ADAPTIVE DISPARITY ESTIMATION FOR AUTO CONVERGENCE OF REGION OF INTEREST IN A VIDEO

Young-Gon Kim, Rae-Hong Park

Department of Electronic Engineering, School of Engineering, Sogang University 35 Baekbeom-ro (Sinsu-dong), Mapo-gu, Seoul 121-742, Korea

ABSTRACT

Recently, various devices for three-dimensional (3-D) effect have been developed. For producing 3-D effect of the scene or the region of interest (ROI), disparity should be accurately estimated. People watching 3-D video feel visual fatigue if magnitude of parallax for the ROI is excessively large because a convergence point is not accurately put on the ROI. For producing 3-D effect, a 3-D formatter overlaps left and right images by shifting horizontally the right image by the estimated disparity of the ROI. In this paper, an adaptive disparity estimation algorithm for auto convergence of the ROI in a video is proposed using the first-order Taylor series expansion of disparity and adaptive disparity search range prediction in a stereoscopic video. First, a stereo video that consists of a number of pairs of left and right images is captured in parallel stereo camera configuration. A window in each frame is selected within the ROI and tracked. Then, for automatically adjusting a convergence point on the ROI, two steps are needed with the previously estimated disparities. The first-order Taylor series expansion is used to approximate disparity search range in similarity measure computation. Subjective evaluation such as visual fatigue, comfort, and 3-D effect of the proposed algorithm and existing algorithms is performed. Experimental results with four test videos and subjective evaluation show that the proposed algorithm gives 3-D effect with visual comfort.

Keywords

Disparity estimation, Region of interest, 3-D effect, Convergence, Disparity search, Subjective evaluation, Stereoscopic video, Taylor series expansion, Parallax, 3-D video, Moving average filter, Similarity measure

1. INTRODUCTION

Various computer vision techniques for three-dimensional (3-D) image and video have been investigated [1]–[6], and some of them have been developed as a commercial product in a market. In order to realize 3-D in a life, two types of devices have been developed: glasses type and non-glasses type. The glasses type is categorized into two classes: the shutter glasses [1], [2] and the polarized glasses types [2]–[4]. The operation of the shutter glasses type is based on the blocking of view for the left-and-right eye alternatively. On the other hand, the polarized glasses type uses a polarized filter to project scene information to each eye. Non-glasses types were also developed such as parallax barrier and lenticular display devices [5], [6].

Two optical axes of towed-cameras intersect at a point (i.e., convergence point), at which disparity or parallax is zero. When the image captured by a left camera is a reference image, a

searching direction to estimate a similarity is left-to-right for a right image. The obtained result corresponds to a region of interest (ROI) in a left image. If a disparity of the ROI is negative (positive), it is called positive (negative) parallax [7]. For a negative parallax ROI, the human eyes recognize the ROI floating in front of a screen. A positive parallax ROI is perceived as an opposite situation where the ROI is in the back side of a screen. If the magnitude of parallax of the ROI is too large, the human eyes feel fatigue easily because a convergence point is not put on the ROI. Any ROI located far from the convergence point produces visual fatigue when people watch the object [8]. It is noted that a convergence point should be put on the ROI for visual comfort while watching the 3-D contents.

As mentioned above, the convergence point should be set to an ROI to reduce visual fatigue and make both eyes more comfortable. There are two approaches to set the convergence point to the ROI. The first approach is a replacement of camera components, such as lens or CCD sensor by sliding along the left or right direction [9]–[11]. Various cameras are used to capture stereo images, such as toed-in camera [12], planar camera [13], [14], and so on.

The second approach is based on a horizontally shift algorithm by the estimated disparity of the ROI. For this goal, rectification is needed [13]. Generally, left and right images are rectified by a transformation of conjugate epipolar lines that become collinear and parallel to one of the image axes. If non-rectified left and right images are used, keystone distortion arises around boundary of the result image by the 3-D formatter, in which eyes can easily get visual fatigue [15]. Zhou and Li proposed a rectification method for two images captured by stereo camera that satisfies constraints of the stereoscopic visualization platforms [16].

Various researches on 3-D processing have been done not only for still image but also for video. Chen *et al.* proposed an algorithm to adjust a convergence point of a region at the bottom in each frame of a video [17], where the ROI is extracted by scale invariant feature transform algorithm [18], [19]. Xu *et al.* proposed an algorithm in which a convergence point is put on the foreground in order to eliminate negative parallax [20], which makes all the objects located behind the display device in each frame of a video (i.e., all the objects have zero or positive parallax). Results of these two existing algorithms produce visual fatigue because the convergence point is not automatically set to the ROI for all frames of a video when the ROI moves towards or backward along the Z-axis in 3-D world coordinates. The convergence point should be set to the ROI by changing the disparity of the ROI to zero for visual comfort [9]–[11], [17]–[20].

As in the two-dimensional (2-D) contents, people can enjoy 3-D contents without visual fatigue if the object of the ROI is always placed on the convergence point. Adjusting the convergence point to the ROI also produces 3-D effect because of the relative disparity changes of the other objects around the ROI. To place the convergence point to the ROI, an algorithm for automatically adjusting a convergence point to the ROI in each frame in a video is proposed using the firstorder Taylor series expansion of disparity and adaptive disparity search range prediction.

The rest of the paper is organized as follows. In Section 2, the proposed disparity estimation algorithm of the ROI is proposed using the first-order Taylor series expansion of disparity and adaptive disparity search range prediction. Experimental results of the proposed and two existing algorithms are given in Section 3, and finally Section 4 concludes the paper.

2. PROPOSED METHOD

For producing the 3-D effect without visual fatigue, the convergence point is set to the ROI because two eyes can get easily visual fatigue if the ROI has large negative or positive parallax (if the convergence point is set to the ROI, parallax is zero). Two existing algorithms set the

convergence point to fixed location with the estimated disparity of each frame in a video [16], [20]. Eyes can get easily visual fatigue by two existing algorithms because the convergence point is not correctly put on the ROI. To give visual comfort, the convergence point should be adjusted to the ROI by shifting the images as much as the estimated disparity of the ROI.

Thus, for visual comfort, the disparity estimated correctly for each frame is to be used for automatically adjusting the convergence point to the ROI. An adaptive disparity estimation algorithm for auto convergence in a video is proposed using the first-order Taylor expansion of disparity and adaptive disparity search range prediction. At first, a user needs to select the ROI. Then, a window of the current frame for disparity estimation is extracted from the selected ROI. Once initial disparity is estimated with a large search range, disparities of the successive frames can be easily calculated by the first-order Taylor series expansion. A moving average filter is used to determine search range adaptively in each frame.

2.1. Window Selection of the *n*th Frame of a Video

A constraint that there is a single ROI of smooth 3-D surface is used under the assumption that one can gaze at a single ROI when one watches 2-D or 3-D contents. In this paper, a ROI represents an object while a window denotes a region within the ROI. Size of the ROI increases (decreases) if the object considered as the ROI moves forward (backward) to the stereo camera. The window Ω_n in the *n*th frame is needed for calculating the similarity measure between a window in the left image and its corresponding window in the right image. The window Ω_n is selected from the ROI, where the shape of the ROI varies according to the algorithm used to detect the ROI. A shape of the window should be decided according to the shape of the ROI. Most shapes commonly used are squares or rectangles. In this case, the window Ω_n is expressed as

$$\Omega_n = \left\{ (i, j) \mid x_n^1 \le i \le \left\lfloor \frac{x_n^1 + x_n^2}{2} \right\rfloor, \\ \left\lfloor \frac{y_n^1 + y_n^2}{2} \right\rfloor - \frac{H}{2} \le j \le \left\lfloor \frac{y_n^1 + y_n^2}{2} \right\rfloor + \frac{H}{2} \right\}$$
(1)

where (x_n^1, y_n^1) denotes coordinates of the top left point of the ROI in the *n*th frame, while (x_n^2, y_n^2) represents coordinates of bottom right point, where $x_n^2 > x_n^1$ and $y_n^2 > y_n^1$. The height of the window *H* is assumed to be an even number and |x| denotes the largest integer not greater than *x*.

If the shape of the ROI is not a square or rectangle, the minimum bounded rectangle is used for representation of the shape of the ROI. Then, a window Ω_n is also described by using (x_n^1, y_n^1) and (x_n^2, y_n^2) extracted from the ROI.

2.2. Disparity Approximation Using the First-Order Taylor Series Expansion

In this step, disparity of the *n*th frame of video using temporal information is approximated. At first, initial disparity estimate \hat{d}_0 without using temporal information is calculated, which is described as

$$\hat{d}_{0} = \arg\min_{d_{0}} \sum_{(p,q)\in\Omega_{n}} |I_{L}(p,q) - I_{R}(p+d_{0},q)|, \quad 0 \le \hat{d}_{0} \le D_{0}$$
(2)

where I_L and I_R denote the rectified left and right images from the stereo configuration, respectively, (p,q) represents the coordinates of point within the window Ω_n . D_0 signifies initial disparity search range for initial disparity estimate \hat{d}_0 . The sum of square differences (SAD) is used as the similarity measure for real-time disparity estimation [21], [22].

Then, disparity \hat{d}_n of the *n*th frame can be recursively estimated, which is calculated as

$$\hat{d}_n = \tilde{d}_n + \Delta \hat{d}_n. \tag{3}$$

The first term \tilde{d}_n approximates the disparity of the *n*th frame obtained from the first-order Taylor series expansion. It can be described as

$$\tilde{d}_n = \hat{d}_{n-1} + d'_{n-1} \tag{4}$$

$$d'_{n} = \hat{d}_{n} - \hat{d}_{n-1} \tag{5}$$

where d'_n represents the difference of disparity estimates between the *n*th and (n-1)th frames. The second term $\Delta \hat{d}_n$ in (3) denotes the correcting disparity that compensates for the first term (approximated disparity), which is calculated by the SAD as

$$\Delta \hat{d}_n = \operatorname{argmin}_{\Delta d_n} \sum_{(p,q)\in\Omega_n} \left| I_L(p + \tilde{d}_n, q) - I_R(p + \tilde{d}_n + \Delta d_n, q) \right|, - D_n \le \Delta \hat{d}_n \le D_n, 1 \le n < N$$
(6)

where *N* represents the total number of frames in a video. The adaptive disparity search range estimate D_n of the *n*th frame is used except for the initial disparity estimate of the first frame. The adaptive disparity search range estimate D_n will be described in more detail in Section 2.3. Figure 1 shows disparity estimation of the *n*th frame using the first-order Taylor series expansion.



Figure 1. Disparity estimation of the nth frame in a video.

2.3. Adaptive Search Range Prediction Using a Moving Average Filter

 D_0 in (2) is used as the maximum search range to estimate \hat{d}_0 , while adaptive search range D_n in (6) is used to specify the disparity search range of $\Delta \hat{d}_n$. Once \tilde{d}_n is obtained with previously estimated disparities, the SAD is only computed around approximated disparity \tilde{d}_n to calculate $\Delta \hat{d}_n$ with D_n . D_n is described as

$$D_n = D^o + \left\lfloor \left| \tilde{D}_n \right| \right\rfloor \tag{7}$$

$$\tilde{D}_{n} = \frac{1}{2k+1} \sum_{i=n-k}^{n+k} d'_{i} = \overline{d'}_{n+k-1} + \frac{d'_{n+k} - d'_{n-k-1}}{2k+1}$$
(8)

where D^o represents the offset of disparity search range and $\lfloor x \rfloor$ denotes the largest integer not greater than x. \tilde{D}_n denotes the smoothed disparity search range obtained by averaging differences of disparity estimates of the previous (2k+1) frames between the (n-k)th and (n+k)th frames. Selecting appropriate k is needed to effectively eliminate the fluctuation caused by varying differences of disparity estimates. $\overline{d'_n}$ denotes average difference of disparity estimates from the (n-k)th to (n+k)th frames. In (8), the first equality is recursively expressed as moving average filtering to reduce the computation time. A moving average filter smoothes data by replacing each data with the average of the neighboring data points within the span.

3. EXPERIMENTAL RESULTS AND DISCUSSIONS

Four test videos are used in our experiments for performance evaluation of the proposed disparity estimation algorithm. Table 1 shows features of videos 1–4 used in experiments. Video 1 is selected from Karlsruhe database [23] and videos 2–4 are captured by a stereo camera. Videos 1 to 4 consist of 340, 499, 600, and 400 frames with image size of 1347×374, 800×600, 960×588, and 960×588, respectively. In videos 1 and 2, the stereo camera takes images of a car driving in front of the car equipped with the stereo camera. Video 3 contains a person moving fast forward

and backward to the stereo camera that is fixed on a table. In video 4, several stationary objects on a table are captured by the stereo camera that moves around the objects.

_	-	Video						
•	Feature	1	2	3	4			
_		Camera,	Camera,	Ohiert	C			
•	Niotion	object	object	Object	Camera			
•	Scene	Outdoor	Outdoor	Indoor	Indoor			
•	Number of	240	400	600	100			
	frames	340	499	600	400			
•	Frame size	1347×374	800×600	960×588	960×588			
•	ROI	Car	Car	Face	Lion doll			

Table 1. Features of videos 1-4 used in experiments.

For objective performance evaluation, the root mean square (RMS) error E_{RMS} is defined as

$$E_{\rm RMS} = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} (\hat{d}_n - d_n)^2}$$
(9)

where \hat{d}_n represents the estimated disparity of the *n*th frame and d_n denotes the ground truth disparity of the *n*th frame that is manually selected for all frames of a video. N signifies the total number of frames in a video.

The ROI is detected as a square in videos 1 and 2 whereas a rectangular in videos 3 and 4. As mentioned in Section 2.1, the window is selected by considering the shape of the ROI, as expressed in (1). As an example for definition of the ROI and the window, Figure 2 shows coordinates of the ROI and the window of the *n*th frame.



Figure 2. Coordinates of the ROI and the window of the *n*th frame.

The adaboost algorithm [24] is used to detect the ROI, for example, a face in the first frame of video 3. Adaboost algorithm consists of three steps. In the first step, a large number of Haar-like features using a number of training sets are extracted, and integral image is used to reduce the computation time. Then, a strong classifier is constructed from a number of weak classifiers. Lastly, using strong classifiers, coarse-to-fine cascade structure is constructed, where a large number of non-face regions are rejected at the first stage of cascade structure. Then, detected face is tracked by mean shift algorithm [25], [26]. Mean shift algorithm iteratively computes the location of the window that maximizes the similarity between a detected ROI and a new candidate ROI in the current frame.

The ROI detected [24] and tracked [25], [26] for all frames is used for disparity estimation. To select appropriate parameters (D° and k) is important for reliable disparity estimation. For example, video 3 containing fast motion is considered. A person whose face is selected as the ROI runs towards the stereo camera and then walks backward twice. An absolute difference of disparity estimates of two consecutive frames decreases if the person runs towards the camera (before the 472th frame) whereas increases if the person walks backward gradually from the camera (after the 472th frame).

Figure 3 shows correcting disparity $\Delta \hat{d}_n$ as a function of the frame index with different offsets of disparity search range D^o , where D^o with 1, 3, and 5 $(D_n = D^o)$ are represented by solid, dashed, and dotted lines, respectively. Results corresponding to ellipses A, B, and C around the 447, 476, and 474th frames, respectively, show that the offset of disparity search range for disparity estimation should be adaptively selected for each frame. Correcting disparity $\Delta \hat{d}_n$ with $D^o = 1$ estimated by using the SAD is clipped as marked by the ellipse A around the 447th frame, while not clipped with $D^o = 3$ or 5.



Figure 3. Correcting disparity as a function of the frame index with different offsets of the disparity search range.

Figure 4 shows experimental results of the 447th frame interlaced with left image and right image shifted by the amount of disparity estimated using the offset of the disparity search range $D^\circ=1$ and 5. Figure 4(a) shows the result interlaced with left image and shifted right image by ground truth disparity, where ground truth disparity is obtained manually. Figures 4(b) and 4(c) show the enlarged results interlaced with left image and right image shifted by the estimated disparity for a dotted rectangular ROI shown in Figure 4(a), where disparity is estimated with offset of the disparity search range $D^\circ=1$ and 5, respectively. A face considered as the ROI is overlapped in

Figure 4(b) because the disparity is not correctly estimated due to a small disparity search range. Note that a face in Figure 4(c) is not overlapped.



Figure 4. Interlaced results by the estimated disparity using two different offsets of the disparity search range (video 3, 447th frame). (a) Result interlaced with left image and shifted right image by ground truth disparity, (b) Enlarged results of a dotted rectangle window using $D^{\circ} = 1$, (c) Enlarged results of a dotted rectangle window using $D^{\circ} = 5$.

However, the search range does not have to be large because the probability to estimate wrong disparity increases, as marked by the ellipse B around the 476th frame. The face considered as the ROI is overlapped with $D^o = 3$ while not overlapped with $D^o = 5$ because the disparity is inaccurately estimated due to a larger search range than necessary.

Table 2 shows comparison of E_{RMS} according to the offset of the disparity search range for video 3. The smallest E_{RMS} is shown in bold with $D^o = 3$ because the estimated disparity is similar to ground truth disparity. The disparities with $D^o = 5$ are inaccurately estimated due to a large disparity search range, as marked by the ellipse C around the 474th frame in Figure 3. However, several estimated disparities are close to ground truth disparities as marked by the ellipse C. From Figures 3 and 4 and Table 2, it is noted that the disparity search range should be adaptively changed according to previous disparity estimates.

			1) °	
-	1	2	3	4	 10
• E _{RMS}	62.92	1.64	1.41	1.54	 1.67

Table 2. Comparison of E_{RMS} according to the offset of the disparity search range for video 3.

Figure 5 shows absolute smoothed disparity search range $\left| \left| \tilde{D}_n \right| \right|$ as a function of the frame index with different k of 1, 2, and 4 in solid, dashed, and dotted lines, respectively. Table 3 shows comparison of E_{RMS} according to k used in a moving average filter with $D^o = 3$. When the value of k is equal to 1, some fluctuations are shown and the estimated disparities are delayed a little bit, which hardly affect the performance of disparity estimation as shown in Table 3. As k increases, the fluctuations become smoother if the estimated disparities are delayed more. Table 3 shows that E_{RMS} slightly decreases as k increases to 4, while significantly increases if k is larger than 4 because absolute smoothed disparity search range is considerably smoothed as shown in Figure 5. So, it is important to select appropriate parameter k for reliable disparity estimation in video. From Figure 5 and Table 3, the search range for disparity estimation should be adaptively determined.

_				k		
•	1	2	3	4	5	 10
 E_{RMS} 	1.41	1.38	1.32	1.29	1.36	 2.15

Table 3. Comparison of E_{RMS} according to k used in a moving average filter with $D^0=3$ for video 3.

The results of the proposed disparity estimation algorithm are compared with two existing algorithms: Chen *et al.*'s algorithm [17] and Xu *et al.*'s algorithm [20]. Chen *et al.*'s algorithm puts the convergence point on the bottom region in each frame of the video. Xu *et al.*'s algorithm puts the convergence point on the foreground in order to eliminate negative parallax for every frames of the video. The proposed algorithm puts the convergence point on the objects that are considered as the ROI.



Figure 5. Absolute smoothed disparity search range as a function of the frame index with different k.

Figures 6–9 show comparisons of result images (top rows) and enlarged ROI images (bottom rows) using the estimated disparity by Chen *et al.*'s algorithm, Xu *et al.*'s algorithm, and the proposed algorithm. Table 4 shows comparisons of estimated disparities of the three algorithms with videos 1–4.

For removing all positive parallaxes, existing algorithms set the convergence point on the nearest object that generally lies at the bottom of the video frame. However, the proposed algorithm sets the convergence point anywhere in the frame.

In Figures 6(a), 6(b), and 6(c), a left image and a shifted right image (video 1, 332th frame) are interlaced by the estimated disparity using Chen *et al.*'s algorithm, Xu *et al.*'s algorithm, and the proposed algorithm, respectively. For better comparison, enlarged results for the ROI (car) are shown in the bottom row. Dotted ellipse on the top row represents the region around convergence point detected by each algorithm. As shown in Figures 6(a) and 6(b), a car in the left image and that in the right image considered as the ROI of two existing algorithms are not correctly overlapped, while those of the proposed disparity estimation algorithm are overlapped correctly as shown in Figure 6(c). When people watch the result image of the proposed disparity estimation algorithm, human eyes can get comfortable 3-D effect without visual fatigue because the

convergence point is set on the ROI by the estimated disparity of 28 (see Table 4), where converged ROI has zero parallax. However, with the result images by two existing algorithms, eyes can get easily visual fatigue because the estimated disparities (81 by Chen et al.'s algorithm and 71 by Xu et al.'s algorithm) produce large positive parallaxes.

•	A1 50	Video			
	Algorithm	1	2	3	4
•	Chen <i>et al.</i> 's algorithm	81	43	30	105
•	Xu et al.'s algorithm	71	57	23	86
•	Proposed algorithm	28	8+	20	72

Table 4. Comparison of estimated disparities used for videos 1-4.



Figure 6. Comparison of result images and enlarged ROI images (car) by the estimated disparity (video 1, 332th frame). Dotted ellipse on the top row represents the region around the convergence point. A left image and a shifted right image are interlaced by the estimated disparity. (a) Estimated disparity using Chen et al.'s algorithm, (b) Estimated disparity using Xu et al.'s algorithm, (c) Estimated disparity using the proposed algorithm.



(a)

(c)

Figure 7. Comparison of result images and enlarged ROI images (car) by the estimated disparity (video 2, 128th frame). Dotted ellipse on the top row represents the region around the convergence point. A left image and a shifted right image are interlaced by the estimated disparity. (a) Estimated disparity using Chen *et al.*'s algorithm, (b) Estimated disparity using Xu *et al.*'s algorithm, (c) Estimated disparity using the proposed algorithm.

Similarly, the result images in Figures 7 (video 2, 128th frame), 8 (video 3, 168th frame), and 9 (video 4, 139th frame) by the proposed disparity estimation algorithm give comfortable 3-D effect because the parallax of the ROIs is to be zero by estimated disparities of 8, 20, and 72 (see Table 4), respectively. However, the result images in Figures 7, 8, and 9 by Chen *et al.*'s and Xu *et al.*'s algorithms give visual fatigue because the estimated disparities shown in Table 4 produce large positive parallaxes. Eyes can get easily visual fatigue because the estimated disparities by the two existing algorithms set the convergence point on locations far from the ROI.



Figure 8. Comparison of result images and enlarged ROI images (face) by the estimated disparity (video 3, 168th frame). Dotted ellipse on the top row represents the region around the convergence point. A left image and a shifted right image are interlaced by the estimated disparity. (a) Estimated disparity using Chen et al.'s algorithm, (b) Estimated disparity using Xu et al.'s algorithm, (c) Estimated disparity using the proposed algorithm.



Figure 9. Comparison of result images and enlarged ROI images (lion doll) by the estimated disparity (video 4, 139th frame). Dotted ellipse on the top row represents the region around the convergence point. A left image and a shifted right image are interlaced by the estimated disparity. (a) Estimated disparity using

Chen et al.'s algorithm, (b) Estimated disparity using Xu et al.'s algorithm, (c) Estimated disparity using the proposed algorithm.

To compare the proposed algorithm with two existing algorithms, subjective evaluation is performed [2], [7]. 45 adults aged from 22 to 32 participated in the subject evaluation. They watched four videos that are randomly ordered at a distance of 2 meter from the 3-D display device. Symptoms of visual fatigue, comfort, and 3-D effect are measured subjectively. Each result is rated by the number 0 to 5, where the score is given highly if the subject feels visual fatigue, comfort, or 3-D effect. Visual fatigue, visual comfort, and 3-D effect represent how much visual fatigue, visual comfort, and 3-D effect are felt by subjects while watching the 3-D contents.

Figure 10 shows subjective evaluation for results of four videos by Chen *et al.*'s, Xu *et al.*'s, and proposed disparity estimation algorithms. Figure 10 shows that the proposed algorithm is more effective to feel visual comfort and 3-D effect than two existing algorithms regardless of the characteristics of the four test videos used in experiments.



Figure 10. Subjective evaluation. (a) Result of video 1, (b) Result of video 2, (c) Result of video 3, (d) Result of video 4.

The performance of the adaptive proposed disparity estimation algorithm is dependent on the accuracy of the object detection and tracking algorithms. In the experiments using object detection and tracking algorithm, the ROI of all frames of each video is accurately detected and tracked.

There is a tradeoff between the performance evaluation E_{RMS} and the computation time. If the size of the window is larger than that used in experiments, the performance will be improved with the computation time increased.

4. CONCLUSIONS

An adaptive disparity estimation algorithm for auto convergence of the ROI in video is proposed using the first-order Taylor expansion and adaptive disparity search range estimate. For accurately estimating disparity using temporal information, disparity of the *n*th frame is approximated by the first-order Taylor series expansion. Next, correcting disparity is obtained by moving average filtering. The estimated disparity of the current frame is obtained by adding approximated disparity and the correcting disparity. In experiments, subjective evaluation of result video shows that the proposed algorithm produces more comfortable 3-D effect than two existing algorithms, in which the convergence point by the proposed algorithm is set on the ROI of each frame. Further work will focus on the development of the window selection method for adaptive representation of ROI shape in disparity estimation.

ACKNOWLEDGEMENTS

This work was supported in part by the MtekVision.

REFERENCES

- Gallagher, A. C.: 'Detecting anaglyph images with channel alignment features'. Proc. IEEE Int. Conf. Image Processing, Hong Kong, Sep. 2010, vol. 5, pp. 2985–2988.
- [2] Xiao, W., Fei, W.: 'A study of stereoscopic display technology and visual fatigue caused by viewing stereoscopic images'. Proc. IEEE Conf. Computing, Control and Industrial Engineering, Wuhan, China, Aug. 2011, vol. 1, pp. 379–382.
- [3] Pastoor, S., Wopking, M.: '3D displays: A review of current technologies', Int. Jour. Display Technology, Apr. 1997, 17, pp. 100–110.
- [4] May, P., 'A survey of 3D display technologies', Int. Jour. Information Display, Apr. 2005, 32, pp. 28–33.
- [5] Kim, C. H., Ra, J. B.: 'Noninteger view multiplexing for 3-D lenticular display'. Proc. IEEE Conf. 3-DTV, Kos Island, Greece, May 2007, pp. 1–4.
- [6] Lipton, L., Feldman, M.: 'A new autostereoscopic display technology: The synthaGram'. Proc. SPIE Photonics West: Electronic Imaging, San Jose, CA, USA, May 2002, vol. 4660, pp. 229–235.
- [7] Kim, D., Sohn, K.: 'Visual fatigue prediction for stereoscopic image', IEEE Trans. Circuits and Systems for Video Technol., Feb. 2011, 21(2), pp. 231–236.
- [8] Lee, J.-H., Song, J.-K.: 'Individual variation in 3D visual fatigue caused by stereoscopic images', IEEE Trans. Consumer Electronics, May 2012, 58(2), pp. 500–504.
- [9] Cookson, C., Joblove, G. H., Hays, D. L.: '3-D auto-convergence camera'. U.S. Patent 8,878,908, Nov. 2014.
- [10] Reza, S.-R., Milivoje, A.: 'Method and apparatus with fast camera auto focus'. U.S. Patent 8,724,013, May 2009.
- [11] Ludwing, L., Joachim, L.: 'Optical stereo device and autofocus method therefor'. U.S. Patent 2 373 043 A1, May 2011.
- [12] Hartley, R. I.: 'Theory and practice of projection rectification', Int. Jour. Computer Vision, Dec. 1999, 36(2), pp. 115–127.
- [13] Papadimitriou, D. V., Dennis, T. J.: 'Epipolar line estimation and rectification for stereo image pairs', IEEE Trans. Image Processing, Apr. 1996, 5(4), pp. 672–676.
- [14] Wu, H.-H. P.: 'Rectification of stereoscopic video for planar catadioptric stereo systems', IEEE Trans. Circuits and Systems for Video Technology, June 2007, 17(6), pp. 686–698.
- [15] Kang, W., Lee, S.: 'Horizontal parallax distortion correction method in toed-in camera with wide-angle lens'. Proc. IEEE Conf. The True Vision–Capture, Transmission and Display of 3D Video, Potsdam, Germany, May 2009, pp. 1–4.
- [16] Zhou, J., Li, B.: 'Image rectification for stereoscopic visualization', Jour. Optical Society of America A, Nov. 2008, 25(11), pp. 2721–2733.
- [17] Chen, W.-C., Hsiao, F.-J., Lin, C.-W.: 'An automatic parallax adjustment method for stereoscopic augmented reality systems'. Proc. Int. Symposium 2010 9th Mixed and Augmented Reality, Seoul, Korea, Oct. 2010, pp. 215–216.
- [18] Lowe, D. G.: 'Distinctive image features from scale-invariant keypoints', Int. Jour. Computer Vision, Oct. 2004, 60(2), pp. 91–110.
- [19] Lindeberg, T.: 'Scale-space theory: A basic tool for analyzing structures at different scales', Jour. Applied Statistics, 1994, 21(2), pp. 225–270.
- [20] Xu, D., Corea, L. E., Nasiopoulos, P.: 'Quality of experience for the horizontal pixel parallax', Proc. 2012 Digest of Technical Papers Int. Conf. Consumer Electronics, Las Vegas, NV, USA, Jan. 2012, pp. 394– 395.
- [21] Lin, C.-W., Fei, E.-Y., Chen, Y.-C.: 'Hierarchical disparity estimation using spatial correlation', IEEE Trans. Consumer Electronics, Aug. 1998, 44(3), pp. 630–637.
- [22] Brown, M. Z., Burschka, D., Hager, G. D.: 'Advances in computational stereo', IEEE Trans. Pattern Analysis and Machine Intelligences, Aug. 2003, 25(8), pp. 993–1008.
- [23] Geiger, A., Roser, M., Urtasun, R.: 'Efficient large-scale stereo matching computer vision'. Proc. 10th Asian Conf. Computer Vision, Lecture Notes in Computer Science, Queenstown, New Zealand, Nov. 2010, vol. 6492, pp. 25–38.
- [24] Viola, P., Jones, M.: 'Robust real-time face detection', Int. Jour. Computer Vision, May 2004, 57(2), pp. 137–154.
- [25] Comaniciu, D., Ramesh, V., Meer, P.: 'Kernel-based object tracking', IEEE Trans. Pattern Analysis and Machine Intelligence, May 2003, 25(5), pp. 564–575.
- [26] Jeyakar, J., Babu, R. V., Ramakrichnan, K. R.: 'Robust object tracking using local kernels and background information'. Proc. 2007 IEEE Int. Conf. Image Processing, San Antonio, TX, USA, Sep. 2007, vol. 5, pp. 49–52.