Free Viewpoint Video Synthesis based on Natural Features Using Uncalibrated Moving Cameras

Songkran Jarusirisawad[1] and **Hideo Saito**[2], Non-members

ABSTRACT

In most of previous systems for free viewpoint video synthesis of a moving object, cameras are calibrated once at initial setting and can not zoom or change view direction during capture. Field of view of each camera in those systems must cover all the area in which the object moves. If the area is large, the object's resolution in the captured video and also in the free viewpoint video will become very low. To overcome this problem, we propose a novel method to synthesize free viewpoint video which allow cameras to be rotated and zoomed during capture. Projective Grid Space (PGS) which is 3D space defined by epipolar geometry of two basis cameras is used for object's shape reconstruction. By using PGS, geometrical relationship among cameras can be obtained from 2D-2D corresponding points between views. We use SIFT (Scale Invariant Transform) for finding corresponding points from natural features for dynamically registering cameras to PGS. In the experiment, free viewpoint video is successfully synthesized from multiple rotated/zoomed cameras without manual operation.

Keywords: Free viewpoint video, Image based rendering, Projective Grid Space, SIFT

1. INTRODUCTION

Free viewpoint image synthesis is the problem of creating an image of the same scene as input images but from the different viewpoint. Past years, this kind of research has become one of popular topics in computer vision.

In most of previous systems for free viewpoint image synthesis of a moving object, cameras are calibrated only once at initial setting, so they can not be zoomed or changed view direction during capture. Field of view of each camera in those systems must wide enough to cover all the area in which the object moves. If the area is large, moving object's resolution in the captured video and also in the free viewpoint video will become very low.

Allowing cameras to be rotated and zoomed during capture is one way for obtaining sufficient resolution of moving objects. During capture input video, cameras can be zoomed to capture only part of a scene where there is interested moving object. However, all cameras must be dynamically calibrated for synthesizing free viewpoint video.

In this paper, we propose a novel method for synthesizing free viewpoint video based on natural features from uncalibrated rotating/zooming cameras. Our method does not require any information about cameras parameters. For obtaining geometrical relationship among the cameras, Projective Grid Space (PGS) [1] which is 3D space defined by epipolar geometry between two basis cameras is used. All other cameras can be related to the PGS by fundamental matrices. Fundamental matrices for relating every cameras are estimated once at initial setting. After that, all cameras are dynamically registered to PGS via homography matrices. SIFT [2] is used for finding corresponding points between initial frame and the other frame for automatic homography estimation. We recover shape of objects by silhouette volume intersection [3] in PGS. The recovered shape in PGS provides dense correspondences among the multiple cameras, which are used for synthesizing free viewpoint images by view interpolation [4].

1.1 Related Works

One of the earliest researches for free viewpoint image synthesis of a dynamic scene is Virtualized Reality [5,6]. In that research, 51 cameras are placed around hemispherical dome called 3D Room to transcribe a scene. 3D structure of a moving human is extracted using multi-baseline stereo (MBS) [7]. Then free viewpoint video is synthesized from the recovered 3D model. Moezzi et al. synthesize free viewpoint video by recovering visual hull of the objects from silhouette images using 17 cameras [8]. Their approach creates true 3D models with fine polygons. Each polygon is separately colored thus requiring no texture-rendering support. Their 3D model can use standard 3D model format such as VRML (Virtual Reality Modeling Language) delivered though the internet and viewed with VRML browsers.

Many methods for improve quality of free viewpoint image have been proposed. Carranza et al. recover human motion by fitting a human shaped model

Manuscript received on February 28, 2007 ; revised on July 21, 2007.

[1,2] The authors are with Department of Information and Computer Science, Keio University 3-4-1 Hiyoshi, Kohoku-ku, Yokohama, 223-8522, JAPAN, Emails: songkran@ozawa.ics.keio.ac.jp and saito@ozawa.ics.keio.ac.jp

to multiple view silhouette input images for accurate shape recovery of the human body [9]. Starck optimizes a surface mesh using stereo and silhouette data to generate high accuracy virtual view image [10]. Saito et al. propose appearance-based method [11], which combines advantage from Image Based Rendering and Model Based Rendering.

In terms of computation time, real-time systems for synthesizing free viewpoint video have also been developed recently [12-14]. In the mentioned systems and most of the previous researches on free viewpoint image synthesis, they propose the systems that use calibrated fix cameras. Cameras in those systems are arranged to the specified positions around a scene and calibrated before capturing. During video acquisition, camera parameters must be the same, so cameras can not be moved or zoomed. Field of view (FOV) of all cameras must be wide enough to cover the whole area in which object moves. If the object moves around a large area, the moving objects resolutions in the captured video will not enough to synthesize a good quality free viewpoint image.

Ito et al. overcome this problem by proposing a system for synthesizing free viewpoint image using moving cameras [15,16]. Using their method high resolution free viewpoint image can be obtained. However, they demonstrate only in the case that a clear homogeneous color background with some artificial markers are placed around the scene because of the difficulty of feature point tracking for the moving camera.

In some situations, high resolution free viewpoint images are desired in a natural scene without any marker, for example for sports or outdoor events. In this paper, we propose a novel method to synthesize free viewpoint video of a moving object together with a background scene which is captured by rotating/zooming cameras [17].

2. OVERVIEW

There are two main difficulties of allowing moving cameras to translate, rotate and zoom during capture input video for free viewpoint synthesis in natural scene. The first one is cameras must be dynamically calibrated where there is no any special marker. Weak calibration technique like Projective Grid Space (PGS) [1], require only 2D-2D correspondences between cameras. Using PGS, there is no need of special markers. However, tracking or finding such corresponding points in 3D complex scene is difficult to acheive robustly as shown in [18]. Two images from different views have very different appearance due to motion parallax. To make the system practical for synthesizing a long video sequence, these calibration task must be done automatically.

The second one is silhouette segmentation of moving object for 3D shape reconstruction. If cameras are static, background scene can be captured beforehand, so it is trivial to get silhouette image using simple background subtraction. In the case that cameras can be translated, rotated and zoomed during capture, background image of these cameras cannot be captured before because it is impossible to recapture the scene with the same trajectory and zoom.

To reduce these problems, we assume that capturing position of cameras are not much changed during capture. Even such assumption, we can still give a flexibility of allowing cameras to be zoomed and/or change view direction to capture moving object, e.g. camera is placed on a tripod and rotated/zoomed freely. By using this assumption we can resolve two stated problems as following.

At initial frame of each input video, we capture the whole background scene without moving object. We select two cameras for defining PGS and weakly calibrate initial frames to PGS manually. To register the other frames to PGS, homographies which relate those frames to initial frames are estimated automatically. Because initial frame and other frame captured from the same position, there is no motion parallax between these images and two images are approximately 2D similarity. Accurate corresponding points can be found automatically using SIFT as will be described in section 4.2.

For silhouette segmentation of moving object, background image of moving cameras are created by warping initial frame where there is no moving object using the same homography for registering moving cameras to PGS.

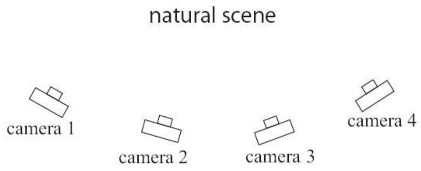


Fig.1: Cameras Configuration.

Our system consist of 4 cameras on the tripods like Fig.1. All cameras are zoomed and rotated independently by man during capture. The overall process is illustrate in Fig.2.

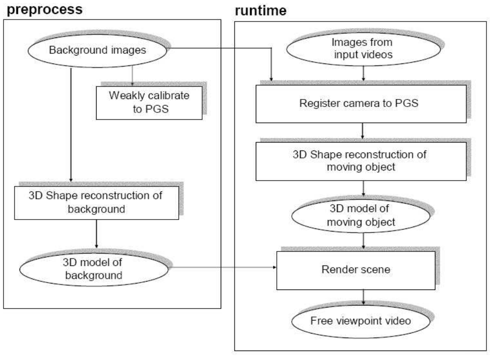


Fig.2: Overall Process.

3. PROJECTIVE GRID SPACE

Reconstructing 3D model for synthesizing free viewpoint image requires a relation between 3D world coordinate and 2D image coordinate. Projection matrix that represents this relation can be estimated by strong camera calibration which requires 3D-2D correspondences. Measuring 3D-2D corresponding points requires a lot of work. Moreover, in case of a large natural scene, it is difficult to precisely measure calibrating points throughout all the area.

To remove effort of obtaining strong calibration data, we use a weak calibration frame work, called Projective Grid Space (PGS) [1], for shape reconstruction. 3D coordinate in PGS and 2D image coordinate is related by epipolar geometry using fundamental matrices. To estimate fundamental matrices between views, only 2D-2D correspondences which can be directly measured from input videos are required.

3D space in PGS is defined by image coordinates of two arbitrarily cameras. These two cameras are called the basis camera1 and the basis camera2. The nonorthogonal coordinate system P-Q-R is used in PGS. The image coordinates x and y of basis camera1 corresponds to the P and Q axis in PGS. Image coordinate x of the basis camera2 corresponds to the R axis.

Fig.3 illustrates how PGS is defined. 3D coordinate A (p,q,r) in PGS is projected on image coordinate a_1 (p,q) of the basis camera1 and on image coordinate a_2 (r,s) of the basis camera2. a_2 is the point on epipolar line of point a_1 where image coordinate x equals to r.

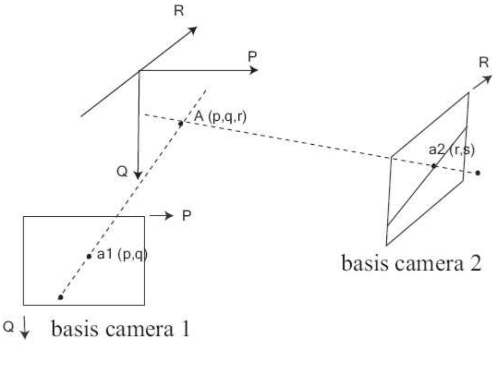


Fig.3: *Projective Grid Space is Defined by Two Basis Cameras.*

Other cameras can be related to PGS by fundamental matrices between 2 basis cameras. Finding such fundamental matrices required only 2D-2D correspondences. So, it is relatively easy comparing to full calibration which required 3D-2D correspondences. 3D coordinate A (p,q,r) in PGS is projected onto non-basis camera at point a_i which is the intersection between epipolar line l_1 and l_2 as shown in Fig.4.

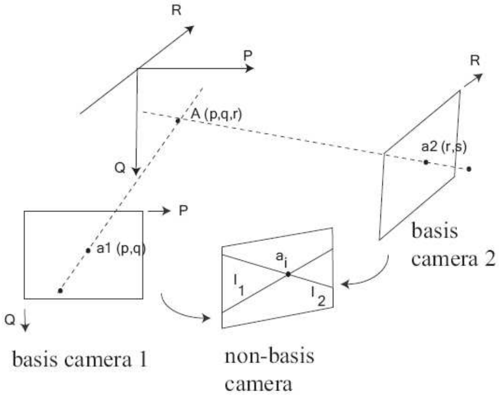


Fig.4: *3D Coordinate in PGS is Projected to Non-Basis Camera Using Fundamental Matrices.*

Epipolar line l_1 and l_2 are calculated from the equations 1 and 2.

$$l_1 = F_{1i} \begin{pmatrix} p \\ q \\ 1 \end{pmatrix} \tag{1}$$

$$l_2 = F_{2i} \begin{pmatrix} r \\ s \\ 1 \end{pmatrix} \tag{2}$$

where F_{1i} and F_{2i} are fundamental matrix from basis camera1 and from basis camera2 to non-basic camera respectively.

In Fig.5, 3D camera position of the basis camera1 in PGS is ($C1_x$, $C1_y$, $e12_x$), where ($C1_x$, $C1_y$) is camera center in the basis camera1, and ($e12_x$, $e12_y$) is epipole of basis camera1 in basis camera2. In the same way, camera position of the basis camera2 is ($e21_x$, $e21_y$, $C2_x$), where ($e21_x$, $e21_y$) is the epipole of the basis camera2 in the basis camera1, and ($C2_x$, $C2_y$) is camera center in basis camera2. For non-basis camera, 3D camera position in the PGS is ($e1_x$, $e1_y$, $e2_x$) where ($e1_x$, $e1_y$) and ($e2_x$, $e2_y$) are epipoles on basis camera1 and basis camera2, respectively.

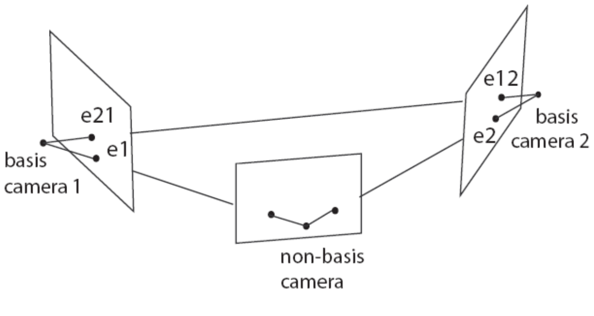


Fig.5: *Camera Position in Projective Grid Space.*

4. WEAK CALIBRATION

4.1 Preprocess

At initial frame, we zoom out all cameras to capture the whole area of a scene without object. We call this background image of camera i as bg$_i$. We select camera1 and camera4 as basis cameras defining PGS. 2D-2D Corresponding points for estimating fundamental matrices between basis cameras and other cameras are assigned manually on bg$_i$ image during preprocess. Once fundamental matrices are estimated, 3D coordinate in PGS can be project to all bg$_i$ images. These images will be used for generate virtual background for background subtraction and also used as reference image for register moving cameras to PGS as will be described in section 4.2. Fig.6 shows background images of our experiment.

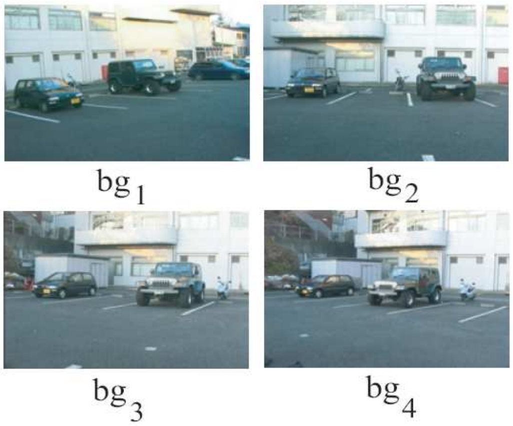


Fig.6: *Background Images.*

4.2 Runtime

During capture input video, object will move around a large space. Each camera is zoomed and rotated to capture moving object with high resolution in the image. View and focal length of each camera are changed from initial frame. Fundamental matrices estimated during preprocess can not be used to project 3D coordinate in PGS to 2D coordinate of the other frames. From the assumption that capturing position of each cameras is not much changed during capture, 2D coordinate of bg$_i$ can be transformed to 2D coordinate of the other frames of the same camera using homography matrix.

To estimate homography matrix, corresponding points between bg$_i$ and the other frame are necessary. We employ SIFT (Scale Invariant Feature Transform) [2], which is the method for extracting features from images that can be used to perform reliable matching, for finding such corresponding points. Coorresponding point initially found by SIFT include some outlier. We employ RANSAC (RAndom Sample Consensus) [19] using homography constraint to remove those outliers. Only inliers are used for finding accurate homography.

3D coordinate $A(p, q, r)$ in PGS which is projected on (x_{bg}, y_{bg}) of bg$_i$ image is projected to the other frame of the same camera at (x_{other}, y_{other}) by equation 3.

$$s\begin{pmatrix} x_{other} \\ y_{other} \\ 1 \end{pmatrix} = H_i \begin{pmatrix} x_{bg} \\ y_{bg} \\ 1 \end{pmatrix} \quad (3)$$

where H_i is homography matrix between bg$_i$ image and the other frame.

Example corresponding points that automatically found using SIFT are shown in Fig.7. In Fig.7, the left image is bg$_i$ image and the right image is the other frame which will be registered to Projective Grid Space. Fig.7(a) shows all corresponding points found by SIFT while Fig.7(b) shows only inliers after outliers are removed by RANSAC.

SIFT is robust for finding corresponding points between images which there are different in scale and 2D rotation. However, reliable of matching will decrease if there are perspective distortion between two images. In our case, two images are captured from approximately the same position. There is no motion parallax between these images. Two images are approximately 2D similarity regardless of complexity of a scene. Therefore, SIFT is very robust for using in our system.

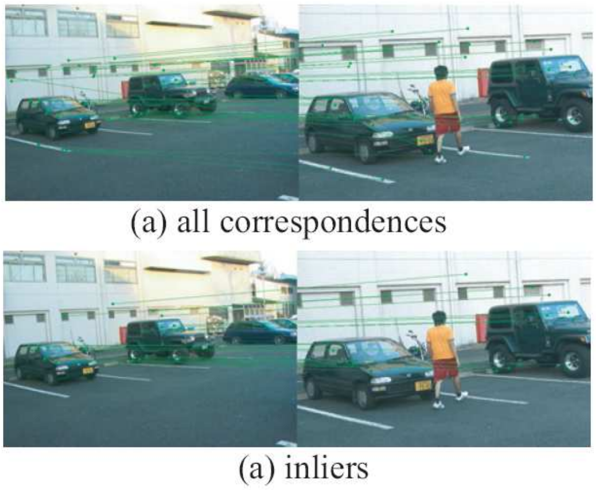


Fig.7: *Corresponding Points Found Using SIFT for Estimating Homography.*

5. 3D RECONSTRUCTION

We consider that objects in input video consist of

- Background planes
- Static objects
- Moving object (Human)

which will have the different way to recover the 3D information for rendering free viewpoint image. Background plane is the real plane like a floor or the scene which is far away so that can be approximated as planar scene. Fig.8 shows how background scene is catagorized. Static objects and background planes are not changed during video capture, so 3D information of these are estimated only once during preprocess while 3D shape of moving object is reconstructed automatically every frame.

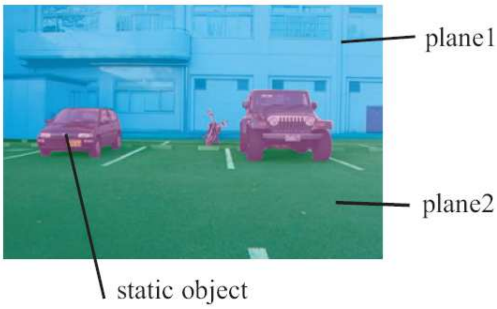


Fig.8: *Background Scene Consist of Planes and Static Objects.*

5.1 Preprocess

From bg_i images of all cameras, 3D shapes of static objects are reconstructed using silhouette volume intersection method [3]. Silhouette images of each static object are segmented manually. Each voxel in PGS is projected onto silhouette images to test voxel occupacy. Surfaces of 3D voxel model are extracted to 3D triangular mesh model using Marching Cube algorithm [20]. This 3D triangular mesh model will be used for making dense correspondences for view interpolation.

During preprocess we reconstruct 3D position of several points which lie on a plane by assigning corresponding points between two basis cameras defining PGS. Let a_1(p,q) is 2D coordinate of point in basis camera1 and a_2(q,r) is 2D coordinate of point in basis camera 2, 3D position of this point is (p,q,r) from definition of PGS in section3. These 3D position in PGS will be used for render planes in free viewpoint video as will be explained in section 6.1.

5.2 Runtime

3D model of moving object is reconstructed using silhouette volume intersection method in the same way as static object. The difference is 3D model of moving object is reconstructed every frame automatically. Silhouette of moving object needs to be segmented from input image. From preprocess, we have background image of a whole scene, called bg_i, captured from each camera. Background image of camera i during runtime is generated by warping bg_i image using homography estimated automatically in section 4.2. Example generated background images for moving camera are shown in Fig.9. After generating virtual background for moving cemera, silhouette image is segmented by background subtraction as in Fig.10.

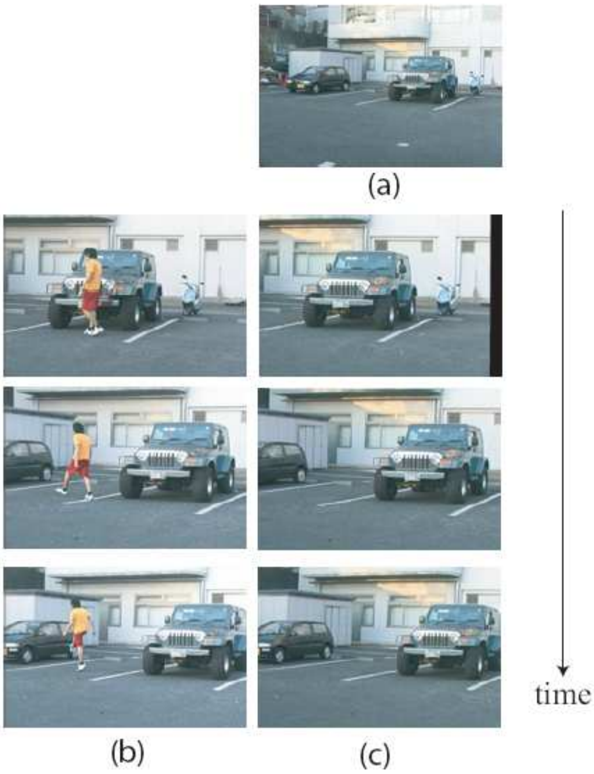


Fig.9: *(a) Background Image. (b) Some Frames from Input Video of the Same Camera as (a). (c) Automatically Generated Background Images from (a).*

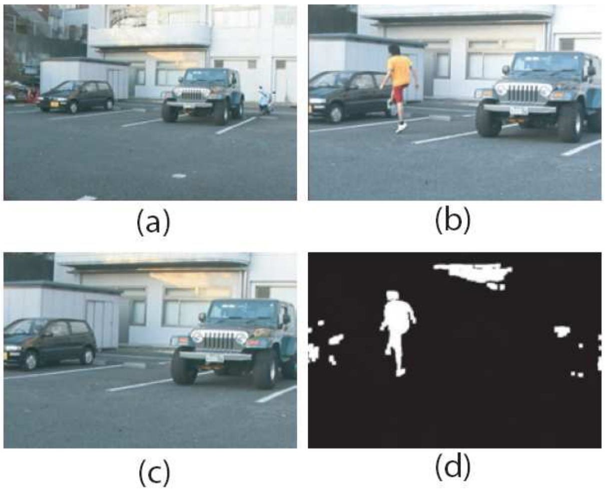


Fig.10: *(a) Background Image. (b) One Frame from Input Video of The Same Camera as (a). (c) Automatically Generated Background from (a). (d) Silhouette Image of (b).*

Voxels in PGS are projected onto bg_i image first, then the projected 2D coordinate is transfered to current frame using equation 3. Voxel is considered to be in a 3D model volume if projected points of all cameras are in silhouette. Surfaces of 3D voxel model are extracted to 3D triangular mesh model using Marching Cube algorithm [20]. Fig.11 shows 3D model of moving object reconstructed in PGS.

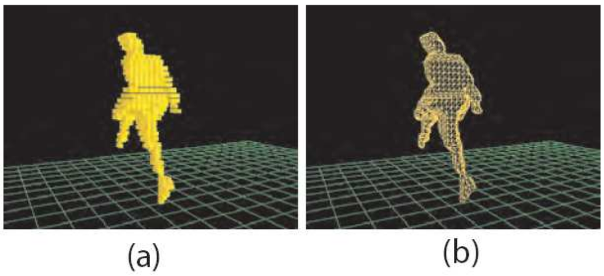


Fig.11: *3D Model of Human in PGS. (a) Volumetric Representation (b) Triangular Mesh Representation.*

6. FREE VIEWPOINT RENDERING

Our method can synthesize free viewpoint between any two reference views. Free viewpoint image are rendered in two steps. Background planes in scene are rendered first. Moving and static objects are then rendered overlay to synthesized planes. The following subsections explain the detail of two rendering phase.

6.1 Planes Rendering

During preprocess, 3D position of points which lie on planes are already reconstructed. These 3D position in PGS are projected on to both reference views. 2D positions of these points on free viewpoint image are determined using linear interpolation as equation

$$\begin{pmatrix} x \\ y \end{pmatrix} = w \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + (1-w) \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \qquad (4)$$

where w is a weight, ranging from 0 to 1, defining the distance from virtual view to second reference view. $(x_1, y_1)^T$ and $(x_2, y_2)^T$ are corresponding points on the first reference view and the second reference view respectively. Corresponding points between background image of reference view and virtual view are used for estimating homography. Plane in background image which is segmented during preprocess is warped to virtual view. Warped planes from two reference views are then merged together. In case that the scene consists of more than one plane, two or more plane in virtual view are synthesized in this way and merge together. Fig.12 illustrates how the plane is rendered in free viewpoint image.

6.2 Objects Rendering

Free viewpoint images of static and moving objects are synthesized by an image-based rendering method. 3D triangular mesh model of static objects and moving object are combined together. Combined 3D model is used for making a dense correspondence and also for testing occlusion between reference images. Each corresponding triangular mesh is warped to virtual viewpoint image based on view interpolation method [4].

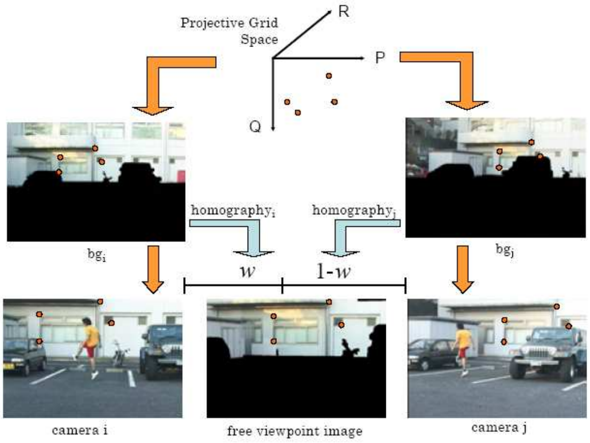


Fig.12: *Rendering Plane on Free Viewpoint Image.*

To test occlusion triangular patches, Z-Buffer of each camera is generated. All triangle patches of a combined 3D model are projected onto Z-Buffer of each camera. Pixel value of Z-Buffer is store the 3D distance from cameras optical center to projected triangle patch. If some pixels are projected by more than one patch, the shortest distance is stored. The distance of point $a(p_1, q_1, r_1)$ and $b(p_2, q_2, r_2)$ in PGS is defined as equation [5].

$$D = \sqrt{(p_1 - p_2)^2 + (q_1 - q_2)^2 + (r_1 - r_2)^2} \qquad (5)$$

To synthesize free viewpoint image, each triangle mesh is projected onto two reference images. Z-Buffer is used to test occlusion. Patch whose distance from input camera focal point is different from the value stored in the Z-Buffer is decided to be occluded. In the case that a patch is occluded in both two input views, this patch will not be interpolated in a free viewpoint image. If a patch is seen from one or both input views, this patch will be warped and merged into a new viewpoint image. Position of a warped pixel in new viewpoint image is determined by equation [4].

To merge two warped triangular patch, RGB colors of the pixel are computed by the weighted sum of the colors from both warped patch. If a patch is seen from both input view, weight that use for interpolating RGB color is the same for determining position of a patch. In case that patch is occluded in one view, weight of occluded view is set to 0 while the other view is set to 1. Fig.13 shows example of free viewpoint image of static and moving objects.

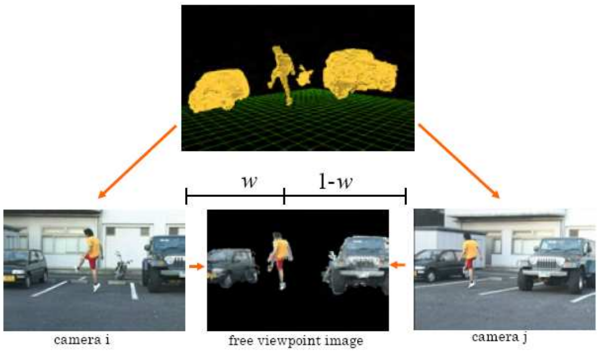


Fig.13: *Rendering Static and Moving Objects on Free Viewpoint Image.*

7. EXPERIMENTAL RESULTS

In this section, we show and evaluate the result from our proposed method. Section 7.1 shows free viewpoint video synthesized by our proposed method, which allow cameras to be rotated and zoomed freely to capture high resolution moving object. To compare our result with conventional system, section 7.2 shows the result free viewpoint video from the same scene but capture by fixed cameras.

The experimental environment is a large natural scene as Fig.6. We use four Sony-DV cameras with 720x480 resolutions in both experiments. All cameras are placed on tripods in front of the scene as in Fig.1.

7.1 Free Viewpoint Video from Our Proposed Method

We synthesize free viewpoint video from consecutive 520 frames of 4 input videos by our proposed method. During 520 frames, moving cameras which are placed on tripods have been zoomed and rotated to capture high resolution moving object independently by man. Approximately, each camera is rotated on tripod around -20 to 20 degree. There is no artificial marker placed in the scene.

Only natural features are used for finding corresponding points. Our method can correctly register all frames to PGS and synthesize free viewpoint video without manual operation. The number of correspondences during 520 frames which are found using SIFT before and after removing outlier as describe in section 4.2 is shown in Fig.14 (only one from four cameras is shown in this graph). To estimate homography matrix, at least four correspondences are necessary. From Fig.14, the number of inliers in every frame is more than four points which is sufficient for estimating homography.

To evaluate accuracy of camera registration, fitting errors of homography before and after removing outliers by RANSAC is shown in Fig.15. Graph in Fig.15 use log scale because the error of homography estimated from all correspondence is relatively large. This means that outliers removal works well for obtaining more correct registration in the proposed method.

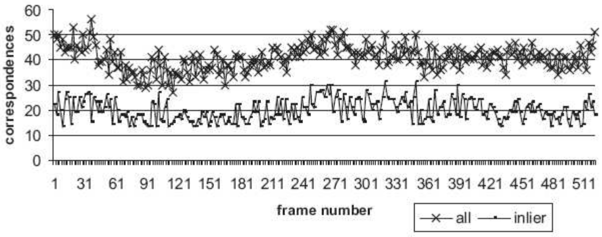


Fig.14: *Number of Correspondence for Registering Camera to PGS.*

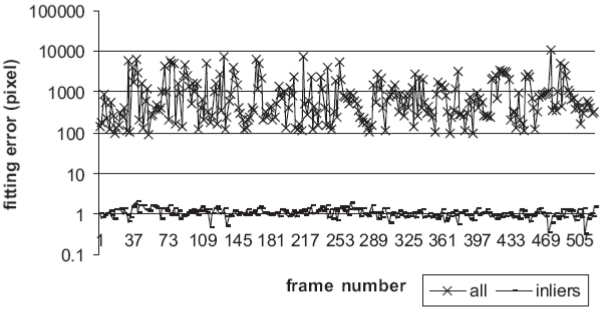


Fig.15: *Fitting error of homography from All Correspondences and Only Inliers.*

Fig.16 shows some example frames from the result free viewpoint video.

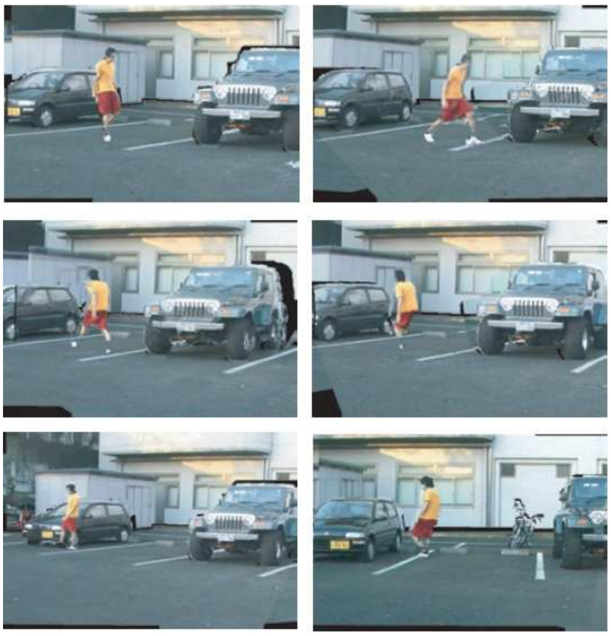


Fig.16: *Example Free Viewpoint Video from Consecutive 520 Frames.*

To compare quality of interpolated images with the original images, we select one frame from the input video as Fig.17 to synthesize free viewpoint images at several weight ratios between camera2 and camera3. The result free viewpoint images are shown in Fig.18. We can see that the rendered background planes, static objects and moving object from both reference views are correctly aligned and merged in the free viewpoint images. Occlusion areas between two reference views, e.g. motorcycle, are also correctly rendered.

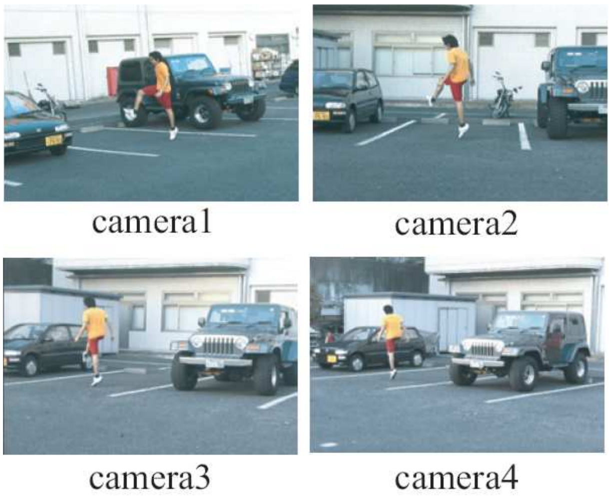


Fig.17: *One Frame from Four Input Videos.*

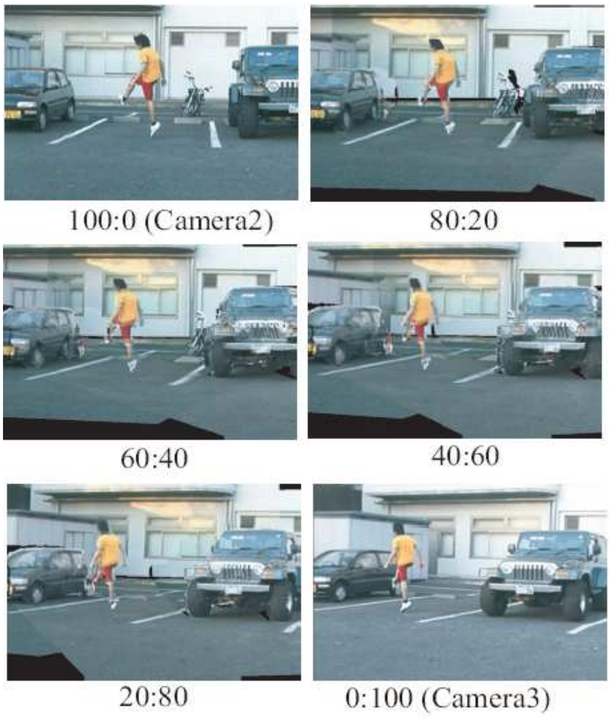


Fig.18: *Free Viewpoint Images from Our System. Ratio Under each Figure is The Weight Ratio between Camera2 and Camera3*

There are some blured texture of static objects and moving object in synthesized image due to inaccuracy of 3D model. This accuracy can be improved easily by increasing number of cameras in the system. In terms of reconstruction algorithm, other technique can be used together with shape from silhouette to improve quality of reconstructed 3D model as well [21].

In Fig.16, some parts of moving object are missing. This happens because of incomplete silhouette image from some views. Now we work in natural scene, a completely clear silhouette from background subtraction is hard to achieve even in case that fixed cameras are used like conventional system.

7.2 Comparison with Fixed Multiple Cameras Input

To compare our proposed method that using moving cameras with fixed cameras system, we synthesize free viewpoint video in the same scene as section 7.1 from 4 fixed cameras. We set a man to move around the same space in both systems. Fig.19 shows the difference between fixed cameras input and moving cameras input. Please note that a man move to the left car and the right car in both input which means that captured area is the same in both systems. Cameras in our system can be zoomed and rotate to capture only part of a scene where there is a man. In contrast, cameras in fixed cameras system must be zoomed out so that field of views are wide enough to cover the area in which the human moves and cannot zoom or rotate cameras during capture.

Resolution of human in input video of fixed cameras system is low compared to moving cameras system. Fig.20 shows the result of free viewpoint images from fixed cameras system while result from our moving cameras system are already shown in section 7.1. Moving object's size in free viewpoint image obtained from fixed cameras system is small and have less detail compared to moving cameras system. This difference can be more visible if the relative size between the scene and moving object become larger.

8. CONCLUSIONS

We proposed a novel method to synthesize free viewpoint video of a moving object in natural scene, which is captured by rotating/zooming cameras. Cameras in our system are all uncalibrated. Neither intrinsic nor extrinsic parameters are known. Projective Grid Space [1] which is 3D space defined by image coordinate of two cameras is used for 3D shape reconstruction. By using PGS, geometrical relationship among cameras can be obtained from 2D-2D corresponding points between views. Every frame, homography is used to register rotating/zooming cameras to PGS. We employ SIFT [2] for finding corresponding points to estimate homography automatically. In our experiment in natural scene without any special marker for demonstrating the efficacy of the proposed method, free viewpoint video is successfully synthesized without manual operation.

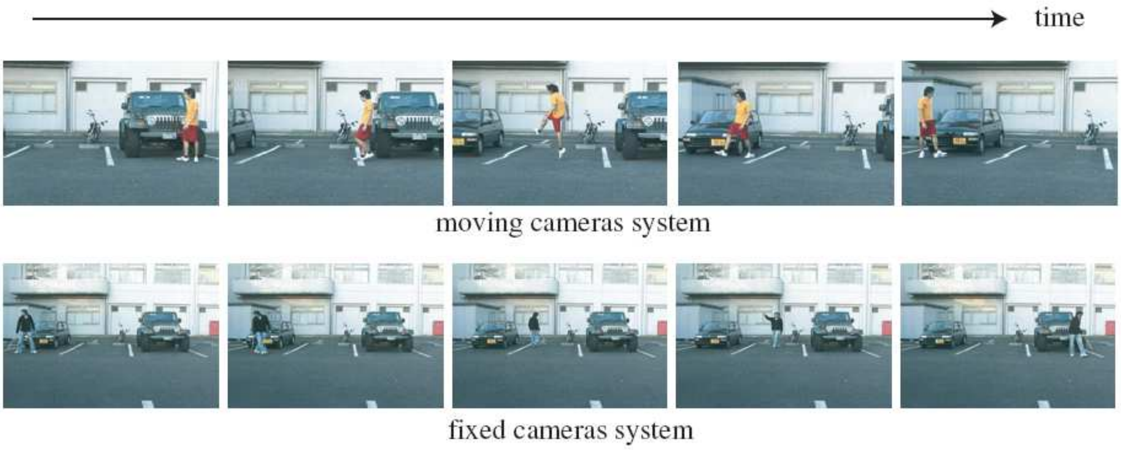
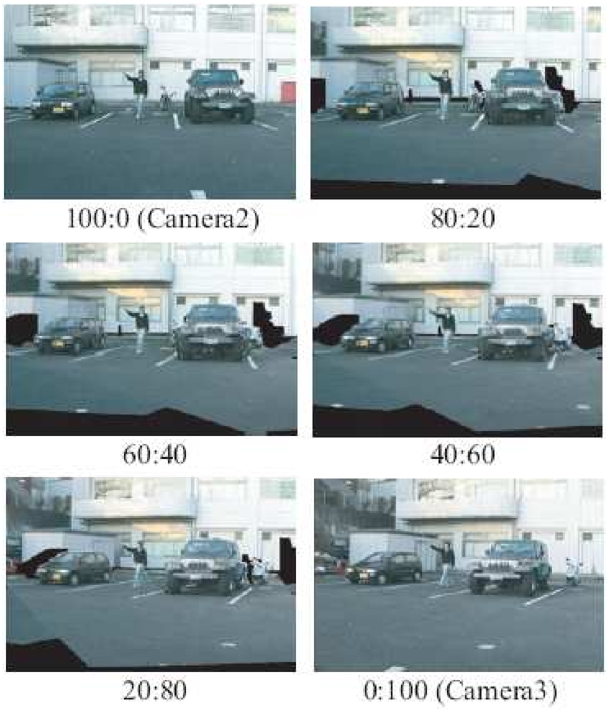


Fig.19: *Example Input Frames from One Camera.*



100:0 (Camera2) 80:20

60:40 40:60

20:80 0:100 (Camera3)

Fig.20: *Free Viewpoint Images from Fixed Cameras System. Ratio Under each Figure is The Weight Ratio between Camera2 and Camera3*

References

[1] H. Saito, T. Kanade, "Shape Reconstruction in Projective Grid Space from Large Number of Images," *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR-99)*, June, 1999, pp. 49-54.

[2] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, 60, 2 (2004), pp. 91-110.

[3] A. Laurentini, "The Visual Hull Concept for Silhouette Based Image Understanding," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.16, no.2, pp. 150-162, 1994.

[4] S. Chen and L. Williams, "View interpolation for image synthesis," *Proceedings of SIGGRAPH'93*, pp. 279288, 1993.

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

[6] S. Vedula, P. W. Rander, H. Saito, and T. Kanade, "Modeling, Combining, and Rendering Dynamic Real-World Events From Image Sequences," *Proceedings. of 4th Conference Virtual Systems and Multimedia*, Vol. 1, pp. 326-322, 1998.

[7] M. Okutomi and T. Kanade, "A multiple-baseline stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(4), pp. 353-363, 1993.

[8] S.Moezzi, L.C.Tai, P.Gerard, "Virtual View Generation for 3D Digital Video," *IEEE Multimedia*, 4, 1, pp. 1826, 1997.

[9] J.Carranza, C.Theobalt, M.Magnor, H.-P. Seidel, "Free-Viewpoint Video of Human Actors," *ACM Trans. on Computer Graphics*, vol. 22, no. 3, pp. 569-577, July 2003.

[10] J. Starck and A. Hilton, "Towards a 3D virtual studio for human appearance capture," *Proceedings of IMA International Conference on Vision*, Video and Graphics, Bath, 2003.

[11] H. Saito, S. Baba, and T. Kanade, "Appearance-Based Virtual View Generation From Multicamera Videos Captured in the 3-D Room," *IEEE Transactions on Multimedia*, Vol. 5, No. 3, September, 2003, pp. 303-316.

[12] B.Goldlucke and M.Magnor, "Real-time microfacet billboarding for freeviewpoint video rendering" *Proceedings of IEEE International Conference on Image Processing (ICIP03)*, Barcelona, Sept. 2003, pp. 713716.

[13] O. Grau, T. Pullen, G.A. Thomas, "A Combined Studio Production System for 3D Captur-

ing of Live Action and Immersive Actor Feedback," *IEEE Trans. Circuits and Systems for Video Technology*, March 2004.

[14] V. Nozick, S. Michelin and D. Arqus, "Real-time Plane-sweep with local strategy," *Journal of WSCG*, Vol.14, No.1-3, pp. 121-128, 2006.

[15] Y. Ito and H. Saito, "Free viewpoint image synthesis using uncalibrated multiple moving cameras," *Computer Vision / Computer Graphics Collaboration Techniques and Applications (MIRAGE2005)*, pp.173-180, March, 2005.

[16] Y. Ito, H. Saito, "Free-viewpoint image synthesis from multiple-viewimages taken with uncalibrated moving cameras," *The IEEE International Conference on Image Processing (ICIP05)*, Italy, Sep, 2005

[17] S. Jarusirisawad and H. Saito, "New Viewpoint Video Synthesis in Natural Scene Using Uncalibrated Multiple Moving Cameras," *International Workshop on Advanced Imaging Technology IWAIT 2007*, pp. 78-83, Bangkok, Thailand, January 2007.

[18] P. Moreels and P. Perona, "Evaluation of features detectors and descriptors based on 3d objects," *International Journal of Computer Vision*, 73(3), 2007, pp. 263284.

[19] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography," *Communication Association and Computing Machine*, 24(6), pp. 381-395, 1981.

[20] Lorensen W.E. and Cline H.E., "Marching cubes: A high resolution 3d surface construction algorithm," *Proceedings of SIGGRAPH 87*,Computer Graphics, Vol. 21, No. 4, pp. 163-169, July 1987.

[21] S.Yaguchi, H.Saito, "Improving Quality of Free-Viewpoint Image by Mesh Based 3D Shape Deformation," *Proceedings of The 14th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision(WSCG2006)*,February, 2006

Songkran Jarusirisawad received the B.E.(first class honors) degree in computer engineering from Chulalongkorn University, Bangkok, Thailand, in 2005 and M.E. degree in information and computer science from Keio University, Yokohama, Japan, in 2007. Currently, he is a Ph.D. student in the Department of Information and Computer Science, Keio University. His research interests include computer vision, image based modeling and rendering.

Hideo Saito received the B.E., M.E., and Ph.D. degrees in electrical engineering from Keio University, Yokohama, Japan, in 1987, 1989, and 1992, respectively.

He has been on the faculty of Department of Electrical Engineering, Keio University, since 1992. From 1997 until 1999, he was a Visiting Researcher with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA. Since 2006, he has been a Professor in the Department of Information and Computer Science, Keio University. Since 2000, he has also been a Researcher with Precursory Research for Embryonic Science and Technology (PRESTO), Japan Science and Technology Corporation (JST), Tokyo, Japan. He has been engaging in the research areas of computer vision, image processing, and human-computer interaction.

Dr. Saito is a member of The Institute of Electronics, Information and Communication Engineers (IEICE), Information Processing Society Japan (IPSJ), The Society of Instrument and Control Engineers (SICE), and The Virtual Reality Society of Japan (VRSJ).