Arbitrary-view image generation from multiple silhouette images in projective grid space

Yaguchi, Satoshi, Saito, Hideo

Satoshi Yaguchi, Hideo Saito, "Arbitrary-view image generation from multiple silhouette images in projective grid space," Proc. SPIE 4309, Videometrics and Optical Methods for 3D Shape Measurement, (22 December 2000); doi: 10.1117/12.410888

Event: Photonics West 2001 - Electronic Imaging, 2001, San Jose, CA, United States

Arbitrary view image generation from multiple silhouette images in projective grid space

Satoshi Yaguchi and Hideo Saito

Dept. of Information and Computer Science, Keio Univ. Yokohama, Japan

ABSTRACT

In this paper, we propose a method for arbitrary view generation from multiple view images taken with uncalibrated camera system. In Projective Grid Space (PGS), that is the three dimensional space which is defined by epipolar geometry between the two basis cameras in the multiple cameras, we reconstruct three dimensional shape model from the silhouette images of the multiple cameras. For the shape reconstruction in the PGS, the multiple cameras do not have to be fully calibrated, but the fundamental matrices of every camera to the two basis cameras must be collected. By using the three dimensional model reconstructed in the PGS, we can obtain the point correspondence between arbitrary pair of images which can generate the image of arbitrary view between the pair of images.

Keywords: Arbitrary View, F-Matrix, Projective Geometry, Projective Grid Space, 3D Reconstruction, Shape from Silhouette

1. INTRODUCTION

The technology of 3D shape modeling and rendering from multiple view images has recently been intensely researched, mainly because of advances of computation power and capacity of data handling. Research in 3D shape reconstruction from multiple view images conventionally been applied in robot vision and machine vision systems, in which the reconstructed 3D shape is used for recognizing the real scene structure and object shape. However, in field of computer vision and computer graphics, these researches applied to arbitrary view generation from multiple view images have been conventionally conducted too.

These works can be broken down into two basic groups: generating new view images from 3D structure models that are reconstructed from the information taken in the computer with some techniques (so called "model based rendering"), and generating new view images directly from multiple input images (so called "image based rendering"). Model based rendering needs 3D structure model as input data. 3D reconstruction using volumetric integration of range images such as Hilton et al.,[7] Curless and Levoy,[4] Masuda and Yokoya,[12] and Wheeler et al.[20] led several approaches to recovering global 3D geometry. Most of this work relies on direct range-scanning hardware, which is relatively slow and costly for dynamic multiple sensor modeling system.

Image-based rendering has also seen significant development. Katayama et. al.[8]demonstrated that images from dense set of viewing positions on a plane can be directly used to generate images for arbitrary viewing positions. Levoy and Hanraha[11] and Gortler et al.[6] extend this concept to construct a four-dimensional field representing all light rays passing through a 3D surface. New view generation is posed as computing the correct 2D cross section of the field of the light rays. A major problem of with these approaches is that thousands of real images may be required to generate new views realistically, therefore making the extension to dynamic scene modeling impractical.

For the purpose of virtualizing the dynamic events, we research on generating arbitrary view images from multiple images, which are taken from several 10 CCD cameras. Generally, 3D reconstruction from multiple camera system requires strong calibration for each camera.[5,9,19] For calibrating cameras, 3D position in Euclidian space of several points and 2D position on each view images of those points must be measured precisely. For this reason, when there are many cameras, much effort is needed to calibrate every camera. In the case of large space, it is difficult to set many calibrating points precisely throughout the large area.

In this paper, we propose the method to reconstruct the 3D shape models from uncalibrated multiple cameras in the "Projective Grid Space ", which is based on the epipolar geometry between each camera.[14] We reconstruct 3D shape model in the projective grid space applying to the shape from silhouette (SS) method, and generate arbitrary

Send correspondence to: Email :yagu@ozawa.ics.keio.ac.jp

Videometrics and Optical Methods for 3D Shape Measurement, Sabry F. El-Hakim, Armin Gruen, Editors,
Proceedings of SPIE Vol. 4309 (2001) © 2001 SPIE · 0277-786X/01/$15.00

view images based on the 3D shape model. Furthermore, we propose a method for merging the floor plane of the background, which is removed when the SS method is performed in the PGS. We demonstrate the proposed framework by showing several virtual image sequences, which is generated by applying to the real images from uncalibrated nine cameras.

2. 3D SHAPE MODEL RECONSTRUCTION

Generally, reconstructing 3D shape model from multiple view images requires correspondence between any 3D scene points and their projected points onto each image plane. Therefore, for corresponding between any 3D scene point and each image plane, current techniques need to estimate the projection matrices for every camera.[9,19] It is equivalent to camera calibration.

In our method, 3D point is related to 2D image point without estimating the projection matrices by "Projective Grid Space (PGS)[14]", which can be determined by only a fundamental matrix[21] representing the epipoler geometry between two basis cameras. Applying PGS enables 3D reconstruction from multiple images without estimating the projection matrices of each camera.

Under the framework of PGS, we reconstruct 3D shape model by shape from silhouette (SS) method. The point on the image plane to which the 3D position in the PGS projects is decided by only the fundamental matrix between each camera. The object region can be determined by checking if the projected PGS point is included in each silhouette or not, and then 3D model is reconstructed from the object region. In the following subsections, we explain the details of the procedure.

2.1. Projective Grid Space

Projective Grid Space is defined as follows. Two view images are selected as the basis of the projective grid space. Each pixel point (p, q) in the first image defines one grid line in the space. On the grid line, grid node points are defined by horizontal position r in the second image. Since fundamental matrix $\boldsymbol{F}_{21}$ limits the position in the second basis view on the epipolar line $\boldsymbol{l}$, r is sufficient for defining the grid point. In this way, the projective grid space can be defined by two basis view images, of which node points are represented by (p, q, r).

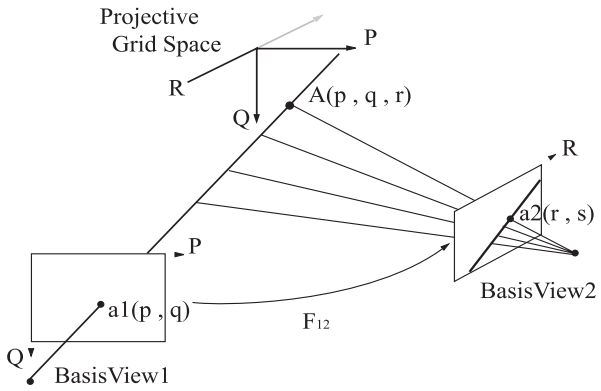

Figure 1. Definition of projective grid. The point $A(p, q, r)$ on the projective grid space is projected to $a_1(p, q)$ and $a_2(r, s)$ on the first and second basisview.

2.2. 3D Shape Model Reconstruction

In our proposed method, 3D shape model is reconstructed by Shape from Silhouette (SS) method. In the conventional SS method, each voxel in a certain Euclidean space must be projected onto every silhouette image with projection matrices, which are calculated by strong Euclidean calibration of every camera,[2,13] for checking if the voxel is included in the object region.

In our method, the SS method is implemented in PGS so that Euclidean calibration is not required. Every point in the PGS is projected onto each silhouette image with fundamental matrices. At least, eight correspondence points

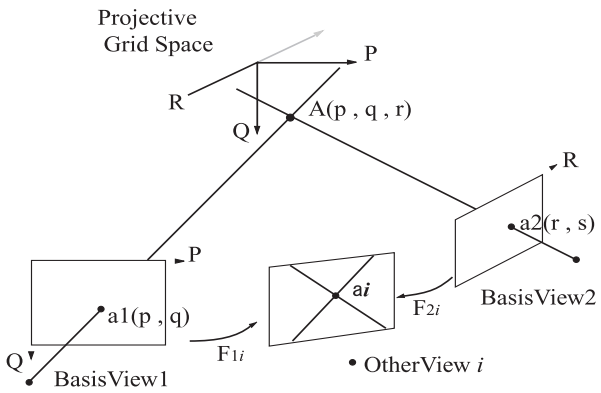

Figure 2. Projection of point in the space onto an image. The point $A(p,q,r)$ on the projective grid space is projected to the cross point of two epipolar lines in the image of view i.

between two images decide the fundamental matrix, therefore we are able to obtain it more easily than the camera parameters.

The fundamental matrix represents epipoler geometry between two images. As described in the previous section, the PGS is defined by two basis views, and the point in the PGS is represented as $A(p,q,r)$. The point $A(p,q,r)$ is projected onto $a_1(p,q)$ and $a_2(s,r)$ in the first basis image and the second basis image, respectively. The point a_1 is projected as the epipoler line l on the second basis view. The point a_2 on the projected line (figure 1), is expressed as

$$\boldsymbol{l} = \boldsymbol{F}_{21} \begin{bmatrix} p \\ q \\ 1 \end{bmatrix} \tag{1}$$

where $\boldsymbol{F}_{21}$ represents the fundamental matrix between the first and second images.

The projected point in ith arbitrary real image is determined two fundamental matrices, $\boldsymbol{F}_{i1}$, $\boldsymbol{F}_{i2}$ between two basis images and ith image. Since $A(p,q,r)$ is projected onto $a_1(p,q)$ in the first basis image, the projected point in the ith image must be on the epipolar line $\boldsymbol{l}_1$ of $a_1(p,q)$, which is derived by the $\boldsymbol{F}_{i1}$ as

$$\boldsymbol{l}_1 = \boldsymbol{F}_{i1} \begin{bmatrix} p \\ q \\ 1 \end{bmatrix} \tag{2}$$

In the same way, the projected point in the ith image must be on the epipoler line $\boldsymbol{l}_2$ of $a_2(r,s)$ in the basis image, which is derived by the $\boldsymbol{F}_{i2}$ as

$$\boldsymbol{l}_2 = \boldsymbol{F}_{i2} \begin{bmatrix} r \\ s \\ 1 \end{bmatrix} \tag{3}$$

The intersection point between the epipolar line $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$ is the projected point $A(p,q,r)$ onto the ith image.(figure2) In this way, every projective grid point is projected onto every image, where the relationship can be represented by only the fundamental matrices between the image and two basis images.

The process of SS method for reconstructing 3D shape model is as follows. A certain region is determined in the projective grid space, and every voxel in that region projects onto each silhouette image with proposed scheme as shown figure 3. The voxel that is projected onto the object silhouette for all images is decided as existent voxel, while others are nonexistent. Thus the volume of the object can be determined in the voxel represented in PGS.

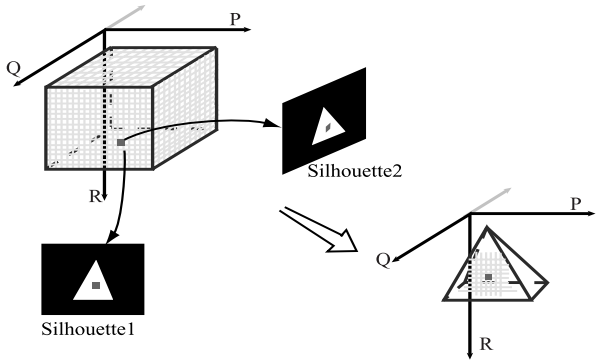

Figure 3. The shape from silhouette process. The voxel projected not onto the object region is removed.

In this process, using the advantage of our method can reduce the amount of calculation. Since the vertical and horizontal coordinate of the first basis view image are equivalent to P and Q coordinate of the PGS, any calculation involving f-matrix is not required to project each voxel onto the first basis view image. In the second basis view image, the projected point is decided by calculating only one f-matrix multiply for determining epipoler line, this implies that projection calculation onto second basis view becomes half compared with projecting the other images. For this reason, checking the voxel and removing the unnecessary voxel one image after another, in order of basis view 1, basis view 2, and the other view image, the amount of calculation can be reduced drastically.

After this existent voxel determination, implicit surface of the voxel representation of the object is extracted by "Marting Cubes". Finally, the object model is reconstructed as the surface representation in the PGS.

3. ARBITRARY VIEW GENERATION

There are two ways to generate the arbitrary view image from 3D shape model, that is texture mapping on the 3D Shape model,[19] and the morphing from correspondence of some images.[1,3,16,15] The reconstructed projective shape model provides dense correspondence maps between arbitrary pairs of input images, so that they can be used for the latter procedure. In the former procedure, the texture of the images are projected onto the 3D shape model, then re-projected onto the image again. In this procedure, however, the generated images are suffered from rendering artifact caused by the inaccuracy of 3D shape.

Therefore we apply the latter procedure to generate arbitrary view images.

3.1. Arbitrary View Generation

Arbitrary view image is synthesized as intermediate images of selected two input view images. For generating arbitrary view image, we take the two-step algorism, where depth image is rendered at intermidiate point between two input views with z-buffer algorithm, and these rendered images are blended with the correspondence maps that is computed by using the reconstructed 3D shape model.

To apply the z-buffer algorithm to render the 3D model at the input views, the surface of 3D model that reflect on the input views is decided. In the z-buffer of each pixel, the distance between the viewpoint and the closest surface point is stored, and the stored distances are used for deciding the occluded region in blending step.

Next, intermediate images of two input views are synthesized by interpolating two rendered images. The interpolation is based on the related concepts of "view interpolation" and "view morphing", in which the position and color of every pixel are interpolated from the corresponding points in two images. The following equations are applied to the interpolation:

$$\boldsymbol{P}_i = w_1 \boldsymbol{P} + w_2 \boldsymbol{P}' \tag{4}$$

$$I_i(\boldsymbol{P}_i) = w_1 I(\boldsymbol{P}) + w_2 I'(\boldsymbol{P}') \tag{5}$$

$\boldsymbol{P}$ and $\boldsymbol{P}'$ are the position of the corresponding points in the two views, $I(\boldsymbol{P})$ and $I'(\boldsymbol{P}')$ are the colors of the corresponding points, and $\boldsymbol{P}_i$ and $I(\boldsymbol{P}_i)$ are the interpolated position and color. The interpolation weighting factors are represented by w_1 and w_2 $(w_1 + w_2 = 1)$.

This interpolating method requires consistent correspondence between two images. However, there is the case that some points in one image cannot be seen in another image. In this case, the position on the interpolating image of such point is derived by equation (5) , and the color is decided on the value of visible point. The distance image stored by the z-buffer algorithm is used for deciding the point visibility.

3.2. Camera Position in Projective Grid Space

In the previous procedure, z-buffer algorithm requires camera position in projective grid space. The fundamental matrix provides sufficient information about the relative camera position of the camera center, which can be derived from that matrix.

We can obtain the position of camera in the coordinate of (p, q, r) as shown in figure 4. In this coordinate, camera position of the first basis camera $C1$ is $(p_c, q_c, e12_r)$,where $(p_c, q_c)$is camera center in the first basis view in second basis view, and $e12_r$ is r component of the epipole of first basis view in second basis view, $e12$. In the same way, camera position of the second basis camera $C2$ is $(e21_p, e21_q, e12_r)$, where $(e21_p, e21_q)$is the epipole of the second basis view in the first basis view $e21$.

For ith camera, we can obtain epipoles $ei1$ and $ei2$ in the first and second basis views, respectively. Therefore, the position of ith camera is $(ei1_p, ei1_q, ei2_r)$, which is derived from fundamental matrices, every camera position in the (p, q, r) coordinate can be obtained from only fundamental matrices.

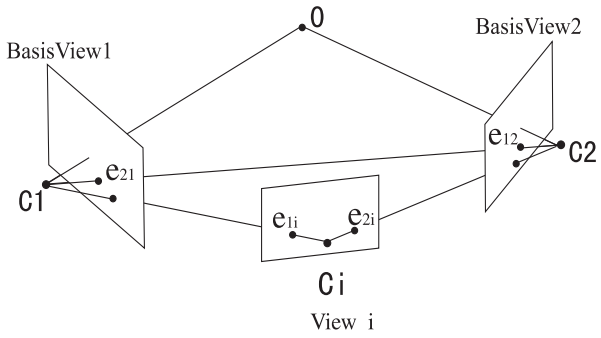

Figure 4. Camera position in the projective grid space.

3.3. Synthesizing the Floor Plane

We also propose the method to synthesize the background such as a floor plane in this projective method. Since floor plane is removed at the step of making the silhouette image, SS method only provides 3D shape of the object without background. For generating more realistic images, we synthesize a floor plane image that is combined with the foreground.

Since the coordinate axes of the projective grid space are defined by two basis cameras, a line and a plane in the PGS can not be represented by the equation of Euclidian grid space. Context and geometric relation are conserved, when we project the image object on to PGS and re-project it onto image. Taking into account this feature, we synthesize the floor plane from more than three points that are picked out from the real background image. In the following, we explain the details of the procedure.

Several correspondence points on the floor region between the two basis view images are picked out as shown figure 5. From the definition of the PGS, the coordinate of a correspondence point A is $A1(x1, y1)$ on the first basis view, and $A2(x2, y2)$ on the second basis view, then the coordinate of A in the PGS becomes $A(x1, y1, x2)$. Since the coordinate of a point in the PGS is fixed, the point can be projected onto every input view images with the fundamental matrices, in the same way stated before.

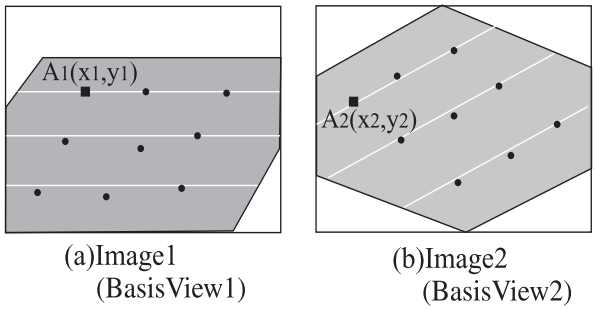

Figure 5. Pointing the correspondence points between the two basis view images.

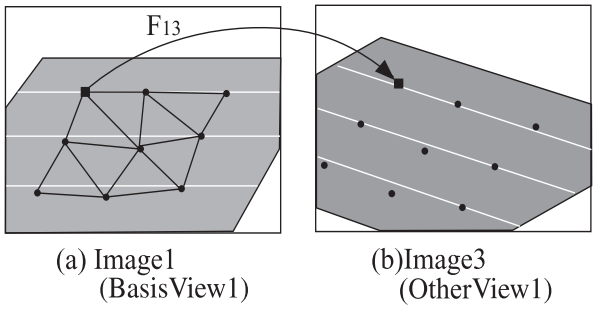

Figure 6. Project the vertex of Delauney triangles onto the other view image by fundamental matrices.

The points on the floor are triangulated for representing the floor plane by using the Delauney triangulation in the first basis view image. The vertices of the triangle mesh are project onto two interpolating background image with fundamental matrices as shown figure 6.

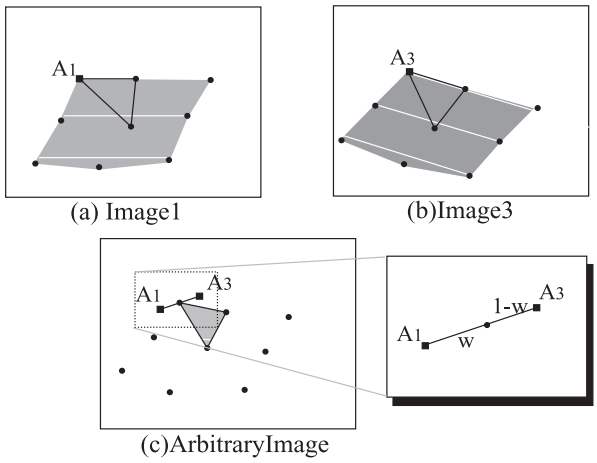

Figure 7. Generating the floor plane on the arbitrary view image.

The positions of vertices of a triangle mesh on the arbitrary view image are decided by the position on the two background images as shown figure 7, which is in the same way expressed in equation4. Affine matrix between arbitrary view image and each background images is calculated for each triangle mesh. Inside the region of each

triangle of the arbitrary view image is interpolated with the affine transformation. The color of each pixel is also determined by interpolating that of the corresponding points on each background images in the same way of equation5. Finally, the floor plane of arbitrary view image is constructed by integrating those triangle meshes.

For making the arbitrary view images, it is requires to synthesize the object region and the floor region separately. Both of them cannot be rendered simultaneously, because each rendering method is different. However, it is clear that the floor plane is behind the object, therefore the floor plane is rendered first, and the object is rendered over there.

4. EXPERIMENTAL RESULT

We applied the proposed method to real image sequence of which we took pictures with nine cameras that settled on the wall and ceiling of the B-con Plaza hall in Beppu, Ooita, Japan. One camera was settled on the ceiling and eight cameras were on the wall. We applied our method to a number of frames out of the image sequence taken those cameras, 3D models were reconstructed each frame, and arbitrary view images around the object were synthesized. The size of an image is 640 times 480 pixels, RGB format. The example of input real images are shown figure 8. The silhouette images were generated background subtraction, and 3D shape model was reconstructed in the PGS by the proposal method. The 3D shape model in the representation of orthographic grid space, which is seen from the arbitrary view, are shown figure 9, and figure 10.

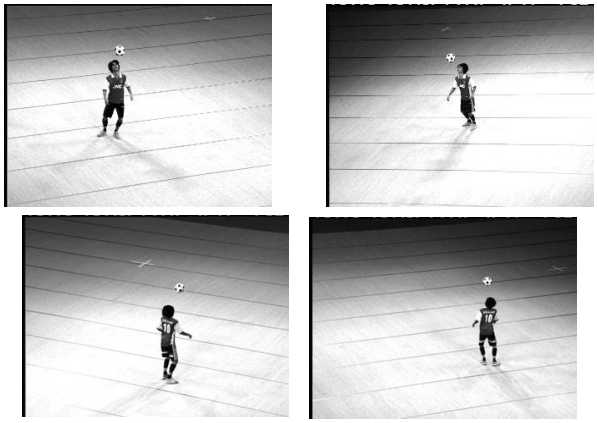

Figure 8. The example of input real images.

The interpolated arbitrary view images of each real view images with changing weighting factor are shown in figure 11. Four images of the top and the bottom are real view images, and number of the images is shown the camera numbers. Other images are interpolated view images of each side images changing the weighting factor.

The sequence of the interpolated view images under the interpolate weight 0.7, are shown figure 12. The real image sequence of the camera2 and camera3 are shown each side of figure 12. The centers are interpolated images of each side images. Although some noises are appeared at the boundary of the objects, most of the texture on the ball and the player in the virtual view are correct. The noise on the boundary, it seems that which is caused by the difference between the silhouette of the object and real object region.

5. CONCLUSIONS

We proposed a method for reconstructing the 3D shape model in the projective grid space and generating the arbitrary view image from the multiple image sequences taken with uncalibrated cameras. The projective grid space can be defined with two basis views, whose relationship is represented by a fundamental matrix. The grid points in the space are related to an arbitrary image by fundamental matrices between the image and the two basis views. In this framework for virtual view generation, we do not take into account the geometrical correctness of the interpolated virtual view because we currently only use simple correspondences interpolation between images. However, as Seitz et .al.[16] pointed out in view morphing such simple correspondence interpolation cannot correctly interpolate the geometry of the views. For more realistic new view generation, such correctness of the geometry has to be considered also.

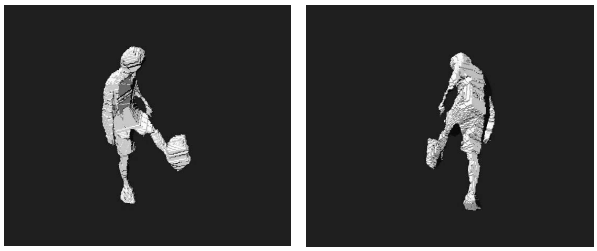

Figure 9. Frame 1. Reconstructed projective shape in the representation of orthographic grid space.

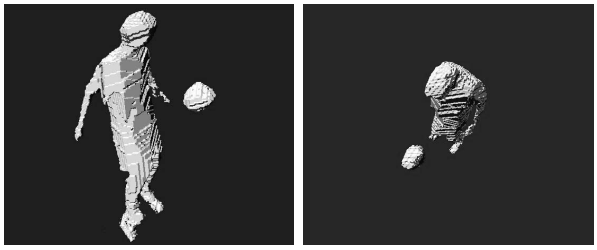

Figure 10. Frame 10. Reconstructed projective shape in the representation of orthographic grid space.

6. ACKNOWLEDGEMENT

The authors thank the members of the consortium for virtualized reality experimental project in Oita, Japan, including Prof. T.Kanade(CMU) and Prof. Y.Ohta (University of Tsukuba), for their effort to capturing the image sequences.

REFERENCES

1. T.Beier, S.Neely : "Feature-Based Image Metamorphosis",Proc. of SIGGRAPH '92, pp.35-42, 1992.
2. C.H.Chein, J.K.Aggarawal : "Identification of 3D Objects from Multiple Silhouettes using Quadtrees / Octrees". Conputer Vision, Graphics and Image Processing.36(1986), 100-113
3. S.Chen, L.Williams : "View Interpolation for Image Synthesis",Proc. of SIGGRAPH '93,pp.279-288, 1993.
4. B. Curless and M. Levoy, "A Volumetric Method for Building Complex Models from Range Images", *Proc.of SIGGRAPH '96*, 1996.
5. D.M.Gavrila, L.S.Davis : "3-D Models Based Tracking of Humans in Action : Multi-View Approach", Proc. Computer Vision and Pattern Recognition 96, pp.73-80, 1996.
6. S.J. Gortler, R. Grzeszczuk, R. Szeliski, and M.F. Cohen, "The Lumigraph", *Proc. of SIGGRAPH'96*, 1996.
7. A. Hilton, J. Stoddart, J. Illingworth, and T. Windeatt, "Reliable Surface Reconstruction From Multiple Range Images", *Proc. of ECCV'96* pp.117-126, 1996.
8. A. Katayama, K. Tanaka, T. Oshino, and H. Tamura, "A view point dependent stereoscopic display using interpolation of multi-viewpoint images", *SPIE Proc.* Vol.2409, Stereoscopic Displays and Virtual Reality Systems II, pp.11-20, 1995.
9. T.Kanade, P.W.Rander, S.Vedula, H.Saito : "Virtulized Reality:Digitizing a 3D Time-Valying Event As Is and in Real Time", International Symposium on Mixed Reality(ISMR99), pp41-57, Yokohama, Japan, March 1999.
10. S.Laveau and O.Faugers : "3-D Scene Representation as a Collection of Images", Proc. Int'l. Conf. Pattern Recognition,1994.
11. M. Levoy and P. Hanrahan, "Light Field Rendering", *Proc. of SIGGRAPH'96*, 1996.
12. T.Masuda, N.Yokoya, "A Robust MEthod for Registration and Segmentation of Multiple Range Images", Computer Vision and Image Understanding, Vol.61, No.3, pp.295-307, 1995.
13. M.Potmesil : "Generating Octree Models of 3D Objects from Their Silhouettes in a Sequense of Images". Conputer Vision, Graphics and Image Processing.40(1987), 277-283

14. H.Saito,T.Kanade : "Shape Reconstruction in Projective Grid Space from Large Number of Images", IEEE Proc. Computer Vision and Pattern Recognition, Vol.2, pp.49-54, 1999.
15. H.Saito, S.Baba, M.Kimura, S.Vedula, T.Kanade : "Apperance-Baced Virtual View Generation of Temporally-Varying Events from Multi-Camera Images in 3D Room",Computer Science Technical Report, CMU-CS-99-127, April 1999.
16. S.M.Seitz, and C.R.Dyer : "View Morphing",proc. of SIGGRAPH '96, pp.21-30, 1996.
17. S.M.Seitz, and C.R.Dyer : "Photorealistic Scene Reconstruction by Voxel Coloring", Proc. Computer Vision and Pattern Recognition (CVPR), pp.1067-1073, 1997.
18. R.Tsai : "A Versatile Camera Caribration Technique for High- Accuracy 3D Machine Vision Metrology Using Off-the-Shelf Tv Cameras and Lenses", IEEE Journal of Robotics and Auto mation RA-3,4, pp323-344, 1987.
19. S.Vedula, P.W.Rander, H.Saito, T.Kanade : "Modeling, Combining, and Rendering Dynamic Real-World Events From Image Sequences", Proc. 4th Conf. Virtual Systems and Multimedia, Vol.1, pp.326-322, 1998.
20. M.D. Wheeler, Y. Sato, and K. Ikeuchi, "Consensus surfaces for modeling 3D objects from multiple range images", *DARPA Image Understanding Workshop*, 1997.
21. Z.Zhang : "Determining the Epipolar Geometry and its Uncertainly: A Review". INRIA reserch report, 2927, 1996.

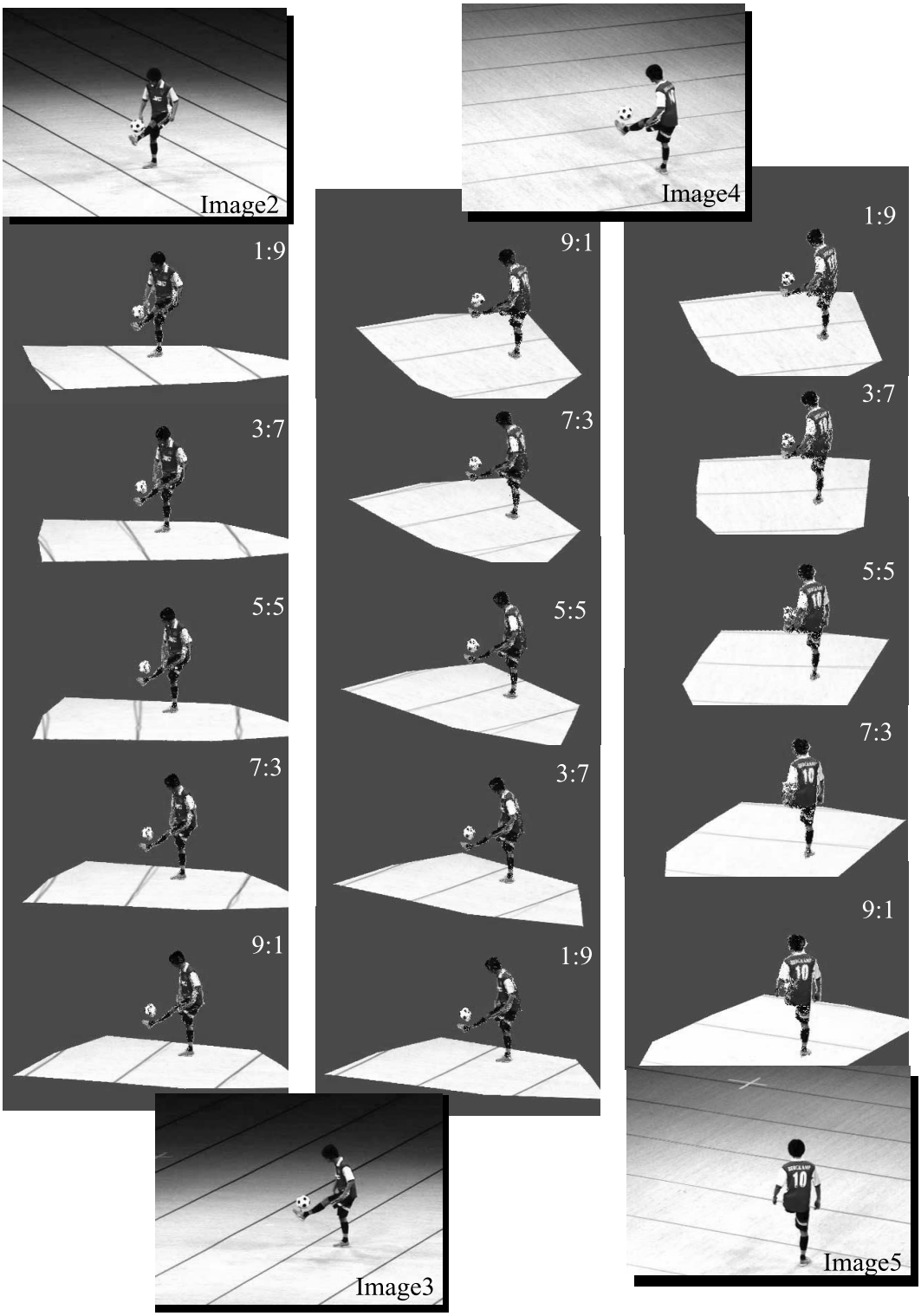

Figure 11. Synthesized intermediate view images between each two images.

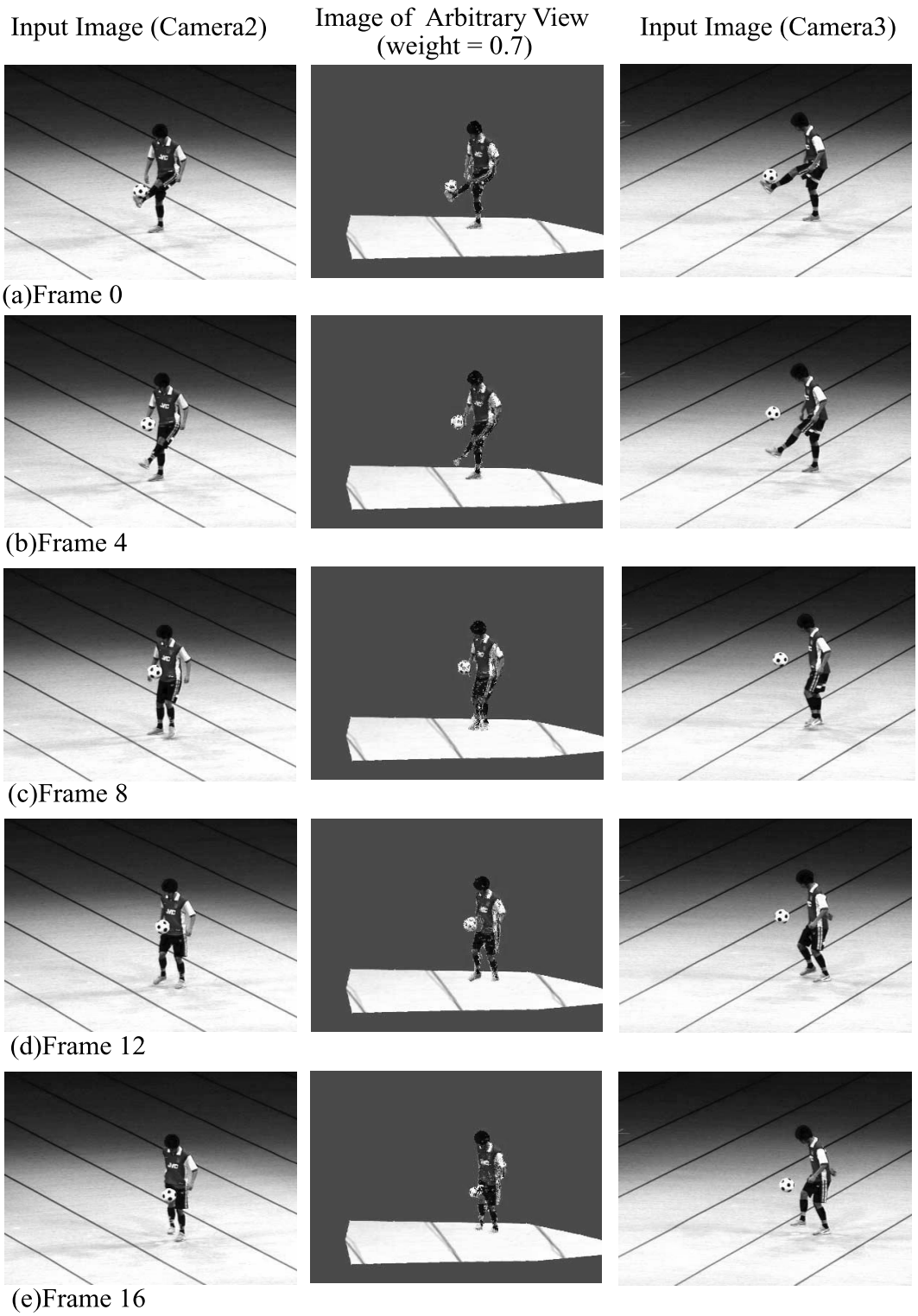

Figure 12. Synthesized intermediate view images between each side images.