Fast correlation-based stereo matching with the reduction of systematic errors

Sukjune Yoon *, Sung-Kee Park, Sungchul Kang, Yoon Keun Kwak

Department of Mechanical Engineering, Korea Advanced Institute of Science & Technology, 373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Republic of Korea

Abstract

The purpose of this work is to develop a stereo matching algorithm that produces a dense disparity map for a real time vision system. Generally, correlation-based correspondence algorithms produce systematic errors at depth discontinuities, including half occluded areas, and require significant computational power. To solve these problems, we identify the properties of these areas containing matching errors and carry out an error compensation process with an effective implementation.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Stereo vision; Dense disparity map; Real time stereo; Reduction of errors

1. Introduction

The needs of a real time stereo vision system have increased in the field of robotics because it can give geometrical information. In this system, stereo matching is one of the most important and time consuming processes in these systems.

A number of stereo matching algorithms are available for finding corresponding points in two image pairs. These algorithms could be classified into two categories: area-based methods (e.g. Kanade and Okutomi, 1994) and feature-based methods (e.g. Chen et al., 1999). The former are superior to the latter for robotic applications, as these methods can find a dense disparity map from two stereo images, in contrast to the feature-based methods, which can give only a sparse disparity map. However, these methods require large computing power and yield systematic errors such as blurred errors

* Corresponding author. Tel.: +82 42 869 5212; fax: +82 42 869 5201.

E-mail addresses: inca4004@kaist.ac.kr, inca4004@daum.net (S. Yoon).

at depth discontinuities. Therefore, an algorithm that can compensate for these defects and calculate a dense disparity map in short time is strongly required in order to apply these area-based stereo matching algorithms to real systems.

Matthies (1992) developed a real time stereo vision system by a large hardware system. Konolige (1997) developed the Small Vision Module (SVM) and Hariyama et al. (2001) proposed the VLSI processor for stereo matching. These developed methods have been implemented by special hardware, and as such they are not suitable for complicated algorithms. Faugeras et al. (1993) have developed a real time correlation-based stereo algorithm. They proposed an optimized technique that reduces the computation time by removing redundant calculations. Mühlmann et al. (2002) and Schmidt et al. (2002) have developed a real time method that calculates a dense disparity map.

The aforementioned methods have been developed based on the assumption that all the points in the correlation window have the same disparities; however, this assumption is unavoidably violated at depth discontinuities. Therefore, several algorithms have been proposed to overcome this systematic error. Kanade and Okutomi (1994) proposed an adaptive window method that selects window size depending on local variations of intensity and disparity. However, this algorithm is inadequate for real time systems due to large computing time. Fusiello et al. (1997) proposed an efficient multiple windowing method applicable to a real time system. In this method, correlation calculations are performed with all nine windows for every pixel and every disparity, and subsequently only the result of the best correlation value disparity is adapted. Hirschmüller et al. (2002) have proposed a real time algorithm that uses multiple supporting windows with border correction. A temporary disparity map is obtained by multi supporting correlation windows and then an error reduction process is performed at the depth discontinuities. However, these methods do not systemically deal with half occlusion areas, wherein mismatched errors mainly arise.

In this paper, we propose an error reduction method that can compensate for the mismatched errors at half occluded regions and blurred errors

of object borders at once. To reduce systematic errors, first the regions containing matching errors in the disparity map are identified and then an error reduction process is then performed with different sized correlation windows. Sum of absolute difference (SAD) with an optimized technique (Faugeras et al., 1993) and single instruction and multiple data (SIMD) techniques are implemented to reduce computation time. In Section 2, the problems of correlation-based stereo matching at depth discontinuities and half occluded regions are analyzed. A proposed algorithm that reduces systematic errors and calculates a dense disparity map at 7–10 frame/s is explained in Section 3. The results of this algorithm and evaluation are presented in Section 4, and conclusions are finally presented in Section 5.

2. Error analysis

2.1. Blurred errors at depth discontinuities

The area-based correlation method assumes that image points in the correlation window have constant disparities, but this assumption is violated at depth discontinuities (Kanade and Okutomi, 1994). Fig. 1(a) and (b) is the stereo image pair and Fig. 1(c) is the disparity map. As shown in Fig. 1(c), the disparity map is blurred at the depth discontinuities.

The range of blurring error at these points depends on the size of the correlation window, and the maximum blurring error is less than the size of the correlation window. To reduce such errors, a small correlation window can be used. However, in this case, the disparity map is sensitive to image noise (Kanade and Okutomi, 1994). Therefore, it produces numerous matching errors in a real image. In this paper, to reduce blurring errors at depth discontinuities, correlation values with different sized windows are calculated. This method effectively utilizes both large window correlation and small window correlation methods.

2.2. Mismatched errors at half occlusion

The half occluded areas appear in only one of the two cameras. Shown in Fig. 2, the half oc-
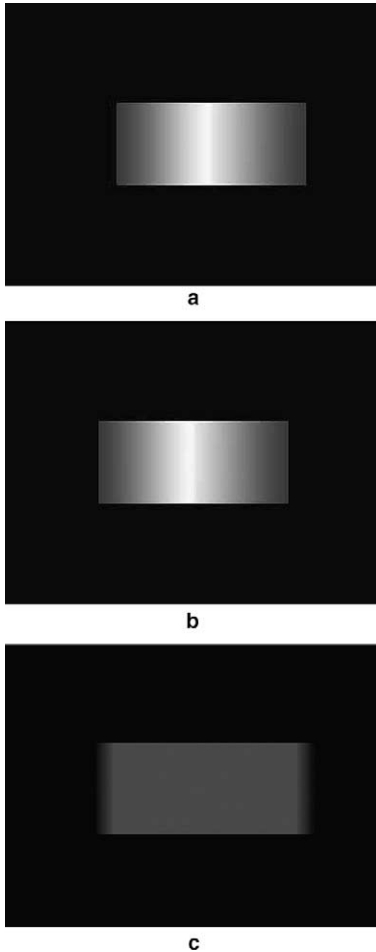
Fig. 1. (a) Synthesized left image, (b) synthesized right image, (c) dense disparity map by 9×9 SAD.

cluded area (A) will appear in the left image only, and vice versa in the case of half occluded area (B).

If these areas are not identified, the points in the half occlusion area will be mismatched, because these points have no corresponding points. In this paper, we identify these erroneous areas, and then refine the disparities in these regions.

3. Proposed correlation-based stereo matching algorithm

The proposed algorithm is composed of four stages. First, we construct a disparity space image (DSI) from stereo image pair, and then we calculate temporary disparity maps by the SAD method, which is implemented by an optimized technique (Faugeras et al., 1993) and the single instruction and multiple data (SIMD) technique to reduce computation time. The errors in this map are filtered by a left/right consistency check (Fua, 1993). The third stage reduces both the blurred errors at depth discontinuities and the mismatched errors at half occluded areas. Finally, we use a median filter to interpolate the dense disparity map. The process of our proposed algorithm is shown in Fig. 3. Left and right images are rectified by a compact rectification (Fusiello et al., 2000).

3.1. Disparity space image

We choose the SAD method to calculate correlation values, because it requires only ± arithmetic operation and performs better than the sum of square difference (SSD) approach in the presence of outliers (Mühlmann et al., 2002). To implement the algorithm effectively, we construct a three-dimensional DSI (e.g. Mühlmann et al., 2002). These cuboid dimensions are given by the width and height of the images and the disparity search range. Every position (x, y, d) in this cuboid memory denotes the absolute difference value between the value at the position (x, y) in the left image and the value at the position $(x + d, y)$ in the right image (Fig. 4).

3.2. Initial dense disparity map

After the DSI is calculated, an initial dense disparity map is computed. An array DSI (x, y, d) has an absolute difference value at the position of (x, y) and disparity d. Therefore, we can easily implement an optimized technique (Faugeras et al., 1993) that reduces computation time. Eq. (1) expresses the mathematical form of SAD.

$$\text{SAD}(x, y, d) = \sum_{i=-\frac{1}{2}(\text{win}_x-1)}^{\frac{1}{2}(\text{win}_x-1)} \sum_{j=-\frac{1}{2}(\text{win}_y-1)}^{\frac{1}{2}(\text{win}_y-1)} |I_{\text{L}}(x + i, y + j) - I_{\text{R}}(x + i + d, y + j)| \quad (1)$$
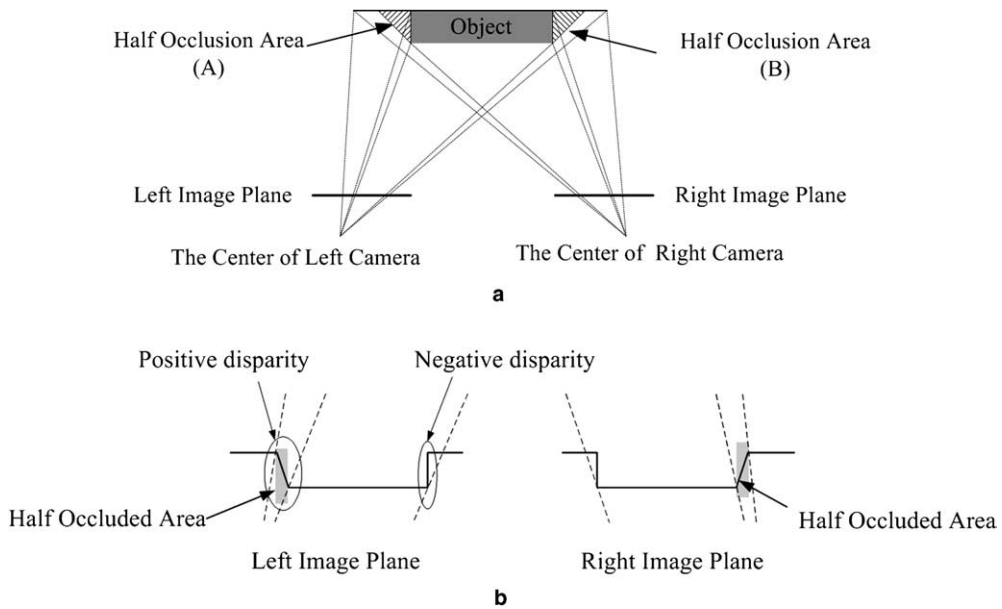
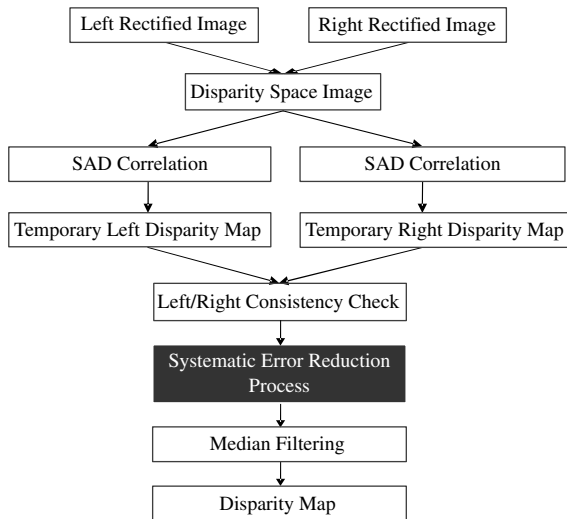Fig. 2. (a) Example of half occlusion, (b) images at left and right image plane.

Fig. 3. Process of calculating dense disparity map from two rectified images with the systematic error reduction.

Fig. 4. Three-dimensional disparity space image (Mühlmann et al., 2002).

$$d(x,y) = \underset{d_{\min} \leqslant d \leqslant d_{\max}}{\arg \min} \ \mathrm{SAD}(x,y,d) \qquad (2)$$

Then, we apply a left/right consistency check (Fua, 1993). A valid correspondence should match in both directions. The left and right disparities should satisfy Eq. (3). $D_{\mathrm{R}}(x,y)$ is the disparity at (x,y) in the right image and $D_{\mathrm{L}}(x,y)$ is the disparity at (x,y) in the left image. If Eq. (3) is not satisfied, the $D_{\mathrm{L}}(x,y)$ is invalid and assigned -1.

$$D_{\mathrm{R}}(x + D_{\mathrm{L}}(x,y), y) = -D_{\mathrm{L}}(x,y) \qquad (3)$$

The disparity at (x,y), $d(x,y)$, is the point where the correlation value of SAD is the smallest. We use a winner-take-all algorithm for a fast running time. Therefore, the disparity value is as follows:
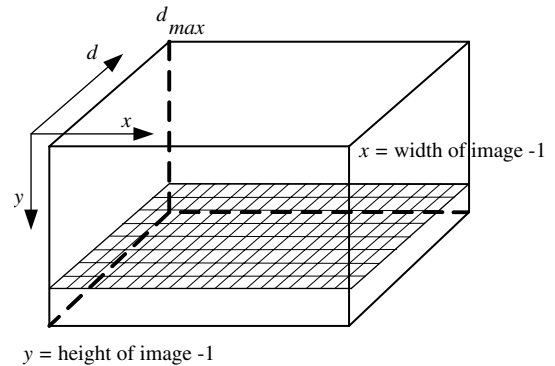
3.3. Reduction of systematic errors

To reduce blurring errors at depth discontinuities and mismatched errors at half occlusions, first we identify the areas where the disparities are refined. As shown in Fig. 2(b), we can analogize that the half occluded areas belong to depth discontinuities areas. After identifying the areas to be modified, we reduced simultaneously the blurred errors by small sized correlation window and the mismatched errors by edge sensitive correlation method. The proposed method is following.

Step 1: Find the positions of positive disparity in the right image to modify the left disparity map.

Step 2: Split the correlation window into two small sized correlation windows, $R_L(i,j)$ and $R_R(i,j)$ at positive disparity in the right image. Fig. 5 shows the small sized sub-windows. In the case of 5×5 sub-window, con-

sidering the blurring errors, the center of $R_L(i,j)$ is $(x - 3 - (\text{win}_x - 1), y)$ and the center of $R_R(i,j)$ is $(x + 3 + (\text{win}_x - 1), y)$.

Step 3: Identify the area to be modified in the left disparity map. The amount of blurring error at depth discontinuities is less than the size of the correlation window. In the case of 5×5 sub-correlation windows, the range to be modified in the left disparity map is determined from $(x - 3 - (\text{win}_x - 1) + dr_b, y)$ to $(x + 3 + (\text{win}_x - 1) + dr_o, y)$, where dr_o is the disparity of the object in the right image, dr_b is the disparity of the background in the right image, and win_x is the size of the correlation window. As shown in Fig. 6, dr_b is $D_R(x - 3 - (\text{win}_x - 1), y)$ and dr_o is $D_R(x + 3 + (\text{win}_x - 1), y)$. Fig. 6 shows the area to be modified and the initial position of left sub-windows, $L_L(i,j)$ and $L_R(i,j)$.
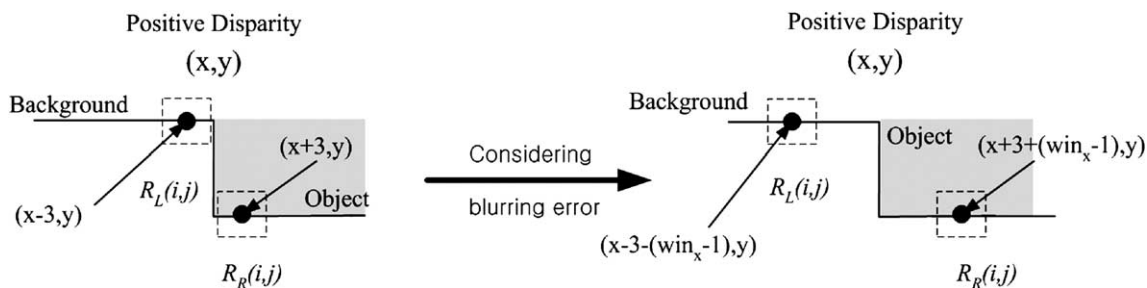

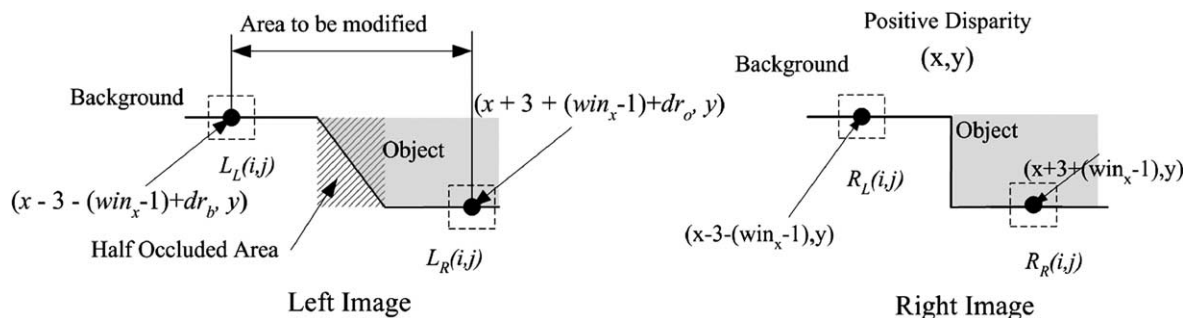Fig. 5. The position of the sub-windows, $R_L(i,j)$ and $R_R(i,j)$, in the right image.

Fig. 6. Identification of the areas to be modified and the initial position of sub-windows $L_L(i,j)$ and $L_R(i,j)$, in the left image.
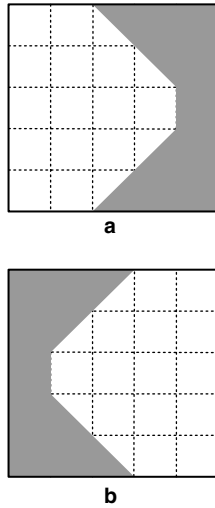
Fig. 7. (a) Error reduction process stops if the position of the mass center is located in the gray area on the left side at positive disparity. (b) Error reduction process stops if the position of the mass center is located in the gray area on the right side at positive disparity.

Step 4: Transformed values by Eq. (3) in the $L_L(i,j)$ and $R_L(i,j)$ to increase sensitivity to object edges and reduce the effect of

noise. $R_L(i,j)$ and $L_L(i,j)$ are transformed to $R'_L(i, j)$ and $L'_L(i, j)$ by Eq. (3), where C_{max} is the maximum value in the window, C_{ave} is the average value in the window. The meaning of the value 128, half of the intensity level, in Eq. (3) is threshold value, we got this value by experimentally.

$$C'_L(i,j) = \exp\left(\frac{C_{max}}{128} \cdot \frac{C(i,j)}{C_{ave}}\right) \tag{3}$$

Step 5: Refine the disparities. Determine the disparity at $(x - 3 - (win_x - 1)dr_b, y)$, the center of $L'_L(i, j)$, as dl_b in the left image, if the position of the mass center of the correlation value between $R'_L(i, j)$ and $L'_L(i, j)$ is not located in the gray area in Fig. 7(a). And then shift $L'_L(i,j)$ to the right and determine the disparity of this position as dl_b in the left image until the position of the mass center of the correlation value between $R'_L(i, j)$ and $L'_L(i, j)$ is located in the gray area in Fig. 7(a). In the case of a 5×5 sub-window, dl_b, the disparity of the background in the left image,
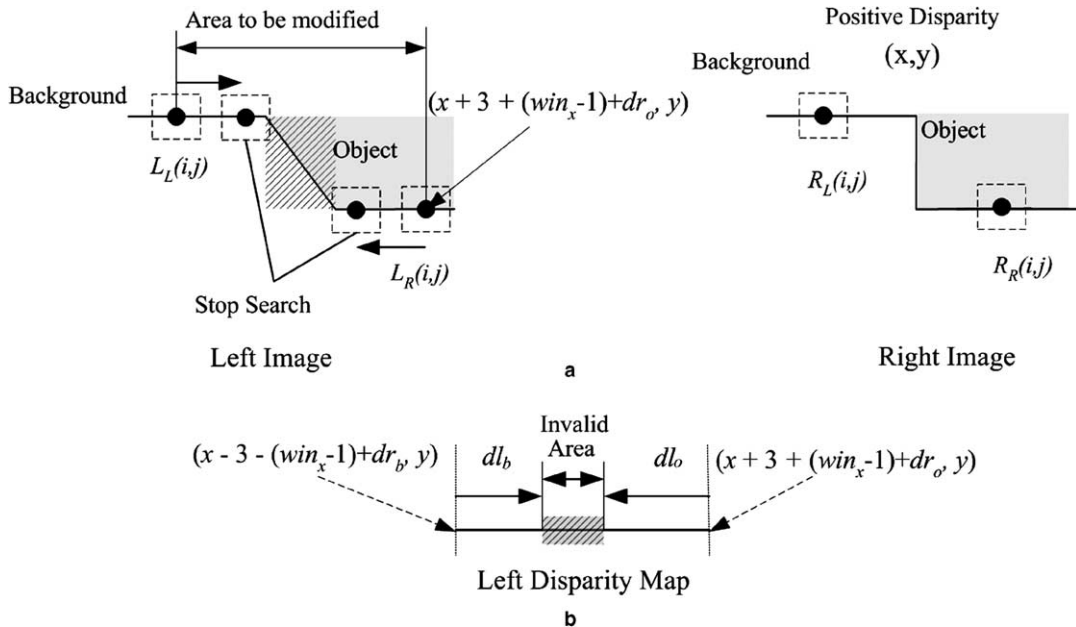

Fig. 8. (a) Process of reducing errors at depth discontinuities, (b) the modified disparities in the left disparity map.
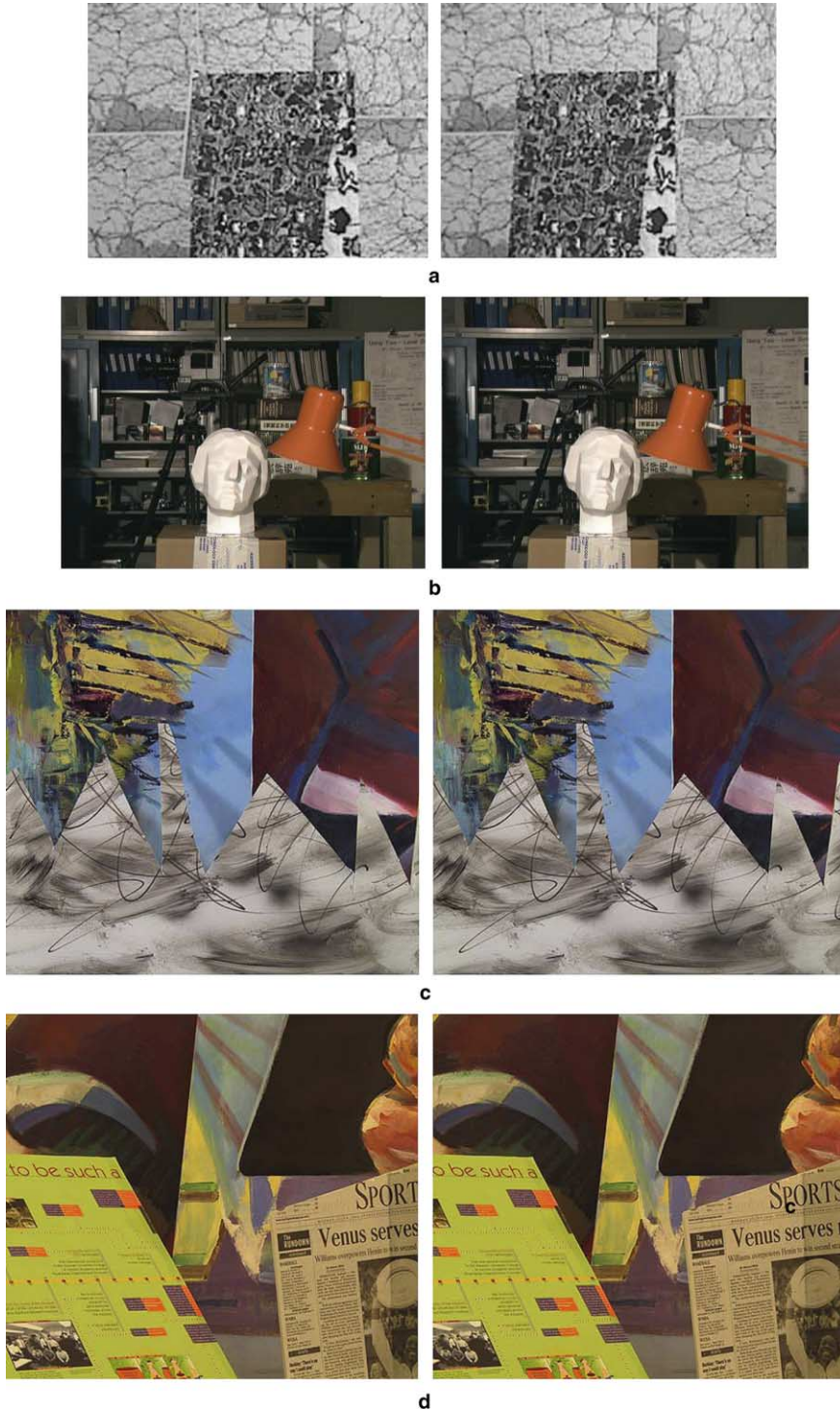
Fig. 9. (a) Map image pair, (b) Tsukuba image pair, (c) sawtooth image pair and (d) venus image pair. (These images were obtained from the Middlebury College stereo vision research page.)

Table 1
Image parameters used for disparity computation

Parameters	Map images	Tsukuba images	Sawtooth images	Venus images
Image size	284×216	384×288	434×380	434×383
Disparity range	0–32	0–32	0–32	0–32
No. of disparity level	32	32	32	32
Size of correlation window	9×9	9×9	9×9	9×9
Size of sub-correlation window	5×5	5×5	5×5	5×5
Size of median filter	5×5	5×5	5×5	5×5

is $D_L(x - 3 - (\text{win}_x - 1) + dr_b, y)$ and, dl_o, the disparity of the object in the left image, is $D_L(x + 3 + (\text{win}_x - 1) + dr_o, y)$.

Then, apply the same process (step 4 and step 5) to the right side of the positive disparity while shifting the $L'_R(i, j)$ to the left until the mass center of the correlation value between $R'_R(i, j)$ and $L'_R(i, j)$ is located in the gray area in Fig. 7(b). The process of step 5 and the modified disparity values in this region are shown in Fig. 8.

Step 6: In the case of negative disparity, the same process (steps 1–5) is performed.

3.4. Spatial interpolation

After the error reduction process, we interpolate the dense disparity map by median filtering (Mühlmann et al., 2002), which is a well-known filtering method to remove salt and pepper noise to interpolate a dense disparity map. This filter improves the quality of the dense disparity map in weakly textured regions and outliers (Mühlmann et al., 2002).

4. Experimental result

4.1. Evaluation of the proposed method for matching performance

We implemented the algorithm using Microsoft Visual C++ 6.0 on Microsoft Windows 2000. The results in this paper are obtained under an Intel Pentium IV 2.66 GHz Processor. Stereo image pairs and ground truth from R. Szeliski and D. Scharstein's Middlebury College stereovision research page[1] are used for the evaluation. Test image pairs are shown in Fig. 9. The parameters used for computing the disparity maps are shown in Table 1.

Fig. 10 shows the disparity maps for test image pairs. In Fig. 10, the disparities are scaled up by a factor of eight for the visual analysis except Tsukuba image pair. For Tsukuba image pair, disparities are scaled up by a factor of 16. The level of invalid area is −1 and is marked black in the disparity maps. And then, these results are compared with the ground truth. The correct matches and the invalid areas are indicated by a level of 256, white, and all errors are marked black in Fig. 11. Table 2 shows the stereo matching result for test image pairs.

For Tsukuba image pair, we compare the proposed algorithm with fast stereo matching algorithm, such as SAD method, multi-windowing method (Fusiello et al., 1997), SAD5 with border correction (Hirschmüller et al., 2002). The results are shown in Table 3. As shown in Table 3, the proposed method shows the best result from the viewpoint of correct matching percentage. Compared with the SAD correlation method, systematic errors are reduced and correct areas are increased. In the case of multi-windowing, the correct areas are increased and invalid areas are decreased and all errors and systematic errors are slightly decreased. Compared with SAD5 with border correction, the errors in all areas and at depth discontinuities by the proposed method are decreased and the invalid area is slightly increased. After error filtering

---

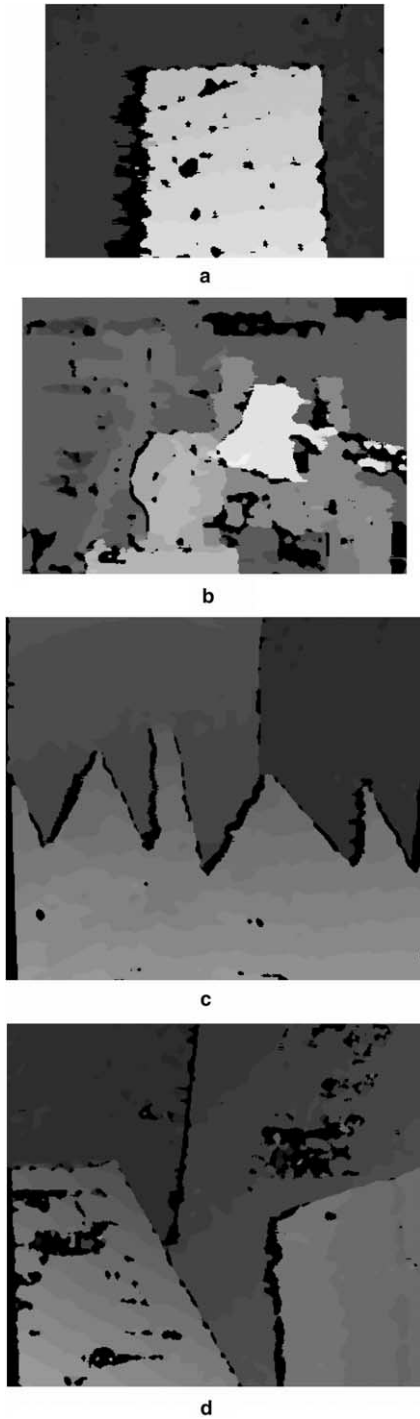Fig. 10. (a) Disparity map for map image pair, (b) disparity map for Tsukuba image pair, (c) disparity map for sawtooth image pair, (d) disparity map for venus image pair.
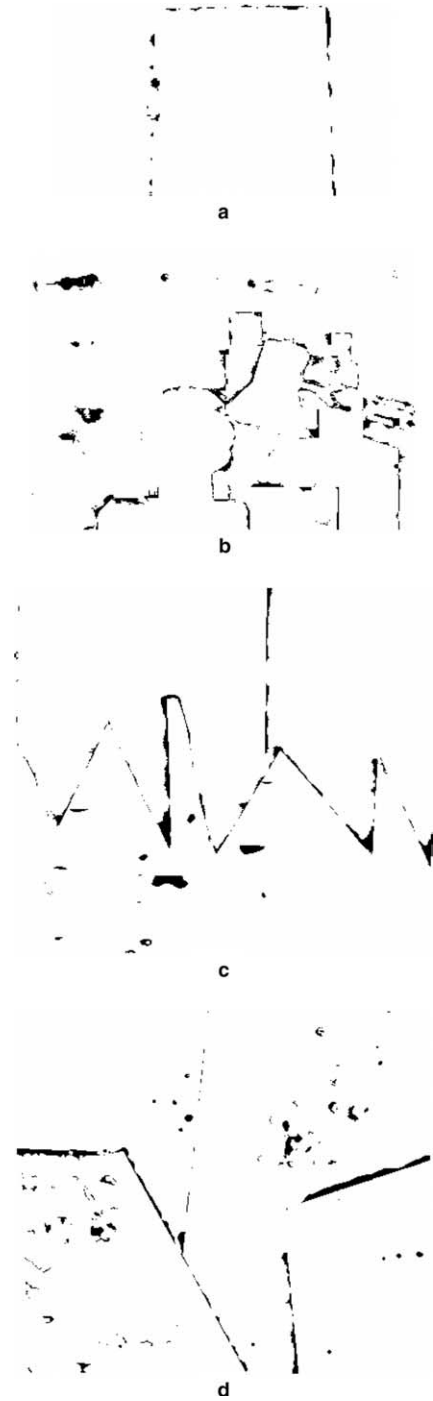

Fig. 11. (a) Error pixels in disparity map for map image pair, (b) error pixels in disparity map for Tsukuba image pair, (c) error pixels in disparity map for sawtooth image pair, (d) error pixels in disparity map for venus image pair.

Table 2
The stereo matching results by the proposed method for test image pairs

Image pair	Windows	Correct (%)	All errors (%)	Error at depth discontinuities (%)	Invalid (%)
Map images	9×9	90.86	0.77	0.73	8.37
	5×5				
Tsukuba images	9×9	87.21	4.25	2.89	8.54
	5×5				
Sawtooth images	9×9	92.30	2.09	1.53	5.61
	5×5				
Venus images	9×9	88.06	2.23	1.33	9.71
	5×5				

Table 3
Results of SAD method, multi-windowing method, Hirschmüller's method, and proposed method

Method	Windows	Log (σ)	Correct (%)	All errors (%)	Error at depth discontinuities (%)	Invalid (%)
Sum of absolute difference	9×9	1.0	82.97	6.00	4.39	11.03
Multi-windowing	11×9		80.88	4.91	2.92	14.21
Hirschmüller's method (SAD5 with border correction)	7×9	1.0	85.63	6.10	4.04	8.26
Hirschmüller's method (SAD5 with border correction and error filtering)	7×9	0.0	82.24	3.26	2.45	14.05
Proposed method	9×9		87.21	4.25	2.89	8.54
	5×5					
Proposed method with error filtering	9×9		81.17	3.27	2.30	15.56
	5×5					

(Hirschmüller et al., 2002), all errors and error pixels at depth discontinuities are decreased and invalid areas are increased. In this case, invalid areas are more than SAD5 with border correction and error filtering, however errors at depth discontinuities is decreased. Therefore, the proposed method shows good performance in the respect of reducing the errors at depth discontinuities.

4.2. Evaluation of the proposed method for time consumption

The proposed algorithm is implemented on Microsoft Windows 2000 by Visual C++ and an inline assembler to take advantage of parallel processing, SSE2 (Streaming SIMD Extension 2), which can be used on an Intel Pentium IV Proces-

Table 4
Required times for computing dense disparity map (in ms)

Calculation step	Map images	Tsukuba images	Sawtooth images	Venus images
DSI matrix computation	7	12	15	15
SAD correlation	19	37	47	47
Disparity selection	21	38	47	47
Left/right consistency check	4	4	5	5
Reducing errors at depth disparities	27	68	105	110
Median filtering	2	2	3	3
Total (ms)	80	161	222	227

sor. We can calculate 16 byte data by one instruction using SSE2.

The main factors to increase computation time are the size of image and the disparity range. And the time to reduce errors at depth discontinuities depends on the amount of depth discontinuities. Table 4 shows the required time to calculate a dense disparity map for the test image pairs.

The proposed error reduction method requires additional computation time and slows down the frame rate. However, in the case of 320×240 size images and a 32 disparity range, we can obtain the dense disparity at a rate of 7 frames per second. This update rate is suitable for real time applications in the field of robotics. More optimized implementation can reduce the computation time.

5. Conclusions

We proposed a new fast stereo matching method that can reduce the systematic errors of correlation-based stereo matching methods. Using the different sized correlation windows, we can reduce the errors at the region containing blurring and mismatched errors, in which the matching process commences from each side of the region and progresses toward the center. Dense disparity maps are obtained at a speed of approximately 7 frames per second in the case of 320×240 sized image pairs. The computation time can be reduced by an optimized implementation, parallel processing by Intel SIMD architecture, and a sliding summation, which prevent redundancy in the computation. A more accurate three-dimensional reconstruction can be obtained by this suggested method, and moreover the proposed method can be used in real time applications with greater accuracy.

References

Chen, T.-Y., Bovik, A.C., Cormack, L.K., 1999. Stereoscopic ranging by matching image modulations. IEEE Trans. Image Process. 8, 785–797.

Faugeras, O., Hotz, B., Mathieu, H., Viéville, T., Zhang, Z., Fua, P., Théron, E., Moll, L., Berry, G., Vuillemin, J., Bertin, P., Proy, C., 1993. Real time correlation-based stereo: Algorithm, implementations and application. INRIA, Technical Report No. 2013.

Fua, P., 1993. A parallel stereo algorithm that produces dense depth maps and preserves image features. Machine Vision Appl. 6, 35–49.

Fusiello, A., Roberto, V., Trucco, E., 1997. Efficient stereo with multiple windowing. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, pp. 858–863.

Fusiello, A., Trucco, E., Verri, A., 2000. A compact algorithm for rectification of stereo pairs. Machine Vision Appl. 12, 16–22.

Hariyama, M., Takeuchi, T., Kameyama, M., 2001. VLSI processor for reliable stereo matching based on adaptive window-size selection. Proc. IEEE Internat. Conf. Robot. Automat. 2, 1168–1173.

Hirschmüller, H., Innocent, P.R., Garibaldi, J., 2002. Real-time correlation-based stereo vision with reduced border errors. Internat. J. Comput. Vision 47 (1–3), 229–246.

Kanade, T., Okutomi, M., 1994. A stereo matching algorithm with an adaptive window: Theory and experiment. IEEE Trans. Pattern Anal. Machine Intell. 16, 920–932.

Konolige, K., 1997. Small vision systems: Hardware and implementation. In: Proceedings of Eighth International Symposium on Robotics Research, Hayama Japan, pp. 203–212.

Matthies, L.H., 1992. Stereo vision for planetary rovers: Stochastic modeling to near real-time implementation. Internat. J. Comput. Vision 8 (1), 71–91.

Mühlmann, K., Maier, D., Hesser, J., Männer, R., 2002. Calculating dense disparity maps from color stereo images, an efficient implementation. Internat. J. Comput. Vision 47 (1–3), 79–88.

Schmidt, J., Niemann, H., Vogt, S., 2002. Dense disparity maps in real-time with an application to augmented reality. In: Proceedings of Sixth IEEE Workshop on Applications of Computer Vision, pp. 225–230.