J. Vis. Commun. Image R. 25 (2014) 1595–1603

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Discontinuity preserving disparity estimation with occlusion handling

Woo-Seok Jang *, Yo-Sung Ho

Gwangju Institute of Science and Technology (GIST), 123 Cheomdan-gwagiro, Buk-gu, Gwangju 500-712, Republic of Korea

ARTICLE INFO

Article history:
Received 28 October 2013
Accepted 23 July 2014
Available online 1 August 2014

Keywords:
Distance transform
Occlusion handling
Stereo vision
Energy optimization
Depth image-based rendering
Depth discontinuity
Hierarchical structure
3D content

ABSTRACT

In this paper, we propose a stereo matching algorithm based on distance transform to generate high-quality disparity maps with occlusion handling. In general, pixel intensities around object edges are smeared due to mixed values located between the object and its background. This leads to problems when identifying discontinuous disparities. In order to handle these problems, we present an edge control function according to distance transform values. Meanwhile, occluded regions occur, i.e., some portions are visible only in one image. An energy function is designed to detect such regions considering warping, cross check, and luminance difference constraints. Consequently, we replace the disparity in the occluded region with the one chosen from its neighboring disparities in the non-occluded region based on color and spatial correlations. In particular, the occlusion hole is filled according to region types. Experimental results show that the proposed method outperforms conventional stereo matching algorithms with occlusion handling.

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 640×480, 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-

* Corresponding author. Tel.: +82 62 715 2258; fax: +82 62 715 3164.
E-mail addresses: jws@gist.ac.kr (W.-S. Jang), hoyo@gist.ac.kr (Y.-S. Ho).
URL: http://vclab.gist.ac.kr (W.-S. Jang).

http://dx.doi.org/10.1016/j.jvcir.2014.07.005
1047-3203/

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 ambiguousness. 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[1]-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

$$E(D) = E_D(D) + \lambda E_s(D), \tag{1}$$

where E_D 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 handing, 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 ambiguity. Thus, we sufficiently refine the results. Based on Yang's work,

[1] For interpretation of color in Figs. 1 and 5–7, the reader is referred to the web version of this article.

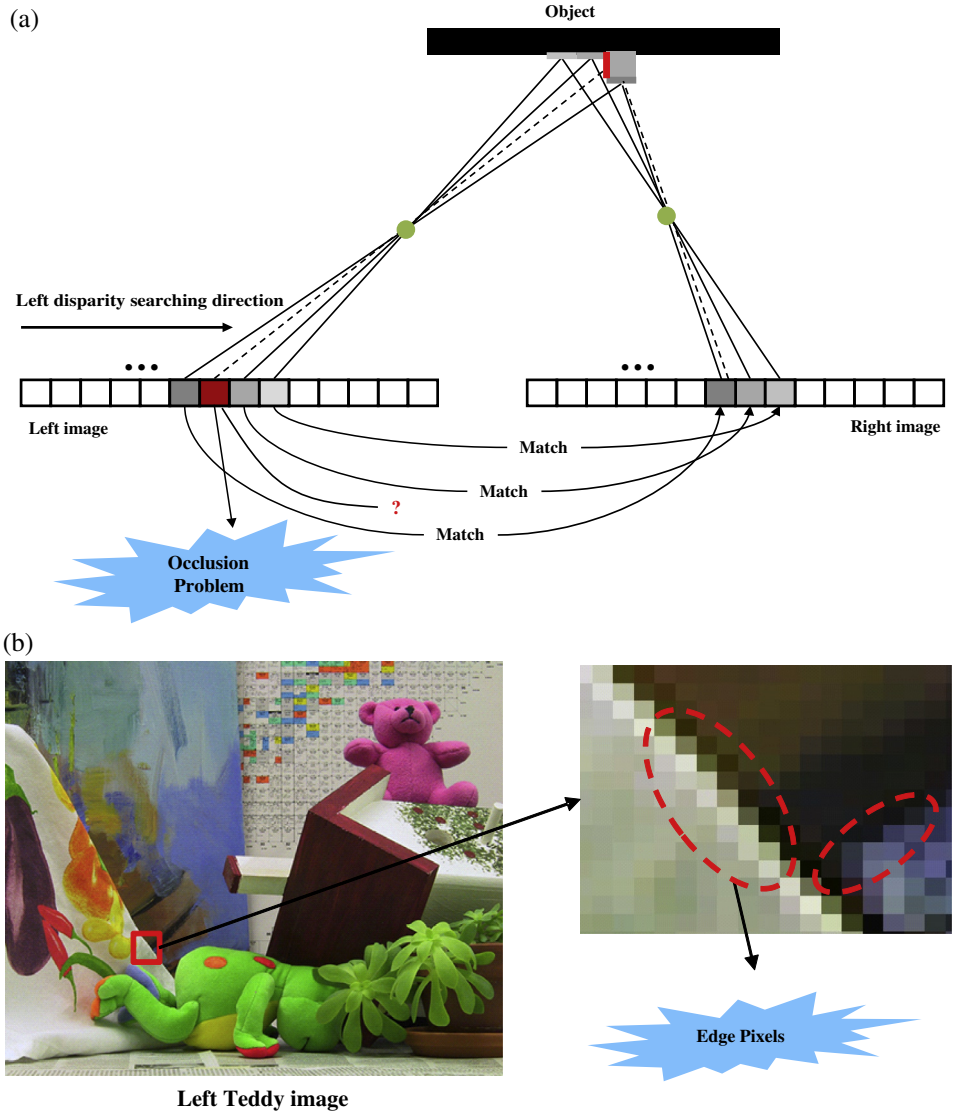

Fig. 1. Occlusion and smeared edge pixel problem. (a) Occlusion region is only visible in one image, causing mismatching. (b) Smeared pixels near the edge affect the discontinuity measure.

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 α–β distance transform (α–β DT), the DT value $r^k_{i,j}$ at iteration k is computed by

$$r^k_{ij} = \min \begin{bmatrix} r^{k-1}_{i-1\,j-1} + \beta & r^{k-1}_{i-1\,j} + \alpha & r^{k-1}_{i-1\,j+1} + \beta \\ r^{k-1}_{i\,j-1} + \alpha & r^{k-1}_{ij} & r^{k-1}_{i\,j+1} + \alpha \\ r^{k-1}_{i+1\,j-1} + \beta & r^{k-1}_{i+1\,j} + \alpha & r^{k-1}_{i+1\,j+1} + \beta \end{bmatrix}, \tag{2}$$

where α and β 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

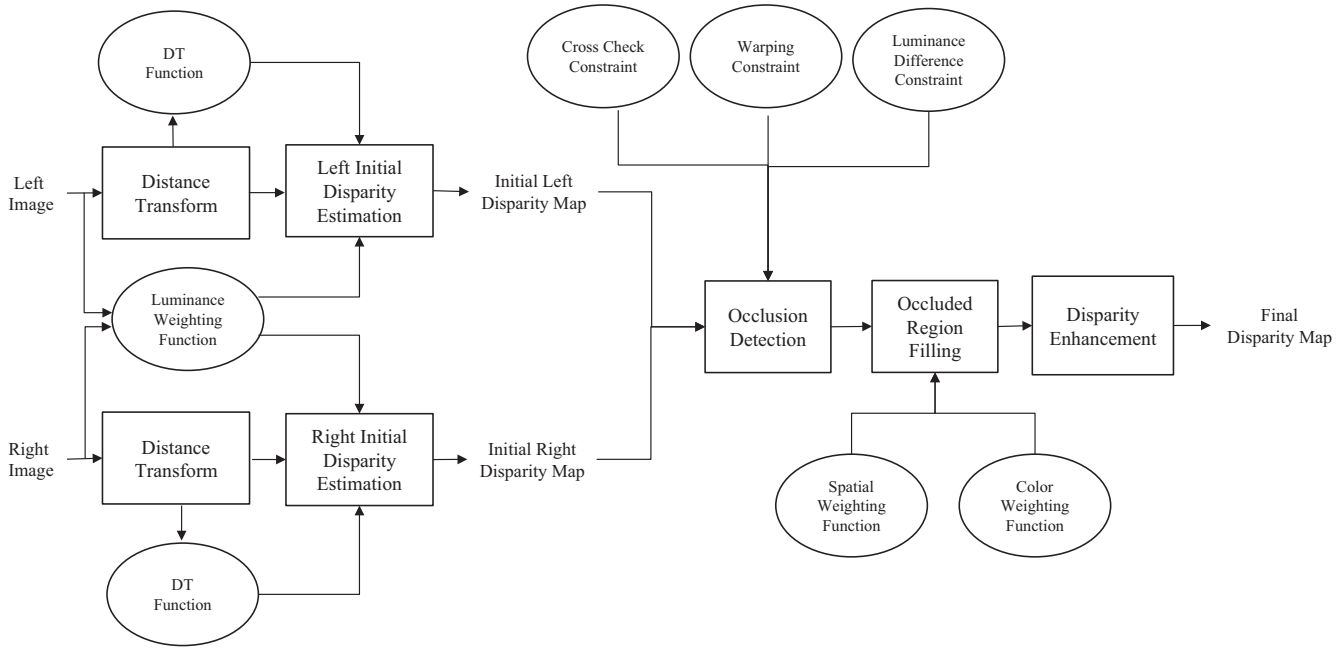

Fig. 2. Overall framework of the proposed method.

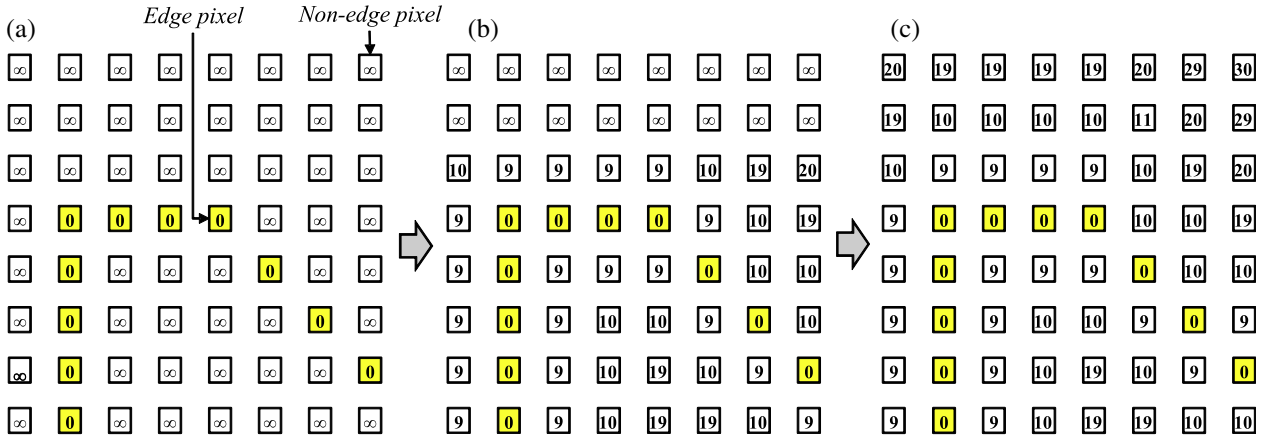

Fig. 3. DT map generation. (a) Initialization; (b) 1st iteration ($k = 1$); and (c) 2nd iteration ($k = 2$).

hand, 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

$$E(d) = \sum_{s} D_s(d_s) + \sum_{s,t \in N(s)} S_{s,t}(d_s, d_t), \tag{3}$$

where $D_s(\cdot)$ indicates the data term of s and $S_{s,t}(\cdot)$ 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 dt_t at t controls the matching cost for better disparity estimation in the edge region. The proposed matching cost is defined by

$$D_s(d_s) = \frac{\sum_{t \in N(s)} W_{s,t}(dt_t) \cdot F_{s,t}(d_s)}{\sum_{t \in N(s)} W_{s,t}(dt_t)}, \tag{4}$$

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

$$F_{s,t}(d_s) = \min(|I_L(x_s, y_s) - I_R(x_t + d_s, y_t)|, T_d), \tag{5}$$

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_t is computed by

$$W_{s,t}(dt_t) = f(dt_t) \cdot g(|I_{L,s} - I_{L,t}|), \tag{6}$$

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

$$f(dt_t) = 1 - e^{-\frac{dt_t^2}{2\sigma_f}}, g(|I_{L,s} - I_{L,t}|) = e^{-\frac{|I_{L,s}-I_{L,t}|^2}{2\sigma_g}}, \quad (7)$$

where σ_f and σ_g are smoothing parameters of f and g, respectively. σ_f and σ_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 dt_t, and $0 \leqslant f \leqslant 1$. 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 $S_{s,t}$ is based on the degree of difference among disparities of neighboring pixels. $S_{s,t}$ is represented by

$$S_{s,t}(d_s, d_t) = \min(\lambda|d_s - d_t|, T_s), \quad (8)$$

where T_s is the constant controlling to deny cost increase. The smoothness strength λ 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

$$E_w(D_L) = \sum_s w_w |o_s - W_L(s, D_L)|, \quad (9)$$

where W_L is the warping constraint map, o_s is the hypothesized occlusion value, and w_w is the weighting factor. $W_L(s,D_L)$ 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_s 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

$$E_c(D_L, D_R) = \sum_s |o_s - C_L(s; D_L, D_R)|, \quad (10)$$

where C_L indicates the cross check constraint map.

$$\left\{ \begin{array}{ll} C_L = 0, & \text{if } D_L(x_s) = D_R(x_s - D_L(x_s)) \\ C_L = 1, & \text{otherwise} \end{array} \right\}. \quad (11)$$

D_L and D_R are the left and right disparity maps respectively. x_s is a pixel in the left image. If the left disparity is equal to its right disparity at the corresponding pixel coordinate, C_L is set to zero in (11). When $C_L = 1$, 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.

$$D_{cd}(s) = |I_L(x_s) - I_R(x_s - D_L(x_s))|. \quad (12)$$

The final energy function for occlusion detection is defined as

$$E_{OD} = \sum_x [(1 - o_s) \cdot D_{cd}(s) + \lambda_o o_s] + \lambda_w E_w(D_L) + \lambda_x E_c(D_L, D_R) + \sum_s \sum_{s,t \in N(s)} \lambda_s |o_s - o_t|. \quad (13)$$

$\lambda_o o_s$ is the cost of penalty for occlusion labeling. This is necessary to balance the luminance difference constraint term. It prevents the

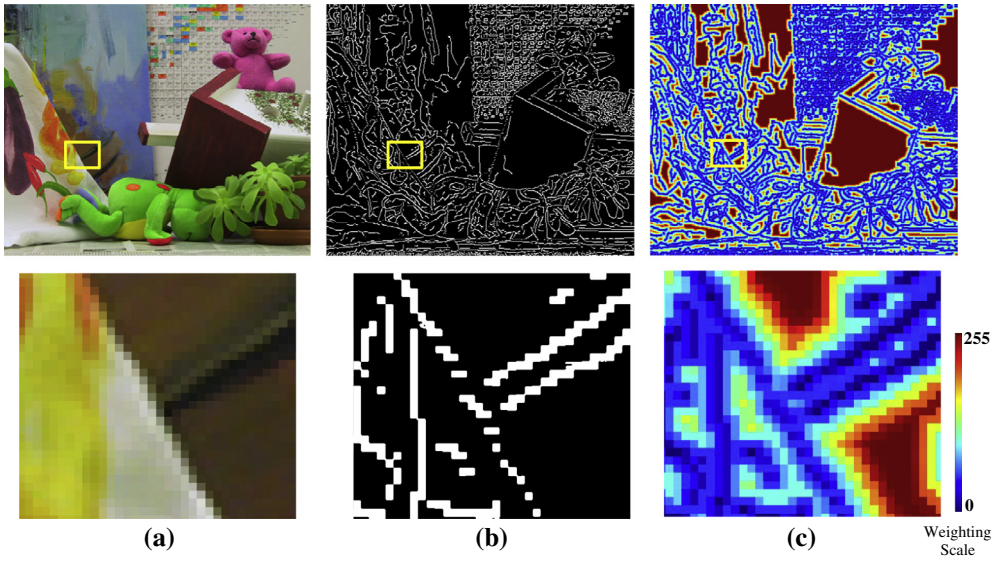

Fig. 4. DT function. (a) Color data; (b) edge data; and (c) DT weighting value.

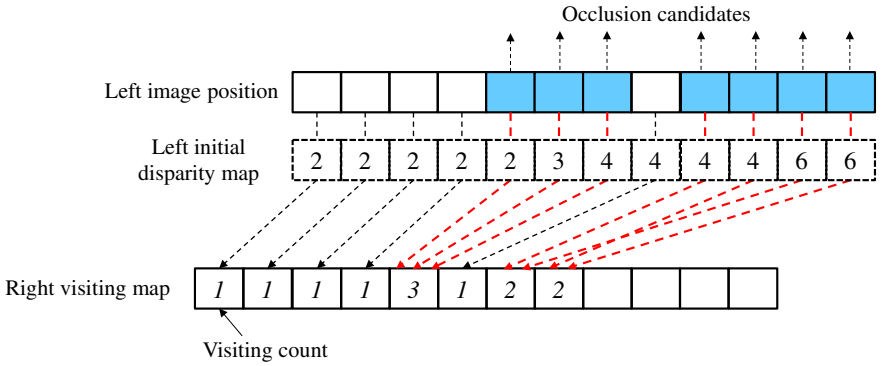

Fig. 5. Warping constraint. Many-to-one mapping pixels are regarded as occlusion.

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 $s \geqslant$ 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).

$$E_{DA}(s,d_s) = \sum_{t \in C \setminus B} (1 - o_t) \frac{1}{dist(s,t)} \exp\left(-\frac{diff_{s,t}}{\sigma_{da}^2}\right), \tag{14}$$

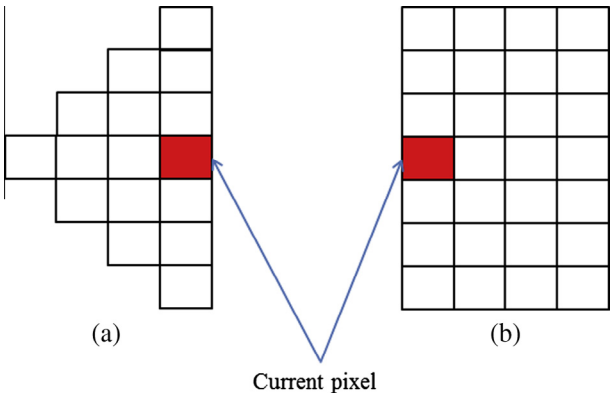

Fig. 7. Mask shapes according to occlusion types. (a) Inner occlusion and (b) leftmost occlusion.

where $dist(s,t)$ is the spatial distance and $diff_{s,t}$ 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

Fig. 6. Two kinds of occlusion. (a) Color image and (b) occlusion map.

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.

$$f(s,t) = \underset{d_t}{argmax}(1-o_t)\frac{1}{dist(s,t)}\exp\left(-\frac{diff_{s,t}}{\sigma_{da}^2}\right). \quad (15)$$

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

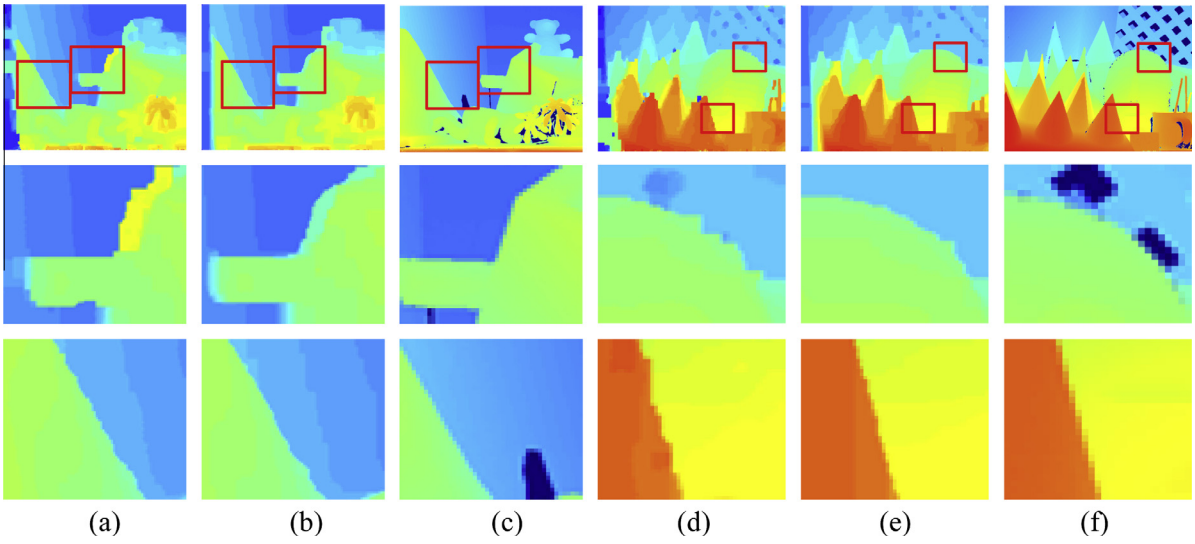

Fig. 8. Comparison of initial work results. (a) CSBP in *Teddy*; (b) DT-based method in *Teddy*; (c) ground truth in *Teddy*; (d) CSBP in *Cones*; (e) DT-based method in *Cones*; and (f) ground truth in *Cones*.

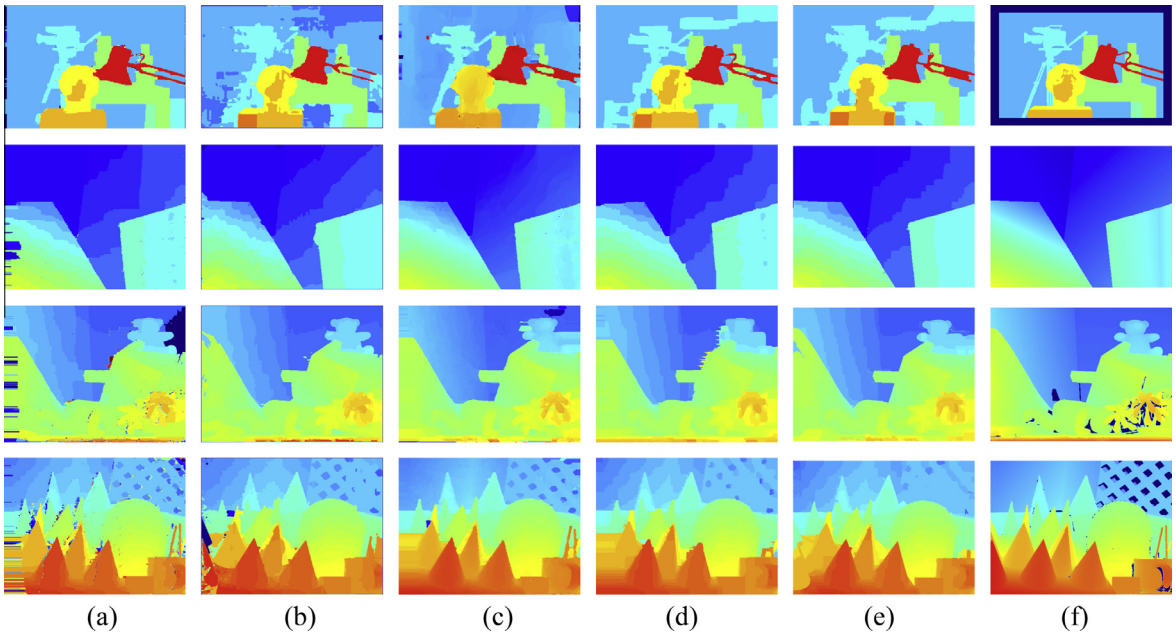

Fig. 9. Comparison of results. (a) GC + occ; (b) CCH + SegAggr; (c) VarMSOH; (d) Jang's; (e) proposed method; and (f) ground truth.

Table 1
Objective evaluation of the proposed method, comparing the percentage of bad pixels in the non-occluded region (nonocc), all regions (all), and regions near depth discontinuities (disc). The subscripts of error rate are the ranking among the presented methods.

Algorithm		CSBP [17]	GC + occ [14]	CCH + SegAggr [20]	VarMSOH [21]	Jang's method [22]	Proposed method
Tsukuba	nonocc	2.00_5	1.19_1	1.74_4	3.97_6	1.42_2	1.67_3
	all	4.17_5	2.01_1	2.11_2	5.23_6	2.30_4	2.25_3
	disc	10.50_5	6.24_1	9.23_3	14.90_6	7.94_2	9.35_4
Venus	nonocc	1.48_5	1.64_6	0.41_2	0.28_1	0.91_4	0.43_3
	all	3.11_6	2.19_5	0.94_3	0.76_2	1.54_4	0.72_1
	disc	17.70_6	6.75_4	3.97_3	3.78_1	12.71_5	3.93_2
Teddy	nonocc	11.10_5	11.20_6	8.08_3	9.34_4	6.34_1	7.19_2
	all	20.20_6	17.40_5	14.3_3	14.30_4	13.62_2	12.33_1
	disc	27.50_6	19.80_3	19.80_4	20.00_5	17.59_1	19.48_2
Cones	nonocc	5.98_5	5.36_3	7.07_6	4.14_1	4.96_2	5.51_4
	all	16.50_6	12.40_2	12.90_5	9.91_1	12.70_4	12.49_3
	disc	16.00_5	13.00_2	16.30_6	11.40_1	14.44_3	15.12_4
Average bad pixels		11.34_6	8.27_5	8.07_3	8.17_4	8.04_2	7.54_1
Average ranking		5.42_6	3.25_4	3.67_5	3.17_3	2.83_2	2.67_1

$$t_1:(x_s-1,y_s), t_2:(x_s+1,y_s), t_3:(x_s,y_s), t_4:(x_s,y_s-1), t_5:(x_s,y_s+1). \tag{16}$$

Each candidate has its own cost based on spatial and color weighting functions φ and ψ. Formally, the cost $C_{s,t}$ at t with respect to s is calculated by

$$C_{s,t} = \varphi(|s-t|)\cdot\psi(|d_s-d_t|), \tag{17}$$

where σ_φ and σ_ψ are smoothing parameters of φ and ψ, respectively.

Then, we seek t_x that has the minimum cost among the five candidate set $Q=\{t_1,t_2,t_3,t_4,t_5\}$. t_x is represented by

$$t_x : argmin\{C_{s,t_1}, C_{s,t_2}, C_{s,t_3}, C_{s,t_4}, C_{s,t_5}\}. \tag{18}$$

Finally, s is assigned by the disparity at t_x.

4. Experimental results

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

In the experiment, α and β 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 σ_f and σ_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. λ_o, λ_w, λ_c, and λ_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, σ_{da} 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 subscript 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 2
Performance comparison in occlusion. The percentage of bad pixels in the occluded region is used as measure. The subscripts mean the accuracy ranking in the occluded region.

Algorithm	CSBP [17]	GC + occ [14]	CCH + SegAggr [20]	VarMSOH [21]	Jang's method [22]	Proposed method
Tsukuba	86.36_6	33.04_3	15.85_1	52.97_5	35.87_4	24.45_2
Venus	89.67_6	31.09_4	28.93_3	26.18_2	34.92_5	16.25_1
Teddy	96.05_6	67.08_4	64.79_3	55.01_1	75.88_5	56.30_2
Cones	94.23_6	73.11_5	63.45_2	63.49_3	68.00_4	62.33_1
Average bad pixels	91.58_6	51.08_4	43.26_2	49.41_3	53.67_5	39.83_1
Average ranking	6.00_6	4.00_4	2.25_2	2.75_3	4.50_5	1.50_1

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

[1] C. Fehn, R.S. Pastoor, Interactive 3-DTV-concepts and key technologies, Proc. IEEE 94 (2006) 524–538.
[2] L. Zhang, T. Wa James, Stereoscopic image generation based on depth images for 3D TV, IEEE Trans. Broadcast. 51 (2005) 191–199.
[3] S.-Y. Kim, J.H. Cho, A. Koschan, M.A. Abidi, 3D video generation and service based on a TOF depth sensor in MPEG-4 multimedia framework, IEEE Trans. Consum. Electron. 56 (2010) 1730–1738.
[4] R. Crabb, C. Tracey, A. Puranik, J. Davis, Real-time foreground segmentation via range and color imaging, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008, pp. 1–5.
[5] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, A. Blake, Real-time human pose recognition in parts from single depth images, in: IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 1297–1304.
[6] Video and Requirement Group, Call for Proposals on 3D Video Coding Technology N12036, ISO/IEC JTC1/SC29/WG11, 2011.
[7] S.-Y. Kim, A. Koschan, M. Abidi, Y.-S. Ho, Three-dimensional video contents exploitation in depth camera-based hybrid camera system, in: M. Mrak, M. Grgic, M. Kunt (Eds.), High-Quality Visual Experience, Springer, Berlin, Heidelberg, 2010, pp. 349–369.
[8] G. Borgefors, Distance transformations in digital images, Comput. Vision Graph. Image Process. 34 (1986) 344–371.
[9] H. Hirschmüller, P. Innocent, J. Garibaldi, Real-time correlation-based stereo vision with reduced border errors, Int. J. Comput. Vision 47 (2002) 229–246.
[10] M. Bleyer, M. Gelautz, Graph-cut-based stereo matching using image segmentation with symmetrical treatment of occlusions, Signal Process.: Image Commun. 22 (2007) 127–143.
[11] J. Sun, N.-N. Zheng, H.-Y. Shum, Stereo matching using belief propagation, IEEE Trans. Pattern Anal. Mach. Intell. 25 (2003) 787–800.
[12] Z.-W. Gao, W.-K. Lin, Y.-S. Shen, C.-Y. Lin, W.-C. Kao, Design of signal processing pipeline for stereoscopic cameras, IEEE Trans. Consum. Electron. 56 (2010) 324–331.
[13] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2001) 1222–1239.
[14] V. Kolmogorov, R. Zabih, Computing visual correspondence with occlusions using graph cuts, in: IEEE International Conference on Computer Vision, 2001, pp. 508–515.
[15] P.F. Felzenszwalb, D.P. Huttenlocher, Efficient belief propagation for early vision, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004, pp. I-261–I-268.
[16] Q. Yang, L. Wang, R. Yang, S. Wang, M. Liao, D. Nister, Real-time global stereo matching using hierarchical belief propagation, in: British Machine Vision Conference, 2006, pp. 989–998.
[17] Q. Yang, L. Wang, N. Ahuja, A constant-space belief propagation algorithm for stereo matching, in: IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 1458–1465.
[18] M. Humenberger, T. Engelke, W. Kubinger, A census-based stereo vision algorithm using modified semi-global matching and plane fitting to improve matching quality, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2010, pp. 77–84.
[19] M. Michael, J. Salmen, J. Stallkamp, M. Schlipsing, Real-time stereo vision: optimizing semi-global matching, in: IEEE Intelligent Vehicles Symposium, 2013, pp. 1197–1202.
[20] T. Liu, P. Zhang, L. Luo, Dense stereo correspondence with contrast context histogram, segmentation-based two-pass aggregation and occlusion handling, in: T. Wada, F. Huang, S. Lin (Eds.), Advances in Image and Video Technology, Springer, Berlin, Heidelberg, 2009, pp. 449–461.
[21] R. Ben-Ari, N. Sochen, Stereo matching with Mumford-shah regularization and occlusion handling, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 2071–2084.
[22] W.-S. Jang, Y.-S. Ho, Efficient disparity map estimation using occlusion handling for various 3D multimedia applications, IEEE Trans. Consum. Electron. 57 (2011) 1937–1943.
[23] J. Ram Rajesh, D. Nagarjunan, R.M. Arunachalam, R. Aarthi, Distance transform based hand gestures recognition for power point presentation navigation, Adv. Comput.: Int. J. 3 (2012) 41–48.
[24] B. Stenger, A. Thayananthan, P.H.S. Torr, R. Cipolla, Model-based hand tracking using a hierarchical Bayesian filter, IEEE Trans. Pattern Anal. Mach. Intell. 28 (2006) 1372–1384.
[25] J. Canny, A computational approach to edge detection, IEEE Trans. Pattern Anal. Mach. Intell. PAMI-8 (1986) 679–698.
[26] S.-Y. Kim, M. Kim, Y.-S. Ho, Depth image filter for mixed and noisy pixel removal in RGB-D camera systems, IEEE Trans. Consum. Electron. 59 (2013) 681–689.
[27] W.-S. Jang, Y.-S. Ho, Disparity map acquisition with occlusion handling using warping constraint, in: IEEE International Symposium on Circuits and Systems, 2012, pp. 600–603.
[28] Y.-S. Ho, W.-S. Jang, Occlusion Detection Using Warping and Cross-Checking Constraints for Stereo Matching, The Era of Interactive Media, Springer, New York, 2013. pp. 363–372.
[29] D. Scharstein, R. Szeliski, R. Zabih, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, in: IEEE Workshop on Stereo and Multi-Baseline Vision, 2001, pp. 131–140.