Proceedings of Mirage 2005, INRIA Rocquencourt, France, March, 1-2 2005

FREE VIEWPOINT IMAGE SYNTHESIS USING UNCALIBRATED MULTIPLE MOVING CAMERAS

Yosuke Ito and Hideo Saito

{y-ito, saito}@ozawa.ics.keio.ac.jp
Department of Information and Computer Science, Keio University, Yokohama, 223-8522, Japan

ABSTRACT

We propose a novel method to synthesize free-viewpoint images for a moving object, which is captured by uncalibrated multiple moving cameras. If multiple fixed cameras are used for capturing a moving object, we must zoom out in order to capture the moving object within FOV of the cameras. Such zooming-out limits the resolution for the moving object in the captured images. In the proposed method, we use multiple moving cameras that capture the moving object in the center of the images with high resolution. For shape reconstruction of the object from the uncalibrated multiple moving camera images, two fixed cameras are employed for determining Projective Grid Space, which defines a projective 3D coordinate in the object space. The coordinate in PGS can be related to every moving camera by fundamental matrices between the moving camera and the fixed cameras. In the experiment that is performed for demonstrating the efficacy of the proposed method, high resolution free-viewpoint images can be successfully synthesized by the proposed method.

1. INTRODUCTION

Free viewpoint video synthesis has recently been studied by many researchers. Eye Vision [1] is known as the free-viewpoint capturing and displaying system that is practically used in Superbowl broadcasting. Bullet-Time system [2], which employs more than 100 cameras around an object scene, is used for realizing a new camera effect to movie production. Those systems generate free viewpoint videos by just switching the fixed multiple cameras, so it is difficult to generate free viewpoint video in which the viewpoint can be completely controlled by user's preference.

One of popular topic in computer vision area is new view synthesis from multiple cameras. In most of researches of new view synthesis, objects are supposed to be captured within FOV of every camera. If the objects moves around the scene, FOV of cameras need to be wide so that the objects can always be captured within the images. Therefore, image resolution for the objects is not sufficient in some cases.

Use of moving cameras is one way for obtaining sufficient resolution for moving objects. Moving cameras can capture moving objects in a constant area in the image by tracking moving objects. However, all the moving cameras need to be dynamically calibrated for synthesizing new view from multiple moving cameras.

In this paper, we propose a new method for synthesizing free-viewpoint video from moving multiple cameras by manually. We suppose that uncalibrated multiple cameras are moved by hand for capturing moving objects in FOVs in the captured images. For obtaining geometrical relationship among the cameras, we put two fixed cameras in addition to the multiple moving cameras. Then we define Projective Grid Space (PGS) [3] based on those two fixed cameras. All the moving cameras can geometrically be related to the PGS by computing fundamental matrices of each moving camera with two fixed cameras. We can compute the fundamental matrices by tracking natural feature points in the image sequences captured with the moving cameras. We recover shape of objects by volume intersection of all the silhouette images captured by the multiple moving cameras in PGS. The recovered shape in PGS provides pixel-wise correspondences among the multiple cameras, which are used for synthesizing free-viewpoint images by view interpolation [4].

In the rest of the paper, we first describe about related works in Section 1.1. Then, we present the theory and the detailed algorithms of the proposed method in Section 2 – 6. Finally, we show experimental results for demonstrating the effectiveness of the proposed method in Section 7 followed by the discussion and the conclusion.

1.1. Related Works

Free viewpoint images can easily be synthesized from multiple view images if 3D shape of objects can be recovered. A basic scheme for free viewpoint image synthesis is new view generation from stereo images [5]. Such methods are applied for synthesizing facial image from user's view direction in tele-conference systems [6, 7]. Increasing the number of cameras will improve recovered 3D shape and quality of free viewpoint images. Virtualied Reality Project by Kanade et al. [8] is one of earlier researches based on such multiple cameras system. They apply multiple baseline stereo for recovering object 3D shape from 50 cameras, and synthesize free-viewpoint video[9]. Moezzi et al. also synthesize free viewpoint video by recovering visual hull of objects from silhouette images of 17 cameras [10]. Saito et al. apply view interpolation to synthesize free viewpoint images for improving quality of images [11]. Carranza et al. recover human motion by fitting human shape model to input multiple view silhouette images for accurate shape recovery of object human body, which provide high quality free viewpoint videos of object human [12]. 3D studios applying such free viewpoint video synthesis have recently been developed [13, 14, 15].

In most of these researches, multiple cameras are fixed and calibrated. For avoiding the effort to fully calibrate multiple cameras, Saito and Kanade have proposed Projective Grid Space [3], which can be defined from just fundamental matrices amount multiple cameras. Such weak calibration of multiple cameras represented by fundamental matrices can be measured much easier than full calibration. PGS is also used for free viewpoint video synthesis [16, 17]. Other method for avoiding effort to full calibration is applying self calibration method to multiple cameras. Self calibration method proposed by Pollefeys [18] is applied in the 3D studio system used in [12, 14].

The proposed method in this paper is based on PGS [3, 16]. In the proposed system, two fixed cameras are used for defining PGS. All moving cameras are geometrically related to the PGS by tracking feature points, which are used for computing fundamental matrices with the fixed cameras.

2. PROJECTIVE GRID SPACE

Estimating the projection matrices or camera parameters is called full calibration. In multiple camera settings, measuring the 3D-2D correspondences in the objective space for all cameras often requires a lot of work. On the other hand, it is relatively easy to measure just 2D-2D correspondences among multiple camera images, because no 3D position of sample points is needed. It is called weak calibration to estimate geometrical relationship among multiple cameras from such 2D-2D correspondences. One representation of weak calibration is fundamental matrices between two cameras.

Projective Grid Space (PGS) [3] is a scheme for easy definition of 3D space by fundamental matrices among cameras. Therefore, the PGS enables 3D reconstruction from multiple images without full calibration of each camera.

The PGS is defined by image coordinates of two basis cameras (basis camera1 and basis camera2), which are arbitrarily selected from multiple cameras. Instead of using the X-Y-Z coordinate system in 3D Euclidian grid space, the P-Q-R coordinate system is used in PGS. The camera-image coordinates x and y in the basis camera1 take on the P and Q coordinates in PGS. The camera-image coordinate x in the basis camera2 corresponds to the R coordinate (Fig.1).

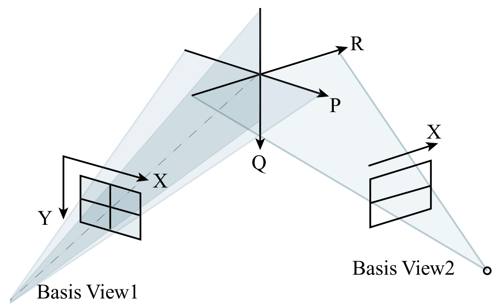


Figure 1: *Definition of the Projective Grid Space.*

In epipolar geometry, each viewpoint appears as the epipoles on the other images (Fig.2). The position of the basis camera1 C_1 in the PGS is determined as $C_1\left(X1_c, Y1_c, e_{21_x}\right)$ where $c_1\left(X1_c, Y1_c\right)$ is the camera center on basis1 and $e_{21}\left(e_{21_x}, e_{21_y}\right)$ is the epipole, which is the projection of the basis view1 onto the basis view2. Similarly, basis camera2 C_2 in the PGS is represented as $C_2\left(e_{12_x}, e_{12_y}, X2_c\right)$ where $c_2\left(X2_c, Y2_c\right)$ is the camera center on basis2 and $e_{12}\left(e_{12_x}, e_{12_y}\right)$ is the epipole, which corresponds to the basis camera1. The non-basis cameras C_i in the PGS are defined as $C_i\left(e_{1i_x}, e_{1i_y}, e_{2i_x}\right)$ where $e_{1i}\left(e_{1i_x}, e_{1i_y}\right)$ and $e_{2i}\left(e_{2i_x}, e_{2i_y}\right)$ are the epipoles projected onto the basis camera1 and 2 respectively.

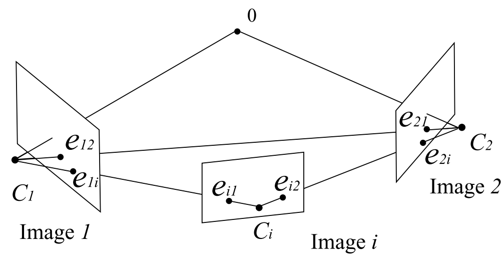


Figure 2: *Epipoles on each view.*

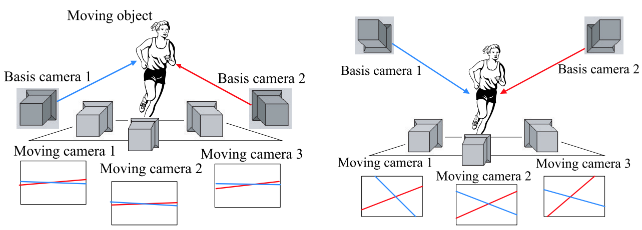
3. SYSTEM ENVIRONMENT

In this paper, we aim to realize free-viewpoint image synthesis from moving multiple cameras. However, it is especially difficult to obtain 3D-2D correspondences at every time instance in the motion of the cameras for full camera calibration, because it is almost impossible to put markers with known 3D positions in the scene. Measuring 2D-2D correspondences at every instance is relatively easy, so we employ PGS for recovering 3D shape of the object from moving cameras.

In the proposed system, in addition to the moving cameras, two fixed cameras are utilized. This two fixed cameras play the role of the basis cameras to define P-Q-R coordinate system in PGS. The two view directions are set to be almost orthogonal so that we can roughly approximate PGS as the Euclidian grid space. Besides, the two cameras are set far from the moving object so that the object can be captured constantly within the cameras' FOVs.

We consider two kinds of camera settings, which are horizontal settings and non-horizontal one shown in Fig.3. On each moving camera image in the horizontal settings the angle between the two epipolar lines that are projections of the two basis views is very small and results in ambiguity of the position where a 3D point in PGS is projected. In the proposed method, the position of the point where the two epipolar lines intersect have to be determined precisely to project a point in PGS onto the image-coordinate accurately. The detail is described in section 5.

The proposed system employs the non-horizontal settings such as Fig.3(b) to avoid such a problem. In that setting, the two basis cameras look down the object for making enough angle between the two epipolar lines on each moving camera image.



(a)horizontal settings (b)non-horizontal settings

Figure 3: *Camera settings.*

4. ESTIMATION OF THE FUNDAMENTAL MATRICES BETWEEN THE CAMERAS

As described in Sec.2, it is required to estimate a fundamental matrix between two basis cameras to define $P-Q-R$ coordinate system in PGS. Each fundamental matrix between a basis camera and a moving camera has also to be computed to project 3D points in PGS onto the moving camera images or to estimate the 3D positions of moving view points in PGS.

In the first frame the fundamental matrices among all moving cameras and two fixed cameras are estimated by 2D-2D correspondences of feature points on each view. The feature points are extracted by the Harris corner detector [19]. We obtain manually 2D-2D correspondences of the feature points on each view. The fundamental matrices are computed from those correspondences by using normalized eight-point algorithm [20].

From the second frame it is required to update each fundamental matrix between a basis camera and a moving camera. Those are updated by tracking the feature points on the moving camera images and map them to the feature points on the basis camera images as in Fig.4. Cross-correlations are computed for the feature points extracted within the search window between N frame and $N+1$ frame. The points that have high correlations are candidates of the tracked points.

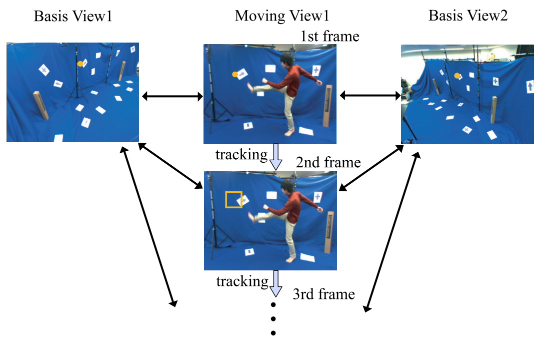


Figure 4: *Tracking of extracted feature points.*

We employ RANSAC (RAndom SAmple Consensus) algorithm to remove mistracked points. The inputs are the candidates of the tracked points and corresponding points on the basis camera images. The outliers are iteratively detected.

The new fundamental matrices are computed by all the tracked points except the mistracked points.

5. 3D RECONSTRUCTION

The 3D shape model from multiple-view images is reconstructed by using a volume intersection method, [21]. In the traditional volume intersection method, voxels in Euclidian grid space are projected onto the silhouette images by projection matrices. While, fundamental matrices between cameras are used to project voxels in PGS.

A certain number of voxels in a PGS are projected onto each silhouette image to check whether the projections are within the silhouette or not. The 3D shape model in PGS is reconstructed as a voxel model that consists of the voxels projected within the silhouette images.

Each silhouette image is synthesized by the chroma-keying and the noise removal. However, some part of background still remains, which is removed manually.

A voxel $A\,(p,q,r)$ in PGS is projected onto the image-coordinate $\boldsymbol{a}_1\,(p,q)$ in the basis camera1 in accordance with the definition of PGS. The point $\boldsymbol{a}_2\,(r,s)$ that is the projection of $A\,(p,q,r)$ to the basis camera2 is estimated with the epipolar line $\boldsymbol{l}$ that is the projection of the point $\boldsymbol{a}_1$ to the basis camera2 (Fig.5). The epipolar line $\boldsymbol{l}$ is represented as

$$\boldsymbol{l} = \mathbf{F}_{12} \begin{bmatrix} p \\ q \\ 1 \end{bmatrix} \quad (1)$$

where $\mathbf{F}_{12}$ is the fundamental matrix between the basis camera1 and the basis camera2. The point $\boldsymbol{a}_2$ is located on the point whose image-coordinate x equals r on the epipolar line $\boldsymbol{l}$.

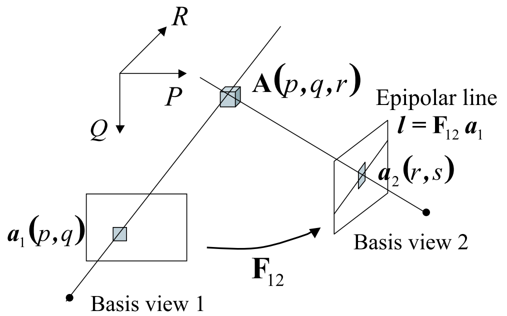


Figure 5: *Projection of a voxel to the two basis view.*

The voxel $A\,(p,q,r)$ is projected to the moving camera i as the point $\boldsymbol{a}_i$ whose position is determined by the two epipolar lines projected from the two basis cameras. The points $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ in the two basis cameras appear as the two epipolar lines $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$ respectively. The point $\boldsymbol{a}_i$ is located on the intersecting point of the epipolar lines $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$ (Fig.6).

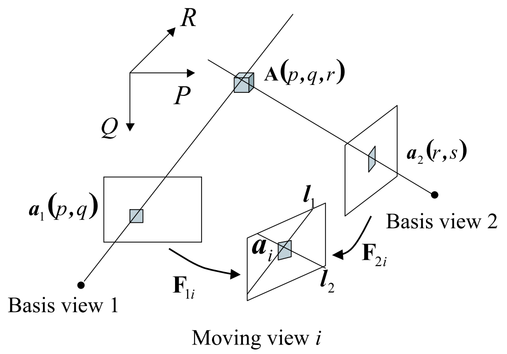


Figure 6: *Projection of a voxel to the moving view.*

The 3D shape model is reconstructed as the voxel model by projecting every voxel to each silhouette image as described above and judging with silhouette images. The voxel model are converted into the surface reconstructed model consists of the triangle patches by using Deformed cubes algorithm [22]. The 3D shape model converted are utilized for dense mapping the textures between the two input images to synthesize the images at free-viewpoint (Sec.6).

6. FREE-VIEW SYNTHESIS

Free-viewpoint images are synthesized by an image-based rendering method using the reconstructed 3D shape model. A method was proposed that synthesizes virtual-view images by interpolating the textures between the two neighboring input images [16], which is based on view interpolation method [4].
Because the location of virtual-view is limited between two input images, we can obtain relatively high quality virtual viewpoint images even if the recovered 3D shape is not accurate by employing such a view-interpolation-based method.

6.1. Generation of Z-Buffer

Z-Buffer of each input image is generated to judge the occlusions of the triangle patches for the input images in the rendering stage.
Each input view allocates the Z-Buffers, which are initialized.
All the tirangle patches on the 3D model surface are projected onto each Z-Buffer in the similar manner in Section 5 to generate Z-Buffer. In each pixel of the Z-Buffer the value is stored, which is the distance between the 3D-position of the input view and the 3D-position of the triangle patch on the 3D model surface which is projected onto the pixel. If some of patches are projected onto the same pixel on the Z-Buffer, the shortest distance is stored. Therefore, the Z-Buffer of each input view equals the range image.
The 3D-position of each view and the definition of the distance in PGS are necessary for the calculation of the distance. The former is estimated by the epipoles on the two basis views as described in Sec.2. The latter is represented as

$$D = \sqrt{(p_1 - p_2)^2 + (q_1 - q_2)^2 + (r_1 - r_2)^2} \quad (2)$$

where (p_1, q_1, r_1) and (p_2, q_2, r_2) are two arbitrary points in the PGS.

6.2. Rendering

Virtual viewpoint images are rendered by warping the two neighboring input images, and merging the two warped

images. The warped images are synthesized by shifting the position of every pixel in the input images. The pixel positions in the warped images are determined by interpolating the corresponding pixel positions in the two heigboring input images.
The correspondences of the pixels between the two neighboring input view-images are determined by projecting all the triangle patches on the 3D model surface onto the two views. Some triangle patches are occluded for either or both of the two input views, which results in the incorrect crrespondences of the pixels.
The Z-Buffer method is employed to detect such occlusions. Patches whose distance from a input view is different from the value stored in the Z-Buffer are judged to be occluded for the input view.
The free viewpoint images are synthesized by warping the two input views and merging them.
The position of a pixel $\boldsymbol{v}_3$ on the virtual viewpoint image is calculated by the weighted sum of the positions of the corresponding pixels $\boldsymbol{v}_1\,(x_1, y_1)$ on the input view1 and $\boldsymbol{v}_2\,(x_2, y_2)$ on the input view2 according to the following equation.

$$\boldsymbol{v}_3 = w \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + (1-w) \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \tag{3}$$

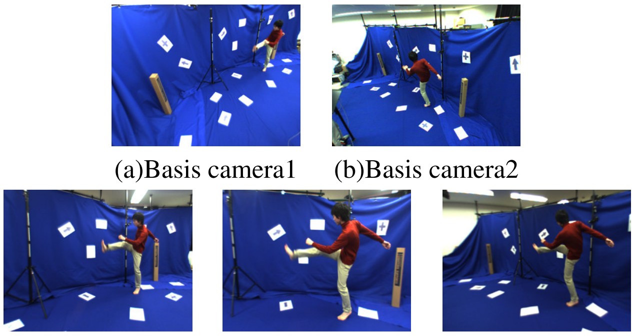
where w is the weight that defines the distance of the virtual view to the two input views.
The two warped images are synthesized by warping from $\boldsymbol{v}_1\,(x_1, y_1)$ to $\boldsymbol{v}_3$ and from $\boldsymbol{v}_2\,(x_2, y_2)$ to $\boldsymbol{v}_3$.
To merge the two warped images, the RGB colors of the pixel at $\boldsymbol{v}_3$ are also computed by the weighted sum of the colors of the two warped images. In the case of the occluded patches for either of the two input view, the weight of the colors of the pixels from the input view is equal to 0 and the weight of another is equal to 1.

7. EXPERIMENTAL RESULTS

In Sec.7.1 the performance of the proposed system is evaluated by synthesizing the free-view images from the captured images. The proposed system is compared with the system that consists of the only multiple fixed cameras in Sec.7.2
In our experiment the fixed cameras and the moving cameras have the same specifications and are synchronized for the synthesis of the free view-viewpoint images with the moving object. The moving cameras are moved by humans as many as the number of the cameras to capture the object within their FOV.

7.1. Evaluation of the performance

The proposed system consists of three moving cameras and two fixed cameras in this experiment. Three persons move each camera to capture a person moving in a laboratory room. The settings are non-horizontal settings as Fig.3(b) described in Sec.3. The captured images in a frame are shown in Fig.7, which have 640×480 resolution.



(a)Basis camera1 (b)Basis camera2
(c)Moving camera1 (d)Moving camera2 (e)Moving camera3

Figure 7: *Input views captured at a frame with the proposed system.*

Sampled free-views of the 3D surface model reconstructed from the input images are shown in Fig.8, which are rendered by OpenGL. To render the 3D model, the coordinates of the triangle patches in Euclidian-grid space are need to be known. The coordinates in the PGS are considered as the coordinates in the Euclidian grid space in Fig.8.

Fig.9 shows the images synthesized at the virtual-views by changing the value of the weight between the moving camera2 and the moving camera3. The free-viewpoint images can successfully be synthesized by the proposed system with the uncalibrated moving cameras. Some corruption and lack of the textures can be observed in the synthesized images, which are caused by the inaccuracy of the reconstructed shape. As long as we employ just a volume intersection method, such inaccuracy of the shape cannot be avoided. We will improve accuracy of shape reconstruction in the future work.

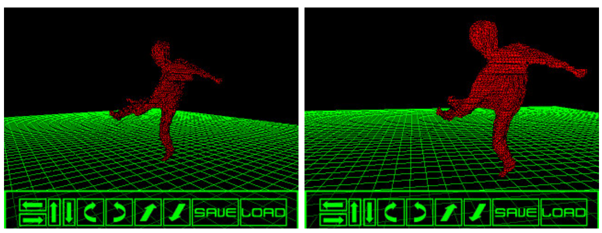
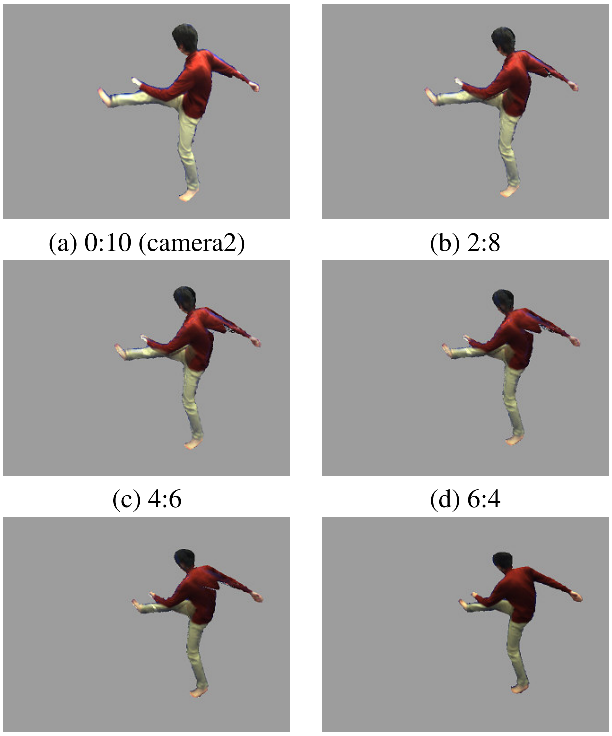


Figure 8: *3D shape model at some views.*

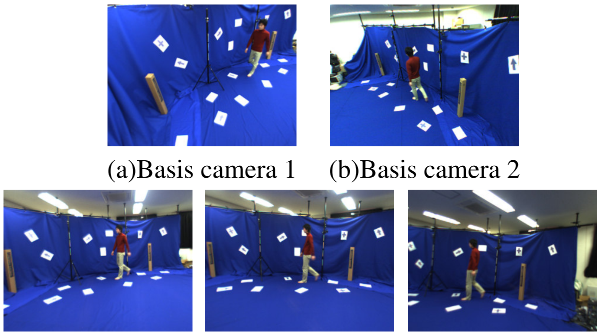


(a) 0:10 (camera2) (b) 2:8 (c) 4:6 (d) 6:4 (e) 8:2 (f) 10:0 (camera3)

Figure 9: *The virtual views between the moving camera 2 and the moving camera 3.*

7.2. Comparison with the proposed system and the conventional system

In this experiment the system that consists of only five fixed cameras is used as the conventional system to be compared with the proposed system. The three moving cameras in the settings as Fig.3(b) noted in Sec.3 are replaced with the three fixed cameras. The three fixed cameras are set far from the moving object to capture it constantly within their FOV. The captured images in a frame are shown in Fig.10 that have 640×480 resolution.
The virtual-views between the fixed camera 1 and the fixed camera 3 are shown in Fig.11, which are synthesized by the method same as the proposed method. The synthesized images in this experiment have same quality as the images synthesized with the proposed system in terms of the corruption and lack of the textures on the images.
By zooming the free-viewpoint images synthesized with the proposed system and with the conventional system, it is indicated that the object on the free-viewpoint images with the conventional system have only half of the vertical and horizontal resolution compared with the object on the images with the proposed system. If the relative size of the environment to the object is larger, the performance of the proposed system can become more visible.



(a)Basis camera 1 (b)Basis camera 2 (c)Fixed camera 3 (d)Fixed camera 4 (e)Fixed camera 5

Figure 10: *Input views captured at a frame with the conventional system.*

(a) 2:8 (b) 5:5 (c) 8:2

Figure 11: *The virtual views between the fixed camera 1 and the fixed camera 3.*

8. DISCUSSION

In each frame, the number of the corresponding points between a moving camera and a basis camera decreases more and more due to the occlusions or the disappearance from FOV. It results in a negative effect on the accuracy in the estimation of the fundamental matrices.
In our experiment, the accuracy of the fundamental matrices was not sufficient to obtain satisfactory reconstruction, when the camera motion was larger than 20% of the width of the FOV, for example 120 pixel horizontal shift in the images with 640 pixel width. Therefore, manual operation was required to pick-up the new feature points that corresponded to the feature points in the basis cameras. The manual corresponding operation had to be done every motion of 20% of the FOV (approximately 20 frames) for the satisfactory quality of the free-viewpoint images.
However, the correspondence procedure can also be achieved automatically by projecting the feature points that have corresponded between the two basis views onto each moving camera-image and making new correspondences. The details are described in the following.
After executing RANSAC algorithm described in 4, fundamental matrices among moving cameras and basis cameras are computed. The corresponding points between the two basis views that have not yet corresponded to the moving cameras are projected onto each moving camera-

image by the fundamental matrices.

The positions where those feature points are projected are close to the candidates of the corresponding points on each moving view. The feature points that are nearest neighbor to those positions are considered as the candidates of the corresponding points.

There are two cases for the candidates. First case is the visible point that is expected to be the new correspondence. The second case is the occluded point.

For deciding the case, a correlation between the candidate on the moving view and the corresponding points on the two basis views is computed. If the correlation is high, the point is regarded as the visible point. On the other hand, the point is regarded as occluded if the correlation is low.

The correlation with the simple rectangular window may not work due to the difference in appearance of the textures around the candidates. The correlation method such as wide baseline stereo matching with the affine invariant regions [23] can be employed to solve that problem.

9. CONCLUSION

We propose a novel method to synthesize free-viewpoint images for a moving object, which is captured by uncalibrated multiple moving cameras. We use multiple moving cameras that are able to capture the moving object in the center of the images with high resolution.

Two fixed cameras are employed for determining Projective grid space that defines a projective 3D coordinate in the object space for 3D reconstruction of the object from the multiple moving cameras without the calibration of them.

In the experiment, the efficacy of the proposed method was demonstrated by presenting high resolution free-viewpoint images that can be successfully synthesized by using the uncalibrated moving cameras and fixed cameras.

10. REFERENCES

[1] "http://www.ri.cmu.edu/events/sb35/tksuperbowl.html," Jan. 2005.

[2] "http://whatisthematrix.warnerbros.com/cmp/sfx-bullet_text.html," Jan. 2005.

[3] H. Saito and T. Kanade, "Shape reconstruction in projective grid space from large number of images," in *Proc. CVPR'99,Vol.2*, 1999, pp. 49–54.

[4] S. Chen and L. Williams, "View interpolation for image synthesis," in *Proc. of SIGGRAPH '93*, 1993, pp. 279–288.

[5] D. Scharstein, "View synthesis using stereo vision," in *Lecture Notes in Computer Science 1583, Springer*, 1999.

[6] M. Ott, J. Lewis, and I. Cox, "Teleconferencing eye contact using a virtual camera," in *INTERCHI93*, 1993, pp. 119–110.

[7] R. Yang and and Z. Zhang, "Eye gaze correction with stereovision for video tele-conferencing," in *In Proc. 7th European Conference on Computer Vision (ECCV2002), Volume I*, Copenhagen, May 2002, pp. 479–494.

[8] T. Kanade, P. W. Rander, and P. J. Narayanan, "Virtualized reality: concepts and early results," in *IEEE Workshop on Representation of Visual Scenes*, 1995, pp. 69–76.

[9] S. Vedula, P. W. Rander, H. Saito, and T. Kanade, "Modeling, combining, and rendering dynamic real-world events from image sequences," in *Proc. 4th Conf. Virtual Systems and Multimedia*, 1998, pp. 326–322.

[10] S. Moezzi, L.C. Tai, and P. Gerard, "Virtual view generation for 3d digital video," in *IEEE Multimedia, Vol 4, Issue 1*, 1997, pp. 18–26.

[11] H. Saito, S. Baba, and T. Kanade, "Appearance-based virtual view generation from multicamera videos captured in the 3-d room," in *IEEE Trans. on Multimedia, vol.5, no.3*, 2003, pp. 303–316.

[12] J. Carranza, C. Theobalt, M. Magnor, and H.-P. Seidel, "Free-viewpoint video of human actors," in *ACM Trans. on Computer Graphics (SIGGRAPH'03), vol. 22, no. 3*, July 2003, pp. 569–577.

[13] O. Grau, T. Pullen, and G.A. Thomas, "A combined studio production system for 3d capturing of live action and immersive actor feedback," in *IEEE Trans. Circuits and Systems for Video Technology*, Mar. 2004.

[14] B. Goldlucke and M. Magnor, "Real-time microfacet billboarding for free-viewpoint video rendering," in *Proc. IEEE International Conference on Image Processing (ICIP'03)*, Barcelona, Sept. 2003, pp. 713–716.

[15] M. Gross, S. Wurmlin, M. Naef, E. Lamboray, C. Spagno, A. Kunz, E. Koller-Meier, T. Svoboda, Luc Van Gool, S. Lang, K. Strehlke, A. Vande Moere, and O. Staadt, "blue-c: A spatially immersive display and 3d video portal for telepresence," in *Proceedings of ACM SIGGRAPH*, 2003, pp. 819–827.

[16] S. Yaguchi and H. Saito, "Arbitrary viewpoint video synthesis from multiple uncalibrated cameras," in *IEEE Trans. on Systems, Man and Cybernetics, PartB, vol. 34, no.1*, 2004, pp. 430–439.

[17] D. Iso and H. Saito, "Modeling and rendering in 3d coordinate defined by two cameras for shared virtual space communication," in *MIRAGE2003*, Mar. 2003, pp. 819–827.

[18] M. Pollefeys, R. Koch, and L. V. Gool, "Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters," in *International Journal of Computer Vision, 32(1)*, 1999, pp. 7–25.

[19] C. J. Harris and M. Stephens, "A combined corner and edge detector," in *In Proc. 4th Alvey Vision Conf.*, Manchester, 1988, pp. 147–151.

[20] R. Hartley, "In defense of the eight-point algorithm," in *IEEE Trans PAMI, Vol. 19, No.6*, 1997, pp. 580–593.

[21] A. Laurentini, "The visual hull concept for silhouette based image understanding," in *IEEE Trans. Pattern Analysis and Machine Intelligence, vol.16, no.2*, 1994, pp. 150–162.

[22] T. Nagae, T. Agui, and H. Nagahashi, "Object surface construction from volume data with appropriate topology," in *EICE D-II, Vol. J76-D-II, No. 8*, 1993, pp. 1704–1711.

[23] T. Tuytelaars and L. Van Gool, "Matching widely separated views based on affine invariant regions," in *Int. Journal of Computer Vision, 59(1)*, 2004, pp. 61–85.