A Probabilistic Theory of Occupancy and Emptiness

Rahul Bhotika[1], David J. Fleet[2], and Kiriakos N. Kutulakos[3]

[1] Computer Science Department, University of Rochester, Rochester, NY 14627, USA
[2] Palo Alto Research Center, Palo Alto, CA 94304, USA
[3] Computer Science Department, University of Toronto, Toronto, ON M5S3G4, Canada
{*bhotika@cs.rochester.edu, fleet@parc.com, kyros@cs.toronto.edu*}

Abstract. This paper studies the inference of 3D shape from a set of n noisy photos. We derive a probabilistic framework to specify what one can infer about 3D shape for arbitrarily-shaped, Lambertian scenes and arbitrary viewpoint configurations. Based on formal definitions of visibility, occupancy, emptiness, and photo-consistency, the theoretical development yields a formulation of the *Photo Hull Distribution*, the tightest probabilistic bound on the scene's true shape that can be inferred from the photos. We show how to (1) express this distribution in terms of image measurements, (2) represent it compactly by assigning an occupancy probability to each point in space, and (3) design a stochastic reconstruction algorithm that draws fair samples (i.e., 3D photo hulls) from it. We also present experimental results for complex 3D scenes.

1 Introduction

Many applications, from robotics [1] to computer graphics [2] and virtual reality, rely on probabilistic representations of occupancy to express uncertainty about the physical layout of the 3D world. While occupancy has an intuitive and well-defined physical interpretation for every 3D point (i.e., a point is occupied when the scene's volume contains it), little is known about what probabilistic information regarding occupancy is theoretically computable from a set of n noisy images. In particular, what information about the occupancy and emptiness of space is intrinsic to a set of n stochastically-generated photos taken at arbitrary viewpoints? Importantly, how can we answer this question for general scenes whose shape is arbitrary and unknown?

This paper outlines a general probabilistic theory of occupancy and emptiness for Lambertian scenes that is intended to handle arbitrary 3D shapes, arbitrary viewpoint configurations and noisy images. Based on formal definitions of several key concepts, namely, visibility, occupancy, emptiness and photo-consistency, our theoretical development yields a formulation of the *Photo Hull Distribution*. The Photo Hull Distribution is the tightest probabilistic bound on the scene's true shape that can be inferred from the n input photos. Importantly, we show that the theory leads directly to a stochastic algorithm that (1) draws fair samples from the Photo Hull Distribution and (2) converges to an optimal conservative estimate of the occupancy probability, i.e., to the lowest upper bound on that probability that is theoretically computable from the photos. Experimental results illustrating the applicability of our framework are given for real scenes.

Our work is motivated by four objectives that we believe are key to designing a general and useful probabilistic theory of occupancy and emptiness:

– **Understanding stereo ambiguities:** Since stereo reconstruction can be ambiguous even in the noiseless case, a clear understanding of the mathematical relationship
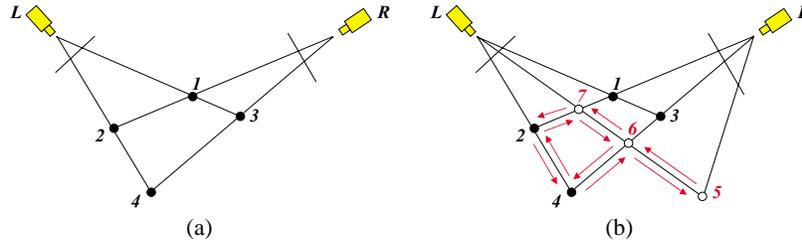
Figure 1. (a) Noiseless images of four Lambertian points with identical albedos afford seven consistent scene interpretations—{2,3},{1,4} {1,2,3}, {1,2,4}, {1,3,4},{2,3,4}, and {1,2,3,4}. Maximizing image consistency in the presence of even small amounts of noise will cause one of these interpretations to be arbitrarily preferred. The resulting shape (e.g., $\{2, 3\}$), however, cannot conclusively rule out the possibility that other points may also be occupied (e.g., $\{1\}$). (b) The probabilistic event "5 is visible from L" implies that points 6 and 7 are in empty space, thereby inducing a cycle of probabilistic dependencies (arrows) that make the events "4 is visible from L" and "5 is visible from L" dependent on each other. Such cycles characterize the visibility and occupancy dependencies between *every* pair of points in $\mathbb{R}^3$ within the cameras' view.

between *shape ambiguity* and *shape uncertainty*, and their effect on occupancy computation, becomes crucial in the design of stable algorithms (Figure 1a).

– **Probabilistic treatment of visibility:** Unconstrained scenes induce arbitrarily complex visibility relationships between cameras and surface points, and arbitrarily long probabilistic dependency cycles between any two 3D points (Figure 1b). One must provide a physically-valid and self-consistent probabilistic interpretation of occupancy (which refers to individual points) and visibility (which depends on a scene's global shape).

– **Algorithm-independent analysis of occupancy:** Without a clear understanding of what is theoretically computable from a set of input images, it is difficult to assess the performance of individual algorithms and the effect of heuristics or approximations they employ.

– **Handling sensor and model errors:** Since computational models are inherently limited in their ability to describe the physical world and the true image formation process, it is important to account for such errors.

A key aspect of our analysis is the explicit distinction we make between the notions of *ambiguity* and *uncertainty*. Intuitively, we interpret shape *ambiguity* as the multiplicity of solutions to the n-view reconstruction problem given *noiseless* images. Shape *uncertainty*, on the other hand, exists in the presence of noise and modeling errors and is modeled as a probability distribution that assigns a probability to every 3D shape.

While ambiguities have long been recognized as an important problem in stereo (e.g., [3]), the question of how these ambiguities affect the estimation of occupancy and emptiness is not well understood. Existing stereo algorithms make one or more assumptions that allow one to estimate a unique shape. Most methods compute the shape that maximizes some form of image consistency likelihood (i.e., that a reconstructed shape is consistent with the image measurements), assuming constraints or prior beliefs about shape and image disparities. Examples include the use of SSD-based likelihoods [4], window-based correlation [5], ordering constraints on correspondence [6, 7], shape priors [8], as well as MRF [9, 10] and parametric models [11, 12]. But these approaches do not estimate the probability that a region is empty or occupied. This is because image

consistency is not equivalent to occupancy, nor is it the opposite of emptiness (Figure 1a). In general, ascertaining the emptiness of a region requires that one consider the likelihood that *no* consistent scene interpretation intersects that region, a problem for which no formal analysis currently exists.

Even though recent formulations of the n-view stereo problem have demonstrated an ability to resolve global visibility relationships among views [13–19], most of these ignore the existence of noise and the resulting model uncertainty. Two notable exceptions are the Roxel Algorithm [20] and the Probabilistic Space Carving Algorithm [21], which attempt to capture shape uncertainty by assigning a "soft" occupancy value between zero and one to every voxel in space. Working under the assumption that an accurate treatment of long-range probabilistic dependencies is intractable, these approaches rely on heuristic approximations to assess a voxel's visibility. For instance, the Roxel Algorithm assesses a voxel's visibility by assuming that voxels in front of it are "semi-transparent," an assumption that is not consistent with image formation models for opaque scenes. Analogously, the Probabilistic Space Carving Algorithm assumes that visibilities of distant voxels are independent from each other and that visibility information can be propagated locally from voxel to voxel without regard to any specific global 3D shape. Neither algorithm takes into account the global nature of visibility, leading to occupancy calculations that do not afford a probabilistic interpretation in a strict mathematical sense. Similar problems would arise in any method that relies on a local model for propagating probabilities since long-distance dependency cycles, like those depicted in Figure 1b, must be taken into account.

To our knowledge, no attempts have been made to define the notions of "occupancy probability" and "emptiness probability" in a mathematically rigorous manner when noisy images are given as input. Instead, existing approaches define their goal procedurally, by giving an algorithm that assigns soft occupancies to 3D points. The question of what probabilistic information is theoretically computable from a set of n noisy images, without regard to specific algorithms or implementations, is therefore still open. In practice, this makes it difficult to characterize the behavior of individual algorithms, to define precise notions of optimality and to compare results from different algorithms.

In this paper, we argue that a probabilistic treatment of occupancy and emptiness that accounts for stereo ambiguities, shape uncertainty and *all* long-range probabilistic dependencies caused by visibility can be formulated. Moreover, it leads to a fairly straightforward reconstruction algorithm. At the heart of our approach lies the realization that we can overcome technical difficulties via a probabilistic modeling of *shape space*, i.e., the space of all 3D reconstructions. Using the photo-consistency theory in [19] as a starting point, we show that shape-space modeling allows one to derive a complete theory of occupancy and emptiness. This theory arises from a small set of "axioms" (a generative image formation model and a radiance independence assumption) and does not rely on heuristic assumptions about scene shape or the input photos.

2   Shape-Space Modeling: Photo Hulls, Occupancy, and Emptiness

In the formulation of shape ambiguity and uncertainty that follows, we represent shapes using a discrete volumetric approximation and we further define a bounding volume $V \subseteq \Re^3$ that is assumed to contain the scene. The *shape space* of V is defined simply as its powerset, 2^V. With these definitions, we use the photo-consistency theory of Kutulakos and Seitz [19] to model shape ambiguities and to introduce the concept of the *Photo Hull Distribution* to model shape uncertainty.
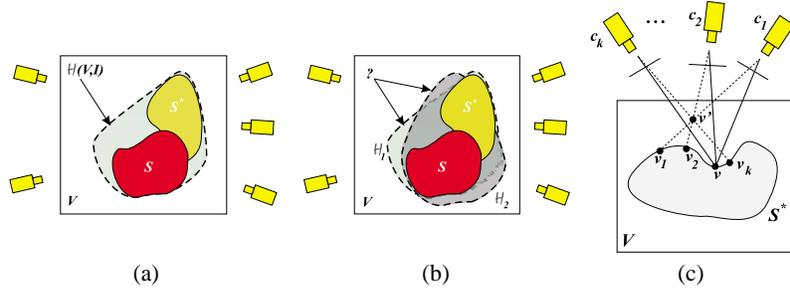
Figure 2. (a) The photo hull $\mathcal{H}(V, I)$ contains all photo-consistent shapes (e.g., S^* and a shape S). (b) When noise is present, the photo hull cannot be established with certainty. The Photo Hull Distribution assigns a probability to every possible hull (e.g., $\mathcal{H}_1$ and $\mathcal{H}_2$). (c) Foreground and background modeling: the foreground and background models are satisfied for voxels $\mathbf{v}$ and $\mathbf{v}'$, respectively.

2.1 The Photo Hull

Photo-consistency is based on the idea that a valid solution to the reconstruction problem should reproduce the scene's input photographs exactly when no image noise is present. When one or more such photographs are available, each taken from a known viewpoint, they define an equivalence class of solutions called *photo-consistent shapes*. For Lambertian shapes, a voxel $\mathbf{v} \in S$ is *photo-consistent* if (1) $\mathbf{v}$ is not visible to any camera, or (2) we can assign a radiance to $\mathbf{v}$ that is equal to the color at $\mathbf{v}$'s projection in all photos where $\mathbf{v}$ is visible. Accordingly, a shape $S \in 2^V$ is *photo-consistent* if all its voxels are photo-consistent.

The equivalence class of photo-consistent shapes captures all the ambiguities present in a set of n noiseless images and is closed under set union (Figure 2a):

Theorem 1 (Photo Hull Theorem [19]) For every volume V that contains the true scene and every noiseless image set I, there exists a unique shape called the *Photo Hull*, $\mathcal{H}(V, I)$, that is the union of all shapes in 2^V that are photo-consistent with I. This shape is a superset of the true scene and is itself photo-consistent.

The photo hull is the tightest possible bound on the true scene that can be recovered from n photographs in the absence of *a priori* scene information. The photo hull is especially important in our analysis for two reasons. First, it is algorithm-independent and is applicable to scenes of arbitrary shape. Second, it suggests that even though photo-consistency is an ambiguous property (i.e., many distinct 3D shapes can possess it), *emptiness* of space is not; the volume $V - \mathcal{H}(V, I)$ cannot intersect *any* photo-consistent 3D shape. Hence, by studying how noise affects the information we can recover about the photo hull and about the emptiness of space (rather than about individual photo-consistent shapes) we can establish a clear mathematical distinction between the analyses of shape ambiguity and shape uncertainty. For simplicity and without loss of generality we assume in this paper that the true scene S^* is equal to the photo hull.

2.2 The Photo Hull Distribution

When the set I of input images is generated stochastically, the emptiness or occupancy of voxels cannot be determined with certainty any more (Figure 2b). To model this uncertainty, we introduce the Photo Hull Distribution:

Definition 1 (Photo Hull Probability) The *Photo Hull Probability* of $S \in 2^V$ is the conditional probability

$$p\left[\textbf{Hull}(S, V, I) \mid V, I\right] : 2^V \to [0, 1],$$

where **Hull**(S, V, I) is the event that S is the photo hull $\mathcal{H}(V, I)$.

The Photo Hull Distribution captures all the information that we can possibly extract from noisy photographs of an arbitrarily-shaped scene in the absence of prior information. This is because (1) scene radiances are not observable and hence shape information cannot be extracted with certainty, and (2) the inherent ambiguities of the (noiseless) n-view stereo problem preclude bounding the scene with anything but its photo hull.

While useful as an analytical tool, the Photo Hull Distribution is impractical as a shape representation because it is a function over 2^V, an impossibly large set. To overcome this limitation, we also study how to reconstruct 3D shapes that have high probability of being the photo hull and how to compute scene reconstructions that represent shape uncertainty in a compact way. To achieve the latter objective, we introduce a probabilistic definition of *occupancy*:

Definition 2 (Occupancy Probability) The *occupancy probability* of a voxel **v** is given by the marginal probability

$$p\left[\textbf{Occ}(\textbf{v}, V, I) \mid V, I\right] \; = \sum_{\{S \mid S \in 2^V \, and \, \textbf{v} \in S\}} p\left[\textbf{Hull}(S, V, I) \mid V, I\right].$$

Intuitively, the occupancy probability defines a scalar field over V whose value at a voxel $\textbf{v} \in V$ is equal to the probability that **v** is contained in the photo hull.

Our goal in the remainder of this paper is to characterize the information that n noisy photos provide about the shape of an arbitrary 3D scene. We achieve this by deriving an image-based expression for the photo hull distribution and a method for drawing fair samples from it. Toward this end, we first factor the photo hull probability of a shape S into two components, namely, the probability that S is photo-consistent conditioned on the emptiness of its exterior $V - S$ (so that the shape is visible), and the probability that its exterior is empty:

$$p\left[\textbf{Hull}(S, V, I) \mid V, I\right] = \\ p\left[\textbf{PhotoC}(S, I) \mid \textbf{Empty}(V - S, I), V, I\right] \; p\left[\textbf{Empty}(V - S, I) \mid V, I\right]. \quad (1)$$

Here, **PhotoC**(S, I) denotes the event that S is photo-consistent with image set I. Similarly, **Empty**$(V - S, I)$ denotes the event that the exterior volume $V - S$ does not intersect any shape in 2^V that is photo-consistent with I. The following sections derive expressions for these two fundamental distributions in terms of image measurements.

3 Probabilistic Modeling of Photo-Consistency

When confronted with the problem of formulating the photo-consistency probability $p\left[\textbf{PhotoC}(S, I) \mid \textbf{Empty}(V - S, I), V, I\right]$ in terms of image measurements two questions arise: (1) how do we express the photo-consistency probability of a 3D shape S in terms of individual pixels (which are projections of individual voxels on the shape), and (2) how do we ensure that all global dependencies between 3D points are correctly accounted for? To answer these questions, we use two key ideas.

First, we assume *radiance independence*, that the radiances of any two surface points on the true scene are independent. This simplifies the probabilistic modeling of the photo-consistency of individual voxels. The second idea, which we call the *visibility-conditioning principle*, states that, with the radiance independence assumption, the photo-consistency of individual voxels are conditionally independent, when conditioned on the voxels' visibility in the true scene, S^*. This conditional independence is derived from our image formation model that is outlined below, and leads to the following factorization of the photo-consistency probability of shape S:[1]

Lemma 1 (Photo-Consistency Probability Lemma)

$$p\left[\textbf{PhotoC}(S, I) \mid \textbf{Empty}(V - S, I), V, I\right] \quad = \quad \prod_{\mathbf{v} \in S} p\left[\textbf{PhotoC}(\mathbf{v}, I, \beta_S) \mid I, \beta_S(\mathbf{v})\right] \quad , \quad (2)$$

where $\textbf{PhotoC}(\mathbf{v}, I, \beta_S)$ denotes the event that voxel $\mathbf{v}$ is photo-consistent, conditioned on the set of cameras from which it is visible, $\beta_S(\mathbf{v})$. The following sections give our simple model of image formation and derive the form of the voxel photo-consistency probability that it entails.

3.1 Image Formation Model

Let S^* be an opaque, Lambertian 3D scene that occupies an arbitrary volume in $\Re^3$. Also assume that the scene is viewed under perspective projection from n distinct and known viewpoints, $\mathbf{c}_1, \ldots, \mathbf{c}_n$, in $\Re^3 - S^*$. Then, the measured RGB value of a pixel depends on three factors: (1) *visibility*; (2) *surface radiance*; and (3) *pixel noise*.

We use the term *visibility state* to denote the set of input viewpoints from which $\mathbf{x}$ is visible. It can be represented as an n-bit vector $\beta_{S^*}(\mathbf{x})$ whose i-th bit, $[\beta_{S^*}(\mathbf{x})]_i$, is one if and only if $\mathbf{x}$ is visible from $\mathbf{c}_i$. Each 3D scene produces a unique *visibility function*, $\beta_{S^*} : \Re^3 \to \{0, 1\}^n$, mapping points to their visibility states.

The radiance, $\mathbf{r}(.)$, is a function that assigns RGB values to surface points on S^*. Let $\mathbf{r}(.)$ be a random variable with a known distribution, $\mathcal{R}$, over $[0, 1]$; i.e., $\forall \mathbf{x} : \mathbf{r}(\mathbf{x}) \sim \mathcal{R}^3 [0, 1]$. Ignoring noise, if a surface point $\mathbf{x}$ is visible from viewpoint $\mathbf{c}_i$, then the RGB values observed from $\mathbf{x}$ in the camera at $\mathbf{c}_i$ are equal to $\mathbf{x}$'s radiance [23]; more formally,
$$(\mathbf{x} \in S^*) \wedge ([\beta_{S^*}(\mathbf{x})]_i = 1) \quad \implies \quad proj_i(\mathbf{x}) = \mathbf{r}(\mathbf{x}),$$
where $proj_i(\mathbf{x})$ denotes the image RGB value at $\mathbf{x}$'s projection along $\mathbf{c}_i$. In practice, image measurements are corrupted by noise. Here, we model pixel noise, $\mathbf{n}$, as mean-zero Gaussian additive noise with variance $\sigma_{\mathbf{n}}^2$. The RGB value $proj_i(\mathbf{x})$ is therefore a random variable given by $proj_i(\mathbf{x}) = \mathbf{r}(\mathbf{x}) + \mathbf{n}$. Assuming that radiances and noise are independent, it follows that (1) the mean of $proj_i(\mathbf{x})$ is $E[\mathbf{r}(\mathbf{x})]$, (2) its variance is $Var[\mathbf{r}(\mathbf{x})] + \sigma_{\mathbf{n}}^2$, and (3) conditioned on radiance, the measured pixel color is distributed according to the noise distribution:

$$proj_i(\mathbf{x}) | \mathbf{r}(\mathbf{x}) \sim \mathcal{N}(\mathbf{r}(\mathbf{x}), \sigma_{\mathbf{n}}^2).$$

In what follows we use $\mathbf{proj}(\mathbf{x})$ to denote the vector of random variables $\{proj_i(\mathbf{x}) : [\beta_{S^*}(\mathbf{x})]_i\}$ of a point $\mathbf{x}$ with visibility state $\beta_{S^*}(\mathbf{x})$. We use I to denote the set of images obtained from the n cameras, and $I(\mathbf{x}, \beta_{S^*})$ to denote a sample of the random vector $\mathbf{proj}(\mathbf{x})$, i.e., the vector of image RGB values observed at $\mathbf{x}$'s projections.

---

[1] See the Appendix for proof sketches of some of the paper's main results. Complete proofs can be found in [22].

3.2 Photo-Consistency

Given only $\beta_{S^*}(\mathbf{v})$ and $I(\mathbf{v}, \beta_{S^*})$, what can we say about whether or not $\mathbf{v}$ belongs to S^*? To answer this question we formulate two generative models that characterize the distribution of the set of observations, $I(\mathbf{v}, \beta_{S^*})$, that corresponds to $\mathbf{v}$.

The foreground model captures the fact that if $\mathbf{v}$ lies on the surface and $\beta_{S^*}(\mathbf{v})$ is $\mathbf{v}$'s correct visibility state, then all radiances in $\mathbf{proj}(\mathbf{v})$ are identical since they arise from the same location on the surface:

Observation 1 (Foreground Model) If $\mathbf{v} \in S^*$ and $k = |I(\mathbf{v}, \beta_{S^*})|$ is the number of cameras for which $\mathbf{v}$ is visible, then (1) the distribution of $\mathbf{proj}(\mathbf{v})$ conditioned on the actual radiance, $\mathbf{r}(\mathbf{v})$, is a multi-variate Gaussian

$$\mathbf{proj}(\mathbf{v}) \mid \mathbf{r}(\mathbf{v}) \sim \mathcal{N}^k(\mathbf{r}(\mathbf{v}), \sigma_{\mathbf{n}}^2), \quad \text{and} \tag{3}$$

(2) the sample variance of the RGB measurements, $Var[I(\mathbf{v}, \beta_{S^*})]$, will converge to the noise variance $\sigma_{\mathbf{n}}^2$ as k tends to infinity.

In practice, k is finite. It follows from Eq. (3) that the observation likelihood, $p[Var[I(\mathbf{v}, \beta_{S^*})] \mid \mathbf{v} \in S^*]$, is a χ^2 density with $3(k-1)$ degrees of freedom.

If $\mathbf{v}$ is not a voxel on S^*, the color at each of its k projections is due to a distinct scene voxel $\mathbf{v}_j$ (Figure 2c), with $proj_j(\mathbf{v}) = \mathbf{r}(\mathbf{v}_j) + \mathbf{n}$ for $j = 1, \ldots, k$. This, along with the radiance independence assumption, that radiances at different voxels are independent, leads to our simple background model:

Observation 2 (Background Model) If $\mathbf{v} \notin S^*$, the distribution of $\mathbf{proj}(\mathbf{v})$ conditioned on the actual radiances is given by

$$\mathbf{proj}(\mathbf{v}) \mid \{\mathbf{r}(\mathbf{v}_j)\}_{j=1}^k \sim \prod_{j=1}^k \mathcal{N}(\mathbf{r}(\mathbf{v}_j), \sigma_{\mathbf{n}}^2), \tag{4}$$

the sample variance of which, $Var[I(\mathbf{v}, \beta_{S^*})]$, converges to $Var[\mathbf{r}] + \sigma_{\mathbf{n}}^2$ as k tends to infinity.

In practice, with finite cameras and an unknown background distribution, we model the likelihood $p[Var[I(\mathbf{v}, \beta_{S^*})] \mid \mathbf{v} \notin S^*]$, using the input photos themselves: we generate 10^6 sets of pixels, each of size $|I(\mathbf{v}, \beta_{S^*})|$, whose pixels are drawn randomly from distinct photos. This collection of variances is used to model the distribution of sample variances for voxels that do not lie on the surface.

We can now use Bayes' rule to derive the probability that a voxel is on the scene's surface given a set of image measurements associated with its visibility state β:

$$p[\mathbf{v} \in S^* \mid Var[I(\mathbf{v}, \beta)]] = \frac{p[Var[I(\mathbf{v}, \beta)] \mid \mathbf{v} \in S^*]}{p[Var[I(\mathbf{v}, \beta)] \mid \mathbf{v} \in S^*] + p[Var[I(\mathbf{v}, \beta)] \mid \mathbf{v} \notin S^*]} . \tag{5}$$

Note that this particular formulation of the foreground and background likelihoods was chosen for simplicity; other models (e.g., see [21]) could be substituted without altering the theoretical development below.

Finally, by defining the voxel photo-consistency probability to be the probability that $\mathbf{v}$ is consistent with the scene surface conditioned on visibility, i.e.,

$$p[\mathbf{PhotoC}(\mathbf{v}, I, \beta_S) \mid I, \beta_S(\mathbf{v})] \stackrel{\text{def}}{=} p[\mathbf{v} \in S^* \mid Var[I(\mathbf{v}, \beta_S(\mathbf{v}))]] ,$$

we obtain the factored form for the shape photo-consistency probability in Lemma 1.
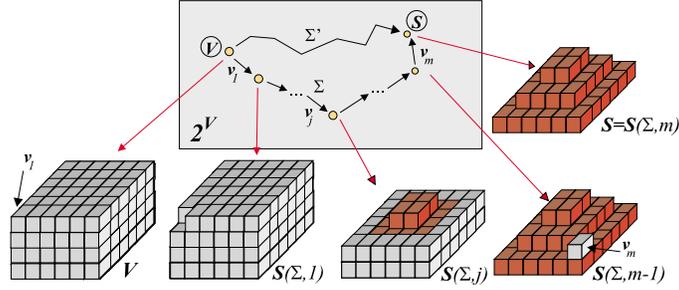
Figure 3. Visualizing voxel visibility strings and carving sequences. The carving sequence $\Sigma = \langle \overline{\mathbf{v}_1} \cdots \overline{\mathbf{v}_m} \rangle$ defines a sequence of m shapes, where the $(i-1)$-th shape determines the visibility of the i-th voxel and $S(\Sigma, i-1) = V - \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{i-1}\}$. Voxels in $V - S$ and S are shown as light and dark gray, respectively. Also shown is the shape-space path defined by Σ.

4 Probabilistic Modeling of Emptiness

Unlike photo-consistency, the definition of emptiness of region $V - S$ does not depend only on the shape S; it requires that no photo-consistent shape in 2^V intersects that region. In order to express this probability in terms of individual pixel measurements, we follow an approach analogous to the previous section: we apply the visibility-conditioning principle to evaluate $p\left[\mathbf{Empty}(V - S, I) \mid V, I\right]$, the emptiness probability of a volume $V - S$, as a product of individual voxel-specific probabilities. But what voxel visibilities do we condition upon? The difficulty here is that we cannot assign a unique visibility state to a voxel in $V - S$ because its visibility depends on which shape in 2^V we are considering. To overcome this difficulty, we introduce the concept of *voxel visibility strings*. These strings provide a convenient way to express the space of all possible visibility assignments for voxels in $V - S$. Importantly, they allow us to reduce the estimation of emptiness and photo hull probabilities to a string manipulation problem:

Definition 3 (Voxel Visibility String) (1) A *voxel visibility string*, $\Sigma = \langle q_1 \cdots q_{|\Sigma|} \rangle$, is a string of symbols from the set $\{\mathbf{v}, \overline{\mathbf{v}} \mid \mathbf{v} \in V\}$. (2) The shape $S(\Sigma, i)$ induced by the first i elements of the visibility string Σ is obtained by removing from V all voxels in the sub-string $\langle q_1 \cdots q_i \rangle$ that occur in their $\overline{\mathbf{v}}$-form. (3) The *visibility state* of the voxel $\mathbf{v}$ corresponding to string element q_i is determined by the shape $S(\Sigma, i-1)$, i.e., it is the bit vector $\beta_{S(\Sigma, i-1)}(\mathbf{v})$. (4) The *value function*, $F(\Sigma)$, of string Σ is the product

$$F(\Sigma) = \prod_{i=1}^{|\Sigma|} f(q_i), \text{ with } f(q_i) = \begin{cases} p\left[\mathbf{PhotoC}(\mathbf{v}, I, \beta_{S(\Sigma, i-1)}) \mid I, \beta_{S(\Sigma, i-1)}(\mathbf{v})\right] & (q_i = \mathbf{v}) \\ 1 - p\left[\mathbf{PhotoC}(\mathbf{v}, I, \beta_{S(\Sigma, i-1)}) \mid I, \beta_{S(\Sigma, i-1)}(\mathbf{v})\right] & (q_i = \overline{\mathbf{v}}) \end{cases}$$

(5) A voxel visibility string that is a permutation of the voxels in $V - S$, each of which appears in its $\overline{\mathbf{v}}$-form, is called a *carving sequence*.

Note that the set of voxel visibility strings is infinite since voxels can appear more than once in a string. Carving sequences, on the other hand, are always of length $|V - S|$ and have an intuitive interpretation (Figure 3): a carving sequence $\Sigma = \langle \overline{\mathbf{v}_1} \cdots \overline{\mathbf{v}_m} \rangle$ can be thought of as a path in 2^V that "joins" V to S. That is, beginning with the initial volume, V, every time we encounter a string element with a bar we remove the

corresponding voxel and thereby generate a new shape. When we reach the end of the string we will have generated the desired shape S. Thus, the set of all carving sequences for a shape S gives us all the possible ways of obtaining S by carving voxels from V.

Carving sequences provide an avenue for relating the emptiness of space to the non-photo-consistency of individual voxels in $V - S$. This is because the order of voxels in a carving sequence assigns a unique visibility state to every voxel it contains, allowing us to apply the visibility-conditioning principle through the following definition:

Definition 4 (Σ-Non-Photo-Consistency Event) We use $\mathbf{NPhotoC}(V - S, I, \Sigma)$ to denote the event that $V - S$ is non-photo-consistent with respect to carving sequence $\Sigma = \langle \overline{\mathbf{v_1}} \cdots \overline{\mathbf{v_m}} \rangle$. This is the event that no voxel $\mathbf{v}$ in Σ is photo-consistent when its visibility state is given by Σ:

$$\mathbf{NPhotoC}(V - S, I, \Sigma) \iff \bigwedge_{i=1}^{m} \neg \mathbf{PhotoC}(\mathbf{v}_i, I, \beta_{S(\Sigma, i-1)}) \, .$$

Using this definition, it is easy to show that the non-photo-consistency of the volume $V - S$, when conditioned on the visibilities induced by a carving sequence, is simply the product of the non-photo-consistency probabilities of its constituent voxels:

$$p\left[\mathbf{NPhotoC}(V - S, I, \Sigma) \mid V, I, \Sigma\right] = F(\Sigma).$$

To fully specify the emptiness probability of the volume $V - S$ it therefore suffices to establish a relation between the events $\mathbf{NPhotoC}(V - S, I, \Sigma)$ and $\mathbf{Empty}(V - S, I)$. The following proposition gives us a partial answer to this problem:

Lemma 2 (Empty-Space Event) If a non-photo-consistent carving sequence exists, $V - S$ is empty:

$$\bigvee_{\Sigma \in \mathcal{C}(V,S)} \mathbf{NPhotoC}(V - S, I, \Sigma) \implies \mathbf{Empty}(V - S, I).$$

where $\mathcal{C}(V, S)$ is the set of all carving sequences that map V to S.

Lemma 2 suggests a rule to establish a volume's emptiness without exhaustive enumeration of all photo-consistent shapes. It tells us that it suffices to find just one carving sequence whose voxels are all non-photo-consistent. While this rule is sufficient, it is not a necessary one: it is possible for $V - S$ to be empty without any of the events $\mathbf{NPhotoC}(V - S, I, \Sigma)$ being true. This seemingly counter-intuitive fact can be proven by contradiction because it leads to an expression for the Photo Hull Probability that is not a distribution,[2] i.e., $\sum_{S \in 2^V} p\left[\mathbf{Hull}(S, V, I) \mid V, I\right] < 1$. From an informal standpoint, this occurs because the set of voxel visibility strings that are carving sequences does not adequately cover the space of all possible visibility assignments for voxels in $V - S$, resulting in an under-estimation of the "probability mass" for the event $\mathbf{Empty}(V - S, I)$. Fortunately, the following proposition tells us that this mass can be correctly accounted for by a slightly expanded set of voxel visibility strings, all of which are mutually disjoint, which consider all sets of visibility assignments for voxels in $V - S$ and contain every voxel in $V - S$ at most $|V - S|$ times:

---

[2] A simple way to verify this fact is to expand the sum in Eq. (7) for a 2-voxel initial volume V. Details are provided in Appendix A.3.

Proposition 1 (Emptiness Probability)

$$p\left[\textbf{Empty}(V-S, I) \mid V, I\right] \;=\; \frac{1}{|V-S|!} \sum_{\Sigma \in \mathcal{C}(V,S)} F(\Sigma) + \sum_{\Sigma \in \mathcal{C}(V,S)} \sum_{P \in \mathcal{P}(\Sigma) - \mathcal{C}(V,S)} R(P)\,G(P) \qquad (6)$$

where $\mathcal{P}(\Sigma)$ is the set of strings $P = \langle p_1 \cdots p_l \rangle$ with the following properties: (1) P is created by adding voxels in their $\mathbf{v}$-form to a carving sequence Σ, (2) each voxel $\mathbf{v}_i$ occurs at most once between any two successive voxels appearing in their $\overline{\mathbf{v}}$-form, (3) no voxel in P can occur in its $\mathbf{v}$-form after it has occurred in its $\overline{\mathbf{v}}$-form; and $G(P)$ and $R(P)$ are given by

$$G(P) = \prod_{i=1}^{|P|} g(p_i)\,, \text{ where } g(p_i) = \begin{cases} f(p_i) & (1^{\text{st}} \text{ occurrence}) \\ f(p_i)/f(p_j) & (\text{voxel occurs in } \mathbf{v}\text{-form,} \\ & \text{previous occurrence at } p_j) \\ \left[f(p_i) + f(p_j) - 1\right]/f(p_j) & (\text{voxel occurs in } \overline{\mathbf{v}}\text{-form,} \\ & \text{previous occurrence at } p_j) \end{cases}$$

$$R(P) = \frac{(m-l_1)!}{m!}\,\frac{(m-1-l_2)!}{(m-1)!}\,\cdots\,\frac{(2-l_{m-1})!}{2!}$$

where $m = |V - S|$; and l_j is the number of voxels, $\overline{\mathbf{v}}_j$ inclusive, that exist in P between two voxels $\overline{\mathbf{v}}_{j-1}$ and $\overline{\mathbf{v}}_j$ that are consecutive in Σ.

Proposition 1 tells us that the "extra" mass needed to establish the emptiness of a region consists of the $G(P)$-terms in Eq. (6). This result is somewhat surprising, since it suggests that we can capture all relevant visibility configurations of voxels in $V - S$ by only considering a very constrained set of voxel visibility strings, whose length is at most $\frac{|V-S|^2+|V-S|}{2}$. Moreover, it not only leads to a probability distribution for the photo hull but, as shown in Section 6, it leads to a simple stochastic algorithm that produces fair samples from the photo hull distribution. We consider the implications of this proposition below.

5 The Photo Hull Distribution Theorem

One of the key consequences of our probabilistic analysis is that despite the many intricate dependencies between global shape, visibility, emptiness, and photo-consistency, the expression of a shape's photo hull probability in terms of image measurements has a surprisingly simple form. This expression is given to us by the following theorem that re-states Eq. (1) in terms of image measurements by bringing together the expressions in Lemma 1 and Proposition 1. It tells us that for a shape S, this probability is a sum of products of individual voxel photo-consistency or non-photo-consistency probabilities; each voxel in S contributes exactly one voxel photo-consistency factor to the product, while each voxel in $V - S$ contributes multiple non-photo-consistency factors.

Theorem 2 (Photo Hull Distribution Theorem) (1) If every pixel observation in I takes RGB values over a continuous domain, the probability that shape $S \subseteq V$ is the photo hull $\mathcal{H}(V, I)$ is

$$p\left[\textbf{Hull}(S, V, I) \mid V, I\right] \;=\; \sum_{\Sigma \in \mathcal{C}(V,S)} \sum_{P \in \mathcal{P}(\Sigma)} G(P \,\|\, \Pi)\,R(P), \qquad (7)$$

where Π is a string corresponding to an arbitrary permutation of the voxels in S, all of which appear in their $\mathbf{v}$-form; $\mathcal{C}(V, S)$, $\mathcal{P}(\Sigma)$ and $G(\cdot), R(\cdot)$ are defined in Proposition 1; and "$\|$" denotes string concatenation. (2) Eq. (7) defines a probability distribution function, i.e., $\sum_{S \in 2^V} p\left[\textbf{Hull}(S, V, I) \mid V, I\right] = 1$.

6 Probabilistic Occupancy by Fair Stochastic Carving

Even though the photo hull probability of a shape S has a relatively simple form, it is impractical to compute because of the combinatorial number of terms in Eq. (7). Fortunately, while the Photo Hull Distribution cannot be computed directly, we can draw fair samples from it (i.e., complete 3D photo hulls) using a stochastic variant of the space carving algorithm [19]. Moreover, by computing a set of 3D photo hulls through repeated applications of this algorithm, we can approximate the occupancy probability of every voxel in V according to Definition 2.

Occupancy Field Reconstruction Algorithm

- Set $trial = 1$ and run the following algorithm until $trial = K$:
  Stochastic Space Carving Algorithm
  Initialization Set $S_{trial} = V$ and repeatedly perform the following two steps.
  Voxel Selection Randomly choose a voxel $\mathbf{v} \in S_{trial}$ that either has not been selected or whose visibility state changed from β to β' since its last selection.
  Stochastic Carving If no such voxel exists, set $trial = trial + 1$ and terminate. Otherwise, carve $\mathbf{v}$ from S_{trial} with probability

$$\frac{p\left[\mathbf{PhotoC}(\mathbf{v}, I, \beta) \mid I, \beta\right] - p\left[\mathbf{PhotoC}(\mathbf{v}, I, \beta') \mid I, \beta'\right]}{p\left[\mathbf{PhotoC}(\mathbf{v}, I, \beta) \mid I, \beta\right]}$$

- For every $\mathbf{v} \in V$, set $p\left[\mathbf{Occ}(\mathbf{v}, V, I) \mid V, I\right] = L/K$ where L is the number of carved shapes in $\{S_1, \ldots, S_K\}$ containing $\mathbf{v}$.

From a superficial point of view, the Stochastic Space Carving Algorithm in the above inner loop is almost a trivial extension of standard space carving—the only essential difference is the stochastic nature of the carving operation, which does not depend on a pure threshold any more. This carving operation and the precise definition of the carving probability are extremely important, however. This is because if carving is performed according to the above rule, the algorithm is guaranteed to draw fair samples from the Photo Hull Distribution of Theorem 2. In the limit, this guarantees that the occupancy probabilities computed by the algorithm will converge to the theoretically-established values for this probability:

Proposition 2 (Algorithm Fairness) When the number of trials goes to infinity, the frequency by which the Stochastic Carving Algorithm generates shape S will tend to $p\left[\mathbf{Hull}(S, V, I) \mid V, I\right]$.

A second feature of the Stochastic Space Carving Algorithm, which becomes evident from the proof of Proposition 2 in the Appendix, is that it provides us with an approximate means for comparing the likelihoods of two photo hulls it generates: since each run of the algorithm corresponds to a single term in the sum of Eq. (7), i.e., a particular combination of the Σ, P, Π strings, and since we can compute this term from the algorithm's trace as it examines, re-examines and removes voxels, we can compare two photo hulls by simply comparing the value of these terms. This effectively allows us to develop an approximate method for selecting a maximum-likelihood photo hull from a set of generated candidate shapes.[3]

---

[3] Note that this method is approximate because we can only compute a single term in the sum in Eq. (7) that defines $p\left[\mathbf{Hull}(S, V, I) \mid V, I\right]$, not the probability itself.
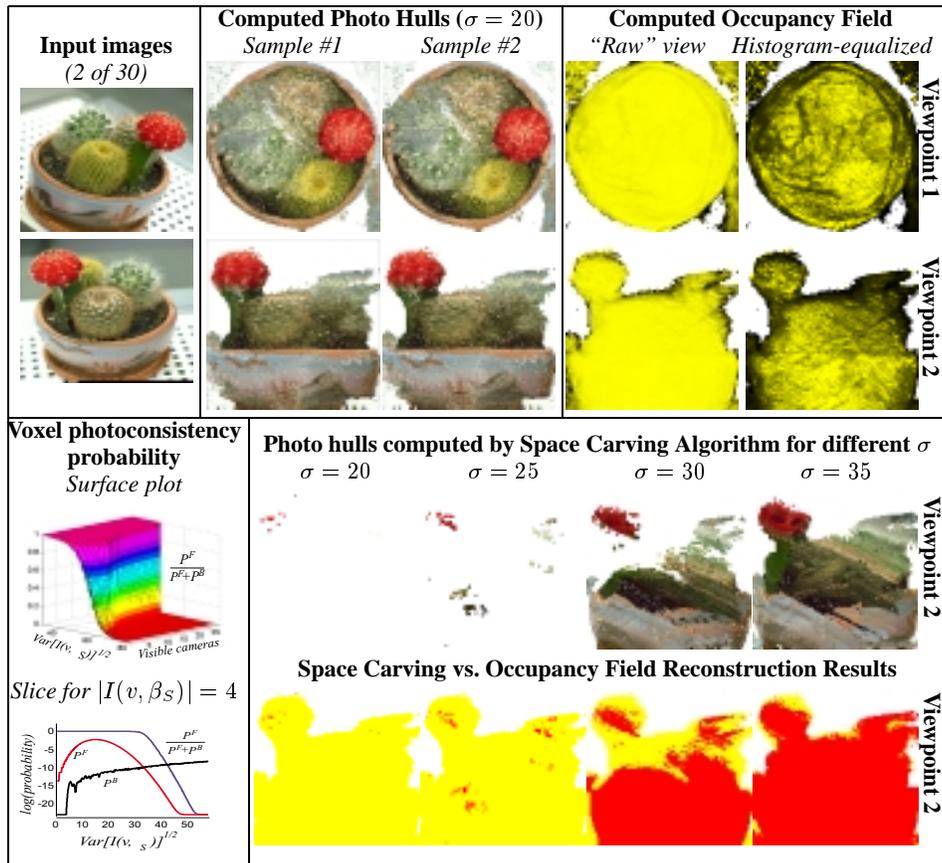
Figure 4. Cactus: Raw views of the occupancy field were created by computing at each pixel, the average over all voxels that project to it and have non-zero occupancy probability. Each image in the last row compares the recovered occupancy field to volumes obtained from standard space carving, shown directly above it. A light (dark) gray pixel indicates that at least one voxel in the occupancy field with non-zero probability projects to it, and that no (at least one) voxel in the carved volume projects to it. Intuitively, light gray pixels mark regions that the standard algorithm "over-carves" , i.e., regions that were carved even though they have a non-negligible probability of being in the photo hull. Here, P^F and P^B correspond to the foreground and background likelihoods in Section 3.2.

7 Experimental Results

To demonstrate the practical utility of our formulation, we applied the algorithm of Section 6 to different scenes of varying complexity and structure. [4]

The cactus dataset (from [18]) is especially challenging because of the drastic changes in visibility between photos, the complex scene geometry, and because the images contain many "mixed" pixels, quantization errors, and 1-4 pixel calibration errors that make

---

[4] The complete input sequences are available at http://www.cs.toronto.edu/˜kyros/research/occupancy/.

pixel matching difficult. The carving probabilities in the Stochastic Carving Step of the algorithm require foreground and background likelihoods (denoted P^F and P^B in Figure 4). The foreground likelihood was given by a χ^2 density with $\sigma = 20$. The background likelihood was modeled from the input photos as described in Section 3.2. The resulting plots in Figure 4 (bottom left) show these likelihoods. When the number of input views is small, the carving probability changes slowly as a function of the variance of a voxel's projections. For many views, where a coincidental pixel match across views is unlikely, it approaches a step function and carving becomes almost deterministic.

We ran the algorithm for $K = 400$ iterations on a cubical, 128^3-voxel initial volume V, generating 400 distinct photo hulls that were used to compute the occupancy field shown in Figure 4 (top right). Each iteration took about 17 minutes on an R12000 SGI O2. Views of two of the 400 photo hulls are shown in Figure 4 (top middle). Despite the large number of stochastic carving decisions made by the algorithm, the generated photo hulls are similar in both shape and appearance, with volumes differing by less than 3% from hull to hull. The number of voxels with non-zero probability in the final occupancy field was only 7% more than the voxels in a single sample, suggesting that despite the high dimensionality of shape-space, the Photo Hull Distribution is "highly peaked" for this dataset. In Figure 4 (bottom right) we compare the occupancy field to photo hulls obtained from the standard space carving algorithm [19] for different variance thresholds. Note that the standard space carving algorithm fails catastrophically for lower values of σ. Even at $\sigma = 35$, standard space carving results in a volume that appears over-carved and has artifacts. This behavior of standard space carving over a wide range of variance thresholds shows that it fails to model the uncertainty in the data set. The occupancy field represents this uncertainty quite well: the volume corresponding to the pot base has voxel occupancies close to 1, while near the top of the tallest cactus, where detail is highest, voxels have lower occupancy probabilities.

Results with $K = 400$ for a second dataset (the gargoyle from [19]) are summarized in Figure 5. Each iteration took about 6 minutes. The gargoyle statuette has more structure than the earlier example and the resulting occupancy field is less diffuse. However, as each voxel is visible in at most 7 of the 16 input images, the voxel probability distribution has a gentler slope, and the occupancy field shows a smoother gradation of voxel occupancy probabilities moving inwards of the object, most noticeably in regions with discontinuities and irregularities on the surface, and in the cavity between the gargoyle's hands and body (Figure 5, bottom row).

8   Concluding Remarks

This paper introduced a novel, probabilistic theory for analyzing the 3D occupancy computation problem from multiple images. We have shown that the theory enables a complete analysis of the complex probabilistic dependencies inherent in occupancy calculations, provides an expression for the tightest occupancy bound recoverable from noisy images, and leads to an algorithm with provable properties and good performance on complex scenes. Despite these features, however, our framework relies on a shape representation that is discrete and on a radiance independence assumption that does not hold for scenes with smoothly-varying textures. Overcoming these limitations will be a major focus of future work. Additional topics will include (1) providing a mechanism for introducing shape priors, (2) developing efficient strategies for sampling the Photo Hull Distribution and for computing likely Photo Hulls, and (3) developing fast, sampling-free approximation algorithms for occupancy field computation.
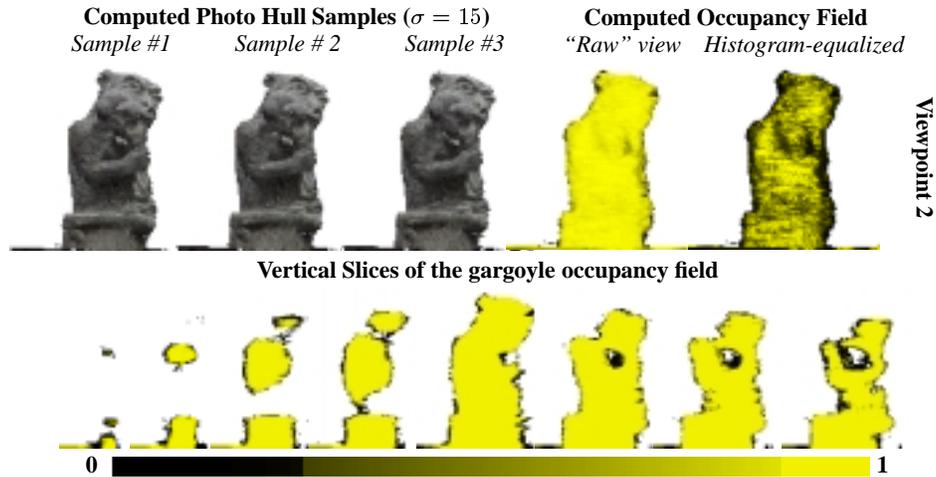
Figure 5. Gargoyle dataset (input images not shown): (Top) 3 of 400 sample hulls generated and the occupancy field obtained by combining them. (Bottom) Vertical slices of the occupancy field. The bar at the bottom shows the mapping from voxel occupancy probabilities to gray levels.

References

1. A. Elfes, "Sonar-based real-world mapping and navigation," *IEEE J. Robotics Automat.*, vol. 3, no. 3, pp. 249–266, 1987.
2. B. Curless and M. Levoy, "A volumetric method for building complex models from range images," in *Proc. SIGGRAPH'96*, pp. 303–312, 1996.
3. D. Marr and T. Poggio, "A computational theory of human stereo vision," *Proceedings of the Royal Society of London B*, vol. 207, pp. 187–217, 1979.
4. L. Matthies, "Stereo vision for planetary rovers: Stochastic modeling to near real-time implementation," *Int. J. Computer Vision*, vol. 8, no. 1, pp. 71–91, 1992.
5. T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: Theory and experiment," *IEEE Trans. PAMI*, vol. 16, no. 9, pp. 920–932, 1994.
6. I. Cox, S. Hingorani, S. Rao, and B. Maggs, "A maximum likelihood stereo algorithm," *CVIU: Image Understanding*, vol. 63, no. 3, pp. 542–567, 1996.
7. H. Ishikawa and D. Geiger, "Occlusions, discontinuities and epipolar line in stereo," in *Proc. 5th European Conf. on Computer Vision*, pp. 232–248, 1998.
8. P. N. Belhumeur, "A Bayesian approach to binocular stereopsis," *Int. J. on Computer Vision*, vol. 19, no. 3, pp. 237–260, 1996.
9. R. Szeliski, "Bayesian modeling of uncertainty in low-level vision," *Int. J. Computer Vision*, vol. 5, no. 3, pp. 271–301, 1990.
10. D. Geiger and F. Girosi, "Parallel and deterministic algorithms from MRF's: surface reconstruction," *IEEE Trans. PAMI*, vol. 13, no. 5, pp. 401–412, 1991.
11. K. Kanatani and N. Ohta, "Accuracy bounds and optimal computation of homography for image mosaicing applications," in *Proc. 7th Int. Conf. on Computer Vision*, vol. 1, pp. 73–78, 1999.
12. P. Torr, R. Szeliski, and P. Anandan, "An integrated bayesian approach to layer extraction from image sequences," *IEEE Trans. PAMI*, vol. 23, no. 3, pp. 297–303, 2001.

13. P. Fua and Y. G. Leclerc, "Object-centered surface reconstruction: Combining multi-image stereo and shading," *Int. J. Computer Vision*, vol. 16, pp. 35–56, 1995.
14. R. T. Collins, "A space-sweep approach to true multi-image matching," in *Proc. Computer Vision and Pattern Recognition Conf.*, pp. 358–363, 1996.
15. 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.
16. S. Roy and I. J. Cox, "A maximum-flow formulation of the N-camera stereo correspondence problem," in *Proc. 6th Int. Conf. on Computer Vision*, pp. 492–499, 1998.
17. O. Faugeras and R. Keriven, "Complete dense stereovision using level set methods," in *Proc. 5th European Conf. on Computer Vision*, pp. 379–393, 1998.
18. K. N. Kutulakos, "Approximate N-View stereo," in *Proc. European Conf. on Computer Vision*, pp. 67–83, 2000.
19. K. N. Kutulakos and S. M. Seitz, "A theory of shape by space carving," *Int. J. Computer Vision*, vol. 38, no. 3, pp. 199–218, 2000. Marr Prize Special Issue.
20. J. D. Bonet and P. Viola, "Roxels: Responsibility-weighted 3D volume reconstruction," in *Prof. 7th. Int. Conf. on Computer Vision*, pp. 418–425, 1999.
21. A. Broadhurst, T. Drummond, and R. Cipolla, "A probabilistic framework for space carving," in *Proc. 8th Int. Conf. on Computer Vision*, pp. 388–393, 2001.
22. R. Bhotika, D. J. Fleet, and K. N. Kutulakos, "A probabilistic theory of occupancy and emptiness," Tech. Rep. 753, University of Rochester Computer Science Department, Dec. 2001.
23. B. K. P. Horn, *Robot Vision*. MIT Press, 1986.

A   Proof Sketches

A.1   Lemma 1: Photo-consistency Probability

We rely on the following result which is stated here without proof (see [22] for details):

Proposition 3 (Conditional Independence of Photo-consistency and Emptiness)   If every pixel observation in I takes RGB values over a continuous domain, the photo-consistency probability of S, conditioned on the visibility function β_S, is independent of the emptiness of $V - S$:

$$p\left[\textbf{PhotoC}(S, I) \mid \textbf{Empty}(V - S, I), V, I\right] = p\left[\textbf{PhotoC}(S, I) \mid I, \beta_S\right]. \qquad (8)$$

To prove the lemma, it therefore suffices to derive an expression for the right-hand side of Eq. (8).

A shape is photo-consistent if and only if every voxel $\mathbf{v} \in S$ is photo-consistent. Considering an arbitrary voxel $\mathbf{v}_1 \in S$ and applying the chain rule gives

$$p\left[\textbf{PhotoC}(S, I) \mid I, \beta_S\right] = p\left[\bigwedge_{\mathbf{v} \in S} \textbf{PhotoC}(\mathbf{v}, I, \beta_S) \mid I, \beta_S\right] =$$

$$p\left[\textbf{PhotoC}(\mathbf{v}_1, I, \beta_S) \mid \bigwedge_{\mathbf{v} \in S - \{\mathbf{v}_1\}} \textbf{PhotoC}(\mathbf{v}, I, \beta_S), I, \beta_S\right] p\left[\bigwedge_{\mathbf{v} \in S - \{\mathbf{v}_1\}} \textbf{PhotoC}(\mathbf{v}, I, \beta_S) \mid I, \beta_S\right].$$

The visibility function β_S defines the visibility state for all voxels in S, including $\mathbf{v}_1$. Now, the photo-consistency probability of $\mathbf{v}_1$, when conditioned on $\mathbf{v}_1$'s visibility state in S, is independent of the photo-consistency probability of all other voxels in S:

$$p\left[\textbf{PhotoC}(\mathbf{v}_1, I, \beta_S) \mid \bigwedge_{\mathbf{v} \in S - \{\mathbf{v}_1\}} \textbf{PhotoC}(\mathbf{v}, I, \beta_S), I, \beta_S\right] = p\left[\textbf{PhotoC}(\mathbf{v}_1, I, \beta_S) \mid I, \beta_S(\mathbf{v}_1)\right].$$

This equation is a precise statement of the visibility conditioning principle. When applied to all voxels in S and combined with Eq. (8), it yields the desired factorization:

$$p\left[\textbf{PhotoC}(S, I) \mid \textbf{Empty}(V - S, I), V, I\right] = \prod_{\mathbf{v} \in S} p\left[\textbf{PhotoC}(\mathbf{v}, I, \beta_S) \mid I, \beta_S(\mathbf{v})\right]. \,\square \qquad (9)$$

Figure 6. A two voxel initial volume, $V = \{\mathbf{v}_1, \mathbf{v}_2\}$, viewed by cameras L and R.

A.2 Lemma 2: Empty Space Events

We prove the lemma by contradiction. In particular, we show that if $V - S$ intersects a photo-consistent shape, i.e., if **Empty**$(V - S)$ is false, no carving sequence $\Sigma \in \mathcal{C}(V, S)$ exists for which **NPhotoC**$(V - S, I, \Sigma)$ is true.

Suppose $V - S$ intersects a photo-consistent shape and let S_1, S_2 be the shape's intersections with $V - S$ and S, respectively. We partition the set of carving sequences, $\mathcal{C}(V, S)$, into two subsets $\mathcal{C}_1, \mathcal{C}_2$ as follows. Let $\mathcal{C}_1$ be the set of those carving sequences in $\mathcal{C}(V, S)$ where every voxel in S_1 appears after the entire set of voxels in $V - (S \cup S_1)$, and let $\mathcal{C}_2 = \mathcal{C}(V, S) - \mathcal{C}_1$.

Now consider any carving sequence $\Sigma^* \in \mathcal{C}_1$. If voxel $\mathbf{v}_k$ is the first voxel in S_1 to appear in Σ^*, this voxel will be photo-consistent with respect to the visibility state β_{S_k}, where $S_k = S_1 \cup S$. This is because the cameras that view $\mathbf{v}_k$ according to β_{S_k} are a subset of the cameras that view it when $\mathbf{v}_k$ lies on the shape $S_1 \cup S_2$, which is both a subset of S_k *and* is photo-consistent. Therefore, **NPhotoC**$(V - S, I, \Sigma^*)$ is not true.

Moreover, voxel $\mathbf{v}_k$ will be photo-consistent in every carving sequence in $\mathcal{C}_2$ in which it appears before every other voxel in S_1. This is because in that case $\mathbf{v}_k$ will be visible from even fewer cameras than those that view $\mathbf{v}_k$ in Σ^*. It follows that the above argument is true for every carving sequence in $\mathcal{C}_1$ and can be extended to cover every sequence in $\mathcal{C}_2$ as well. It follows that the event **NPhotoC**$(V - S, I, \Sigma)$ is not satisfied for any carving sequence in $\mathcal{C}(V, S)$. $\square$

A.3 Emptiness and Inadequacy of Carving Sequences

If it were possible to define the emptiness of a volume $V - S$ simply by the event that at least one carving sequence is not photo-consistent, i.e., **Empty**$(V - S, I) \iff \bigvee_{\Sigma \in \mathcal{C}(V,S)}$ **NPhotoC**$(V - S, I, \Sigma)$, the expression for the photo-hull probability of S would reduce to

$$p\left[\mathbf{Hull}(S, V, I) \mid V, I\right] = \sum_{\Sigma \in \mathcal{C}(V,S)} G(\Sigma \parallel \Pi) R(\Sigma), \tag{10}$$

where Π is a voxel visibility string corresponding to an arbitrary permutation of voxels in S, all of which appear in their **v**-form. Eq. (10) is obtained by setting $\mathcal{P}(\Sigma) = \{\Sigma\}$ in Eq. (7). Despite the fact that Eq. (10) may seem intuitively correct, this equation does not hold in general. This is because the emptiness of $V - S$ does not imply the existence of a non-photo-consistent carving sequence. We prove this fact by contradiction, by showing that the photo-hull probability defined by Eq. (10) is not a distribution in general, i.e., $\sum_{S \in 2^V} p\left[\mathbf{Hull}(S, V, I) \mid V, I\right] < 1$.

Consider the volume and viewing geometry illustrated in Figure 6. Let $V = \{\mathbf{v}_1, \mathbf{v}_2\}$, viewed by two cameras L and R. Voxel $\mathbf{v}_1$ is always visible from L and is occluded from R by $\mathbf{v}_2$, while $\mathbf{v}_2$ is always visible from R and is occluded from L by $\mathbf{v}_1$. The shape space consists of $S_1 = V$, $S_2 = \{\mathbf{v}_1\}$, $S_3 = \{\mathbf{v}_2\}$, and $S_4 = \emptyset$. From Eq. (10), we get the following expressions for the photo-hull probability of these shapes:

$$p\left[\mathbf{Hull}(S_1, V, I) \mid V, I\right] = G(\emptyset \parallel \mathbf{v}_1 \mathbf{v}_2) R(\emptyset),$$
$$p\left[\mathbf{Hull}(S_2, V, I) \mid V, I\right] = G(\overline{\mathbf{v}_2} \parallel \mathbf{v}_1) R(\overline{\mathbf{v}_2}),$$
$$p\left[\mathbf{Hull}(S_3, V, I) \mid V, I\right] = G(\overline{\mathbf{v}_1} \parallel \mathbf{v}_2) R(\overline{\mathbf{v}_1}), \text{ and}$$
$$p\left[\mathbf{Hull}(S_4, V, I) \mid V, I\right] = G(\overline{\mathbf{v}_1}\,\overline{\mathbf{v}_2} \parallel \emptyset) R(\overline{\mathbf{v}_1}\,\overline{\mathbf{v}_2}) + G(\overline{\mathbf{v}_2}\,\overline{\mathbf{v}_1} \parallel \emptyset) R(\overline{\mathbf{v}_2}\,\overline{\mathbf{v}_1})$$

Using Definition 3, the equations above, and the following notational shorthand

$$p_{\mathbf{v}}^{\beta_S} = p\left[\mathbf{PhotoC}(\mathbf{v}, I, \beta_S) \mid I, \beta_S(\mathbf{v})\right],$$

the sum of photo-hull probabilities of all shapes in the shape space is given by

$$\sum_{S \in 2^V} p\left[\mathbf{Hull}(S,V,I) \mid V,I\right] = p_{\mathbf{v_1}}^{\beta_{S_1}} p_{\mathbf{v_2}}^{\beta_{S_1}} + (1 - p_{\mathbf{v_2}}^{\beta_{S_1}})p_{\mathbf{v_1}}^{\beta_{S_2}} + (1 - p_{\mathbf{v_1}}^{\beta_{S_1}})p_{\mathbf{v_2}}^{\beta_{S_3}}$$

$$+ \frac{1}{2}(1 - p_{\mathbf{v_1}}^{\beta_{S_1}})(1 - p_{\mathbf{v_2}}^{\beta_{S_3}}) + \frac{1}{2}(1 - p_{\mathbf{v_2}}^{\beta_{S_1}})(1 - p_{\mathbf{v_1}}^{\beta_{S_2}})$$

$$= 1 - \frac{1}{2}(1 - p_{\mathbf{v_1}}^{\beta_{S_1}})(p_{\mathbf{v_2}}^{\beta_{S_1}} - p_{\mathbf{v_2}}^{\beta_{S_3}}) - \frac{1}{2}(1 - p_{\mathbf{v_2}}^{\beta_{S_1}})(p_{\mathbf{v_1}}^{\beta_{S_1}} - p_{\mathbf{v_1}}^{\beta_{S_2}}).$$

The second and third "residual" terms in the final expression of the above sum will be non-zero in general. This means that $\sum_{S \in 2^V} p\left[\mathbf{Hull}(S,V,I) \mid V,I\right] < 1$, implying that the photo-hull probability of shapes $S \subseteq V$ in Eq. (10) is underestimated. Hence, the emptiness probability of $V - S$ is not completely explained by the non-photo-consistency of the carving sequences in $\mathcal{C}(V,S)$. The two residual terms correspond to the voxel visibility strings $\mathbf{v_2} \overline{\mathbf{v_1}} \overline{\mathbf{v_2}}$ and $\mathbf{v_1} \overline{\mathbf{v_2}} \overline{\mathbf{v_1}}$, respectively. These strings belong to $\cup_{\Sigma \in \mathcal{C}(V,S_4)} \mathcal{P}(\Sigma)$, but were omitted when Eq. (7) was reduced to Eq. (10).

A.4 Proposition 2: Algorithm Fairness

Proof (Sketch). The proof involves four steps: (1) showing that each run of the algorithm is described by a voxel visibility string T of bounded length; (2) defining an onto mapping μ between voxel visibility strings that characterize a run of the algorithm and voxel visibility strings $\langle P \parallel \Pi \rangle$ in Theorem 2; (3) showing that $G(T) = G(\mu(T))$; and (4) using the above steps to prove fairness.

Sketch of Step 1: Define string T to be the exact trace of voxels selected by the algorithm's Voxel Selection Step, where a voxel appears in its $\mathbf{v}$-form if it was selected and not carved in the Stochastic Carving Step immediately following it, or in its $\overline{\mathbf{v}}$-form if it was carved in that step. The algorithm terminates when all surface voxels have been examined and no visibility transitions have occurred for these voxels. Since carving can only increase the number of cameras viewing a voxel, the total number of visibility transitions is $n|V|$. This gives an upper bound on $|T|$.

Sketch of Step 2: Let S be the shape computed by the algorithm and let T be its trace. To construct the string $\mu(T)$ we change T as follows: (i) delete from T all but the last occurrence of every voxel in S, creating string T_1; (ii) move the voxels in S to the end of the string, creating T_2; and (iii) define $\mu(T) = T_2$. String $\mu(T)$ will be of the form $\langle P \parallel \Pi \rangle$, with $P \in \mathcal{P}(\Sigma)$ for some $\Sigma \in \mathcal{C}(V,S)$. This mapping is onto since every string in $\mathcal{P}(\Sigma)$ is a valid trace of the algorithm.

Sketch of Step 3: (i) The visibility state of points in Π that is induced by the string $\mu(T) = \langle P \parallel \Pi \rangle$ is identical to the one assigned by the visibility function β_S. It follows that $G(\Pi) = F(\Pi) = p\left[\mathbf{PhotoC}(S,I) \mid \mathbf{Empty}(V - S, I), V, I\right]$. (ii) Since the voxels in T that are contained in Π always occur in their $\mathbf{v}$-form, their position in T does not affect the visibilities induced by T to those voxels in T that belong to P. Hence, $G(T_2) = G(T_1)$. (iii) If a voxel in $V - S$ appears multiple times in T, this means that it was examined and left uncarved for all but its last occurrence in T. It now follows from the definition of $G(.)$ that $G(T) = G(T_2)$.

Sketch of Step 4: The frequency that the algorithm arrives at shape S is equal to the probability that the algorithm produces a trace in $\cup_{\Sigma \in \mathcal{C}(V,S)} \mathcal{P}(\Sigma)$. It can be shown that Steps 1-3 now imply that the probability the algorithm produces a trace T is equal to $G(\mu(T))R(\mu(T))$, where $R(.)$ is defined in Proposition 1. Intuitively, the $R(\mu(T))$ factor accounts for the portion of all traces T that map to the same $\mu(T)$. Since each trial of the algorithm is independent of all previous ones, it follows that the probability that the algorithm generates shape S is equal to the shape's photo hull probability, as defined in Theorem 2. □