

3-D RECONSTRUCTION OF REAL-WORLD OBJECTS USING EXTENDED VOXELS

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 present a voxel-based 3-D reconstruction technique that computes a set of non-transparent object surface voxels from a given set of calibrated camera views. We show that the quality of the reconstruction strongly depends on the accuracy of the computed voxel projection in the image plane and discuss different approximations of the exact projection. The most simple and computationally least demanding approximation is obtained when projecting point voxels, i.e., voxels without spatial extent. However, correct occlusion handling is not possible for point voxel volumes leading to reconstruction artifacts. The most accurate projection is obtained by computing the exact outline of the projected voxels. This projection is computationally most demanding but allows correct occlusion handling during reconstruction. Experimental results that compare the reconstruction quality for point and exact voxel projection show that it is worthwhile computing the exact image plane 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]. Here, 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 aforementioned 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 noticeable artifacts in the 3-D computer models. In [10] and [11] the footprint of the voxels is approximated by a square mask.

In this paper we compare the reconstruction results obtained for point voxel projection and exact voxel footprint computation. The visibility handling for voxels with spatial extent differs from [10] and [11] in that it is not assumed that voxels in the same voxel layer cannot occlude each other, leading to exact visibility handling during volume traversal [13].

2. VOXEL PROJECTION

The projection of voxels without spatial extent into the image plane leads to a single point. This is the most simple and computationally least demanding projection. Here, only the center of the voxel is projected and the image plane pixel hit by the projected voxel center is assumed to be entirely covered by this voxel. This simple voxel projection leads to the following two problems:

1. Due to the rounding operation that is used to identify the image plane pixel that corresponds to a projected voxel, it can happen that more than one surface voxel projects to the same pixel. The first voxel projected to this pixel is assumed to occlude all following voxels projected into the same pixel. As a result, some voxels that are actually partially visible are classified as being occluded.
2. In those views where the voxel volume resolution and the image plane resolution do not match, holes in the projection occur in between the projected voxel center points of a closed surface. As a result, voxels become visible in this view which are not part of the object surface turned to the camera but lie behind the surface voxels and should be classified as occluded. As will be shown later, this leads to artifacts in the 3-D reconstruction.

In order to illustrate the first of the aforementioned problems of point voxel projection, we use the view of a cassette in Fig. 1 and assign the pixel colors to those voxels that after projection are determined as being unoccluded in this view. The result is shown in Fig. 2 where white pixels correspond to voxels that are classified as being occluded although they lie on the surface turned to the camera. Increasing the resolution of the projection buffer, i.e., upsampling the camera views reduces this problem, but at the same time increases the effect of the second problem now causing even bigger holes that are filled by voxels behind the object surface misclassified as being unoccluded. The second problem can be reduced by increasing the resolution of the voxel volume. This,



Fig. 1. Original view of the *Cassette* test sequence.

however, rapidly leads to a prohibitive number of voxels to store and process and in turn amplifies the first problem.



Fig. 2. Illustration of point voxel based occlusion determination. White voxels are misclassified as being occluded although they are actually partially visible. Please note that parts of the background are visible because the voxel volume employed is larger than the extent of the cassette.

In comparison to point voxel projection, for extended voxels a small footprint in the image plane, potentially covering more than one pixel, is obtained. One computationally attractive technique to approximate the exact footprint is to use the bounding rectangle that can be computed from the projection of the eight corner points of a voxel. All pixels in the projection buffer that are covered by the bounding rectangle are assigned to the corresponding voxel. This is illustrated in Fig. 3. This approximation solves the

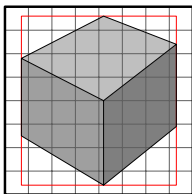


Fig. 3. Bounding rectangle approximation of the voxel projection.

problem of having holes in the projection buffer since the bounding box is always bigger than the exact footprint of the voxel. As

can be seen from Fig. 3, especially in the corners of the bounding rectangle the approximation of the voxel footprint can be poor. Since the bounding rectangle is bigger than the exact footprint it can again happen that partially visible surface voxels are classified as being occluded. Fig. 4 shows the same experiment as in Fig. 2 for the bounding rectangle approximation. The number of misclassified surface voxels is considerably reduced but some artifacts are still visible. Only an exact computation of the voxel footprint can solve this problem. The exact footprint of a voxel



Fig. 4. White pixels correspond to voxels that are misclassified as being occluded although they are actually partially visible as a result of the bounding rectangle approximation of the exact voxel footprint.

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,

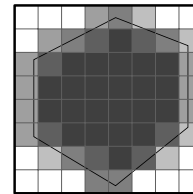


Fig. 5. Voxel footprint and pixel contribution.

it can be determined which pixels fall inside the polygon. Fig. 5 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].

In the following we briefly describe the reconstruction algorithm that is based on the multi-hypothesis voxel coloring strategy in [12]. For those parts depending on the accuracy of the voxel projection we explicitly mention the differences for point voxel projection and exact footprint computation.

3. 3-D OBJECT RECONSTRUCTION

The first step of the 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 reconstruction algorithm, color hypotheses are assigned to the voxels of the predefined volume. The k^{th} hypothesis H_{lmn}^k for a voxel V_{lmn} with voxel index (l, m, n) is

$$H_{lmn}^k = (R(X_i, Y_i), G(X_i, Y_i), B(X_i, Y_i)), \quad (1)$$

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

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

with

$$(x_{li}, y_{mi}, z_{ni})^T = \mathbf{R}_i(x_l, y_m, z_n)^T + \mathbf{T}_i. \quad (3)$$

\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_{lmn}^k 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 that is less than a predefined threshold Θ

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

with

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

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 footprint for exact voxel projection 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 occlusion by other voxels. 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 that do not correspond to the correct color of the object's surface.

3.2. Consistency Check and Hypothesis Removal

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. In case of exact footprint computation, the decision of 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. For extended 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.
- Else, the threshold Θ in (4) is modified according to

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

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

For point voxel projection the area measure above can obviously not be used and we reuse the color test criterion in (4), which corresponds to $F_o = F_n$.

Only those voxels that 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 fixed number of different volume traversal orders can be identified depending on the relative position of the camera and the object volume. Processing the voxels in this order ensures correct occlusion handling for the particular view. Views leading to the same traversal order can be processed simultaneously. For cameras that are outside the voxel volume, it is sufficient to consider 48 different traversal orders [12] for point voxel projection, while the exact footprint projection requires the identification of one out of 78 different traversal orders [13].

In order to decrease the computational complexity of the hypothesis generation step in Section 3.1, we generate color hypotheses only for those voxels that potentially become visible during reconstruction. We keep track of potentially visible voxels by storing and updating a surface voxel list. Fig. 6 illustrates the update of the surface list after removal of one surface voxel.

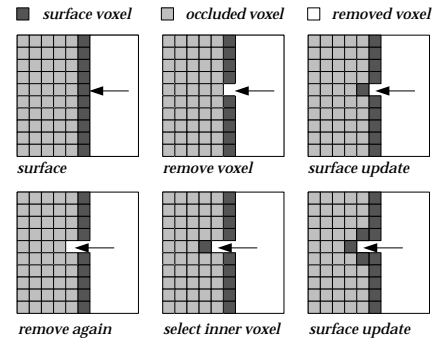


Fig. 6. Surface list update after voxel removal.

4. EXPERIMENTAL RESULTS

The following experiment compares the reconstruction quality that is obtained when considering point voxel or exact voxel projection. A 24 view sequence (352×288 pixels) of a plant is used. Fig. 7 shows two different original views of the sequence. Fig. 8 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 improves for extended voxels. The main reason for this is the inherent deficiency in occlusion handling for point voxel projection described in Section 2. Some voxels of the surface that is turned away from the camera become visible as a result of the second problem described in Section 2. These voxels shine through the object surface turned towards the camera since the spatial extent of the voxels is not taken into account. This produces the shining-through artifacts observed in the upper images of Fig. 8. Additionally, for point voxels, the color consistency test is not performed for all views where the voxel is visible since voxels are sometimes misclassified as being occluded although they are partially visible. Therefore, the coloring strategy leads to different colors for some voxels in comparison to the color assigned when computing the exact footprint of the voxel. These problems are eliminated when considering the exact footprint of the projected voxels. However, using point voxels decreases the computational complexity in our implementation by an order of magnitude.



Fig. 7. Two original views of the *Plant* sequence.

5. CONCLUSIONS

In this paper we discuss voxel-based 3-D reconstruction using either computationally efficient point voxel projection or computationally demanding exact voxel footprint determination. We show that the inherent deficiency in occlusion handling for point voxel projection leads to noticeable artifacts in the reconstruction that can be eliminated using exact voxel projection. Our experimental results suggest that it is worthwhile considering spatially extended voxels despite the higher computational complexity involved.

6. 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.



Fig. 8. Reconstructed views. The upper images show the results obtained when considering point voxels. The lower images show the results for extended voxels. New viewing positions are selected that are not part of the set of original views.

- [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.M. 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.
- [13] E. Steinbach, B. Girod, P. Eisert, and A. Betz, "3-D Object Reconstruction Using Spatially Extended Voxels and Multi-Hypothesis Voxel Coloring", *Proc. ICPR '00*, Barcelona, Spain, September 2000.