MVA2000 IAPR Workshop on Machine Vision Applications, Nov. 28–30, 2000, The University of Tokyo, Japan

13—32 3D Reconstruction based on Epipolar Geometry

Makoto Kimura* Hideo Saito *
Department of Information and Computer Science, Keio University

Abstract

In this paper, we propose a new method for 3D reconstruction from three cameras based on the projective geometry. If the subject is just synthesizing images from new viewpoint, 3D shape reconstruction in Euclidean space is not required, and projective reconstruction gives enough information to synthesize new viewpoint images. This means that full calibration is not needed but recovery of epipolar geometry between input cameras is sufficient. 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 axis is aligned with the camera's projection direction. Such alignement simplifies the procedure for projection and back projection between the 3D space and the image planes. Taking advantage of this PVS, we apply shape-from-silhouette in the PVS to acquire bounding space of the object. The consistency of color value between the projected pixel in the camera images is evaluated for final determination of the object surface voxel.

1 Introduction

For synthesizing 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 which is performed by checking the correspondence between 3D geometry in world coordinates and 2D geometry in image coordinates [10, 5]. However reconstructed 3D geometry itself is not required explicitly [8], the reconstruction of 3D geometry has some advantages. One of the advantages is handling all data in a common coordinate frame.

Recently, projective geometry has been often used in the field of computer vision [6, 7, 8, 1], because projective geometry can be determined easier than Euclidean geometry. While determination of Euclidean geometry requires a map of correspondences between the points in the image and Euclidean geometry of those points, determination of projective geometry requires only some matching points in each image [11]. 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 make common coordinates like Euclidean geometry.

In this paper, we propose an approach to reconstruct 3D shape in a voxel space from three images and weak calibration. We call this voxel space "Projective Voxel Space"(PVS). The coordinate axes are not orthogonal in PVS, so the reconstructed 3D shape itself is not equal to 3D shape in Euclidean geometry. However, PVS makes it possible to handle all data in a common coordinate frame, and the reconstructed shape describes enough 3D information for synthesizing images from new point of view.

2 Projective Voxel Space

First, 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 arbitrarily points in the other image. In this way, fundamental matrix can restrict the searching area of a matching point in a line in each image, because the correct matching point must exist in the corresponding epipolar line in each image.

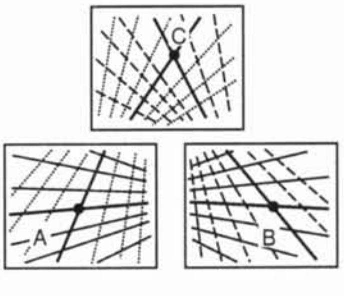

Figure 1: Epipolar lines in three images (A,B,C : example of matching points)

*3–14–1 Hiyoshi Kouhoku-ku Yokohama 223-8522, Japan. {kimura,saito}@ozawa.ics.keio.ac.jp

As mentioned, fundamental matrix can solve epipolar lines between a pair of images. Then, epipolar lines in three images have many intersections as show in Figure 1. Considering these intersections as re-sampled pixels, a matching information between two of the images can identify the matching point in the third (the other) image, because we have epipolar information in all re-sampled pixels.

This feature enables to rectify three images so that the vertical and horizontal directions can be completely aligned to the epipolar planes between the cameras. The rectified images provides Projective Voxel Space (PVS), in which the three axes is aligned with the camera's projection direction. In PVS, all lines parallel to one of the axes is parallel to projection direction of one of three cameras, and all planes vertical to one of axes is epipolar plane between two cameras. Such alignment simplifies the procedure for projection and back projection between the 3D space and the image planes, because the 3D index in PVS implies the projected point onto each camera image as shown in Figure 2.

The shape of the object in this voxel space doesn't have similarity to the shape in the real world. However, the surface of the reconstructed shape is equal to the correct matching point information in each image, and PVS is enough to detect correct matching points.

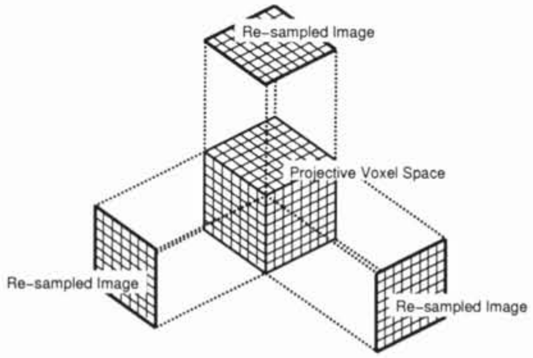

Figure 2: Projective Voxel Space

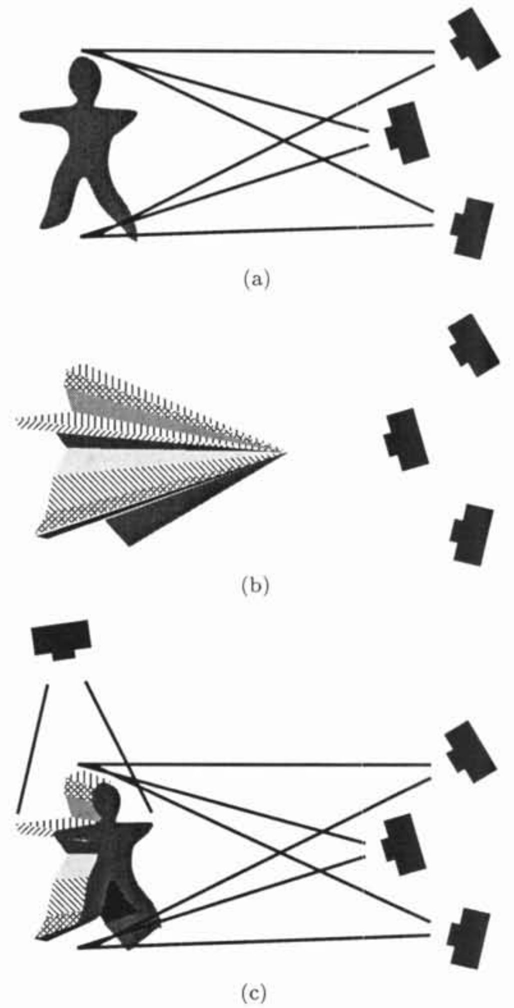

Figure 3: Bounding space acquired by shape-from-silhouette method: (a) Shape-from-silhouette method with three cameras ; (b) Acquired shape by (a) ; (c) Acquired shape with additional camera

3 Shape-from-silhouette method

Our reconstruction method is based on shape-from-silhouette method. We apply shape-from-silhouette method in the PVS using silhouette information in the three input images. The accuracy of the obtained 3D shape is dependent on the location of three cameras. If cameras are set in almost parallel direction, the acquired shape has ambiguity in a direction of center of three cameras in Euclidean space. This is because the silhouette information of three images are almost identical as shown in Figure 3-(a),(b). In this case, we should 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 3-(c). In this way, bounding 3D space of the object is acquired from silhouette information.

4 Surface Voxel Determination

As described in previous sections, shape-from-silhouette method acquires bounding area of the object. This bounding area does not have enough accuracy in most of the cases. The purpose of our method is to synthesize images from free point of view, and it requires topological information of the occlusion. To obtain accurate 3D model with consideration of occlusion, complex or iterative method has been proposed [9, 3]. In this paper we adopt a simple method, because our purpose does not require detailed 3D information and rough shape with correct topological information is enough.

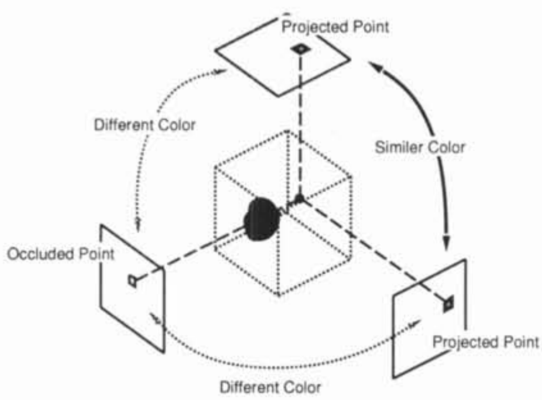

Figure 4: Occlusion in Projective Voxel Space

From the definition of PVS, correct matching point is equal to surface of the object. In most cases of the occlusion, the invisible point in a image might be visible in the other two images as shown in Figure 4. Then, pixels which are projected from surface voxel have similar color information in two images. When a point is not in occlusion, the projected pixels in three images have similar color naturally. In both cases, the projected pixels have same color at least in two images. Then, correct matching point is searched by a assumption that the projected surface voxel has same color in two of the images. The obtained matching points are equal to the surface of the object. Since projection of each voxel in PVS onto each camera images is very simple (i.e. voxel index implies the projected point as shown in Figure 4), the consistency check of color value described above can be performed by the very simple way.

5 Experiments

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 silhouette of the target objects were obtained by preprocessing. A $512 \times 512 \times 512$ voxel space was constructed from three 640×480 images.

Figure 5 shows input images and Figure 6 shows the depth map based on the obtained surface of the model in PVS. Figure 7 shows synthesized images based on the obtained matching information. Occluded points (e.g. left leg of the person) are synthesized plausibly, because the matching information beyond some other points exists in PVS.

6 Conclusion

This paper focused on 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 is also proposed. The accuracy of proposed method depends on the arrangement of cameras, and the arrangement needs more examinations.

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 5: Input images

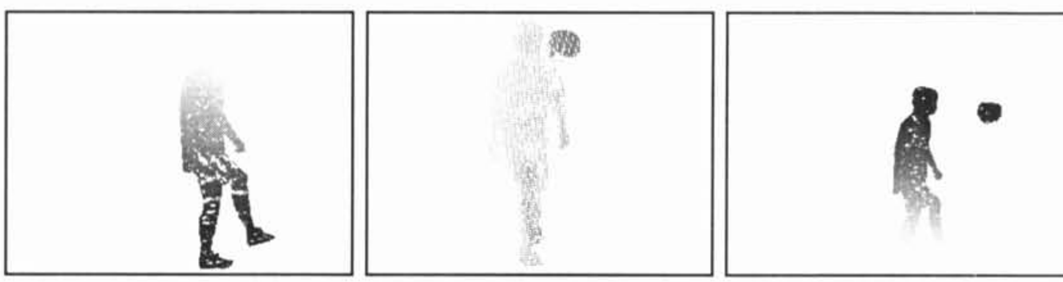

Figure 6: Depth maps based on the reconstructred shape in PVS
(depth value contains the depth in PVS)

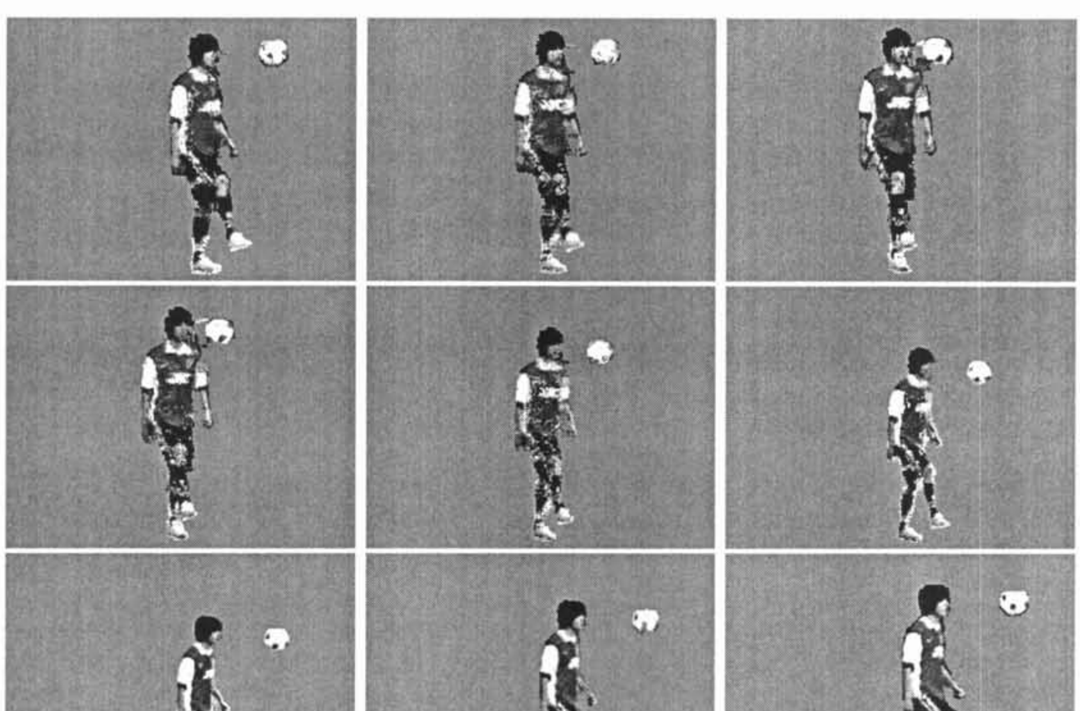

Figure 7: Synthesized images based on the reconstructred shape