3-D Object Reconstruction Using Spatially Extended Voxels and Multi-Hypothesis Voxel Coloring

Eckehard Steinbach and Bernd Girod
Information Systems Laboratory
Stanford University
Email: {steinb,bgirod}@stanford.edu

Peter Eisert and Arnulf Betz
Telecommunications Laboratory
University of Erlangen-Nuremberg
Email: {eisert,betz}@LNT.de

Abstract

In this paper we describe a voxel-based 3-D reconstruction technique from multiple calibrated camera views that makes explicit use of the finite size footprint of a voxel when projected into the image plane. We derive a class of computationally efficient axis-aligned volume traversal orders that ensure that a processed voxel cannot occlude previously processed voxels. For each view, one out of 79 different cases of volume traversal is identified depending on the relative position between camera and voxel volume. Views belonging to the same visibility class can be processed simultaneously. Our voxel coloring strategy is based on a color hypothesis test that ensures the consistency of the projected reconstruction with the original images. A surface voxel list is constantly updated during reconstruction ensuring that only a minimum number of voxels has to be processed. Experimental results that compare the reconstruction quality for voxels with and without spatial extent underline that it is worthwhile taking into account the exact footprint of the projected voxels.

1. Introduction

The automatic acquisition of photorealistic 3-D computer models from many camera views is a very active research area with applications in virtual reality and multimedia. The large body of work devoted to this problem can basically be divided in two different classes of algorithms. The first class of 3-D model acquisition techniques computes depth maps from two or more views of the object and then registers the depth maps into a single 3-D surface model. The depth map recovery often relies on sparse or dense matching of image points with subsequent 3-D structure estimation [1, 2, 3] or is supported by additional depth information from range sensors [4, 5]. Other approaches are based on volume intersection, and are often referred to as *shape-from-silhouette* algorithms [6, 7, 8]. The object shape is typically computed as the intersection of the outline cones which are back-projected from all available views of the object. This requires the reliable extraction of the object contour in all views which restricts the applicability to scenes where the object can be easily segmented from the background.

Recently, techniques for 3-D object reconstruction from multiple calibrated views have been proposed that combine the advantages of the two beforementioned classes. Using a volumetric representation of the objects, volume elements (voxels) are colored by comparing the pixel color of the projected voxel in all views where the voxel is visible [10],[11],[12]. These techniques avoid image correspondence problems by working in a discretized scene space whose elements are traversed in a fixed order during reconstruction for correct visibility handling. In [12] voxels are assumed to be 3-D points without spatial extent leading to low computational complexity at the expense of shining-through artifacts in the 3-D computer models. In [10] and [11] the footprint of the voxels is approximated by a square mask. During volume traversal, voxels in the same voxel layer are assumed not to occlude each other. The voxel coloring stategy in [10],[11] assigns mean color values to the non-transparent voxels while the voxel coloring in [12] is based on a multi-hypothesis test of the projected voxel model.

In this paper we combine the explicit consideration of the finite extent of the volume elements with the multi-hypothesis voxel coloring strategy in [12]. In comparison to [10] and [11] it is not assumed that voxels in the same voxel layer cannot occlude each other leading to exact visibility handling during volume traversal.

2. Voxel Projection

The projection of voxels without spatial extent into the image plane leads to a single point. In comparison, for extended voxels a small footprint in the image plane, potentially covering more than one pixel, is obtained. The exact footprint of a voxel has to consider its cubic shape. The projection leads to a convex 2-D polygon in the image plane, either 4-sided or 6-sided. In order to obtain the exact outline of the projected voxel we first project the eight corner

points into the image plane and then compute their closed convex hull. Once the voxel footprint has been computed,

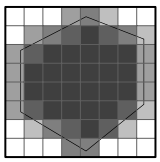

Figure 1. Voxel footprint and pixel contribution.

it can be determined which pixels fall inside the polygon. Fig. 1 shows an example where the contribution of a pixel, measured as the percentage of the area intersecting with the polygon, is represented as gray values. The intersecting area for the border pixels is determined using *polygon clipping* of the voxel polygon with respect to the rectangular pixel [9].

3. 3-D Object Reconstruction

The first step of the proposed reconstruction algorithm is to define a volume in the reference coordinate system that encloses the 3-D object to be reconstructed. The volume extensions are determined from the calibrated camera parameters and its surface represents a conservative bounding box of the object. The volume is discretized in all three dimensions leading to an array of voxels with associated color, where the position of each voxel in the 3-D space is defined by its indices (l, m, n). Initially, all voxels are transparent.

3.1. Hypothesis Generation

In the second step of the proposed reconstruction algorithm, color hypotheses are assigned to each voxel of the predefined volume. The k^{th} hypothesis H^k_{lmn} for a voxel V_{lmn} with voxel index (l, m, n) is

$$H^k_{lmn} = (R(X_i, Y_i), G(X_i, Y_i), B(X_i, Y_i)), \qquad (1)$$

with (X_i, Y_i) being the pixel position of the perspective projection of the voxel center (x_l, y_m, z_n) into the i^{th} camera view. R, G, and B are the three color components. The projection of the voxel center for view i is obtained as

$$X_i = -f_x \frac{x_{li}}{z_{ni}}, \quad Y_i = -f_y \frac{y_{mi}}{z_{ni}}, \qquad (2)$$

with

$$(x_{li}, y_{mi}, z_{ni})^T = \mathbf{R_i}(x_l, y_m, z_n)^T + \mathbf{T_i}. \qquad (3)$$

$\mathbf{R_i}$ and $\mathbf{T_i}$ are the object rotation and translation in view i with respect to the reference coordinate system. The parameters f_x and f_y describe the camera geometry and the scaling that relates pixel coordinates to world coordinates.

Hypothesis H^k_{lmn} is associated to voxel V_{lmn} if the projection of V_{lmn} into at least one other camera view $j \neq i$ leads to an absolute difference of the color channels

$$\left| \frac{R_i(X_i, Y_i)}{N_i(X_i, Y_i)} - \frac{R_j(X_j, Y_j)}{N_j(X_j, Y_j)} \right| + \left| \frac{G_i(X_i, Y_i)}{N_i(X_i, Y_i)} - \frac{G_j(X_j, Y_j)}{N_j(X_j, Y_j)} \right| + \left| \frac{B_i(X_i, Y_i)}{N_i(X_i, Y_i)} - \frac{B_j(X_j, Y_j)}{N_j(X_j, Y_j)} \right| < \Theta \qquad (4)$$

with

$$N_i(X, Y) = R_i(X, Y) + G_i(X, Y) + B_i(X, Y)\ . \qquad (5)$$

that is less than a predefined threshold Θ. The normalization of the color components in (4) is used to increase the robustness of the reconstruction algorithm with respect to varying illumination conditions. For hypothesis generation, the finite size of the voxel footprint is not exploited. Of course it could be, but experiments show that it is sufficient for this step to use the voxel center. This leads to a smaller number of color hypotheses that have to be stored. The voxel need not be visible in all views due to occlusions and it might not be visible in any view at all if it is inside the object. At this stage of the algorithm we do not know the geometry of the object and cannot decide whether a voxel is visible. We therefore have to remove those hypotheses of the overcomplete set that do not correspond to the correct color of the object's surface.

3.2. Consistency Check and Hypothesis Rejection

The color hypotheses in Section 3.1 are extracted from 2 or more consistent views but might contradict other views where the voxel is visible as well. We now start to refine our volume by iterating over all views. We start from the outermost voxel layer of the volume and remove voxels until the 3-D shape of the object is recovered. The decision which color hypotheses to eliminate for each voxel considers the finite area of the projected voxel. Let F_o be the area of the projection of a voxel obtained without considering occlusion and F_n the area which is obtained when considering occlusion by other voxels. The color test criterion in (4) is modified as follows

- if the projected area F_n is small, the consistency test is not performed since either the voxel is very far from the camera or heavily occluded by other voxels.
- the threshold Θ in (4) is modified according to

$$\Theta_{new} = (\frac{3}{2} - \frac{1}{2}\frac{F_n}{F_o})\Theta \qquad (6)$$

which leads to an increase of the threshold value for heavily occluded voxels of about 50 %.

Color hypotheses have to be generated only for those voxels that become visible during reconstruction. We keep track of potentially visible voxels by storing and updating a surface voxel list. Fig. 2 illustrates the update of the surface list after removal of one surface voxel. Only those voxels that

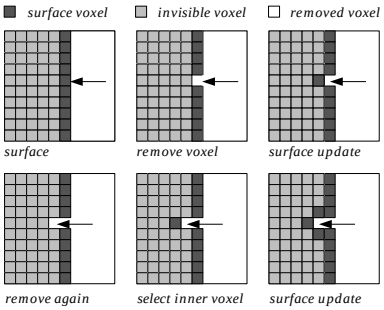

Figure 2. Surface list update after voxel removal.

are visible in a particular view have to be tested for color hypothesis consistency. This requires the determination of their visibility. Since the object model consists of a structured set of voxels, a traversal order similar to the one described in [12] can be derived. Processing the surface voxels according to this volume traversal order ensures correct occlusion handling for the particular view.

We now derive the different volume traversal order classes that can be identified when considering all possible mutual voxel occlusion. Enlarging the bounding planes of the object volume in all coordinate directions divides the 3-D space outside the volume into 26 regions. These different regions are classified into three cases:

1. The closest element of the volume is a corner voxel.
2. The closest element of the volume is an edge voxel.
3. The closest element of the volume is a face voxel.

Fig. 3 illustrates these different cases. The simplest traversal order is obtained for case 1 as can be seen from Fig. 4. The camera is represented by a sphere and the arrow points to the closest voxel. The voxel indices l, m, and n run over all defined values. The closest corner of the bounding box

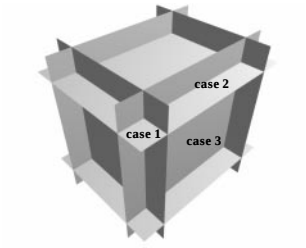

Figure 3. Partitioning of the 3-D space into three different cases.

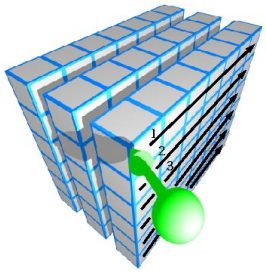

Figure 4. Volume traversal order for case 1.

determines, whether l, m or n have to be incremented or decremented. In which order this has to be done depends on the relative position of the camera with respect to the corner. Fig. 5 illustrates the selection of a particular order in 2-D. A total of $3! = 6$ different permutations of the

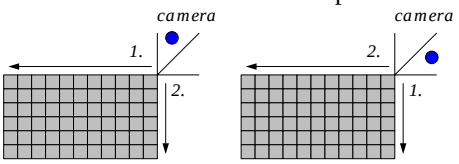

Figure 5. 2-D illustration of the index permutation selection.

voxel indices can be identified. Since these permutations apply for all 8 corners of the bounding volume we obtain $8 \times 6 = 48$ different traversal orders. These 48 cases correspond to the traversal orders in [12]. For case 2 in Fig. 3, the closest voxel is found on an edge of the bounding volume. Here, $2! = 2$ different index permutations exist. Since these permutations apply for all 12 edges, we obtain $2 \times 12 = 24$ different traversal orders. The traversal order for this case is illustrated in Fig. 6. Case 3 in Fig. 3 leads to the traversal

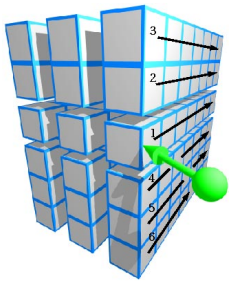

Figure 6. Volume traversal order for case 2.

order shown in Fig. 7. For a camera that lies in the interior of the voxel volume (e.g., when reconstructing a room), a different traversal order has to be used. This is illustrated in Fig. 8 for one slice of the object volume. The total number of different traversal orders can now be summarized. For the 8 corners we obtain 48 different cases. The twelve edges of the volume add 24 cases and the 6 faces another 6 cases. Including the traversal order for the camera in the interior of the volume we obtain $48 + 12 * 2 + 6 + 1 = 79$ different volume traversal order cases.

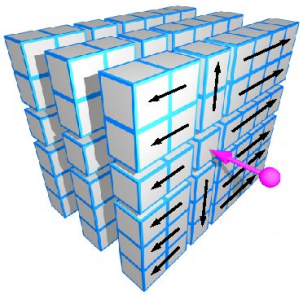

Figure 7. Volume traversal order for case 3.

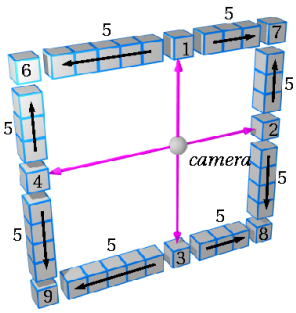

Figure 8. Volume traversal order for a camera that falls inside the voxel volume.

4. Experimental Results

The following experiment illustrates the reconstruction quality that is obtained when considering extended voxels and the exact occlusion handling presented in this paper. A 24 view sequence (352×288 pixels) of a plant is used. Fig. 9 shows two different original views of the sequence. Fig. 10 compares the reconstruction result for new views of the object for the case of point voxels (top) and extended voxels (bottom). New viewing positions are selected that are not part of the original set of views. It can be seen that the reconstruction quality considerably improved for extended voxels. The main reason for this is that for point voxels shining-through artifacts in rendered views from new viewing positions occur.

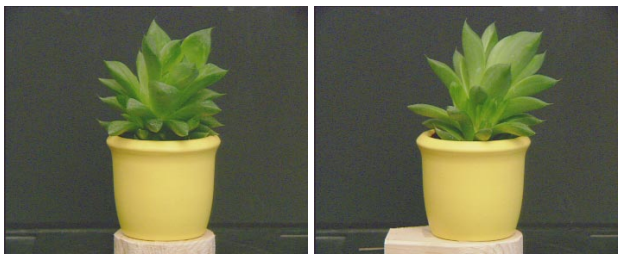

Figure 9. Two original views of the *plant* sequence.

Figure 10. Reconstructed views of the *plant*. The lower images show the results obtained when considering extended voxels. The upper images show the results for point voxels.

References

[1] P. Beardsley, P. Torr, and A. Zisserman, "3D Model Acquisition from Extended Image Sequences," *Proc. ECCV '96*, pp. 683-695, Cambridge, UK, 1996.

[2] R. Koch, M. Pollefeys, and L. Van Gool, "Multi Viewpoint Stereo from Uncalibrated Sequences," *Proc. ECCV '98*, pp. 55-71, Freiburg, Germany, 1998.

[3] M. Pollefeys, R. Koch, M. Vergauwen, and L. Van Gool, "Flexible Acquisition of 3D Structure from Motion," *Proc. Tenth IMDSP Workshop '98*, pp. 195-198, Austria, 1998.

[4] B. C. Vemuri, J. K. Aggarwal, "3-D Model Construction from Multiple Views Using Range and Intensity Data," *Proc. CVPR '86*, pp. 435-437, Miami Beach, 1986.

[5] B. Curless and M. Levoy, "A Volumetric Method for Building Complex Models from Range Images," *ACM Siggraph '96*, pp. 303-312, June 1993.

[6] E. Boyer, "Object models from contour sequences," *Proc. ECCV '96*, pp. 109-118, Cambridge, UK, 1996.

[7] W. Niem, J. Wingbermühle, "Automatic Reconstruction of 3D Objects Using a Mobile Monoscopic Camera", *Proc. International Conference on Recent Advances in 3D Imaging and Modelling*, Ottawa, Canada, May 1997.

[8] R. Szeliski, "Rapid octree construction from image sequences," *CVGIP 93*, pp. 23-32, July 1993.

[9] J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes, "Computer Graphics: Principles and Practice," *Addison-Wesley, Second Edition*, 1990.

[10] S.M. Seitz and C.R. Dyer, "Photorealistic Scene Reconstruction by Voxel Coloring," *Proc. CVPR '97*, pp. 1067-1073, Puerto Rico, 1997.

[11] K.N. Kutulakos and S. Seitz, "A Theory of Shape by Space Carving," Proc. ICCV '99, pp. 307-314, 1999.

[12] P. Eisert, E. Steinbach, and B. Girod, "Multi-hypothesis, Volumetric Reconstruction of 3-D Objects from Multiple Calibrated Camera Views," *Proc. ICASSP '99*, pp. 3509-3512, Phoenix, March 1999.