Shape Reconstruction in Projective Grid Space from Large Number of Images

Hideo Saito* Takeo Kanade
Robotics Institute
Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213, USA
hideo@cs.cmu.edu tk@cs.cmu.edu

Abstract

This paper proposes a new scheme for multi-image projective reconstruction based on a projective grid space. The projective grid space is defined by two basis views and the fundamental matrix relating these views. Given fundamental matrices relating other views to each of the two basis views, this projective grid space can be related to any view. In the projective grid space as a general space that is related to all images, a projective shape can be reconstructed from all the images of weakly calibrated cameras. The projective reconstruction is one way to reduce the effort of the calibration because it does not need Euclid metric information, but rather only correspondences of several points between the images. For demonstrating the effectiveness of the proposed projective grid definition, we modify the voxel coloring algorithm for the projective voxel scheme. The quality of the virtual view images re-synthesized from the projective shape demonstrates the effectiveness of our proposed scheme for projective reconstruction from a large number of images.

1. Introduction

Vision systems using a large number of cameras have recently been developed for a wide variety of applications. Kanade et al. [5, 13] have developed the system using a number of cameras for digitizing whole real world events including 3D shape information of the dynamic scene. Davis et al. [3] have developed a multi-camera system for human motion capturing without any sensors on the human body. Jain et al. [4] proposed Multiple Perspective Interactive (MPI) Video, which attempts to give viewers control of what they see, by computing 3D environments for view generation. This is accomplished by combining a priori environment models and dynamic pre-determined motion models. Even in the case at one camera, 3D structure recovery from a moving camera involves a number of images around the object [11] [8].

3D shape reconstruction from multi-view images has recently become an intensively researched area because of recent advances in computation power and capacity of data handling. In such algorithms, registration between the images is the key issue so that the correspondence can provide 3D information about the scene. To achieve accurate correspondence, camera calibration plays an important role. If we have only two cameras, the calibration is not so difficult, but increasing the number of cameras (or views) increases labor of the calibration because, as the number of DOFs of the camera parameters is increased, it becomes increasingly more difficult to get consistent calibration throughout every camera.

Recently, projective reconstruction [2][15] has been introduced as a means for reducing camera calibration efforts. For projective reconstruction, no metric information is required. Only epipolar geometry between cameras, which is represented in the fundamental matrix, is required.

In this paper, we propose a new scheme for taking advantage of the projective reconstruction in the case of a large number of images. In this scheme, we select two basis views for defining a "projective grid space", which can be determined by a fundamental matrix relating the two basis images. The projective grid points can be re-projected onto an arbitrary image by fundamental matrices which relate that image to the two basis images. Since this re-projection relates every grid point to every view image, projective shape reconstruction from number of images can be performed in the projective grid space.

The proposed shape reconstruction scheme in the projective grid space requires only fundamental matrices that relate every view image to two basis view images. Because the fundamental matrix can be obtained only from the correspondence of several calibrating points in the images, much effort for camera calibration can be reduced in the projec-

* Dr. Saito is a Visiting Scientist from Department of Information and Computer Science, Keio University, Yokohama 223-8522, Japan.

0-7695-0149-4/99 $10.00 © 1999 IEEE

tive shape reconstruction as opposed to the case of Euclidean shape reconstruction [12], which basically requires 3D position in Euclidian space of several calibrating points. While 3D positioning accuracy of the calibrating points is a significant factor for accurate Euclidian calibration, such 3D positioning accuracy does not affect the projective reconstruction. This advantage of the projective reconstruction can be realized for shape reconstruction from a number of images by applying the proposed projective grid space.

For demonstrating the effectiveness of the proposed projective grid space, we employ the voxel coloring algorithm proposed by Seitz et al. [10, 6] under the proposed scheme. We also obtain a correspondence point map from the projective reconstruction, and then synthesize virtual view images by applying view interpolation [14] to the correspondence map between images.

2. Projective Grid Space

Two view images are selected as the basis of the projective grid space. Figure 1 explains the scheme for defining one point in the space by the two reference views. 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 either horizontal position r or vertical position s in the second image. Since the fundamental matrix $\boldsymbol{F}_{21}$ limits the position in the basis view 2 on the epipolar line $\boldsymbol{l}$, either r or s is sufficient for defining the grid point. In this way, the projective grid space can be defined by the two basis view images, of which node points are represented by (p, q, r). A node of the grid (p, q, r) is projected onto (p, q) and (r, s) in the first reference image and the second reference image, respectively, where s depends on (p, q, r) as

$$s = -(l_x r + l_z)/l_y \tag{1}$$

where the vector, $\boldsymbol{l} = (l_x, l_y, l_z)$, represents the grid line projected onto the second basis view, which is expressed as

$$\boldsymbol{l} = \boldsymbol{F}_{21} \begin{bmatrix} p \\ q \\ 1 \end{bmatrix} \tag{2}$$

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

The relationship between the projective grid and an arbitrary image is determined by the two fundamental matrices of the image with the two basis images, $\boldsymbol{F}_{i1}$ and $\boldsymbol{F}_{i2}$ for ith image. The projected point from a grid point (p, q, r) is derived by the following procedure. Since (p, q, r) is projected onto (p, q) in the first reference image, the projected grid point in the ith image must be on the epipolar line $\boldsymbol{l}_1$ of

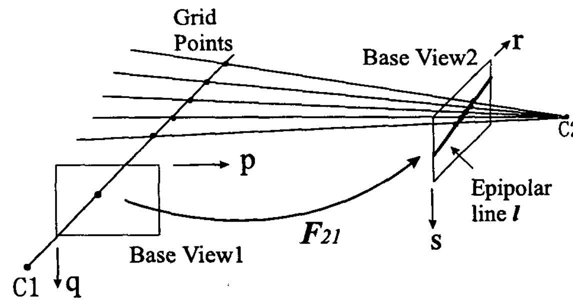


Figure 1. Definition of projective grid. Two basis views define the grid point in the projective space, where each grid position is represented by (p, q, r).

(p, q), which is derived by the $\boldsymbol{F}_{i1}$ as

$$\boldsymbol{l}_1 = \boldsymbol{F}_{i1} \begin{bmatrix} p \\ q \\ 1 \end{bmatrix} \tag{3}$$

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

$$\boldsymbol{l}_2 = \boldsymbol{F}_{i2} \begin{bmatrix} r \\ s \\ 1 \end{bmatrix} \tag{4}$$

The intersection point between the epipolar lines $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$ is the projected grid point (p, q, r) onto the ith image. 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 reference images. Figure 2 shows this scheme.

In the proposed scheme of the definition of the projective grid, all the geometrical relationships are sufficiently represented by only fundamental matrices between two images, because every grid line defined by the position in the first image can be uniquely determined by the fundamental matrix. In this way, every projective grid point can be projected onto every image plane.

3. Voxel Coloring in Projective Grid Space

3.1. Sweeping of plane

In the previous section, we proposed the new scheme for defining the projective grid space for a large number of images. For demonstrating the effectiveness of the proposed scheme, we employ the voxel coloring method for reconstruction of the shape from the input images. The voxel

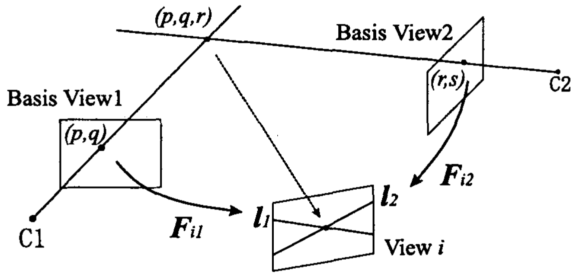
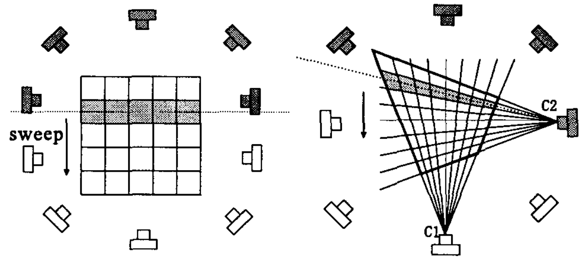


Figure 2. Projection of point in the space onto an image. The point (p, q, r) on the projective grid is projected to (p, q) and (r, s) on the first basis view and second basis view, respectively. In the image of view i, the cross point of two epipolar lines is the projected point of (p, q, r).



(a) Sweeping in Euclidean grid space. (b) Sweeping in projective grid space.

Figure 3. Sweeping plane strategy in voxel coloring. Only the cameras before the sweeping plane (gray colored cameras in the figures) are used for checking the occupancy of the voxle on the plane.

coloring algorithm basically uses color consistency from every viewpoint for determining each voxel's occupancy and color.

The original algorithm requires every camera be strongly calibrated in the sense of Euclidean calibration for determining the geometrical relationship between each voxel and each image. By applying the proposed projective grid definition, Euclidean calibration is not required. Rather projective calibration is sufficient for shape reconstruction because every projective grid point is completely related to every image location under the proposed projective grid space as described in the previous section.

In the voxel coloring algorithm, occupancy of each voxel must be checked by only visible cameras from the voxel. To ensure that only visible cameras are selected for occupancy checking, each voxel is visited in the order of sweeping plane as shown in Figure 3(a), where only the cameras before sweeping plane in the space are used for the voxel occupancy checking. The sweeping strategy ensures that occupancy of each voxel can be checked by only visible cameras [6].

To make the sweeping strategy possible in the projective grid scheme as shown in Figure 3(b), the camera position must be represented in the projective grid space. Although we have only fundamental matrices relating every image to basic view images, the position of every image's camera can be derived from the fundamental matrices as described in the following section.

3.2. Camera position in projective grid space

The fundamental matrix provides sufficient information about the relative camera position because the epipole, which is the projected position of the camera center, can be derived from that matrix.

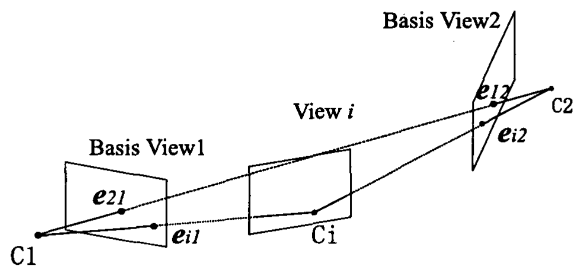


Figure 4. Camera position in projective grid space. Epipolar derived from fundamental matrix provides the camera position in projective grid space.

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, 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 the epipoles. Since the epipoles can easily be derived from fundamental matrices, every camera position in the (p, q, r) coordinate can be obtained from only fundamental matrices.

4. Experimental Results

We applied the projective voxel coloring to the basketball images taken with the 3D Dome system [13]. There are 51 view color images, which are distributed on a hemisphere of 5m in diameter. We chose two views for projective basis from these 51 views.

4.1. Calibration

We collect calibration images for obtaining fundamental matrices between the images. The images are collected while straight bar with LED point light sources as a calibration rig is swept in the reconstructed region. The collection for every camera is simultaneously performed, so that the same dot in the same position can be seen from every camera.

From the calibration images, we have a number of lines on which the dots of LED are aligned. Because of the radial distortion of the lenses, those lines are curved in the images, even though they are actually straight lines. We estimate the radial distortion parameter by using those line segments. The basic idea of this distortion correction method is based on [9]. The radial distortion is corrected by the following equation

$$\begin{aligned} x' &= x + (x - x_c)\{k_1 R^2 + k_2 R^4\} \\ y' &= y + (y - y_c)\{k_1 R^2 + k_2 R^4\} \end{aligned} \tag{5}$$

where $R = \sqrt{(x - x_c)^2 + (y - y_c)^2}$

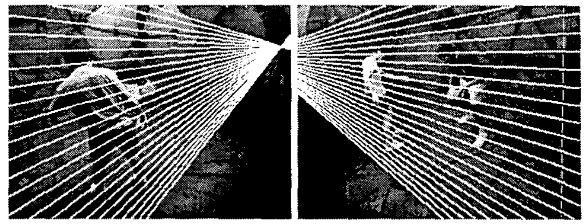
(x, y)	distorted point
(x', y')	distortion corrected point
k_1 and k_2	distortion coefficients
(c_x, c_y)	optical center of the lens

We first define evaluation of linearity of the LED points, then search the distortion parameters of k_1, k_2, c_x, and c_y which give the most linear evaluation.

The convergence is good because we have a sufficient number of lines in the image. This correction does not need any 3D information because the radial distortion has an affect in a 2D image plane.

For estimating fundamental matrices, the LED dots are labeled so that they are matched with the other LED dots in the other images. We obtain approximately 500 correspondence dots between the images. By using these correspondence points, we estimate the fundamental matrix by using Zhang's method [15].

In this calibration process, we use only line segments for radial distortion, and dot correspondences for fundamental matrix estimation. We do not need any relationship of 3D world coordinates to 2D image coordinates, which is generally required for Euclidean calibration as shown in Tsai's method [12]. We do not need accurate positioning of the calibration rig but rather sweeping of the calibration rig in the reconstructed region. In this way, the calibration effort can be reduced in the proposed projective reconstruction scheme.



(a) Basis view 1 (b) Basis view 2

Figure 5. Basis view images for defining projective grid space. On each image, some epipolar lines between the basis images are displayed. The intersection points of the epipolar lines are some of the projective grid.

4.2. Projective grid space

Figure 5 shows the basis view images on which some epipolar lines between the images are displayed. The intersection point of two epipolar lines in each image is a projective grid. In the actual implementation, 320 × 240 epipolar lines for all pixels in the basis view 1 are projected onto the basis view 2 for generating a projective grid space. Then 320 grid points are defined on every epipolar line by the horizontal pixel position in the basis view 2. In this way, 320 × 240 × 320 grid points are totally defined.

Even though we only consider the projective grid without any metric, it is better to make each projective grid closer to cubic shape in Euclidean space. In this sense, we choose the two basis view images because their optical axes are almost perpendicular. Choice of the basis view affects the reconstruction quality, but this issue will be studied in our future research.

4.3. Projective shape reconstruction

Although the projective axes, p, q, and r are not orthogonal to each other, we show the projective shape in the space of orthogonal p, q, and r space by in Figure 6. Even though the actual voxel is an arbitrarily shaped hexahedron, each voxel is assumed to be a cube in this representation. The top two images show the occupied voxels that are detected by the voxel coloring algorithm. Note that the displayed voxel size in those two images are 4 times larger in length than the voxel size actually used for projective shape reconstruction. The other images are rendered images from the

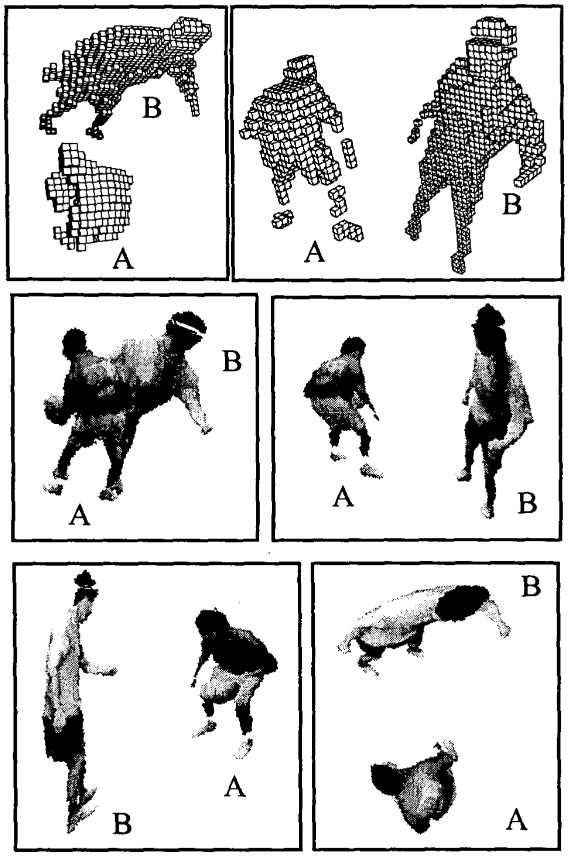


Figure 6. Reconstructed projective shape in the representation of orthographic grid space.

colored voxel representation of the reconstructed projective shape. Since the center of the basis view 1 is located behind person B, the real size of the voxel around the person A is larger than that around the person B. This is because that the real size of the voxel is increased in proportional to the distance from the center of the basis view 1. This fact affects the size of the object in this representation, i.e, person A is reduced in size.

4.4. Synthesizing virtual view images

We synthesized the virtual viewpoint images from the projective reconstruction. The reconstructed projective shape provides dense correspondence maps between arbitrary pairs of images, so that they can be used for view-interpolation [14] to synthesize intermediate images. Since the shape is reconstructed by using all 51 images, the correspondence map does not suffer by the occlusion, while typical stereo matching algorithms between a few images give wrong correspondences in the occluded region. Figure 7 shows the example of synthesized intermediate view images between two images.

The occluded region on the person in the background is completely recovered in the intermediate view images because projective shape is successfully recovered from all view images via projective voxel coloring as demonstrated in Figure 6. Since the proposed scheme of projective grid space is related to all view images by the set of fundamental matrices, projective shape structure of the scene can be successfully reconstructed by the use of all view images.

5. Conclusion

We proposed a new scheme for reconstructing shape in projective grid space for a large number of input images. 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.

For demonstrating the effectiveness of the proposed projective grid voxel space scheme, we apply the voxel coloring algorithm. The quality of the re-synthesized virtual view images from the projective shape demonstrates the effectiveness of our proposed scheme for projective reconstruction from large number of images. The view interpolated images are also generated by using the reconstructed projective shape. In the generated images, the occlusion region is reasonably interpolated because the 3D structure can be recovered.

Such regions of occlusion can not be interpolated by using the image-based method without recovery of 3D structure of the scene [1, 7], because those methods implicitly assume the consistent correspondence among all input images. The occlusion region can not be correctly interpolated without 3D structure information of the scene. Such 3D structure information must be recovered by integrating multiple view image information in the space that can represent the 3D structure by any means. The proposed projective grid space provides the 3D space in the framework of the projective geometry.

For demonstrating this feature of the proposed scheme, the voxel coloring method is applied to reconstruct projective shape from a number of images. Since the reconstructed projective shape provides the occlusion structure, the interpolated new view images can take account into the existing occlusion.

Acknowledgements The authors thank Peter Rander and Sundar Vedula for their help in collecting the multiple view images.

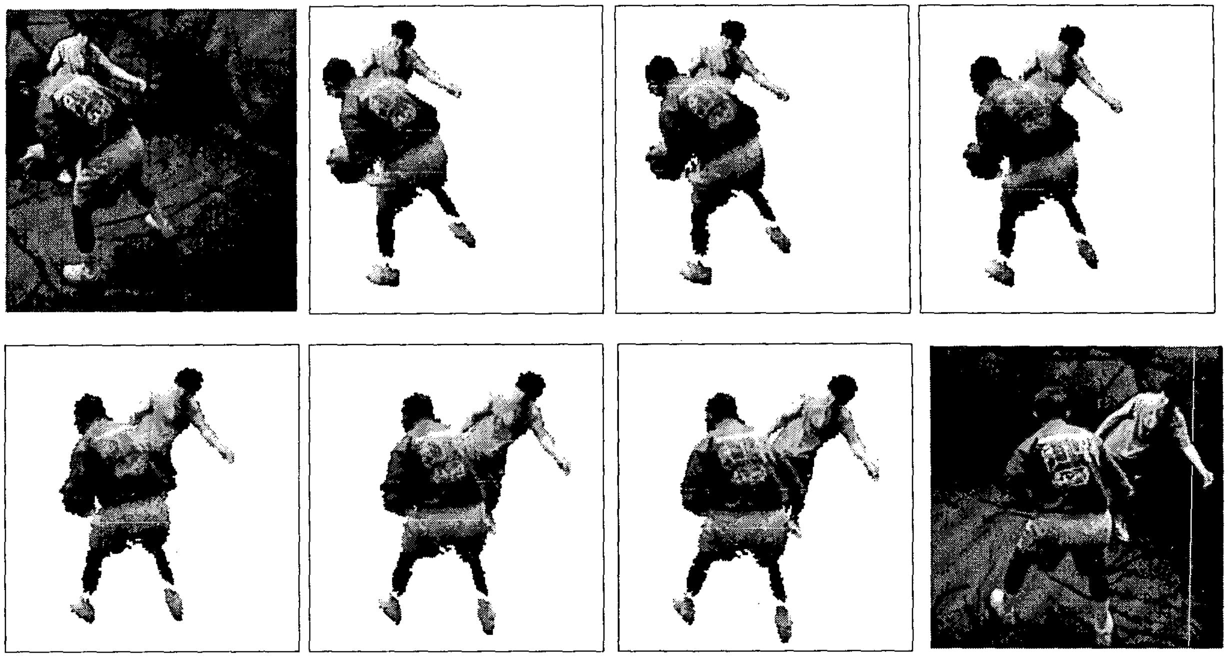


Figure 7. Synthesized intermediate view images between two images from correspondence map obtained by the projective reconstructed shape.

Shigeyuki Baba helped by generating the interpolated view images.

References

[1] S. Avidan and A. Shashua. "Novel View Synthesis by Cascading Trilinear Tensors". *IEEE Trans. Visualization and Computer Graphics*, 4(4):293–306, 1998.

[2] O. Faugeras. "Stratification of Three-Dimensional Vision: Projective, Affine, and Metric Representations". *Journal of OSA, A*, 12(3):465–484, 1995.

[3] D. Gavrila and L. Davis. "3-D Model Based Tracking of Humans in Action : Multi-View Approach". *Proc. Computer Vision and Pattern Recognition 96*, pages 73–80, 1996.

[4] R. Jain and K. Wakimoto. "Multiple Perspective Interactive Video". *Proceedings of IEEE Conference on Multimedia Systems*, 1995.

[5] T. Kanade, P. Rander, and P. Narayanan. "Virtualized Reality: Constructing Virtual Worlds from Real Scenes". *IEEE MultiMedia*, 4(1), 1997.

[6] K. Kutulakos and S. Seitz. "What Do N Photographs Tell Us About 3D Shape?". *Tech. Report of Computer Science Dept. of U. Rochester*, TR680, 1998.

[7] S. Laveau and O. Faugeras. "3-D Scene Representation as a Collection of Images". *Proc.Int'l. Conf. Pattern Recognition*, 1994.

[8] M. Pollefeys, R. Koch, and L. Gool. "Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters". *Proc. ICCV98*, pages 90–95, 1998.

[9] B. Prescott and G. McLean. "Line-Based Correction of Radial Lens Distortion". *Graphical Models and Image Processing*, 59(1):39–47, 1997.

[10] S. Seitz and C. Dyer. "Photorealistic Scene Reconstruction by Voxel Coloring". *Proc. Computer Vision and Pattern Recognition 97*, pages 1067–1073, 1997.

[11] C. Tomasi and T. Kanade. "Shape and Motion from Image Streams Under Orthography: A Factrization Method". *Int'l J. Computer Vision*, 2(9):137–154, 1992.

[12] R. Tsai. "A Versatile Camera Calibration 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):323–344, 1987.

[13] S. Vedula, P. Rander, H. Saito, and T. Kanade. "Modeling, Combining, and Rendering Dynamic Real-World Events From Image Sequences". *Proc. 4th Conf. Virtual Systems and MultiMedia*, 1:326–332, 1998.

[14] T. Werner, R. Hersch, and V. Hlavac. "Rendering Real-World Objects Using View Interpolation". *Proc. ICCV95*, pages 957–962, 1995.

[15] Z. Zhang. "Determining the Epipolar Geometry and its Uncertainly: A Review". *INRIA research report*, 2927, 1996.