Interpolation of three views based on epipolar geometry

Kimura, Makoto, Saito, Hideo

Makoto Kimura, Hideo Saito, "Interpolation of three views based on epipolar geometry," Proc. SPIE 4310, Visual Communications and Image Processing 2001, (29 December 2000); doi: 10.1117/12.411799

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

Interpolation of three views based on epipolar geometry

Makoto Kimura and Hideo Saito

Department of Information and Computer Science, Keio University
3–14–1 Hiyoshi Kouhoku-ku Yokohama 223-8522, Japan

ABSTRACT

In this paper, we propose a method for generating arbitrary view image by interpolating images between three cameras using epipolar geometry. Projective geometry has recently been used in the field of computer vision, because projective geometry can be easily determined compararing with Euclidean geometry. In the proposed method, three input camera images are rectified so that the vertical and horizontal directions can be completely aligned to the epipolar planes between the cameras. This rectification provides Projective Voxel Space (PVS), in which the three axes are aligned with the direction of camera's projection. Such alignment simplifies the procedure for projection and back projection between the 3D space and the image planes. First, we apply shape-from-silhouette with taking advantage of PVS. The consistency of color value between the images is evaluated for final determination of the object surface voxel. Therefore, consistent matching in three images is estimated and images can be interpolated from the matching information. Synthesized images are based on 3D shape in PVS, so the occlusion of the object is reproduced in the generated images, however it requires only weak calibration.

Keywords: projective geometry, fundamental matrix, shape reconstruction, interpolation

1. INTRODUCTION

For interpolating images from new point of view based on real multiple images, one approach is to reconstruct the 3D shape of object in the scene, so that images can be generated from 3D shape and texture data. In general, 3D reconstruction requires camera calibration that is performed by checking the correspondence between 3D geometry in world coordinates and 2D geometry in image coordinates.[10,5] Although reconstructed 3D shape itself is not required explicitly,[8] the reconstruction of 3D geometry has some advantages. With handling multiple images in a common 3D coordinate, the geometry of occluded point in an image can be determined by the other images.

Recently, projective geometry has often been used in the field of computer vision,[6–8,1] because projective geometry can be determined easier than Euclidean geometry. While determination of Euclidean geometry requires a map of correspondences between the position in image, which is represented in 2D coordinate, and the position in Euclidean space, which is represented in 3D coordinate, determination of projective geometry requires only several correspondences of position in image, that is represented in 2D coordinate. Then we call the traditional camera calibration "strong calibration" and calibration for projective geometry "weak calibration". Projective geometry makes it possible to determine epipolar line for any point in image, however it has no notion of a world coordinate.[11] Thus, projective geometry is easy to calibrate, but it doesn't determine world coordinate such as Euclidean geometry which determines 3D space for all cameras and objects in the scene.

In this paper, we propose an approach to reconstruct 3D shape in a voxel space from three images and weak calibration. We name this voxel space "Projective Voxel Space" (PVS), which makes it possible to handle all data in a common three-dimensional coordinate frame. The coordinate axes in PVS are not orthogonal in the real world, so the reconstructed 3D shape itself is not equal to 3D shape in Euclidean geometry. However the reconstructed shape describes enough 3D information for generating arbitrary new images by interpolating input images.

E-mail addresses :
Makoto Kimura: kimura@ozawa.ics.keio.ac.jp
Hideo Saito: saito@ozawa.ics.keio.ac.jp

Visual Communications and Image Processing 2001, Bernd Girod, Charles A. Bouman,
Eckehard G. Steinbach, Editors, Proceedings of SPIE Vol. 4310 (2001) © 2001 SPIE · 0277-786X/01/$15.00

2. PROJECTIVE VOXEL SPACE (PVS)

2.1. EPIPOLAR GEOMETRY

Epipolar geometry is one form of handling the projection function. As shown in the Figure 1, a point, which is visible in an image, must exist on a line, which is back-projected from the camera in Euclidean geometry. Therefore, a point in a scene must be equal to a cross-point of back-projected line from each camera. The epipolar plane is a plane going through the line connecting each focus point of the camera, and the epipolar line is a line in an image obtained by projecting the back-projection line of the other camera.

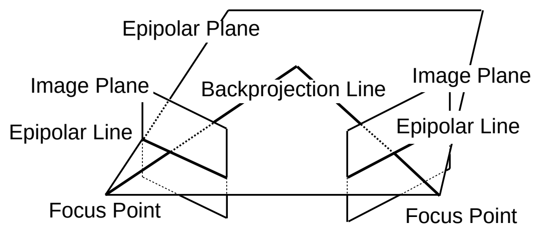


Figure 1. Epipolar lines and an epipolar plane in Euclidean geometry

,

2.2. FUNDAMENTAL MATRIX

First of all, we apply fundamental matrix, which is one of the description forms of projective geometry in two images. Fundamental matrix is a 3×3 matrix, and it theoretically requires only seven matching points in it's determination.[2] Using the fundamental matrix, epipolar lines can be solved from arbitrary points in the other image as shown in Figure 2. Conversely, a point in the solved epipolar line and the fundamental matrix can solve corresponding epipolar line in the original image. Thus, fundamental matrix can solve arbitrary epipolar lines in a pair of images.

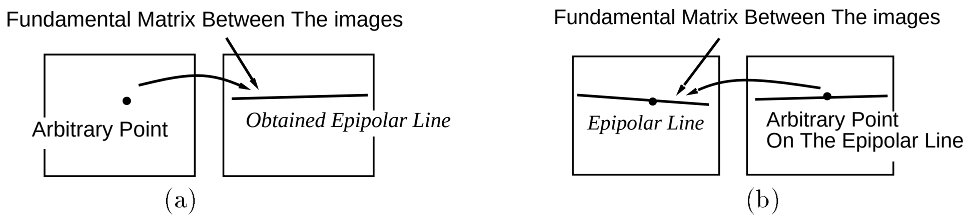


Figure 2. Obtainment of epipolar line : (a) An arbitrary point and fundamental matrix can obtain an epipolar line in the other image ; (b) Obtained epipolar line in the original image

2.3. RECTIFICATION OF EPIPOLAR LINES

As mentioned, the correct matching point theoretically exists on the corresponding epipolar line in each image. In this way, fundamental matrix can restrict the searching area of a matching point just on a line in each image. Having more than two images and fundamental matrices between all pairs of the images, we can apply this restriction more efficiently. Because the fundamental matrix and a point in the image can determine epipolar line in the other image, matching point in the other image must be equal to the crosspoint of the epipolar lines which are determined by matching points in two images. Figure 3 shows this framework of epipolar line. Thus, a correct matching point in two of the images can automatically determine the matching point in the other image.

Generally, epipolar lines between two images are distributed like fan shape. Then, rectification of epipolar lines is reasonable for processing of searching matching points. In the same way, Figure 4-(a) shows the distribution of epipolar lines between each pairs of three images, and rectification works as Figure 4-(b). As a matter of course, the ability of epipolar lines, which restricts the searching area of matching points, is still valid in the rectified images. As show in Figure 4-(b), a pair of matching points in two images can determine the matching point in the other image.

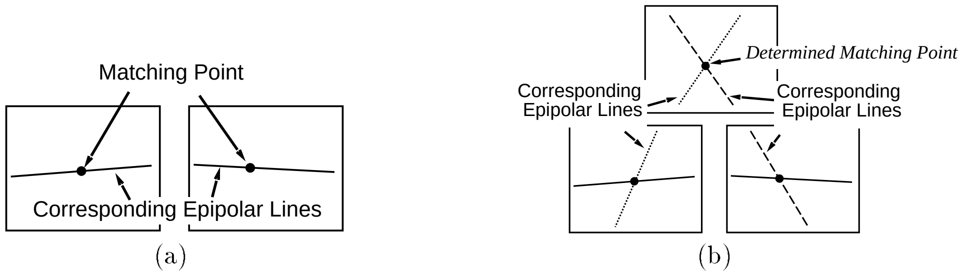


Figure 3. Determination of matching point in the other image using epipolar lines : (a) matching points in two images ; (b) determined matching point using epipolar lines

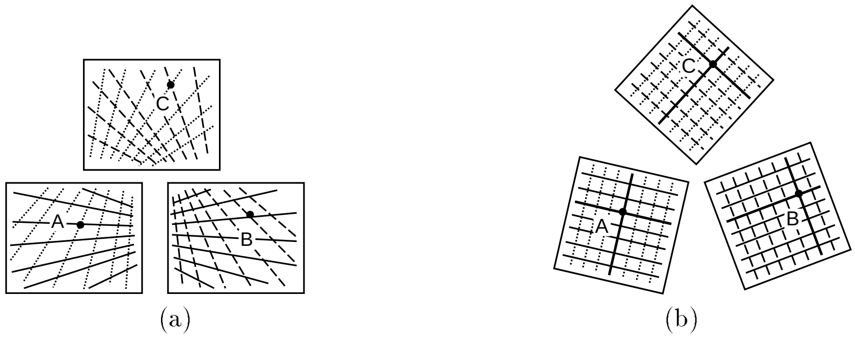


Figure 4. Rectification of epipolar lines between three images and an example of matching points: (a) epipolar lines in original images ; (b) rectified epipolar lines and examples of matching points

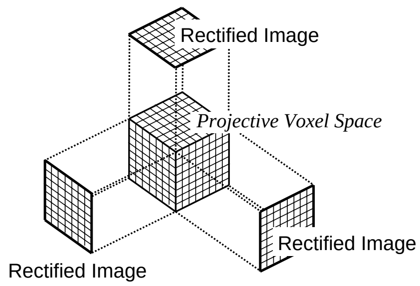
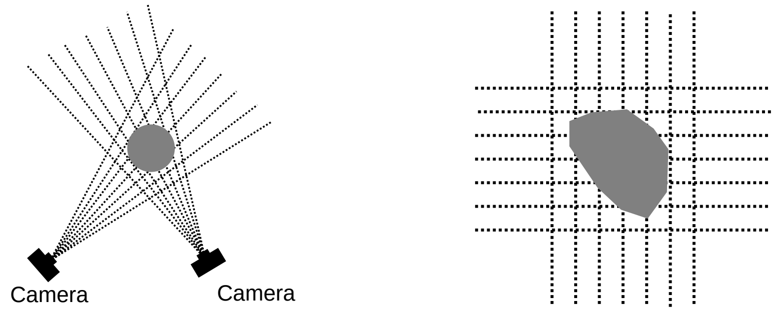


Figure 5. Projective Voxel Space

2.4. DETERMINATION OF PROJECTIVE VOXEL SPACE

In the section 2.3, three images are rectified and epipolar lines are aligned to the epipolar lines between images. Setting the three rectified images as shown in Figure 5, a voxel space is defined. The searching of matching points in three images is equal to 3D reconstruction in this voxel space. We name this space "Projective Voxel Space" (PVS) , because the axes in this space are aligned to the direction of the camera projection.

Thus, PVS is a 3D space that has distorted coordinates in Euclidean geometry, and the shape of the object in this voxel space doesn't have similarity to the shape in the real world. The concept of this distortion between PVS and the real world is shown in Figure 6. However, the surface of the shape in PVS denotes the matching points in all images. Thus, the concept of PVS enables correct matching points in two images to determine the occluded matching points in the other images.



(a) (b)

Figure 6. Distortion caused by rectification : (a) epipolar lines and shape of an object in original images ; (b) rectified epipolar lines and distorted shape of the object

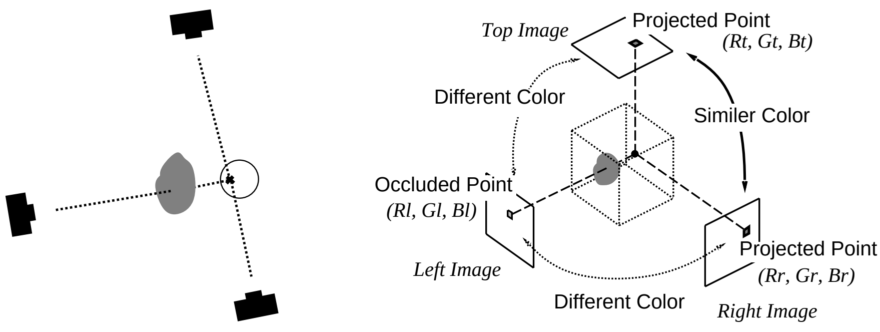
3. SHAPE-FROM-SILHOUETTE METHOD

As mentioned in the previous sections, matching points in all pairs of three images can be handled in the common coordinate system, which is named PVS. The relation between PVS and three rectified images is orthographical projection as show in Figure 5. We apply shape-from-silhouette method in the PVS using silhouette in the three input images, and bounding 3D space of the object in PVS is acquired.

4. SURFACE VOXEL DETECTION

As described in the previous section, shape-from-silhouette method can acquire bounding space of the object in PVS. This bounding shape does not have enough accuracy in most of the cases. The purpose of our method is to synthesize images from free point of view. Synthesis of images requires occluding relationship of voxels that is notion about which point is hidden and which point hides that hidden point. To obtain accurate 3D model with consideration of occlusion, several methods have been proposed.[9,3] In this paper we adopt a simple method, because our purpose does not require detailed 3D shape, but rough shape with correct occluding relationship is sufficient for new view generation.

As described in Section 2, correct matching point is equal to the surface of the object in PVS. The correct matching point might have similar color in the image, so our searching method is based on the color. All voxels inside of the bounding space that is acquired by the previous sections are evaluated by the color of the projected pixels in three images.



(a) (b)

Figure 7. Occlusion in Projective Voxel Space : (a) Occlusion in the real world ; (b) Occlusion in PVS

When a point of the surface is not in occlusion, the projected pixels in three images naturally have similar color. If the point is occluded as shown in Figure 7, a point on the surface of the object can be seen from top and right images, and occluded in left image by some other object in this case. Using color of the left image in this case, this voxel can't be detected as the correct matching point. Thus, searching has to be done with consideration of occlusion.

In this paper, we assume that all surface of the object can be seen in at least two images. With this assumption, a voxel, which contains correct matching points, must be projected to the similar color pixel in at least two images with or without occlusion. So, we evaluate the voxel using the most similar pair in the projected pixel in three images. This evaluation is denoted in equation (1), in which color of projected pixel is (R_t, G_t, B_t) in top image, (R_l, G_l, B_l) in left image, and (R_r, G_r, B_r) in right image as shown in Figure 7. The minimum value of E_{tl}, E_{lr}, and E_{rt} is adopted for the voxel evaluation. After evaluating all voxels in the bounding space acquired by shape-from-silhouette, matching points are finally obtained by taking the local minimum voxels in PVS.

$$\begin{aligned} E_{tl} &= \sqrt{(R_t - R_l)^2 + (G_t - G_l)^2 + (B_t - B_l)^2} \\ E_{lr} &= \sqrt{(R_l - R_r)^2 + (G_l - G_r)^2 + (B_l - B_r)^2} \\ E_{rt} &= \sqrt{(R_r - R_t)^2 + (G_r - G_t)^2 + (B_r - B_t)^2} \end{aligned} \tag{1}$$

Conversely, once the surface of the object is acquired anyhow in PVS, then it is easy to detect occluded point in an image. As shown in Figure 7-(b), occlusion in the rectified image must happen in the direction of axis. Thus, all occlusions are equal to the situation that two or more voxels exist on the axis directions in PVS, because all projection lines are orthographic and three camera directions are vertical each other. In this way, the geometric feature in PVS makes occlusion detection very easy. Therefore, the projected position of the hidden voxel in an image can be easily found, and hiding relation between voxels can also be obtained in PVS. As described later in Section 5, such occluding relation is important for synthesis of images.

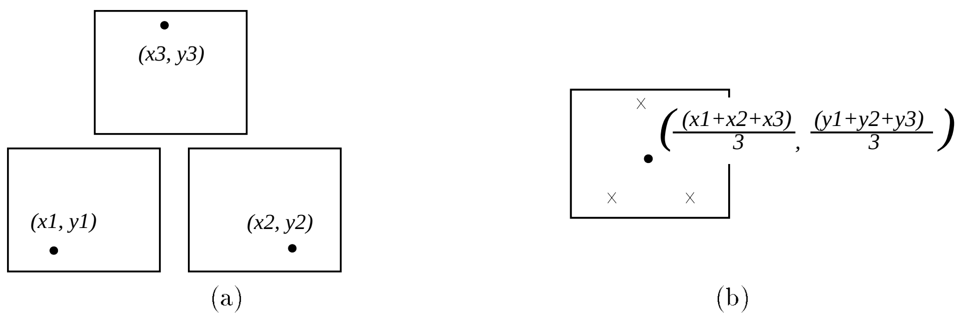


Figure 8. Synthesis of the position of a point by interpolation : (a) Position of matching points in three input images ; (b) Position of the point in the synthesized image in which the virtual viewpoint is the center or three cameras

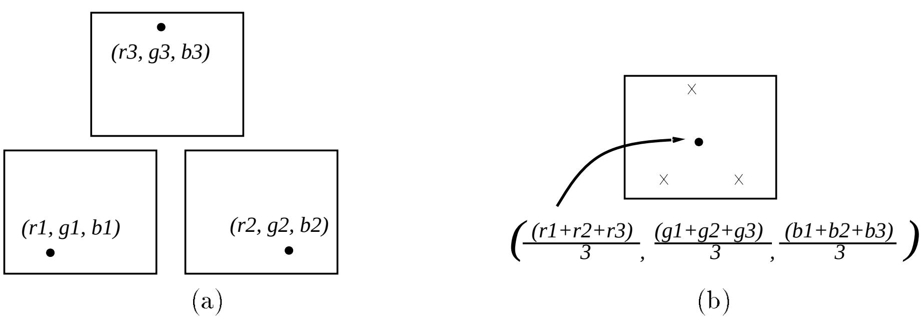


Figure 9. Synthesis of the color of a point by interpolation : (a) Color of matching points in three input images ; (b) Color of the point in the synthesized image in which the virtual viewpoint is the center or three cameras

5. SYNTHESIS OF IMAGES

Simple 2D morphing method can synthesize images based on the matching points in input images. The virtual viewpoint for the synthesized image is described with the ratio of input images, and we call this ratio “view-ratio” here. All matching points are placed in the synthesized image based on the position of matching points in original images and view-ratio. Figure 8 shows an example of this interpolation of geometry. The color of points in the synthesized image is blended from the pixel color in original images based on view-ratio. Figure 9 shows an example of this interpolation of color value.

When a point is occluded in an image, blending from only two visible images solves the color of the interpolated point as shown in Figure 10. Also, the occluded point must appear or disappear based on the occluding relationship of the voxels and location of the virtual viewpoint. The acquired model in PVS contains the occluding relationship between voxels that is notion about which voxel is hidden and which voxel hides that hidden voxel. Figure 11 shows the concept of reproduction of occlusion. In Figure 11-(a), almost all points in the circle region are occluded by the square region in the right image. As described, we have the notion “which point is hidden and which point hides”. This notion can obtain the geometry of points in the occluded circle region in the right image. Then, the position of occluded points in the circle area can be interpolated in the same way of non-occluded points as shown in Figure 8. The position of the points in the square region can be naturally interpolated because they are not occluded.

Figure 11-(b) shows examples of interpolated points in the circle region and interpolated points in the square region. The occluding relationship between voxels also can determine the fact that the points in the square region

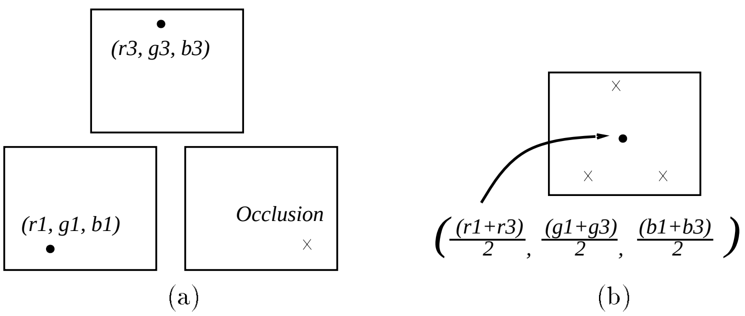


Figure 10. Synthesis of the color for a point in occlusion : (a) Color of matching points in three input images ; (b) Color of the point in the synthesized image in which the virtual viewpoint is the center or three cameras

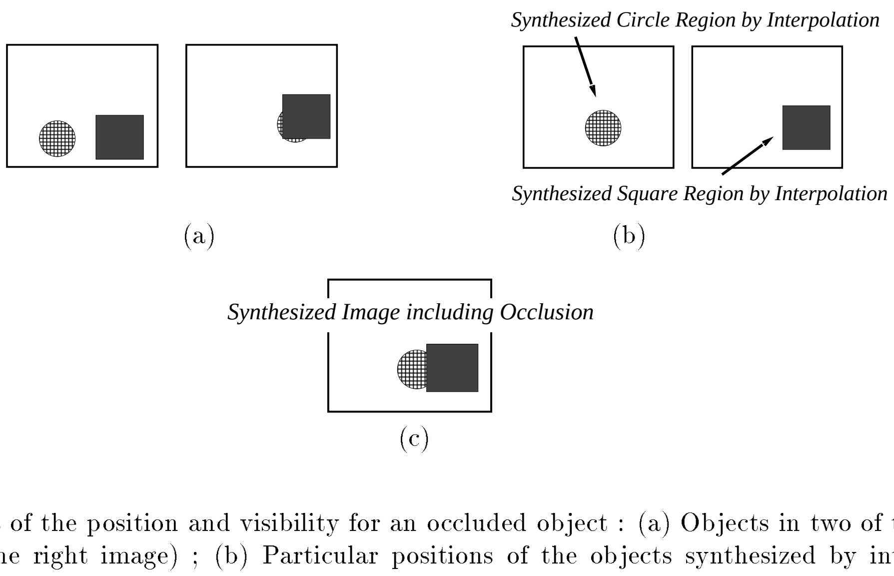


Figure 11. Synthesis of the position and visibility for an occluded object : (a) Objects in two of the input images (occlusion exists in the right image) ; (b) Particular positions of the objects synthesized by interpolation ; (c) Reproduced occlusion in the interpolated image

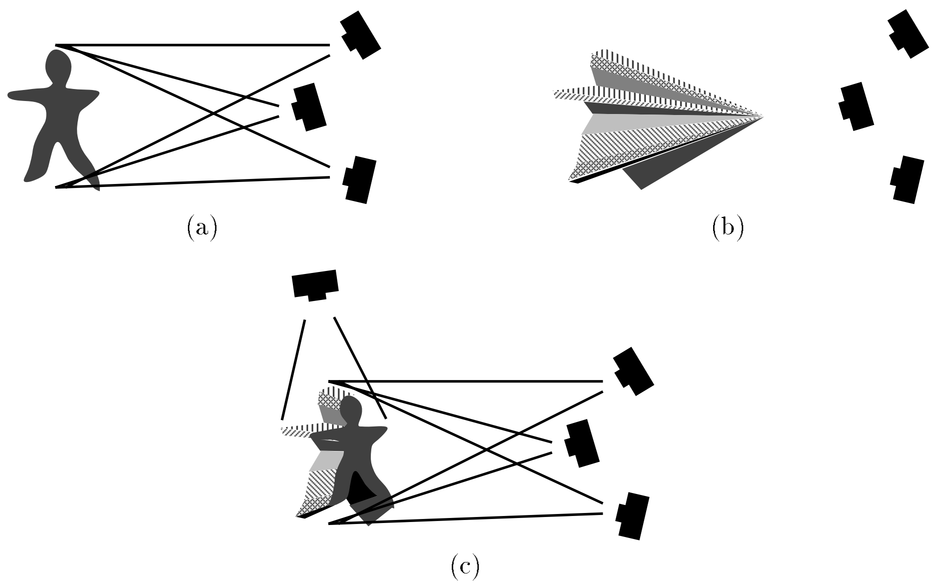


Figure 12. Bounding space acquired by shape-from-silhouette method: (a) Shape-from-silhouette method with three cameras in almost parallel directions; (b) Acquired shape by (a) ; (c) Acquired shape with additional camera

hide the points in the circle region. So, the interpolated point in the square region must be in the forefront of the points in the circle region. Therefore, the occlusion can be reproduced in the interpolation as show in figure 11-(c).

6. EXPERIMENTS AND DISCUSSION

To demonstrate the effectiveness of our proposed method, we tested the method using real images. The fundamental matrices were solved with about 30 matching points in each image. The silhouettes of the target objects were obtained by preprocessing.

A 384 $\times$ 384 $\times$ 512 sized voxel space was constructed from three 640 $\times$ 480 sized input images shown in Figure 13. Figure 14 show depth maps based in the obtained matching in PVS. Figure 15 shows synthesized images based on the obtained matching. Occluded points are synthesized plausibly, because the matching for those occluded points was obtained in PVS. For example, right leg of the person appears or disappears by moving the virtual viewpoint.

Figure 16 shows another input images. A 256 $\times$ 512 $\times$ 512 sized voxel space was constructed from three 640 $\times$ 480 sized images. Three cameras were set in almost parallel direction in this case, and our proposed method required an additional camera. The reason for the additional camera is as below.

In PVS, three rectified images are set in orthogonal directions to each other, however the directions of the cameras are unknown in the real world. Turning back to Euclidean geometry, the shape-from-silhouette method is just a projection of silhouette images to the 3D space. If cameras are set in almost parallel direction in the real world, the acquired shape has huge ambiguity in a direction of center of three cameras in Euclidean space. This is because the silhouette in three images contains almost identical information as shown in Figure 12-(a),(b). Even in PVS, the result of shape-from-silhouette from such camera distribution also contains this kind of ambiguity. In such cases, it is effective to apply one more camera which is set in vertical direction of three cameras, because this vertical camera can reduce this ambiguity as shown in Figure 12-(c).

Figure 17 shows the additional camera image, and Figure 18 shows the depth maps based on the obtained matching in PVS. Figure 19 shows synthesized images based on the obtained matching. Thus, occluded points (e.g. left leg of the person) are synthesized plausibly again, because the additional camera works effectively.

7. CONCLUSION

This paper focuses on interpolation method using 3D voxel concept in projective geometry. PVS is constructed with three images and weak calibration, and shape-from-silhouette method can be applied in this voxel space. A simple method for detect the surface of objects in PVS is also proposed, and a method for synthesis of images is denoted. Interpolated images are based on occluding relationship that makes the synthesis in better quality, because the occluding relationship enables synthesis of occlusion plausibly. Experiments show the effectiveness of our proposed method. In spite of using only simple methods, our proposed method use three images effectively and can interpolate images from real images.

REFERENCES

1. Shai Avidan and Amnon Shashua : "Novel View Synthesis by Cascading Trilinear Tensors" *IEEE Transactions on Visualization and Computer Graphics, Vol.4, No.4, pp. 293-306, 1998*
2. *R. Hartley : "In Defense of the 8-point Algorithm"* IEEE ICCV95 *pp. 1065-1070 1995.*
3. *M. Kimura, H. Saito and T. Kanade : "3D Voxel Construction based on Epipolar Geometry"* Proc. of ICIP '99 *27AO2 1999.*
4. *Stéphane Laveau, and Olivier Faugeras : "3-D Scene Representation as a Collection of Images and Fundamental Matrices"* Research Report 2205, *INRIA Sophia-Antipolis, France, February 1994.*
5. *P.J. Narayanan, Peter W. Rander, and Takeo Kanade : "Constructing Virtual Worlds Using Dense Stereo"* Proceedings of Sixth IEEE International Conference on Computer Vision (ICCV'98), *Bombay, India, pp. 3-10, January 1998.*
6. *Phillip Pritchett and Andrew Zisserman : "Wide Baseline Stereo Matching"* Proceedings of Sixth IEEE International Conference on Computer Vision (ICCV'98), *Bombay, India, pp. 754-760, January, 1998.*
7. *Phillip Pritchett and Andrew Zisserman : "Proceedings of Matching and Reconstruction from Widely Separated Views"* 3D Structure from Multiple Images of Large-Scale Environments European Workshop (SMILE'98), *p-viii+346, 78-92, 1998*
8. *Steven M. Seitz and Charles R. Dyer : "View Morphing"* Proceedings of SIGGRAPH '96, *pp. 21-30, 1996.*
9. *Steven M. Seitz and Charles R. Dyer : "Photorealistic Scene Reconstruction by Voxel Coloring"* Proceedings of Computer Vision and Pattern Recognition Conference, *pp. 1067-1073, 1997.*
10. *Sundar Vedula, Peter Rander, Hideo Saito, and Takeo Kanade : "Modeling, Combining, and Rendering Dynamic Real-World events from Image Sequences"* Proceedings of Fourth International Conference on Virtual Systems and Multimedia, *Gifu, Japan, November 1998.*
11. *Zhengyou Zhang : "Determining the Epipolar Geometry and its Uncertainty : A Review"* Research Report 2927, *INRIA Sophia-Antipolis, France, July 1996.*

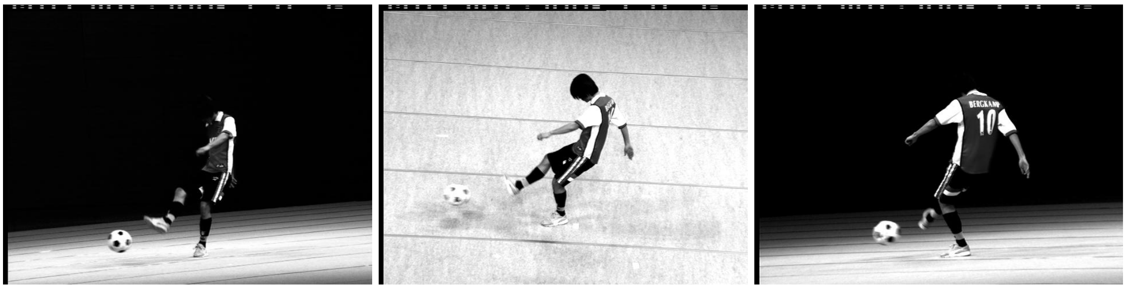


Figure 13. Input images

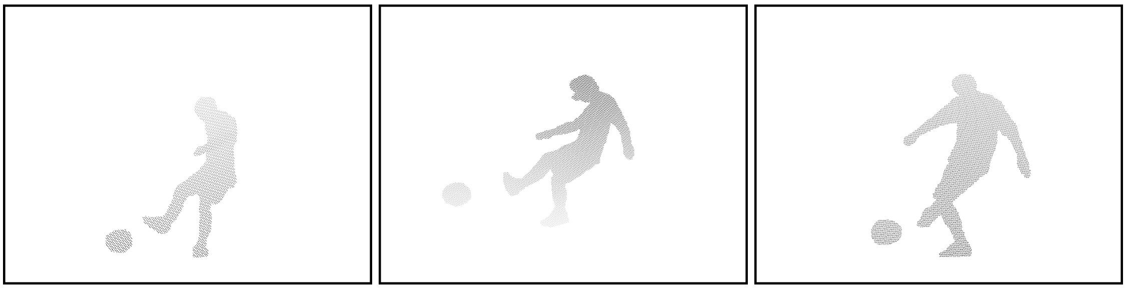


Figure 14. Depth maps based on the reconstructed shape in PVS

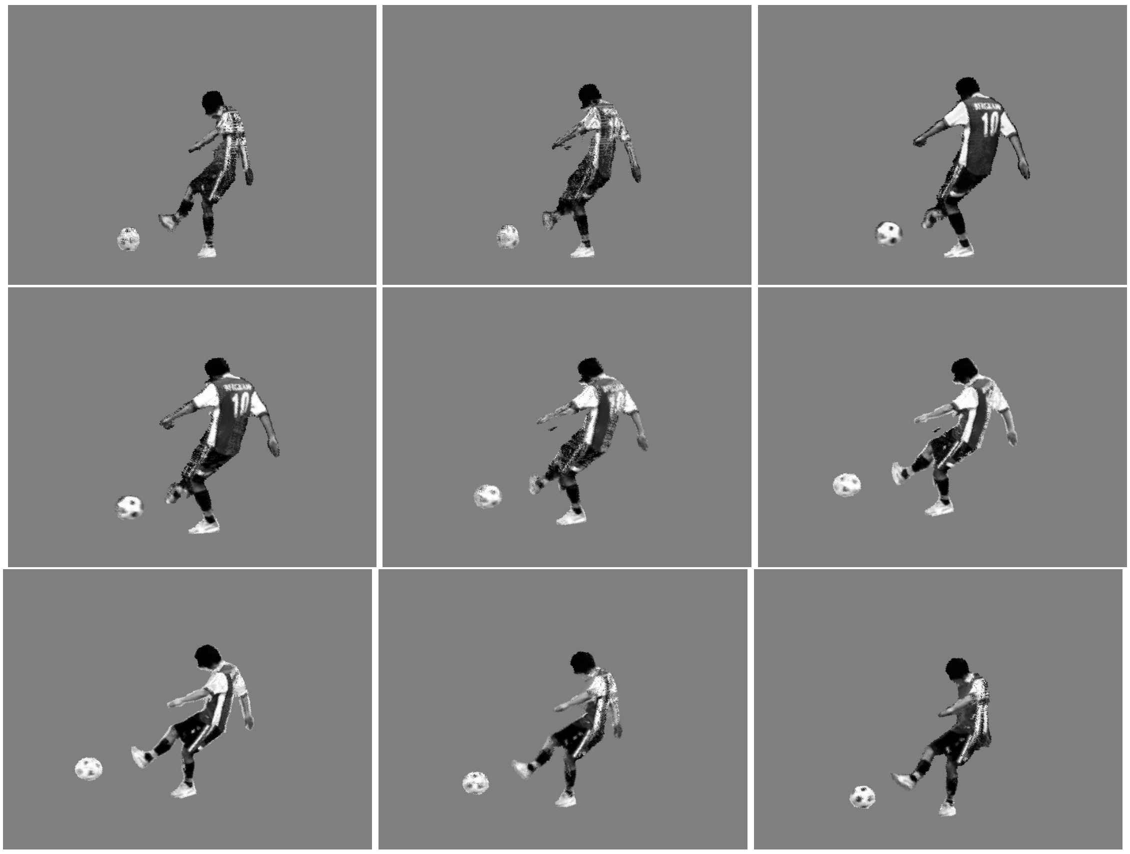


Figure 15. Synthesized images based on the resolved matching

Figure 16. Input images

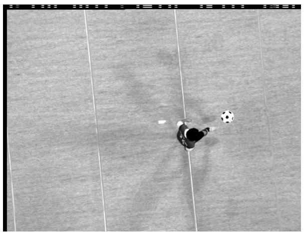


Figure 17. Additional image

Figure 18. Depth maps based on the reconstructed shape in PVS

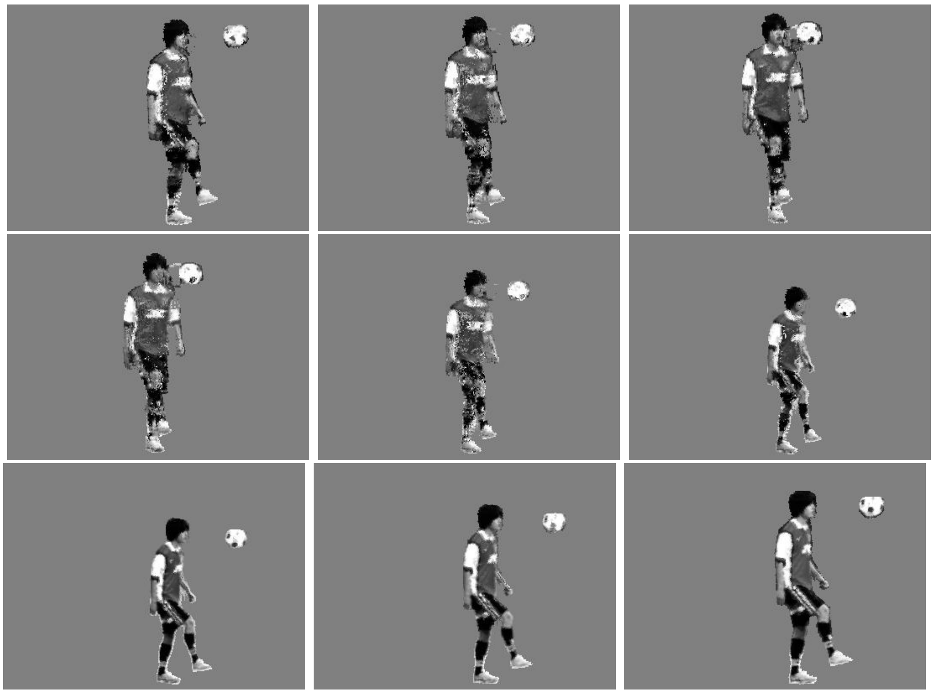


Figure 19. Synthesized images based on the resolved matching