1. Introduction

Depth images represent distance information between the camera and objects in the captured scene. The depth map is usually provided with its corresponding color image as a pair, often called video-plus-depth [1]. Recently, efficient image generation methods for arbitrary view positions have been vital due to the development of multi-view display devices and three-dimensional (3D) contents. In particular, depth image-based rendering (DIBR) is one of the most widely used methods which create a virtually-synthesized image by projecting color and depth data onto a target-view image plane [2]. The performance of DIBR mainly depends on the quality of depth information.

In general, active sensor-based and passive sensor-based methods exist for measuring depth information from a natural scene. The former employs physical sensors, e.g., infrared ray (IR) sensor, to directly acquire depth data based on the principles of time-of-flight [3]. Usually, the active sensor is more effective in producing high-quality depth images than the passive sensor.

However, active sensors suffer from three inherent problems. First, depth data acquisition is difficult if the object is far from the sensor; off-the-shelf sensors allow measuring distances of within 10 m. Second, they are not applicable to outdoor environments. Finally, they produce low-resolution depth images, i.e., less than 640 x 480, due to challenging real-time distance measuring systems. Such inherent problems make active sensors not practical for various applications. In the industry, their usage is limited to applications mainly involving foreground extraction [4] and motion tracking [5] in indoor environment.

On the other hand, passive sensor-based methods indirectly estimate depth information from 2D images captured by cameras. Such methods can measure depth information of all objects in the captured scene unlike active sensor-based methods. In addition, indirect depth sensing of passive methods is applicable to both indoor and outdoor environments. Another advantage is that the depth image resolution depends on camera resolution, which is not limited to low resolution as in the active sensor. Due to such benefits, the ISO/IEC JTC1/SC29/WG11 Moving Picture Experts Group (MPEG) has utilized passive depth sensing rather than active depth sensing in the 3D video system standardization [6].

Stereo matching is one of the most widely used passive sensor-based methods. This process extracts 3D information from left and right images captured by a stereoscopic camera. In stereo matching, 3D information is calculated by examining the different perspective distortions of objects in the scene of two images. Consequently, in stereo normal case, the different image positions of corresponding image points called disparity is directly related to depth information based on camera parameters.

Over the past several decades, a variety of stereo matching methods have been developed to obtain high-quality disparity maps. However, accurate measurement of depth information from a natural scene still remains problematic due to difficult corre-
spondence matching in three types of regions: textureless, discontinuous depth, and occluded areas [7]. First, since color data of the textureless region in left and right images are so similar each other in a wide range, correspondence matching often fails because of its ambiguosness. Second, in case of the depth discontinuous region, i.e., the edge region, smeared color values exist, which leads to ineffective correspondence matching. Lastly, in the occluded region, some pixels may appear in the left image but not in the right image; accordingly, there is no corresponding pixel in the right image.

In this paper, we propose a distance transform-based disparity estimation method with occlusion handling to solve the important problems of stereo matching. Distance transform (DT) [8] calculates the distance to the closest edge for each pixel of an input image. DT values of left and right images control the luminance weighting term for better correspondence matching in edge regions. In addition, an energy function is modeled with three constraints to detect occluded regions. Occlusion hole filling based on color and spatial weighting functions is presented as well. In particular, the proposed hole filling method utilizes different shape of referred windows according to occlusion types, i.e., leftmost occlusion and inner occlusion.

The contributions of our work are as follows: (a) DT-based stereo matching is proposed to increase the accuracy of disparities in the edge region, (b) a new occlusion detection function is designed based on three constraints, and (c) occlusion hole filling is performed adaptively according to occlusion types.

The remainder of this paper is organized as follows. In Section 2, we state the problem in question and briefly introduce the related works. Then, Section 3 presents the proposed method in detail. Section 4 discusses the experimental results followed by conclusions in Section 5.

2. Problem statement

2.1. Occlusion and edge pixel problems

Over the past several decades, occlusion handling has been a challenging task in stereo matching. For left disparity map estimation, the occlusion region represents certain parts of an object that are visible in the left image but not in the right image, and vice versa. Fig. 1(a) illustrates the occlusion problem in left disparity map generation case. The red $\text{-}$marked region appears in the left image only, which means occlusion. The occlusion problem leads to failure of finding corresponding pixels in the right image.

Accurate measurement of depth information in the edge region is important in stereo matching, because depth data of object borders are usually distinguishable. However, as shown in Fig. 1(b), pixels around edges in the left and right images have smeared color values. This affects measuring of discontinuous disparities in the associated area. For reduction of ambiguity in discontinuous regions, several approaches employ variable window sizes or adaptive window shapes via segmentation or pixel-wise similarity measures. The proposed method produces similar effects compared with the approaches using variable window size or adaptive window shapes. While the classical approaches alter the window to determine the pixels, the proposed approach keeps the window and controls the influence of pixels within the regular scope. The proposed method can reduce the effect of inaccurate pixel determination and reflect enough edge influence by distance transform.

In the approaches using segmentation, the influence on the segmentation quality is greater than the algorithm itself. On the other hand, the proposed method can simply calculate the disparity map without prior work such as segmentation. Pixel-wise similarity measure enables the acquisition of scene details. However, this produces poor results in textureless areas and is very sensitive to image noise.

2.2. Previous work

In general, stereo matching can be categorized into local and global methods. Local methods are processed by windows based on correlation where the disparity is assumed to be equal for all pixels within the correlation window [9]. Nevertheless, at discontinuities, this assumption generates blurred object borders and removes small details depending on the size of the correlation window. Thus, such an assumption should be disregarded for depth discontinuities.

In global methods [10], the task of computing disparities is cast as an energy minimization problem. Typically, an energy function for obtaining a disparity map $D$ is formulated as

$$ E(D) = E_0(D) + \lambda E_s(D), $$

where $E_0$ is a data term which measures the pixel similarity and $E_s$ is called the smoothness term which penalizes disparity variations. Belief propagation [11], dynamic programming [12] and graph cuts [13,14] are well-known methods for solving this energy function. Generally, global methods are computationally complex even for low resolution images with a small disparity range. Thus, they are not practical. Recently, several methods have been introduced to reduce the complexity of global methods [15–19]. However, the performance of the algorithms considering the practical use is insufficient. Thus, further refinement process is necessary.

In regards to occlusion handling, Kolmogorov and Zabih [14] have proposed an additional occlusion term for the energy function to penalize occluded pixels. Then, the energy function is optimized via graph cuts to compute final disparities. The drawback is that the penalty of the occlusion term depends on only the uniqueness constraint. Liu et al. [20] have presented a two-step local method; the initial matching cost is computed using contrast contest histogram descriptors. Consecutively, disparity estimation is performed via two-pass weighted cost aggregation considering segmentation-based adaptive support weights. In this algorithm, disparity similarities of neighboring pixels which prevent disparity variations are inapplicable to localized results. Ben-Ari and Sochen [21] have introduced a variational approach to find corresponding points. Two coupled energy functions are included for half-occlusion handling and discontinuity map generation. Since optimization is repeated, high complexity is induced. Even though Jang’s method [22] generates high quality disparity maps, disparity information in edge regions are not estimated accurately due to its ambiguity. Furthermore, some errors in the non-occlusion region may propagate to the occlusion region during the disparity assignment process.

3. Proposed method

3.1. Overall framework

The proposed method is initially motivated by Yang’s work [17] based on hierarchical belief propagation. Due to the hierarchical structure, the previous work computes disparities accurately in the textureless region. Execution speed-wise, their work is one of the most effective global algorithms. For practical use, we adopt this method. However, the quality is insufficient, especially in regards to occlusion and depth discontinuity due to their ambiguosness. Thus, we sufficiently refine the results. Based on Yang’s work,
the proposed method uses distance transform to improve the disparity quality in the edge region. Furthermore, the proposed method includes occlusion handling.

Fig. 2 represents the overall framework of the proposed distance transform-based stereo matching with occlusion handling. For initial left and right disparity map generation, the proposed method is implemented as it follows: (1) distance transform (DT) including edge extraction is performed, (2) DT-based weighting function is computed, (3) luminance weighting function is calculated, (4) block-based stereo matching is carried out based on such weighting functions, and (5) disparity enhancement is performed.

For the occlusion handling process, (1) occluded regions are detected by cross check, warping, and luminance difference constraints, (2) color and spatial weighting functions are calculated, and (3) vacant pixels in the occluded region are filled by neighboring disparities chosen by the two weighting functions.

3.2. Distance transform-based stereo matching

In computer vision, DT is usually beneficial in tracing human motions, for example hand tracking [23, 24]. In this paper, we apply this to disparity estimation. Prior to the distance transform, the Canny edge operator [25] is used for extraction of color edge map from the image. The application of Canny edge operator to the input image may generate excessive edge information. Unnecessary isolated edge points may obstruct the purpose of improving the depth accuracy in discontinuity regions. In order to remove these, we apply a median filter to the original image prior to edge detection.

In order to obtain DT map, DT values in edge pixels are set to zero, while infinity is assigned to non-edge pixels, initially. Then, based on \( \alpha - \beta \) distance transform \((\alpha - \beta \ DT)\), the DT value \( r_{i,j}^k \) at iteration \( k \) is computed by

\[
 r_{i,j}^k = \min \left\{ r_{i,j-1}^{k-1} + \beta, r_{i-1,j}^{k-1} + \alpha, r_{i+1,j}^{k-1} + \beta, r_{i,j-1}^{k-1} + \alpha, r_{i,j+1}^{k-1} + \beta, r_{i+1,j}^{k-1} + \alpha, r_{i,j+1}^{k-1} + \beta \right\}, \tag{2}
\]

where \( \alpha \) and \( \beta \) control the strength of distance transform [26]. Fig. 3 illustrates the DT map generation procedure using 9–10 DT. As shown in Fig. 3, if the DT value of a pixel is close to zero, the pixel may belong to a textured area, i.e., the edge region. On the other hand, for pixels where the DT value is infinity, the pixel may belong to a non-textured area, i.e., the occluded region.
In case of a large DT value, i.e., the pixel is far from the edges, it belongs to a homogeneous region which is textureless.

In order to find corresponding points between left and right images, stereo matching defines an energy function composed of a data term and a smoothness term. When the energy function has the minimum value via energy optimization techniques such as graph cuts [13] and belief propagation [11], the optimal disparity value is determined.

Suppose there exists a left image $I_L$ and a right image $I_R$. Let $s$ and $t$ denote coordinates of pixels. $s$ is the center pixel of the local window $N(s)$ and $t$ is the neighboring pixel of $s$ within the window where $t \in N(s)$. The goal of stereo matching is to find the disparity $d_s$ of $s$. The energy function is formulated as

$$ E(d) = \sum_s D_s(d_s) + \sum_{s \in N(s)} S_{st}(d_s, d_t). $$

where $D_s(\cdot)$ indicates the data term of $s$ and $S_{st}(\cdot)$ represents the smoothness term between $s$ and $t$.

In this paper, for matching cost calculation, we employ the weighted absolute luminance difference between two blocks as the data term. In particular, the distance transform value $d_{tt}$ at $t$ controls the matching cost for better disparity estimation in the edge region. The proposed matching cost is defined by

$$ D_s(d_s) = \sum_{t \in N(s)} W_{st}(d_{tt}) \cdot F_{st}(d_s), $$

where $W_{st}$ is the weighting function at $t$ considering its DT value $d_{tt}$, and $F_{st}(\cdot)$ is the absolute luminance difference at $t$ with respect to $s$. In case of left disparity map generation, $F_{st}$ is represented by

$$ F_{st}(d_s) = \min(\|I_L(x_s, y_s) - I_R(x_t + d_s, y_t)\|, T_d), $$

where $(x_s, y_s)$ and $(x_t, y_t)$ are coordinates of $s$ and $t$, respectively. $T_d$ controls the data cost limit. The proposed DT-based weighting function $W_{st}$ is computed by

$$ W_{st}(d_{tt}) = f(d_{tt}) \cdot g(\|I_{st} - I_{tt}\|), $$

where $f(\cdot)$ is the DT function and $g$ is the luminance weighting function. $\|\|$ is the operator for calculating Euclidean distance between the luminance value $I_{st}$ at $s$ and the luminance value $I_{tt}$ at $t$ in the left image. In this work, $f$ and $g$ are modeled as

![Fig. 2. Overall framework of the proposed method.](image-url)
\[ f(d_t) = 1 - e^{-\frac{1}{\sigma_f}g(|l_{t,s} - l_{t,r}|)} = e^{-\frac{|l_{t,s} - l_{t,r}|^2}{2\sigma^2_f}}, \]  

where \( \sigma_f \) and \( \sigma_g \) are smoothing parameters of \( f \) and \( g \), respectively. \( \sigma_f \) and \( \sigma_g \) are usually defined as the standard deviation of the Gaussian function.

In (6), the DT function \( f \) is inversely proportional to the DT value \( d_t \), and \( 0 < f \leq 1 \). Since the smeared edge pixel problem makes correspondence searching difficult, \( f \) imposes small weighting values on them, i.e., less than 0.5. Fig. 4 exhibits the DT function. Fig. 4(a) shows the left image of Teddy and a magnified part. Fig. 4(b) shows its edge information and Fig. 4(c) represents the associated DT function. As shown in Fig. 4(c), the closer the pixel is located to edges, the smaller the DT weighting value is assigned to the pixel to reduce the smeared edge pixel problem.

The smoothness term \( S_s \) is based on the degree of difference among disparities of neighboring pixels. \( S_s \) is represented by

\[ S_s(d_t, d_r) = \min(\lambda|d_t - d_r|, T_s), \]  

where \( T_s \) is the constant controlling to deny cost increase. The smoothness strength \( \lambda \) is a scalar constant. We employ the smoothness term in Yang’s work [17].

### 3.3 Disparity map refinement considering occlusion and post-processing

Prior to final disparity generation, occluded regions should be extracted. For occlusion detection, we present three constraints: warping constraint, cross check constraint, and luminance difference constraint. In case that we find occluded regions in the left disparity map with the warping constraint, all pixels in the left image are projected to the right image coordinates using the left disparity map.

For occlusion determination, we introduce a right visiting map. If a projected pixel from the left image is matched with the coordinate of the right visiting map as a manner of one-to-one mapping, the left disparity is regarded as a reliable; its location does not belong to occluded regions. In contrast, if more than two projected pixels are assigned to the same coordinate of the right visiting map as a manner of many-to-one mapping, the corresponding disparity locations are assumed to be belonged to occluded regions. Fig. 5 illustrates the warping constraint. Since the number of visiting counts in the right visiting map is greater than one, the blue-marked pixels are regarded as candidates of occluded pixels.

For the warping constraint, we define an energy function \( E_w \) to cover aforementioned characteristics by

\[ E_w(D_t) = \sum_s W_w|o_t - W_t(s, D_t)|, \]  

where \( W_t \) is the warping constraint map, \( o_t \) is the hypothesized occlusion value, and \( w_w \) is the weighting factor. \( W_t(s, D_t) \) is a binary map constructed by the warping constraint. Multiple matching pixels in the left image are set to ‘1’. If pixel \( s \) is assumed to be an occluded pixel, the occlusion value \( o_t \) is set to ‘1’.

Second, the cross check constraint evaluates the mutual consistency of both disparity maps. If a particular pixel in the image is not an occluded pixel, the disparity values from both maps should be consistent. The corresponding points in both images have the same disparity value. The energy function \( E_c \) for the cross check constraint is calculated by

\[ E_c(D_l, D_r) = \sum_s |o_t - C_t(s; D_l, D_r)|, \]  

where \( C_t \) indicates the cross check constraint map.

\[
\begin{align*}
C_t & = 0, & \text{if } D_t(x_t) = D_l(x_t) - D_r(x_t) \\
C_t & = 1, & \text{otherwise}
\end{align*}
\]

\( D_l \) and \( D_r \) are the left and right disparity maps respectively. \( x_t \) is a pixel in the left image. If the left disparity is equal to its right disparity at the corresponding pixel coordinate, \( C_t \) is set to zero in (11). When \( C_t = 1 \), the possibility of its disparity location being included in occluded regions is high.

Lastly, the luminance difference constraint is defined by (12). We use the luminance difference as the matching cost. This comes from the assumption that the large difference of luminance generates wrong matching even if a particular pixel is regarded as a visible pixel by warping and cross check constraint.

\[ D_{ld}(s) = |l_t(x_t) - l_l(x_t) - D_l(x_t)|. \]  

The final energy function for occlusion detection is defined as

\[
E_{00} = \sum_s (1 - o_t) \cdot D_{ld}(s) + \lambda_o o_t + \lambda_w E_w(D_t) + \lambda_e E_c(D_l, D_r)
\]

\[ + \sum_{l \in \mathcal{N}(s)} \lambda_{l2} |o_t - o_l|. \]

\( \lambda_o o_t \) is the cost of penalty for occlusion labeling. This is necessary to balance the luminance difference constraint term. It prevents the
whole occlusion map from being labeled as occlusion. In (13), the last term represents the smoothness term for the energy function of occlusion detection and it uses Sum of Absolute Difference (SAD) among the neighboring pixels of pixel $s$. This final function is optimized by belief propagation [17].

After occlusion detection, the reasonable disparity value should be assigned to the occluded pixel. Since occlusion is only visible in one image, it is impossible to determine the accurate disparity value by means of conventional stereo matching. The vacant disparity of a pixel in the occluded region can be filled with the disparities of its four neighboring pixels with the assumption that disparity values in occluded pixels are similar to those of near non-occluded pixels. The proposed method propagates the disparity values of non-occluded pixels to occluded pixels. First, we classify occlusion regions into leftmost and inner occlusion parts. Fig. 6 shows the left image and the corresponding occlusion map. The red part in Fig. 6(b) is the leftmost and the rest of the occlusion is the inner part.

The reason why occlusion in the inner part is occurred is as follows. In the right image, the object occludes the background which exists at the left-side of the object in the left image. Thus, the reasonable disparity value in the inner occlusion can be obtained from the left-side background of the occlusion.

In order to assign the proper data to inner occlusion, a potential energy function is defined. Let $L(s)$ be the neighboring pixels whose distance from occluded pixel $s$ is smaller than the predefined distance and $C = \{s, t\}$ horizontal coordinate of $s >$ horizontal coordinate of $t$, $t \in L(s)$ be the set of all nearby pixels which affect pixel $s$. $B = \{s, t\}$ $d_s \neq d_t, t \in C$ and $o_t$ is the occlusion value from the obtained occlusion map. Formally, the potential energy function for disparity assignment is defined in (13).

$$E_{dm}(s, d_s) = \sum_{t \in C, B} (1 - o_t) \frac{1}{\text{dist}(s, t)} \exp \left( \frac{\text{diff}_{st}}{\sigma_{dm}} \right).$$

where dist$(s, t)$ is the spatial distance and diff$_{st}$ is the color difference between occluded pixel $s$ and visible pixel $t$. The disparity value, which has the maximum value of (14), is determined as the disparity for the pixel $s$. This process assigns the optimal disparity by finding the similar region to the occlusion part according to the weighting of distance.

The occlusion handling process in the inner part works at only occluded pixels which are near visible pixels. Thus, it completely handles thin or small occlusion. However, wide and large occlusion is processed at only near visible pixels. In order to solve this problem, we apply the potential energy function for occlusion handling repeatedly until all occluded pixels are removed.

Occlusion in the leftmost part is generated due to the non-existence of this occlusion region in the right image. Thus, it is useless to estimate the disparity using left-side neighboring region of
occlusion in the leftmost part. In addition, disparity extension of the leftmost visible pixels to this occlusion part for each horizontal line is also risky [22].

In order to handle the leftmost part, we search the analogous region to current pixel at neighboring of the leftmost occlusion. Mask shape for search is different from inner occlusion. Fig. 7 shows the mask shape of inner and leftmost occlusion, respectively. The red pixel is the current pixel in the occlusion region and the others are the pixels that affect disparity assignment of the current pixel.

The measure of likeness for finding the analogous region in the leftmost occlusion is defined by (15) according to the distances and color differences from neighbor pixels. The disparity value of the most analogous region is selected as optimal disparity value of the current occlusion.

\[
f(s, t) = \arg\max_{d_i} \left( 1 - o_i \right) \frac{1}{\text{dist}(s, t)} \exp \left( -\frac{\text{diff}_{st}}{\sigma^2_{st}} \right).
\]  

(15)

Some papers consider occlusion types [27,28]. However, they do not consider the mask shape according to the occlusion characteristics. In the leftmost part, the disparity values of the leftmost visible pixels are simply extended to the leftmost occlusion part for each horizontal line. In the inner part, small and large occlusion regions are handled separately.

After occlusion handling, we enhance the disparity map based on Yang’s work as a post-processing. For disparity enhancement, five candidate pixels \( t_1, t_2, t_3, t_4, t_5 \) are selected in the local window; \( t_1, t_2, t_3, t_4, t_5 \) are the left, right, center, top and bottom pixels in the local window. When \( s \) is \((x_s, y_s)\) coordinate, the candidates are defined by

Fig. 8. Comparison of initial work results. (a) CSBP in Teddy; (b) DT-based method in Teddy; (c) ground truth in Teddy; (d) CSBP in Cones; (e) DT-based method in Cones; and (f) ground truth in Cones.

Fig. 9. Comparison of results. (a) GC + occ; (b) CCH + SegAggr; (c) VarMSOH; (d) Jang’s; (e) proposed method; and (f) ground truth.
Each candidate has its own cost based on spatial and color weighting functions \( \varphi \) and \( \psi \). Formally, the cost \( C_{st} \) at \( t \) with respect to \( s \) is calculated by
\[
C_{st} = \varphi(|s - t|) \cdot \psi(|d_t - d_s|),
\]
where \( \sigma_\varphi \) and \( \sigma_\psi \) are smoothing parameters of \( \varphi \) and \( \psi \), respectively.

Then, we seek \( t_s \) that has the minimum cost among the five candidate set \( Q = \{t_1, t_2, t_3, t_4, t_5\} \). \( t_s \) is represented by
\[
t_s = \arg \min \{C_{s,t_1}, C_{s,t_2}, C_{s,t_3}, C_{s,t_4}, C_{s,t_5}\}.
\]
Finally, \( s \) is assigned by the disparity at \( t_s \).

### 4. Experimental results

In order to evaluate the performance of the proposed method, we tested with four stereo image sets with different image size. These reference test data are Tsukuba, Venus, Teddy, and Cones provided by Middlebury Stereo [29].

In the experiment, \( \alpha \) and \( \beta \) values for distance transform are set to 9 and 10, respectively. The larger the strength of DT, the lesser the effect of the edge. Our contribution is to reduce the effect of smeared pixels near the edges for accurate disparity estimation. Thus, we use large strength. For DT-based weighting function, smoothing parameters \( \sigma_f \) and \( \sigma_g \) in (7) are set to 0.3 and 0.2, respectively. For the proposed occlusion detection, each parameter is the weighting of each term in (13). However, the luminance difference constraint term is not weighted. Thus, numerical values of parameters are determined to achieve similar impact by each parameter according to the luminance difference constraint term. \( \sigma_o, \sigma_{on}, \sigma_{c}, \text{ and } \sigma_s \) in (13) are set to 7.5, 12, 12, and 4.2 to balance each term of the energy function. For the occlusion hole filling, \( \sigma_{ds} \) in (14) and (15) is set to 7.

Fig. 8 illustrates the visual comparison of Yang’s work [17] with the initial disparity map using the proposed DT-based stereo matching. The result of Fig. 8 demonstrates that the proposed initial disparity generation improves the quality in edge regions. The final results of our method adding occlusion handling are presented in Fig. 9. Fig. 9 also includes the results of the above other methods including occlusion handling and ground truth disparity maps.

In the proposed method, we adopted Yang’s work which generates unsatisfactory disparity quality, but extremely fast. We did not apply iterative process for occlusion handling. Thus, the computational complexity of the proposed method depends on that of Yang’s work which is one of the fastest algorithms among global methods. In fact, the proposed method for all stereo pair runs in less than five seconds on a 2.67 GHz Intel Core machine.

In order to evaluate our final disparity map, we compare our proposed method with other methods which have good performance with occlusion handling. Table 1 shows the objective evaluation which measures the percentages of bad matching pixels [29]. This measure is computed for three subsets of the image: non-occluded, whole, and discontinuity regions, denoted as “nonocc”, “all”, and “disc”, respectively. When the absolute disparity error is greater than one pixel, the pixel is regarded as a bad pixel. The subscripts of error rate in Table 1 represents rankings among the presented methods. These results indicate that the proposed method outperforms other comparative methods by 3.80%, 0.73%, 0.53%, 0.63%, and 0.50% on average.

Table 2 shows the percentages of bad matching pixels in the occlusion region. The quality of the proposed method in occlusion outperforms other comparative methods by 51.75%, 11.25%, 3.43%, 9.58%, and 13.84% on average. These results show that our method is highly effective in occlusion handling. The main contribution of the proposed method, i.e., disparity map refinement, can be

---

**Table 1**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Tsukuba</td>
<td>2.00</td>
<td>1.91</td>
<td>1.74</td>
<td>3.97</td>
<td>1.42</td>
<td>1.67</td>
</tr>
<tr>
<td>all</td>
<td>4.17</td>
<td>2.01</td>
<td>2.11</td>
<td>5.23</td>
<td>2.30</td>
<td>2.23</td>
</tr>
<tr>
<td>disc</td>
<td>10.50</td>
<td>6.24</td>
<td>9.23</td>
<td>14.90</td>
<td>7.94</td>
<td>9.35</td>
</tr>
<tr>
<td>Venus</td>
<td>1.48</td>
<td>1.64</td>
<td>0.41</td>
<td>0.28</td>
<td>0.91</td>
<td>0.43</td>
</tr>
<tr>
<td>all</td>
<td>3.11</td>
<td>2.19</td>
<td>0.94</td>
<td>0.76</td>
<td>1.54</td>
<td>0.72</td>
</tr>
<tr>
<td>disc</td>
<td>17.70</td>
<td>6.75</td>
<td>3.97</td>
<td>3.78</td>
<td>12.71</td>
<td>3.93</td>
</tr>
<tr>
<td>Teddy</td>
<td>11.10</td>
<td>11.26</td>
<td>8.03</td>
<td>9.34</td>
<td>6.34</td>
<td>7.19</td>
</tr>
<tr>
<td>disc</td>
<td>27.50</td>
<td>19.80</td>
<td>19.80</td>
<td>20.00</td>
<td>17.59</td>
<td>19.48</td>
</tr>
<tr>
<td>Cones</td>
<td>5.98</td>
<td>5.36</td>
<td>7.07</td>
<td>4.14</td>
<td>4.96</td>
<td>5.14</td>
</tr>
<tr>
<td>all</td>
<td>16.50</td>
<td>12.40</td>
<td>12.90</td>
<td>9.91</td>
<td>12.70</td>
<td>12.49</td>
</tr>
<tr>
<td>disc</td>
<td>16.00</td>
<td>13.00</td>
<td>16.30</td>
<td>11.40</td>
<td>14.44</td>
<td>15.12</td>
</tr>
<tr>
<td>Average bad pixels</td>
<td>11.34</td>
<td>8.27</td>
<td>8.07</td>
<td>8.17</td>
<td>8.04</td>
<td>7.54</td>
</tr>
<tr>
<td>Average ranking</td>
<td>5.42</td>
<td>3.25</td>
<td>3.67</td>
<td>3.17</td>
<td>2.63</td>
<td>2.67</td>
</tr>
</tbody>
</table>

---

**Table 2**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Tsukuba</td>
<td>86.36</td>
<td>33.04</td>
<td>15.85</td>
<td>52.97</td>
<td>35.87</td>
<td>24.42</td>
</tr>
<tr>
<td>Venus</td>
<td>89.67</td>
<td>31.09</td>
<td>28.93</td>
<td>26.18</td>
<td>34.92</td>
<td>16.25</td>
</tr>
<tr>
<td>Teddy</td>
<td>96.05</td>
<td>67.08</td>
<td>64.79</td>
<td>55.01</td>
<td>75.88</td>
<td>56.30</td>
</tr>
<tr>
<td>Cones</td>
<td>94.23</td>
<td>73.11</td>
<td>63.45</td>
<td>63.49</td>
<td>68.00</td>
<td>62.33</td>
</tr>
<tr>
<td>Average bad pixels</td>
<td>91.58</td>
<td>51.08</td>
<td>43.26</td>
<td>49.41</td>
<td>53.67</td>
<td>39.83</td>
</tr>
<tr>
<td>Average ranking</td>
<td>6.00</td>
<td>4.00</td>
<td>2.25</td>
<td>2.75</td>
<td>4.50</td>
<td>1.50</td>
</tr>
</tbody>
</table>
applied to other methods. The final quality depends on the performance of the algorithm that it is based on. The performance improvement from reference methods can be examined by applying the proposed method to other approaches.

5. Conclusions

This paper proposes a disparity estimation method solving discontinuity and occlusion issues which cause inherent problems of stereo matching. The proposed method exploits key techniques: distance transform based discontinuity preserving disparity estimation, occlusion detection via three constraints and occluded region filling. These techniques significantly improve the disparity quality maintaining the practicality. Experimental results show that the proposed method produces more accurate disparity maps compared to widely used other methods that incorporate occlusion handling.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2013-067321).

References