Journal of Information Display, vol. 6, no. 4, 2005

A New Rectification Scheme for Uncalibrated Stereo Image Pairs and Its Application to Intermediate View Reconstruction

Jung-Hwan Ko, Yong-Woo Jung, and Eun-Soo Kim*

Abstract

In this paper, a new rectification scheme to transform the uncalibrated stereo image pair into the calibrated one is suggested and its performance is analyzed by applying this scheme to the reconstruction of the intermediate views for multi-view stereoscopic display. In the proposed method, feature points are extracted from the stereo image pair by detecting the corners and similarities between each pixel of the stereo image pair. These detected feature points, are then used to extract moving vectors between the stereo image pair and the epipolar line. Finally, the input stereo image pair is rectified by matching the extracted epipolar line between the stereo image pair in the horizontal direction. Based on some experiments done on the synthesis of the intermediate views by using the calibrated stereo image pairs through the proposed rectification algorithm and the uncalibrated ones for three kinds of stereo image pairs; 'Man', 'Face' and 'Car', it is found that PSNRs of the intermediate views reconstructed from the calibrated images improved by about 2.5 ~ 3.26 dB than those of the uncalibrated ones.

Keywords : rectification, epipolar geometry, fundamental matrix

1. Introduction

Recently, to improve the shortcomings of the conventional binocular stereoscopic display method, many researches have been done on the multi-view stereoscopic 3D imaging and display systems. Basically, in this system, multiple cameras are used to obtain the multi-view images of an object and the sophisticated displaying system is used also to spatially project the multi-view images to the corresponding directions[1-2]. However, problems such as having excessive amounts of data to be processed and transmitted due to the increase in the number of views, and having a discontinuity between the viewpoints, can occur [3-5]. Accordingly, an intermediate view reconstruction (IVR) technique is suggested as an alternative to the practical multi-view stereoscopic 3D display system, where the IVR technique makes it possible to digitally synthesize numerous intermediate views at as many times as required, but using only the limited views of an object. The existing IVR technique [6-8] has the capability of accurately synthesizing in standard stereo images that do not need any calibration, but is significantly caught trouble which can result in a synthesis error and a image distortion in case of stereo images obtained from the uncalibrated camera. That is, in case of the real image captured from stereo camera at real environments, the input stereo image becomes distorted by the mismatching and discrepancy between two cameras this results, in the intermediate views synthesized by using these images to have many distortion, which in turn cause a falling-off in PSNR and consequently a defective 3D display. Recently, as a new approach to overcome the limitation of these existing IVR techniques, an image rectification and calibration methods for the uncalibrated stereo image pair has been suggested.

Al-Shalfan et al presented a direct algorithm for rectifying pairs of uncalibrated images [2]. Isgr`o and Trucco presented a robust algorithm performing uncalibrated rectification which does not require explicit computation of the epipolar geometry [3]. Pollefeys et al proposed a simple and efficient rectification method for general two view stereo images [4]. Loop and Zhang proposed a technique for computing rectification homo-

Manuscript received November 7, 2005; accepted for publication December 5, 2005.

* Member, KIDS.

This research was supported by the MIC(Ministry of Information and Communication), Korea, under the ITRC(Information Technology Research Center) support program supervised by the IITA(Institute of Information Technology Assessment)(IITA-2005-C1090-0502-0038)

Corresponding Author : Jung-Hwan Ko
Dept. of Electronic Eng., Kwangwoon University 447-1 Wolge-Dong, Nowon-Gu, Seoul, 139-701, Korea.
E-mail : misagi@explore.kw.ac.kr **Tel** : +02 940-5118 **Fax** : +02 941-5979

graphies for stereo vision [5]. In addition, Papadimitriou and Dennis presented an algorithm for rectifying stereo images in cases where images are taken with convergence geometry (coplanar X and Z axes and parallel Y axes) [6]. Ayache and Hansen presented a technique for calibrating and rectifying pairs or triplets images [7]. In their case, generally a camera matrix needs to be estimated, and thus their rectification algorithm works only for the calibrated cameras.

Therefore, in this paper, a stereo image rectification algorithm based on corner detection and epipolar geometry is proposed, in which the image's corner is extracted from stereo image pair and their correspondences are determined by imposing a number of matching constraints. Also, feature points are extracted from the stereo image pair through detection of the corners and similarities between each pixel of the stereo image. These detected feature points are used to extract, the global motion factor between stereo image and the epipolar line. Finally, the input stereo image is rectified by matching the extracted epipolar line between the stereo image in the horizontal direction and intermediate views are reconstructed by using these rectified stereo images.

2. Stereo Image Rectification Scheme

Image rectification is an important step in the three dimensional analysis of scenes. For stereo vision, image rectification can increase both the reliability and the speed of the disparity estimation process. This is because, in the rectified images, the relative rotation among the original images are removed and the disparity search is performed along the image horizontal or vertical scan lines. The rectification process requires certain camera calibration parameters or weakly calibrated (uncalibrated) epipolar geometries of the image pair or image triplet.

Fig. 1 shows the overall flowchart of the proposed stereo image rectification and IVR technique and a transformation process of the parameter for the stereo image rectification, respectively. Generally, the main problem in stereo image rectification is finding the corresponding points in images taken from different perspectives, i.e., is not easy to estimate the background difference occurring between images. These differences are called global motion and can be seen as a vector field moving one stereo image into the other. Accordingly, in this paper, the strength of the initial corresponded candidate point is detected by auto-correlation of image intensity over two half size × half size pixel patches centered on each feature as shown in Eq. (1).

$$C = \sum_{i,j \in patch} [I_2(i,j) - I_1(i,j)]^2 \tag{1}$$

Where $I_n(i,j)$ is the image intensity at coordinates (i,j) in the n th image. The match with the maximum strength is stored for each corner from the first to the second image. The same process is then applied in reverse from the second to the first image. Matches are accepted into the initial set if they exhibit a maximum in both comparisons. This has the effect of removing corners that have multiple candidate matches, which causes ambiguity.

Fig. 2 shows the feature displacement between left and right image after detecting a corresponding point through this process. Generally, the major problem with the perspective model is that it requires the calculation of the epipoles which tend to move away from common objects on the image plane. This is an ill-conditioned problem that brings about image triction very in the computation process.

Fig. 3 shows the fundamental epopolar geometric relationship between two perspective cameras in reference

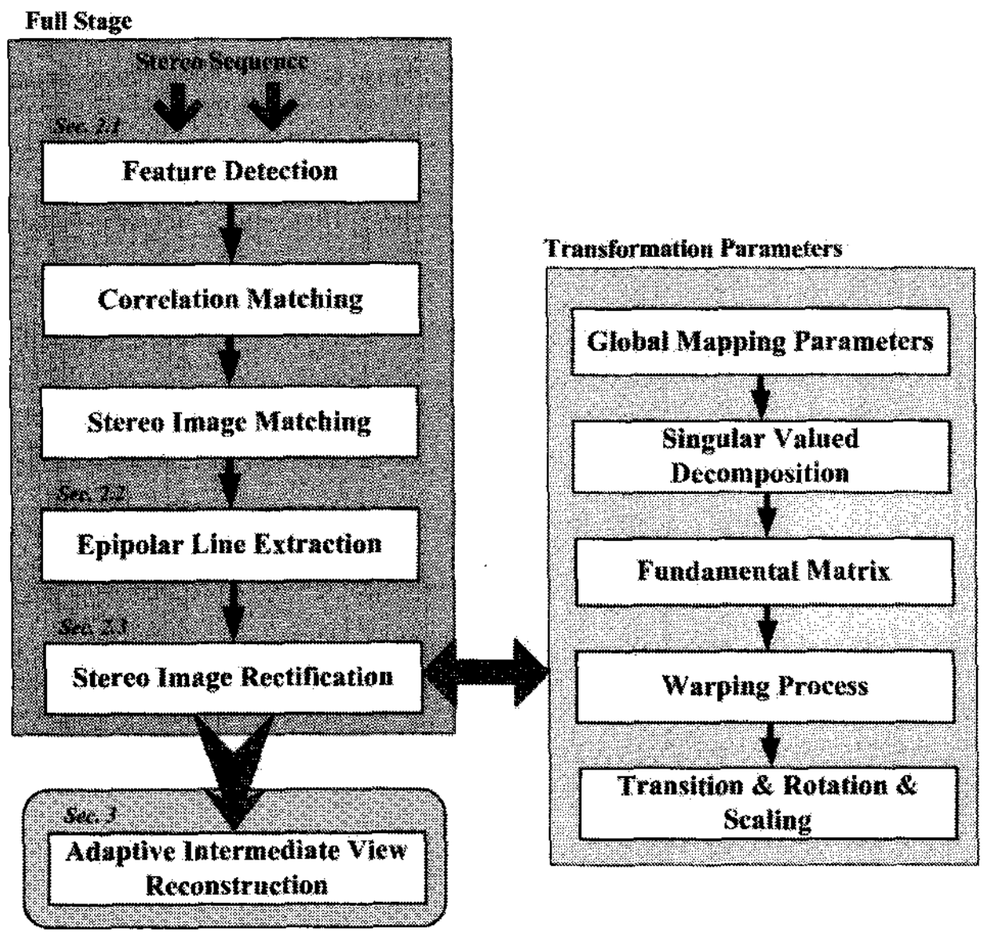


Fig. 1. Overall flowchart of the proposed stereo image rectification and IVR technique.

(a) Left image (b) Right image

Fig. 2. Detection of the feature displacement between left and right image.

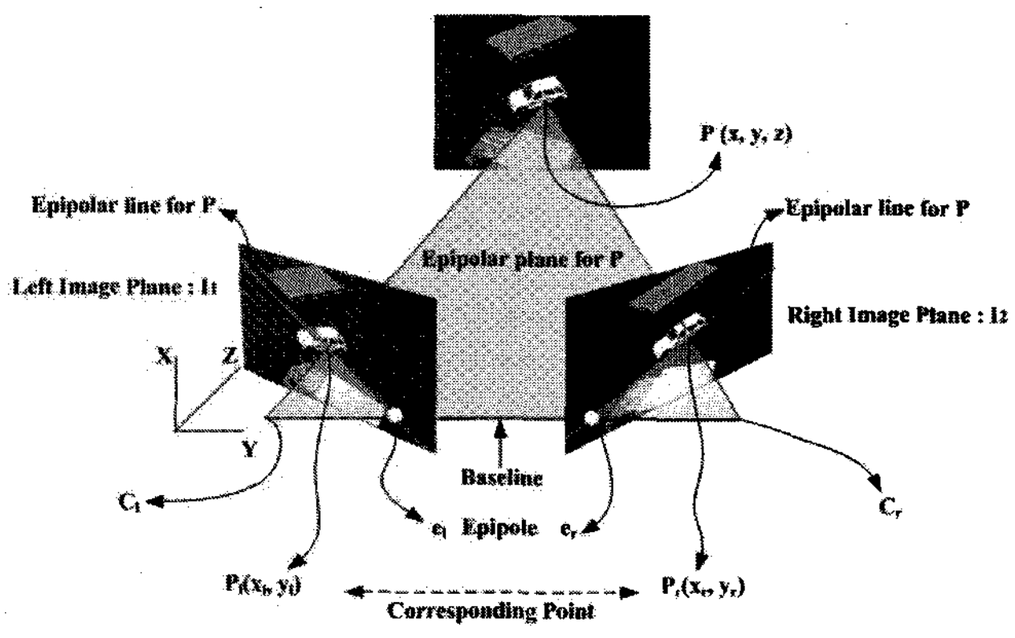


Fig. 3. Fundamental epopolar geometric relationship between two perspective cameras.

corresponding to one point. In Fig. 3, the epipolar lines $\{P_l, e_l\}$ and $\{P_r, e_r\}$ are defined by the intersection of the left and right planes with the epipolar plane (C_l, C_r, P). The points e_l and e_r are the epipoles and are defined by the intersection of the image planes with the baseline $\{C_l, C_r\}$. Note that the epipolar lines are parallel when the stereo cameras are arranged in parallel configuration [8].

As shown in Fig. 3, in order to project the stereo image plane containing a coordinate of certain point P on the image plane of the same dimension, it makes the stereo image parallel using epipolar lines. Accordingly, an epipolar line so that can materialize the detected feature from left and image can be detected by finding the so-called epipole using a fundamental matrix that is the algebraic representation of epipolar geometry.

In this paper, a fundamental matrix is detected by using the SVD (singular valued decomposition) of a square matrix, in which this decomposition uses some things that come out of eigen vector in the square symmetric matrix D^TD which is m × m [8]. Also, the relationship between a

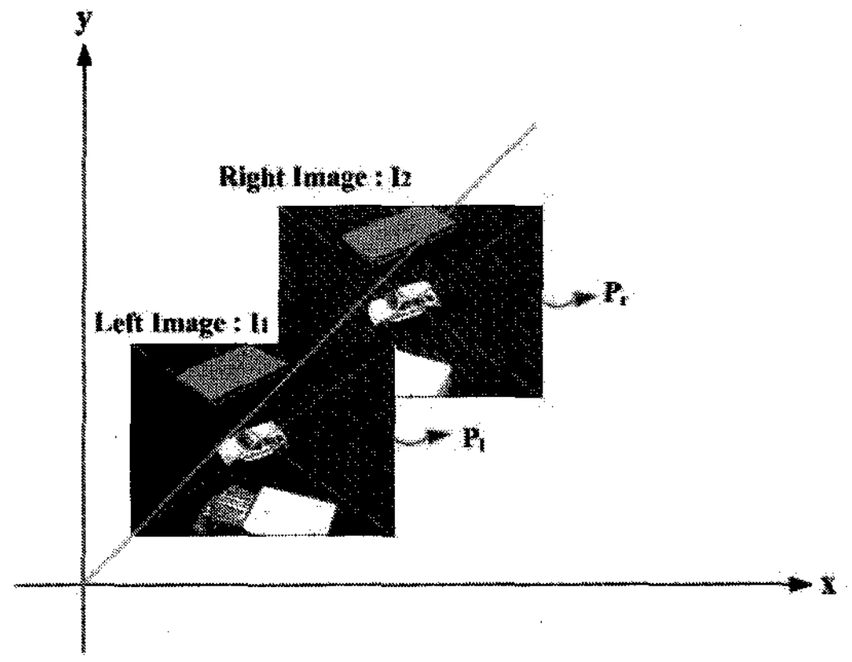


Fig. 4. 2D position relationship between stereo images.

point ***P*** in the observed scene and a point ***I*** of the camera retina is modeled as a linear transformation in projective coordinates.

Eq. (2) shows the SVD form as a linear and homogeneous equation at 9 unknown coefficients of matrix F by using a corresponding point between stereo images.

$$D_i = [uu', vu', wu', uv', vv', wv', uw', vw', ww'] \quad (2)$$

Where, if the coordinates of P is (x, y, z), (U, V, W) represents the projective coordinates of I.

The fundamental matrix obtained from the result of the mathematical modeling between left and right image only includes information of the feature between the stereo images [9-10]. Thus, the best fit polynomial transformation necessary to perform a rectification procedure, which horizontally gets epipolar line of each image detected through feature points. Accordingly, in order to align the epipolar lines of this stereo pair, some transformation necessary parameter that can transform to fit a reference image into a geometry transformation of such transition, rotation and scaling, is needed Fig. 4 shows a two-dimensional position relationship between stereo images established on the assumption here is that focal length of the camera (z-axis) is equivalent. Since the right image I_2 is compared with the left image I_1 meaning the reference image and, P_r which performs an affine transformation such as a transmission, a rotation for P_1 is given by Eq. (3) when any point of I_1, P_1, corresponds to one point of I_2, P_r.

$$P_r = TRP_l \quad (3)$$

Where, T and R represent a transmission and rotation matrix, respectively. Furthermore, in the case of the coordinates of P_r that was obtained through this, they after performing a scaling inverse transformation through a focal length transformation is equivalent to P_r. Thus, if the coordinates of P_r denote (P_x, P_y, λ_1), then the relationship between P_r and P_r can be expressed as in Eq. (4).

$$p_x^{''} = \lambda_1 \frac{p_x^{'}}{p_z^{'}}, \qquad p_y^{''} = \lambda_1 \frac{p_y^{'}}{p_z^{'}} \tag{4}$$

If Eq. (4) is noted as a matrix form after substituting the result of Eq. (3) for each P_r, then Eq. (4) is given by Eq. (5).

$$\begin{bmatrix} p_x^{''} \\ p_y^{''} \\ 1 \end{bmatrix} = \begin{bmatrix} k_1 & k_2 & k_3 \\ k_4 & k_5 & k_6 \\ k_7 & k_8 & k_9 \end{bmatrix} \cdot \begin{bmatrix} p_x \\ p_y \\ 1 \end{bmatrix} \tag{5}$$

Each k_i consists of the formula of transformation parameters, If all k_i are calculated, then each parameter is yielded through the inverse transformation of this formula. Thus, the transformation matrix through this can be written as Eq. (6).

$$\begin{bmatrix} k_1 & k_2 & k_3 \\ k_4 & k_5 & k_6 \\ k_7 & k_8 & k_9 \end{bmatrix} = \begin{bmatrix} \sum x_j^2 & \sum x_j y_j & \sum x_j \\ \sum x_j y_j & \sum y_j{}^2 k_3 & \sum y_j \\ \sum x_j & \sum y_j & n \end{bmatrix} \cdot \begin{bmatrix} \sum x_j x_j^{'} & \sum x_j y_j^{'} & \sum x_j \\ \sum y_j x_j^{'} & \sum y_j y_j^{'} & \sum y_j \\ \sum x^{'}{}_j & \sum y^{'}{}_j & n \end{bmatrix} \tag{6}$$

By using each matrix element k_i and parameters of the global motion calculated as shown in Eq. (6), the global corresponding parameter is obtained, where a transformation matrix $[k_i]$ is a transformation parameter which can change location vectors of the feature on the original stereo image. Furthermore, the parameter obtained from Eq. (6) is a transformation formula which can transform SVD value obtained from the feature of the left and right image, which is parallel to the horizontal line of the epipole as x-axis. If the distance between the x and x' is constantly maintained, the image rectification is successfully completed.

A general concept of the intermediate view reconstruction is shown in Fig. 5. It shows the intersection of the desired coordinate positions for the intermediate view. The left view-plane, 'L', corresponds to the right view-plane, 'R', through the plane of intermediate view, 'I'. The position of the corresponding intermediate view is defined with a normalized distance α from the left view.

The distance from the left to the right view plane is normalized to 1, such that $0 \leq \alpha \leq 1$. The intermediate view images can be synthesized by a linear combination of the left and right images using interpolation. Interpolation is known as a technique to represent an arbitrary continuous function as a discrete sum of weighted and shifted synthesis functions. In this paper, more natural intermediate views were acquired through the interpolation with a weighted mean value [7-8]. Eq. (7) shows the case of interpolation with a weighted mean by using the position value α of the viewpoints[8], where I_P denotes the synthesized intermediate-view image.

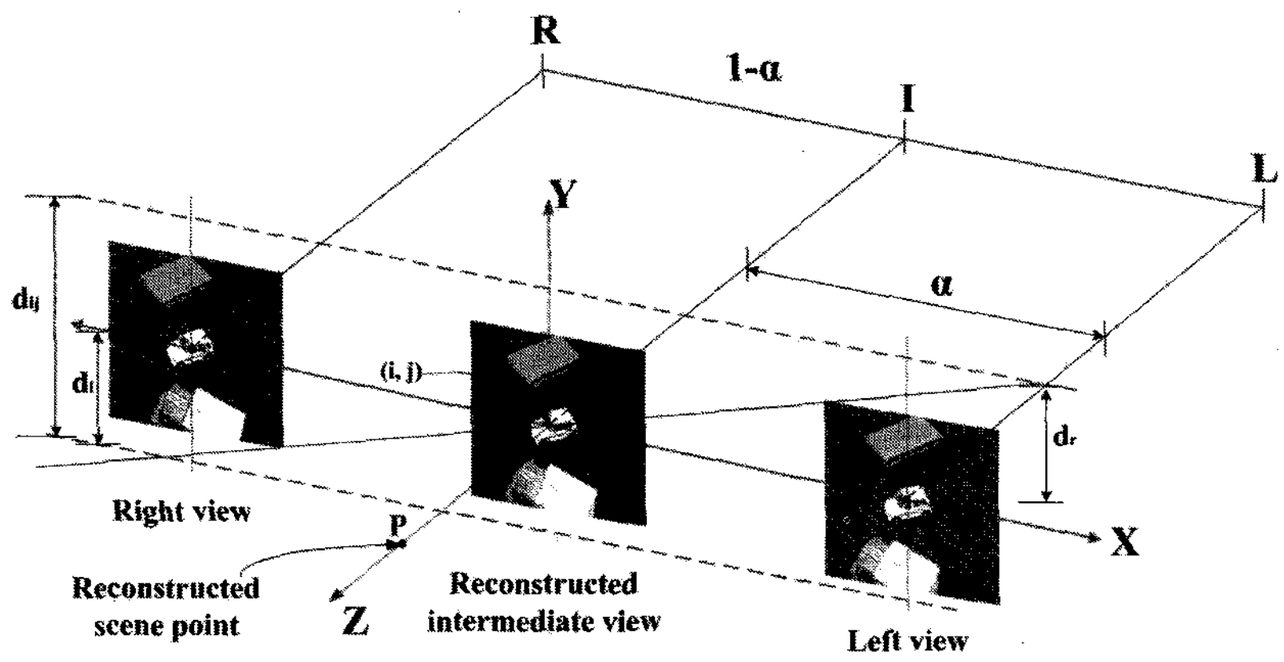
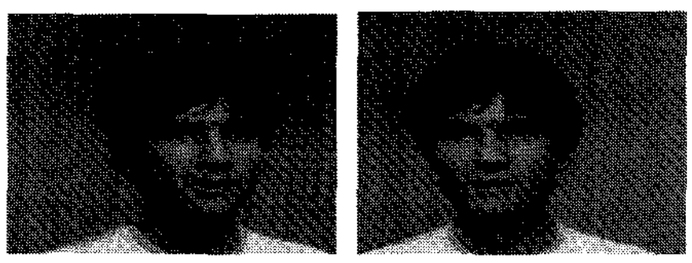
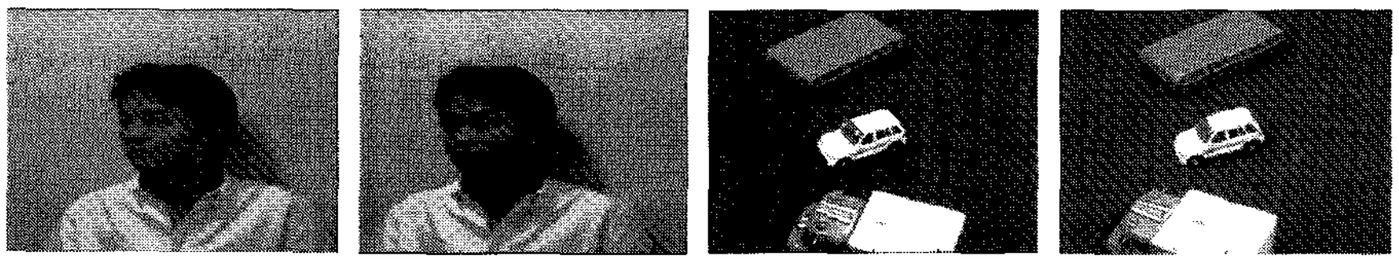


Fig. 5. Corresponding point of intermediate view from left and right view-plane.

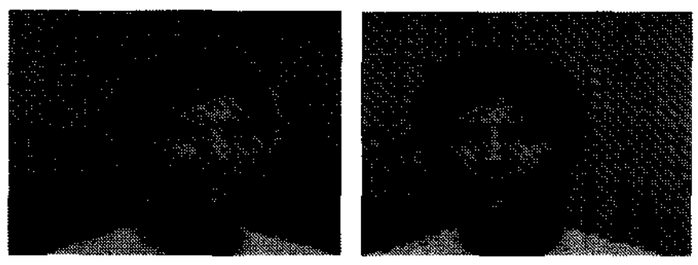


(a) Left image (b) Right image

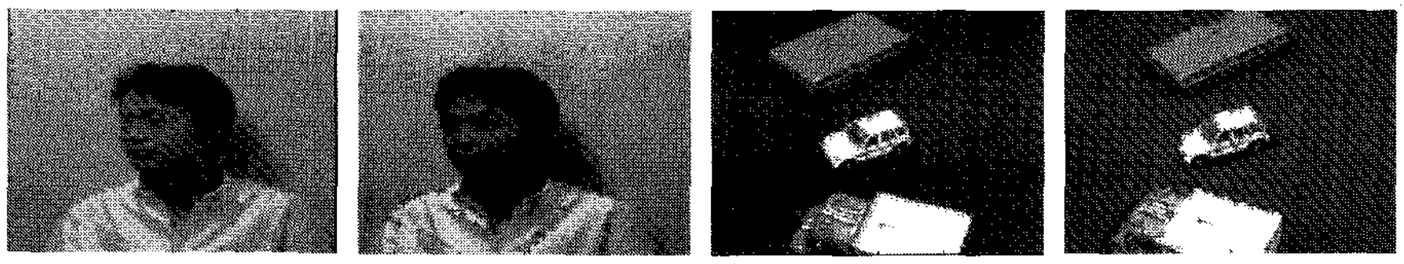


(c) Left image (d) Right image (e) Left image (f) Right image

Fig. 6. 3 stereo images pairs of 'Man', 'Face' and 'Car'.



(a) Left image (b) Right image



(c) Left image (d) Right image (e) Left image (f) Right image

Fig. 7. Extracted feature by using corner detection method.

$$I_P(i,j) = (1-\alpha)\cdot I_R(i+\hat{d}(i,j),j) + \alpha \cdot I_L(i-\hat{d}(i,j),j) \quad (7)$$

Generally, during the intermediate view reconstruction process, some occluded regions might form because one of the stereo cameras lose visibility while the other does not and some regions the allocation of the disparity vector is overlapped. Therefore, in this paper, the disparity vectors for these regions were substituted with the mean values of the disparity vectors of the nearby regions by disparity regularization. If the viewpoint is not occluded, then the disparity is defined as the distance between the image points in both images. Equation (8) shows the disparity in the horizontal direction and the relationship between left and right images, where I_R, I_L represent the right and left images in the coordinates of $(i_R + j_R)$ and $(i_L + j_L)$, respectively.

$$I_R = \begin{bmatrix} i_R \\ j_R \end{bmatrix} = \begin{bmatrix} i_L + \hat{d}\ (i_L, j_L) \\ j_L \end{bmatrix} \approx I_L + \begin{bmatrix} \hat{d}(i_L, j_L) \\ 0 \end{bmatrix} \quad (8)$$

3. Experiments and Results

In the experiment, CCETT's one kinds of images, 'Man' is used as the stereo image pairs by formulating into the 'raw' files of 640×480 pixels, as shown in Figs. 7(a) and (b). Also, a stereo image pairs of 640×640 pixels captured by using two Sony(XC-ST50) cameras fitted with

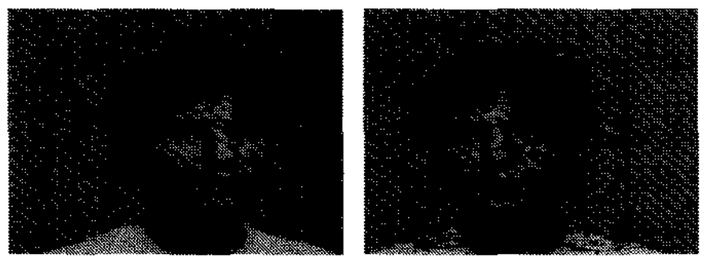
6 mm lenses and digitized by a frame grabber (Meteor II MC/4), 'Face' and 'Car' were used as shown in Figs. 7(c), (d), (e), and (f). That is, Figs. (c) and (d) are the face images of a simple background, which has a background silhouette as compared with CCETT's 'Man' image, and stereo image after controlling convergence in the stereo object tracking system, Figs. (e) and (f) have a fore/ background. Firstly, point match across the two images were carried out for three kinds of stereo image pairs used in the experiment, Fig. 7 shows 200 extracted features

Table 1. Comparison results of execution time

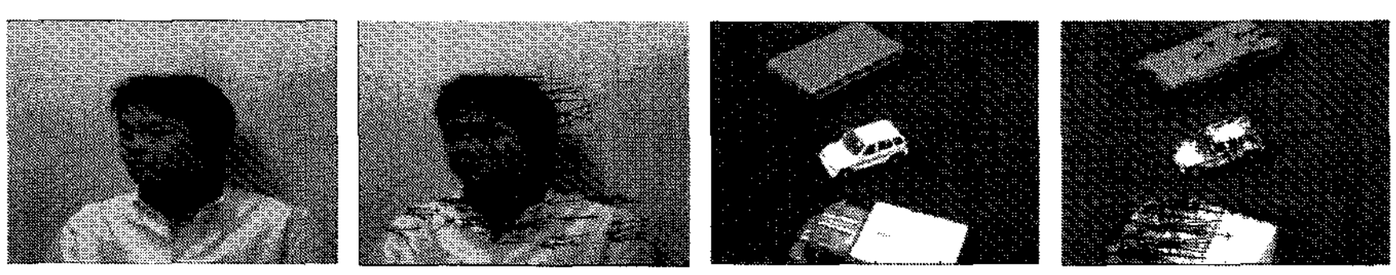
Feature number \ Image	'Man'	'Face'	'Car'
200	0.2002 [sec]	0.3557 [sec]	0.4527 [sec]
500	0.2443 [sec]	0.5117 [sec]	0.7121 [sec]
1000	0.8998 [sec]	0.9917 [sec]	1.2329 [sec]

Table 2. Comparison results of PSNR

Rectification \ Image	PSNR (dB)		
	'Man'	'Face'	'Car'
Before rectification	26.25	18.18	16.47
After rectification	29.85	20.77	17.94

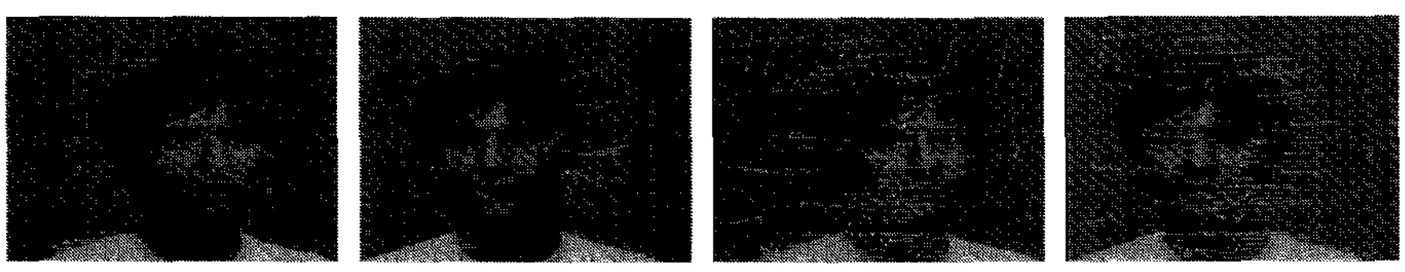


(a) Left image (b) Right image

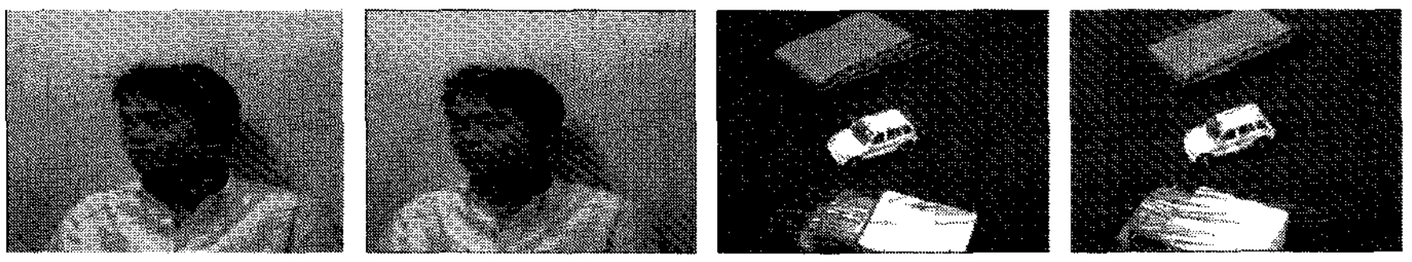


(c) Left image (d) Right image (e) Left image (f) Right image

Fig. 8. Motion vector extracted from the stereo image.



(a) Left image (b) Right image (c) Left image (d) Right image



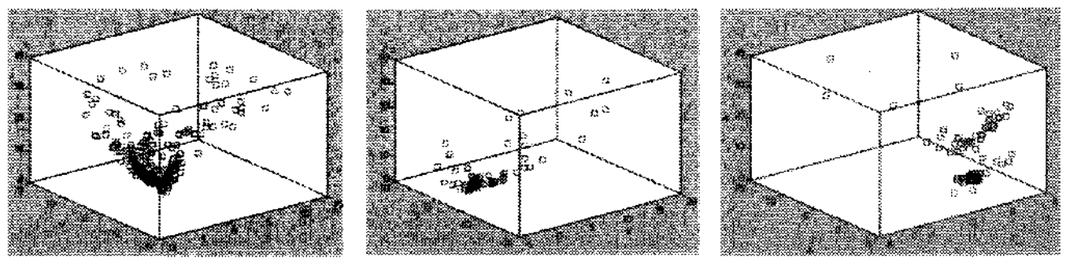
(e) Left image (f) Right image (g) Left image (h) Right image

Fig. 9. Epipolar line extraction.

obtained by Birchfield/Tomasi method [11], in which original stereo image is applied to the auto-correlation among the existing corner detectors. The comparison of the detection time as the number of feature increases for three kinds of the stereo image are summarized in Table 3. As shown in Table 1, as the number of features increases, executing time also significantly increases. Fig. 8 illustrates the motion vector of the disparity position for the stereo image matching. In Fig. 8, the motion vectors of the disparity position are superimposed over the right image for each corresponding point immediately detecting the feature between the stereo image pairs. That is, by using the feature extracted from the stereo image pairs as difference information, the motion vector between the left and right image is shown in Fig. 8. Figs. 9 shows the epipolar lines of the stereo image the extracted with 200 features for each image by using the detected motion vector and SVD, respectively. However, in case of Figs. 9 (a) and (b), there is a vertical error in the detection of the epipolar line due to detection error of the feature, which can occurr due to the similarity in the brightness between background and human face.

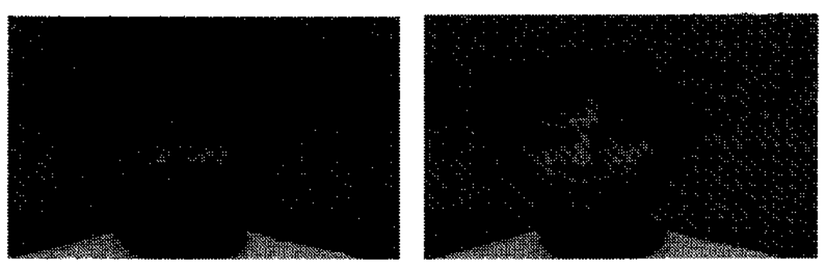
Figs. 10 (a), (b) and (c) show a motion disparity vector obtained by the corner detector described in Sec. 2.1 and epipolar lines detected by the fundamental matrix and SVD described in Sec. 2.2 on the right image to estimate the rectification information between the uncalibrated stereo image pairs. Figs. 10 (d), (e) and (f) also show the 3D plot for the experimental result of Figs. 10 (a), (b) and (c), respectively, in which the scattered square dots represents

(a) 'Man' (b) 'Face' (C) 'Car'

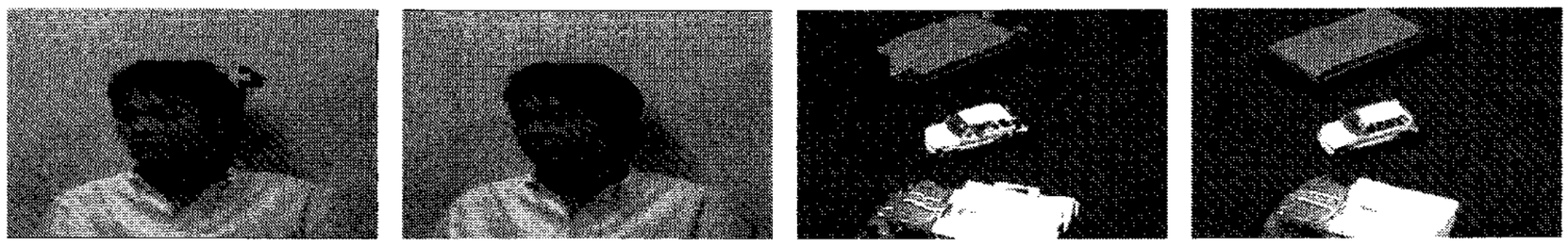


(d) 3D plot (Man) (e) 3D plot (Face) (f) 3D plot (Car)

Fig. 10. Rectification information.



(a) Before rectification (b) After rectification



(c) Before rectification (d) After rectification (e) Before rectification (f) After rectification

Fig. 11. Intermediate views synthesized.

the obtained rectification information between stereo image pairs. In Figs. 10 (d), (e) and (f), the rectification information, marked with red square dots, represents feature information of the stereo image with the disparity. This means that as the number of features increases, the distance between the features also increases and this needs to be rectified more for each disparity. Fig. 11 shows the intermediate view synthesis before rectification and after rectification for the uncali-brated stereo image, respectively. Figs. 11 (b), (d) and (f) represent the intermediate view synthesis immediately after applying the adaptive disparity estimation algorithm the rectified stereo image by using the rectification information and transformation parameter obtained in Eq. (6), meanwhile Figs. 11 (a), (c) and (e) represent the intermediate view synthesis by using the adaptive disparity estimation algorithm for the uncalibrated stereo image pairs. In case of the uncalibrated stereo image, it was found that some mismatch still occurs and some disparities are improperly assigned around the edges of an object, as shown in Figs. 11 (a), (c), and (e). But, on the whole, the intermediate view images rectified by the proposed algorithm look finer not only on the edges of an object because of its fine matching in those regions, but also in the background regions because of its coarse matching at that particular area. Table 2 shows peak-signal-to-noise (PSNR) of the intermediate view synthesized by the adaptive disparity estimation algorithm for the uncalibrated stereo image and the rectified stereo image, respectively. PSNR is computed by using Eq. (9), in which the root mean squared error (RMSE) means the square root of MSE.

$$PSNR = 20\log_{10}\left[\frac{255}{RMSE}\right] \tag{9}$$

It is found that the rectified images improve PSNR by 3.6 dB, 2.59 dB and 1.47 dB over that of the uncalibrated images for each of the 'Man', 'Face' and 'Car' images, respectively. In Table 2, in case of the uncalibrated image of 'Face' captured by stereo camera, PSNR is improved by as much as 2.59 dB after image rectification, and 'Car' image also showed improvements in PSNR by 1.47 dB, which means that the intermediate view synthesis is achieved under the objects existing in the for/background having a disparity with no error. Moreover, in case of the calibrated image of 'Man', PSNR improved after image rectification as well. From Table 2, PSNR can be seen to be a very high value of about 2.55dB on average.

These very satisfactory experimental results, shows that the proposed feature-based image rectification and intermediate view reconstruction using these rectified images can effectively synthesize the intermediate view image with very high PSNR and it can be applied to the practical implementation of the true-view stereoscopic 3D display system.

4. Conclusions

In this paper, a new intermediate view reconstruction method employing a stereo image rectification algorithm by which an uncalibrated input stereo image can be transformed into the calibrated one was suggested and its performance was analyzed. Based on the result of some experiments done on synthesis of the intermediate views with three kinds of stereo images; a CCETT's stereo image pair of 'Man' and two stereo image pairs of 'Face' & 'Car' captured by stereo camera, PSNRs of the intermediate views reconstructed from the calibrated images was analyzed by using the proposed rectification algorithm and showed an improvement by 2.5dB for 'Man', 4.26dB for 'Face' and 3.85dB for 'Car' compared of the uncalibrated ones. This very satisfactory experimental result shows that the proposed stereo image rectification algorithm can be applied to the intermediate view reconstruction scheme with improved PSNR.

References

[1] N. A. Dodgson, J. R. Moore, and S. R. Lang, *IBC* (1999), p. 497.

[2] W. Niem and M. Steinmetz, *ICIP-96* (1996), p. 16.

[3] J. R. Ohm and K. Muller, *IEEE Trans. on Circuits and Systems for Video Tech.*, **9**, 389 (1999).

[4] C. L. Pagliari, M. M. Perez, and T. J. Dennis, *ICIP-9* (1998), p. 627.

[5] A. Redert, E. Hendriks, and J. Biemond, *ICASSP-97*(1997), p. 2749.

[6] D. Tzovaras, N. Grammalidis, and M. Strintzis, *IEEE Trans. On C.S.V.T.*, **7**, p. 312 (1997).

[7] J. S. McVeigh, M. W. Siegel, and A. G. Jordan, *Signal Processing: Image Communication*, **9**, 21 (1996).

[8] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes in C*, (The Art of Scientific Computing 2nd, 1992), p. 20.

[9] K. H. Bae, J. J. Kim, and E. S. Kim, *Optical Engineering*, **42**, 1778 (2003).

[10] R. I. Hartley, *International Journal of Computer Vision*, **35**, 115 (1999).

[11] N. Ayache and C. Hansen, *Proc. of International Conference on Pattern Recognition* (1988), p. 11.

[12] D. Salzmann, *Atomic Physics in Hot Plasmas* (Oxford University Press, Oxford, 1998), p. 345.