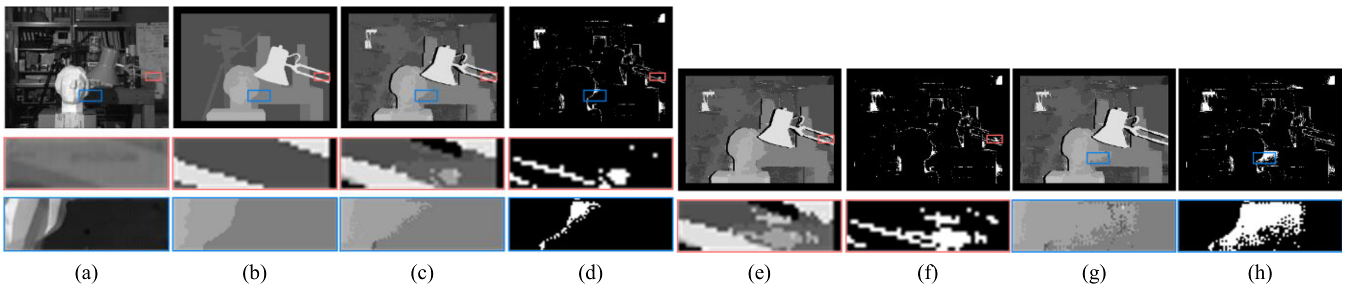


FIGURE 10. Experimental data and results of "Tsukuba" from the Middlebury stereo datasets. The magnified parts of the images with disparity discontinuity area are shown in the bottom row. (a) Input left image. (b) ground truth disparity map; (c) disparity map obtained by the QCT; (d) error map of (c) in the non-occlusion area; (e) disparity map obtained by QCT_{SP}; (f) error map of (e) in non-occlusion area; (g) disparity map obtained by QCT_{CSP}; (h) error map of (g) in the non-occlusion area.

the QCT achieved an accuracy improvement of nearly 10% with "Wood1" and "Tsukuba," as shown in the left two columns of Table 4.

In addition, compared to QCT_{CSP}, the QCT showed a 25.93% performance improvement with "Tsukuba" as shown in the right two columns of Table 4. Fig. 10 demonstrates the experimental data and results of "Tsukuba," which showed significant accuracy improvements in both cases. At the bottom of each image, the area where depth values are discontinuous in the ground truth disparity map is displayed. As shown in Figs. 10(d) and (f), the VSP mitigated the disparity error in the red-boxed region as compared to the SP by properly adjusting the size of the patch for the QEF. In addition, Figs. 10(d) and (h) demonstrate that the VSP substantially alleviated the disparity error derived from the CSP in the blue-boxed region.

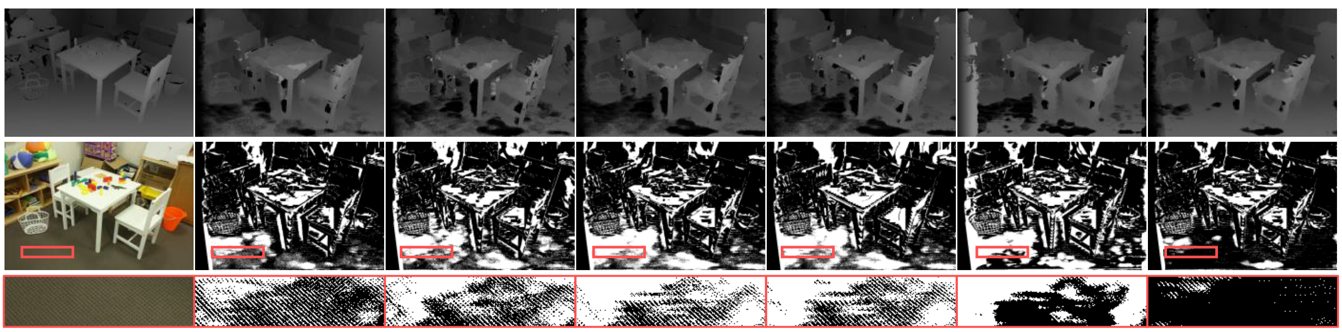
D. VARIANTS OF CENSUS TRANSFORM VS. QCT

In this subsection, the comparison between the proposed QCT and conventional census transform-based methods was made on extended stereo image pairs and the results are listed in Table 5. In comparison with the AD-Census and RG-Census that use the original census transform with other similarity measures, the proposed method improved the average accuracy by 8.03% and 27.37%, respectively. The AD-Census successfully improved the conventional census transform and even showed higher accuracy compared with the proposed QCT in "Baby2", "Rocks2" and "Vintage". However, the AD-Census suffered from numerous disparity error on the image pairs, where the left and right images were captured in different environments such as "ArtL", "MotorcycleE", and "PianoL" due to the absolute difference which is utilized with the census transform. In comparison with the

TABLE 5. APBP (%) on the 37 Stereo Image Pairs and the Average APBP (%).

Data	non-occlusion						all					
	AD Census	RG Census	Census transform	TCC Census	4M Census	QCT	AD Census	RG Census	Census transform	TCC Census	4M Census	QCT
Art	13.11	13.82	13.48	12.55	24.22	**11.71**	29.94	**29.60**	30.84	30.52	39.44	29.73
Baby1	4.73	7.68	4.33	4.37	8.44	**3.82**	12.88	14.46	11.97	11.84	15.61	**11.10**
Baby2	**4.26**	7.12	4.95	7.94	8.67	4.31	**11.77**	14.33	12.68	15.05	16.56	11.93
Books	13.15	15.38	13.79	13.24	16.76	**12.17**	22.59	24.94	23.55	22.97	25.49	**22.09**
Bowling2	9.73	10.63	9.15	9.60	14.60	**8.17**	24.29	25.19	24.44	24.76	28.96	**23.25**
Cloth1	2.14	7.40	2.76	2.30	6.35	**2.13**	**11.49**	16.55	12.36	11.88	14.88	11.58
Cloth2	5.82	9.59	5.91	5.48	12.17	**5.44**	18.48	21.66	18.84	18.46	24.04	**18.39**
Cloth3	3.29	7.38	3.63	3.18	7.42	**2.91**	12.92	16.98	13.73	13.18	16.41	**12.90**
Cloth4	2.45	5.88	2.62	2.49	6.63	**2.36**	17.17	19.91	17.38	17.26	20.03	**16.84**
Dolls	8.54	10.49	8.24	6.97	15.69	**6.82**	19.36	21.33	19.81	18.55	26.08	**18.43**
Flowerpots	11.69	13.44	12.31	12.31	16.26	**10.89**	**28.99**	32.72	32.05	31.28	35.42	30.98
Midd1	42.17	**40.24**	42.43	46.95	46.34	41.42	47.45	**45.99**	47.94	51.90	51.13	47.04
Moebius	11.59	11.92	11.10	11.65	17.05	**9.84**	21.86	21.65	21.62	22.00	26.93	**20.64**
Plastic	36.70	39.32	36.56	51.66	38.19	**35.30**	42.76	44.92	42.86	56.70	44.76	**42.18**
Reindeer	12.35	16.45	12.39	14.23	21.77	**10.20**	26.29	29.26	26.73	28.09	34.51	**25.17**
Rocks1	6.28	10.10	6.49	6.11	9.99	**5.46**	17.09	20.77	17.67	17.28	19.94	**16.46**
Rocks2	**3.43**	7.29	4.07	3.51	7.28	3.47	**16.48**	19.98	17.33	16.76	19.42	16.72
Wood1	8.08	15.25	8.60	9.77	15.04	**7.71**	**20.09**	25.76	20.43	21.27	26.30	20.26
Wood2	4.75	6.22	5.06	8.07	7.95	**4.47**	17.09	17.79	17.02	19.62	19.41	**16.78**
Cones	5.19	7.31	5.33	4.35	9.46	**4.25**	17.46	19.11	17.75	16.86	20.90	**16.80**
Teddy	8.81	10.87	9.03	9.85	15.23	**8.31**	19.39	21.09	19.63	20.43	24.85	**19.12**
Tsukuba	3.84	6.99	4.35	5.96	4.38	**3.41**	21.58	**21.34**	23.55	24.85	23.52	22.87
Venus	2.33	3.42	2.02	2.17	2.87	**1.84**	5.26	6.61	5.11	5.18	6.24	**4.90**
Adirondack	14.73	14.08	11.65	18.52	18.51	**10.90**	22.43	21.00	18.63	26.08	26.52	**18.22**
ArtL[1]	26.42	12.67	11.84	11.87	20.63	**11.69**	46.76	31.46	**30.98**	31.52	41.08	31.24
Jadeplant	21.48	20.56	**19.71**	20.01	32.35	21.09	44.14	42.33	**41.69**	42.13	55.63	43.96
Motorcycle	12.44	14.74	12.09	13.19	18.62	**11.87**	24.05	26.15	23.57	24.67	30.54	**23.45**
MotorcycleE[2]	22.04	13.52	11.33	12.54	18.13	**11.22**	34.04	24.78	22.77	23.90	30.03	**22.74**
Piano	24.44	19.92	17.47	19.81	21.13	**17.44**	36.00	30.79	28.58	30.69	33.44	**28.48**
PianoL[1]	43.28	**29.50**	30.93	30.12	36.32	30.33	54.70	**40.34**	41.97	41.00	48.68	41.31
Pipes	13.22	11.48	11.19	**10.86**	15.44	10.90	31.90	28.96	29.13	**28.55**	34.30	28.99
Playroom	26.77	23.20	21.70	19.93	22.75	**17.46**	44.96	41.14	39.67	37.75	40.82	**35.49**
Playtable	35.78	44.26	40.49	43.18	42.77	**35.73**	48.17	56.17	52.45	55.14	55.38	**47.77**
PlaytableP[3]	24.42	31.85	30.26	31.91	32.05	**18.56**	36.65	43.60	41.95	43.49	44.49	**30.33**
Recycle	14.07	13.70	12.99	17.63	17.68	**12.45**	22.60	22.02	21.38	26.21	26.15	**20.94**
Shelves	35.17	37.98	34.03	34.68	35.64	**33.65**	51.01	53.72	49.78	50.43	51.51	**49.44**
Vintage	**33.69**	36.48	36.76	39.99	38.81	34.32	**43.63**	46.34	46.64	49.93	48.74	43.99
Average	15.47	16.44	14.35	15.65	19.02	**13.08**	27.67	28.13	26.61	27.79	31.03	**25.47**

The best and second-best results are boldfaced and underlined, respectively.
[1] Stereo image pairs that right images were captured under different **L**ighting.
[2] Stereo image pair that right image was captured under different **E**xposure.
[3] Stereo image pair that was **P**erfectly calibrated.



(a) (b) (c) (d) (e) (f) (g)

FIGURE 11. Experimental data and results of "PlaytableP" from the Middlebury stereo datasets. The magnified parts of the images with repetitive patterns are shown in the bottom row. (a) Ground truth disparity map and input left image; (b) disparity map obtained by the AD Census and its error map; (c) disparity map obtained by the RG Census and its error map; (d) disparity map obtained by the census transform and its error map; (e) disparity map obtained by the TCC Census and its error map; (f) disparity map obtained by the 4M Census and its error map; (g) disparity map obtained by the QCT and its error map.

census transform, the proposed QCT significantly improved the performance of 38.64% in "PlaytableP," and exhibited an average performance improvement of 8.85%. In comparison with the TCC-Census, the proposed QCT improved the performance from at least 0.77% in "Cloth2" to over 45.69% in "Baby2," and exhibited an average improvement of 16.42%. Compared to the the 4M-Census, the proposed QCT improved the accuracy from at least 7.57% in "Plastic" to over 66.51% in "Cloth1," and exhibited an average accuracy improvement of 31.23%. As shown in Fig 11, the pro-

TABLE 6. APBP (%) on the 23 Stereo Image Pairs and the Average APBP (%) in Non-Occlusion Region.

Data	LBP & Variants of LBP					QEF
	LBP ($CLBP_S$)	LTP	LQP	$CLBP_M$	$CLBP_{M+S}$	
Art	13.48	26.76	16.02	18.50	14.15	**12.65**
Baby1	4.33	12.79	5.40	5.83	4.88	**4.04**
Baby2	4.95	18.71	8.76	6.89	5.40	**4.48**
Books	13.79	23.16	15.01	14.97	**12.65**	13.11
Bowling2	9.15	45.84	13.64	10.60	9.33	**8.54**
Cloth1	2.76	4.01	4.10	4.81	3.73	**2.18**
Cloth2	5.91	11.70	7.17	8.16	6.83	**5.57**
Cloth3	3.63	6.82	3.73	5.41	4.20	**3.00**
Cloth4	2.62	6.77	2.46	4.68	3.10	**2.45**
Dolls	8.24	16.35	9.27	11.07	8.70	**7.00**
Flowerpots	12.31	38.57	18.41	12.94	12.13	**11.51**
Midd1	42.43	60.44	48.54	52.88	**42.15**	42.18
Moebius	11.10	21.67	12.63	12.34	**10.12**	10.19
Plastic	36.56	71.88	56.59	50.89	38.58	**35.38**
Reindeer	12.39	38.43	24.11	13.51	11.44	**10.49**
Rocks1	6.49	8.75	6.45	7.61	6.76	**5.59**
Rocks2	4.07	7.16	3.87	5.85	4.85	**3.52**
Wood1	8.60	26.66	17.44	9.46	**7.72**	8.49
Wood2	5.06	37.48	10.54	5.44	**4.80**	4.85
Cones	5.33	9.03	4.65	7.89	5.86	**4.52**
Teddy	9.03	19.19	11.35	11.91	9.74	**8.58**
Tsukuba	4.35	6.30	3.92	4.40	4.29	**3.77**
Venus	2.02	15.63	4.21	2.51	**1.88**	1.98
Average	9.94	23.22	13.40	12.55	10.14	**9.31**

The best and second-best results are boldfaced and underlined, respectively.

posed method successfully alleviates disparity error on the floor where the other methods suffer from large disparity error due to repetitive patterns. Consequently, the proposed QCT exhibited promising accuracy consistently over the entire experimental data.

E. VARIANTS OF LBP VS. QEF

As mentioned in Section II, the LBP has been actively researched for texture classification. Among the variants of the LBP, some can be applied to stereo matching. In this subsection, we compared the proposed QEF with the LTP, local quinary patter (LQP) [24] and completed-LBP (CLBP) [36]. The CLBP consists of three methods: CLBP_Sign ($CLBP_S$), CLBP_Magnitude ($CLBP_M$), and CLBP_Center gray level ($CLBP_C$). The $CLBP_C$ converts an input image to a binary image by global thresholding, resulting in excessive information loss for stereo matching. On the other hand, since the $CLBP_S$ and $CLBP_M$ use the same method of encoding the code stream as the census transform, these methods can be directly employed for stereo matching. Also, the $CLBP_S$ is identical to the LBP/census transform, and thus we conducted experiments with two types of the CLBP: $CLBP_M$ and $CLBP_{M+S}$ whose output is obtained by concatenating the outputs of the $CLBP_M$ and $CLBP_S$. As in the previous experiments, we used a square patch of radius 4, the similarity aggregation method [6], and the encoding functions as follows: LBP, LTP, LQP, $CLBP_M$, $CLBP_{M+S}$, and QEF.

Table 6 lists the APBP of the stereo matching methods using the LBP variants and QEF on the stereo image pairs. The proposed QEF recorded significantly better performance than LTP in the stereo matching. Compared with the LBP ($CLBP_S$), LQP, $CLBP_M$, and $CLBP_{M+S}$, the QEF showed higher accuracy of 6.36%, 25.81%, 30.55%, and 8.23%, respectively.

F. NOISE ROBUSTNESS OF QEF (vs. BEF & TEF)

To evaluate the noise robustness of the proposed QEF compared to the conventional BEF and TEF, we conduct additional experiments in noise environment. Following the experiment in [27], we considered different level of additive

TABLE 7. APBP (%) on the 23 Stereo Image Pairs with AWGN and the Average APBP (%) in Non-Occlusion Region.

	AWGN with $\sigma = 5$			AWGN with $\sigma = 10$			AWGN with $\sigma = 15$			AWGN with $\sigma = 25$		
Data	BEF	TEF	QEF	BEF	TEF	QEF	BEF	TEF	QEF	BEF	TEF	QEF
Art	24.19	23.56	**18.63**	45.99	44.89	**33.54**	63.34	62.61	**50.75**	81.50	81.22	**74.47**
Baby1	12.58	11.15	**10.03**	32.07	29.78	**22.77**	50.90	49.63	**44.86**	77.06	76.21	**75.05**
Baby2	21.65	17.09	**15.95**	39.18	36.58	**28.83**	55.65	54.00	**45.92**	79.63	78.81	**74.68**
Books	24.13	21.71	**19.22**	37.64	35.10	**28.92**	55.73	53.60	**47.90**	77.40	76.38	**72.83**
Bowling2	36.45	34.21	**29.91**	56.53	54.43	**47.58**	71.52	69.91	**62.32**	87.38	86.79	**80.96**
Cloth1	5.47	5.46	**3.98**	8.39	8.22	**6.23**	15.93	15.46	**14.03**	43.27	**42.46**	46.88
Cloth2	11.25	10.13	**9.32**	22.86	21.11	**18.06**	43.16	40.76	**36.91**	75.63	74.22	**70.94**
Cloth3	7.21	7.08	**5.22**	15.40	14.61	**10.64**	31.68	30.42	**26.13**	62.28	61.05	**59.47**
Cloth4	6.28	6.28	**4.81**	12.81	11.96	**9.35**	24.51	23.26	**19.32**	54.26	52.48	**51.43**
Dolls	19.17	18.52	**13.92**	32.84	31.94	**23.13**	46.96	46.09	**36.28**	70.13	69.35	**61.73**
Flowerpots	43.23	39.50	**31.51**	67.68	66.55	**51.65**	82.20	81.47	**67.66**	91.72	91.63	**82.11**
Midd1	58.28	57.87	**55.76**	69.53	68.70	**63.78**	77.36	76.84	**72.91**	86.14	85.93	**84.55**
Moebius	20.52	19.45	**17.65**	35.73	34.84	**27.80**	57.59	56.63	**48.31**	80.00	79.48	**72.77**
Plastic	70.21	66.49	**63.88**	80.19	77.96	**72.85**	86.74	85.90	**80.19**	93.14	92.85	**89.15**
Reindeer	27.61	27.20	**24.82**	54.92	53.17	**45.14**	69.30	68.46	**60.10**	82.91	82.65	**76.26**
Rocks1	10.99	10.79	**9.67**	22.05	21.46	**17.94**	40.17	39.45	**34.87**	71.60	71.04	**68.88**
Rocks2	7.08	6.95	**5.82**	15.49	15.09	**12.27**	31.88	31.05	**26.41**	67.45	66.65	**62.88**
Wood1	24.67	23.16	**22.06**	47.63	46.57	**40.37**	74.47	73.59	**67.43**	90.98	90.69	**86.93**
Wood2	39.04	37.43	**36.66**	63.63	62.16	**58.22**	80.68	80.16	**77.57**	91.79	91.63	**91.08**
Cones	7.90	7.59	**5.86**	16.70	15.63	**12.59**	32.93	31.54	**27.18**	62.13	61.06	**57.98**
Teddy	17.65	16.95	**15.29**	36.86	35.67	**28.05**	53.86	52.92	**46.75**	74.33	73.69	**70.34**
Tsukuba	11.14	10.25	**6.95**	20.91	20.12	**14.06**	34.61	33.93	**28.47**	51.18	51.07	**50.17**
Venus	15.23	14.46	**11.75**	33.83	32.53	**25.72**	45.93	44.98	**41.33**	64.59	**63.96**	65.47
Average	22.69	21.45	**19.07**	37.78	36.48	**30.41**	53.35	52.29	**46.24**	74.63	73.97	**70.74**

The best and second-best results are boldfaced and underlined, respectively.

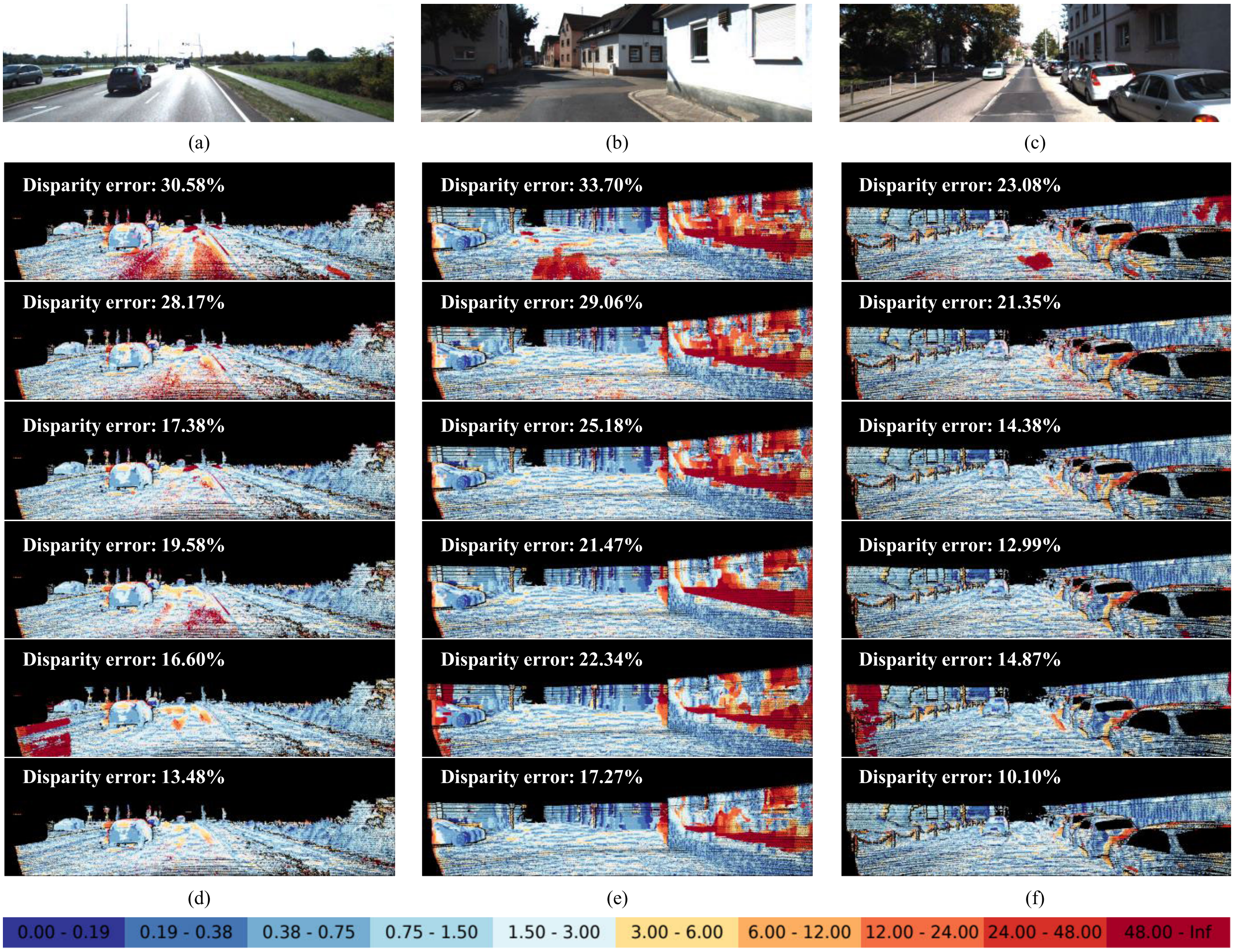


FIGURE 12. Experimental data and disparity error images of "No.79", "No.85", and "No.132" from the KITTI2015 datasets. The disparity error images are visualized using the color scheme depicted in the legend. The disparity error images are displayed from top-to-bottom: result of AD-Census, RG-Census, Census transform, TCC-Census, 4M-Census, and the proposed method. (a) Input "No.79" left image; (b) input "No.85" left image; (c) input "No.132" left image; (d) disparity maps of (a); (e) disparity maps of (b); (f) disparity maps of (c).

white Gaussian noise (AWGN). For this experiment, we used a square patch of radius 4 and the similarity aggregation method [6] with BEF, TEF, and QEF.

Table 7 lists the APBP of the stereo matching methods using the three encoding functions on the stereo image pairs with AWGN. In the case of AWGN with $\sigma = 5$, 10, and 15, the proposed QEF achieved 16.26% and 13.09% higher accuracy on average compared to the BEF and TEF, respectively. In the extreme case of AWGN with $\sigma = 25$, compared to the BEF and TEF, the QEF still showed higher accuracy of 5.21% and 4.37%, respectively.

V. CONCLUSION

This paper presented a quaternary census transform for similarity measure in stereo matching. To overcome the limitations of the conventional transforms, the proposed transform adopts two properties of the human visual system: the minimum visibility threshold of the human eye, the JND, and the varying area of the human retina activated depending on the luminance. Therefore, the proposed transform summarizes more detailed local structural information in a variable-sized patch by using both the JND and relative intensity ordering as compared to the conventional methods. Experimental results demonstrate that the proposed transform significantly improves the accuracy of stereo matching as compared to the conventional methods.

APPENDIX A
PERFORMANCE EVALUATION IN THE OUTDOOR ENVIRONMENTS

Since Middlebury stereo datasets were captured in the indoor environments, we conducted additional experiments on KITTI2015 datasets [37] to evaluate the performance of stereo matching methods in the outdoor environments. Unlike Middlebury stereo datasets that use an error threshold of 1 pixel, KITTI2015 stereo datasets count errors if the disparity exceeds 3 pixels and 5% of its true value. We use

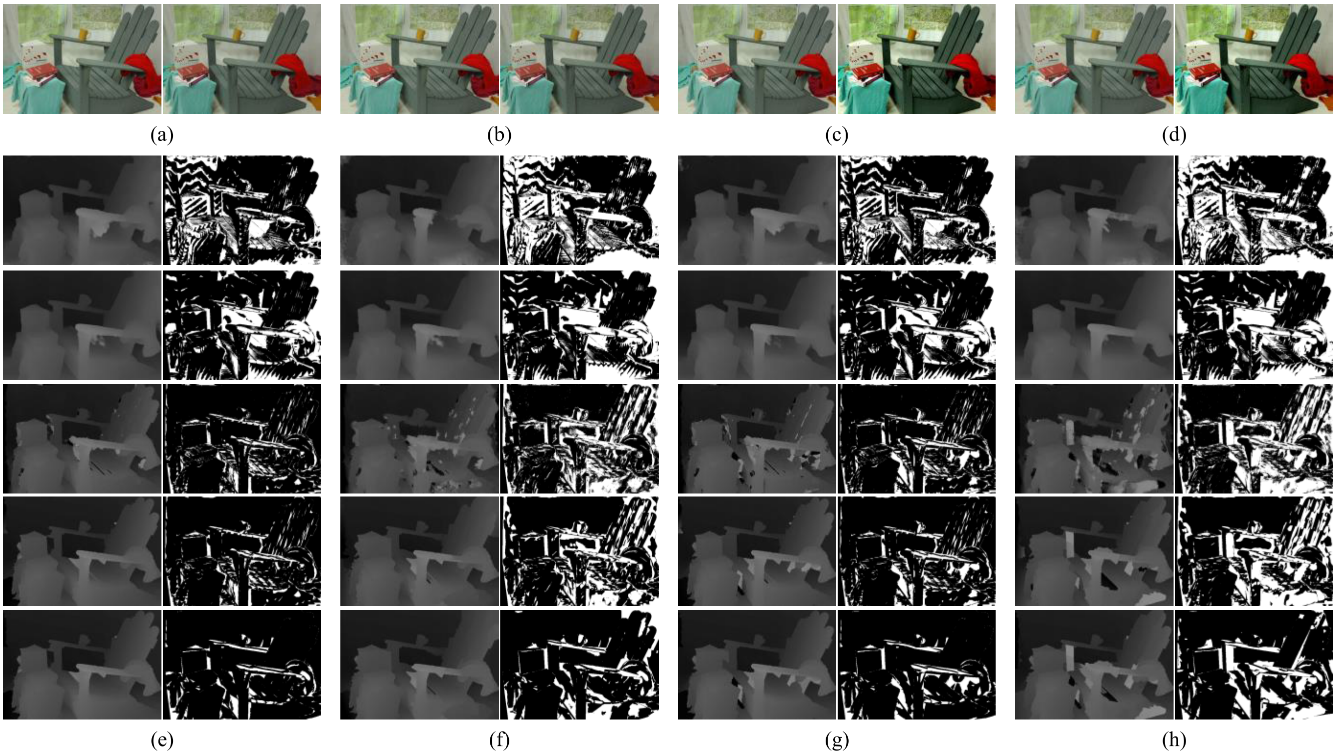


FIGURE 13. Input stereo pairs, disparity maps and error maps of "Adirondack" from the Middlebury stereo datasets with various conditions. The disparity maps and error images are displayed from top-to-bottom: result of HD3, GWC, QCT, QCT*, and QCT$^+$. (a) Original input image pair; (b) input image pair with AWGN; (c) input image pair that contrast of right image is adjusted; (d) input image pair with AWGN and contrast adjustment; (e) disparity and error maps of (a); (f) disparity and error maps of (b); (g) disparity and error maps of (c); (h) disparity and error maps of (d).

the development kit provided with KITTI2015 datasets to calculate the disparity error and visualize the disparity error maps. The experiments were conducted with the same setup as the previous experiments for Middlebury datasets, and the proposed QCT was compared with the census transformation, AD-Census, RG-Census, TCC-Census, and 4M-Census.

Table 8 lists the average disparity error of the stereo matching methods on 200 stereo image pairs from KITTI2015 datasets. The proposed method still recorded the highest accuracy the same as Middlebury datasets, and other methods except for the AD-Census recorded similar rank compared to the previous experiments. The AD-Census which achieved the highest accuracy in the indoor environments except for the proposed method recorded the second-lowest accuracy in the outdoor environments. This is because the absolute difference which is used with the census transform degrades the performance of the AD-Census in the outdoor environments with large illumination variations. Fig. 12 shows the disparity error and visual quality comparison of disparity images on KITTI2015 datasets obtained by using the provided development kit. Figs. 12(a), (b), and (c) show the input left images and Figs. 12(d), (e), and (f) show resultant disparity error images of the input images. The disparity error images were visualized using the color scheme depicted in the legend (bottom row in Fig. 12). As shown in the last row in Figs. 12(d), (e), and (f), the proposed QCT produced disparity maps with lower error compared to other methods.

TABLE 8. Average Disparity Error (%) on KITTI2015 Datasets.

	AD Census	RG Census	Census Transform	TCC Census	4M Census	QCT
Avg. error	21.29	23.67	15.81	<u>14.47</u>	16.34	**13.43**

The best and second-best results are boldfaced and underlined, respectively.

APPENDIX B
COMPARISON OF THE GENERALIZATION ABILITY

Owing to a rapid development of neural networks, neural network-based stereo matching methods have shown promising results. However, these learning-based methods face a generalization problem when there is a domain gap between the training and test datasets. Although simple, the census transform is illumination-invariant and morphologically invariant, and thus the census transform-based stereo matching techniques have showed robustness to different environments [11], [38]. To compare the generalization ability of the proposed and neural network-based methods, we conducted additional experiments. In particular, we simulated two challenging conditions: noisy image pairs and image pairs with different global contrast. The accuracy of the neural network-based methods trained on the KITTI datasets and the proposed QCT were evaluated on the Middlebury 2014 dataset.

TABLE 9. Average APBP (%) and RMSE on Middlebury datasets.

	APBP							
	non-occlusion				all			
Conditions	HD3	GWC	QCT	QCT*	HD3	GWC	QCT	QCT*
N/A	32.52	29.55	19.05	**18.47**	45.82	42.16	32.97	**32.39**
N_1[1]	38.68	34.69	31.21	**29.26**	52.27	47.67	45.48	**43.49**
N_2[2]	44.99	40.96	41.88	**36.96**	58.79	54.52	56.40	**51.40**
C_1[3]	38.61	39.61	28.67	**27.53**	52.19	53.11	42.77	**41.61**
C_2[4]	45.66	51.73	38.67	**37.09**	59.64	66.01	52.85	**51.28**
N_1 & C_1	43.05	43.20	37.40	**35.40**	56.78	56.95	51.78	**49.76**
N_1 & C_2	49.30	53.86	45.16	**43.02**	63.33	68.26	59.65	**57.50**
N_2 & C_1	50.39	48.31	44.12	**39.77**	64.43	62.57	58.68	**54.25**
N_2 & C_2	55.78	57.49	50.11	**45.86**	70.07	72.15	64.74	**60.45**
	RMSE							
	non-occlusion				all			
Conditions	HD3	GWC	QCT	QCT+	HD3	GWC	QCT	QCT+
N/A	9.64	**7.16**	8.56	7.75	13.68	**12.72**	15.44	13.12
N_1	9.06	10.55	10.85	**8.81**	**12.96**	15.64	16.55	14.04
N_2	11.59	**10.80**	14.09	12.47	**15.20**	15.74	19.33	18.42
C_1	12.92	**12.45**	15.99	16.94	**16.44**	17.08	20.29	20.44
C_2	15.60	**13.03**	18.60	17.40	19.43	**18.15**	22.52	20.90
N_1 & C_1	**10.73**	13.07	15.16	10.83	**14.37**	17.81	20.10	15.68
N_1 & C_2	13.39	**13.21**	17.36	16.24	**17.17**	18.37	21.94	24.15
N_2 & C_1	15.68	12.90	15.96	**11.08**	18.82	17.66	20.79	**15.73**
N_2 & C_2	17.36	**13.52**	17.65	14.12	20.43	18.82	22.24	**18.28**

The best and second-best results are boldfaced and underlined, respectively.
[1] N_1 – AWGN with $\sigma = 5$
[2] N_2 – AWGN with $\sigma = 10$
[3] C_1 – Contrast adjustment by saturating the bottom and top 20% of pixels.
[4] C_2 – Contrast adjustment by saturating the bottom and top 30% of pixels.
QCT^* – QCT with median filtering (11x11)
QCT^+ – QCT with SDR

The accuracy of the QCT with additional disparity refinements was also measured because the neural network-based methods implicitly perform the disparity refinement as well as cost computation and disparity estimation in an end-to-end manner. The accuracy of stereo matching was evaluated in terms of the average percent of bad pixel (APBP) and root mean square error (RMSE).

Table 9 lists the APBP and RMSE of the hierarchical discrete distribution decomposition (HD3) [39], group-wise correlation stereo network (GWC) [40], QCT, and QCT with two disparity refinements: median filtering and segment based disparity refinement (SDR) [41]. For the generation of noisy image pairs, we added AWGN to left and right images. For the simulation of image pairs with different contrast, we changed the global contrast of the right image by linearly mapping the range between the top and bottom 20% (or 30%) of the pixels to the total dynamic range, i.e., [0, 255]. In terms of the APBP, the proposed QCT showed lower error rate compared to the HD3 and GWC. When combined with simple median filtering, the proposed method achieved the lowest error rate. In terms of the RMSE, the HD3 and GWC performed generally better than the proposed method, but the QCT with SDR showed the best and second-best results on several conditions. In other words, the QCT generated disparity maps with a higher number of accurately estimated pixels, whereas the HD3 and GWC generated disparity maps closer to the ground-truth on average. This can also be found in Fig. 13. As shown in Figs. 13 (e), (f), (g), and (h), the resultant disparity maps of the QCT contain a fewer number of erroneous pixels, but the HD3 and GWC produce visually plausible results.

REFERENCES

[1] I.-L. Jung, T.-Y. Chung, J.-Y. Sim, and C.-S. Kim, "Consistent stereo matching under varying radiometric conditions," *IEEE Trans. Multimedia*, vol. 15, no. 1, pp. 56–69, Jan. 2013.

[2] G. A. Kordelas, D. S. Alexiadis, P. Daras, and E. Izquierdo, "Content-based guided image filtering, weighted semi-global optimization, and efficient disparity refinement for fast and accurate disparity estimation," *IEEE Trans. Multimedia*, vol. 18, no. 2, pp. 155–170, Feb. 2016.

[3] B. Chen, C. Jung, and Z. Zhang, "Variational fusion of time-of-flight and stereo data for depth estimation using edge-selective joint filtering," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 2882–2890, Nov. 2018.

[4] Q. Yang, "A non-local cost aggregation method for stereo matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1402–1409.

[5] G. A. Kordelas, D. S. Alexiadis, P. Daras, and E. Izquierdo, "Enhanced disparity estimation in stereo images," *Image Vis. Comput.*, vol. 35, pp. 31–49, Mar. 2015.

[6] D. Min, J. Lu, and M. N. Do, "Joint histogram-based cost aggregation for stereo matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 10, pp. 2539–2545, Oct. 2013.

[7] X. Zhou and P. Boulanger, "Radiometric invariant stereo matching based on relative gradients," in *Proc. 19th IEEE Int. Conf. Image Process.*, Sep. 2012, pp. 2989–2992.

[8] K. Zhang, J. Lu, G. Lafruit, R. Lauwereins, and L. V. Gool, "Robust stereo matching with fast normalized cross-correlation over shape-adaptive regions," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 2357–2360.

[9] S. Kim, B. Ham, B. Kim, and K. Sohn, "Mahalanobis distance cross-correlation for illumination-invariant stereo matching," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 11, pp. 1844–1859, Nov. 2014.

[10] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 1994, pp. 151–158.

[11] H. Hirschmuller and D. Scharstein, "Evaluation of stereo matching costs on images with radiometric differences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 9, pp. 1582–1599, Sep. 2009.

[12] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang, "On building an accurate stereo matching system on graphics hardware," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Nov. 2011, pp. 467–474.

[13] J. Shi, F. Fu, Y. Wang, W. Xu, and J. Wang, "Stereo matching with improved radiometric invariant matching cost and disparity refinement," in *Proc. Int. Conf. Intell. Comput.* Cham, Switzerland: Springer, 2016, pp. 61–73.

[14] T.-A. Chang, X. Lu, and J.-F. Yang, "Robust stereo matching with trinary cross color census and triple image-based refinements," *EURASIP J. Adv. Signal Process.*, vol. 2017, no. 1, p. 27, Dec. 2017.

[15] Z. Lee, J. Juang, and T. Q. Nguyen, "Local disparity estimation with three-moded cross census and advanced support weight," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 1855–1864, Dec. 2013.

[16] Y. Men, N. Ma, G. Zhang, X. Li, and C. Men, "A stereo matching algorithm based on four-moded census and relative confidence plane fitting," *Chin. J. Electron.*, vol. 24, no. 4, pp. 807–812, Oct. 2015.

[17] Y. Yang, D. Xu, S. Rong, G. Xie, and F. Chen, "An efficient stereo matching algorithm based on four-moded census transform for high-resolution images," *3D Res.*, vol. 9, no. 3, p. 33, Sep. 2018.

[18] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, nos. 1–3, pp. 7–42, Apr. 2002.

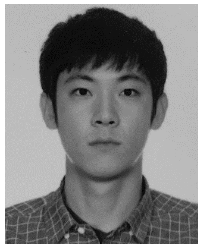
[19] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, 2003, p. 1.

[20] D. Scharstein and C. Pal, "Learning conditional random fields for stereo," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[21] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[22] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, Jan. 1996.
[23] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, Jun. 2010.
[24] L. Nanni, A. Lumini, and S. Brahnam, "Local binary patterns variants as texture descriptors for medical image analysis," *Artif. Intell. Med.*, vol. 49, no. 2, pp. 117–125, Jun. 2010.
[25] K. Wang, C.-E. Bichot, C. Zhu, and B. Li, "Pixel to patch sampling structure and local neighboring intensity relationship patterns for texture classification," *IEEE Signal Process. Lett.*, vol. 20, no. 9, pp. 853–856, Sep. 2013.
[26] T. Song, L. Xin, C. Gao, G. Zhang, and T. Zhang, "Grayscale-inversion and rotation invariant texture description using sorted local gradient pattern," *IEEE Signal Process. Lett.*, vol. 25, no. 5, pp. 625–629, May 2018.
[27] T. Song, H. Li, F. Meng, Q. Wu, B. Luo, B. Zeng, and M. Gabbouj, "Noise-robust texture description using local contrast patterns via global measures," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 93–96, Jan. 2014.
[28] T. Song, H. Li, F. Meng, Q. Wu, and J. Cai, "LETRIST: Locally encoded transform feature histogram for rotation-invariant texture classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 7, pp. 1565–1579, Jul. 2018.
[29] T.-L. Ji, M. K. Sundareshan, and H. Roehrig, "Adaptive image contrast enhancement based on human visual properties," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 573–586, Dec. 1994.
[30] M. K. Kundu and S. K. Pal, "Thresholding for edge detection using human psychovisual phenomena," *Pattern Recognit. Lett.*, vol. 4, no. 6, pp. 433–441, Dec. 1986.
[31] S. C. Nercessian, K. A. Panetta, and S. S. Agaian, "Non-linear direct multi-scale image enhancement based on the luminance and contrast masking characteristics of the human visual system," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3549–3561, Sep. 2013.
[32] T. S. Curry, J. E. Dowdey, and R. C. Murry, *Christensen's Physics of Diagnostic Radiology*. Philadelphia, PA, USA: Lippincott Williams & Wilkins, 1990.
[33] J. Barbur and A. Stockman, "Photopic, mesopic and scotopic vision and changes in visual performance," *Encyclopedia Eye*, vol. 3, pp. 323–331, Jan. 2010.
[34] D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nešić, X. Wang, and P. Westling, "High-resolution stereo datasets with subpixel-accurate ground truth," in *Proc. German Conf. Pattern Recognit.* Cham, Switzerland: Springer, 2014, pp. 31–42.
[35] D. R. J. Laming, *The Measurement of Sensation*. Oxford, U.K.: OUP Oxford, 1997.
[36] Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1657–1663, Jun. 2010.
[37] M. Menze and A. Geiger, "Object scene flow for autonomous vehicles," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3061–3070.
[38] D. Hafner, O. Demetz, and J. Weickert, "Why is the census transform good for robust optic flow computation?" in *Proc. Int. Conf. Scale Space Variational Methods Comput. Vis.* Cham, Switzerland: Springer, 2013, pp. 210–221.
[39] Z. Yin, T. Darrell, and F. Yu, "Hierarchical discrete distribution decomposition for match density estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 6044–6053.
[40] X. Guo, K. Yang, W. Yang, X. Wang, and H. Li, "Group-wise correlation stereo network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3273–3282.
[41] T. Yan, Y. Gan, Z. Xia, and Q. Zhao, "Segment-based disparity refinement with occlusion handling for stereo matching," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 3885–3897, Aug. 2019.

SEOWON JI received the B.S. degree in electrical engineering from Korea University, Seoul, South Korea, in 2015, where he is currently pursuing the Ph.D. degree in electrical engineering. His research interests include image processing, computer vision, and deep-learning.



SEUNG-WOOK KIM (Member, IEEE) received the B.S. and Ph.D. degrees in electronics engineering from Korea University, Seoul, South Korea, in 2012 and 2019, respectively. He is currently a Research Professor with the School of Electrical Engineering, Korea University. His research interests include deep-learning-based applications, image processing, and computer vision.

DONGPAN LIM received the B.S. degree in electronic engineering from Busan University, in 2011, and the M.S. degree in electrical engineering from Korea University, in 2019. He is currently working as a Senior Research Engineer with the S. LSI Division, Samsung Electronics.

SEUNG-WON JUNG (Senior Member, IEEE) received the B.S. and Ph.D. degrees in electrical engineering from Korea University, Seoul, South Korea, in 2005 and 2011, respectively. He was a Research Professor with the Research Institute of Information and Communication Technology, Korea University, from 2011 to 2012. He was a Research Scientist with the Samsung Advanced Institute of Technology, Yongin, South Korea, from 2012 to 2014. He was an Assistant Professor with the Department of Multimedia Engineering, Dongguk University, Seoul, from 2014 to 2020. He is currently an Assistant Professor with the School of Electrical Engineering, Korea University. He has published over 50 peer-reviewed articles in international journals. His current research interests include deep learning and computer vision applications. He received the Hae-Dong Young Scholar Award from the Institute of Electronics and Information Engineers, in 2019.

SUNG-JEA KO (Fellow, IEEE) received the B.S. degree in electronic engineering from Korea University, in 1980, and the M.S. and Ph.D. degrees in electrical and computer engineering from the State University of New York at Buffalo, in 1986 and 1988, respectively.

From 1988 to 1992, he was an Assistant Professor with the Department of Electrical and Computer Engineering, University of Michigan-Dearborn. In 1992, he joined the Department of Electronic Engineering, Korea University, where he is currently a Professor. He has published over 210 international journal articles. He also holds over 60 registered patents in fields, such as video signal processing and computer vision.

Dr. Ko was a 1999 recipient of the LG Research Award. He received the Best Paper Award from the IEEE Asia Pacific Conference on Circuits and Systems, in 1996, the Hae-Dong Best Paper Award from the Institute of Electronics and Information Engineers (IEIE), in 1997, the Research Excellence Award from Korea University, in 2004, and the Technical Achievement Award from the IEEE Consumer Electronics (CE) Society, in 2012. He received the 15-Year Service Award from the TPC of ICCE, in 2014, and the Chester Sall Award (First Place Transaction Paper Award) from the IEEE CE Society, in 2017. He was honored with the Science and Technology Achievement Medal from the Korean Government, in 2020. He has served as the General Chairman of ITC-CSCC 2012 and the IEICE 2013. He was the President of the IEIE, in 2013, the Vice-President of the IEEE CE Society, from 2013 to 2016, and the Distinguished Lecturer of the IEEE, from 2015 to 2017. He is a member of the National Academy of Engineering of Korea. He is a member of the editorial board of the IEEE Transactions on Consumer Electronics.