What Do N Photographs Tell Us About 3D Shape?

Kiriakos N. Kutulakos
kyros@cs.rochester.edu

Steven M. Seitz
sseitz@microsoft.com

Abstract

In this paper we consider the problem of computing the 3D shape of an unknown, arbitrarily-shaped scene from multiple color photographs taken at known but arbitrarily-distributed viewpoints. By studying the equivalence class of all 3D shapes that reproduce the input photographs, we prove the existence of a special member of this class, the maximal photo-consistent shape*, that (1) can be computed from an arbitrary volume that contains the scene, and (2) subsumes all other members of this class. We then give a provably-correct algorithm for computing this shape and present experimental results from applying it to the reconstruction of a real 3D scene from several photographs. The approach is specifically designed to (1) build 3D shapes that allow faithful reproduction of all input photographs, (2) resolve the complex interactions between occlusion, parallax, shading, and their effects on arbitrary collections of photographs of a scene, and (3) follow a "least commitment" approach to 3D shape recovery.*

1 Introduction

Little is currently known about how the combination of occlusion, shading, and parallax in N arbitrary photographs of a scene constrain the scene's shape. The reason is that most approaches to shape recovery have been following a *principle of most commitment*: they employ as many *a priori* assumptions as needed to ensure that shape-from-stereo, shading, or occluding contours is well-posed when considered in isolation. Examples include the use of smoothness constraints for regularization [1], the use of small stereo baselines for minimizing the effect of occlusions [2], and the use of texture-less surfaces for contour-based reconstruction [3, 4]. Unfortunately, since no shape information about the scene is available when these assumptions are made, it is impossible to predict their effect on the final reconstruction.

In contrast to these approaches, we consider the shape recovery problem from a completely different perspective, following a *least commitment principle* to recover 3D shape: our goal is to extract as much information as possible about the scene from arbitrary photographs without making any assumptions about the scene's 3D shape, and without relying on the existence of relevant features in the input photographs. We achieve this by re-examining the shape recovery problem from first principles and answering three questions:

- How can we analyze the family of *photo-consistent shapes*, i.e., the shapes that, when assigned appropriate reflectance properties and re-projected into all the original photographs, reproduce those photographs?
- Is it possible to compute a shape from this family and if so, what is the algorithm?
- What is the relationship of the computed shape to all other photo-consistent shapes?

The key observation used in our work is that these questions become particularly easy to answer when scene radiance belongs to a general class of radiance functions we call *locally computable*. This class characterizes scenes for which global illumination effects such as shadows, transparencies and inter-reflections can be ignored, and is sufficiently general to include scenes with parameterized radiance models. Using this observation as a starting point, we show how to compute a shape, starting from an arbitrary volume that bounds the scene, that is photo-consistent with the N photographs. We show that the computed shape is precisely the shape defined by the occlusion, shading, and parallax cues and that it can be computed without detecting image features as an intermediate step. The only requirements are that (1) the cameras are calibrated, (2) their position is known for all input photographs, and (3) scene radiance follows a known, locally computable radiance function. Experimental results illustrating our method's performance are given for a geometrically-complex real scene.

One of the main difficulties in recovering shape from multiple photographs of a 3D scene is that radiance, self-occlusion, as well as parallax all contribute to a photograph's ultimate dependence on viewpoint. A wealth of research on stereo [5], shape-from-shading [6–8] and shape-from-contour [3, 9–11] illustrates the usual paradigm for dealing with this difficulty—studying a single 3D shape cue under the assumptions that (1) other sources of variability

can be safely ignored, and (2) the input photographs contain features relevant to that cue. Implicit in this work is the view that untangling parallax, self-occlusion and shading effects in N arbitrary photographs of a scene leads to a problem that is either under-constrained or intractable. Here we challenge this view by showing that shape recovery from N arbitrary photographs of an unknown scene is not only a tractable problem but has a simple solution as well.

Reconstruction by least commitment has several advantages over existing methods. First, it allows us to delay the application regularization techniques until the very end of the reconstruction process, when their implications on photo-consistency can be thoroughly resolved. This approach is similar in spirit to "stratification" approaches of shape recovery [12, 13], in which 3D shape is first recovered modulo an equivalence class of reconstructions and is then refined *within* that class at subsequent stages of processing. Second, because the reconstructed scene is photo-consistent, it ensures that its projections will closely resemble photographs of the true scene. This property is especially important in computer graphics, virtual reality, and tele-presense applications [2, 11, 14, 15] where the photo-realism of constructed 3D models is of primary importance. Third, it is the only method, to our knowledge, that does not put any restrictions on the camera positions from which the photographs can be acquired.

Our approach extends and generalizes previous work by providing a detailed geometrical analysis of the family of shapes that are photo-consistent with N arbitrary photographs. This family defines the shape ambiguity inherent in the input photographs and tells us precisely what 3D shape information about the scene we can extract without imposing *a priori* assumptions, aside from a scene radiance model. Our analysis therefore shares similar objectives with studies of the ambiguities in structure-from-motion algorithms [16], even though the geometry and techniques employed here are completely different. Moreover, the special geometrical properties of this family give a way to extract this 3D shape information using discrete volumetric algorithm that iteratively "carves" space. In this respect, our approach bears strong similarity to previous volume-based reconstruction methods. Indeed, the results presented here represent a direct generalization of silhouette-based techniques like volume intersection [4, 10] to the case of grayscale and full-color images, and extend voxel coloring [17] and plenoptic decomposition [18] to the case of arbitrary camera geometries.

The remainder of this paper is structured as follows. Section 2 analyzes the constraints that a set of photographs place on scene structure, given a known model of scene radiance. Based these constraints, a theory of photo-consistency is developed that provides a basis for characterizing the space of all reconstructions of a scene. Section 3 uses this theory to derive the notion of a *maximal photo-consistent shape* which provides the tightest possible bound on the space of all photo-consistent scene reconstructions. Section 4 presents our *least commitment* approach to scene reconstruction and Section 5 concludes with experimental results on real images.

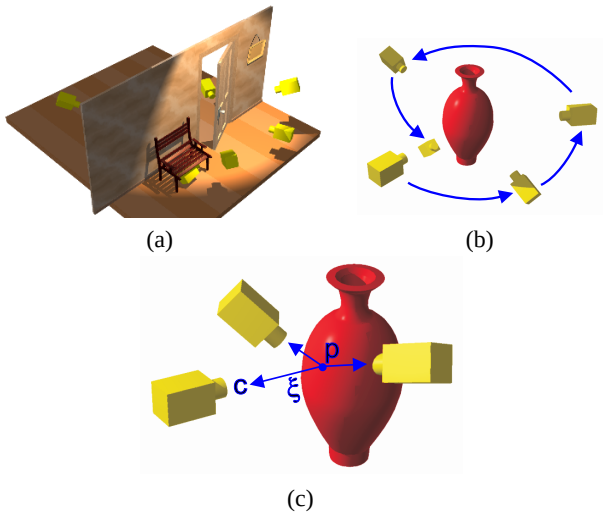


Figure 1. The scene volume and camera distribution covered by our analysis can both be completely arbitrary. Examples include (a) a 3D environment viewed from a collection of cameras that are arbitrarily dispersed in free space, and (b) a 3D object viewed by a single camera moving around it. (c) Viewing geometry.

2 Picture Constraints

Let $\mathcal{V}$ be a 3D scene defined by a finite, opaque, and possibly disconnected volume in space. We assume that $\mathcal{V}$ is viewed under perspective projection from N known positions $c_1, \ldots, c_N$ in $\Re^3 - \mathcal{V}$ (Figure 1). The *radiance* of a point p on the scene's surface is a function $rad_p(\xi)$ that maps every oriented ray ξ through the point to the color of light reflected from p along ξ. We use the term *shape-radiance scene description* to denote the scene $\mathcal{V}$ together with an assignment of a radiance function to every point on its surface. This description contains all the information needed to reproduce a photograph of the scene for any camera position. In general, such a photograph will contain a potentially empty set of background pixels that are not images of any scene point.

Every photograph of a 3D scene taken from a known location partitions the set of all possible shape-radiance scene descriptions into two families, those that reproduce the photograph and those that do not. We characterize this constraint for a given shape and a given radiance assignment by the notion of *photo-consistency*:

Definition 1 (Point Photo-Consistency) *A point p that is visible from c is* photo-consistent with the photograph at c *if (1) p does not project to a background pixel, and (2) the color at p's projection is equal to* $\text{rad}_p(\vec{pc})$.

Definition 2 (Shape-Radiance Photo-Consistency) *A shape-radiance scene description is* photo-consistent with the photograph at c *if all visible points are photo-consistent and every non-background pixel is the projection of a point in $\mathcal{V}$.*

Definition 3 (Shape Photo-Consistency) *A shape $\mathcal{V}$ is* photo-consistent with the photograph at c *if there is an assignment of radiance functions to the visible points of $\mathcal{V}$ that makes the resulting shape-radiance description photo-consistent.*

Our goal is to provide a concrete characterization of the family of all scenes that are photo-consistent with N input photographs. We achieve this by making explicit the two ways in which photo-consistency with N photographs can constrain a scene's shape.

2.1 Background Constraints

Photo-consistency requires that no point of $\mathcal{V}$ projects to a background pixel. If a photograph taken at position c contains identifiable background pixels, this constraint restricts $\mathcal{V}$ to a cone defined by c and the photograph's non-background pixels. Given N such photographs, the scene is restricted to the *visual hull*, which is the volume of intersection of their corresponding cones [4, 10].

When no *a priori* information is available about the scene's radiance, the visual hull defines all the shape constraints in the input photographs. This is because there is always an assignment of radiance functions to the points on the surface of the visual hull that makes the resulting shape-radiance description photo-consistent with the N input photographs.[1] The visual hull can therefore be thought of as a "least commitment reconstruction" of the 3D scene—any further refinement of this volume must necessarily rely on additional assumptions about the scene's shape or radiance.

While visual hull reconstruction has often been used as a method for recovering 3D shape from photographs [10], the picture constraints captured by the visual hull only exploit information from the background pixels in these photographs. Unfortunately, these constraints become useless when photographs contain no background pixels (i.e., the visual hull degenerates to $\Re^3$) or when background identification cannot be performed accurately. Below we study the picture constraints provided by non-background pixels when the scene's radiance is restricted to a special class of radiance models. The resulting picture constraints will in general lead to photo-consistent scenes that are strict subsets of the visual hull.

[1] For example, set $rad_p(\vec{pc})$ equal to the color at p's projection.

2.2 Radiance Constraints

The color of light reflected in different directions from a single scene point usually exhibits a certain degree of coherence for physical scenes that are not transparent or mirror-like. This coherence provides additional picture constraints that depend entirely on non-background pixels. Here we exploit this idea by focusing on scenes whose radiance satisfies the following criterion:

> **Consistency Check Criterion:** An algorithm $\texttt{consist}_K()$ is available that takes as input at least $K << N$ colors $col_1, \ldots, col_K$, K vectors $\xi_1, \ldots, \xi_K$, and the light source positions (non-Lambertian case), and decides whether it is possible for a single surface point to reflect light of color col_i in direction ξ_i simultaneously for all $i = 1, \ldots, K$.

Given a shape $\mathcal{V}$, the Consistency Check Criterion gives us a way to establish the photo-consistency of every point on $\mathcal{V}$'s surface. This criterion defines a general class of radiance models, which we call *locally computable*, that are characterized by a locality property: the radiance at any point is independent of the radiance of all other points in the scene. The class of locally-computable radiance models therefore restricts our analysis to scenes where global illumination effects such as transparency, inter-reflection, and shadows can be ignored. This class subsumes Lambertian radiance ($K = 2$) as well as radiance models that can be expressed in closed form by a small number of parameters.[2]

When an *a priori* locally computable radiance model is established for a physical 3D scene, the model provides sufficient information to determine whether a given shape $\mathcal{V}$ is not photo-consistent with a collection of photographs. The use of radiance models that are locally consistent is important in this context because the *non*-photo-consistency of a shape $\mathcal{V}$ tells us a great deal about the shape of the underlying scene. This in turn imposes a very special structure on the family of photo-consistent shapes. We use the following two lemmas to make this structure explicit. These lemmas provide the analytical tools needed to describe how the non-photo-consistency of a shape $\mathcal{V}$ affects the photo-consistency of its subsets (Figure 2):

Lemma 1 (Visibility Lemma) *Let p be a point on $\mathcal{V}$'s surface,* $\text{Surf}(\mathcal{V})$*, and let* $\text{Vis}_{\mathcal{V}}(p)$ *be the collection of input photographs in which $\mathcal{V}$ does not occlude p. If $\mathcal{V}' \subset \mathcal{V}$ is a shape that also has p on its surface,* $\text{Vis}_{\mathcal{V}}(p) \subseteq \text{Vis}_{\mathcal{V}'}(p)$.

Proof: Since $\mathcal{V}'$ is a subset of $\mathcal{V}$, no point of $\mathcal{V}'$ can lie between p and the cameras corresponding to $Vis_{\mathcal{V}}(p)$. *QED*

[2] Specific examples include (1) using a mobile camera mounted with a light source to capture photographs of a scene whose reflectance can be expressed in closed form, and (2) using multiple cameras to capture photographs of an approximately Lambertian scene under arbitrary unknown illumination (Figure 1).

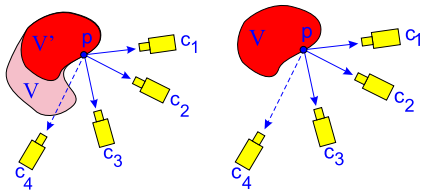


Figure 2. (a) Illustration of the Visibility Lemma. (b) Illustration of the Non-Photo-Consistency Lemma. If p is non-photo-consistent with the photographs at c_1, c_2, c_3, it is non-photo-consistent with the entire set $Vis_{\mathcal{V}}(p)$, which also includes c_4.

Lemma 2 (Non-Photo-Consistency Lemma) *If $p \in \text{Surf}(\mathcal{V})$ is not photo-consistent with a subset of $\text{Vis}_{\mathcal{V}}(p)$, it is not photo-consistent with $\text{Vis}_{\mathcal{V}}(p)$.*

Intuitively, Lemmas 1 and 2 suggest that both visibility and non-photo-consistency exhibit a certain form of "monotonicity:" the Visibility Lemma tells us that the collection of photographs from which a surface point is visible strictly expands for nested subsets of $\mathcal{V}$ that contain the point (Figure 2(a)). Analogously, the Non-Photo-Consistency Lemma, which follows as a direct consequence of the definition of photo-consistency, tells us that each new photograph can be thought of as an additional constraint on the photo-consistency of surface points—the more photographs are available, the more difficult it is for those points to maintain their photo-consistency. Furthermore, once a surface point looses its photo-consistency no new photographs of the scene can re-establish it.

The key consequence of Lemmas 1 and 2 is given by the following theorem which shows that *non*-photo-consistency at a point rules out the photo-consistency of an entire family of shapes:

Theorem 1 (Subset Theorem) *If $p \in \text{Surf}(\mathcal{V})$ is not photo-consistent, no photo-consistent subset of $\mathcal{V}$ contains p.*

Proof: Let $\mathcal{V}' \subset \mathcal{V}$ be a shape that contains p. Since p lies on the surface of $\mathcal{V}$, it must also lie on the surface of $\mathcal{V}'$. From the Visibility Lemma it follows that $Vis_{\mathcal{V}}(p) \subseteq Vis_{\mathcal{V}'}(p)$. The theorem now follows by applying the Non-Photo-Consistency Lemma to $\mathcal{V}'$ and using the locality property of locally computable radiance models. *QED*

We explore the ramifications of the Subset Theorem in the next section where we provide an explicit characterization of the shape ambiguities inherent in the input photographs.

3 The Maximal Photo-Consistent Shape

The family of all shapes that are photo-consistent with a collection of N photographs defines the ambiguity inherent in the problem of recovering 3D shape from those photographs. This is because it is impossible to decide, based on those photographs alone, which photo-consistent shape is the shape of the true scene. When using photographs to recover the shape of a 3D scene, this ambiguity raises two questions:

- Is it possible to compute a shape that is photo-consistent with N photographs and, if so, what is the algorithm?
- If a photo-consistent shape can be computed, how can we relate that shape to all other photo-consistent 3D interpretations of the scene?

Our answer to both questions rests on the following theorem. Theorem 2 shows that for any shape $\mathcal{V}$ there is a unique photo-consistent shape that subsumes all other photo-consistent shapes in $\mathcal{V}$ (Figure 3):

Theorem 2 (Maximal Photo-Consistent Shape Theorem) *Let $\mathcal{V}$ be an arbitrary set and let $\mathcal{V}^*$ be the union of all photo-consistent subsets of $\mathcal{V}$. The shape $\mathcal{V}^*$ is photo-consistent and is called the* maximal photo-consistent shape.

Proof: (By contradiction) Suppose $\mathcal{V}^*$ is not photo-consistent and let p be a non-photo-consistent point on its surface. Since $p \in \mathcal{V}^*$, there exists a photo-consistent shape, $\mathcal{V}' \subset \mathcal{V}^*$, that also has p on its surface. It follows from the Subset Theorem that $\mathcal{V}'$ is not photo-consistent. *QED*

Theorem 2 provides an explicit relation between the maximal photo-consistent shape and all other possible 3D interpretations of the scene: the theorem guarantees that every such interpretation is a refinement of the maximal photo-consistent shape. The maximal photo-consistent shape therefore represents a least-commitment reconstruction of the scene. We describe a volumetric algorithm for computing this shape in the next section.

4 Least-Commitment Reconstruction

An important feature of the maximal photo-consistent shape is that it can actually be computed using a simple, discrete algorithm that "carves" space in a well-defined way. Given an initial volume $\mathcal{V}$ that contains the scene, the algorithm proceeds by iteratively removing (i.e. "carving") portions of that volume until it becomes identical to the maximal photo-consistent shape, $\mathcal{V}^*$. The algorithm can therefore be fully specified by answering four questions: (1) how do we select the initial volume $\mathcal{V}$, (2) how should we represent that volume to facilitate carving, (3) how do we carve at each iteration to guarantee convergence to the maximal photo-consistent shape, and (4) when do we terminate carving?

The choice of the initial volume has a considerable impact on the outcome of the reconstruction process (Figure

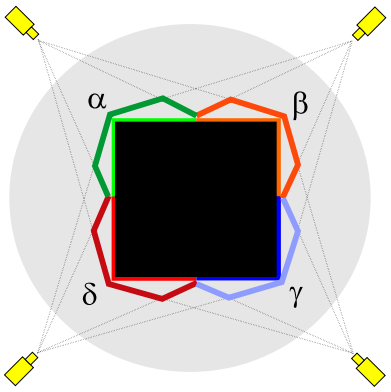


Figure 3. Illustration of the Maximal Photo-Consistent Shape Theorem. The scene is viewed by four cameras and consists of a black square whose sides are "painted" diffuse red, blue, orange, and green. The gray-shaded region corresponds to a shape $\mathcal{V}$ containing the scene. In this example, $\mathcal{V}^*$ is a polygonal region that extends beyond the true scene and whose boundary is defined by the polygonal segments α, β, γ, and δ. When these segments are colored as shown, $\mathcal{V}^*$'s projection is indistinguishable from that of the true scene and *no* photo-consistent shape can contain points outside $\mathcal{V}^*$. The goal of the algorithm in Section 4 is to compute this shape, given $\mathcal{V}$, the photographs, and the camera positions. Note that $\mathcal{V}^*$'s shape depends on the specific scene radiance model and could be significantly different for a similarly-colored, non-diffuse scene viewed from the same positions.

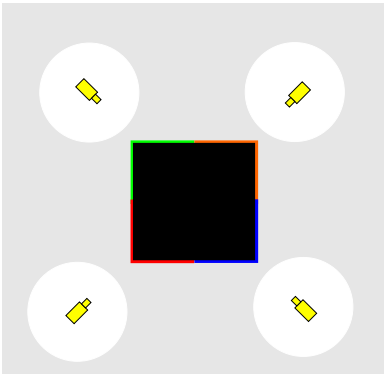


Figure 4. Choosing the initial volume $\mathcal{V}$. If $\mathcal{V}$ is chosen to simply exclude a small circle around each camera in the scene of Figure 3, the maximal photo-consistent shape becomes equal to $\mathcal{V}$. This is because no point on $\mathcal{V}$'s surface is visible by more than one camera and, hence, $\mathcal{V}$ is photo-consistent.

4). Nevertheless, selection of this volume is beyond the scope of this paper; it will depend on the specific 3D shape recovery application and on information about the manner in which the input photographs were acquired.[3] Below we consider a general algorithm that, given N photographs and *any* initial volume that contains the scene, is guaranteed to find the (unique) maximal photo-consistent shape contained in that volume.

4.1 Reconstruction by Space Carving

Let $\mathcal{V}$ be an arbitrary finite volume that contains the scene. We represent $\mathcal{V}$ as a finite collection of voxels $v_1, \ldots, v_M$ whose surface conforms to a radiance model defined by a consistency check algorithm $\texttt{consist}_K()$. Using this representation, each carving iteration removes a single voxel from $\mathcal{V}$.

The Subset Theorem leads directly to a method for selecting the voxel to carve away from $\mathcal{V}$ at each iteration. Specifically, the proposition tells us that if a voxel v on the surface of $\mathcal{V}$ is not photo-consistent, the volume $\mathcal{V} = \mathcal{V} - \{v\}$ must still contain the maximal photo-consistent shape. Hence, if only non photo-consistent voxels are removed at each iteration, the carved volume is guaranteed to converge to the maximal photo-consistent shape. The order in which non-photo-consistent voxels are examined and removed is not important for guaranteeing correctness. Convergence to this shape occurs when no non-photo-consistent voxel can be found on the surface of the carved volume. These considerations lead to the following algorithm for computing the maximal photo-consistent shape:

Space Carving Algorithm

Step 1: Initialize $\mathcal{V}$ to a superset of the scene.

Step 2: Repeat the following steps until a non-photo-consistent voxel v is found on the surface of $\mathcal{V}$:

- **a.** Project v to all photographs in $Vis_{\mathcal{V}}(v)$. Let $col_1, \ldots, col_j$ be the colors at v's projection in each photograph and let $\xi_1, \ldots, \xi_j$ be the optical rays connecting v to the corresponding optical centers.
- **b.** Determine the photo-consistency of v using $\texttt{consist}_K(col_1, \ldots, col_j, \xi_1, \ldots, \xi_j)$.

Step 3: If no non-photo-consistent voxel is found, set $\mathcal{V}^* = \mathcal{V}$ and terminate. Otherwise, set $\mathcal{V} = \mathcal{V} - \{v\}$ and continue with Step 2.

The key step in the space carving algorithm is the search and voxel consistency checking of Step 2. The following proposition gives an upper bound on the number of voxel photo-consistency checks that must be performed during space carving:

[3] Examples include defining $\mathcal{V}$ to be equal to the visual hull or, in the case of a camera moving through an environment , $\Re^3$ minus a tube along the camera's path.

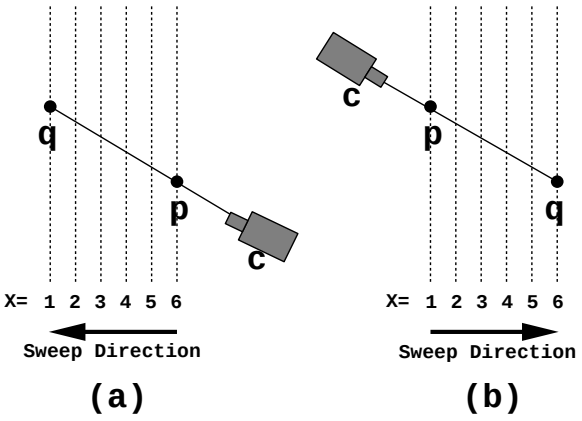


Figure 5. Visiting voxels in order of visibility. Sweeping a plane in the direction of increasing (decreasing) x coordinate ensures that a voxel p will be visited before every voxel q that it occludes, for all cameras to the left (right) of p.

Proposition 1 *The total number of required photo-consistency checks is bounded by $N * M$ where N is the number of input photographs and M is the number of voxels in the initial (i.e., uncarved) volume.*

Proof: Since (1) the photo-consistency of a voxel v that remains on $\mathcal{V}$'s surface for several carving iterations can change only when $Vis_{\mathcal{V}}(v)$ changes due to $\mathcal{V}$'s carving, and (2) $Vis_{\mathcal{V}}(v)$ expands monotonically as $\mathcal{V}$ is carved (Visibility Lemma), the photo-consistency of v must be checked at most N times. *QED*

4.2 Multi-Pass Visibility Computation

In order to implement the Space Carving Algorithm, the following three operations, performed in Step 2 of the algorithm, must be supported: (1) determine $Surf(\mathcal{V})$, (2) compute $Vis_{\mathcal{V}}(v)$ for each voxel $v \in \mathcal{V}$, and (3) check to see if v is photo-consistent. Because carving a single voxel can affect global visibility, it is essential to be able to keep track of visibility information in a way that may be efficiently updated.

To reduce visibility computations, we propose a multi-pass algorithm for space carving. A pass consists of sweeping a plane through the scene volume and testing the photo-consistency of voxels on that plane. The advantage of the plane sweep approach is that voxel are always visited in a visibility-compatible order—if voxel v occludes v', then v will necessarily be visited before v'. This property is achieved by considering only the cameras that are on one side of the plane in each pass, as shown in Figure 5.

Specifically, consider two points $p = (p_x, p_y, p_z)$ and $q = (q_x, q_y, q_z)$, such that p occludes q from a camera centered at c. p therefore lies on a line segment with endpoints q and c. One of the following inequalities must therefore hold:

$$c_i <= p_i <= q_i \quad for\ i = x, y,\ or\ z \tag{1}$$

$$c_i >= p_i >= q_i \quad for\ i = x, y,\ or\ z \tag{2}$$

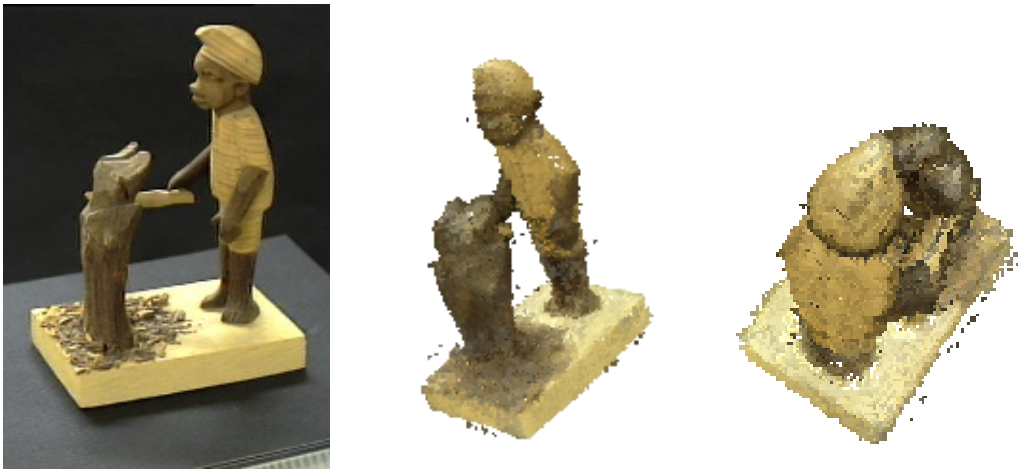
These inequalities suggest the following order for visiting voxels: consider voxels in order of increasing x coordinate, i.e., for a series of planes $x = x_1, x = x_2, \ldots, x = x_n$ with x_i increasing. For a given plane $x = x_i$, consider all cameras centered at c such that $c_x < x_i$. Eq. (1), ensures that points will be visited in order of occlusion (i.e., p before q with respect to any camera c). By Eq. (2), the same holds true when the plane is swept in order of *decreasing* x coordinate.

Observe that sweeping planes in increasing or decreasing x order does not treat the case where $c_x = p_x = q_x$. This problem may be solved by adding an additional sweep through the volume in either one of the positive or negative y or z direction. A more serious objection is that this approach considers only a limited set of cameras at a time. Consequently, it is possible that a voxel will not be carved even though it is not photo-consistent. This problem may be ameliorated by sweeping the space in multiple directions, i.e., in increasing/decreasing x, y, and z directions, to minimize the chance that a non-photo-consistent voxel is not detected. Alternatively, a more sophisticated data structure could be used to keep track of the cameras considered for each voxel.

5 Experimental Results

We ran the Space Carving Algorithm on several images of a wooden sculpture to evaluate the performance of the multi-pass algorithm. The images were acquired by placing the object on a calibrated pan-tilt head and rotating it in front of a camera. To facilitate the carving process, the images were also thresholded to remove background pixels. While this step is not strictly necessary, it is useful to eliminate spurious background-colored voxels scattered throughout the scene volume.

A Lambertian model was used for the Consistency Check Criterion, i.e., it was assumed that a voxel projects to pixels of the same color in every image. In particular, we thresholded the standard deviation of the image pixels to determine whether or not a voxel should be carved. A high threshold (18% average RGB component error) was used to compensate for changes in illumination due to object rotation. Consequently, some fine details were lost in the final reconstruction. The initial volume, $\mathcal{V}$, was chosen to be a solid cube containing the sculpture. Figure 6 shows selected input images and new views of the reconstruction, $\mathcal{V}^*$. As can be seen from these images, the reconstruction captures



(a) (b) (c)

Figure 6. Reconstruction of a wood sculpture. One of 21 input images is shown (a), along with views of the reconstruction from similar (b) and overhead (c) views.

the shape of the sculpture quite accurately, although fine details like the wood grain are blurred. With the exception of a few stray voxels, the reconstruction is very smooth, in spite of the fact that no smoothness bias was used by the algorithm. The reconstruction, which required a total of 24 sweeps through the volume, contains 22000 voxels. It took 8 minutes to compute $\mathcal{V}^*$ on a Silicon Graphics Indy workstation.

6 Concluding Remarks

The use of photo-consistency as a criterion for recovering 3D shape brings about a qualitative change in the way in which shape recovery algorithms are designed and analyzed. We have shown that it allows us to (1) build 3D shapes that allow faithful reproduction of all input photographs, (2) resolve the complex interactions between occlusion, parallax, shading, and their effects on arbitrary collections of photographs of a scene, and (3) follow a "least commitment" approach to 3D shape recovery.

Current research directions include (1) the use of surface-based methods for shape recovery, (2) use of *a priori* shape constraints to refine the maximal photo-consistent shape, (3) analysis of the topological structure of the family of photo-consistent shapes, and (4) development of optimal space carving algorithms.

References

[1] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, vol. 317, no. 26, pp. 314–319, 1985.

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

[3] R. Cipolla and A. Blake, "Surface shape from the deformation of apparent contours," *Int. J. Computer Vision*, vol. 9, no. 2, pp. 83–112, 1992.

[4] A. Laurentini, "The visual hull concept for silhouette-based image understanding," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 2, pp. 150–162, 1994.

[5] M. Okutomi and T. Kanade, "A multiple-baseline stereo," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, no. 4, pp. 353–363, 1993.

[6] R. Epstein, A. L. Yuille, and P. N. Belhumeur, "Learning object representations from lighting variations," in *Object Representation in Computer Vision II* (J. Ponce, A. Zisserman, and M. Hebert, eds.), pp. 179–199, Springer-Verlag, 1996.

[7] P. N. Belhumeur and D. J. Kriegman, "What is the set of images of an object under all possible lighting conditions," in *Proc. Computer Vision and Pattern Recognition*, pp. 270–277, 1996.

[8] R. J. Woodham, Y. Iwahori, and R. A. Barman, "Photometric stereo: Lambertian reflectance and light sources with unknown direction and strength," Tech. Rep. 91-18, University of British Columbia, Laboratory for Computational Intelligence, August 1991.

[9] R. Vaillant and O. D. Faugeras, "Using extremal boundaries for 3-d object modeling," *IEEE Trans. Pat-*

tern Anal. Machine Intell., vol. 14, no. 2, pp. 157–173, 1992.

[10] R. Szeliski, "Rapid octree construction from image sequences," *CVGIP: Image Understanding*, vol. 58, no. 1, pp. 23–32, 1993.

[11] 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.

[12] O. Faugeras, "Stratification of three-dimensional vision: projective, affine, and metric representations," *J. Opt. Soc. Am. A*, vol. 12, no. 3, pp. 465–484, 1995.

[13] J. J. Koenderink and A. J. van Doorn, "Affine structure from motion," *J. Opt. Soc. Am.*, vol. A, no. 2, pp. 377–385, 1991.

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

[15] Z. Zhang, "Image-based geometrically-correct photorealistic scene/object modeling (ibphm): A review," in *Proc. Asian Conf. Computer Vision*, 1998. To appear.

[16] O. D. Faugeras and S. Maybank, "Motion from point matches: multiplicity of solutions," *Int. J. Computer Vision*, vol. 4, pp. 225–246, 1990.

[17] S. M. Seitz and C. R. Dyer, "Photorealistic scene reconstruction by voxel coloring," in *Proc. Computer Vision and Pattern Recognition Conf.*, pp. 1067–1073, 1997.

[18] S. M. Seitz and K. N. Kutulakos, "Plenoptic image editing," in *Proc. 6th Int. Conf. Computer Vision*, 1998. To appear.