J. Vis. Commun. Image R. 40 (2016) 118–127

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Disparity map enhancement in pixel based stereo matching method using distance transform ☆

Yong-Jun Chang, 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 6 February 2016
Revised 24 May 2016
Accepted 20 June 2016
Available online 21 June 2016

Keywords:
Stereo matching
Disparity map
Distance transform
Disparity error detection
Disparity error correction

ABSTRACT

With the great success in three-dimensional (3D) movies, a lot of 3D content have been generated. Depth information is one of the important elements in 3D content generation. Stereo matching methods obtain depth information using the characteristic of binocular disparity. These methods find corresponding points between two images which have different viewpoints to calculate the disparity value. However, these methods have difficulties computing accurate disparity values in the textureless region. Smeared pixels near the edge region also make difficult for the stereo matching. In this paper, we propose a pixel based cost computation for the cross-scale stereo matching using the distance transform to improve these problems. In addition, the disparity error detection and correction methods are also proposed as a post-processing step. As a result, we obtain the enhanced disparity map which is robust to the textureless region and the edge region.

1. Introduction

In these days, 3D content is used in various fields such as 3D films, medical images, and 3D games. Depth information is an important requisite for 3D content generation. There are several ways to acquire depth information from the target image. Depth measurement using a depth camera is one of the methods to get the depth value from the object [1]. This method uses infrared rays to measure the distance between the depth camera and the object. Therefore, it can acquire the depth value of the object quickly and accurately. However, the depth camera is vulnerable to the outside because of sunlight. It has an effect on infrared rays of the depth camera.

On the other hand, a depth estimation from captured scenes is not restricted to the place. One of the ways to estimate the depth value from captured scenes is using stereo images. Most of 3D movies use stereo images to generate a 3D effect. These images are acquired by a stereo camera which captures scenes having two different viewpoints. Both images have same objects each other. Each object in stereo images has a disparity value. This value is determined by the distance between the camera and the object. If the object is located near the camera, it has a large disparity value. If the object is far from the camera, it has a small disparity value. For this reason, stereo images allow people to feel the 3D effect.

Stereo matching methods are typical ways to get depth information from stereo images. These methods acquire the disparity value of each pixel in both images. The disparity value is calculated by two corresponding points in stereo images. In order to find the disparity value easily, the image rectification is applied to captured images as a pre-processing algorithm [2]. Therefore, two corresponding points in both rectified images are searched in the same scan line by the epipolar geometry. The result of stereo matching methods is represented as a disparity map. There are two kinds of stereo matching methods. One is a local method and the other one is a global method.

The local method calculates a matching cost of each pixel in stereo images to estimate the optimal disparity value. The matching cost is computed by using similarity measures such as sum of absolute differences (SAD), sum of squared differences (SSD), and normalized cross correlation (NCC). This method considers a limited number of pixels in a specific region to acquire the disparity value of one pixel. Therefore, the local method generally has fast matching results. However, it usually has lower disparity accuracy in the disparity estimation than that of the global method.

On the other hand, the global method considers whole pixels in the image to determine the disparity value of one pixel. It uses an energy function which is based on Markov random field (MRF) to compute the energy between two corresponding points. The

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author.
E-mail addresses: yjchang@gist.ac.kr (Y.-J. Chang), hoyo@gist.ac.kr (Y.-S. Ho).
URLs: http://vclab.gist.ac.kr (Y.-J. Chang), http://vclab.gist.ac.kr (Y.-S. Ho).

http://dx.doi.org/10.1016/j.jvcir.2016.06.017
1047-3203/

energy function is composed of a data term and a smoothness term. The data term calculates the correlation between two corresponding points using similarity measures. This term is similar to the matching cost computation in the local method. The smoothness term checks the disparity consistency among neighboring pixels. The energy function is optimized by several optimization concepts such as belief propagation and graph cuts [3,4]. The global method generally estimates more accurate disparity values than the local method. However, this method is generally slower than the local method.

Both stereo matching methods have problems in the textureless region. Since this region does not have any textures, it is very difficult to find corresponding points in stereo images. For this reason, the stereo matching in this region is problematic. Even though the global method estimates more accurate disparity values than the local method in the textureless region, these problems still remain to be solved as homework yet.

In this paper, we propose a pixel based cost computation using the distance transform to improve the accuracy of disparity values in the textureless region. The distance transform gives the distance value from the edge region to the pixel [5]. Therefore, pixels in the textureless region have specific values by using this transform. This transform also gives a large weighting on the pixel in the edge region. Thus, this transform helps to estimate more accurate disparity value in both regions. Matching costs of the proposed method are aggregated by a cross-scale cost aggregation method [6]. As a result, we acquire an initial disparity map. In addition to this method, we also apply a disparity error correction algorithm to remove remaining disparity errors in the initial disparity map.

This paper is organized as follows. In Section 2, we introduce a disparity accuracy problem in the stereo matching. We also introduce a conventional method which is relevant to this problem. In Section 3, we explain a problem of the conventional algorithm and show a proposed method which has better experiment results than the conventional algorithm. After that, the experiment results are analyzed in Section 4 and we conclude this paper in Section 5.

2. Problem statement

2.1. Disparity accuracy problem in textureless region

In stereo matching methods, a feature detection is an essential step to estimate disparity values between two corresponding points in stereo images. The result of feature detection is affected by the characteristic of regions in the image. The textureless region in the captured scene does not have any features. Therefore, it causes matching ambiguities in the stereo matching. Fig. 1(a) shows the textureless region in the image. Fig. 1(b) shows disparity errors in that region.

In the local method, there is a pixel based stereo matching method to search the matching pixel. This matching method generally has the matching ambiguity problem. Since this method checks the pixel similarity in stereo images using only one pixel, a lot of similar pixels may be existed in the same scan line. For this reason, it is very weak for the textureless region and even some textured regions. Thus, the pixel based matching method generates noises in the disparity map. Fig. 2 shows the matching ambiguity problem of the pixel based matching.

In order to avoid this problem, the local method generally uses a window based matching. This matching method uses the window to find the corresponding pixel in stereo images. It checks the similarity of all pixels in the window. Hence, it usually finds more accurate disparity values in most regions than the pixel based matching method. The result of this method has different qualities depending on the window size. The larger window size is used, the more accurate disparity values in the textureless region are estimated. However, using the large sized window causes inaccurate discontinuity depth values in the edge region.

On the contrary, the global method has great matching results in many regions including the textureless region. However, this method sometimes has a smudged effect near the edge region because of the smoothness term. The smoothness term checks the disparity consistency among neighboring pixels. If the disparity value of current pixel has a large difference with that of neighboring pixel, then this term gives a penalty to the energy function to avoid choosing this disparity value.

2.2. Relevant work

The cross-scale cost aggregation method was proposed by Zheng et al. to improve the disparity accuracy in the textureless region [6]. This method uses multi-scale images to aggregate matching costs among different scale images. It bases on a coarse-to-fine (CTF) strategy [7]. In the process of stereo matching, there are cost noises in the cost computation result. All regions in stereo images have cost noises because of a mismatching problem. Especially, the textureless region has a lot of cost noises. These cost noises lead to estimate inaccurate disparity values in the disparity map.

The low-scale image usually has less cost noises than those of the large-scale image. Since the low-scale image has a lower resolution than the large-scale image, there are few pixel candidates which are used to the similarity measure. Therefore, the low-scale image is more likely to search exact corresponding points in the textureless region than the large-scale image. The work of Zheng et al. uses this characteristic to relive mismatching problem in that region.

In Zheng's method, the consistency checking is used to aggregate refined matching costs [6]. There are two types of consistencies. First one is an intra-scale cost consistency and the other one is a cross-scale cost consistency. The intra-scale cost consistency compares the matching cost of the current pixel with that of neighboring pixels to reduce the cost noise. In this consistency checking step, the least square optimization is used to find the optimal

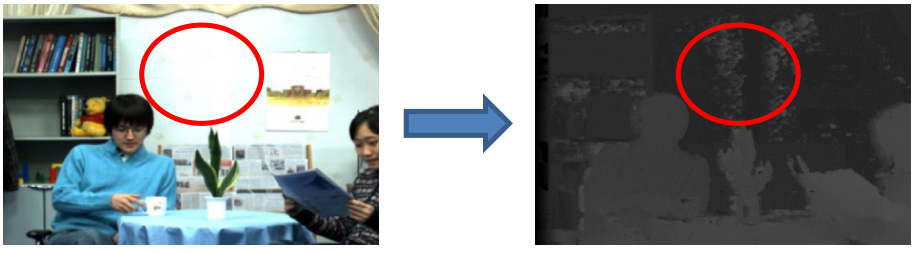

Fig. 1. Stereo matching in the textureless region. (a) The textureless region in the captured scene. (b) Disparity errors in the textureless region.

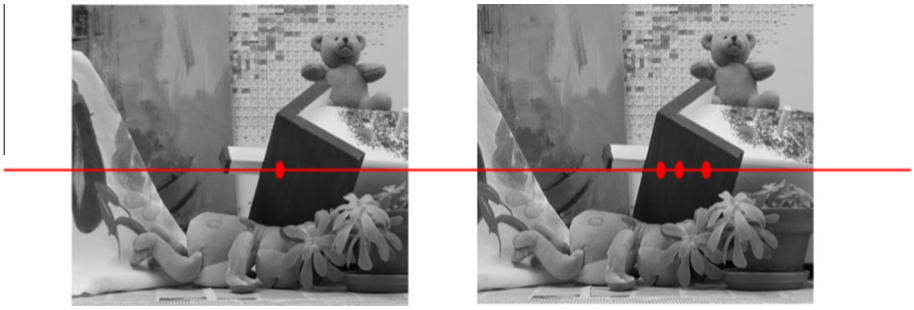

Fig. 2. Matching ambiguity problem of the pixel based matching.

refined cost. Based on the intra-scale cost consistency step, matching costs among different scale images are also refined using the cross-scale cost consistency. Both cost consistency checking steps change the cost to the refined cost. More detail things of this method are explained in Section 3.

In the cross-scale cost aggregation method [6], many ways of the cost computation and aggregation methods are used [8–10]. In terms of the cost computation, Zheng et al. used a pixel based cost function as an equation of the initial matching. This cost function measures the pixel similarity using color information and gradient information [9]. The initial matching cost C is formulated as

$$C(i,d) = (1-\alpha)\cdot \min(\|I_L(x_i,y_i) - I_R(x_i-d,y_i)\|,\tau_1) + \alpha \cdot \min(\|\nabla_x I_L(x_i,y_i) - \nabla_x I_R(x_i-d,y_i)\|,\tau_2), \quad (1)$$

where i is a position of the current pixel and d is a disparity candidate. I and $\nabla_x I$ are color information and gradient information, respectively. The gradient value is calculated in the x direction. In order to control the maximum cost value, τ_1 and τ_2 are used. The ratio between color information and gradient information is determined by a weighting value α.

3. Proposed method

3.1. Problems of conventional method

The conventional cross-scale cost aggregation method using the pixel based cost computation [6] has two problems: the matching ambiguity problem and the edge preserving problem. The conventional method has quite accurate disparity values in the textureless region. However, the matching ambiguity problem in the conventional method still exists because of the equation of initial matching. The initial cost function in the conventional method bases on the pixel-wise matching. This function computes the matching cost quickly and simply. It is also robust to the radiometric variation of stereo images. However, this cost function still has mismatching problem in the textureless region.

The difficulty of the stereo matching in the textureless region of color image is a well-known problem. The gradient image as well as the color image also have matching ambiguity in the same region. Fig. 3 shows the matching ambiguity problem in the gradient image. In Fig. 3, the corresponding point of the current pixel should be the pixel in a circle according to the similarity measure. However, that pixel is not the real corresponding pixel. This problem can cause matching errors in the textureless region. Fig. 4 represents the disparity map of the conventional method which uses the pixel based cost function for the cross-scale cost aggregation. In Fig. 4, there are disparity errors in the textureless region. Therefore, the pixel based matching using both information still has room for improvement in the textureless region.

The other problem of the conventional method is the edge preserving problem. The cross-scale cost aggregation for the stereo matching [6] uses a lot of images which have different scale levels. Since this method reduces the resolution of the image to estimate accurate disparity values in the textureless regions, it is sometimes difficult to preserve textured and edge regions when the image resolution is too small. Fig. 5 shows disparity errors in textured regions.

In order to improve these problems, we propose a new term for the pixel based cost computation using the distance transform. In addition, we also propose a disparity error detection and correction method to complement remaining disparity errors in occlusion regions or other regions. The overall scheme of the proposed method is depicted in Fig. 6.

First, the initial matching cost is calculated using the distance transform. After that, matching costs are aggregated by the cross-scale cost aggregation method to acquire initial disparity maps [6]. In order to enhance initial disparity maps, we also apply the disparity error correction step as a post-processing algorithm. Finally, we acquire optimal disparity maps.

3.2. Initial cost computation using distance transform

The matching ambiguity problem in the textureless region is caused by many similar pixels. They have no distinguishable features each other. A motivation of the proposed method is inspired by giving specific values to pixels in the textureless region. The distance transform [5] calculates the pixel distance from the edge region. Therefore, each pixel in the textureless region can have the specific distance value by using this transform. A distance transformed map (DT map) is the result of the distance transform. In order to acquire the DT map, the color image is transformed to an edge image using the edge detection. After that, a kernel of the distance transform is applied to the edge image. The kernel equation is defined by

$$r_{i,j}^{k} = \min \begin{bmatrix} r_{i-1,j-1}^{k-1} + \beta & r_{i,j-1}^{k-1} + \alpha & r_{i+1,j-1}^{k-1} + \beta \\ r_{i-1,j}^{k-1} + \alpha & r_{i,j}^{k-1} & r_{i+1,j}^{k-1} + \alpha \\ r_{i-1,j+1}^{k-1} + \beta & r_{i,j+1}^{k-1} + \alpha & r_{i+1,j+1}^{k-1} + \beta \end{bmatrix}, \quad (2)$$

where $r_{i,j}^{k}$ represents a DT value at iteration k. The minimum DT value in the previous iteration step is determined as the current DT value. In order to control the strength of DT value, α and β are used as weighting parameters. Fig. 7 shows the process of distance transform. In Fig. 7(a), there are edge pixels which have zero values. Fig. 7(b) is a transformed result of the first iteration. In Fig. 7(b), pixels are close to the edge region have small DT values. On the contrary, if pixels are far from the edge, then they have large DT values.

Jang et al. proposed the edge preserving method using the distance transform [11]. This method uses DT values in the DT map as weighting scales of the matching cost function. It gives the smallest weighting scale to the edge region and gives the large weighting scale to non-edge regions. Therefore, this method

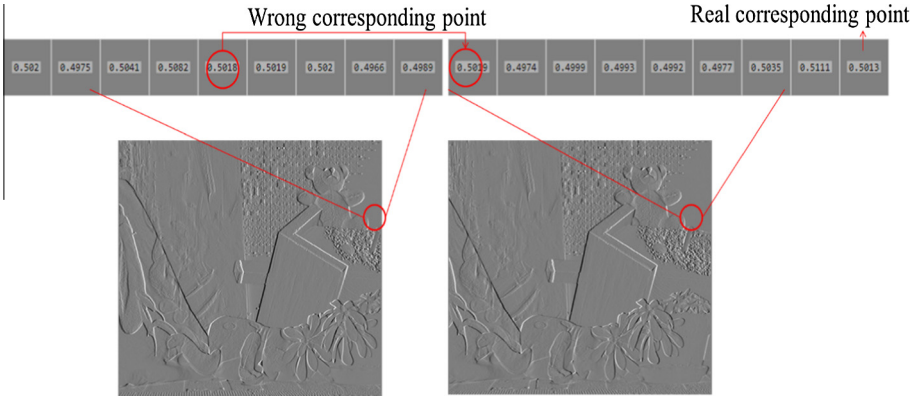

Fig. 3. Matching ambiguity problem in gradient images.

Fig. 4. Disparity map of the conventional method.

Fig. 5. Disparity errors in the texture region.

preserves the edge region in the disparity map by emphasizing non-edge regions.

In terms of our proposed method, we use the DT value as a new cost term in the pixel based cost function to improve the matching accuracy in the textureless region. Pixels of color and gradient images have similar pixel values in the textureless region. However, DT values in that region have different pixel values depending on the distance from the edge region. Hence, these values can help to estimate more accurate disparity values in the textureless region. The distance transform is also useful to the edge preserving problem of the conventional method. Since the DT map bases on the edge image, DT values help the stereo matching in the low scale image to find the accurate corresponding point near the edge boundary. Therefore, using the DT value in the stereo matching has a better edge preserving in the disparity map than that of the conventional method.

In order to use the DT value as the new cost term, the kernel of the distance transform is modified as

$$r_{ij}^{k} = \min\left[r_{i-1j}^{k-1} + \alpha \quad r_{ij}^{k-1} \quad r_{i+1j}^{k-1} + \alpha \right], \tag{3}$$

where α controls the strength of DT value. In (3), the DT value is calculated in the x-direction. The conventional distance transform considers eight different directions to calculate the DT value. However, stereo matching methods usually use rectified images. These images make possible to search the corresponding point in the same scan line. Therefore, it is enough to compute the DT value in the x direction. Fig. 8 shows the result of the modified DT. We also apply the Canny edge detection to acquire the edge image [12].

We define a DT term based on the modified DT map. The DT term is added to the conventional pixel based cost function. The conventional cost function means the equation using color

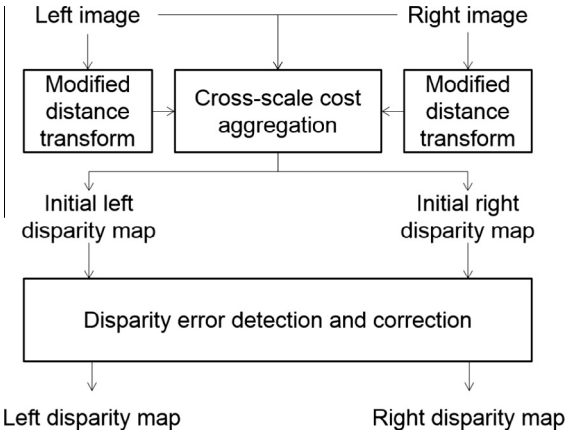

Fig. 6. The overall scheme of the proposed method.

information and gradient information [9]. A new cost function C' is represented by

$$\begin{aligned} C'(i,d) = {} & \alpha \cdot \min(\|I_L(x_i,y_i) - I_R(x_i-d,y_i)\|, \tau_1) \\ & + \beta \cdot \min(\|\nabla_x I_L(x_i,y_i) - \nabla_x I_R(x_i-d,y_i)\|, \tau_2) \\ & + \gamma \cdot \|dt_L(x_i,y_i) - dt_R(x_i-d,y_i)\|, \end{aligned} \tag{4}$$

where dt means the DT value in the modified DT map. α, β, and γ represent weighting values. We complement the matching ambiguity problem of the conventional cost function by adding the DT term. In this paper, we use this equation as an initial matching cost function.

A pixel based cost function using the distance transform was already by Chang et al. [13]. They proposed the same cost function which uses the DT value as the third cost term. In (4), Chang's method used constant weighting values to compute the matching cost. However, two corresponding points in left and right DT maps sometimes do not have same DT values. This problem is caused by two elements: the occlusion region, and the slanted object. Occlusion regions in stereo images have different edge detection results each other. Therefore, these results have an effect on both DT maps to have different DT values.

The slanted object in the captured scene also generates different DT values in both DT maps. Even though there are same slanted objects in stereo images, disparity values inside these objects are different because of the characteristic of binocular disparity. However, the distance transform calculates DT values inside these objects regardless of this characteristic. Therefore, these elements reduce effects of the DT term in the matching result. For this reason, we propose using adaptive weighting values instead of using constant weighting values. Adaptive weighting values are formulated as

$$\begin{aligned} \alpha &= \left(1 - \lambda_2 \cdot e^{-\frac{dt}{\sigma^2}}\right) \cdot (1 - \lambda_1), \\ \beta &= \left(1 - \lambda_2 \cdot e^{-\frac{dt}{\sigma^2}}\right) \cdot \lambda_1, \\ \gamma &= \lambda_2 \cdot e^{-\frac{dt}{\sigma^2}}, \end{aligned} \tag{5}$$

where λ_1 and λ_2 are regulation parameters. In (5), where dt represents the DT value in the DT map. Therefore, we can give the large weighting value to the DT term, if the current pixel is near the edge region. These adaptive weighting values restrain the discordance problem of DT values between stereo DT maps.

Distance information can also be used in motion and flow detection [27,28]. Since this information uses the distance transform, pixels near the edge region can be weighted. For this reason, if both color and distance information is used for the matching cost, then we can find the flow or the motion of objects easily. However, if motion and flow are too large, the distance transform may give certain incorrect information about both matching tasks. Therefore, the distance information near the edge region should be used for the matching.

3.3. Cross-scale cost aggregation

In order to aggregate initial matching costs, we apply the cross-scale cost aggregation method [6]. This method checks the intra-scale cost consistency and the cross-scale cost consistency to refine cost noises. The refined matching cost of the intra-scale is obtained by using a weighted least square (WLS) optimization. It is defined by

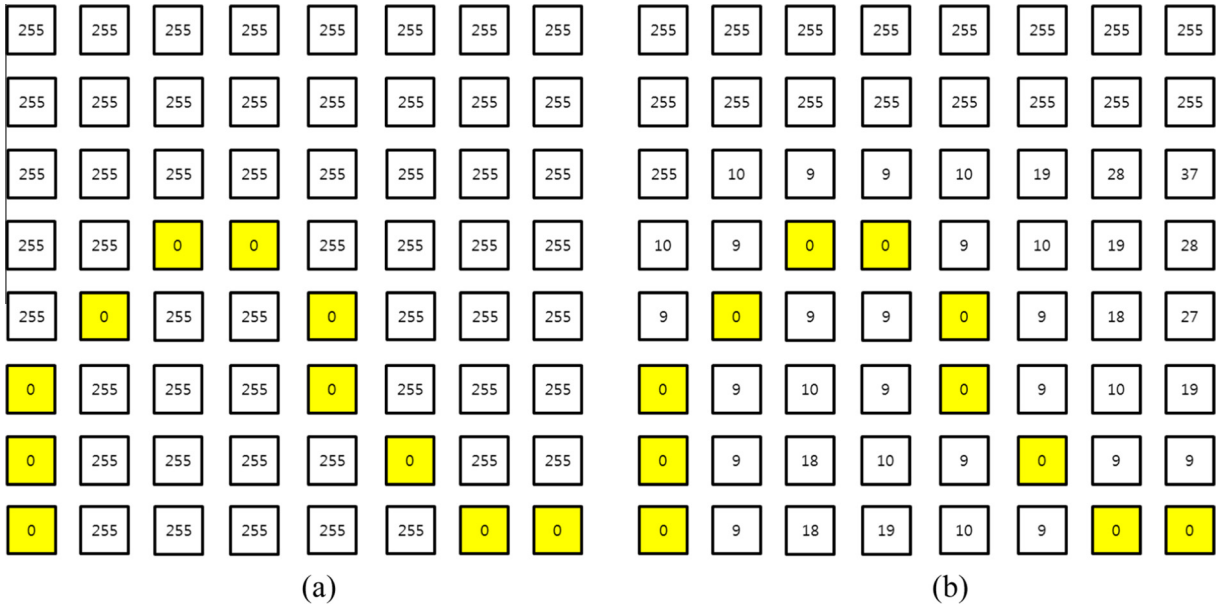

Fig. 7. Process of the distance transform. (a) Initialization. (b) $k = 1$.

Fig. 8. Result image of modified DT. (a) Left image. (b) Right image.

$$\widetilde{C}'(i,d) = \arg\min_{s} \frac{1}{Z_i} \sum_{j \in N_i} K(i,j) \| s - C'(j,d) \|^2, \tag{6}$$

where $\widetilde{C}'$ is the refined matching cost and N_i is a set of neighboring pixels in the refinement kernel K. In this paper, we use the kernel of bilateral filter to refine cost noises [14]. In (6), where Z_i is a sum of weighting values in the kernel. This equation is solved by using the partial differential. The solution of this equation is represented by

$$\widetilde{C}'(i,d) = \frac{1}{Z_i} \sum_{j \in N_i} K(i,j) C'(j,d), \tag{7}$$

In (7), the refined matching cost is calculated in a single-scale image. This cost function is extended for the multi-scale images. A vector $\tilde{v}$ which represents the set of the refined matching costs in multi-scale levels is defined by

$$\tilde{v} = \arg\min_{\{s^l\}_{l=0}^{L}} \sum_{l=0}^{L} \frac{1}{Z_{i^k}^k} \sum_{j^k \in N_{ik}} K(i^l, j^l) \| s^l - C^l(j^l, d^l) \|^2, \tag{8}$$

where l represents a scale level and L is the maximum scale level. This vector set contains refined matching costs which are checked using the intra-scale cost consistency at each scale level. The refined matching cost of vector $\tilde{v}$ is solved using the same optimization way with (6). It is formulated as (9).

$$\forall_l, \quad \widetilde{C}'^l(i^l, d^l) = \frac{1}{Z_{i^l}^l} \sum_{j^l \in N_{il}} K(i^l, j^l) C'^l(j^l, d^l), \tag{9}$$

Refined matching costs in (8) and (9) are results of the intra-scale cost consistency. In order to check the cost consistency among different scale images, a vector $\hat{v}$ is formulated as

$$\hat{v} = \arg\min_{\{s^l\}_{l=0}^{L}} \left(\begin{array}{c} \sum_{l=0}^{L} \frac{1}{Z_{i^l}^l} \sum_{j^l \in N_{il}} K(i^l, j^l) \| s^l - C^l(j^l, d^l) \|^2 \\ + \lambda \sum_{l=1}^{L} \| s^l - s^{l-1} \|^2 \end{array} \right), \tag{10}$$

where λ is a constant parameter which controls the strength of consistency checking. The vector $\hat{v}$ is composed of refined matching costs which are applied the cross-scale cost consistency. Elements of this vector can be obtained using the same solution with (8).

From the solution of (10), we can induce the relationship between $\tilde{v}$ and $\hat{v}$. This relationship is represented by

$$A\hat{v} = \tilde{v}, \tag{11}$$

where A is a $(L+1) \times (L+1)$ tridiagonal matrix. This matrix is composed of the constant parameter λ. Therefore, the final refined matching cost in the finest scale level is defined as follows.

$$\widehat{C}^0(i^0, d^0) = \sum_{l=0}^{L} A^{-1}(0,l) \widetilde{C}^l(i^l, d^l), \tag{12}$$

This cost aggregation method estimates the optimal disparity value of each pixel in the finest scale level by minimizing the matching cost in (12).

3.4. Disparity error detection and correction

Stereo matching methods have an occlusion problem because of different viewpoints in stereo images. Occluded pixels in that region cannot find their corresponding points in stereo images. This problem causes disparity errors in the disparity map. Not only the occlusion problem but also some mismatching problems generate disparity errors in the result image. In order to detect disparity errors in the disparity map, the cross checking method was proposed [15]. This method checks the accordance of disparity values between two corresponding points in stereo images. If two corresponding points have different disparity values, then these two pixels are regarded as error pixels.

The conventional cross-checking method uses two disparity maps to detect error pixels. This method finds disparity errors simply and quickly. However, it sometimes has a low accuracy in the disparity error detection because this method uses just two disparity maps of stereo images. If both corresponding points have same wrong disparity values, the cross checking method cannot detect these pixels as error pixels. In order to prevent this problem, we use two error detection steps. First, color images as well as disparity maps are used to detect error pixels [16]. In this process, we acquire an initial error map. The error map is a binary image and it has zero values in the error region. Second, we check the error consistency between the current pixel and neighboring pixels in the initial error map to detect remaining error pixels. Therefore, we can detect more accurate disparity errors than the conventional cross-checking method.

In order to acquire the initial error map, we calculate the matching error of color pixels after the cross checking method. If the current pixel does not have the same disparity value with that of the corresponding point, we define that pixel as an initial error pixel. If not, that pixel is checked again using the matching error. The matching error is computed with color pixels. The matching error M between two color pixels is formulated as

$$M(I_L, I_R(d)) = \frac{\frac{1}{3}\sum_{I\in R,G,B}|I_L - I_R(d)|}{255}, \tag{13}$$

where $I_R(d)$ represent RGB color values which correspond to color values I_L in the different viewpoint. If two corresponding points of color images are similar each other, then the matching error closes to zero. If the matching error of pixel is larger than the threshold, then we determine that pixel as an initial error pixel.

We also apply the error consistency check to the initial error map in order to detect remaining disparity errors [16]. This method calculates the smoothness cost to remove outliers in the error map. The smoothness cost S is represented by

$$S(i,j) = \sum_{j\in N_i} f(j)o(i), \tag{14}$$

where $o(i)$ determines whether the current pixel i is the error pixel or not. $f(j)$ counts the number of error pixels around the current pixel. The smoothness cost is computed using the initial error map. Equations of $f(j)$ and $o(i)$ are defined by

$$f(j) = \begin{cases} 1, & \text{if } I_{E,j} = 0 \\ 0, & \text{otherwise} \end{cases}, \quad o(i) = \begin{cases} 1, & \text{if } I_{E,i} = 1 \\ 0, & \text{otherwise} \end{cases}, \tag{15}$$

where $I_{E,i}$ and $I_{E,j}$ represent pixel values of i and j in the error map, respectively. Therefore, the smoothness cost which is defined by (14) and (15) counts the number of error pixels around the current non-error pixel in the initial error map. If the smoothness cost is larger than the threshold value, then it means that the current non-error pixel is more likely to be the error pixel. For this reason, we change this non-error pixel to the error pixel.

The final error map can be obtained using above steps. Fig. 9 describes results of the disparity error detection. Fig. 9(a) is results of the conventional cross checking method. Fig. 9(b) is results of the proposed error detection method. As you can see in these figures, our method detects more error pixels than the conventional method.

We change detected error pixels in the initial disparity map to holes which have zero pixel values. After that, we fill those holes using the disparity error correction method. The proposed disparity error correction method is inspired by Min et al. [17]. They applied an iterative support-and-decision process using probability functions to fill holes in the disparity map. In terms of our proposed method, we correct error pixels without the iterative process. The proposed method finds the most probable disparity value near the hole.

In order to find the optimal disparity value of the error pixel, we calculate the hole filling cost H. This cost function is formulated as

$$H(i,j) = k(i,j)o(j), \tag{16}$$

where k is the probability function using the distance difference and the color difference. The hole filling cost is calculated using the window based method. Therefore, neighboring pixels around the current error pixel i are used to compute the hole filling cost. The probability function k is defined by

$$k(i,j) = \exp\left(-\frac{D(i,j)}{\sigma_D^2} - \frac{D(I_i, I_j)}{\sigma_I^2}\right), \tag{17}$$

where $D(i,j)$ and $D(I_i, I_j)$ are the Euclidean distance of pixel positions and that of color values, respectively. In (17), the probability function uses the window kernel of bilateral filter [14].

The optimal disparity value is chosen by the neighboring pixel in the window. If the neighboring pixel j has the maximum hole filling cost, then the current hole will be filled by the disparity value of that neighboring pixel. Therefore, we can correct disparity errors using (18) which is formulated as

$$d_i = \arg\max_{d_j} H(i,j), \tag{18}$$

where d_i and d_j are disparity values of pixel i and j, respectively. The error pixel is filled by the neighboring disparity value d_j.

4. Experimental results

The proposed method was tested using 4 different images: *Teddy*, *Cones*, *Tsukuba*, and *Venus* [18]. We also tested 6 test images in 'Middlebury 2005' and 21 test images in 'Middlebury 2006' [19,20]. In addition, the KITTI dataset including 10 training images was used for the implementation [21]. In order to implement the proposed method, we set parameters first. In (3), we set the control parameter α to 1. If this parameter is too big, then DT values cannot

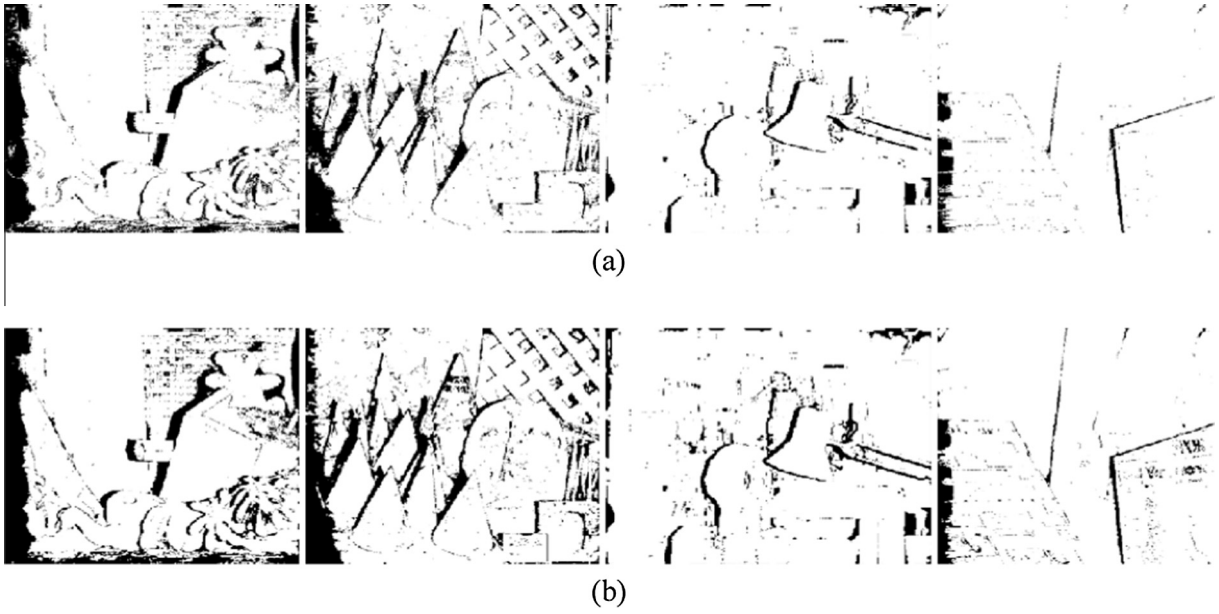

Fig. 9. Error map comparison between the conventional cross checking method and the proposed disparity error detection method. (a) Cross checking method. (b) Proposed disparity error detection.

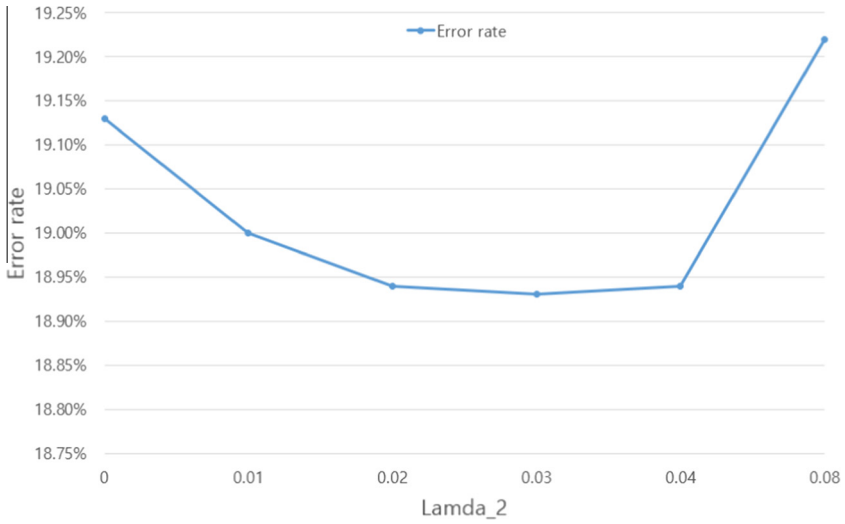

Fig. 10. Error rate of changes depending on λ_2.

fill the textureless region. For this reason, we set small value to the parameter. In (5), there are some parameters to define the initial cost function. We set λ_1 and σ to 0.89 and 0.1, respectively. In order to find the optimal λ_2, we checked error changes depending on the value of λ_2. Fig. 10 shows the error rate of changes for 31 Middlebury datasets. In Fig. 10, the error rate is the lowest value when λ_2 is set to 0.03.

In terms of the conventional pixel based cost function, we set α to 0.89. In the process of cross-scale cost aggregation, we set the constant parameter in (1) to 0.3 for Middlebury datasets and 1.0 for KITTI datasets. We also set the maximum scale level L to 5. In the disparity detection step, the threshold value of the matching error M is set to 0.9 and that of the smoothness cost is set to 4. We make the smoothness cost check eight neighboring pixels around the current pixel. In order to correct disparity errors, σ_D and σ_I in (17) are set to 17.5 and 100, respectively. Based on these parameters, we tested the proposed method.

Fig. 11 shows experiment results for four test images (*Teddy*, *Cones*, *Tsukuba*, *Venus*). In Fig. 11, the order of the pictures from top to bottom is as follows: *Teddy*, *Cones*, *Tsukuba*, and *Venus*. In Fig. 11(d), the results images show enhanced disparity values than those of Fig. 11(b). However, it is difficult to distinguish differences

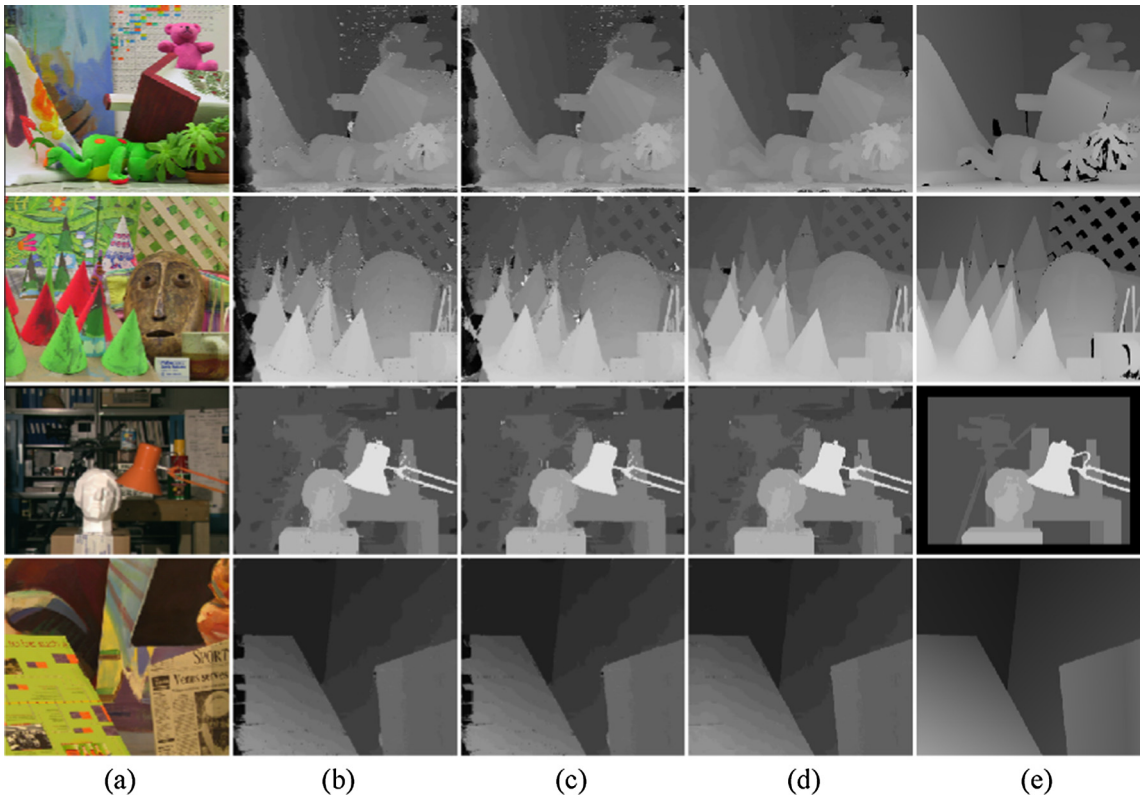

Fig. 11. Result images of the conventional method and the proposed method. (a) Original image. (b) Results of the conventional method. (c) Results of the proposed method without the disparity error correction. (d) Results of the proposed method with the disparity error correction. (e) Ground truth.

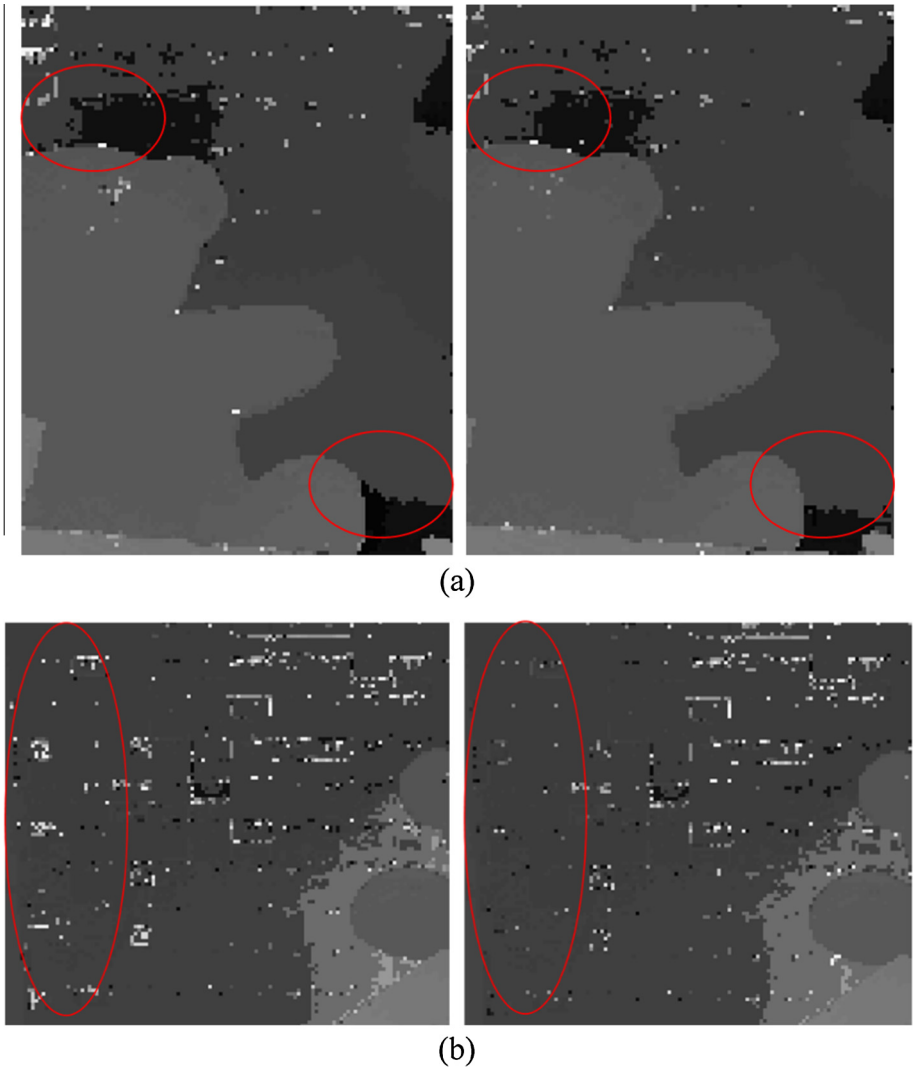

Fig. 12. Enlarged result images. (a) Disparity error reduction in the textureless region. (b) Disparity error reduction in the textured region.

Table 1
Error rate comparison of the Middlebury datasets in all regions, non-occlusion regions, and discontinuity regions. The number in this table represents the percentage of bad pixels.

Algorithm		ANCC [22]	Census [23]	Intensity + gradient [6,9]	Proposed method (without error correction)	Proposed method (with error correction)
MI-31	Nonocc.	19.58	11.99	11.81	**11.72**	**10.79**
	All	28.71	22.23	21.38	**21.12**	**16.62**
	Disc.	33.12	26.08	24.21	**23.94**	**22.9**

Table 2
Error rate comparison of the KITTI datasets in all regions, non-occlusion regions, and discontinuity regions. The number in this table represents the percentage of bad pixels.

Algorithm		ANCC [22]	Census [23]	Intensity + gradient [6,9]	Proposed method (without error correction)	Proposed method (with error correction)
K-10	Nonocc.	42.87	11.43	21.37	**20.59**	**20.07**
	All	43.98	13.15	22.85	**22.05**	**20.95**

between Fig. 11(b) and (c). Thus, enlarged images of *Teddy* are depicted in Fig. 12. In Fig. 12, we can check that the proposed method estimates disparity values in the textureless region more accurately than the conventional method. Fig. 12(b) shows enlarged disparity map in the textured region. The result of proposed method also has a better matching result in the textured region than the conventional method.

We measured the bad pixel rate (BPR) of result images which were applied different algorithms including the proposed method. The BPR checks the error rate of the disparity map. In terms of the Middlebury datasets, if the disparity difference between the pixel of the result image and that of the ground truth is larger than 1, then the pixel in the result image is regarded as an error pixel. On the other hands, we set the error threshold to 3 for the KITTI datasets and calculate the error rate using our evaluation program. Therefore, the BPR makes us compare matching results objectively. Table 1 shows the BPR comparison of the Middlebury datasets between the conventional method and the proposed method. In this table, MI-31 and K-10 represent the results of 31 Middlebury datasets and 10 KITTI datasets, respectively. We measured the BPR in all regions, non-occlusion regions, and discontinuity regions. In order to compare our results with other cost computation algorithms, ANCC [22], Census [23], intensity + gradient [9] cost functions are used. The proposed method in this table is not

Table 3
Error rate comparison with other stereo matching algorithms. The proposed method is compared with other stereo matching methods.

Algorithm		AdaptAggrDP [24]	BitPlaneNLF [25]	LCVB-DEM [26]	Conventional method [6,9]	Proposed method (with error correction)
Teddy	Nonocc.	6.79	8.3	9.99	8.71	**7.1**
	All	14.3	13.6	16.3	17.32	**13.81**
	Disc.	16.2	17.1	26.1	21.53	**20.03**
Cones	Nonocc.	5.53	3.68	6.56	6.37	**4.42**
	All	13.2	9.68	13.6	15.98	**12.08**
	Disc.	14.8	9.91	18.2	15.73	**11.99**
Tsukuba	Nonocc.	1.57	1.76	4.49	2.33	**2.3**
	All	3.5	2.33	5.23	2.67	**2.6**
	Disc.	8.27	8.83	21.3	9.71	**9.57**
Venus	Nonocc.	1.53	3.82	1.32	1.3	**0.43**
	All	2.69	4.16	1.67	3.19	**1.06**
	Disc.	12.4	5.65	11.5	4.11	**2.9**
Average		8.4	7.4	11.4	9.08	**7.36**

applied the error correction algorithm. In Table 1, the error rate of the proposed method has lower error rates compared with other algorithms [6,9,22,23].

Table 2 shows the error rate comparison result with the KITTI datasets. 10 stereo pairs were used for the experiment. We also implemented using other cost functions that are same as the experiment of the Middlebury datasets. In addition, we also compared our algorithm with other recently stereo matching algorithms in Table 3 [24–26]. In Table 3, four test images were used to compare the results with other algorithms. The number in Table 3 represents the percentage of bad pixels.

5. Conclusions

In this paper, we proposed a pixel based cost computation using the distance transform to improve disparity values in the textureless region and the edge region. We added the distance term to the conventional pixel based cost function which uses color information and gradient information. Matching costs are aggregated by the cross-scale cost aggregation method. In addition, we also used the disparity error correction method as a post-processing algorithm to enhance the final disparity map. As a result, experiment results indicated that the proposed method with the post-processing decreases error rates by 1.72% on average compared with the conventional method.

Acknowledgement

This research was supported by the 'Cross-Ministry Giga KOREA Project' of the Ministry of Science, ICT and Future Planning, Republic of Korea (ROK). [GK16C0100, Development of Interactive and Realistic Massive Giga-Content Technology].

References

[1] 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.
[2] Y.-S. Kang, Y.-S. Ho, An efficient image rectification method for parallel multi-camera arrangement, IEEE Trans. Consum. Electron. 57 (2011) 1041–1048.
[3] J. Sun, N.-N. Zheng, H.-Y. Shum, Stereo matching using belief propagation, IEEE Trans. Pattern Anal. Mach. Intell. 25 (2003) 787–800.
[4] 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.
[5] G. Borgefors, Distance transformations in digital images, Comput. Vision Graph. Image Process. 34 (1986) 344–371.
[6] K. Zheng, Y. Fang, D. Min, L. Sun, S. Yang, S. Yan, Q. Tian, Cross-scale cost aggregation for stereo matching, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1590–1597.
[7] M.D. Menz, R.D. Freeman, Stereoscopic depth processing in the visual cortex: a coarse-to-fine mechanism, Nat. Neurosci. 6 (2003) 59–65.
[8] X. Mei, X. Sun, W. Dong, H. Wang, X. Zhang, Segment-tree based cost aggregation for stereo matching, in: IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 313–320.
[9] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, M. Gelautz, Fast cost-volume filtering for visual correspondence and beyond, in: IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 3017–3024.
[10] Q. Yang, A non-local cost aggregation method for stereo matching, in: IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 1402–1409.
[11] W.-S. Jang, Y.-S. Ho, Discontinuity preserving disparity estimation with occlusion handling, J. Vis. Commun. Image Represent. 25 (2014) 1595–1603.
[12] J. Canny, A computational approach to edge detection, IEEE Trans. Pattern Anal. Mach. Intell. 8 (1986) 679–698.
[13] Y.-J. Chang, Y.-S. Ho, Pixel based cost computation using weighted distance information for cross-scale stereo matching, Electron. Imaging (2016) 1–6.
[14] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, in: IEEE Conference on Computer Vision, 1998, pp. 839–846.
[15] G. Egnal, R.P. Wildes, Detecting binocular half-occlusions: empirical comparisons of five approaches, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 1127–1133.
[16] Y.-J. Chang, Y.-S. Ho, Disparity error detection and correction in depth map using stereo images, in: Korean Institute of Smart Media Fall Conference, 2015, pp. 259–262.
[17] D. Min, S. Yea, A. Vetro, Occlusion handling based on support and decision, in: IEEE International Conference on Image Processing, 2010, pp. 1777–1780.
[18] 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.
[19] D. Scharstein, C. Pal, Learning conditional random fields for stereo, in: IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
[20] H. Hirschmuller, D. Scharstein, Evaluation of cost functions for stereo matching, in: IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
[21] A. Geiger, P. Lenz, R. Urtasun, Are we ready for autonomous driving? The KITTI vision benchmark suite, in: IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 3354–3361.
[22] Y.S. Heo, K.M. Lee, S.U. Lee, Illumination and camera invariant stereo matching, in: IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.
[23] R. Zabih, J. Woodfill, Non-parametric local transforms for computing visual correspondence, in: European Conference on Computer Vision, 1994, pp. 151–158.
[24] L. Wang, R. Yang, M. Gong, M. Liao, Real-time stereo using approximated joint bilateral filtering and dynamic programming, J. Real-Time Image Proc. 9 (2014) 447–461.
[25] C.-C. Kao, H.-Y. Lin, Stereo Matching Bit-plane Slicing, Technical Report, 2013, pp. 1–8.
[26] J. Martins, J. Rodrigues, H. du Buf, Luminance, colour, viewpoint and border enhanced disparity energy model, PLoS ONE 10 (2015) 1–24.
[27] L. Liu, W. Lin, Y. Zhong, Traffic flow matching with clique and triplet cues, in: IEEE International Workshop on Multimedia Signal Processing, 2015, pp. 199–207.
[28] W. Lin, Y. Mi, W. Wang, J. Wu, J. Wang, T. Mei, A diffusion and clustering-based approach for finding coherent motions and understanding crowd scenes, IEEE Trans. Image Process. 25 (2016) 1674–1687.