https://doi.org/10.2352/ISSN.2470-1173.2018.13.IPAS-386

Disparity Estimation Using Fast Motion-Search Algorithm and Local Image Characteristics

Yong-Jun Chang, Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)
123 Cheomdangwagi-ro, Buk-gu, Gwangju, 61005, South Korea
Email: {yjchang, hoyo}@gist.ac.kr

Abstract

A stereo matching method estimates the disparity value between two correspondences in both stereo images. The disparity value represents the depth information of objects obtained from stereo images which have two different viewpoints. In many papers, the stereo matching method is tested under limited disparity conditions. These conditions follow configurations of test images. However, the disparity range in real applications is not known for those conditions and configurations. In this case, we have to check all pixels in the scan line to find the correspondence. Therefore, it is a time consuming task. Thus, we propose a fast disparity estimation method using the limited search range. The proposed method limits the disparity search range with the fast motion-search algorithm and local image characteristics.

Keywords: Stereo matching, disparity map, fast motion-search, local image characteristic

1. Introduction

A depth value is one of the important cues for three-dimensional (3D) content, e.g., 3D videos and models [1]. A stereo matching method is one way to obtain this value. Although a depth value can be estimated by a depth camera, it works poorly in outdoor environments, e.g., in sunlight. On the other hand, the stereo matching method is unaffected by sunlight. The stereo matching method searches for correspondences between two viewpoint stereo images using similarity measures such as sum of absolute differences (SAD), normalized cross-correlation (NCC), and so on. Then, a disparity value is calculated using those correspondences. If an object is located near the camera, it has a large disparity value. On the other hand, an object that is far from the camera has a small disparity value. Thus, the disparity value represents the depth information.

Several pre-processes are executed before the stereo matching method is used. In real applications, stereo images are misaligned in the vertical direction. This problem causes a mismatch in the stereo matching process. Therefore, the image should be rectified for the stereo matching [2, 3]. In order to rectify the image, we must know the intrinsic and extrinsic parameters of the camera. These parameters can be obtained through camera calibration [4]. After performing the camera calibration process, camera parameters are used for acquiring aligned stereo images.

The stereo matching method uses rectified stereo images to calculate the disparity value. The image rectification is performed by camera parameters that are obtained from the camera calibration. Thus, the disparity value of each pixel is searched in the same scan line. In the disparity estimation step, the disparity range is an important key. Generally, stereo matching algorithms follow the disparity configuration of the test image [5]. Therefore, the disparity map is obtained under the limited disparity range. However, in practice, we do not know that disparity range. For this reason, a full search range is used for disparity estimation. Stereo matching with the full search range can accurately find the disparity value in textured regions because of some features in those regions. However, it is time-consuming and less effective in homogeneous regions. To enhance the weak points of the full search range, we propose a fast disparity estimation method using a limited search range with the fast motion-search algorithm and local image characteristics.

2. Conventional Disparity Search Ranges

2.1 Full Search Range

The stereo matching method with the full search range estimates the disparity value using all pixels in the scan line. Fig. 1 shows an example of the full search range matching. In Fig. 1, to find the disparity value of $I_L(x, y)$ in the left viewpoint image, the same pixel point $I_R(x, y)$ is selected in the right viewpoint image. Then, the full search range of this example is from 0 to *x*. If the value of *x* is small, the search range will be relatively small. As *x* increases, the search range and matching time also increases.

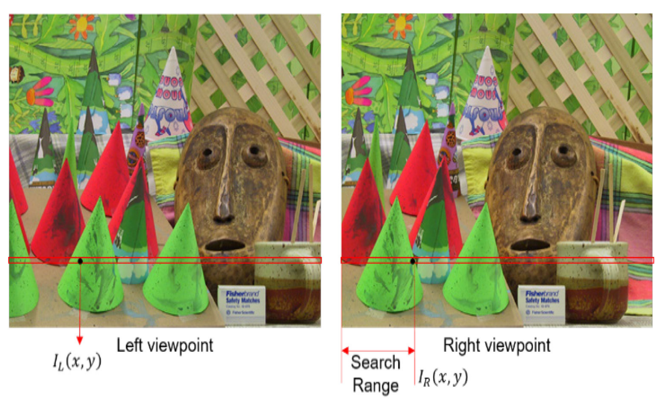
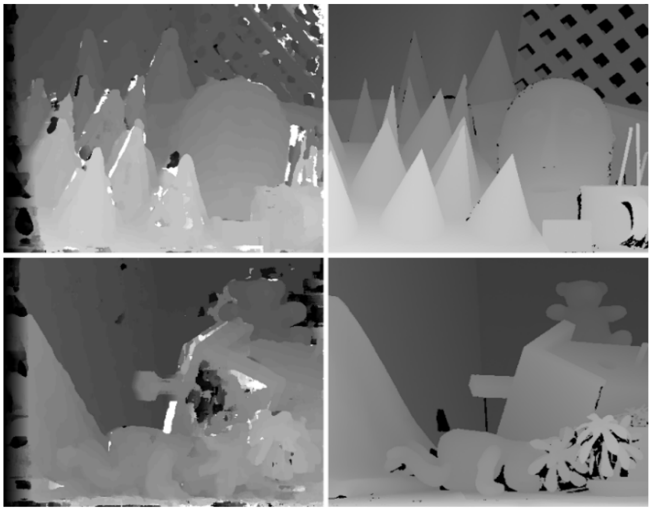


Figure 1. Full Search Range Matching

Fig. 2 shows the results of stereo matching with the full search range. Fig. 2(a) shows some disparity errors compared with the ground truth in Fig. 2(b). Since the search range is not limited in Fig. 2, the disparity error is caused by occluded and homogeneous regions. As depicted in Fig. 2, stereo matching with the full search range estimates quite accurate disparity value near the texture region. However, this method shows poor results in occluded and textureless regions because there are no features in the textureless region and there are no information in the occluded region. Therefore, stereo matching with the full search range is not an effective method for disparity estimation.

For convenience of algorithm evaluations, many papers use a maximum disparity value as the disparity search range. Hence, the performance of the stereo matching algorithm is evaluated more quickly and effectively than the stereo matching with the full search range.



(a) Full Search Range (b) Ground Truth

Figure 2. Matching Result of Full Search Range

2.2 Stereo Matching with Maximum Disparity Value

If we limit the search range to the maximum disparity configuration, the resulting disparity map has better quality than that of the full search range [5]. Because the limited search range provides more feasible disparity candidates than the full search range does, disparity errors as well as matching time can be reduced. Fig. 3 shows the results of a limited search range using the maximum disparity configuration. In Fig. 3(b), disparity errors in occluded and homogeneous regions are reduced compared to Fig. 2(a). However, stereo matching with the maximum disparity value still has some disparity errors in those regions.

(a) Original Image (b) Matching Results

Figure 3. Matching Results of Limited Search Range

Stereo matching with the maximum disparity configuration gives rapid and high quality results. However, it is difficult to use the maximum disparity value in practical situations because there is no prior information about the disparity range. Therefore, a new approach to limit the search range is needed for the stereo matching method.

3. Stereo Matching Using Fast Motion-Search Algorithm and Local Image Characteristics

The motion-search algorithm is commonly used for video coding. In video coding, the motion-search reduces the temporal redundancies by searching for the most similar motion between the current frame and the previous frame. Several fast and efficient motion-search algorithms are available. One of these is the three step search (TSS) algorithm [6]. This algorithm searches for the best motion using a coarse to fine search pattern. Our method applies this concept to the disparity estimation process for the efficient disparity estimation.

3.1 Fast Motion-Search Algorithm

The TSS algorithm in video coding uses a two-dimensional (2D) coordinate system. Since the stereo matching method searches for disparity values in the same scan line, we modified the 2D TSS to a one-dimensional (1D) TSS [7]. Fig. 4 shows an example of the modified TSS.

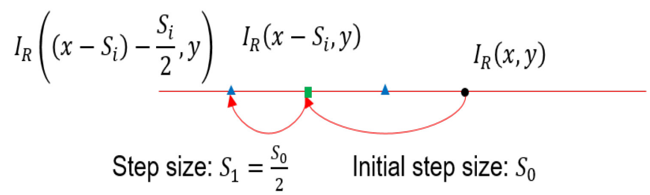


Figure 4. Modified TSS

This example shows the method of finding the correspondence of the left viewpoint pixel $I_L(x,y)$. It was used in [7]. First, we set an initial center point in the right image. In Fig. 4, the initial center point is set to $I_R(x,y)$. Next, the initial center point is moved by an initial step size S_i to obtain the second point, $I_R(x-S_i,y)$. At this point, the matching cost between $I_L(x,y)$ and $I_R(x-S_i,y)$ is calculated. After calculating the matching cost, other two matching costs are also calculated. In order to calculate other two matching costs, we need to reduce the initial step size. In Fig. 4, the second step size S_1 is set to $\frac{S_0}{2}$. Using this step size, the matching costs between $I_L(x,y)$ and $I_R\left((x-S_i)-\frac{S_i}{2},y\right)$, as well as $I_L(x,y)$ and $I_R\left((x-S_i)+\frac{S_i}{2},y\right)$ are calculated. Then, the three matching costs are compared with each other. If the matching cost at $I_R\left((x-S_i)-\frac{S_i}{2},y\right)$ has the optimal value, then the center pixel is moved to this point; otherwise, it is moved to the other point that has the optimal matching cost. The next step size S_i is changed to $\frac{S_{i-1}}{2}$. This process is repeated until the step size S_i is smaller than 1. By using this coarse to fine search pattern, the disparity value of $I_L(x,y)$ is found more efficiently than with the full search method.

3.2 Determining Initial Step Size with Local Image Characteristics

The initial step size has a significant effect on the disparity estimation. If the initial step size is too large, a disparity error can occur because of the local minimum problem as shown in Fig. 5.

In Fig. 5, the real correspondence point of $I_L(x,y)$ is $I_R(x-d,y)$. If the initial step size is very large as depicted in Fig. 5, the real correspondence point cannot be detected because of the wrong search range. Therefore, in this case, there is the local minimum problem in which the optimal correspondence point is detected in the region having no correlation.

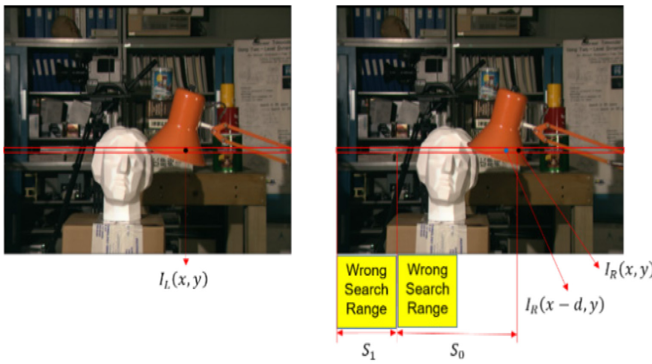


Figure 5. Local Minimum Problem

To avoid this local minimum problem, an equation of the initial step size was proposed in [7]. The equation of the initial step size used in [7] is defined in (1).

$$S_{i,(x,y)} = e^{-\frac{var(x,y)}{\varepsilon}} \cdot d_{(x-1,y)} + \left(1 - e^{-\frac{var(x,y)}{\varepsilon}}\right) \cdot d_{(x,y-1)} \quad (1)$$

In (1), where $var(x,y)$ is the local image variance at (x,y), $d_{(x-1,y)}$ is the disparity value at $(x-1,y)$, and $d_{(x,y-1)}$ is that of pixel at $(x,y-1)$. Hence, this equation gives a larger weight value to the disparity value of the previous pixel when the local variance is small. If not, it also uses the disparity value of the upper pixel.

As depicted in (1), the conventional fast stereo matching method uses disparity values of neighboring pixels for determining the initial step size. However, it does not consider disparity and color continuities in the textureless region. Therefore, this method can choose wrong disparity information from neighboring pixels. Our method focuses on solving these problems. The proposed method also uses the disparity value of the previous pixel in order to determine the initial step size. The initial step size is redefined as follows:

$$S_{i,(x,y)} = \begin{cases} \alpha \cdot \left(d_{(x-1,y)} + 1\right), & if\ d_{(x-1,y)} < \tau \\ e^{-\frac{C(x,y)}{\varepsilon_v}} \cdot d_{(x-1,y)} + \left(1 - e^{-\frac{C(x,y)}{\varepsilon_v}}\right) \cdot PDV, & if\ d_{(x-1,y)} \geq \tau \end{cases} \quad (2)$$

In (2), where $S_{i,(x,y)}$ is the initial step size at (x,y), $d_{(x-1,y)}$ is the disparity value of the pixel at $(x-1,y)$, $C(x,y)$ is the variance value of the matching block at the centered pixel (x,y), and *PDV* is the predicted disparity value at pixel (x,y). Both $d_{(x-1,y)}$ and *PDV* are balanced by the weighting function $e^{-\frac{C(x,y)}{\varepsilon}}$. Therefore, if $d_{(x-1,y)}$ is smaller than τ, we assume that $d_{(x-1,y)}$ is the disparity error caused by the local minimum problem. Thus, we increase the value of the initial step size to avoid this problem. Otherwise, the initial step size is defined by the weighted sum of $d_{(x-1,y)}$ and *PDV*.

Since $C(x,y)$ in (2) means the variance of the matching block, a high weight value is generated in the homogeneous region and a small weight value is generated in the textured region. Therefore, if the matching block is located in the homogeneous region, the high weight value is applied to the previous disparity value. Otherwise, it assigns a large weight value to the *PDV*. Variance $C(x,y)$ is calculated as follows:

$$C(x,y) = \frac{\sum_{i=-\left\lfloor\frac{N}{2}\right\rfloor}^{\left\lfloor\frac{N}{2}\right\rfloor} \sum_{j=-\left\lfloor\frac{N}{2}\right\rfloor}^{\left\lfloor\frac{N}{2}\right\rfloor} |G(x+i,y+j) - G(x,y)|}{N^2} \quad (3)$$

where N^2 is the size of the matching window and $G(x,y)$ is the pixel value of the gray image. Because the equation used to obtain the real variance requires the mean value of all pixels in the matching block, its calculation takes time. If we use a local transform as depicted in (3), the variance of the pixel values can be acquired easily [10].

Fig. 6 shows candidates for the *PDV* at (x,y). This method was inspired by the motion vector predictor [8, 9].

$$\begin{aligned} SAD(d_a) &= \textstyle\sum_{i\in r,g,b} X_i \cdot |I_i(x,y) - I_i(x-1,y)| \\ SAD(d_b) &= \textstyle\sum_{i\in r,g,b} X_i \cdot |I_i(x,y) - I_i(x-1,y-1)| \\ SAD(d_c) &= \textstyle\sum_{i\in r,g,b} X_i \cdot |I_i(x,y) - I_i(x,y-1)| \end{aligned} \quad (4)$$

In Fig. 6, there are three *PDV* candidates: d_a, d_b, and d_c, where d_a is the disparity value at $(x-1,y)$, d_b is the disparity value at $(x-1,y-1)$, and d_c is the disparity value at $(x,y-1)$. To choose only one *PDV*, the matching error is calculated between the current pixel and the three neighboring pixels using the sum of absolute differences. Equations for the computation of matching error is defined in (4).

In (4), where *I* is a pixel value, *i* is one of the color channels. Three weighting values were used: X_r, X_g, and X_b, where X_r is 0.2126, X_g is 0.7152, and X_b is 0.0722. Thus, the *PDV* is determined by (5).

$$PDV = \underset{d\in d_a,d_b,d_c}{\text{argmin}}\ SAD(d) \quad (5)$$

In (5), the disparity value of the pixel which has the smallest matching error is selected as the disparity candidate. It means that the disparity information of a reliable pixel should be used.

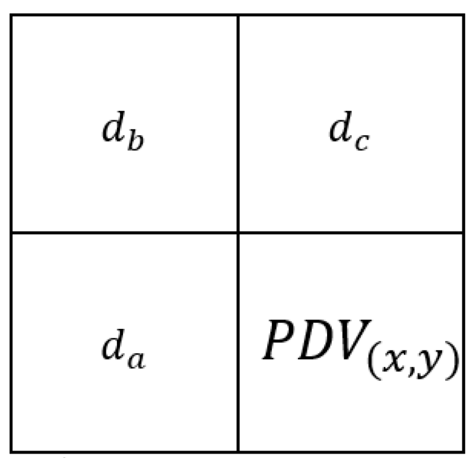


Figure 6. PDV Candidates

3.2 Disparity Cost Term

The proposed method uses the fast motion-search algorithm with the local image characteristic. In this method, only the color difference between the current pixel and the compared pixel is used as the matching cost for searching the corresponding point. The proposed method also checks the disparity continuity in the cost computation step. All the step sizes in the fast motion-search algorithm represent the disparity candidate. In this step, we regard the current step size as one of the disparity candidates. If the pixel value of the current pixel is similar to the previous one, there is a high probability of having similar disparity value. Hence, we add the disparity cost term to the matching cost computation step as defined in (6).

$$Cost(x,y) = e^{-\frac{\nabla I}{\varepsilon_c}} \cdot DC(x,y) + \left(1 - e^{-\frac{\nabla I}{\varepsilon_c}}\right) \cdot CC(x,y) \qquad (6)$$

In (6), where ∇I means the color difference between the current pixel and the previous pixel, DC is the disparity cost term, and CC is the color cost term. If ∇I is small, the disparity value of the current pixel may be like that of the previous pixel. Therefore, the large weight value is given to CC. Otherwise, the large weight value is applied to DC. The color cost term CC is the matching cost between correspondence points of original color images. The disparity cost term DC calculates the matching cost between the previous disparity value and the current step size. The equation of DC is defined in (7).

In (7), where S_c means the current step size. Therefore, this term checks the disparity similarity between the current pixel and the previous pixel. This term tries to find the disparity value that is similar to that of the previous pixel.

$$DC(x,y) = \left|d_{(x-1,y)} - S_c\right| \qquad (7)$$

The equation of CC is represented in (8). In (8), $M(x,y)$ is a pixel value at (x,y). This value can be calculated as a pixel-wise or a block-wise. In our experiments, we calculate the matching cost using the block-wise matching.

$$CC(x,y) = |M_L(x,y) - M_R(x - S_c, y)| \qquad (8)$$

4. Experiment Results

In the experiment, four test images were used [5]. The results of the proposed method are compared with those of the full search range, the limited search range with the maximum disparity configuration, and the conventional fast stereo matching algorithm [7]. The sum of absolute differences for an 11×11 sized block was used for the matching cost computation. Parameters α, τ, ε_v, and ε_c in (2) were set to 8, 3, 100, and 2 respectively. The disparity value at $x = 0$ was set to 0. If x was larger than 1 and y was equal to 0, the initial step size was set to $d_{(x-1,y)}$.

Fig. 7 shows the images resulting from the stereo matching with various search ranges. In Fig. 7, the test images represent *Teddy*, *Cones*, *Tsukuba*, and *Venus*, from top to bottom, respectively. In Fig. 7, the results of the proposed method have better quality than those of stereo matching with the full search range. Our results also have better quality than stereo matching with the maximum disparity value and the conventional method.

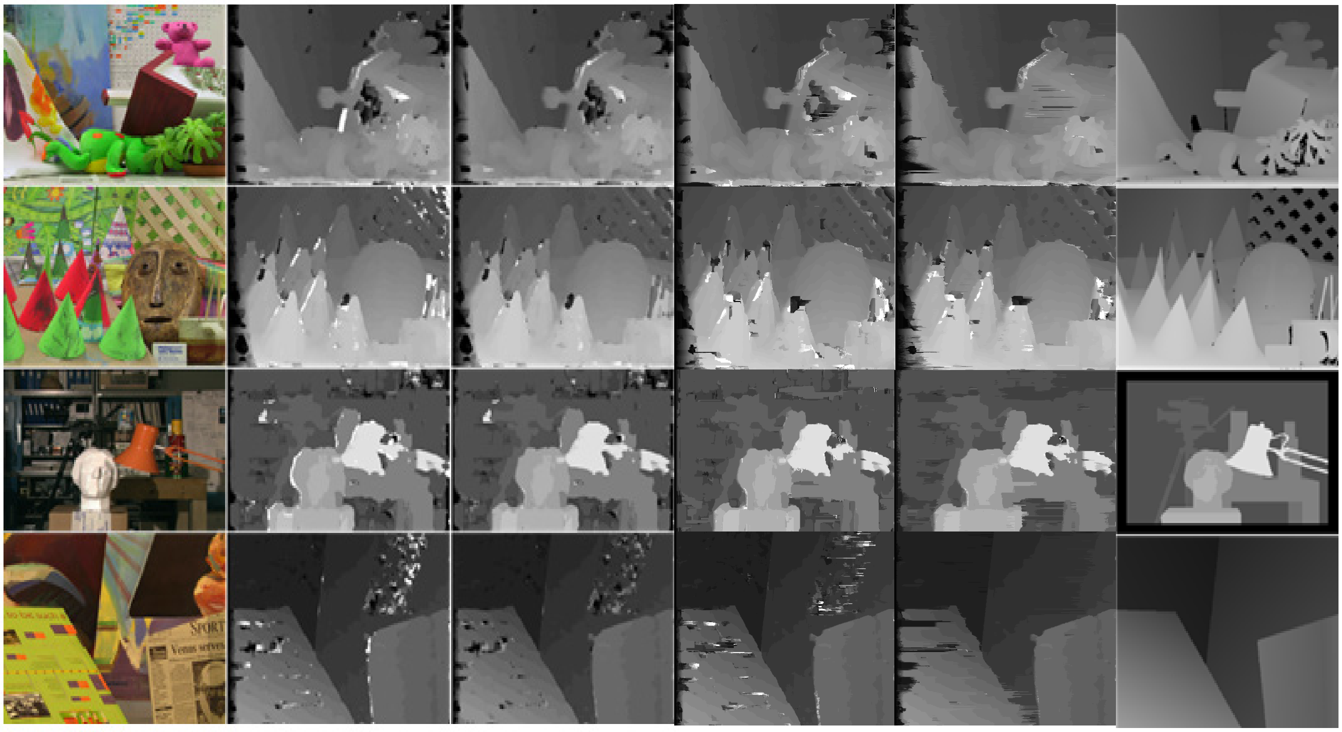
For comparing the results of our method with other methods, we also analyzed the resulting images with the enlarged pictures. Fig. 8 represents enlarged images of *Teddy* and *Venus* near homogeneous regions. Since the proposed method considers disparity continuity, our method shows better quality than other methods in the homogeneous region.

Table 1 shows the comparison of the error rate. The error rate was calculated as being the rate of the bad pixels. We measured error rates in two regions: the non-occluded region and the textureless region. In Table 1, the full search method showed the worst results in both regions. The stereo matching with the maximum disparity value estimated the disparity value better than the full search method. The conventional fast stereo matching method had better results in the textureless region compared with both the full search and the maximum disparity methods.

However, the conventional fast stereo matching method has worse average error rate than that of the maximum disparity method in the non-occluded region. On the other hand, our method showed the best error rates compared to those of the full search algorithm, the maximum disparity algorithm, and the conventional method. In Table 1, other conventional methods used only the color similarity measure for the cost computation. On the other hand, the proposed method used the disparity cost term that was depicted in Section 3.

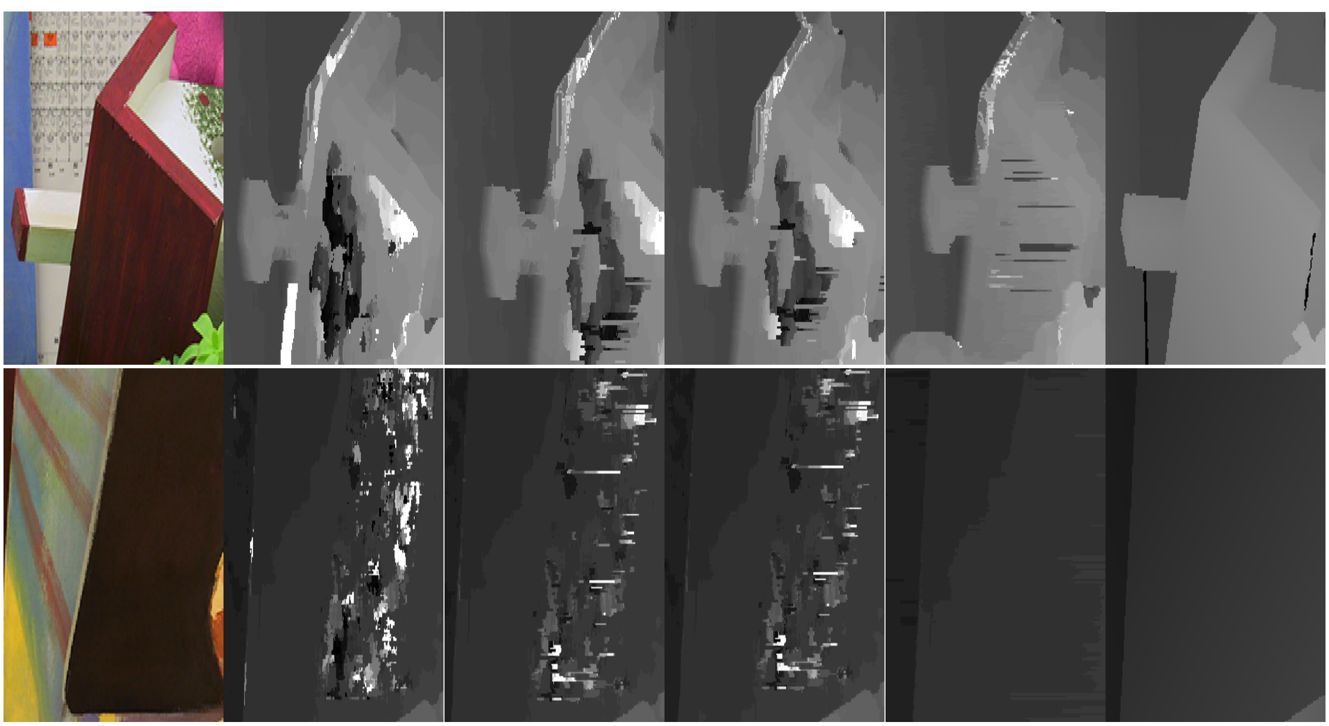
Table 1: Error Rate Comparison

	Full search		Max disparity		Conventional method [7]		Our method	
	Error rate (%)							
	Non-occ	Texture-less	Non-occ	Texture-less	Non-occ	Texture-less	Non-occ	Texture-less
Teddy	20.18	36.81	19.94	36.47	19.64	35.98	17.49	31.8
Cones	14.99	28.7	14.34	26.8	17.07	29.26	15.37	16.26
Tsukuba	8.77	10.46	8.41	10.07	6.84	6.31	7	6.96
Venus	9.93	20.23	9.19	18.48	9.57	18.02	7.33	10.48
Avg.	13.47	24.05	12.97	22.96	13.28	22.39	11.8	16.38



(a) Original Image (b) Full Search (c) Max Disparity (d) Conventional Method (e) Our method (f) Ground Truth

Figure 7. Comparison of Resulting Images



(a) Original Image (b) Full Search (c) Max Disparity (d) Conventional Method (e) Our method (f) Ground Truth

Figure 8. Enlarged Images

Table 2 represents a comparison of the implementation time. In Table 2, the full search method took the longest time for the disparity estimation. The stereo matching with the maximum disparity value was faster than the full search method. However, this method still took a lot of time. On the other hand, both the conventional fast stereo matching method and the proposed method have similar implementation time. However, both methods were faster than the other methods.

Table 2: Implementation Time Comparison

	Full search	Max disparity	Conventional method [7]	Our method
	Time (sec.)			
Teddy	66.26	18.63	3.49	3.29
Cones	70.43	18.56	3.62	3.54
Tsukuba	37.43	2.98	1.56	1.47
Venus	62.88	5.66	2.52	2.28
Avg.	59.25	11.46	2.8	2.64

5. Conclusion

The full disparity search range is generally used for stereo matching. However, it is time-consuming and has poor matching results in occluded and homogeneous regions. To search for the disparity value efficiently, the proposed method used a modified three step search algorithm. This search pattern is affected by the size of the initial search step. To find a reasonable step size, we used disparity values from the neighboring pixels and predicted disparity value with local image characteristics. For this reason, our method limits the search range of the disparity value unlike the full search algorithm. As a result, the proposed method had faster and better matching results than those of the full search algorithm.

Acknowledgement

This research was partially supported by the 'Brain Korea 21 Plus Project' of the Ministry of Education & Human Resources Development, Republic of Korea (ROK) [F16SN26T2205], and partially supported by the 'Cross-Ministry Giga Korea Project' of Ministry of Science, ICT and Future Planning, Republic of Korea (ROK). [GK17C0100, Development of Interactive and Realistic Massive Giga – Content Technology]

References

[1] C. Fehn, R. Barre, and S. Pastoor, "Interactive 3-DTV-concepts and key technologies," Proceeding of the IEEE, vol. 94, no. 3, pp. 524-538, Mar. 2006.

[2] Y. S. Kang and Y. S. Ho, "An efficient image rectification method for parallel multi camera arrangement," IEEE Transactions on Consumer Electronics, vol. 57, issue 3, pp. 1041-1048, Aug. 2011.

[3] N. Ayache and C. Hansen, "Rectification of images for binocular and trinocular stereovision," International Conference on Pattern Recognition, pp. 11-19, Nov. 1988.

[4] Z. Zhang, "A flexible new technique for camera calibration," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, issue 11, pp. 1330-1334, Nov. 2000.

[5] http://vision.middlebury.edu/stereo/data/

[6] Y. J. Chang and Y. S. Ho, "Fast stereo matching method using motion estimation and disparity information of neighboring pixels," Korean Institute of Broadcast and Media Engineers Summer Conference, pp. 186-187, June 2017.

[7] T. Koga, K. Iinuma, A. Hirano, Y. Iijima, and T. Ishiguro, "Motion compensated inter frame coding for videoconferencing," Proceedings of the National Telecommunications Conference, pp. 1-5, Dec. 1981.

[8] M. C. Chen and A. N. Willson, "A spatial and temporal motion vector coding algorithm for low-bit-rate video coding," IEEE International Conference on Image Processing, pp. 791-794, Oct. 1997.

[9] G. Laroche, J. Jung, and B. Pesquet-Popescu, "RD optimized coding for motion vector predictor selection," IEEE Transactions on Circuits and Systems for Video Technology, vol. 18, issue 9, pp. 1247-1257, Sept. 2008.

[10] R. Zabih and L. Woodfill, "Non-parametric local transforms for computing visual correspondence," European Conference on Computer Vision, pp. 151-158, May 1994.

[11] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," International Journal of Computer Vision, vol. 47, issue 1, pp. 7-42, Apr. 2002.

Author Biography

Yong-Jun Chang received his B.S. in electronic engineering and avionics from the Korea Aerospace University, Gyeonggi-do, Korea (2014) and his M.S. in electrical engineering and computer science from the Gwangju institute of Science and Technology, Gwangju, Korea (2016). Since then he has studied in the Gwangju Institute of Science and Technology in Gwangju, Korea for Ph.D. courses. His research interests are stereo matching, video coding, and image processing.

Yo-Sung Ho received his B.S. and M.S. degrees in electronic engineering from Seoul National University, Seoul, Korea (1981, 1983) and his Ph.D. in electrical and computer engineering from University of California, Santa Barbara, USA (1990). He worked at ETRI from 1983 to 1995, and Philips Laboratories from 1990 to 1993. Since 1995, he has been with Gwangju Institute of Science and Technology, Gwangju, Korea, where he is currently a professor. His research interests include video coding, 3D image processing, 3DTV, AR/VR, and realistic broadcasting systems.