Real-Time Stereo Using Foreground Segmentation and Hierarchical Disparity Estimation

Hansung Kim, Dong Bo Min, and Kwanghoon Sohn

Dept. of Electrical and Electronics Eng., Yonsei University,
134 Shinchon-dong, Seodaemun-gu, Seoul 120-749, Korea
khsohn@yonsei.ac.kr
http://diml.yonsei.ac.kr

Abstract. We propose a fast disparity estimation algorithm using background registration and object segmentation for stereo sequences from fixed cameras. Dense background disparity information is calculated in an initialization step so that only disparities of moving object regions are updated in the main process. We propose a real-time segmentation technique using background subtraction and inter-frame differences, and a hierarchical disparity estimation using a region-dividing technique and shape-adaptive matching windows. Experimental results show that the proposed algorithm provides accurate disparity vector fields with an average processing speed of 15 frames/sec for 320x240 stereo sequences on a common PC.

1 Introduction

One of the most important problems in 3D image processing is to locate corresponding points in the images, a process referred as disparity estimation. As shown in Fig. 1, stereo imaging involves two separate image views of a single world point w. The objective is to find the corresponding pair I_1 and I_2 in the

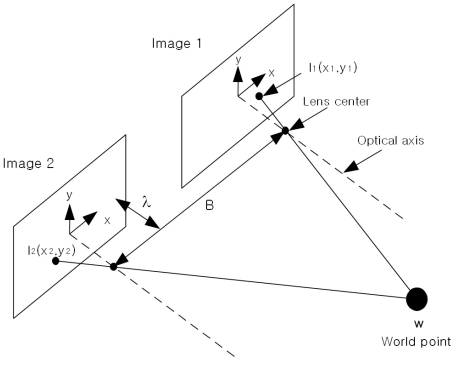


Fig. 1. Stereo geometry

Y.-S. Ho and H.J. Kim (Eds.): PCM 2005, Part I, LNCS 3767, pp. 384–395, 2005.
© Springer-Verlag Berlin Heidelberg 2005

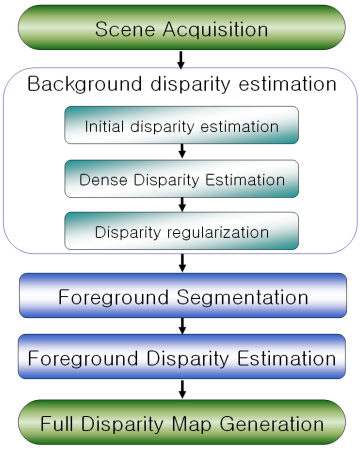


Fig. 2. Block diagram of the proposed algorithm

image pair. If we assume that the cameras are identical and the coordinate systems of both cameras are aligned in parallel, the determination of disparity from I_1 to I_2 becomes finding a function $d(x,y)$ such that:

$$I_2(x, y) = I_1(x + d(x, y), y) \tag{1}$$

A number of studies have been reported on the correspondence problem since the 1970's. D. Scharstein and R. Szeliski recently discussed the taxonomy of existing stereo algorithms [1] and a test bed for the quantitative evaluation of the algorithms [2]. However, most of them have serious limitations on being applied to common applications since they do not work in real-time. Several real-time methods were recently proposed [3][4][5], but they were implemented on DSP for acceleration or show poor quality to be used for wide-ranging applications.

We have previously proposed a two-stage algorithm to find smooth and precise disparity vector fields in a stereo image pair [6]. The algorithm has consisted of a dense disparity estimation and edge-preserving regularization. It results in such a clean disparity map with good discontinuity localization, but the computational cost is so high that it does not work in real-time. In this paper, we propose a fast disparity estimation algorithm using background registration and object segmentation. We assume that a stereo camera set does not move, and there is no moving object for a few seconds in an initialization step for generating background information. Accurate and detailed disparity information for background is estimated in advance, then only disparities of moving foreground regions are calculated and merged into background disparity fields.

Fig. 2 shows a block diagram of the proposed system. As a preprocessing, acquired image sequences are low-pass filtered to reduce noise effect and rectified since we assume that stereo images are captured in parallel stereo cameras in disparity estimation. We use a real-time stereo rectification function provided by Triclops SDK [7].

2 Foreground Segmentation

Real-time foreground segmentation is one of the most important components of the proposed system, since the performance of the segmentation decides the efficiency and quality of the final disparity fields. We propose a foreground segmentation technique using background subtraction and inter-frame differences based on the technique which we have previously proposed [8]. Fig. 3 shows overall segmentation process. At first, the background masks $I_{min}(x,y)$and $I_{max}(x,y)$ are modeled with minimum and maximum intensities of the first N frames, respectively, because the background information is very sensitive to noise and change of illumination. Then, the frame difference mask $I_{fd}(x,y)$ is calculated by the difference between two consecutive frames. In the third step, an initial foreground mask is constructed from the frame difference and background difference masks by the OR process, that is, if a pixel of current frame satisfies one of Eq. (2), it is determined to be belonged to an initial foreground region. Th_{tol} and Th_{fd} mean threshold values for background and frame difference regions, respectively.

$$\begin{aligned} I_{cur}(x,y) &< I_{min}(x,y) - Th_{tol} \\ I_{cur}(x,y) &> I_{max}(x,y) + Th_{tol} \\ I_{fd}(x,y) &> Th_{fd} \end{aligned} \tag{2}$$

However, due to the camera noise and irregular object motion, there exist some noise regions in the initial mask. One of the conventional ways to eliminate the noise regions is using the morphological operations to filter out small regions. Therefore, we refine the initial mask by a closing process and eliminate small regions with a region-growing technique.

Finally, in order to smooth the boundaries of foreground and to eliminate holes inside the regions, we propose a profile extraction technique. This technique is remodeled from Kumar's profile extraction technique [9]. A weighted one pixel thick drape moves from one side to the opposite side. The adjacent pixels of the drape are connected by elastic spring, so it covers object but does

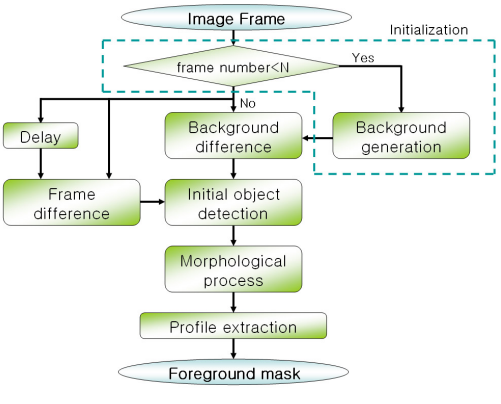


Fig. 3. Real-time segmentation algorithm

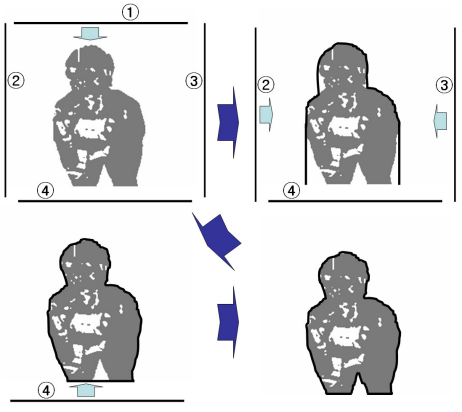
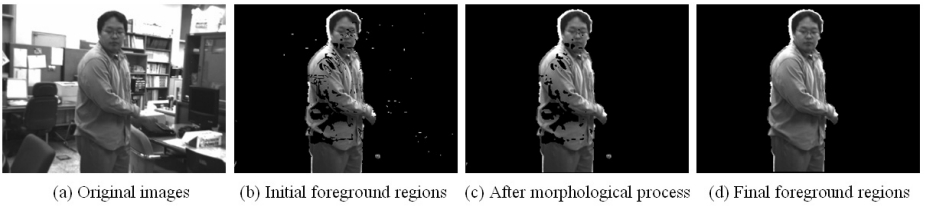


Fig. 4. Profile extraction process



(a) Original images (b) Initial foreground regions (c) After morphological process (d) Final foreground regions

Fig. 5. Segmentation results

not infiltrate into gaps whose widths are smaller than a threshold M. This process is performed from all quarters and the region wrapped by four drapes is decided as a final foreground region. Fig. 4 shows the profile extraction process applied to an initial object.

Segmentation results by the proposed method are shown in Fig. 5. The image is captured in typical office environment without any special lighting equipment. Fig. 5 (b) is the result of initial object detection from Fig. 5 (a). Main object are detected well, but they include noises on background and object boundaries. In Fig. 5 (c), we can see that noises are eliminated and object surfaces are smoothed by a morphological process. However, many holes still exist inside the objects. Fig. 5 (d) is the final segmentation result. After applying the profile extraction technique, good semantic foreground regions are obtained.

3 Disparity Estimation

3.1 Background Disparity Estimation

In windows-based algorithms, the reliability and efficiency depend on the size of a matching window. Large window sizes provide reliable but not detailed results. Moreover, employing a large window for each pixel in dense disparity estimation

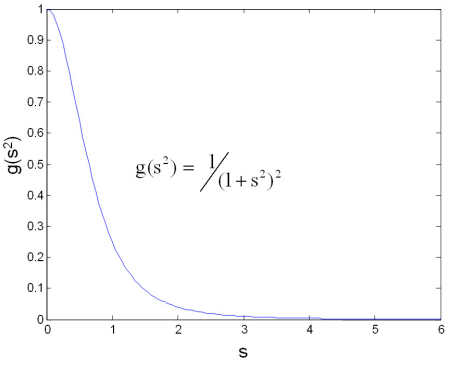


Fig. 6. Behavior of the diffusivity function

increases the computational load. Therefore, in the proposed algorithm, dense disparity information of background is initially estimated in a hierarchical way.

The first step in hierarchical estimation is a BxB block-based initial disparity estimation. In the second step, dense disparity vectors for each pixel are estimated based on the initial block vectors. In order to cover all the probable disparity candidates, 9 initial vectors (1 from the current block and 8 from neighboring blocks) are tested within a small search range α from the vector. In order to improve computational efficiency in disparity estimation, we use a region-dividing technique which we have previously proposed [6]. The technique performs point matching in the order of the possibility of correct matching and divides the region into sub-regions at the true matching point. After the region splits into two sub-regions in matching process, the search ranges of points in each sub-region are restricted to the corresponding sub-region.

However, in the disparity vectors estimated by the above-described method, spatial correlation of the estimated vector fields is not considered. In order to provide more accurate and reliable background disparity fields, we refine the fields by regularizing them by means of the following nonlinear diffusion equation with an additional fidelity term [6].

$$\begin{aligned} \frac{\partial d}{\partial t} &= \lambda \; div\left(g(|\; \triangledown I_l(x,y)\;|^2)\triangledown d(x,y)\right) \\ &\quad + (I_l(x,y) - I_r(x+d,y))\frac{\partial I_r(x+d,y)}{\partial x} \\ where \;\; g(s^2) &= 1/(1+s^2)^2 \end{aligned} \tag{3}$$

$g(|\triangledown I_l|^2)$ is a diffusivity function which plays the role of discontinuity marker. Fig. 6 shows the behavior of the function $g(|\triangledown I_l|^2)$. Therefore, this function reduces smoothing on object boundaries to preserve their discontinuities. In order to solve Eq. (3), we discretize the parabolic system by finite differences, and find the regularized disparity field in recursive manner by updating the field using Eq. (4).

$$\frac{d^{k+1}(x,y)-d^{k}(x,y)}{\tau} = \lambda \left\{ \frac{\partial}{\partial x}\left(g\left(\left| \frac{\partial I_l(x,y)}{\partial x} \right|^2 \right) \times \frac{\partial d^k(x,y)}{\partial x} \right) + \frac{\partial}{\partial y}\left(g\left(\left| \frac{\partial I_l(x,y)}{\partial y} \right|^2 \right) \times \frac{\partial d^k(x,y)}{\partial y} \right) \right\} + \left(I_l(x,y) - I_r(x+d^k,y) \right) \times \frac{\partial I_r(x+d^k,y)}{\partial x} + \left(d^k(x,y) - d^{k+1}(x,y) \right) \times \left(\frac{\partial I_r(x+d^k,y)}{\partial x} \right)^2 \quad (4)$$

3.2 Foreground Disparity Estimation

The most important requirement of foreground disparity estimation is a processing speed because the fields of foreground must be updated in every frame. Hierarchical disparity estimation used in background disparity estimation is applied to the blocks which include foreground regions except a regularization step. Initial search ranges are also restricted by the neighbor background disparities since the foreground objects always exist in front of background region. Eq. (5) shows the search range decision where SR_{Max} and SR_{Min} mean maximum and minimum search range, respectively, and d_{ln} and d_{rn} are left and right neighboring background disparities of the foreground region on the same scanline.

$$\begin{aligned} &\textbf{for L} \rightarrow \textbf{R disparity} \\ &\qquad SR_{max} = Min(d_{ln}, d_{rn}) \\ &\textbf{for R} \rightarrow \textbf{L disparity} \\ &\qquad SR_{min} = Max(d_{ln}, d_{rn}) \end{aligned} \quad (5)$$

As a result, search ranges are restricted by three factors: background disparity, region-dividing technique and hierarchical estimation. Thus, the processing time of foreground estimation is greatly reduced.

In matching process, however, conventional rectangular window yield false result around object boundaries because the result is highly influenced by strong feature. In background disparity estimation, wrong disparities around the regions are corrected by regularization, but it can result in errors in foreground estimation. For example, in the cases of points A and B in Fig. 7, although they belong to different regions, the same disparity vectors are assigned because of the strong edge between them. In order to avoid this type of problem, we propose a new matching window which provides a high degree of reliability around the boundary region by deforming its shape according to the flow of the features. Let Ω denote the contour of the matching window. Starting from a sufficiently small contour Ω_0, the contour expands in the direction of non-increasing $|\nabla I|$ until a maximum size NxN is reached. Fig. 8 shows an example of window generation in the 1D case. The window does not cross strong feature so that the correct sharp boundary of disparity vectors can be obtained, as shown in Fig. 9, where

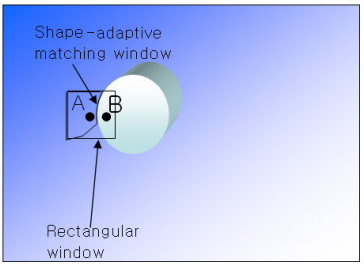


Fig. 7. Rectangular window and the proposed window

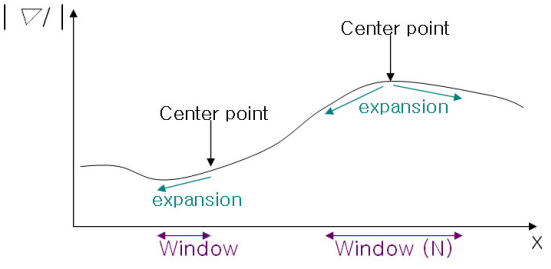


Fig. 8. Window generation in 1D case

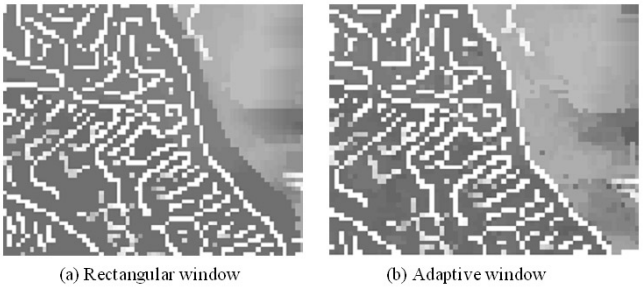


Fig. 9. Matching results using a rectangular window and the proposed window

white lines represent the real edges of the object. However, the adaptive window may decrease the matching power in highly textured regions. Thus, the shape-adaptive window is applied only for pixels in the block, where the maximum difference of disparity to other surrounding blocks is larger than ϵ.

Finally, estimated foreground disparity fields are merged into background disparity fields. We check the reliability of the disparity for the pixels in boundary blocks which include the boundary between background and foreground in order to compensate errors induced by wrong foreground segmentation. Final disparities of the pixels in boundary blocks are determined by the following conditions.

$$\textbf{if}(|\mathbf{I_r}(\mathbf{x},\mathbf{y}) - \mathbf{I_l}(\mathbf{x}+\mathbf{d_{fore}},\mathbf{y})| < |\mathbf{I_r}(\mathbf{x},\mathbf{y}) - \mathbf{I_l}(\mathbf{x}+\mathbf{d_{back}},\mathbf{y})|)$$
$$d_{final}(x,y) = d_{fore}(x,y) \quad (6)$$
$$\textbf{else}$$
$$d_{final}(x,y) = d_{back}(x,y)$$

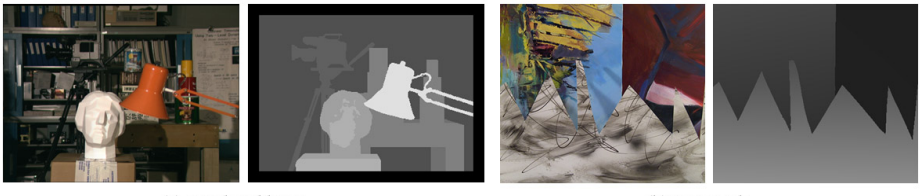
4 Simulation Results

The proposed algorithm is applied to stereoscopic sequences captured by Digiclops which provides a rectified stereo sequence with a speed of 30 frames/sec [7]. The size of images is 320x240 and we used a PC with a Pentium IV 3.0 GHz CPU and 512 Mbytes memories. The parameters used in the simulation are listed in Table 1.

At first, we compared the performance of the proposed algorithm with other 4 fast algorithms in Table 2. For the objective evaluation, we applied the algorithm to the still images of Fig. 10 provided on Scharstein's homepage with ground truth disparity maps [2], and compared accuracy of the estimated disparity fields. We used two measures of quality. The first is BMP (bad matching percentage) of the estimated disparity map employed by Zitnick and Kanade [10], which is defined as:

Table 1. Parameters used in simulation

Stage	Parameter	Values
Foreground segmentation	Background generation	N = 50
	Background difference	$\mathrm{Th}_{tol} = 10$
	Frame difference	$\mathrm{Th}_{fd} = 5$
Disparity estimation	Block size	B = 8
	Dense disparity range	$\alpha = 2$
	Shape-adaptive window	$\varepsilon = 2$
Disparity regularization	Lagrange multiplier	$\lambda = 2000$
	Time step size	$\tau = 0.0001$
	Number of iteration	T = 150



(a) Head and lamp (b) Sawtooth

Fig. 10. Test images and true disparity fields

Table 2. Comparative performance of algorithms

	Bad Matching Percentage (%)		RMSE (pixel)	
	Head and lamp	Sawtooth	Head and lamp	Sawtooth
Multi-window [11]	4.48	2.18	1.3980	1.2973
Max-Surface [12]	9.25	6.72	1.5294	1.6933
Real-time-DP [4]	4.22	6.11	1.1255	1.7542
MMHM [5]	8.00	3.03	1.6242	1.3069
Hierarchical	5.22	2.46	1.1047	1.3028
Final disparity	4.07	2.25	0.9193	0.9094

$$B = \frac{1}{N} \sum_{x,y} \delta \left(d_e(x,y), d_T(x,y) \right) \tag{7}$$

$$\text{where } \delta(a,b) = \begin{cases} 1 & , \quad \text{if } |a-b| > 1 \\ 0 & , \quad \text{else} \end{cases}$$

The second is RMSE (Root-Mean-Squared Error) of the estimated map. The RMSE between the estimated map $d_e(x,y)$ and the ground truth map $d_T(x,y)$ can be calculated by:

$$RMSE = \left(\frac{1}{N} \sum_{x,y} \left(d_e(x,y) - d_T(x,y) \right)^2 \right)^{1/2} \tag{8}$$

The proposed algorithm does not deal with a boundary problem, thus a border of 20 pixels was excluded from the evaluation.

In Table 2, the "Hierarchical" row means the results before regularization, that is, we can regard them as a performance of foreground estimation though the effect from the segmentation is not considered. In the BMP evaluation, the results of applying the proposed algorithm are somewhat inferior to several algorithms in the "Head and lamp" images, and it is a good second to the graph cut algorithm in the "Sawtooth." However, the proposed algorithm gives the best results in the RMSE category. Figs. 11 and 12 show the disparity maps of the "Sawtooth" and the "Head and lamp," respectively. In examining the results, the multi-window and the real-time DP algorithms are superior in terms of finding discontinuities, but they have problems in error propagation in the horizontal direction. The max-surface and the MMHM algorithms show a good result with the "Sawtooth," but produces prominent errors in some regions in the case of the "Head and lamp." The proposed algorithm results in reasonably clean maps with good discontinuity localization. However, the algorithm fails to find disparity in a narrow background such as the area between the arms of the lamp.

Table 3 shows the average running time analysis of our algorithm when one person moves in a scene. The system requires about 6-7 seconds for initialization before it works. After that, our algorithm shows an average speed of 15 frames/sec. According to referenced papers, Multi-window shows about 5 frames/sec, Max-Surface 2 frames/sec, Real-time-DP 8 frames/sec without

Table 3. Processing speed (msec)

Stage	Step	Time
Initialization	Background generation	1852
	Background disparity Estimation	5156
Main processing	Capturing and rectification	28.26
	Initial segmentation	9.69
	Morphological process	4.64
	Silhouette extraction	6.45
	Disparity estimation	17.81
	Total	66.85

MMX optimization, and MMHM 5 frames/sec. Considering both processing speed and quality of disparity fields, the proposed algorithm shows the best results.

Fig. 13 is the snapshot of test sequence and estimated background disparity fields. The image sequences are captured in natural condition without any special lighting equipment or any arrangement of objects for extracting good results. We can see that the proposed algorithm results in such a clean map with good discontinuity localization. Fig. 14 show several frames from the resulting sequences; the left one is segmented foregrounds and the right one final disparity fields in each pair. In the final disparity fields, we can easily imagine a 3D structure of the scene.

5 Conclusion

In this paper, we propose a real-time disparity estimation algorithm using background registration and foreground segmentation. Dense background disparity

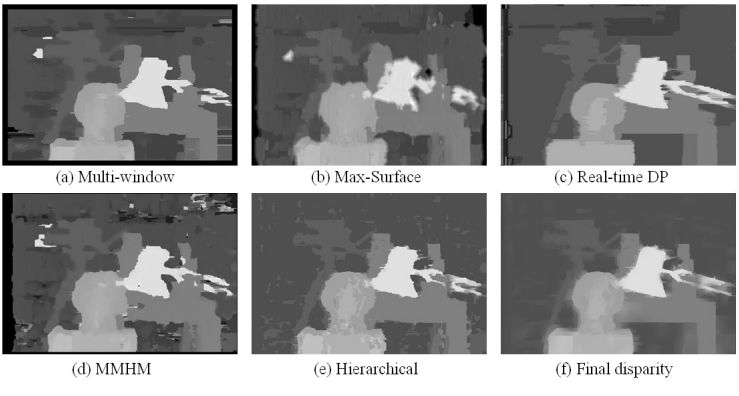


Fig. 11. Disparity fields of "Head and lamp"

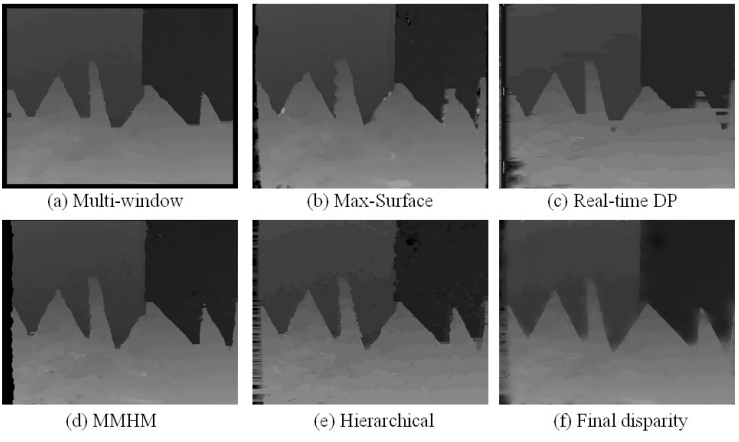


Fig. 12. Disparity fields of "Sawtooth"

Fig. 13. Test sequence and estimated background disparity

Fig. 14. Results of foreground segmentation and final disparity

information is calculated in advance and only disparities of moving object regions are updated in the main process. For efficient and accurate estimation, a real-time segmentation algorithm, hierarchical disparity estimation and shape-adaptive windows are proposed. The performance of the proposed algorithm was

evaluated in objective and subjective ways. Computation time mainly depends on the image size, and it was about 15 frames/sec for image pairs having a resolution of 320x240 on a common PC.

As a future work, we have to develop more powerful segmentation algorithm. The performance of the segmentation decides the efficiency and quality of the final disparity fields. Especially, foreground regions classified into background due to wrong segmentation make serious errors in final fields since the fields are not updated. The second perspective of our work will be to improve accuracy of disparity fields at object boundary regions. It is also planned to develop a complete 3D modeling algorithm from multiple stereo cameras. We are currently investigating a depth fields merging algorithm with camera calibration.

Acknowledgements

We would like to thank Dr. D. Scharstein and Dr. R. Szeliski for supplying the ground truth data on their homepage, and Dr. Y. Ohta and Dr. Y. Nakamura for the imagery from the University of Tsukuba. This work was supported by the Ministry of Information and Communication, Korea, under Information Technology Research Center (ITRC) Support Program.

References

1. Scharstein, D., Szeliski, R.: A Taxonomy and Evaluation of Dense Two-frame Stereo Correspondence Algorithms. IJCV, Vol.47, (2002) 7-42
2. http://www.middlebury.edu/stereo
3. Schreer, O., Brandenburg, N., Kauff, P.:Real-time Disparity Analysis for Applications in Immersive Teleconference Scenarios - a Comparative Study. Proc. ICIAP (2001) 346-351
4. Forstmann, S., Ohya, J., Kanou, Y., Schmitt, A., Thuering, S.: Real-time Stereo by Using Dynamic Programming. Proc. CVPR (2004) pp.29
5. Muhlmann, K., Maier, D., Hesser, J., Manner, R.: Calculating Dense Disparity Maps from Color Stereo Images, an Efficient Implementation. IJCV, Vol.47 (2002) 79-88
6. Kim, H., Choe, Y., Sohn, K.: Disparity Estimation Using Region-dividing Technique with Energy-based Regularization. Optical Engineering, Vol.43, No.8 (2004) 1882-1890
7. http://www.ptgrey.com/
8. Kim, H., Kitahara, I., Kogure, K., Hagita, N., Sohn K.: Sat-Cam: "Personal Satellite Virtual Camera". Proc. PCM, Vol.3 (2004) 87-94
9. Kumar, P.,Sengupta, K., Ranganath, S.: Real Time Detection and Recognition of Human Profiles using Inexpensive Desktop Cameras. Proc. ICPR, Vol.1 (2000) 1096-1099
10. Zitnick, L., Kanade, T.: A Cooperative Algorithm for Stereo Matching and Occlusion Detection. IEEE Trans. PAMI, Vol.22, No.7 (2000) 675-684
11. Hirschmuller, H.: Improvements in Real-time Correlation-based Stereo Vision. Proc. CVPR Stereo Workshop (2001) 141-148
12. Sun, C.: Fast Stereo Matching Using Rectangular Subregioning and 3D Maximum-surface Techniques. IJCV, Vol.42, No.1 (2002) 7-42