In Proc. Computer Vision and Pattern Recognition Conf., pp. 1067-1073, 1997.

Photorealistic Scene Reconstruction by Voxel Coloring

Steven M. Seitz   Charles R. Dyer

Department of Computer Sciences
University of Wisconsin, Madison, WI 53706
E-mail: {seitz,dyer}@cs.wisc.edu
WWW: http://www.cs.wisc.edu/computer-vision

Abstract

A novel scene reconstruction technique is presented, different from previous approaches in its ability to cope with large changes in visibility and its modeling of intrinsic scene color and texture information. The method avoids image correspondence problems by working in a discretized scene space whose voxels are traversed in a fixed visibility ordering. This strategy takes full account of occlusions and allows the input cameras to be far apart and widely distributed about the environment. The algorithm identifies a special set of invariant voxels which together form a spatial and photometric reconstruction of the scene, fully consistent with the input images. The approach is evaluated with images from both inward- and outward-facing cameras.

1 Introduction

We consider the problem of acquiring photorealistic 3D models of real environments from widely distributed viewpoints. This problem has sparked recent interest in the computer vision community [1, 2, 3, 4, 5] as a result of new applications in telepresence, virtual walkthroughs, and other graphics-oriented problems that require realistic textured object models.

We use the term *photorealism* to describe 3D reconstructions of real scenes whose reprojections contain sufficient color and texture information to accurately reproduce images of the scene from a broad range of target viewpoints. To ensure accurate reprojections, the input images should be representative, i.e., sparsely distributed throughout the target range of viewpoints. Accordingly, we propose two criteria that a photorealistic reconstruction technique should satisfy:

- Photo Integrity: The reprojected model should accurately reproduce the input images, preserving color, texture and pixel resolution
- Broad Viewpoint Coverage: Reprojections should be accurate over a large range of target viewpoints. This requires that the *input images* are widely distributed about the environment

The photorealistic scene reconstruction problem, as presently formulated, raises a number of unique challenges that push the limits of existing reconstruction techniques. Photo integrity requires that the reconstruction be dense and sufficiently accurate to reproduce the original images. This criterion poses a problem for existing feature- and contour-based techniques that do not provide dense shape estimates. While these techniques can produce texture-mapped models [1, 3, 4], accuracy is ensured only in places where features have been detected. The second criterion means that the input views may be far apart and contain significant occlusions. While some stereo methods [6, 7] can cope with limited occlusions, handling visibility changes of greater magnitude appears to be beyond the state of the art.

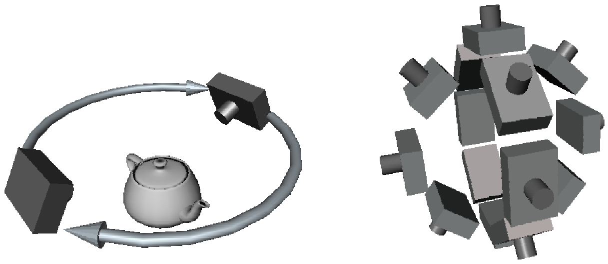
Instead of approaching this problem as one of shape reconstruction, we instead formulate a *color reconstruction* problem, in which the goal is an assignment of colors (radiances) to points in an (unknown) approximately Lambertian scene. As a solution, we present a *voxel coloring* technique that traverses a discretized 3D space in "depth-order" to identify voxels that have a unique coloring, constant across all possible interpretations of the scene. This approach has several advantages over existing stereo and structure-from-motion approaches to scene reconstruction. First, occlusions are explicitly modeled and accounted for. Second, the cameras can be positioned far apart without degrading accuracy or run-time. Third, the technique integrates numerous images to yield dense reconstructions that are accurate over a wide range of target viewpoints.

The voxel coloring algorithm presented here works by discretizing scene space into a set of voxels that is traversed and colored in a special order. In this respect, the method is similar to Collins' Space-Sweep approach [8]

The support of DARPA and the National Science Foundation under Grant No. IRI-9530985 is gratefully acknowledged.

which performs an analogous scene traversal. However, the Space-Sweep algorithm does not provide a solution to the occlusion problem, a primary contribution of this paper. Katayama et al. [9] described a related method in which images are matched by detecting lines through slices of an epipolar volume, noting that occlusions could be explained by labeling lines in order of increasing slope. Our voxel traversal strategy yields a similar scene-space ordering but is not restricted to linear camera paths. However, their algorithm used a reference image, thereby ignoring points that are occluded in the reference image but visible elsewhere. Also related are recently developed panoramic stereo [10, 11] algorithms which avoid field of view problems by matching 360° panoramic images directly. Panoramic reconstructions can also be achieved using our approach, but without the need to build panoramic images (see Figs. 1 (b) and 4).

The remainder of the paper is organized as follows: Section 2 formulates and solves the voxel coloring problem, and describes its relations to shape reconstruction. Section 3 presents an efficient algorithm for computing the voxel coloring from a set of images, discussing complexity and related issues. Section 4 describes experiments on real and synthetic image sequences.



(a) **(b)**

Figure 1: Compatible Camera Configurations. Both of the following camera configurations satisfy the ordinal visibility constraint: (a) a downward-facing camera moved 360 degrees around an object, and (b) a rig of outward-facing cameras distributed around a sphere.

2 Voxel Coloring

The voxel coloring problem is to assign colors (radiances) to voxels (points) in a 3D volume so as to maximize *photo integrity* with a set of input images. That is, rendering the colored voxels from each input viewpoint should reproduce the original image as closely as possible. In order to solve this coloring problem we must consider the following two issues:

- Uniqueness: Multiple voxel colorings can be consistent with a given set of images. How can the problem be well-defined?
- Computation: How can a voxel coloring be computed from a set of input images?

This section formalizes the voxel coloring problem and explores geometrical constraints on camera placement that enable an efficient solution. In order to address the uniqueness and computation issues, we introduce a novel visibility constraint that greatly simplifies the analysis. This *ordinal visibility constraint* enables the identification of certain *invariant* voxels whose colorings are uniquely defined. In addition, the constraint defines a depth-ordering of voxels by which the coloring can be computed in a single pass through the scene volume.

2.1 Notation

We assume that both the scene and lighting are stationary and that surfaces are approximately Lambertian. Under these conditions, the radiance at each point is isotropic and can therefore be described by a scalar value which we call *color*. We also use the term color to refer to the irradiance of an image pixel. The term's meaning should be clear by context.

A 3D scene $\mathcal{S}$ is represented as a finite[1] set of opaque voxels (volume elements), each of which occupies a finite homogeneous scene volume and has a fixed color. We denote the set of all voxels with the symbol $\mathcal{V}$. An image is specified by the set $\mathcal{I}$ of all its pixels. For now, assume that pixels are infinitesimally small.

Given an image pixel p and scene $\mathcal{S}$, we refer to the voxel $V \in \mathcal{S}$ that is visible and projects to p by $V = \mathcal{S}(p)$. The color of an image pixel $p \in \mathcal{I}$ is given by $color(p, \mathcal{I})$ and of a voxel V by $color(V, \mathcal{S})$. A scene $\mathcal{S}$ is said to be *complete* with respect to a set of images if, for every image $\mathcal{I}$ and every pixel $p \in \mathcal{I}$, there exists a voxel $V \in \mathcal{S}$ such that $V = \mathcal{S}(p)$. A complete scene is said to be *consistent* with a set of images if, for every image $\mathcal{I}$ and every pixel $p \in \mathcal{I}$,

$$color(p, \mathcal{I}) = color(\mathcal{S}(p), \mathcal{S}) \quad (1)$$

2.2 The Ordinal Visibility Constraint

For concreteness, a pinhole perspective projection model is assumed. To simplify the analysis, we introduce a constraint on the positions of the cameras relative to the scene. This constraint simplifies the task of resolving visibility relationships by establishing a fixed depth-order enumeration of points in the scene.

Let P and Q be scene points and $\mathcal{I}$ be an image from a camera centered at C. We say P *occludes* Q if P lies on the line segment $\overline{CQ}$. We require that the input cameras are positioned so as to satisfy the following constraint:

[1] It is assumed that the visible scene is spatially bounded.

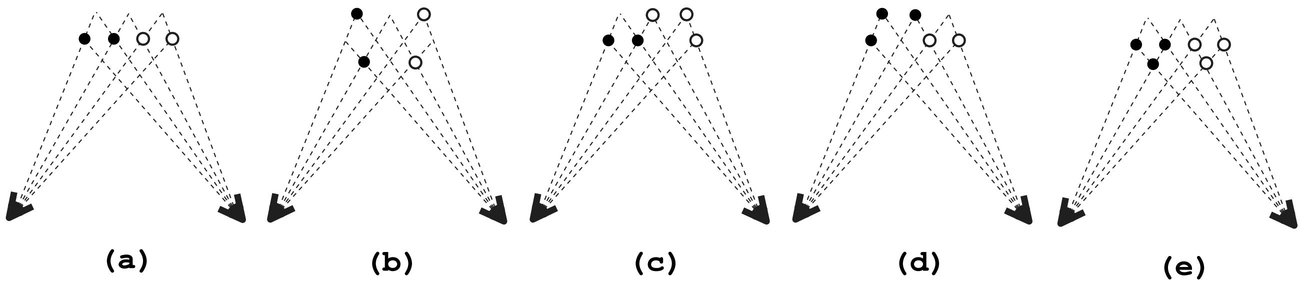


Figure 2: Ambiguity in Scene Reconstruction. All five scenes are indistinguishable from these two viewpoints. Shape ambiguity: scenes (a) and (b) have no points in common—no hard points exist. Color ambiguity: (c) and (d) share a point that has a different color assignment in the two scenes. Color invariants: each point in (e) has the same color in every consistent scene in which it is contained. These six points constitute the *voxel coloring* for these two views.

> **Ordinal visibility constraint:** There exists a norm $\|\cdot\|$ such that for all scene points P and Q, and input images $\mathcal{I}$, P occludes Q in $\mathcal{I}$ only if $\|P\| < \|Q\|$.

We call such a norm *occlusion-compatible*. For some camera configurations, it is not possible to define an occlusion-compatible norm. However, a norm *does* exist for a broad range of practical configurations. For instance, suppose the cameras are distributed on a plane and the scene is entirely below that plane, as shown in Fig. 1(a). For every such viewpoint, the relative visibility of any two scene points depends entirely on which point is closer to the plane, so we may define $\|\cdot\|$ to be distance to the plane. More generally, the ordinal visibility constraint is satisfied whenever **no scene point is contained within the convex hull $\mathcal{C}$ of the camera centers**. Here we use the occlusion-compatible norm $\|P\|_{\mathcal{C}}$, defined to be the Euclidean distance from P to $\mathcal{C}$. For convenience, $\mathcal{C}$ is referred to as the *camera volume*. Fig. 1 shows two useful camera configurations that satisfy this constraint. Fig. 1(a) depicts an inward-facing overhead camera rotating 360° around an object. Ordinal visibility is satisfied provided the camera is positioned slightly above the object. The constraint also enables panoramic reconstructions from outward-facing cameras, as in Fig. 1(b).

2.3 Color Invariance

The ordinal visibility constraint provides a depth-ordering of points in the scene. We now describe how this ordering can be used in scene reconstruction. Scene reconstruction is complicated by the fact that a set of images can be consistent with more than one rigid scene. Determining a scene's spatial occupancy is therefore an ill-posed task because a voxel contained in one consistent scene may not be contained in another. (see Fig. 2(a),(b)). Alternatively, a voxel may be part of two consistent scenes, but have different colors in each (Fig. 2(c),(d)).

Given a multiplicity of solutions to the color reconstruction problem, the only way to recover intrinsic scene information is through *invariants*— properties that are satisfied by *every* consistent scene. For instance, consider the set of voxels that are contained in every consistent scene. Laurentini [12] described how these invariants, called *hard points*, could be recovered by volume intersection from binary images. Hard points are useful in that they provide absolute information about the true scene. However, such points are relatively rare; some images may yield none (see, for example, Fig. 2). In this section we describe a more frequently occurring type of invariant relating to color rather than shape.

> A voxel V is a **color invariant** with respect to a set of images if, for every pair of scenes $\mathcal{S}$ and $\mathcal{S}'$ consistent with the images, $V \in \mathcal{S} \cap \mathcal{S}'$ implies $color(V, \mathcal{S}) = color(V, \mathcal{S}')$

Unlike shape invariance, color invariance does not require that a point be contained in every consistent scene. As a result, color invariants are more prevalent than hard points. In particular, any set of images satisfying the ordinal visibility constraint yields enough color invariants to form a complete scene reconstruction, as will be shown.

Let $\mathcal{I}_1, \ldots, \mathcal{I}_m$ be a set of images. For a given image point $p \in \mathcal{I}_j$ define V_p to be the voxel in $\{\mathcal{S}(p) \mid \mathcal{S}\ consistent\}$ that is closest to the camera volume. We claim that V_p is a color invariant. To establish this, observe that $V_p \in \mathcal{S}$ implies $V_p = \mathcal{S}(p)$, for if $V_p \neq \mathcal{S}(p)$, $\mathcal{S}(p)$ must be closer to the camera volume, which is impossible by the definition of V_p. It follows from Eq. (1) that V_p has the same color in every consistent scene; V_p is a color invariant.

> The **voxel coloring** of an image set $\mathcal{I}_1, \ldots, \mathcal{I}_m$ is defined to be:
> $\overline{\mathcal{S}} = \{V_p \mid p \in \mathcal{I}_i,\ 1 \leq i \leq m\}$

Fig. 2(e) shows the voxel coloring for the pair of images in Fig. 2. These six points have a unique color interpretation, constant in every consistent scene. They also comprise the closest consistent scene to the cameras in the following sense—every point in each consistent scene is either contained in the voxel coloring or is occluded by points in the voxel coloring. An interesting consequence of this closeness bias is that neighboring image pixels of the same color produce cusps in the voxel coloring, i.e., protrusions toward the camera volume. This phenomenon is clearly shown in Fig. 2(e), where the white and black points form two separate cusps. Also, observe that the voxel coloring is not a minimal reconstruction; removing the two closest points in Fig. 2(e) still leaves a consistent scene.

2.4 Computing the Voxel Coloring

In this section we describe how to compute the voxel coloring from a set of images that satisfy the ordinal visibility constraint. In addition it will be shown that the set of voxels contained in the voxel coloring form a complete scene reconstruction that is consistent with the input images.

The voxel coloring is computed one voxel at a time in an order that ensures agreement with the images at each step, guaranteeing that all reconstructed voxels satisfy Eq. (1). To demonstrate that voxel colorings form consistent scenes, we also have to show that they are complete, i.e., they account for every image pixel as defined in Section 2.1.

In order to make sure that the construction is incrementally consistent, i.e., agrees with the images at each step, we need to introduce a weaker form of consistency that applies to incomplete voxel sets. Accordingly, we say that a set of voxels with color assignments is *voxel-consistent* if its projection agrees fully with the subset of every input image that it overlaps. More formally, a set $\mathcal{S}$ is said to be voxel-consistent with images $\mathcal{I}_1, \ldots, \mathcal{I}_m$ if for every voxel $V \in \mathcal{S}$ and image pixels $p \in \mathcal{I}_i$ and $q \in \mathcal{I}_j$, $V = \mathcal{S}(p) = \mathcal{S}(q)$ implies $color(p, \mathcal{I}_i) = color(q, \mathcal{I}_j) = color(V, \mathcal{S})$. For notational convenience, define $\mathcal{S}_V$ to be the set of all voxels in $\mathcal{S}$ that are closer than V to the camera volume. Scene consistency and voxel consistency are related by the following properties:

1. If $\mathcal{S}$ is a consistent scene then $\{V\} \cup \mathcal{S}_V$ is a voxel-consistent set for every $V \in \mathcal{S}$.

2. Suppose $\mathcal{S}$ is complete and, for each point $V \in \mathcal{S}$, $V \cup \mathcal{S}_V$ is voxel-consistent. Then $\mathcal{S}$ is a consistent scene.

A consistent scene may be created using the second property by incrementally moving farther from the camera volume and adding voxels to the current set that maintain voxel-consistency. To formalize this idea, we define the following partition of 3D space into voxel layers of uniform distance from the camera volume:

$$\mathcal{V}_{\mathcal{C}}^{d} = \{V \mid \|V\|_{\mathcal{C}} = d\} \tag{2}$$

$$\mathcal{V} = \bigcup_{i=1}^{r} \mathcal{V}_{\mathcal{C}}^{d_i} \tag{3}$$

where $d_1, \ldots, d_r$ is an increasing sequence of numbers.

The voxel coloring is computed inductively as follows:

$$\begin{aligned}
\mathcal{SP}_1 &= \{V \mid V \in \mathcal{V}_{d_1}, \{V\} \text{ voxel-consistent}\} \\
\mathcal{SP}_k &= \{V \mid V \in \mathcal{V}_{d_k}, \{V\} \cup \mathcal{SP}_{k-1} \text{ voxel-consistent}\} \\
\mathcal{SP} &= \{V \mid V = \mathcal{SP}_r(p) \text{ for some pixel } p\}
\end{aligned}$$

We claim $\mathcal{SP} = \overline{\mathcal{S}}$. To prove this, first define $\overline{\mathcal{S}}_i = \{V \mid V \in \overline{\mathcal{S}}, \|V\|_{\mathcal{C}} \leq d_i\}$. $\overline{\mathcal{S}}_1 \subseteq \mathcal{SP}_1$ by the first consistency property. Inductively, assume that $\overline{\mathcal{S}}_{k-1} \subseteq \mathcal{SP}_{k-1}$ and let $V \in \overline{\mathcal{S}}_k$. By the first consistency property, $\{V\} \cup \overline{\mathcal{S}}_{k-1}$ is voxel-consistent, implying that $\{V\} \cup \mathcal{SP}_{k-1}$ is also voxel-consistent, because the second set includes the first and $\mathcal{SP}_{k-1}$ is itself voxel-consistent. It follows that $\overline{\mathcal{S}} \subseteq \mathcal{SP}_r$. Note also that $\mathcal{SP}_r$ is complete, since one of its subsets is complete, and hence consistent by the second consistency property. $\mathcal{SP}$ contains all the voxels in $\mathcal{SP}_r$ that are visible in any image, and is therefore consistent as well. Therefore $\mathcal{SP}$ is a consistent scene such that for each pixel p, $\mathcal{SP}(p)$ is at least as close to $\mathcal{C}$ as $\overline{\mathcal{S}}(p)$. Hence $\mathcal{SP} = \overline{\mathcal{S}}$. □

In summary, the following properties of voxel colorings have been shown:

- $\overline{\mathcal{S}}$ is a consistent scene
- Every voxel in $\overline{\mathcal{S}}$ is a color invariant
- $\overline{\mathcal{S}}$ is computable from any set of images satisfying the ordinal visibility constraint

3 Reconstruction by Voxel Coloring

In this section we present a voxel coloring algorithm for reconstructing a scene from a set of calibrated images. The algorithm closely follows the voxel coloring construction outlined in Section 2.4, adapted to account for image discretization and noise. As before, it is assumed that 3D space has been partitioned into a series of voxel layers $\mathcal{V}_{\mathcal{C}}^{d_1}, \ldots, \mathcal{V}_{\mathcal{C}}^{d_r}$ increasing in distance from the camera volume. The images $\mathcal{I}_1, \ldots, \mathcal{I}_m$ are assumed to be discretized into finite non-overlapping pixels. The cameras are assumed to satisfy the ordinal visibility constraint, i.e., no scene point lies within the camera volume.

If a voxel V is not fully occluded in image $\mathcal{I}_j$, its projection will overlap a nonempty set of image pixels, π_j. Without noise or quantization effects, a consistent voxel should project to a set of pixels with equal color values. In the presence of these effects, we evaluate the correlation of the pixel colors to measure the likelihood of voxel consistency. Let s be the standard deviation and n the cardinality of $\bigcup_{j=1}^{m} \pi_j$. Suppose the sensor error (accuracy of irradiance measurement) is approximately normally distributed with standard deviation σ_0. If σ_0 is unknown, it can be estimated by imaging a homogeneous surface and computing the standard deviation of image pixels. The consistency of a voxel can be estimated using the likelihood ratio test: $\lambda_V = \frac{(n-1)s^2}{\sigma_0^2}$, distributed as χ^2 [13].

3.1 Voxel Coloring Algorithm

The algorithm is as follows:

$\mathcal{S} = \emptyset$
`for` $i = 1, \ldots, r$ `do`
  `for every` $V \in \mathcal{V}_{\mathcal{C}}^{d_i}$ `do`
    `project to` $\mathcal{I}_1, \ldots, \mathcal{I}_m$`, compute` λ_V
    `if` $\lambda_V < thresh$ `then` $\mathcal{S} = \mathcal{S} \cup \{V\}$

The threshold, $thresh$, corresponds to the maximum allowable correlation error. An overly conservative (small) value of $thresh$ results in an accurate but incomplete reconstruction. On the other hand, a large threshold yields a more complete reconstruction, but one that includes some erroneous voxels. In practice, $thresh$ should be chosen according to the desired characteristics of the reconstructed model, in terms of accuracy vs. completeness.

Much of the work of the algorithm lies in the computation of λ_V. The set of overlapping pixels depends both on the shape of V's projection and the set $\mathcal{S}$ of possibly occluding voxels. To simplify the computation, our implementation used a square mask to approximate the projected voxel shape. The problem of detecting occlusions is solved by the scene traversal ordering used in the algorithm; the order is such that if V occludes V' then V is visited before V'. Therefore, occlusions can be detected by using a one-bit mask for each image pixel. The mask is initialized to 0. When a voxel V is processed, π_i is the set of pixels that overlap V's projection in $\mathcal{I}_i$ and have mask values of 0. These pixels are marked with masks of 1 if $\lambda_V < thresh$.

Voxel traversal can be made more efficient by employing alternative occlusion-compatible norms. For instance, using the axis-aligned bounding box of the camera volume instead of $\mathcal{C}$, and L_∞ instead of L_2, gives rise to a sequence of axis-aligned cube-shaped layers.

3.2 Discussion

The algorithm visits each voxel exactly once and projects it into every image. Therefore, the time complexity of voxel coloring is: $O(voxels * images)$. To determine the space complexity, observe that evaluating one voxel does not require access to or comparison with other voxels. Consequently, voxels need not be stored in main memory during the algorithm; the voxels making up the voxel coloring will simply be output one at a time. Only the images and one-bit masks need to be allocated. The fact that the space and time complexities of voxel coloring are linear in the number of images is essential in that large numbers of images can be processed at once.

The algorithm differs from stereo and optical-flow techniques in that it does not perform window-based image correlation in the reconstruction process. Correspondences are found during the course of scene traversal by voxel projection. A disadvantage of this searchless strategy is that it requires very precise camera calibration to achieve the triangulation accuracy of stereo methods. Accuracy and run-time also depend on the voxel resolution, a parameter that can be set by the user or determined automatically to match the pixel resolution, calibration accuracy, and computational resources.

Importantly, the approach reconstructs only one of the potentially numerous scenes consistent with the input images. Consequently, it is susceptible to aperture problems caused by image regions of near-uniform color. These regions cause cusps in the reconstruction (see Fig. 2(e)), since voxel coloring yields the reconstruction closest to the camera volume. This is a bias, just like smoothness is a bias in stereo methods, but one that guarantees a consistent reconstruction even with severe occlusions.

4 Experimental Results

The first experiment involved 3D reconstruction from twenty-one views spanning a 360° object rotation. Our strategy for calibrating the views was similar to that in [14]. Instead of a turntable, we placed the objects on a software-controlled pan-tilt head, viewed from above by a fixed camera (see Fig. 1(a)). Tsai's method [15] was used to calibrate the camera with respect to the head, by rotating a known object and manually selecting image features for three pan positions. The calibration error was approximately 3%.

Fig. 3 shows the voxel colorings computed from a complete revolution of a dinosaur toy and a rose. To facilitate reconstruction, we used a black background and eliminated most of the background points by thresholding the images. While background subtraction is not strictly necessary, leaving this step out results in background-colored voxels scattered around the edges of the scene volume. The threshold may be chosen conservatively since removing most of the background pixels is sufficient to eliminate this background scattering effect. The middle column in Fig. 3 shows the reconstructions from a viewpoint corresponding to one of the input images (shown at left), to demon-

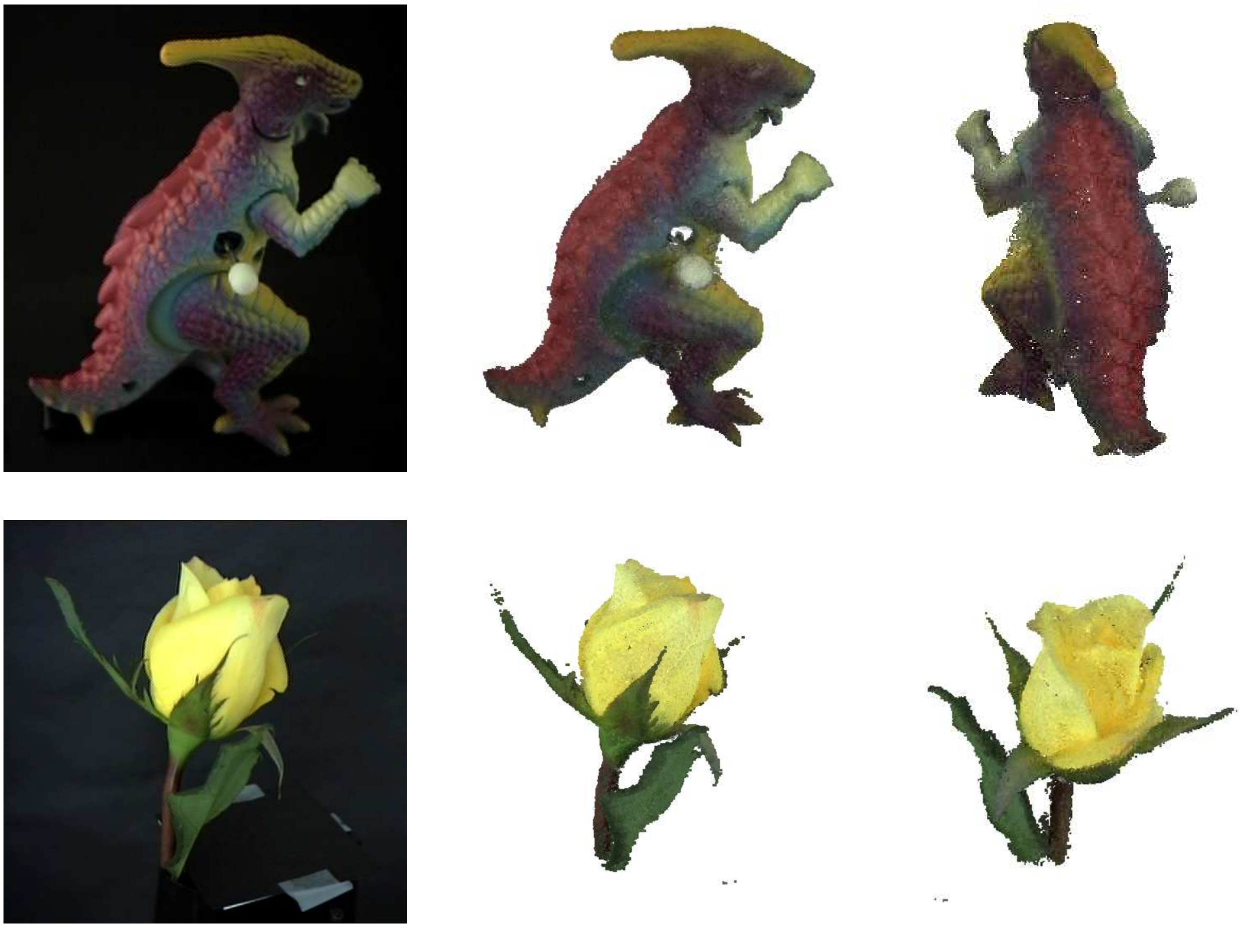
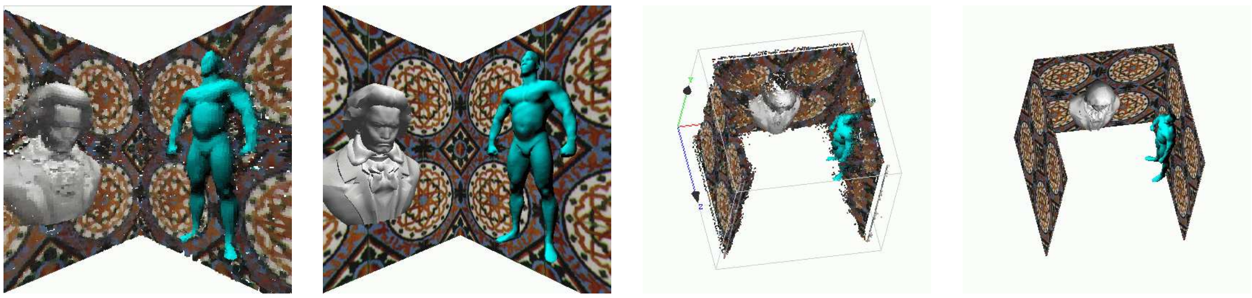


Figure 3: Voxel Coloring of Dinosaur Toy and Rose. The objects were rotated 360° below a camera. At left is one of 21 input images of each object. The other images show different views rendered from the reconstructions.



(a) **(b)** **(c)** **(d)**

Figure 4: Reconstruction of Synthetic Room Scene. The input images were all taken from cameras located inside the room. (a) shows the voxel coloring and (b) the original model from a new interior viewpoint. (c) and (d) show the reconstruction and original model, respectively, from a new viewpoint outside of the room.

strate photo integrity. Note that even fine details such as the wind-up rod on the dinosaur and the leaves of the rose were reconstructed.

We experimented with different voxel resolutions to determine the effects of voxel sampling on reconstruction quality. Increasing the sampling rate improved the reconstruction quality, up to the limits of image quantization and calibration accuracy, at the cost of increased run-time. A low-resolution model can be built very quickly; a reconstruction (not shown) containing 980 voxels took less than a second to compute on a 250 MHz SGI Indigo2. In contrast, the 72,497-voxel dinosaur reconstruction shown in Fig. 3 required evaluating a volume of 7 million voxels and took roughly three minutes to compute.

The next experiment involved reconstructing a synthetic room from cameras *inside* the room. The room interior was highly concave, making reconstruction by volume intersection or other contour-based methods impractical. The room consisted of three texture-mapped walls and two shaded models. The models, a bust of Beethoven and a human figure, were illuminated diffusely from above. 24 cameras were placed at different positions and orientations throughout the room. The optical axes were parallel to the horizontal (XZ) plane.

Fig. 4 compares the original and reconstructed models from new viewpoints. The voxel coloring reproduced images from the room interior quite accurately (as shown in (a)), although some fine details were lost due to quantization effects. The overhead view (c) more clearly shows some discrepancies between the original and reconstructed shapes. For instance, the reconstructed walls are not perfectly planar, as some points lie just off the surface. This point drift effect is most noticeable in regions where the texture is locally homogeneous, indicating that texture information is important for accurate reconstruction. Not surprisingly, the quality of image (c) is worse than that of (a), since the former view was much farther from the input cameras. On the whole, Fig. 4 shows that the overall shape of the scene was captured quite well in the reconstruction. The recovered model contained 52,670 voxels and took 95 seconds to compute.

5 Conclusions

This paper presented a new scene reconstruction technique that incorporates intrinsic color and texture information for the acquisition of photorealistic scene models. Unlike existing stereo and structure-from-motion techniques, the method *guarantees* that a consistent reconstruction is found, even under large visibility differences across the input images. The method relies on a constraint on the input camera configuration that enables a simple solution for determining voxel visibility. A second contribution was the constructive proof of the existence of a set of color invariants. These points are useful in two ways: first, they provide information that is intrinsic, i.e., constant across all possible consistent scenes. Second, together they constitute a spatial and photometric reconstruction of the scene whose projections reproduce the input images.

References

[1] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: A factorization method," *Int. J. of Computer Vision*, vol. 9, no. 2, pp. 137–154, 1992.

[2] T. Kanade, P. J. Narayanan, and P. W. Rander, "Virtualized reality: Concepts and early results," in *Proc. IEEE Workshop on Representation of Visual Scenes*, pp. 69–76, 1995.

[3] S. Moezzi, A. Katkere, D. Y. Kuramura, and R. Jain, "Reality modeling and visualization from multiple video sequences," *IEEE Computer Graphics and Applications*, vol. 16, no. 6, pp. 58–63, 1996.

[4] P. Beardsley, P. Torr, and A. Zisserman, "3D model acquisition from extended image sequences," in *Proc. European Conf. on Computer Vision*, pp. 683–695, 1996.

[5] L. Robert, "Realistic scene models from image sequences," in *Proc. Imagina 97 Conf.*, (Monaco), pp. 8–13, 1997.

[6] Y. Nakamura, T. Matsuura, K. Satoh, and Y. Ohta, "Occlusion detectable stereo–occlusion patterns in camera matrix," in *Proc. Computer Vision and Pattern Recognition Conf.*, pp. 371–378, 1996.

[7] P. N. Belhumeur and D. Mumford, "A Bayesian treatment of the stereo correspondence problem using half-occluded regions," in *Proc. Computer Vision and Pattern Recognition Conf.*, pp. 506–512, 1992.

[8] R. T. Collins, "A space-sweep approach to true multi-image matching," in *Proc. Computer Vision and Pattern Recognition Conf.*, pp. 358–363, 1996.

[9] A. Katayama, K. Tanaka, T. Oshino, and H. Tamura, "A viewpoint dependent stereoscopic display using interpolation of multi-viewpoint images," in *Proc. SPIE Vol. 2409A*, pp. 21–30, 1995.

[10] L. McMillan and G. Bishop, "Plenoptic modeling," in *Proc. SIGGRAPH 95*, pp. 39–46, 1995.

[11] S. B. Kang and R. Szeliski, "3-D scene data recovery using omnidirectional multibaseline stereo," in *Proc. Computer Vision and Pattern Recognition Conf.*, pp. 364–370, 1996.

[12] A. Laurentini, "How far 3D shapes can be understood from 2D silhouettes," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 17, no. 2, pp. 188–195, 1995.

[13] J. E. Freund, *Mathematical Statistics*. Englewood Cliffs, NJ: Prentice Hall, 1992.

[14] R. Szeliski, "Rapid octree construction from image sequences," *Computer Vision, Graphics, and Image Processing: Image Understanding*, vol. 1, no. 58, pp. 23–32, 1993.

[15] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf cameras and lenses," *IEEE Trans. Robotics and Automation*, vol. 3, no. 4, pp. 323–344, 1987.