Dealing with Textureless Regions and Specular Highlights—A Progressive Space Carving Scheme Using a Novel Photo-consistency Measure

Ruigang Yang[*][†], Marc Pollefeys[†], and Greg Welch[†]
Department of Computer Science, University of North Carolina at Chapel Hill
Chapel Hill, North Carolina, USA
ryang@cs.uky.edu, marc@cs.unc.edu, welch@cs.unc.edu

Abstract

We present two extensions to the Space Carving framework. The first is a progressive scheme to better reconstruct surfaces lacking sufficient textures. The second is a novel photo-consistency measure that is valid for both specular and diffuse surfaces, under unknown *lighting conditions.*

1 Introduction

There has been a considerable amount of work on volumetric scene reconstruction from multiple views [22, 8, 9, 4, 1, 15, 3]. Most of this work can be considered variations of the Space Carving framework by Kutulakos and Seitz [14]. Under this framework, an initial bounding volume is divided into a regular 3D voxel grid, then inconsistent voxels are removed until the remaining voxels are *photo-consistent* with a set of input images. That is, rendered images of the resulting voxels from each input viewpoint should reproduce the actual image as closely as possible [22].

Because of the flexibility of the volumetric representation and the elegant treatment of visibility, space carving approaches have been used to achieve strong results on a variety of both natural and artificial scenes. However such approaches typically run into difficulty when applied to scenes with textureless or specular surfaces. For one of our driving applications—reconstruction of real surgical procedures for training, such surfaces are the norm rather than the exception. See Figure 1.

At the heart of the space carving algorithm is the photo-consistency test to determine whether or not a voxel should be removed. Most methods make this decision based solely on the input image color samples corresponding to visible voxels. In textureless regions, false positives resulting from ambiguities in *front* of the true surface typically result in extraneous voxels that "fatten" the reconstruction. This effect is particularly pronounced when the model is viewed from an oblique angle, far away from any of the input viewpoints. Additional constraints are often applied in an attempt to resolve the ambiguity. For example a typical approach in stereo vision is to increase the support of the reconstruction kernel. However the accompanying smoothing effect undermines a unique feature of these voxel based methods: the ability to reconstruct highly complex shapes. Instead we want to apply additional constraints *only* when there is ambiguity. To this end, we present a *progressive space carving scheme*. Starting from a few reliable voxels we incrementally add voxels using photo consistency measures, progressively updated visibility information, uniqueness constraints, and smoothness constraints.

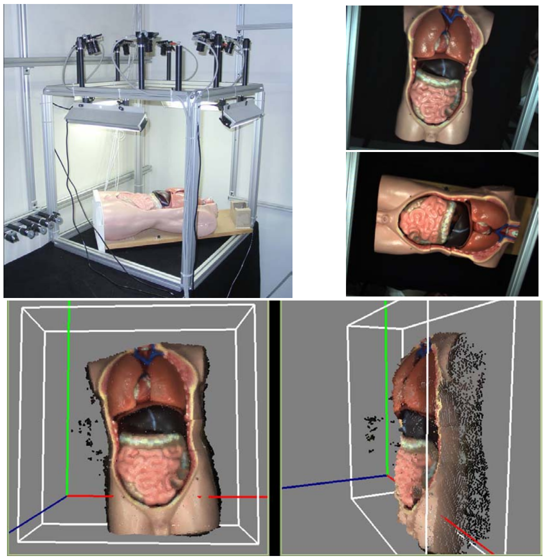


Figure 1. Top left: our one meter-cubed camera rig with eight cameras looking down at a human patient model. Top right: two camera images. Bottom: two views of the reconstructed voxel model. The white bounding box shows the initial volume. Each voxel is rendered as a simple point with color. No interpolation is performed to fill holes.

In addition, most existing space carving methods assume a scene with Lambertian surfaces. This limitation prevents the application of these powerful methods to scenes with

[*] Currently at the University of Kentucky (ryang@cs.uky.edu).

[†] This work was supported in part by the NSF grant ISS-0237533, 0121657, and by a generous 2002-2003 Link Foundation fellowship.

Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 2003) 2-Volume Set
0-7695-1950-4/03 $17.00 © 2003 **IEEE**

specular highlights. Based on the observation that the reflected colors for most real-world surfaces are co-linear in the RGB color space [11], we have developed a *novel photo-consistency measure* that is valid for both specular and Lambertian surfaces. This new measure does not require light calibration or surface normal estimation, thus can be incorporated into *any* existing space carving method to facilitate the reconstruction of highly specular surfaces.

We have implemented our extensions and tested the framework on a number of data sets. We are encouraged by the results. We are able to reconstruct textureless and highly specular surfaces, such as those shown in Figure 1.

2 Previous Work

The problem of multi-view reconstruction has received significant attention during the last few years. In particular, many voxel-based photo-consistency methods have been proposed. Dyer [10] and Slabaugh et.al. [25] each provide comprehensive reviews of recent efforts in this area.

As mentioned in the introduction, there is considerable previous work related to space carving methods [22, 23, 8, 14, 15]. Typically voxels are traversed in a *visibility compatible order*, where only previously committed voxels are allowed to occlude a current voxel. Consider dividing the volume with a plane that separates the cameras from the scene, and then sweeping that plane from near to far (away from the cameras). Any voxel within the plane cannot occlude another. Thus when a voxel v is visited, its visibility in every input image has been uniquely determined. We can then use a photo-consistency measure to decide if v should be carved away or retained. A popular choice is to threshold the variance of the color samples, collected from v's projections in all the visible camera images. Assuming a Lambertian scene, a large variance implies an "inconsistent" voxel. While these methods are very efficient, they are typically sensitive to whatever global threshold was chosen for the photo-consistency evaluation. In practice, as a result of random noise and quantization effects, a single threshold rarely achieves optimal results for a complex scene.

To overcome this problem, some researchers have formulated the voxel reconstruction problem as an energy minimization problem. For example, Slabaugh et al. introduced an iterative post-processing approach [24]. They add or remove surface voxels until the sum of squared differences between each camera image and corresponding model image (rendered from the camera's viewpoint) is minimized. More recently, Kolmogorov and Zabih introduced a graph-cut approach to optimize the volume reconstruction *directly* [13]. In order to make the optimization tractable they approximate the visibility test.

Several probabilistic space carving methods have been introduced [9, 4, 1, 3]. Instead of making "hard" decisions about voxel existence, these approaches compute a per-voxel probability based on appropriate likelihoods. In theory such formulations should consider *all* possible visibility configurations for a voxel. To avoid the combinatorial search, the visibility tests are typically approximated based on heuristics [9, 1, 4] or solved in a stochastic manner through hundreds of iterations [3]. We believe that an accurate treatment of visibility is crucial for a multi-view reconstruction. Our approach progressively reconstructs a voxel model, typically in a few iterations, with visibility computed deterministically and exactly at each iteration.

Although the use of more sophisticated lighting models was envisioned in the original space carving work [14], almost all existing methods use a photo-consistency measure based on a diffuse (Lambertian) surface assumption. Two notable exceptions are the *Surfel* (surface element) sampling algorithm by Carceroni and Kutulakos [6] and the color caching algorithm by Chhabra [7]. The former differs substantially from traditional voxel-based methods. The scene is divided into a very coarse voxel grid, with each voxel represented as a parametric surface referred to as a *surfel*. Under calibrated lighting, additional properties such as surface normal and reflectance parameters can be estimated. Only results from scenes with point light sources were demonstrated. In practice, light calibration is not always possible, especially for area light sources. In the color caching algorithm, Chhabra tries to characterize the reflected light from specular surfaces in the color space. While this analysis is very similar to our thinking, it is restricted to the case where the reflected light passes through the origin of the color cube. This simplification is only valid in some very limited cases, such as monochromatic surfaces under white light. Based on an analysis of the surface color response under the Phong lighting model [19] we arrived at a *general* photo-consistency measure, that when conditioned on some typical surface–light interactions can serve as a maximum likelihood indicator.

There is also some work in stereo vision literature to recover a disparity map in the presence of specular reflections [2, 12, 16, 17]. These methods typically try to first *detect* specular reflections and then *reject* them as outliers or occluders. A notable exception is the work from Magda et al. [18], in which they propose techniques to recover shapes with arbitrary surface reflectance properties under controlled lighting. Our method treats specular reflections as "inliers" and accounts for them inherently, without the need to control lighting.

3 Progressive Space Carving

Kutulakos and Seitz showed that even without additional constraints, the space carving framework will provide the tightest reconstruction using color information alone [14]. They called the recovered shape the *photo hull*. The idea of using color information *alone* has its pros and cons. On the one hand, in regions with rich textures, arbitrarily complex shapes can be recovered. On the other hand, the lack of additional constraints (regularization terms) makes space carving more susceptible to image noise and quantization problems. In addition, the fattening effect in regions that lack color variations sometimes can be disconcerting. In an attempt to preserve the positive and remove the negative characteristics, we employ an iterative approach to progressively refine the shape estimates. The basic idea is to defer decisions about ambiguous voxels until there is enough supporting evidence. The refinement includes new photo-consistency measures under an updated visibility configuration and local smoothness constraints.

Algorithm 1 Pseudo code for a progressive space carving Scheme.

```
 Clear visibility mask $M_I$ for every image I;
 for every voxel v {
    label[v] = UNKNOWN;
    weight[v] = 1.0;
}
while (true) {
    for every UNKNOWN voxel v {
        $s = \varnothing$ ; // s is the visible pixel set
        for every image I {
            $\{p_i\}$ = v's projection in I
            if ($p_i$ visible) $s = s \cup p_i$
        }
        if ($\|s\| < 2$) label[v] = INVISIBLE
        else {
            score[v]=photo_consistency($s$)
                *weight[v];
        }
    }
    n = select_consistent_voxels();
    // no more voxels can be selected, stop
    if (n == 0) break;
    update weight;
    update visibility masks;
}
```

Referring to Algorithm 1 we outline our progressive space carving approach. Each voxel can have one of four labels, UNKNOWN, EMPTY, SURFACE, INVISIBLE. In the beginning, every voxel is labelled UNKNOWN, has a unit weight, and is visible in every input image. At each iteration, we compute the likelihood for each UNKNOWN voxel as the product of its weight and its photo-consistency score. If a voxel can not been seen from at least two views, it is immediately labeled as INVISIBLE. (Note that we can traverse the voxels in any order since the visibility configuration is fixed during an iteration.) When this calculation is complete for all voxels, we find the most unambiguous voxels, changing their labels to EMPTY or SURFACE. The selected voxels are then used to update the visibility configuration and the weights of the other UNKNOWN voxels based on a smoothness constraint. This process is repeated until there are no more UNKNOWN voxels that can be selected.

In the following sub-sections we explain the two central components of our framework: the "best voxel" selection process and the smoothness constraint. Before proceeding we want to point out that the smoothness constraint is optional. Without it the iterative method can be thought of as a space "peeling" algorithm. After the first iteration, any surface voxels visible in all cameras are selected and removed, exposing more surface voxels that are partially occluded. These voxels will be selected in the next iteration, exposing voxels more deeply occluded, etc. After a limited number of iterations, the reconstructed shape will converge to the photo hull. In configurations where a plane-sweep visibility order exists [22], the maximum number of iterations equals the number of voxel planes in the sweep direction.

3.1 Finding the most consistent voxels

Rather than using a simple threshold to decide if a voxel is consistent or ambiguous, we look at the *profile* of a set of related voxels. A pixel p in an image defines a line of sight l; l will intersect a set of candidate voxels, denoted as $\{v_p^i\}$ where i is the index to each voxel along the line of sight. Assuming an opaque scene, at least one voxel in $\{v_p^i\}$ will be the surface voxel reflecting the light that imaged in p, thus it will have the best photo-consistency score if the visibility is solved correctly. Considering the photo-consistency curve of $\{v_p^i\}$, if there is a *single* local maximum, i.e., its consistency value is better than its left and right neighbors (assuming a higher photo-consistency score means better consistency), then the corresponding voxel v is considered consistent and labeled as SURFACE. In addition, any voxel in $\{v_p^i\}$ and in front of v will be labeled as EMPTY, i.e. carved away. If no SURFACE voxel can be found, all voxels in $\{v_p^i\}$ are ambiguous and their occupancies are left to be resolved in later iterations. This scheme does not need a threshold, since we only look for the "best" for each pixel in the input images. It also guarantees the uniqueness constraint, i.e., one pixel only corresponds to one surface voxel.

After the likelihood values for all voxels have been updated, we project the resulting voxel grid onto every input image. (This step can be accelerated using the graphics hardware.) We then find the best voxel for each pixel in every input image. A voxel is labelled SURFACE if and only if it is the best voxel in all views visible. Once a SURFACE voxel is declared, it will be projected into visible views to mask the corresponding pixels—these pixels (in v's footprint) will not participate in future photo-consistency computation, nor will new best voxels be selected for them, i.e., every pixel can only have one corresponding SURFACE voxel. Once a voxel is labeled as SURFACE, it will not be removed in subsequent carving.

3.2 Applying a smoothness constraint

One needs to be careful defining smoothness under a multi-view reconstruction framework. First we note that smoothness is a view-dependent property. Think about a thin sheet of paper: when viewed from the front, the paper is smooth everywhere; when viewed from a 90 degree angle the sheet barely exists. In any case, we believe it makes sense to use smoothness constraints that favor frontal-parallel surfaces, since cameras are more likely to see such surfaces compared to ones at oblique angles. Under a multi-view setting, a different surface assumption can be derived for each input view, but the assumptions need to be consolidated. In this work, we apply a smoothness constraint with respect to every input view and the resulting assumptions are combined in the voxel space, assuming each one is equally likely and valid.

While there are many possible smoothness constraints, we choose to use the *disparity gradient principle* because of its relevance to the human vision system, its simplicity, and its successful use in stereo algorithms. Before we get into the details of our formulation, we first present a brief overview of the disparity gradient principle.

Disparity Gradient Principle Disparity is defined between a pair of rectified stereo images. Given a pixel (u, t) in the first image and its corresponding pixel (u', t') in the second image, disparity is defined as $d = u' - u$. Disparity is inversely proportional to the distance of the 3D point to the cameras. A disparity of zero implies that the 3D point is at infinity.
For two 3D points the disparity gradient can be defined as

$$DG = \left| \frac{\Delta d}{\Delta u - \Delta d/2} \right| \tag{1}$$

where $\Delta u = u_2 - u_1$ and $\Delta d = d_2 - d_1$. Experiments in psychophysics have provided evidence that human perception imposes the constraint that the disparity gradient DG is limited to an upper bound. In [5] the limit $DG < 1$ was reported. The theoretical limit for opaque surfaces is 2, to ensure that the surfaces are visible to both eyes [20]. Also reported in [20], is that under normal viewing conditions most surfaces are observed with a disparity gradient well below the theoretical value of 2.

Applying the Disparity Gradient Principle Manipulating Equation 1 we can arrive at

$$\parallel \Delta d \parallel \leq \frac{\Delta u \cdot DG}{1 - DG/2}, \tag{2}$$

This equation tells us that if one pixel's disparity (depth) is known and DG is limited, then the disparity range of a neighboring pixel is also limited. The closer these two pixels are, the smaller the disparity variation can be. This is why the disparity gradient principle has been used as a smoothness constraint in several stereo matching algorithms [5, 20, 28, 27]. However it has not been applied in a volumetric representation. One problem is that the disparity gradient is defined between a pair of images, so it is not directly applicable in a volumetric setting. Here we introduce a way to relate the disparity gradient to a 3D voxel grid. For each input image, we can assume that there is an imaginary image, taken from a parallel viewpoint some distance away, as shown in Figure 2. This distance should be related to the average distance between neighboring cameras. Assuming a pixel $\vec{m}_1$ corresponding to the voxel v_1, the allowable disparity range for $\vec{m}_2$ given a limit on DG can be obtained from Eq. (2). Then we can back-project the disparity range onto the voxels $\{v_2^i\}$ that intersect the line of sight from $\vec{m}_2$ (the slant-fill voxels in Figure 2). A voxel that is both in the disparity range and in $\{v_2^i\}$ is more likely to be the voxel reflecting light to $\vec{m}_2$. In other word, because v_1 is known to be a surface voxel, the disparity gradient principle tells us that neighboring voxels are also likely to be surface voxels.

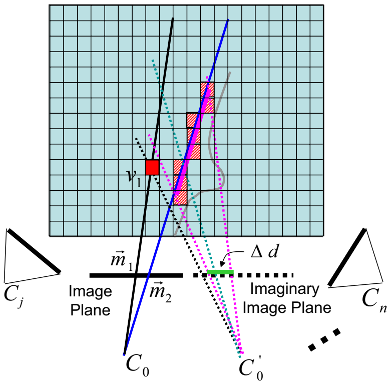


Figure 2. Applying the disparity gradient principle in a volumetric setting. An imaginary image is introduced to define the disparity range. The back-projection of the disparity range limits the voxel search ranges. The gray curve illustrates the weight for voxels.

In practice, there is no need to compute the imaginary image since the location of the iso-disparity planes can be computed directly using $d = fb/Z$ with f the focal length, b the virtual baseline and Z the depth (more precisely the Z-coordinate in a camera centered coordinate frame). We want to favor voxels that have the same depth as v_1, as well as allow small possibility that voxels may fall beyond the disparity range at occlusion boundaries. So we use a weighting function similar to a normal distribution. Let a voxel v_2's projection in image C_0 be $\vec{m}_2$; $\vec{m}_1$ is the closest pixel to $\vec{m}_2$ with a known depth. The disparity difference can be obtained by projecting v_2 and v_1 into the imaginary image C_0' or more directly as $\Delta d = bf(1/Z_2 - 1/Z_1)$.
The weight for v_2 is then given by

$$W_{v_2} = \frac{1}{\sqrt{2\pi}\sigma} \exp(\frac{-(\Delta d/\Delta u)^2}{2\sigma^2}), \tag{3}$$

where $\sigma = \frac{DG}{1-DG/2}$ and $\Delta u = \|\vec{m}_1 - \vec{m}_2\|$. If we know or choose to assume some value for σ, it will be possible to formulate the weighting process using Bayes' rule similar to [28]. If the voxel v_2 is visible in multiple images (as it should be), then the weights from different images can be summed to arrive at a final weight.

4 A New Photo-consistency Measure

Studies in photometry have shown that the reflected light (radiance) from many real-world surfaces can be approximated as the sum of diffuse and specular components [11, 19]. This can be modeled as

$$I = \text{diffuse}(I_p, O_d, N, L) + \text{specular}(I_p, R, V) \quad (4)$$

where I_p is the intensity of a light source, O_d is the object color (albedo), N is the normal vector, L is the lighting vector, R is the reflection vector, and V is the viewing vector.

Now let us examine the change of intensities for a given surface point under different viewing directions. Without lose of generality, we assume that the scene and lighting are both static when images are taken, i.e., N and L are constant. Under this condition, from Equation 4, we can see that the diffuse term remains a constant from any view direction. If the specular effect can be ignored, the color samples from input images will cluster into a point in the color space. That is the basis for the original photo-consistency check proposed by Seitz and Dyer [23]. To check if a voxel exists or not, we simply need to compute the variance of the color samples. We call this the *variance* measure. A large variance indicates they are not likely to be from the same surface point, and thus that voxel should be carved away.

On the other hand, if the specular highlights cannot be ignored, then for a broad class of surfaces such as plastic and glass, the reflected light is only modulated by the incident light. For these surfaces the color values observed from different viewpoints are co-linear in the color space. They form a half-line originating from the diffuse term and extending toward the color of the light I_p. Note that the direction of the line is independent of the object color. The basis of our photo-consistency measure is to detect such a "signature" in the color space. If the surface is diffuse, its signature will be a point; if the surface is specular, its signature will be a line. Because we do not have *a priori* knowledge about whether a voxel represents a specular point or a diffuse point, we want to design a measure that is valid for both cases, while simultaneously providing as much disambiguating power as possible.

Maximum Likelihood Estimation We assume that a color sample can be classified in one of three ways: a diffuse color c_d , an "onset" color c_o, or a saturated color c_s. Each case has a different *a priori* likelihood, denoted $P_d, P_o, and P_s$ respectively. We also assume that the color samples from the images are corrupted by zero-mean gaussian noise with a variance of σ_s^2. For simplicity and robustness, we assume the color of the light is known. (It can be measured by imaging a white object.)

Given an "object" color $\hat{C}$, the likelihood of observing a particular color sample C_j is

$$p(C_j|\hat{C}) = max\begin{pmatrix} N(C_j|\hat{C}, \sigma_s^2)P_d, \\ N(dis(C_j, line(\hat{C}))|0, \sigma_s^2)P_o, \\ N(C_j|C_s, \sigma_s^2)P_s \end{pmatrix}, \quad (5)$$

where N denotes a normal distribution, C_s is the saturation color, usually [1, 1, 1] for normalized RGB images, and $dis(C_j, line(\hat{C}))$ is a function to compute the distance between C_j and a ray defined by $\hat{C}$ and the color of the light. (Note that we interchange C and I because we are dealing with color images.)

Thus the maximum likelihood estimation for an "object" color $\hat{C}$ is

$$max(\prod_{j=1}^{m} p(C_j|\hat{C})). \quad (6)$$

The photo consistency "cost" is the residual after maximum likelihood estimation. The estimated $\hat{C}$ is assigned as the diffuse color of the voxel. We call this measure the *MLE* photo-consistency measure.

Note that in Equation 5, we assume that the distribution along the line is uniform. This is an approximation derived from the Phong lighting model [19]. If we ignore ambient light and the atmospheric attenuation factor (i.e., no fog), the Phong lighting model becomes

$$I = I_p[k_d O_d(NL) + k_s(RV)^n] \quad (7)$$

where k_d is the diffuse coefficient, k_s is the specular coefficient, and n is the object's shininess.

Let the angle between the reflection vector R and the viewing vector V be Θ. It is reasonable to model Θ as a random variable with uniform distribution in its valid range, meaning that a surface point is likely to be viewed from any direction in the hemisphere. Thus the probability density function of Θ is

$$f_\Theta(\theta) = \begin{cases} 1/\pi, & -\pi/2 \le \theta \le \pi/2; \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

Now we need to find the density function of the random variable I. This will tell us how the color samples are distributed along the line. The final result is in Equation 9 (details of the derivation can be found in [26]).

$$f_I(i) = \begin{cases} \frac{2}{\pi} \cdot \frac{k^{\frac{1-n}{n}}}{n\sqrt{1-k^{\frac{2}{n}}}} \cdot \frac{1}{I_p k_s}, & \hat{I} \le i \le \hat{I} + I_p k_s; \\ 0, & \text{otherwise,} \end{cases} \quad (9)$$

where $\hat{I} = I_p k_d O_d(NL)$ and $k = \frac{i-\hat{I}}{I_p k_s}$. In Figure 3, we plot several density curves of I under different shininess (n) settings. In our formulation of our MLE measure, we basically divide the density function into three blocks: c_d corresponds to the left part; c_o corresponds to the middle transition part, which is relatively flat; and c_s corresponds to highlights towards the end of X axis. Note that due to the limited dynamic range of a camera, the saturated class is likely to include more of the middle flat part.

Simplified linearity test The maximum likelihood estimation presented above needs to know the *a priori* likelihoods of the three color sample classes. In case the *a priori* likelihood is unknown or inaccurate, we can use a simplified approach by assuming all samples belong to the onset class, i.e., $P_d = 0, P_o = 1, P_s = 0$. In addition, we assume that less than half of the pixel samples are in specular highlights. Thus we can use the median of the color samples is the object color $\hat{C}$ and simply compute the sum of distances to the ray defined by $\hat{C}$ and the color of the light as

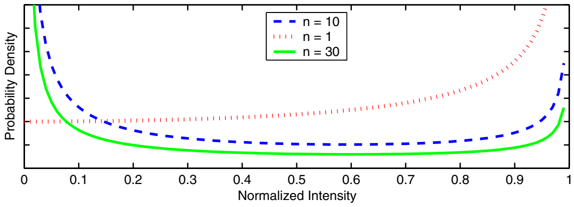


Figure 3. Probability Density Functions under different shininess (n) settings. The X axis is the normalized intensity value (k), while the Y axis is probability density with $I_p k_s = 1$.

the photo-consistency measure. Note that this simplification will fail in smoothly shaded diffuse surfaces with uniform colors. In this case, everywhere is consistent, similar to the result of applying a simple variance test to textureless regions. In practice, we find it still works well on scenes with moderate textures. We call this measure the *LMF* (Line-Model-Fitting) photo-consistency measure. Similarly, if we set $P_d = 1, P_o = 0, P_s = 0$, i.e., only diffuse colors are possible, then the MLE measure reduces to the standard variance measure.

Multiple Stationary Lights The above analysis is valid for multiple fixed lights as long as they have the same color. This is true even if their intensities are different or there are area light sources. If the light colors *are* different, then there will be *multiple* lines in the color space, one for each unique light source. In this case, it would be interesting to investigate if a multiple-line fitting and clustering method would be practical.

Moving Lights and Cameras It is also interesting to consider data captured on a turntable. Typically in this case both the lights and the camera are moving together. If the lights have the same color, the image color samples will form a *plane* instead of a line in the color space. At the same time, if the surface is primarily diffuse, then the color samples will form a half-line starting from the origin and extending toward $I_p k_d O_d$. Thus it is possible to reconstruct diffuse scenes under moving lights and moving cameras. We provide some preliminary results using this finding.

5 Implementation and Results

We have implemented our progressive space carving scheme. We also constructed a capture rig we call the *camera cube* (Figure 1). It consists of eight digital video cameras with VGA resolution (640 × 480). All cameras are fully calibrated and synchronized. We are able to capture and store VGA videos at 15 frames/second. Lens distortions are removed after capture. At this point our implementation only computes the smoothness constraint for a single view. This simplification is justified given our camera arrangement where all 8 cameras look down at the scene from above. In the results shown here, we computed the smoothness constraint from the vertical direction, looking down; similarly, surface voxels were extracted from a top-down vertical sweep. We further define a robust measure to reject potential outliers: if in the color cube space the average distance/pixel to the color ray is over a threshold, we reject that voxel. We used a very generous number—100 levels (assuming 8 bit/channel), and this number was fixed throughout all experiments. Experiments have shown that VDPC is not sensitive to this threshold.

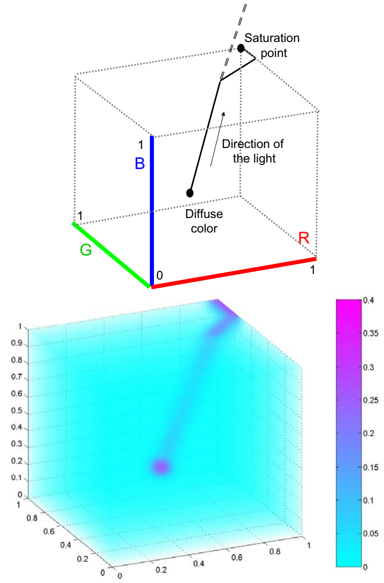


Figure 4. Left: Reflected colors from a surface point forms a line; due to limit dynamic range, it becomes three connected line segments in the RGB cube space . Right: Sample density distribution with object color [0.3, 0.5, 0.1], $\sigma_s = 10/255$, and a priori probability $[0.4, 0.4, 0.2]$

We use *Powell's method* [21] to optimize the function in Equation 6, assuming the color of light is white (which it is). Note that due to the limited dynamic range provided by the cameras, the half-line defined by $\hat{C}$ has three segments in general (left in Figure 4). One channel will first saturate, then the second, and finally the last channel. In Figure 4, we show a sample probability density distribution on the right. Similarly, we compute the distance to the three line segments in the LMF measure.

There are two places where we exploit the computer graphics hardware to accelerate computation. The first place is the visibility update. We render each surface voxel as a cube for every input image, thus we can get the exact footprint to update the visibility mask. The second place is the computation of the smoothness weight. For each voxel, we need to project it to each input view and find the closest pixel that has a corresponding surface voxel. We use *Delaunay* triangulation to create a 2D mesh of marked pixels in each view, and render the color-coded mesh. The rendered image serves as a look-up table to find the closest pixel in $O(1)$ time. We found orders of magnitude speedup after this optimization.

Experiments We captured and reconstructed a variety of real-world scenes. Unless otherwise indicated, all were reconstructed at a resolution of $256 \times 256 \times 128$. We first captured a teapot with rich textures (Figure 5) and tested our photo-consistency measure without applying the smoothness constraint. See Figure 6. Since we know nothing about the surface materials or the scene lighting, except that the color of lights is white, we tried different settings of the *a priori* likelihood (denoted as $\vec{P}$) for the MLE measure. With a 0.1 granularity, there are about 50 different combinations. The most visually pleasing result is shown in Figure 6(a), where $\vec{P} = [0.5, 0, 0.5]$. We also show the result with $\vec{P} = [1.0, 0, 0]$ in Figure 6(b), which is equivalent to using the standard variance measure. Figure 6(c) shows the result with $\vec{P} = [0, 1.0, 0]$, which is very similar to the result from the LMF measure. Comparing to the best one, it has more stray voxels when viewed from the side.

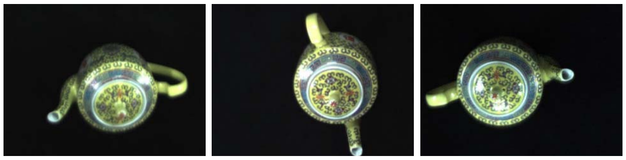
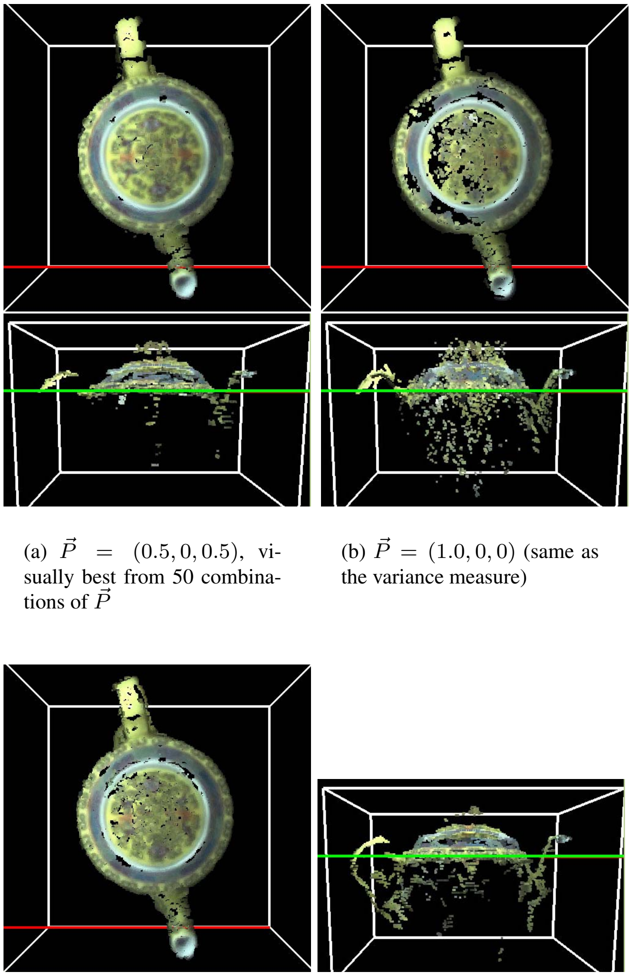


Figure 5. Three of the eight images captured simultaneously by our camera cube (see top left in Figure 1). They are cropped to show more details. The teapot roughly took a 300×300 area in every image.

Our second data set consists of a teapot and a book with substantial textureless regions (Figure 7). We used the LMF measure and applied the smoothness constraint through a few iterations, shown in Figure 8. We stopped at the fourth iteration when newly selected `SURFACE` voxels were less than 2% of the total `SURFACE` voxels. We compare it with the result using the variance measure in Figure 9.

In Figure 10, we show a high resolution reconstruction of a hand. It is computed using the variance consistency measure through four iterations. Note the textureless regions are nicely reconstructed. We further captured a dynamic sequence in which a surgeon was explaining a medical procedure. The torso model was constructed using the LMF measure, while the hand and other moving parts were constructed using the variance measure. They are composited together and shown in Figure 11.

Our last data set, courtesy of the Mitsubishi Electric Research Lab, is a teapot captured on a turntable. The light was not static with respect to the teapot. We did not know this when we first tried our method. In Figure 12, we show the results using the LMF and the variance measure. On the top of the teapot where highlights exist, neither method produces meaningful results. However on the side of the teapot where there are virtually no highlights, the LMF measure performs much better than the variance measure, since under moving lights the reflected lights of a diffuse point form a line (not a point) in the RGB color space.



(a) $\vec{P} = (0.5, 0, 0.5)$, visually best from 50 combinations of $\vec{P}$

(b) $\vec{P} = (1.0, 0, 0)$ (same as the variance measure)

(c) $\vec{P} = (0, 1.0, 0)$(similar to the LMF measure)

Figure 6. Reconstruction results from data in Figure 5. We used the MLE measure under different a priori assumptions, no smoothness weight was applied.

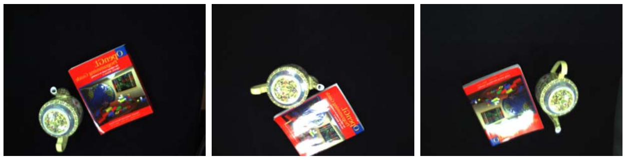


Figure 7. Three of eight images (teapot and book).

6 Conclusions

We present two major extensions to the Space Carving framework. The first is a progressive scheme to better reconstruct surfaces lacking sufficient textures. The second is a novel photo-consistency measure that is valid for both specular and diffuse (Lambertian) surfaces, without the need of light calibration. We applied our method suc-

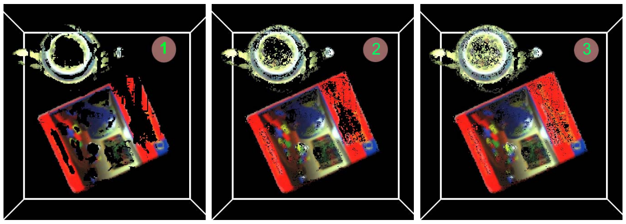


Figure 8. Reconstruction results using data in Figure 7. From left to right, we show the progressively refined results after iteration one to three. We used the LMF consistency measure and set the disparity gradient limit to 0.8.

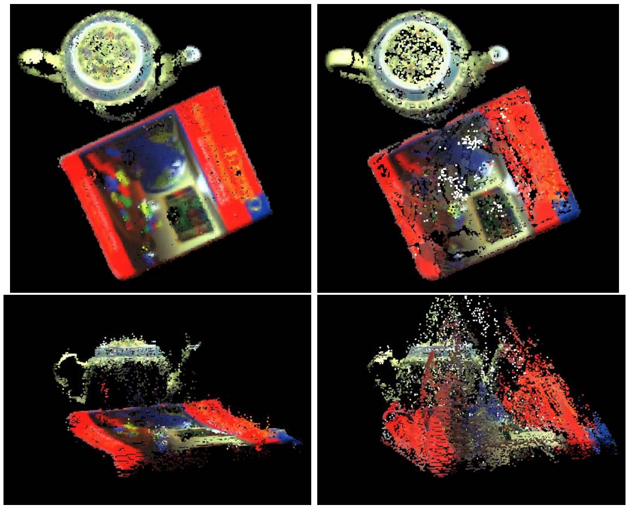


Figure 9. Comparison of different consistency measures. Left column (top and side views): LMF measure; right column: variance measure. Both results were obtained after four iterations. All other parameters were kept the same.

cessfully to a variety of real-world scenes.

Looking to the future, we plan to further investigate the photo-consistency under more complicated lighting conditions, such as moving lights with moving cameras, or multiple light sources with different colors. We also want to explore the possibility of estimating more parameters per voxel, such as surface normals and surface materials, so as to make "re-lighting" possible.

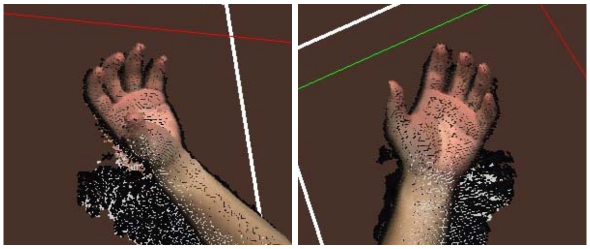


Figure 10. Hi-resolution progressive reconstruction on a $512 \times 512 \times 256$ **grid using the variance measure.**

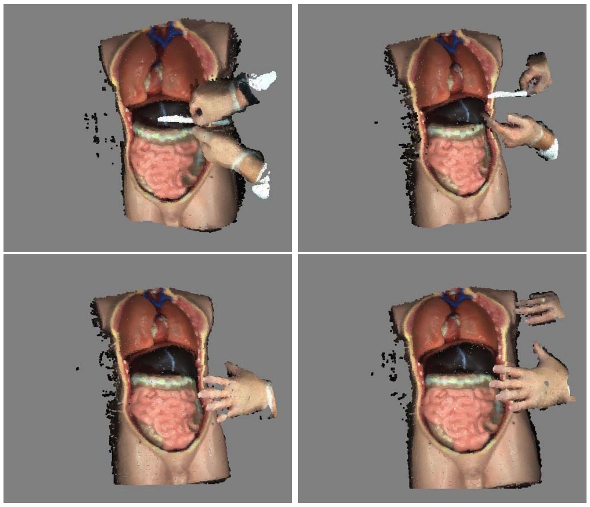


Figure 11. A dynamic sequence captured by the camera cube. A surgeon was explaining a medical procedure. The moving part was reconstructed and rendered with the static model shown in Figure 1.

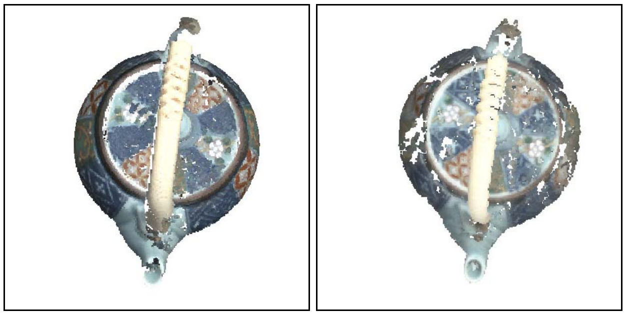


Figure 12. Experiment with moving lights. Left: LMF measure; right: variance measure.

References

[1] M. Agrawal and L. Davis. A Probabilistic Framework for Surface Reconstruction from Multiple Images. In *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001.

[2] D. Bhat and S. Nayar. Stereo in the presence of specular reflection. In *Proceedings of International Conference on Computer Vision (ICCV)*, page 1086C1092, 1995.

[3] R. Bhotika, D. J. Fleet, and K. N. Kutulakos. A Probabilistic Theory of Occupancy and Emptiness. In *Proceedings of European Conference on Computer Vision (ECCV)*, pages 112–132, 2002.

[4] A. Broadhurst, T. Drummond, and R. Cipolla. A Probabilistic Framework for Space Carving. In *Proceedings of International Conference on Computer Vision (ICCV)*, pages 388–393, 2001.

[5] P. Burt and B. Julesz. A Gradient Limit for Binocular Fusion. *Science*, 208:615–617, 1980.

[6] R. L. Carceroni and K. N. Kutulakos. Multi-View Scene Capture by Surfel Sampling: From Video Streams to Non-Rigid 3D MotionShape and Reflectance. In *Proceedings of International Conference on Computer Vision (ICCV)*, 2001.

[7] V. Chhabra. Reconstructing specular objects with image based rendering using color caching. Master's thesis, Worcester Polytechnic Institute, 2001.

[8] B. Culbertson, T. Malzbender, and G. Slabaugh. *Generalized Voxel Coloring*, volume 1883 of *Lecture Notes in Computer Science*, pages 100–115. Springer-Verlag, 1999.

[9] J. de Bonet and P. Viola. Poxels: Probabilistic Voxelized Volume Reconstruction. In *Proceedings of International Conference on Computer Vision (ICCV)*, pages 418–425, 1999.

[10] C. R. Dyer. Volumetric scene reconstruction from multiple views. In L. S. Davis, editor, *Foundations of Image Understanding*, pages 469–489. Kluwer, 2001.

[11] D. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*, chapter 6, page 119. Prentice Hall, 2003.

[12] H. Jin, A. Yezzi, and S. Soatto. Variational multiframe stereo in the presence of specular reflections. Technical Report TR01-0017, UCLA, 2001.

[13] V. Kolmogorov and R. Zabih. Multi-camera Scene Reconstruction via Graph Cuts. In *Proceedings of European Conference on Computer Vision (ECCV)*, pages 82–96, 2002.

[14] K. Kutulakos and S. M. Seitz. A Theory of Shape by Space Carving. *International Journal of Computer Vision (IJCV),*, 38(3):199–218, 2000.

[15] K. N. Kutulakos. Approximate N-view Stereo. In *Proceedings of European Conference on Computer Vision (ECCV)*, pages 67–83, 2000.

[16] Y. Li, S. Lin, H. Lu, S. Kang, and H.-Y. Shum. Multibaseline Stereo in the Presence of Specular Reflections. In *International Conference on Pattern Recognition*, pages 573–576, 2002.

[17] S. Lin, Y. Li, S. B. Kang, X. Tong, and H.-Y. Shum. Diffuse-Specular Separation and Depth Recovery from Image Sequences. In *Proceedings of European Conference on Computer Vision (ECCV)*, pages 210–224, 2002.

[18] S. Magda, T. Zickler, D. Kriegman, and P. Belhumeur. Beyond Lambert: Reconstucting Surfaces with Arbitrary BRDFs. In *Proceedings of International Conference on Computer Vision (ICCV)*, pages 297–302, 2001.

[19] B.-T. Phong. Illumination for computer generated pictures. *CACM*, 18(6):3111–317, June 1975.

[20] S. Pollard, J. Porrill, J. Mayhew, and J. Frisby. Disparity Gradient, Lipschitz Continuity, and Computing Binocular Correspondance. In O. Faugeras and G. Giralt, editors, *Robotics Research: The Third International Symposium*, volume 30, pages 19–26. MIT Press, 1986.

[21] M. Powell. A Fast Algorithm for Nonlinearly Constrained Optimization Calculations. *Numerical Analysis*, 630, 1978. Lecture Notes in Mathematics, Springer Verlag.

[22] S. M. Seitz and C. R. Dyer. Photorealistic scene reconstruction by voxel coloring. In *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1067–1073, 1997.

[23] S. M. Seitz and C. R. Dyer. Photorealistic Scene Reconstruction by Voxel Coloring. *International Journal of Computer Vision (IJCV),*, 35(2):151–173, 1999.

[24] G. Slabaugh, B. Culbertson, T. Malzbender, and R. Schafer. Improved Voxel Coloring via Volumetric Optimization. Technical Report 3, Center for Signal and Image Processing, Georgia Institute of Technology, 2000.

[25] G. Slabaugh, B. Culbertson, T. Malzbender, and R. Schafer. A Survey of Methods for Volumetric Scene Reconstruction from Photographs. 1, Center for Signal and Image Processing, Georgia Institute of Technology, 2001.

[26] R. Yang. *View-Dependent Pixel Coloring – A Physically-Based Approach for 2D View Synthesis*. PhD thesis, Depart of Computer Science, University of North Carolina at Chapel Hill, 2003.

[27] R. Yang and Z. Zhang. Eye Gaze Correction with Stereovision for Video-Teleconferencing. In *Proceedings of European Conference on Computer Vision (ECCV)*, pages 479–494, 2002.

[28] Z. Zhang and Y. Shan. A Progressive Scheme for Stereo Matching. In M. P. et al, editor, *Springer LNCS 2018: 3D Structure from Images - SMILE 2000*, pages 65–85. Springer-Verlag, 2001.