2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission

An approach for shape from surface normals with local discontinuity detection

Yilin Wang Enrique Dunn Jan-Michael Frahm
Department of Computer Science
University of North Carolina
Chapel Hill, USA
Email: {ylwang,dunn,jmf}@cs.unc.edu

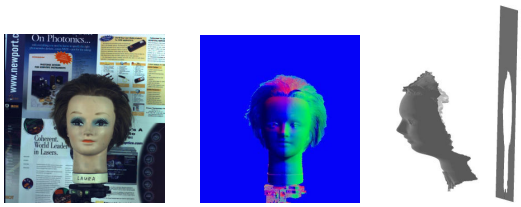
***Abstract*—We present a multi-modal surface reconstruction approach, which utilizes direct surface orientation measurements along with luminance information to obtain high quality 3D reconstructions. The proposed approach models local surface geometry as a set of intersecting natural cubic splines estimated through least squares fitting of our input pixel-wise surface normal measurements. We use this representation to detect discontinuities and segment our scene into disjoint continuous surfaces, which are constructed by an aggregation of connected local surface geometry elements. In order to obtain absolute depth estimates, we introduce the concept of multi-view patch sweeping, where we search for the most photo-consistent patch displacement along a viewing ray. Our approach improves on existing shape from normals methods by enabling absolute depth estimates for scenes with multiple objects. Furthermore, in contrast to existing multi-view stereo methods, we are able to reconstruct textureless regions through the propagation of relative surface orientation measurements. Experiments on synthetic and real data are presented to validate our proposal.**

***Keywords*-shape from surface normal, discontinuity detection, multi-view patch sweeping.**

I. Introduction

State of the art reconstruction approaches can utilize spatial properties, such as surface normals and depth gradients, to recover fine shape details. In photometric stereo [1], [2], [3], [4], surface normals are usually obtained from illumination cues, and then used to generate high quality surfaces. However, photometric stereo relies on knowing (or estimating) the directions of the light sources while making assumptions regarding their relative intensity. Light source calibration errors or other deviations from the assumed source properties may result in unacceptable distortions [5].

Another practical limitation of many photometric approaches is that they work with a single orthogonal image as input, so the reconstructed surfaces have no information about the absolute depths. Moreover, the surface integrability assumption used in many surface from normal algorithms imposes a bias towards single object scene geometries. Accordingly, their application to scenes comprising multiple objects is not straightforward, given that depth discontinuities are difficult to model in this context. In contrast, plane-sweep-based stereo [6], [7], [8] is an efficient approach to recover absolute depths based on photo-consistency measurements. However, the availability of sufficiently textured surfaces is a requirement not always fulfilled in real-world scenarios. Nevertheless, the integration of both of these complementary surface estimation approaches is a promising research path.



(a) (b) (c)

Figure 1. Multi-modal data fusion. Our approach uses (a) color images and (b) normal maps to estimate (c) reconstructed surfaces.

Recently, a novel spatial phase camera has been developed [9], which is capable of passively recording an object's surface normals with an accuracy of two 2 degrees of orientation (as shown in Figure 1b). This new sensing device measures the orientation of the last surface a viewing ray was reflected from as a function of the phase of the light wavelength captured by each CCD array element. Accordingly, we obtain a 3D orientation vector for each pixel in addition to the luminance information. In this work, we leverage this new sensing modality by developing a multi-modal surface reconstruction approach, which efficiently integrates measurements of the orientation and texture of multiple disjoint surfaces.

The proposed approach models local (patch) surface geometry as a set of intersecting natural cubic splines estimated through least squares fitting of our surface orientation measurements. The least squares formulation enables us to identify discontinuities in our local surface estimates, which are then used to obtain an online scene depth segmentation. From this segmentation, each disjoint scene region is then generated by aggregating connected patches into a continuous surface. In order to obtain absolute depth measurements, we introduce the concept of multi-view patch sweeping, where we search for the most photo-consistent patch displacement along a viewing ray. In this way, our approach is able to 1) detect the surface discontinuities, 2) recover the absolute depths based on local information

978-0-7695-4369-7/11 $26.00 © 2011 IEEE
DOI 10.1109/3DIMPVT.2011.31

and 3) retain the fine surface detail given by pixel-wise orientation measurements. A reconstructed sample is shown in Figure 1c.

The remainder of the paper is organized as follows: first, a review of related surface from normal methods is presented. Then, we propose a method to recover the local surface patch from surface normals using natural cubic splines. Next, we discuss how to expand the estimated local patches to obtain global relative surfaces. Afterwards, we introduce a patch sweeping approach to estimate absolute surfaces. We finalize the paper by presenting our experimental results and conclusions.

II. Related work

Given the normal map $N = \{(n_x, n_y, n_z)\}$ defined over a pixel grid, we can obtain a gradient map $G = \{(g_x, g_y)\}$ where $g_x = -\frac{n_x}{n_z}$ and $g_y = -\frac{n_y}{n_z}$. The generating surface $Z(x, y)$ may then be obtained by minimizing the functional [2]

$$\int\int ((Z_x - g_x)^2 + (Z_y - g_y)^2) dxdy \tag{1}$$

Similar integral approaches [10], [11] share the common underlying assumption that the surface is uniformly integrable ($Z_{xy} = Z_{yx}$), i.e. the second partial derivatives are independent of the order of differentiation. Without this constraint, Equation 1 may have an infinite number of solutions. However, the integrability constraint is rarely satisfied in practice due to the presence of sharp edges and occlusion boundaries.

A number of approaches [3], [12], [13], [14] project the gradient field onto a finite set of integrable basis functions $\phi(\cdot)$ in order to apply an integral method on a possibly non-integrable surface. In this case, the surface $Z(x, y)$ is represented as

$$Z(x, y) = \sum_{\omega \in \Omega} C(\omega)\phi(x, y, \omega) \tag{2}$$

where $\omega = (\omega_x, \omega_y)$ is a two-dimensional index, Ω is a finite set of indexes and the integrability of the surface $Z(x, y)$ is contingent on each $\phi(x, y, \omega)$ being integrable. In this scenario, the problem becomes to find the coefficients $C(\omega)$ that minimize the distances between the given possibly non-integrable surface and the approximated integrable surface. Frankot and Chellappa [3] used Fourier basis functions, which are fast and robust to noise. Later, wavelet basis functions were adopted by Karacali and Snyder [13], who point out that the uniform integrability assumption will become invalid and cause significant distortions due to the impact of unknown discontinuities (edges and occlusions). They relaxed the constraint to partial integrability, and provide a global method to detect and localize unknown edges in the gradient field. The key to the edge detection is to examine the difference between the given gradient field and the gradient field obtained by the uniform integrability assumption.

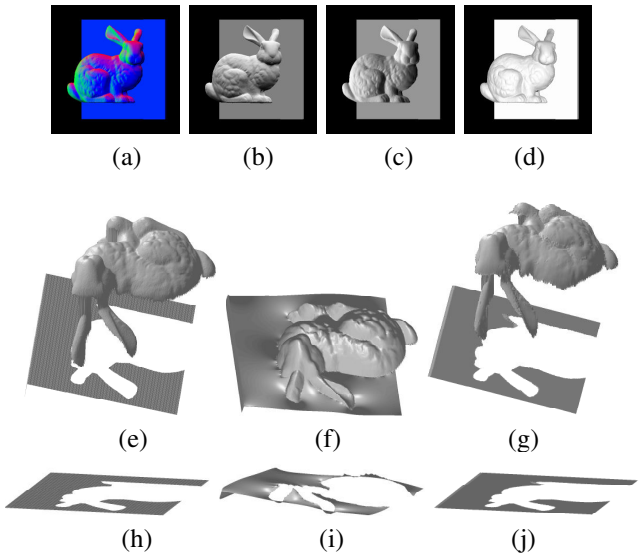


Figure 2. Distortion caused by discontinuities. Top: (a-d) normal map and its 3D components. Middle: scene geometry for (e) the ground truth, (f) the approach presented in [19] and (g) our approach. Bottom: (h-j) corresponding geometry for the background plate. Note the distortions in (i) generated by the integration approach [19].

However, the method is complicated and the computational load in constructing the feasible gradient space descriptors grows rapidly with increasing image size. Kovesi [14] uses shapelets (a redundant non-orthogonal set of basis functions) to enhance sharp transitions in the surface. However, the reconstructed shape is sensitive to the scale of the shapelet function being used, whose optimal determination is still an open problem.

More recent work [4], [15], [16] uses multiple images to obtain the shape with absolute depths. Nehab et al. [15] concluded that the prevailing errors in measured normals are low-frequency in nature, whereas positions measured by stereo triangulation contain mostly high frequency noise. The method first corrects the bias in the normals by low-pass filtering, then it formulates the corrected normals and measured positions as an optimization problem, where the objective function is a large sparse linear system. Chow et al. [4] estimate the absolute depths for the singular points in intensity images by a novel sparse matching technique, then the surface is expanded from these singular points by the fast matching method. Lee and Kuo [16] establish a unified framework for the integration of photometric ratio and stereo by employing perspective projection on a parametric surface via minimizing a cost functional, which consists of a weighted sum of shading and stereo errors. Also, multi-view photometric stereo [17], [18] utilizes illumination conditions, silhouettes and the shading cues to recover the absolute depths.

The aforementioned methods do not provide an explicit and efficient mechanism to detect and treat discontinuities

in the scene. Besides loosing the relative depth between the objects, it leads to heavy distortions near the joint boundaries. To exemplify, a simple scene consisting of a bunny and a plate is shown in Figure 2, and the shape generated by the diffusion method for shape from normals [19] is compared with the ground truth and our method. We can see the plate reconstructed by the traditional method is very jagged. Moreover, since the two objects are treated as a uniform surface, it is impossible to recover the true relative distances between them. We also note that viewing geometry and object topology are other possible causes of surface ambiguity, as some object self occlusions (and the corresponding depth discontinuities) can not be properly reconstructed by a surface integration approach. For example, for a fronto-parallel view of our scene, the bunny's ears in the ground truth model will not be directly connected to the head because of occlusions, but in the diffusion reconstructed surface the two components are smoothly connected. The methods proposed in this paper addresses all these limitations through a multi-modal approach.

Traditional stereo approaches rely on finding dense correspondences among images using local photo-consistency to discriminate among alternative matching hypotheses. For most stereo approaches, the photo-consistency evaluation is performed over an arbitrary neighborhood surrounding a pixel in the reference and a displaced pixel neighborhood in the matching image. The work of Yang et al. [7] applies this concept to multiple images by applying an affine texture transformation to the matching images in order to efficiently model a plane sweeping through the space of depth hypotheses. The affine transformation corresponds to a plane orthogonal to the camera viewing axis. Gallup et al. [8] later improved upon this approach by adaptively selecting a reduced set of main sweeping directions in accordance to sparse depth estimates from an independent structure from motion process. In that work, it was argued that aligning the sweeping direction with the main surface orientation improved accuracy by reducing the depth discretization effect caused by slanted planar surfaces not complying with the local fronto-parallel assumption. We go a step further by modeling the local surface of the matching neighborhood and using this parametric surface to sweep the space of depth hypotheses. This provides more accurate and robust depth estimates than the aforementioned plane sweeping approaches, as will be presented later in Figure 4.

III. Natural Cubic Spline Patches

This section presents the method by which we use normal surface measurements to estimate local surface geometry in the neighborhood of a single pixel. We model local surface as a continuous parametric function, which can be approximated by a set of intersecting 3D cubic splines defined over an image patch. Moreover, we use the surface normal measurements to formulate a linear system of equations from which we can determine least square depth estimates for each pixel in the image patch being considered.

For an image patch of size $k \times k$, the pixels in the same row or same column form k horizontal and k vertical curves. Note that the terms horizontal and vertical refer to the orientation on the image, as these are actually 3D curves generated by the object surface. Suppose the ith horizontal curve C_i^h consists of points $P_{i,j} = [x_{i,j}, y_{i,j}, z_{i,j}]$, $(j = 1, ..., k)$. Then, the segment between point $P_{i,j}$ and $P_{i,j+1}$ can be approximated by a natural cubic spline as follows:

$$\begin{aligned} X_{P_{i,j},P_{i,j+1}}(\lambda) &= a_{i,j,0} + a_{i,j,1}\lambda + a_{i,j,2}\lambda^2 + a_{i,j,3}\lambda^3 \\ Y_{P_{i,j},P_{i,j+1}}(\lambda) &= b_{i,j,0} + b_{i,j,1}\lambda + b_{i,j,2}\lambda^2 + b_{i,j,3}\lambda^3 \\ Z_{P_{i,j},P_{i,j+1}}(\lambda) &= c_{i,j,0} + c_{i,j,1}\lambda + c_{i,j,2}\lambda^2 + c_{i,j,3}\lambda^3, \end{aligned} \quad (3)$$

where $a_{i,j,m}, b_{i,j,m}, c_{i,j,m}$ are the coefficients for the spline function, and $\lambda \in [0, 1]$ is a scalar that parameterizes the traversal of a 3D point along a curve segment with $P_{i,j}$ and $P_{i,j+1}$ as its end points.

According to the spline continuity constraints (0th, 1st, and 2nd order), and assuming the second derivatives at the end points are 0, the spline coefficients $a_{i,j,1}$ can be obtained from the following system of linear equations:

$$\begin{bmatrix} 2 & 1 & & & \\ 1 & 4 & 1 & & \\ & \ddots & & & \\ & & 1 & 4 & 1 \\ & & & 1 & 2 \end{bmatrix} \begin{bmatrix} a_{i,1,1} \\ a_{i,2,1} \\ \vdots \\ a_{i,k-1,1} \\ a_{i,k,1} \end{bmatrix} = 3 \begin{bmatrix} x_{i,2} - x_{i,1} \\ x_{i,3} - x_{i,1} \\ \vdots \\ x_{i,k} - x_{i,k-2} \\ x_{i,k} - x_{i,k-1}. \end{bmatrix} \quad (4)$$

The coefficients $b_{i,j,1}$ and $c_{i,j,1}$ are obtained by similar formulations. Accordingly, the surface tangent vector $\overrightarrow{t_h}$ at point $P_{i,j}$ on the ith horizontal curve C_i^h is defined by

$$\begin{aligned} & \overrightarrow{t_h}(P_{i,j}) \\ = \; & [\frac{\partial X_{P_{i,j},P_{i,j+1}}}{\partial \lambda}, \frac{\partial Y_{P_{i,j},P_{i,j+1}}}{\partial \lambda}, \frac{\partial Z_{P_{i,j},P_{i,j+1}}}{\partial \lambda}] \mid_{\lambda=0} \\ = \; & [a_{i,j,1}, b_{i,j,1}, c_{i,j,1}] \\ = \; & f(x_{i,1}, y_{i,1}, z_{i,1}, ..., x_{i,k}, y_{i,k}, z_{i,k}) \end{aligned} \quad (5)$$

where f is a linear function and $x_{i,j}, y_{i,j}, z_{i,j}$ are the coordinates of the control points $P_{i,1}, ..., P_{i,k}$. For perspective images with a known intrinsic calibration matrix K, the 3D position of the pixel is determined by the depth $z_{i,j}$. Hence, $\overrightarrow{t_h}(P_{i,j})$ can be simplified to a function with k unknowns $z_{i,1}, ..., z_{i,k}$.

Let $\overrightarrow{n}(P_{i,j})$ be the normal measured at point $P_{i,j}$, there are two linear functions for depths $z_{i,1}, ..., z_{i,k}$ and $z_{1,j}, ..., z_{k,j}$:

$$\begin{aligned} \overrightarrow{n}(P_{i,j}) \cdot \overrightarrow{t_h}(P_{i,j}) &= \overrightarrow{n}(P_{i,j}) \cdot f_h(z_{i,1}, ..., z_{i,k}) = 0 \\ \overrightarrow{n}(P_{i,j}) \cdot \overrightarrow{t_v}(P_{i,j}) &= \overrightarrow{n}(P_{i,j}) \cdot f_v(z_{1,j}, ..., z_{k,j}) = 0 \end{aligned} \quad (6)$$

Equation 6 can be extended to a linear function for all the depths in the patch (k^2 unknowns: $z_{1,1}, ..., z_{k,k}$). Given

that each surface point is found at the intersection of two curves, there are two perpendicular tangents per point. So we can obtain $2k^2$ linear equations:

$$\underbrace{\begin{bmatrix} m_{1,1} & m_{1,2} & \dots & m_{1,k^2} \\ m_{2,1} & m_{2,2} & \dots & m_{2,k^2} \\ & \vdots & & \\ m_{2k^2,1} & m_{2k^2,2} & \dots & m_{2k^2,k^2} \end{bmatrix}}_{M} \begin{bmatrix} z_{1,1} \\ z_{1,2} \\ \vdots \\ z_{k,k} \end{bmatrix} = 0 \quad (7)$$

where $m_{i,j}$ are constants for given intrinsic matrix and patch size.

Let M denote the coefficient matrix in equation 7, and $M = USV^T$, where S is a diagonal matrix, and U,V are orthogonal matrices. The last column of V is a normalized nonzero solution for equation 7. Then, if we know the absolute depth for one point, we can recover the absolute depths for the whole patch.

In our approach, the relationship between depth and normals is approximated by a linear function defined over a set of cubic spline coefficients. This modeling offers better accuracy and robustness than using an explicit linear relation among depth and normals. Moreover, by solving an over-determined system of equations in a least square manner, we are able to readily obtain a measure of the reliability of our estimate. In turn, unreliable relative depth estimates (i.e. with a poor fit) are identified as discontinuous surface patches. An examination of the minimum eigenvalue of the coefficient matrix in equation 7 is a straightforward criterion for the correctness of our estimates. However, a more geometrically meaningful criterion for discontinuity detection is to compute the difference between the estimated and observed normals. We define the normal difference by

$$D_{norm} = \sum_{1 \le i,j \le k} \frac{\| \overrightarrow{n}(P_{i,j}) - \overrightarrow{t_v}(P_{i,j}) \times \overrightarrow{t_h}(P_{i,j}) \|}{k^2}, \quad (8)$$

and use it as the quantitative criterion to evaluate the recovered depths. If the normal difference is greater than a pre-defined threshold, it means the patch is not able to be approximated by our natural cubic spline model and it is highly probable that the current image patch contains a discontinuity.

IV. Relative surfaces through patch aggregation

This section presents an efficient method to generate continuous 3D surfaces in the scene by a process of selectively aggregating and segregating local surface patches. The method introduced in Section III estimates the local geometry of a 3D surface patch by solving a linear system with k^2 unknowns. The quadratic growth of our linear system of equations renders the use of the patch estimation method across an entire high resolution image as not computationally tractable. Moreover, any violations of the surface continuity assumption will cause severe distortions in our depth estimation.

In order to efficiently estimate an image-wide relative surface, we propose a region growing procedure that aggregates connected cubic spline patches to form disjoint surfaces whose relative depth is consistent with the input surface orientation measurements. Starting from one *seed* pixel, a local cubic spline patch is estimated and the surface is expanded by attaching neighboring patches sequentially. The expansion of a given surface will stop at discontinuous regions (boundaries, edges or occlusions) and once a given surface can no longer be expanded in any direction, a new *seed* is selected. This process is repeated until all pixels in the image have been examined, yielding a segmentation of the scene into disjoint relative surfaces.

Suppose a is an arbitrary pixel in the image, and its corresponding $k \times k$ surface patch is A. Notice that in this step, we only consider continuous patches, so if A is discontinuous, we drop a and randomly select a new pixel. We assume the depth of a is 1, and use it as the initial element in the fixed point set $F = \{a\}$. Suppose b is one of a's neighbors and its corresponding patch is B. Let $d(x, Y)$ be a function that returns the estimated depth for a point x in Patch Y. By attaching patch B to A, the depth of point b will be scaled by s, where $s = \frac{d(b,A)}{d(b,B)}$. Then the attachment is validated according to the depth difference of a: $V_a = d(a, A) - s \cdot d(a, B)$. If V_a is small, which means the depth ratio of a to b estimated by two patches are similar, then the attachment is valid; otherwise at least one approximated patch is incorrect, so the attachment is invalid and point b is dropped. From these elements we define our depth scaling function as

$$s = \frac{1}{|F \cap P|} \sum_{x \in F \cap P} \frac{d(p, X)}{d(p, P)}, \quad (9)$$

where F is the current fixed point set, p is a new point to attach, and P is the patch generated with p as the patch center. In addition, the function used to validate the correctness of the aggregation of a particular patch to our existing surface estimate is defined by

$$V(s) = \frac{1}{|F \cap P|} \sum_{x \in F \cap P} (d(x, F) - s \cdot d(x, P))^2. \quad (10)$$

The attachment process repeats until all the points are added to the fixed point set or dropped at the end of this process, and then all the points in the fixed set form a continuous surface segment. In the same way, all the pixels in the image can be clustered to certain disconnected segments. The entire procedure is listed in Algorithm 1.

An example for a relative depth map and surface is shown in Figure 3. Compared to existing integral approaches, our method is more robust because the attaching operation usually involves more than one neighbors' information, and the validation operation guarantees that the attached patch will

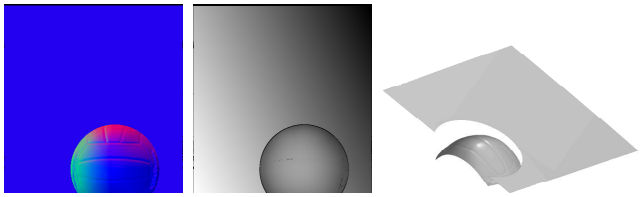


Figure 3. Relative depth map and surface generated from a single normal map.

Algorithm 1 Relative Surface Generation

$U = \{p_1, ..., p_N\}$: initial set of all pixels to be attached
$D = \{0, ..., 0\}$: initial set of depth for all the pixels
$Segs = \{\}$: initial set of recovered surface segments
$si = 0$: index for the current segment
while $U \neq \{\}$ **do**
  $p = U\{1\}$
  $U = U - \{p\}$
  generate patch P for p
  if P is discontinuous **then**
    $continue$
  end if
  $si+ = 1$
  $S_{si} = \{p\}$: set of fixed pixels for current segment
  $D_p = 1$
  N_{si}: set of neighbor pixels for S_{si}
  $C_{si} = U \cap N_{si}$
  while $C_{si} \neq \{\}$ **do**
    $q = C_{si}\{1\}$
    $C_{si} = C_{si} - \{q\}$
    generate patch Q for q
    if Q is discontinuous **then**
      $U = U - \{q\}$
      $continue$
    end if
    computer depth d for q by function (9)
    validate d by function (10)
    if d is valid **then**
      $S_{si} = S_{si} \cup \{q\}$
      $D_q = d$
      $U = U - \{q\}$
      update N_{si} and C_{si}
    end if
  end while
  $Segs\{si\} = S_{si}$
end while
return D and $Segs$

coincide with existing patches. Additionally, the validation operation will detect the discontinuity on the surface to prevent integrating disconnected segments. Although the generated related surface will vary for different expanding paths, the variation is bounded because of the validation operation.

V. Absolute Depth through Patch Sweeping

After obtaining the connected surfaces, the next step is to obtain their absolute depth within the scene. An estimate of the absolute depth of a single pixel is sufficient to propagate depth information through an entire relative surface. In our approach, multiple 3D points are used to make the absolute depth estimate for a given surface more robust

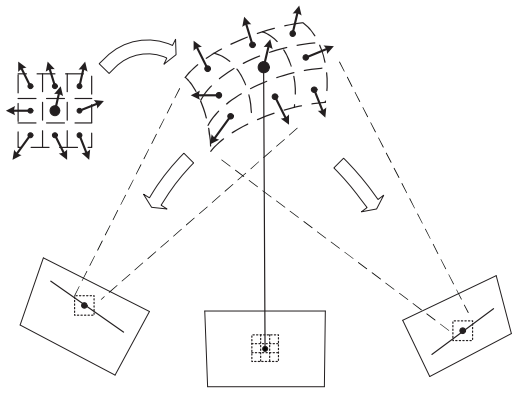


Figure 4. Patch sweeping.

and mitigate any error propagation effects introduced in our surface generation method. Moreover, we estimate the absolute depth for a subset of points in the scene and adjust the relative surface to which they belong accordingly.

Feature points in the luminance image are identified using SIFT [20]. We generate the corresponding cubic spline patch to the SIFT feature and sweep the patch along its viewing ray, as shown in Figure 4. By sweeping the patch along the ray and projecting back to other images, we can accurately estimate the depth for the patch. The patch sweeping approach proposed here is similar to the plane sweeping [6], [7], [8], but points are no longer required to lie on a common plane. Moreover, by sweeping a parametric surface patch instead of a plane, we can model self-occlusions caused by local surface geometry. To reduce the search space, we assume the pixel of interest (not limited to the patch center) can be seen in all the images, then the search space is limited by the boundaries of all the images. The projection of the reduced search space on each image is a segment along the epipolar line, and each pixel on the line segment delivers a depth interval.

The absolute depths obtained by patch sweeping are then used to refine the surface segments. We foresee that patch sweep stereo may provide erroneous estimates due to the limitations common to local stereo estimation approaches (e.g. foreground over-extension, multi-modal photo-consistency function). We deploy a RANSAC based data filtering approach to discard outlier depth estimates that would corrupt the absolute depth of our continuous relative surfaces. In fact, we use the pre-computed relative surface estimate as our fitting model. In this way, we obtain a filtered set of n feature points $p_1, ..., p_n$ with known depths $d_1, ..., d_n$ in the same segment. For each point x with relative depth d_r, its absolute depth is computed by equation 11.

$$d(x) = \frac{1}{\sum_{j=1}^{n} \frac{1}{\| x - p_j \|}} \sum_{i=1}^{n} \frac{d_i \cdot d_r}{\| x - p_i \|} \tag{11}$$

VI. Experiments

We evaluated our approach on both synthetic and real data containing disconnected objects. The synthetic data contains the Stanford Bunny and the Happy Buddha models [21], whose surfaces are formed by about 1 million triangles. Each endpoint of the triangles is assigned a random value $[0, 1]$, so the color of the triangle is the mean of its three endpoints' color, and the normal is computed by the cross product of two edges. We set 36 viewpoints surrounding the models, with a fixed intrinsic matrix and known extrinsic matrix (rotation and translation of the camera). At each viewpoint we generate 1 perspective grayscale image and 1 normal map by projecting the models to the image plane. In this experiment we choose six typical views, three front and three back images (Figure 5). The experiments on real data are captured by the spatial phase camera provided by Photon-X [9], and consist of three grayscale images (left, middle,right) and one normal map for the middle image (Figure 6).

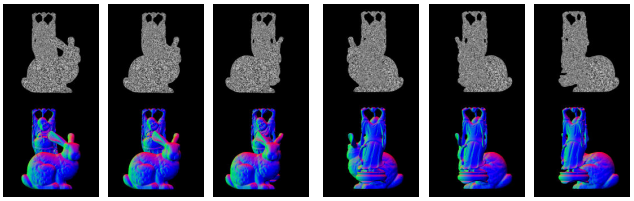


Figure 5. Grayscale and normal images for synthetic data.

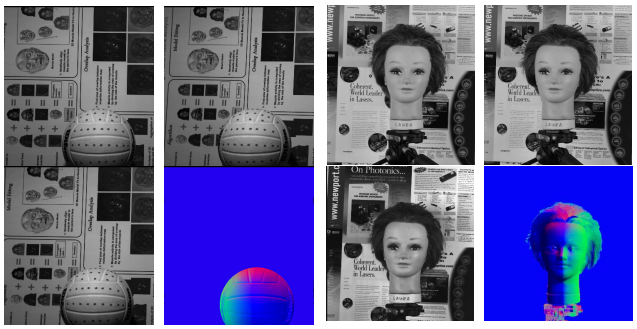


Figure 6. Grayscale and normal images for real data.

We initially evaluate the generated depth maps using synthetic data. The model size is $215mm \times 115mm \times 205mm$, and the program runs several times with different patch sizes: 3×3, 5×5, 7×7, and 9×9. The depth maps and errors for synthetic data are shown in Figure 7 and Table 1, where we observe that greater overall reconstruction accuracy can be achieved by using the two smallest patch sizes. This is mainly attributed to improved local surface approximations near discontinuity boundaries. We also note that small patch sizes are computationally less demanding as the number of equations is proportional to the number of pixel in the patch. Accordingly, in following experiments, the default patch size is 3×3.

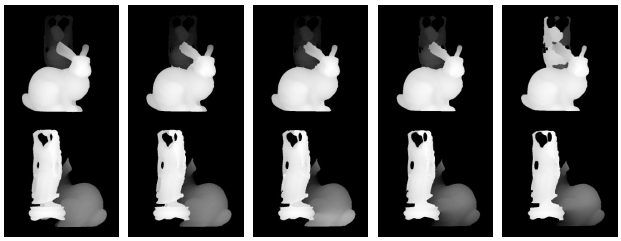


Figure 7. Depth maps for synthetic data. From left: Ground truth, patch size s=3,5,7, and 9.

Patch size	depth error (front)		depth error (back)	
	mean	median	mean	median
3	2.5mm	0.38mm	2.0mm	0.4mm
5	2.5mm	0.30mm	6.4mm	1.1mm
7	2.0mm	0.39mm	7.9mm	2.0mm
9	20.9mm	0.47mm	9.7mm	1.5mm

Table I
Mean and median depth error for synthetic data with different patch size.

Next, we investigate the capability of recovering discontinuous regions by comparing our method with Kovesi's shapelets approach [14] and the diffusion method proposed in [19]. Note that the shapelets and the diffusion algorithms treat normal maps as orthogonal images, so the reconstructed surfaces differ slightly. The first row of Figure 8 shows the reconstructed ear of the bunny in the first (front) synthetic image. Since the ear is overlapped with Buddha's body, we can see the ear reconstructed by either shapelets or diffusion method has a sharp distortion near the tip. Also in the second row of Figure 8, where the Bunny's body is overlapped with Buddha's pedestal, we find that surfaces obtained by shapelets and diffusion methods become concave around the overlapping region. Moreover, only the surfaces reconstructed by our method maintain a consistent geometry w.r.t. the ground truth.

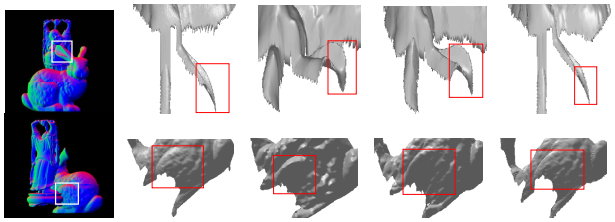


Figure 8. Comparison for discontinuous region reconstruction. The first column shows investigated overlapping regions (bounded by boxes), followed by surfaces reconstructed by Ground truth, Shapelets [14], Diffusion[19], and our approach.

Figure 9 shows the estimated surfaces with their corresponding (estimated) normal maps. We can see that the front and back objects are clearly separated, and neither is distorted by the other, especially for the regions near

the boundaries. However, for the real data there are still a number of jagged points along the shape boundary. Since we only use three images for photo-consistency based stereo matching, we expect this issues to be further reduced by considering additional images. The existence of small gaps in our face model (i.e. in the hair region) is due to the detection of disconnected surface regions (obtained through the analysis of normal information) where no SIFT features are found. While these artifacts can be improved by adjusting the discontinuity threshold, it is evident that the development of an improved adaptive scene segmentation mechanism is an important aspect to be addressed in future works.

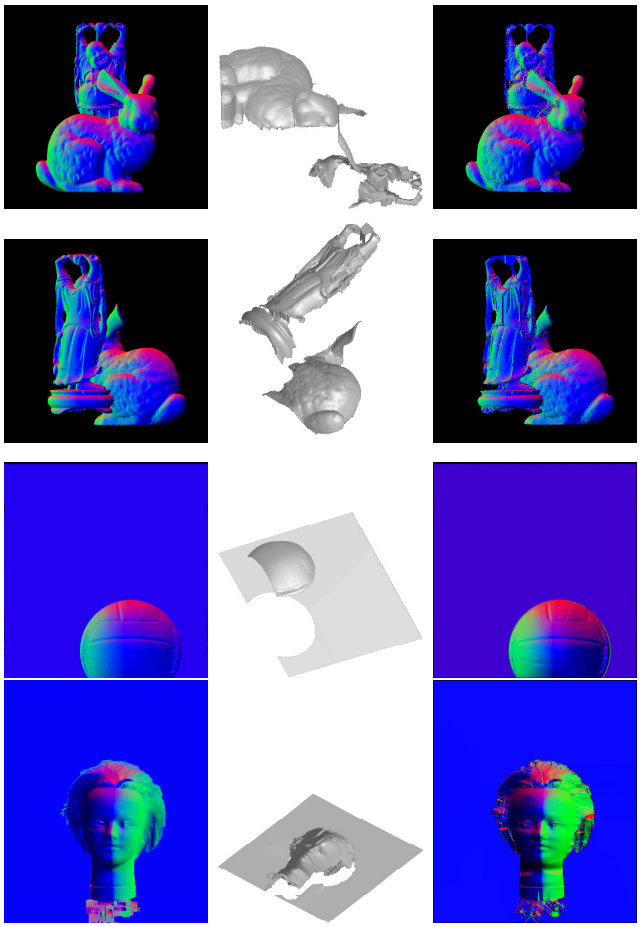


Figure 9. Recovered surfaces and normal maps. Left column: original normal maps used as input; middle column: recovered surfaces; right column: normal maps estimated from recovered depth maps.

In addition to evaluating the overall performance of our approach, we also tested the behavior of our multi-view patch sweeping module. Namely, we compared it against a fronto-parallel plane sweep [7] and a surface aligned oriented plane sweep [8]. Our results (Figure 10) illustrate how modeling the local surface geometry consistently provides a better depth estimate based on photo-consistency. As expected, patches with a set normal surface measurements which diverge from the fronto parallel assumption are increasingly better estimated by our approach. Also, as the size of the matching template increases, patch sweeping does a better job of modeling arbitrary surfaces.

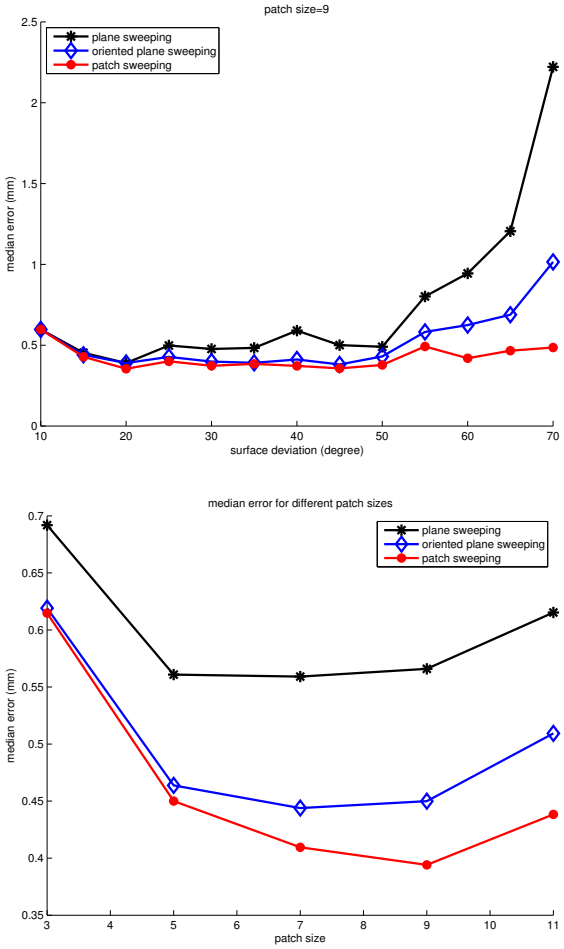


Figure 10. Patch sweep accuracy comparison. At top, depth error as a function of the surface deviation from the fronto-parallel assumption. At bottom, depth error as a function of varying patch size.

VII. Discussion and Future Work

We have proposed a novel method to recover the absolute surfaces of a scene using as input surface normal and texture information. In contrast to previous methods, our method can efficiently detect surface discontinuities, enabling its application to scenes containing multiple objects. Our approach detects surface discontinuities by modeling local surface topology as a parametric surface obtained through least square data fitting of a set of measured surface normals. Accurate depth estimates are obtained through multi-view patch sweeping. The proposed stereo approach outperforms existing plane sweep methods by being able to better model local surface geometry and incorporating this information into our photo-consistency estimates. We have emphasized

the development of computationally efficient local mechanisms throughout our processing pipeline. Accordingly, future work includes the development a real-time GPU implementation of our method.

An important research goal is to incorporate an explicit error propagation analysis for the different stages of our processing pipeline. The statistical characterization of our estimates will eliminate the need for predefined discontinuity thresholds and enable us to obtain relative surfaces invariant to *seed* selection. Future work may also utilize a global optimization framework to fuse our sparse absolute depth estimates and our relative surfaces. Note that the aforementioned improvements can be achieved at the expense of computational efficiency, so the cost/benefit analysis of their implementation is application dependent.

ACKNOWLEDGMENT

This material is based upon work supported by the The National Center for Manufacturing Sciences (US Department of Defense) under contract number 69-200910 and the US Department of Energy under contract number DE-FG52-08NA28778. We also like to thank SPAWAR and Photon-X for supporting parts of this research.

REFERENCES

[1] R. J. Woodham, "Photometric method for determining surface orientation from multiple images," *Optical Engineering*, vol. 19, pp. 139–144, 1980.

[2] B. K. P. Horn and M. J. Brooks, "The variational approach to shape from shading," *Comput. Vision Graph. Image Process.*, vol. 33, pp. 174–208, February 1986.

[3] R. T. Frankot and R. Chellappa, "Shape from shading," B. K. P. Horn, Ed. Cambridge, MA, USA: MIT Press, 1989, ch. A method for enforcing integrability in shape from shading algorithms, pp. 89–122.

[4] C. K. Chow and S. Y. Yuen, "Recovering shape by shading and stereo under lambertian shading model," *Int. J. Comput. Vision*, vol. 85, pp. 58–100, October 2009.

[5] I. Horovitz and N. Kiryati, "Bias correction in photometric stereo using control points," in *Proceedings of the Vision Modeling and Visualization Conference 2001*, ser. VMV '01. Aka GmbH, 2001, pp. 391–398.

[6] R. T. Collins, "A space-sweep approach to true multi-image matching," in *Proceedings of the 1996 Conference on Computer Vision and Pattern Recognition (CVPR '96)*, Washington, DC, USA, 1996, pp. 358–.

[7] R. Yang and M. Pollefeys, "Multi-resolution real-time stereo on commodity graphics hardware," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '03)*, 2003.

[8] D. Gallup, J.-M. Frahm, P. Mordohai, Q. Yang, and M. Pollefeys, "Real-time plane-sweeping stereo with multiple sweeping directions," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '07)*, 2007.

[9] "Photon-x," *http://www.photon-x.com*.

[10] R. Klette and K. Schlns, "Height data from gradient fields," in *Proceedings of SPIE (the international Society for Optical Engineering) on Machine Vision Applications, Architectures, and Systems Integration*, 1996, pp. 204–215.

[11] L. Noakes and R. Kozera, "Nonlinearities and noise reduction in 3-source photometric stereo," *J. Math. Imaging Vis.*, vol. 18, pp. 119–127, March 2003.

[12] J.-W. Hsieh, H.-Y. M. Liao, M.-T. Ko, and K.-C. Fan, "Wavelet-based shape from shading," *Graph. Models Image Process.*, vol. 57, pp. 343–362, July 1995.

[13] B. Karaçali and W. Snyder, "Reconstructing discontinuous surfaces from a given gradient field using partial integrability," *Comput. Vis. Image Underst.*, vol. 92, pp. 78–111, October 2003.

[14] P. Kovesi, "Shapelets correlated with surface normals produce surfaces," in *Proceedings of the Tenth IEEE International Conference on Computer Vision - Volume 2*, ser. ICCV '05. Washington, DC, USA: IEEE Computer Society, 2005, pp. 994–1001.

[15] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi, "Efficiently combining positions and normals for precise 3d geometry," in *ACM SIGGRAPH 2005 Papers*, ser. SIGGRAPH '05. New York, NY, USA: ACM, 2005, pp. 536–543.

[16] K. M. Lee and C. C. J. Kuo, "Shape from photometric ratio and stereo," *Journal of Visual Communication and Image Representation*, vol. 7, pp. 155–162, 1996.

[17] D. Vlasic, P. Peers, I. Baran, P. Debevec, J. Popović, S. Rusinkiewicz, and W. Matusik, "Dynamic shape capture using multi-view photometric stereo," in *ACM SIGGRAPH Asia 2009 papers*. New York, NY, USA: ACM, 2009, pp. 174:1–174:11.

[18] C. Hernandez Esteban, G. Vogiatzis, and R. Cipolla, "Multiview photometric stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, pp. 548–554, March 2008.

[19] A. Agrawal and R. Raskar, "What is the range of surface reconstructions from a gradient field," in *In ECCV*. Springer, 2006, pp. 578–591.

[20] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the International Conference on Computer Vision-Volume 2 - Volume 2*, ser. ICCV '99. Washington, DC, USA: IEEE Computer Society, 1999.

[21] "Stanford 3d scanning repository," *http://www.graphics. stanford.edu/data/3Dscanrep*.