3D Voxel Construction based on Epipolar Geometry

Makoto Kimura[†‡] Hideo Saito[†‡] Takeo Kanade[†]

[†] The Robotics Institute, Carnegie Mellon University
500 Forbes Avenue, Pittsburgh, PA 15123 USA
[‡] Department of Electrical Engineering, Keio University, Yokohama 223-8522, Japan

Abstract

In this paper, we propose an approach for construction of a projective 3D voxel space based on epipolar geometry obtained with weak calibration. This concept of voxel space defines the epipolar lines following to the coordinate axis. In the field of computer vision, it is common to reconstruct 3D geometry data based on camera calibration data and disparities of matching points in each image. In the case that the goal of the system is generating the images from another point of view, complete 3D geometry reconstruction is not presently required. However, detecting the consistent matching points in several pairs of images without complete 3D geometry information is difficult. This paper describes the concept of projective 3D voxel, which makes it possible to handle 3D geometric data without complete 3D geometry information.

1 Introduction

For synthesizing images from new point of view based on real multiple images, one approach is to reconstruct the 3D shape of objects in the scene, so that images can be generated from 3D shape and texture data. This method has been applied in Virtualized Reality [10] [4], which requires calibration of about 50 cameras. This calibration for each camera is performed by checking the correspondence between 3D geometry in world coordinates and 2D geometry in image coordinates. One of the purpose of Virtualized Reality is synthesizing images of real scene from free point of view. Considering this purpose, reconstructed 3D geometry is not required explicitly [8].

The reconstruction of 3D geometry has some advantages, and one of the advantages is handling all data in a common coordinate frame. 3D geometry reconstruction itself is a common subject in the field of computer vison, and in addition, it requires camera calibration information which has difficulty and complexity of acquisition.

One of the common problem in 3D geometry reconstruction is occlusion in the images. 3D geometry is reconstructed by camera calibration information and matching points in the images, while no matching points can be detected for occluded regions. There are researches for 3D geometry reconstruction using more than two cameras to solve this problem. The main purpose of using many cameras is to reconstruct accurate 3D geometry, as the occluded region might be seen from another camera [9] [10] [7].

The detection of matching points is also difficult, even excepting the problem of occlusion. Especially in the case of the baseline of the cameras being wide, it is more difficult to obtain the correct matching points in each image [5] [6].

Recently, projective geometry has been made use of in the field of computer vision [5] [6] [8] [1], because projective geometry can be determined easily. 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, and the 3D geometry information of the matching points is not required in projective geometry [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 construct a projective 3D voxel space from three images. With this method, a 3D voxel space and three orthographically projected images are generated from three images and weak calibration information. This 3D voxel is not a kind of Euclidean or world coordinates 3D geometry, however we can handle the voxel space like a Euclidean one.

2 Epipolar Geometry

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

To get the epipolar lines from arbitrarily points in the other image, the epipolar geometry of the images has to be determined, and it is described by a 3×3 matrix called "fundamental matrix" [11] [3]. Fundamental matrix has only 7 degrees of freedom. Therefore, fundamental matrix is solved with 7 point matches in the images with non-linear method, and 8 point matches are enough for linear solutions [2]. In addition, more matching data makes it more accurate.

Once the fundamental matrix is obtained, weak calibration for that camera pair is finished. The weak calibration is different from traditional calibration which we call "strong calibration". While the strong calibration is identification of all camera parameters, weak calibration identifies only epipolar geometry of a camera pair. As described above, fundamental matrix makes it possible to get the epipolar lines from arbitrarily points in the other image. In the other words, fundamental matrix can limit searching area of matching points.

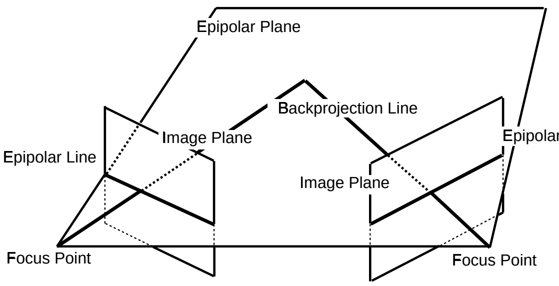


Figure 1: epipolar geometry

3 Projective 3D voxel construction

In this section, a method for projective 3D voxel construction is shown. The outline of the method is as follows.

1. definition of 3D voxel space based on three images and epipolar geometry of each pair of the images
2. substitution of each 3D voxel with correlation value
3. improvement of each voxel value
4. smoothing each voxel value
5. (if necessary) bask to 3.

The definition of 3D voxel is mentioned in Section 3.1, the substitution of each 3D voxel is mentioned in Section 3.2, and the improvement and smoothing are mentioned in Section 3.3.

3.1 Definition of 3D Voxel Space

The first step in this approach is to solve for the fundamental matrices from the matching points data in three images. With this fundamental matrices, it is possible to get an epipolar line from arbitrary points in the other image. Therefore, epipolar lines in each image are obtained with the fundamental matrices. By arranging the interpolated value of each intersection point of epipolar lines, new images are generated. A projective 3D voxel space is constructed by setting these three generated images like Figure 2,

This 3D voxel is different from Euclidean one (Figure 3). Because the direction of each axis is based on epipolar geometry, the shape in this space is not like the shape in the real world. However, this voxel space has a useful property like Euclidean one, and also it has one better feature than real Euclidean voxel.

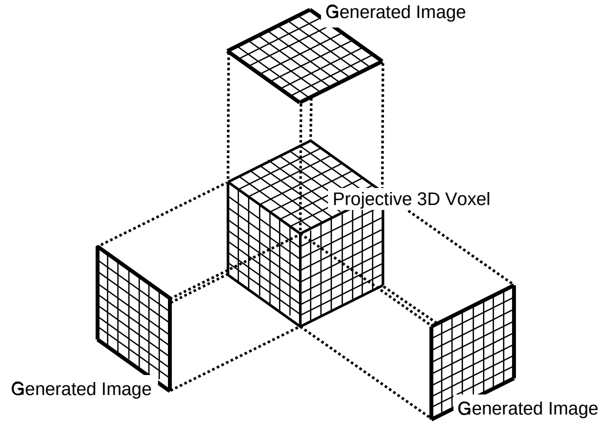


Figure 2: The connection of the generated images and projective 3D voxel

A correct matching point is corresponding a voxel in this 3D space, thus a voxel hidden by another existing voxel is not visible in the image. If the purpose of the system is synthesizing interpolated images from multiple images, knowing the hiding information is important. Hiding information makes it possible to synthesize a image of occluded region [7]. In this way, 3D shape reconstruction in the projective 3D voxel space works for synthesizing images as well as that in the Euclidean space.

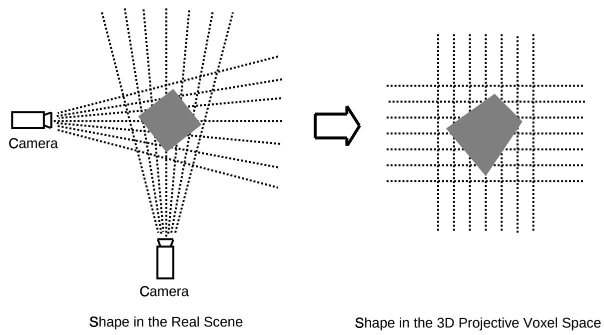


Figure 3: The relation between real world and constructed 3D voxel space

Each spatial axis of the projective 3D voxel is corresponding to the projection lines for each image in the real world. Therefore, the generated images behave like orthographic projections in this projective 3D voxel space (Figure 2). This behavior simplifies geometrical transformation between the voxel and the images.

3.2 Initialization of Voxel Value

The prior section describes the method to construct a projective 3D voxel space based on weak calibration.

This projective 3D voxel is determined by the projective geometry between thee images. Conversely, the relation between three images and voxel space can be described as follows : there is a scene in this 3D voxel space, and these three sampled images are the orthographic projected images of this scene. Thus, to detect matching points in each pair of the images is equal to detect surface of the object in this voxel space. In other words, the value of a voxel should show how certain is the matching of the points in three images. Hereafter, the voxel value is called "likelihood" of the matching point.

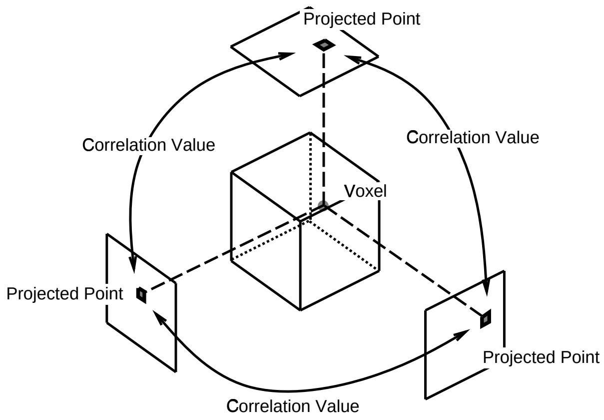


Figure 4: Three correlations between three images

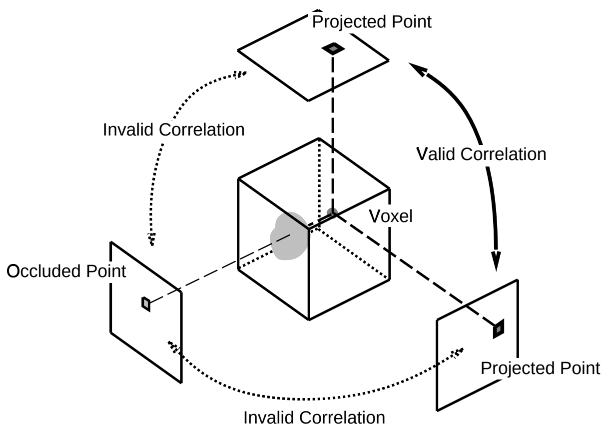


Figure 5: The voxel in occluded area

It is reasonable to fill each voxel with the values of correlation between pairs of points in the projected images (Figure 4). There are three correlation values because there are three projected images. However, all three values should not be adopted. In the case of that a voxel is in a occluded area in a image, that voxel is visible in only two images. Figure 5 shows a example of this condition that a voxel is visible in top image and right image, and invisible in left image. In this example, correlation between left image and another image is not valid for the purpose of determination of likelihood, and correlation between top and right image is valid for the purpose. In practice, we don't know the voxel is visible or invisible, but the valid correlation value is anticipated to be higher than the invalid correlation values. Thus, the voxel should be filled with not the average or total value of three correlation values, but maximum value of three.

3.3 Iterative Improvement of Voxel Value

The previous section applied initial likelihood value in each voxel. This likelihood value is not enough for detecting matching points because it contains many mismatched points values. In this section, an iterative improvement method is proposed.

This iterative improvement method is explained in equation (1).

$$V_{n+1}(x,y,z) = \left(\frac{V_n(x,y,z)}{M(S_x(x,y,z), S_y(x,y,z), S_z(x,y,z))}\right)^{\alpha} \quad (1)$$

Where $S_x(x,y,z)$ is a sum total of squared value of all voxels in the projection line from voxel (x,y,z) to the direction of x axis (Figure 6), and $S_y(x,y,z)$ and $S_z(x,y,z)$ are the same kind of values of the direction y and z axes. $M(value1, value2, value3)$ is a sum total of two smaller values of $value1$, $value2$, and $value3$. $V_n(x,y,z)$ is a voxel value in the geometry (x,y,z) at iteration n.

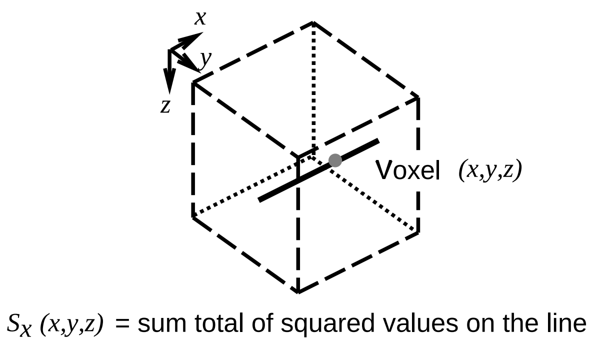


Figure 6: Definition of $S_x(x,y,z)$

This iterative improvement is based on the following concept: if a voxel is a right matching point, then there will be no other matching points on a projection line passing through that voxel. However, this is not always correct in all projection lines. As shown in Figure 5, in the case that a voxel is in occluded area in an image, this concept is acceptable only in two projection lines going towards images which can see that voxel. Additionally, the visible voxel points must be handled in the same way because there may be another invisible voxel which is occluded by this voxel. Thus, the role of $M()$ which adopts two values from three is important in this method.

Furthermore, a smoothing method considering 3D direction is also used in this iteration method. This method is a 3D smoothing method using 3D mask data, and this 3D mask shape is shown in Figure 7.

This shape is determined in the purpose of smoothing to directions of the same disparities.

Using these iterative methods, only real matching voxel remains as a high likelihood value in this 3D voxel space.

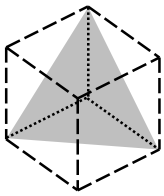


Figure 7: Shape of the smoothing mask

4 Experiments

The effectiveness of our method is shown using three real images. The parameter α was set to be 3, and the number of iteration times was 2. The fundamental matrices were solved with about 500 matching points in each image. The error of the fundamental matrices, which were calculated as the distance between the calibration points and the epipolar lines, were less than 1 pixel range.

The regions of the target objects were obtained by preprocessing. A $256 \times 256 \times 256$ voxel space was constructed from three 320×240 images. Figure 8 shows epipolar lines on the images. The depth maps based on the obtained correspondence in the 3D voxel space is shown in Figure 10. There are many occluded areas in any pair of the images, however the disparity maps of those areas are obtained by using a pair of the other pair of the images. Figure 10 shows a synthesized image based on the matching data acquired by this method. The point of view was set at center of three images.

For the evaluation of iterative improvement method, the likelihood before iterations (the raw correlation data) and the likelihood after each iteration in sub-volume of the voxel space are shown in the Figure 11. In the raw correlation, many points show high value. As shown in the figure, ambiguity of the matching is reduced by the iterative improvement method.

5 Conclusion

This paper focused on 3D voxel concept in epipolar geometry. The projective 3D voxel space is constructed with three images and fundamental matrices of all pair of the three images. Thus, the projective 3D voxel space requires only weak calibration. However the images can be handled as orthographic projected images of this projective 3D voxel space. In this paper, correspondence between three images is solved with correlation data and iterative improvement method based on simple conditions.

However, some subjects still remain in this method. The most important subject is the limitation of the number of input images. This method can handle only three images in a voxel space, and the other three images will make another voxel space. Even using three images, there are occluded regions in each images, so, it is needed to find a way to merge multiple voxel spaces or handle more images in one voxel space.

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] 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.

[4] 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.

[5] 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.

[6] 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

[7] Hideo Saito, Shigeyuki Baba, Makoto Kimura, Sundar Vedula, and Takeo Kanade : "Appearance-Based Virtual View Generation of Temporally-Varying Events from Multi-Camera Images in the 3D Room" *Computer Science Technical Report, CMU-CS-99-127,* April 1999.

[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.

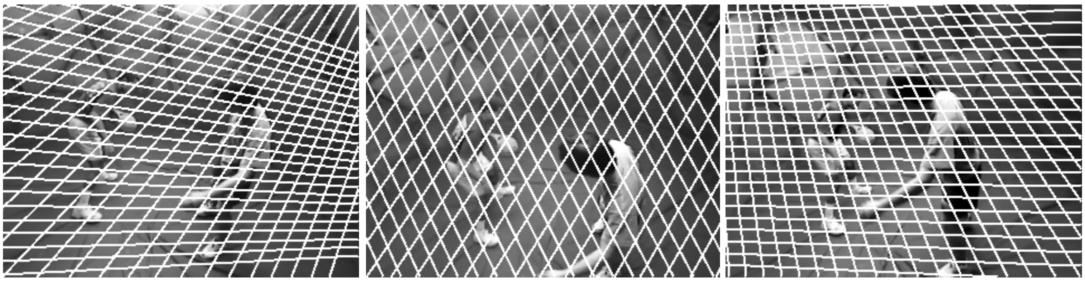


Figure 8: Epipolar lines of each input images

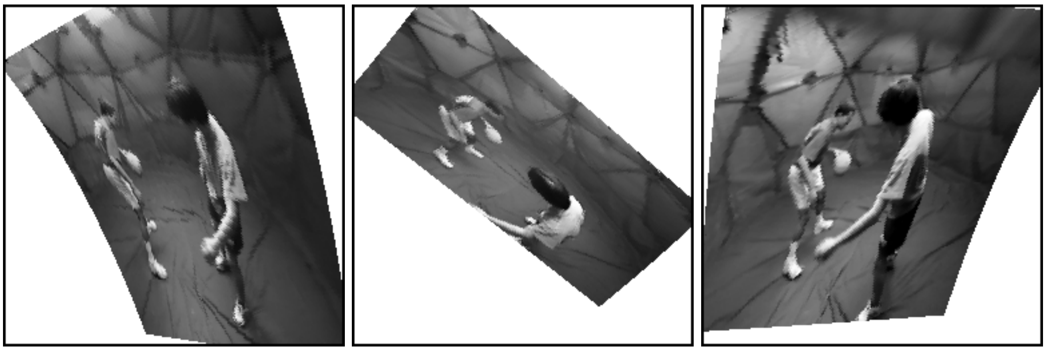


Figure 9: The images converted from the real images based on epipolar geometry

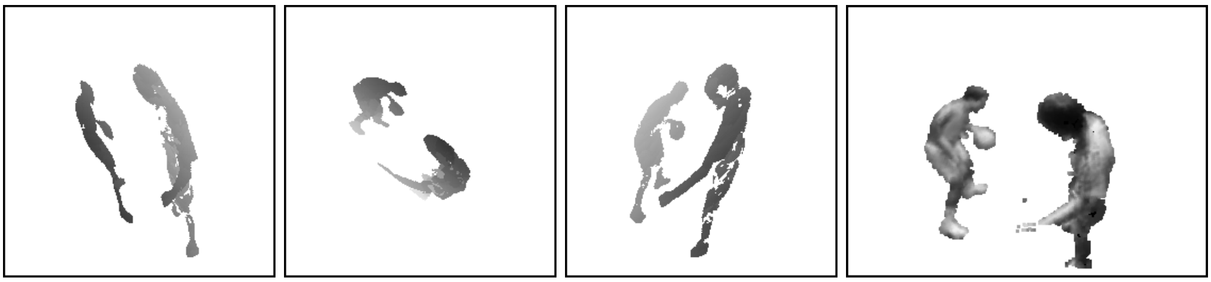
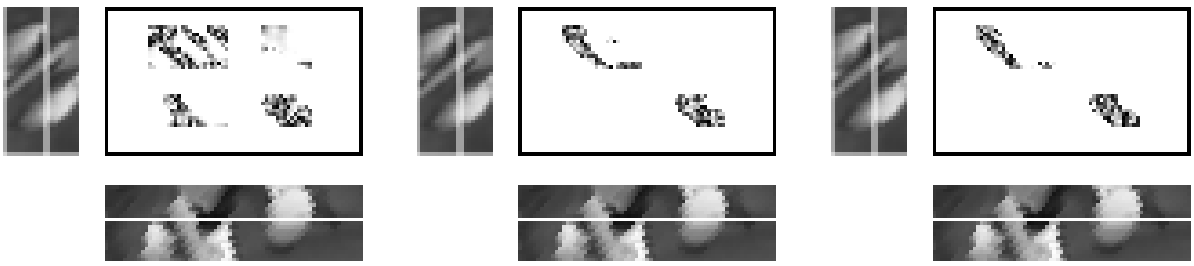


Figure 10: Disparity maps in the projective 3D voxel (three left images), and Synthesized image (right)



(a) iteration=0 (b) iteration=1 (c) iteration=2

Figure 11: The likelihood after two iterations (right) is imprved from raw correlation values (left). Top-right of each figure indicates the cross-section of the likelihood volume. Top-left and bottom-right of each figure show the sub-region of the generated images (Figure 9).