Discontinuity Preserving Depth Estimation Using Distance Transform

Woo-Seok Jang and Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)
123 Cheomdan-gwagiro, Buk-ku, Gwangju 500-712, Republic of Korea

ABSTRACT

Image interpolation methods at arbitrary view positions have become quite important due to the development of three-dimensional multi-view image display devices. Accurate depth information is required for natural image generation. Over the past several decades, a variety of stereo-image-based depth estimation methods have been developed to obtain high-quality depth data. However, obtaining accurate depth information still remains problematic due to difficult correspondence matching in image occlusion regions. In particular, for the discontinuous depth edge region, unclear color values exist, which lead to ineffective of corresponding matching. Thus, we propose a discontinuity preserving depth estimation method to solve the problem. The distance transform (DT) calculates the distance to the closest edge for each pixel of the input image. By controlling the color weighting term using DT values of stereo images, we carry out better correspondence matching in discontinuous regions. Experimental results indicate that the proposed method outperforms other methods. Visual comparison of the experimental results demonstrates that the proposed stereo-image-based depth estimation method improves the quality of the depth map in discontinuous edge regions.

Keywords: depth map, distance transform, stereo matching, depth discontinuity

1. INTRODUCTION

Depth images represent the 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, which is often called video-plus-depth [1]. Recently, with the appearance of multi-view display devices for three-dimensional (3D) image contents, it has been required to develop efficient algorithms to generate new images at arbitrary or designated view positions. Depth image-based rendering (DIBR) is a popular method to create a synthesized image by projecting color and depth data onto a target-view image plane [2]. The performance of DIBR mainly depends on quality of the depth information.

In order to measure depth information from a natural scene, we can use active sensor-based and passive sensor-based methods. Active sensor-based methods employ physical sensors, e.g., infrared ray (IR) sensor, to directly acquire depth data based on the principles of time-of-flight [3]. The active depth sensor-based method produces relatively higher quality depth images than the passive sensor-based method.

However, most active depth sensors suffer from three inherent problems; first, it is difficult to obtain depth data of the object far away from the sensor. Off-the-shelf depth sensors are only available to measure distances within 10*m*. Second, depth sensors are not applicable to outdoor environments. Third, active depth sensors provide low-resolution depth images, less than 640×480, due to many challenges in real-time distance measuring systems. Such inherent problems make it difficult to employ active depth sensors in practical applications. In the industry, use of active depth sensors are limited to applications mainly involving foreground extraction [4] and motion tracking [5] in indoor environments.

On the other hand, passive sensor-based methods indirectly calculate depth information from 2D images captured by stereoscopic video cameras. Although passive depth sensing produces relatively lower quality depth images, it can measure depth information of all objects in the captured scene, unlike active sensor-based methods. In addition, indirect depth sensing is applicable to both indoor and outdoor environments. Furthermore, the depth image resolution depends on the resolution of video cameras. Because of these advantages, advanced 3D video service systems using video-plus-depths employ passive depth sensing.

Among the passive sensor-based methods, stereo matching is widely used to estimate depth information of the scene. Stereo matching is the process to extract 3D information from left and right images captured by the stereoscopic camera

Stereoscopic Displays and Applications XXV, edited by Andrew J. Woods, Nicolas S. Holliman, Gregg E. Favalora,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 9011, 90111S · © 2014 SPIE-IS&T
CCC code: 0277-786X/14/$18 · doi: 10.1117/12.2038449

[6]. In stereo matching, 3D information, called disparity, is calculated by relative positions of objects in the two images of the scene. The disparity information can directly be converted to the depth information based on camera parameters.

The objective of this paper is to generate a more accurate depth map from stereoscopic images. Over the past several decades, a variety of stereo-image-based depth estimation methods have been developed to obtain high-quality depth maps. However, accurate measurement of depth information from natural scene still remains problematic due to difficult correspondence matching in some regions. Especially, in case of the depth discontinuous region, i.e., the edge region, unclear color values exist, which lead to ineffective of corresponding matching. Thus, in this work, we propose a discontinuity preserving depth estimation method to solve the problems. Distance transform (DT) calculate the distance to the closest edge for each pixel of an input image. By controlling the color weighting term using DT values of left and right images, we carry out better correspondence matching in edge regions.

2. WEAKNESS OF STEREO MATCHING

In general, stereo matching can be categorized into two approaches: local and global methods. Local methods are processed by windows based on correlation. The correlation assumed that the disparity is equal for all pixels of a correlation window. This assumption is violated at depth discontinuities. It results in blurred object borders and the removal of small detail depending on the size of the correlation window.

In contrast, global methods [7] define an energy function using Markov Random Field (MRF) and optimize this via several optimization algorithms such as belief propagation [8], dynamic programming [9] and graph cut [10], [11]. However, global methods are too computationally complex even for low resolution images and a small number of disparity levels. Thus, they are not suitable for practical use. Recently, several methods have been introduced to reduce the complexity of belief propagation [12, 13, 14]. When image size is N, the number of disparity levels is L, and the number of iterations is T, the computational complexity is originally $O(4TNL^2)=O(TNL^2)$ in standard belief propagation [8]. One of the fast belief propagation reduce the complexity to $O(TNL)$ by using hierarchical coarse-to-fine manner [12, 13]. This algorithm facilitates real-time computation if it uses GPU implementation. CSBP [14] is the fastest and takes the smallest memory among fast belief propagation algorithms. Its complexity depends on only constant space, that is $O(1)$.

Accurate measurement of depth information on the edge region is important in stereo matching, because depth data on the border between two objects are usually distinguishable. However, as shown in Figure 1, pixels around edges in the left and right images do not have clear values. These unclear pixels around edges make it difficult to measure discontinuous disparities on the associated area.

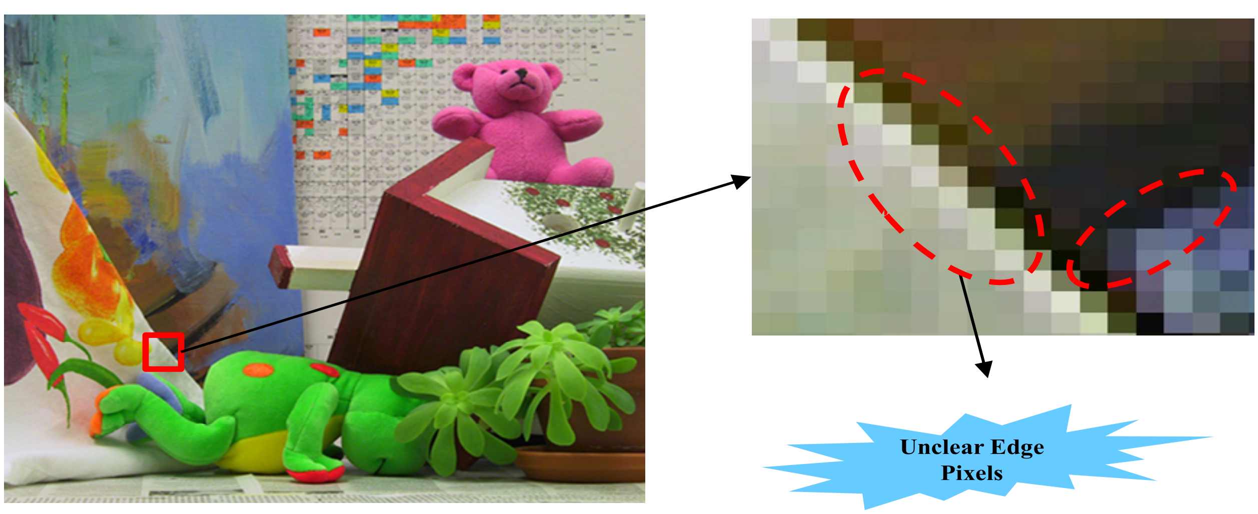


Figure 1. Image rectification in the parallel camera array

3. DISCONTINUITY PRESERVING DEPTH ESTIMATION

3.1 Overall framework

The proposed method is initially motivated by CSBP [14] based on hierarchical belief propagation. Due to the hierarchical structure, the previous work computes disparities accurately in the textureless region. Based on CSBP work, the proposed method adds an algorithm to improve disparity qualities in the edge region using distance transform. Furthermore, the proposed method includes occlusion handling for high-quality disparity map generation.

Figure 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) is performed on the left and right image, 2) DT-based weighting function is computed, 3) Color weighting function is calculated, 4) Block-based stereo matching is carried out based on the DT-based and color weighting functions.

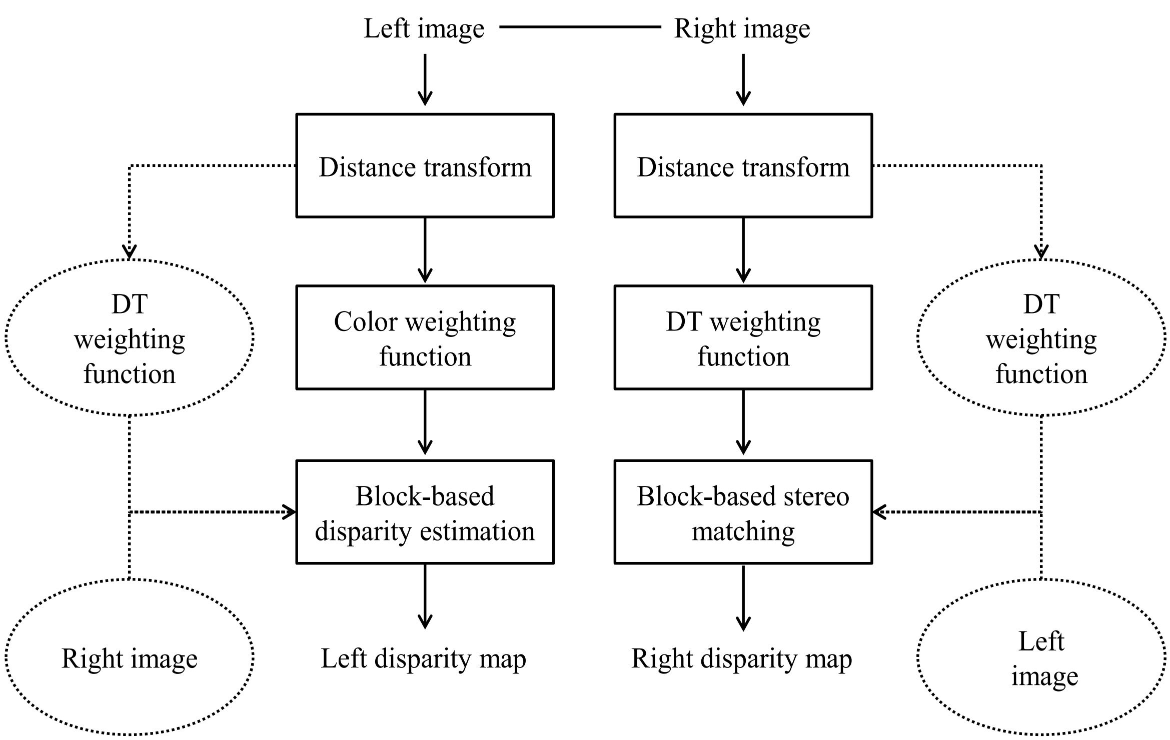


Figure 2. Overall framework of proposed method

3.2 Distance transform-based disparity estimation

Distance transform (DT) represents the shortest distance between edges extracted from an image and all pixels in it. In the field of computer vision, DT is usually adopted to trace human motions such as a hand tracking. In this paper, we present its usability for disparity estimation.

Prior to the distance transform, the Canny edge operator [15] is used to extract a color edge map E from an image. Note that isolated edges are ignored by applying a median filter to input data prior to edge detection. Based on α-β distance transform (α-β DT), the DT value $r^{k}_{i,j}$ at iteration k is computed by

$$r_{i,j}^{k}=\begin{bmatrix} r_{i-1,j-1}^{k-1}+\beta & r_{i-1,j}^{k-1}+\alpha & r_{i-1,j+1}^{k-1}+\beta \\ r_{i,j-1}^{k-1}+\alpha & r_{i,j}^{k-1} & r_{i,j+1}^{k-1}+\alpha \\ r_{i+1,j-1}^{k-1}+\beta & r_{i+1,j}^{k-1}+\alpha & r_{i+1,j+1}^{k-1}+\beta \end{bmatrix}, \tag{1}$$

where α and β control the strength of distance transform. Figure 3 illustrates a DT map generation procedure using 1-2 DT. Initially, edge pixels in E_C and E_D are set to zero, while non-edge pixels are assigned infinity values, as shown in Figure 3(a). As shown in Figure 3, if the DT value of a pixel is close to zero, the pixel may belong to a textured area, i.e., edge regions. On the other hand, if another pixel has a large DT value, the pixel may belong to a homogenous area.

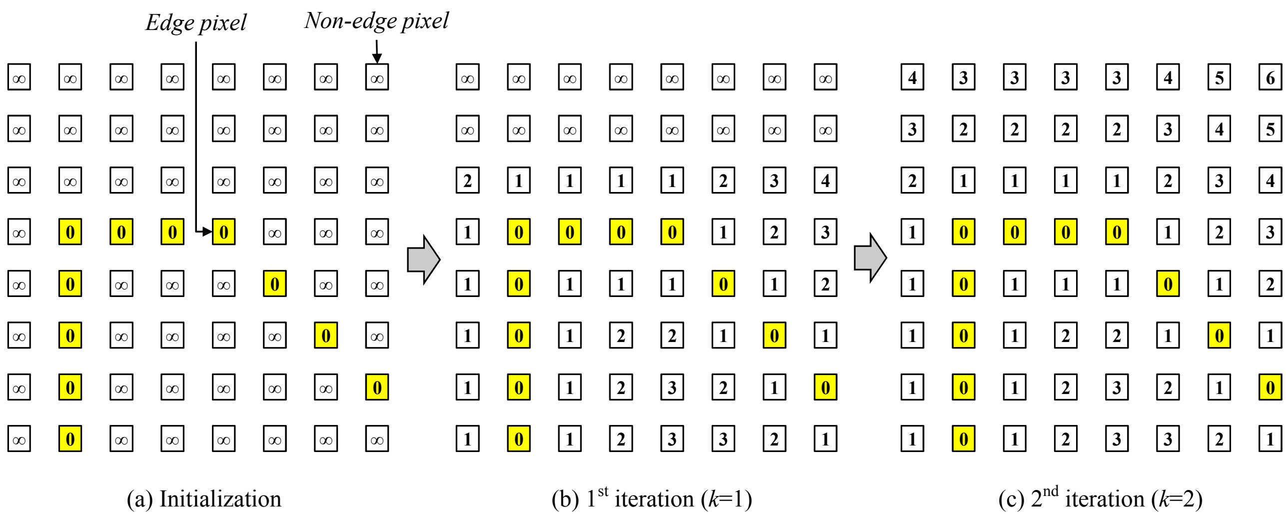


Figure 3. DT map generation.

In order to find corresponding points between left and right images, stereo matching defines an energy function composed of data term and smoothness term. When the energy function has the minimum value via energy optimization techniques such as graph cut [10] and belief propagation [8], optimal disparity value is determined.

Suppose that there are the left image I_L and the right image I_R. Let s and t denote coordinates of pixels. s is the center pixel in a local window $N(s)$. t is the neighboring pixel of s in 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

$$E(d)=\sum_{s} D_s(d_s)+\sum_{s,t\in N(s)} S_{s,t}(d_s,d_t), \tag{2}$$

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, as the data term, we employ the weighted absolute luminance difference between two blocks for the matching cost calculation. In particular, 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

$$D_s(d_s)=\frac{\sum_{t\in N(s)} W_{s,t}(dt_t)\cdot F_{s,t}(d_s)}{\sum_{t\in N(s)} W_{s,t}(dt_t)}\ , \tag{3}$$

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

$$F_{s,t}(d_s)=\min(|I_L(x_s,y_s)-I_R(x_t+d_s,y_t)|,T_d)\ , \tag{4}$$

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

$$W_{s,t}(dt_t)=f(dt_t)\cdot g(|I_{L,s}-I_{L,t}|)\ , \tag{5}$$

where $f(\cdot)$is the DT weighting function and g is the color weighting function. $|\cdot|$ is the operator to calculate 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

$$f(dt_t)=1-e^{-\frac{dt_t^2}{2\sigma_f}},g(|I_{L,s}-I_{L,t}|)=e^{-\frac{|I_{L,s}-I_{L,t}|^2}{2\sigma_g}}\ , \tag{6}$$

where σ_f and σ_g are smoothing parameters of f and g, respectively. σ_f and σ_g are usually defined as standard deviation of the Gaussian function.

In Eq. (6), the DT weighting function f is inversely proportional to the DT value dt_t, and $0 \le f \le 1$. Since the unclear edge pixel problem makes correspondence searching difficult, f imposes small weighting values, less than 0.5, on them. The smoothness term $S_{s,t}$ is based on the degree of difference among disparities of neighboring pixels. $S_{s,t}$ is represented by

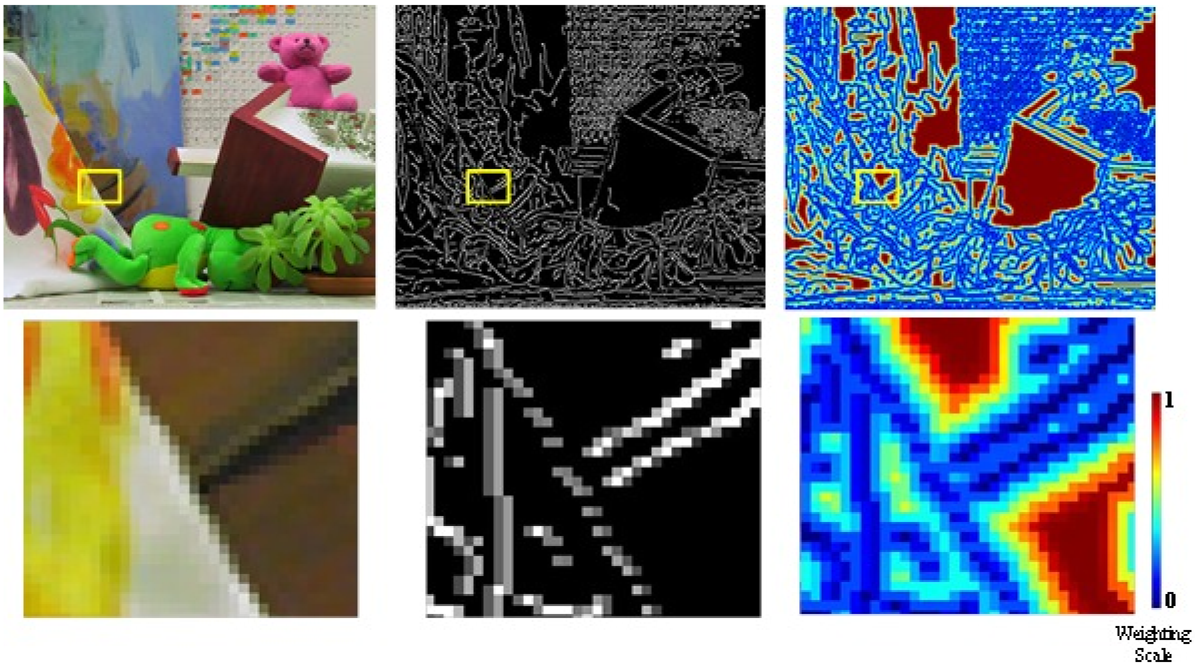
$$S_{s,t}(d_s,d_t)=\min(\lambda|d_s-d_t|,T_s)\ , \tag{7}$$

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 CSBP work [14].

4. EXPERIMENTAL RESULTS

In order to evaluate the performance of the proposed method, we tested on two stereo images sets with the same size. These test data are Teddy, and Cones. The resolution of the test images are 450×375. In the experiment, α and β values for distance transform were set to 9 and 10, respectively. For DT-based weighting function, smoothing parameters σ_f and σ_g in (6) were set to 0.3 and 0.2, respectively.

Figure 4 exhibits the DT weighting function. Figure 4(a) shows the left image of Teddy image and magnified image of a part of the left image. Figure 4(b) shows edge information and Figure 4(c) is the associated DT weighting function. As shown in Figure 4(c), the closer a pixel is to edges, the smaller the DT weighting value are assigned to the pixel to reduce the unclear edge pixel problem.



(a) Color data (b) Edge data (c) DT weighting value

Figure 4. DT weighting function.

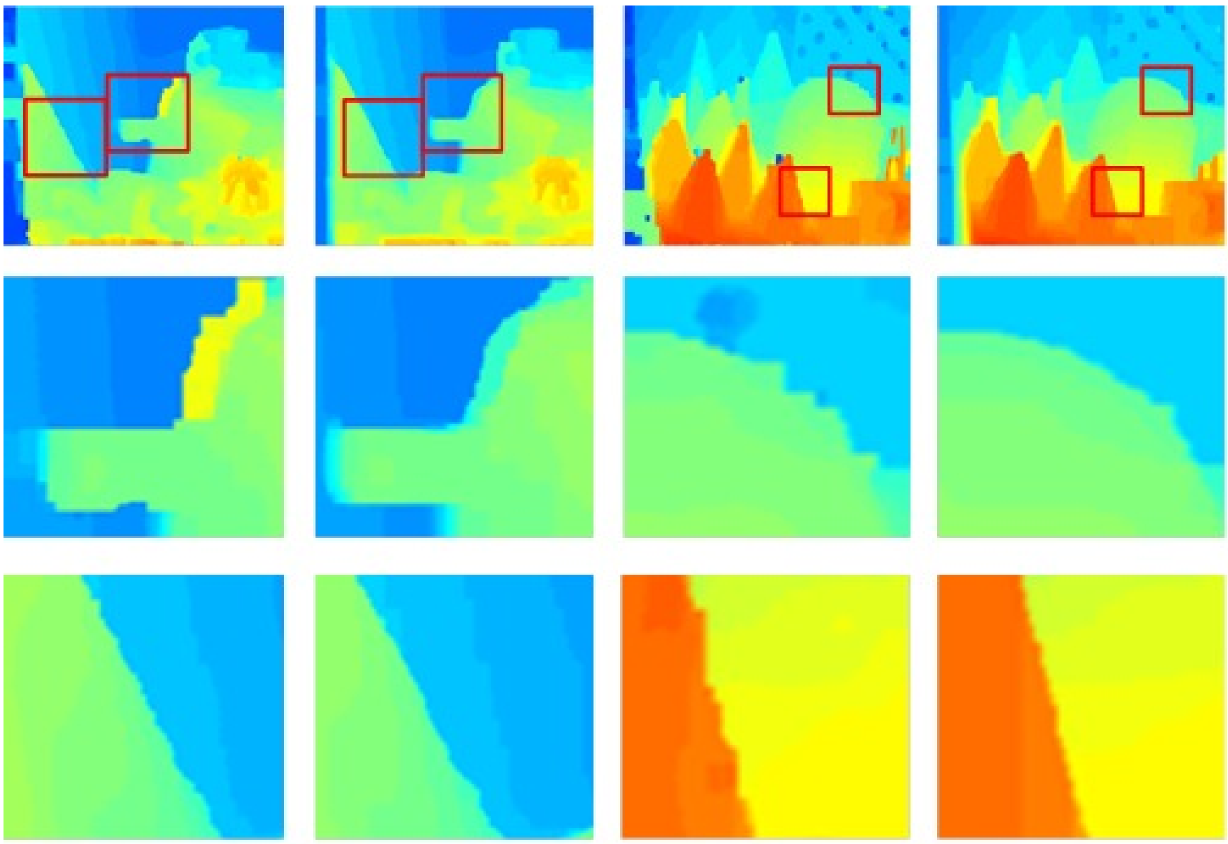
Figure 5 illustrates the visual comparison of CSBP work [14] with the initial disparity map using the proposed DT-based stereo matching. Figure 5 demonstrates that the proposed initial disparity generation improves the quality in the edge regions. The results indicate that the proposed method outperforms other methods. Furthermore, visual comparison of the experimental results demonstrates that the proposed stereo-image-based depth estimation method improve the quality in edge regions.

5. CONCLUSION

This paper presents a new disparity estimation method exploiting edge. This paper presents a new depth estimation method exploiting edge information. In general, pixel intensities around object edges are not clear due to mixed values located between the object and background. This causes problems when identifying discontinuous depth in object borders. In order to handle this, the proposed method generates depth information based on distance transform and color weighting constraint. This increases the depth quality at edges. Experimental results show that our method using distance transform produces better performance compared to the previous methods.

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). [GK130100, Development of Interactive and Realistic Massive Giga-Content Technology].



(a) CSBP in Teddy (b) DT-based method in Teddy (c) CSBP in cones (d) DT-based method in Cones

Figure 5. Depth Comparison of initial work results.

REFERENCES

[1] C. Fehn, R. De La Barré, and S. Pastoor, “Interactive 3DTV-Concepts and key Technologies,” Proceedings of the IEEE, vol. 94, Issue 3, pp. 524-538, March 2006.

[2] L. Zhang and W.J. Tam, “Stereoscopic Image Generation Based on Depth Images for 3DTV,” IEEE Transactions on Broadcasting, vol. 51, no. 2, pp. 191-199, June 2005.

[3] S.Y. Kim, J.H. Cho, and A. Koschan, “3D video generation and service based on a TOF depth sensor in MPEG-4 multimedia framework,” IEEE Transactions on Consumer Electronics, vol. 56, no. 3, pp. 1730-1738, Aug. 2010.

[4] R. Crabb, C. Tracey, A. Puranik, and J. Davis, “Real-time foreground segmentation via range and color Imaging,” Computer Vision and Pattern Workshop pp. 1-5, June 2008.

[5] J. Shotton, A. Fitzgibbon, M. Cook, and A. Blake, “Real-time human pose recognition in parts from single depth images,” Communications of the ACM, vol. 56, Issue 1, pp. 116-124, Jan. 2013.

[6] W.S. Jang and Y.S. Ho, “Efficient Disparity Map Estimation Using Occlusion Handling for Various 3D Multimedia Applications”, IEEE Transactions on Consumer Electronics, Vol. 57, No. 4, pp. 1937-1943, Nov. 2011.

[7] O. Veksler, “Fast variable window for Stereo Correspondence Using Integral Images,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. I-556 - I-561 vol.1, June 2003.

[8] J. Sun, N.N. Zheng, and H.Y. Shum, “Stereo Matching Using Belief Propagation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 7, pp. 787-800, July 2003.

[9] Z.W. Gao, W.K. Lin, Y.S. Shen, C.Y. Lin, and W.C. Kao, “Design of Signal Processing Pipeline for Stereoscopic Cameras,” IEEE Transactions on Consumer Electronics, vol. 56, no. 2, pp. 324-331, May 2010.

[10] Y. Boykov, O. Veksler, and R. Zabih, “Fast Approximate Energy Minimization Via Graph Cuts,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 11, pp. 1222-1239, Nov. 2001.

[11] V. Kolmogorov, and R. Zabih, “Computing Visual Correspondence with Occlusions Using Graph Cuts,” IEEE International Conference on Computer Vision, pp. 508-515, July 2001.

[12] P. F. Felzenszwalb and D. P. Huttenlocher, “Efficient Belief Propagation for Early Vision”, International Journal of Computer Vision, vol. 70, no. 1, pp.41-54, Oct. 2006.

[13] Q. Yang, L. Wang, R. Yang, S Wang, M. Liao, and D. Nister, “Real-Time Global Stereo Matching Using Hierarchical Belief Propagation” British Machine Vision Conference, pp. 989-998, Sept. 2006.

[14] Q. Yang, L. Wang, and N. Ahuja, “A Constant-Space Belief Propagation Algorithm for Stereo Matching”, Computer Vision and Pattern Recognition, pp. 1458-1465, June 2010.

[15] J. Canny, “A Computational Approach to Edge Detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-8, Issue 6, pp. 679-698, Nov 1896.