FREE-VIEWPOINT IMAGE SYNTHESIS FROM MULTIPLE-VIEW IMAGES TAKEN WITH UNCALIBRATED MOVING CAMERAS

Yosuke Ito and Hideo Saito

Department of Information and Computer Science, Keio University Yokohama, Japan, 223-8522

ABSTRACT

In this paper, we propose methods for free-viewpoint image synthesis from multiple-view images taken with uncalibrated cameras. In our method, two viewpoints are selected as basis images for defining a projective 3D coordinate of the object scene in which the scene structure is recovered from the input multiple-view images. The 3D coordinate is defined by the image coordinates of the selected two basis images according to the epipolar geometry between those two images, which is represented by a fundamental matrix. The multiple-view images are also related to the projective 3D coordinate by their epipolar geometry to the basis images. Based on such a framework, we do not need to strongly calibrate the cameras, so we can recover 3D structure of the scene without effort for the camera calibration. In addition to that, we can also render free-viewpoint images from hand-held moving cameras. We will demonstrate the effectiveness of the proposed method by showing free-viewpoint images from multi-view images taken with hand-held moving cameras.

1. INTRODUCTION

One popular topic in computer vision area is new view synthesis from multiple cameras. In most of the researches on new view synthesis, objects are supposed to be captured within the FOV of every camera. If the objects move around the scene, the FOV of each camera 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.

Using 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 views from multiple moving cameras.

In this paper, we propose a new method for synthesizing free-viewpoint video by moving multiple cameras. 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 a Projective Grid Space (PGS) [10] based on those two fixed cameras. All the moving cameras can be geometrically related to the PGS by computing fundamental matrices of each moving camera with the 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 free-viewpoint synthesis by view interpolation [2].

1.1. Related Work

The Virtualized Reality Project by Kanade et.al. [6] is one of the earlier research attempts for image-based rendering from multiple cameras system. Moezzi et.al. also synthesize free viewpoint video by recovering visual hull of objects from silhouette images [8]. Carranza et.al. recover human motion by fitting a human shaped model to multiple view silhouette input images for accurate shape recovery of the human body. This provides high quality free viewpoint videos of a human [1].

In most of these research efforts, multiple cameras are fixed and calibrated. In order to avoid the effort to fully calibrate multiple cameras, Saito and Kanade have proposed the Projective Grid Space [10], which can be defined from just fundamental matrices among 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 [11]. Another method for avoiding calibration is applying self calibration method to multiple cameras. Self calibration method was proposed by Pollefeys [9] and is applied in the 3D studio system used in [1, 4].

The proposed method in this paper is based on PGS [10, 11]. 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.

0-7803-9134-9/05/$20.00 ©2005 IEEE

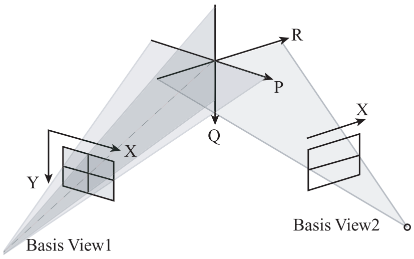


Fig. 1. Definition of the Projective Grid Space.

2. SYSTEM ENVIRONMENT

To reconstruct a 3D model without full camera calibration, we employ PGS (Projective Grid Space)[10] and recover a 3D model in PGS.

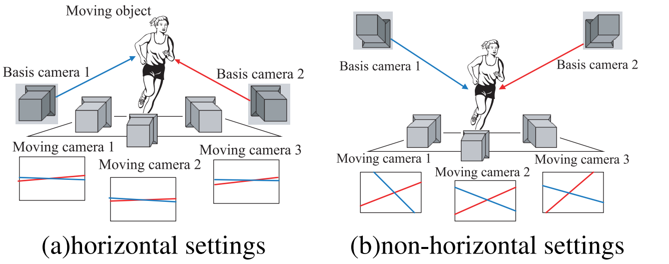
In our method, it is difficult to fully calibrate cameras for 3D reconstruction methods. The reason is that it is almost impossible to put markers with known 3D positions in the scene and to obtain 3D-2D correspondences at every time instance in the motion of the cameras.

An alternative way is 3D reconstruction in PGS, which is a scheme for easy definition of 3D space by fundamental matrices among cameras. The fundamental matrices are computed by 2D-2D correspondences, which is relatively easily measured. A PGS is defined by the image coordinates of two arbitrarily selected cameras from a set of multiple cameras from a set of multiple cameras. These two cameras are called basis camera1 and basis camera2. The nonorthogonal coordinate system P-Q-R is used in PGS. As shown in Fig.1, the image coordinates x and y of basis camera1 correspond to the P and Q coordinates in PGS. The image coordinate x of basis camera2 corresponds to the R coordinate. For instance, if a 3D point in PGS is projected onto (x_1, y_1) and (x_2, y_2) on the image of basis camera1 and the image of basis camera2 respectively, the location of the 3D point in PGS is determined as (x_1, y_1, x_2).

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

We consider two kinds of camera settings, horizontal settings and non-horizontal settings as shown in Fig.2. On the images of each moving camera 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 our proposed method (Fig.2(b)), the position of the point where the two epipolar lines intersect can be determined to project a point in PGS to the image-coordinates accurately. More details regarding the non-horizontal settings are described in section 4.

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

Fig. 2. Camera settings.

3. ESTIMATION OF THE FUNDAMENTAL MATRICES BETWEEN THE CAMERAS

Each fundamental matrix between a basis camera and a moving camera must be computed to project 3D points in PGS onto the moving camera images and estimate the 3D positions of moving cameras in PGS.

3.1. Initial correspondences

At the first frame, each fundamental matrix between a moving camera and a basis camera is estimated by the 2D-2D correspondences of the feature points. The feature points are extracted by the Harris corner detector. We manually obtain 2D-2D correspondences of the feature points between a moving camera image and a basis camera image. The fundamental matrices are computed from those correspondences by using normalized eight-point algorithm [5].

From the second frame on, for each moving camera image, we track the feature points mapped manually at the first frame and estimate each fundamental matrix between a moving camera and a basis camera at each frame. The feature points are tracked by cross-correlation method. After candidates of tracked points are determined, we employ the the RANSAC (RAndom SAmple Consensus) [3] algorithm to remove mistracked points.

3.2. Addition of correspondences

At each frame, the number of the corresponding points between a moving camera and a basis camera decreases more and more due to occlusions or disappearance from the FOV. This results in a negative effect on the accuracy in the estimation of the fundamental matrices.

To solve this problem, at each frame, it is necessary to make new correspondences between feature points in a moving camera and feature points in a basis camera.

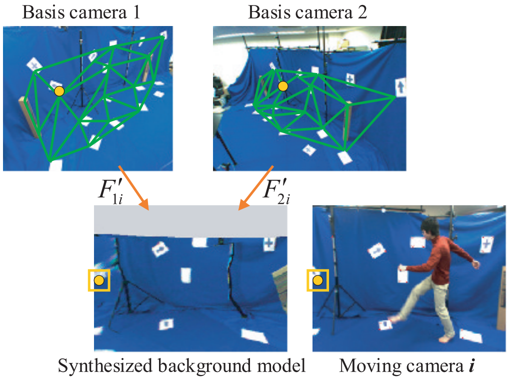


Fig. 3. Making new correspondences based on background model.

Our method makes such correspondences based on the background model. In pre-processing, backgrounds in the two basis camera images are divided manually into mutual triangle patches as shown in the top two images in Fig.3. At each frame, the background model can be rendered onto a moving camera image by view-interpolation of the corresponding triangle patches between the two basis cameras.

After executing the RANSAC algorithm, the initial estimations of the fundamental matrices F'_{1i}, F'_{2i} are computed between the two basis cameras and a moving camera i. The background model for the moving camera i can be synthesized by view-interpolation, which uses F'_{1i} and F'_{2i} to project the corresponding triangle patches onto the moving camera i. All the feature points which have correspondences between the two basis cameras can be projected onto the moving camera i by the initial estimations of the fundamental matrices. As shown in Fig.3, around the projected position in the moving camera i , the local pattern of the feature point in the background model is serched by block matching (yellow blocks), so that the feature point can be corresponded in the moving camera i.

4. 3D RECONSTRUCTION

The 3D shape model is reconstructed from multiple-view images by using a volume intersection method [7].

A certain number of voxels in a PGS is projected onto each silhouette image to check whether the projections are within the silhouette or not. A voxel $A\,(p, q, r)$ in a PGS is projected onto the image-coordinates, $\boldsymbol{a}_1\,(p, q)$ and $\boldsymbol{a}_2\,(r, s)$, in the basis cameras, 1 and 2, respectively in accordance with the definition of PGS. In the case of a moving camera i, the location $\boldsymbol{a}_i$ is at the intersecting point of the two epipolar lines that correspond to $\boldsymbol{a}_1\,(p, q)$ and $\boldsymbol{a}_2\,(r, s)$ in the two basis cameras.

The 3D shape model in PGS is reconstructed as the voxel model that consists of voxels projected within the silhouette images. The voxel model are converted into the surface reconstructed model with triangle patches.

5. RENDERING

Free-viewpoint images are synthesized by using an image-based rendering method based on the reconstructed 3D shape model. In this paper, we emply a virtual viewpoint rendering method based on view interpolation of the textures between two input images [11].

The corresponding textures between the two input view-images are determined by projecting triangle patches on the 3D model surface onto the two input view-images. If some triangle patches are occluded for either of the two input views, the correspondences of the textures are incorrect. The Z-Buffer method is employed to check the occlusions for each triangle patch.

The Z-Buffer is allocated at each input view. All the triangle patches are projected onto each input view to construct the Z-Buffers. Each value of distance between the 3D-position of the input view and the 3D-position of the triangle patch is stored in corresponding pixels on the Z-Buffer at each input view. If some of patches are projected onto the same pixel on the Z-Buffer, the shortest distance is stored. Therefore, the Z-Buffer at each input view equals the range image. After Z-Buffers have been generated, a patch whose distance from a input view is different from the value stored in the Z-Buffer is judged to be occluded on the input view.

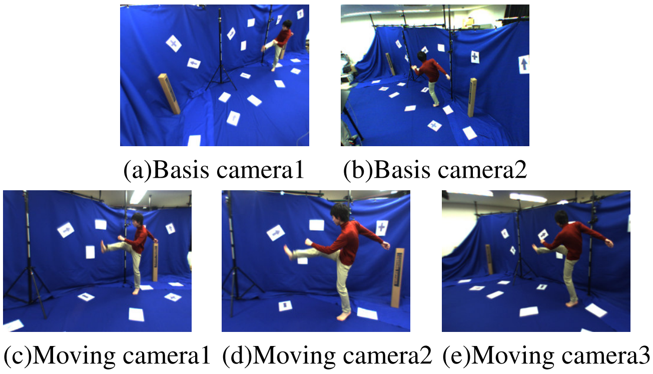
The virtual viewpoint images are synthesized by using the morphing of all the correctly corresponding textures.

6. EXPERIMENTAL RESULTS

In this section, the performance of the proposed system is evaluated by synthesizing the free-view images from the captured images. In our experiment, we mount two fixed cameras in a laboratory room as the two basis cameras. We also use three moving cameras which are controlled by human hands, so that the object is captured constantly within their FOV. The fixed cameras and the moving cameras have the same specifications, and are synchronized for the synthesis of the free-view with the moving object.

The camera setting is non-horizontal Fig.2(b) as described in Sec.2. The captured images of the first frame, which were captured at a 640×480 resolution are shown in Fig.4.

The number of correspondences between the basis camera1 and the moving camera, which are labeled A,B,C is shown in Fig.5 to compare the refined method of tracking correspondences described in Sec.3.2 with the pure tracking method. Camera 1_A to 1_C indicate the refined method. 1_A' to 1_C' represent the pure tracking method. The decline in performance is more moderate in the former than



(a)Basis camera1 (b)Basis camera2

(c)Moving camera1 (d)Moving camera2 (e)Moving camera3

Fig. 4. Input views captured at a frame.

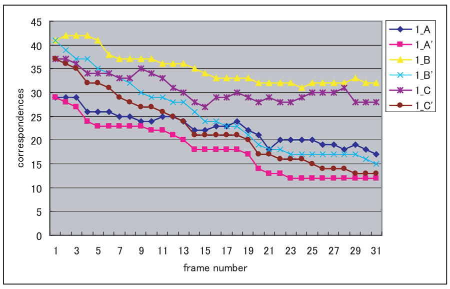


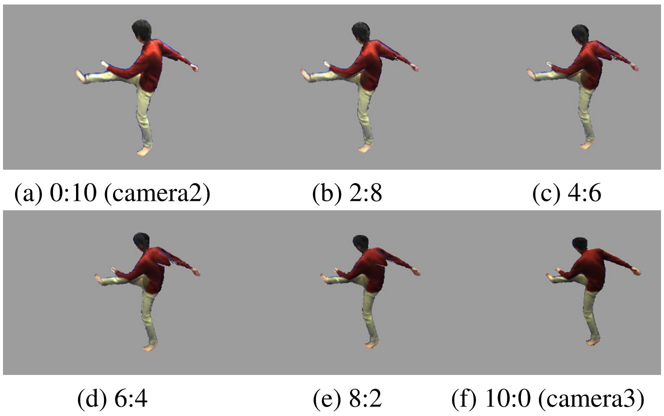
Fig. 5. Performance of the refinement method of correspondences.

in the latter. This means that the refinement of correspondences is effective for maintaining the number of correspondences overtime.

Fig.6 shows the images synthesized at the virtual-views which were between moving camera2 and moving camera3. Free-viewpoint images can successfully be synthesized by our proposed system with uncalibrated moving cameras. Some corruption and lack of the textures can be observed in the synthesized images. These 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 the accuracy of shape reconstruction in future work.

7. CONCLUSION

We proposed a novel method to synthesize free-viewpoint images for a moving object, which is captured by uncalibrated multiple cameras. We use multiple moving cameras that are able to capture a moving object in a constant area. Two fixed cameras are employed for determining a Projective Grid Space which defines a projective 3D coordinate in the object space for 3D reconstruction of the object from multiple moving cameras without calibration.



(a) 0:10 (camera2) (b) 2:8 (c) 4:6

(d) 6:4 (e) 8:2 (f) 10:0 (camera3)

Fig. 6. The virtual views between the moving camera 2 and the moving camera 3.

8. REFERENCES

[1] 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.

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

[3] 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.

[4] B. Goldlucke, M. Magnor, "Real-Time Microfacet Billboarding for Free-Viewpoint Video Rendering," Proc. IEEE International Conference on Image Processing (ICIP'03), Barcelona, Spain, vol. 3, pp.713-716, September 2003.

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

[6] 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.

[7] A. Laurentini, "The visual hull concept for silhouette based image understanding," in *IEEE Trans. Pattern Analysis and Machine Intelligence,* 16, 2, pp.150–162,1994.

[8] S. Moezzi, L.C.Tai, P.Gerard, "Virtual View Generation for 3D Digital Video," IEEE Multimedia, 4, 1, pp.18–26, 1997.

[9] M. Pollefeys, R. Koch, and L. V. Gool, "Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters," International Journal of Computer Vision, 32(1), pp.7-25, 1999.

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

[11] S. Yaguchi and H. Saito, "Arbitrary viewpoint video synthesis from multiple uncalibrated cameras," in *IEEE Trans. on Systems, Man and Cybernetics, B*,34, 1, pp.430–439, 2004.