

Towards a Complete Dense Geometric and Photometric Reconstruction under Varying Pose and Illumination

Martin Weber^{*1}, Andrew Blake², and Roberto Cipolla¹

¹Department of Engineering, University of Cambridge, UK (mw232@cam.ac.uk)

²Microsoft Research, Cambridge, UK

Abstract

This paper proposes a novel framework to construct a geometric and photometric model of a viewed object that can be used for visualisation in arbitrary pose and illumination. The method is solely based on images and does not require any specialised equipment. We assume that the object has a piece-wise smooth surface and that its reflectance can be modelled using a parametric bidirectional reflectance distribution function. Without assuming any prior knowledge on the object, geometry and reflectance have to be estimated simultaneously and occlusion and shadows have to be treated consistently. We exploit the geometric and photometric consistency using the fact that surface orientation and reflectance are local invariants. In a first implementation, we demonstrate the method using a Lambertian object placed on a turn-table and illuminated by a number of unknown point light-sources. A discrete voxel model is initialised to the visual hull and voxels identified as inconsistent with the invariants are removed iteratively. The resulting model is used to render images in novel pose and illumination.

1 Introduction

The reconstruction of real objects from a number of images is a key problem in Computer Vision and its applications. The objective of this paper is to work towards the construction of complete and dense models of objects that contain all the *geometric* and *photometric* information that is required for an artificial re-illumination in arbitrary pose and illumination. The method attempts to extract all necessary information solely based on images without using any assisting special equipment e.g. laser range scanner [12] and without restricting the positions of light-sources or cameras used for the acquisition. We use three switchable point light-sources which increase the number of independent observations and improve the *condition number* for the reflectance estimation (Section 3.2.1).

The fact that the light-sources are allowed to move is of importance to this paper since we apply the method to turn-table sequences. Here the object of interest performs a circular motion relative to the fixed light-sources and the camera. We therefore use the surface

*Supported by the EPSRC, the Cambridge European Trust and the DAAD (Germany)

orientation and the reflectance as quantities that are invariant with respect to lighting and viewing conditions to induce correspondences. Since orientation and reflectance are not directly related to the observations, we formulate a consistency criterion which is used in a space-carving framework to exclude locations that do not have consistent invariants.

2 Previous Work

Besides previous work on shape recovery, the estimation of illumination and reflectance properties [11, 12, 13, 14] are relevant for our method since they are part of the quantities we have to estimate.

2.1 Shape from Shading and Photometric Stereo Methods

Shape from shading and related methods employ the dense intensity structure provided by image data. The problems in applying methods of this type can be seen on one hand in the restriction to one view-point, which excludes a complete reconstruction, and the limitations deriving from its differential structure on the other hand. The differential nature of the reflectance information expresses itself in a partial differential equation and is referred to as *irradiance equation* [8]. As a consequence, the reconstruction requires a prior on shape [7, 1] and regularisation [16]. This problem can be reduced by using several light-sources as in photometric stereo [15] but the technique is limited by the restriction to one view. Our method is related to shape from shading techniques in that we are exploiting the dense irradiance values for geometric constraints on the surface orientation. However, we use multiple-view information (which is required for a complete model anyway) to avoid the integration of the irradiance equation.

2.2 Multiple-View Techniques

One way of avoiding the problem of integrating surface orientations is to use stereo. Multiple view techniques offer direct access to 3D structure by triangulation [6] and the framework of epipolar geometry is a source for several important techniques for calibration and reconstruction [4, 6]. While these methods use salient features that can be localised in and tracked across views, they cannot be used to obtain dense information. Therefore, more recent work has focused on dense multiple view methods [5, 2, 17]. The general problem one faces is to relate intensities that correspond to the same surface location but are in general not invariant due to a relative motion of the light-sources or, for view-dependent reflectance models, also due to the motion of the camera. An interesting approach to overcome this problem is suggested in [10] where special configurations of light-source and camera are explored that can serve to eliminate the view dependency and hence lead to invariants that can be directly correlated. However their methods require either that camera and light-source be interchanged to form a *stereopsis pair* or that a large number of images (> 100) is recorded while the light-source is moved in a prescribed way. These impractical requirements can be avoided if a parametric reflectance model is used to predict intensities of corresponding points. A step in this direction has been taken [2] using *surfels* which comprise geometric and photometric information. Its main problem occurs because although estimates within each surfel give good local estimates, there are in general discontinuities along surfel boundaries. Reflectance estimation can also constitute a

difficult optimisation problem. We propose instead to use geometric continuity as a hard constraint implicit to our formulation as a voxel model and to enforce the agreement of the geometric normal field and the photometric orientations at a later stage.

3 Theoretical Framework

The surface of an object, in the context of this paper is the two-dimensional manifold $S = \partial O$ obtained as the boundary of the set of object points $O \subset \mathbb{R}^3$. Furthermore, we assume the surface S to be smooth (differentiable) so that a uniquely defined outward pointing surface normal is associated with each point on S .

3.1 Lambertian Reflectance Map

In order to model the intensities (irradiance values) observed in the images, one has to complete the purely geometric model by a reflectance function. While in principle one can use any parametric *Bidirectional Reflectance Distribution Function (BRDF)*, we initially restrict ourselves in this paper to the Lambertian reflectance function which leads to a linear least square problem in the estimation. From the Lambertian reflectance law we obtain an expression for the predicted intensity values $y(p)$ for a surface position p :

$$y(p) = R_\rho(p)S(p) \quad (3.1)$$

where R_ρ denotes the Lambertian reflectance factor and S the light-source:

$$R_\rho(l, n) = \rho \langle l, n \rangle, \quad S(p) = \frac{b}{|q - p|^2} \quad (3.2)$$

here ρ is the albedo factor, l and n are the unit vectors at the surface location pointing towards the light-source and the surface-normal respectively and b denotes the strength of the source (located at $q \in \mathbb{R}^3$) measured at the surface point $p \in S$.

3.2 Photometric Consistency Constraint

This section defines the residual map which expresses the constraints that derive from the observed intensities in a point-wise fashion. Assuming¹ the surface location p is visible and not in shade in image j of the sequence, equation 3.1 yields the prediction

$$y_j(p) = R_{\rho_j}(p)S_j(p) \quad (3.3)$$

for the observed irradiance y_j . While one expects generally that neighbouring locations on the object have similar orientation, this relation does not extend to neighbouring pixels in images of the object: here, due to the projection, points of different depth can become neighbours without having any neighbourhood relation on the object. Therefore we postpone the exploitation of neighbourhood relations to the stage where we have estimated an approximate 3D model and can identify which locations are indeed near to each other. Preparing the introduction of the residual map, we define the following cost function as a measure of how well a particular hypothesis agrees with the observations:

¹We assume furthermore that inter-reflections are weak and can be neglected.

Assuming orientation and reflectance are described by $\theta = (n, \rho) \in S^2 \times \mathbb{R}$, we express the cost $f(\theta, p)$ of position $p \in \mathbb{R}^3$ to be on the surface:

$$f(\theta, p) = \frac{1}{2} \sum_{j=1}^m \left(\frac{1}{\Delta y} [R_\rho(l_j, n) S_j(p) - y_j] \right)^2 \quad (3.4)$$

where y_j is the measured irradiance value in image j corresponding to p and Δy denotes its uncertainty. The sum is over all images in which the point p is visible and not in shade with respect to the point light-source described by S_j .

Whether the solution of the optimisation problem related to f can be used to determine all parameters depends on the number of independent images m that observe p :

- $m \geq 3$: The surface normal n and the albedo ρ can be calculated which minimise f and if $m > 3$ the remaining error serves as consistency measure (Section 3.2.1).
- $m < 3$: Observations are not sufficient to determine all parameters.

We now use f to define the *residual map* which assigns the positive value

$$r(p) = \inf_{\theta \in S^2 \times \mathbb{R}} f(\theta, p) \quad (3.5)$$

corresponding to the assumption that location $p \in \mathbb{R}^3$ is on the surface. In the absence of noise, r vanishes on the surface S . Note that unlike methods that directly match intensities, one is not restricted here in the choice of illumination or viewing conditions as long as these conditions are either known or one is able to estimate them independently.

3.2.1 Robust Calculation of the Lambertian Residual Map

The Lambertian reflectance model simplifies the estimation considerably when the combined vectors $\theta = \rho n \in \mathbb{R}^3$ and $\tilde{l}_j := S(p) l_j \in \mathbb{R}^3$ are introduced. The Lambertian expression for the image irradiance (Equation 3.1) is linear in θ : $y_j = \langle \tilde{l}_j, \theta \rangle$. It can therefore be rewritten in Matrix form:

$$A\theta = y, \text{ with } A := (\tilde{l}_1, \dots, \tilde{l}_m)^t \in L(\mathbb{R}^3, \mathbb{R}^m) \quad (3.6)$$

Given the vector of measured irradiance $y \in \mathbb{R}^m$, the optimum parameter combination θ_* can be expressed using the *pseudo inverse*: $\theta_* = A^+ y$. The pseudo inverse can be written using a *truncated QR-decomposition*: Decomposing $A = Q_s R_s$ with $R_s \in L(\mathbb{R}^3, \mathbb{R}^3)$ and $Q_s \in L(\mathbb{R}^3, \mathbb{R}^m)$ one has $A^+ = R_s^{-1} Q_s^t$. Hence one can also write $AA^+ = Q_s Q_s^t$ and using this, one finds for the norm of the residual vector $A\theta_* - y$:

$$r(p) = \frac{1}{2} (A\theta_* - y)^2 = \frac{1}{2} (Q_s Q_s^t y - y)^2 = \frac{1}{2} (y^2 - w^2) \quad (3.7)$$

where $w := Q_s^t y$ got introduced. Hence, the numerical calculation of the residual map, used as carving criteria in Section 4, does not require a matrix inversion and allows for a numerically stable computation. The full inversion (back-substitution) is only required when the reflectance and orientation parameter θ itself have to be calculated using $R_s \theta = w$ in order to supply the photometric data for the carved model. Note that $\{\tilde{l}_i\}$ determines the *condition* of (3.6) and that a non-planar set of vectors \tilde{l}_i is required.

3.2.2 Predicting Geometric Accuracy

The geometric constraint is expressed in the fact that the residual map vanishes if p is a point on the surface. Figure 1 explains the situation when p is off the surface:

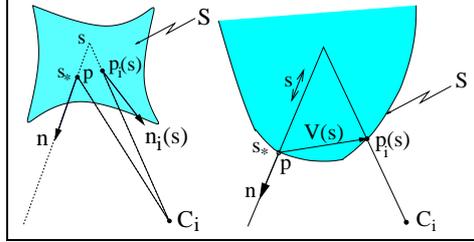


Figure 1: Dependence of the observed irradiance value registered in camera C_i : The left diagram illustrates the true surface location p with surface normal n and depth-parameter s_* . Indicated is also the surface location $p_i(s)$ that is observed in C_i instead of p if s differs from s_* . The right diagram shows the tangential motion induced by a change in the depth hypothesis. Drawn is the slice defined by the surface normal and the focal point C_i . Each depth hypothesis s leads to a corresponding surface-location $p_i(s)$ in the slice. For $s \rightarrow s_*$, $V_i(s)$ converges to a tangent vector $V_i' \in T_p S$.

In general, the camera C_i will observe $p_i(s)$ and record the corresponding irradiance value $y_i(s)$. If $p(s)$ does not correspond to a location on the surface we therefore expect a non-vanishing value of the residual map that can be approximated to second order in the following Taylor expansion: $r(s) \approx \frac{1}{2} \sum_i \left(\frac{y_i'}{\Delta y} \right)^2 (s - s_*)^2$. As is illustrated in Figure 1, one can expect to observe deviations y_i' even in the case of homogeneous surfaces. These deviations are related to the local curvature of the surface and can be expressed using the Weingarten map W (sometimes referred to as *shape operator*): $y_i' = \nabla_n R(l_i, n) W V_i'$, where V_i' denotes the tangent vector indicated in the slice of Figure 1. Finally, deriving from the previous equation, we can predict the uncertainty of the geometric estimation (distance Δs off the surface) in the presence of noise (Δy):

$$\Delta s = \Delta y \frac{1}{\sqrt{r'}} \propto \Delta y \frac{1}{\sqrt{m}} \frac{1}{\kappa |\nabla_n R|} \quad (3.8)$$

We list a few relevant conclusions that can be drawn from this expression:

- The accuracy increases with the square root of the number of available observations.
- Curvature (κ) enters the equation in first order. That implies that the residual parameter does not deliver any constraint in the case of flat surfaces.

4 Experimental Set-Up

The experimental set-up consists of a turn-table, three lamps and a digital camera. The turn-table is used to generate a circular motion of the object while the camera and point light-sources are fixed. The situation is equivalent to a fixed object and simultaneously

rotating camera and light-sources. As light-sources we use controlled halogenic bulbs without any reflector to approximate point-like behaviour. The bulbs have $20W$ power each and are powered by a constant $12V$ DC supply unit. For each of the sequences (Figures 2 and 3) we recorded images corresponding to the 30 combinations resulting from 10 rotational configurations and one of the three light-sources being turned on at a time. We use a *Canon PowerShot G2*, as a commercial digital camera that can be operated in manual mode and record non-compressed image files. In order to obtain sensible irradiance values, we checked the photometric linearity of the camera (response function) and removed the thermal noise (offset) from the generated images. In addition to the recording of the object-sequence, we recorded a second sequence of a cube of homogeneous Lambertian material used for calibration purposes.

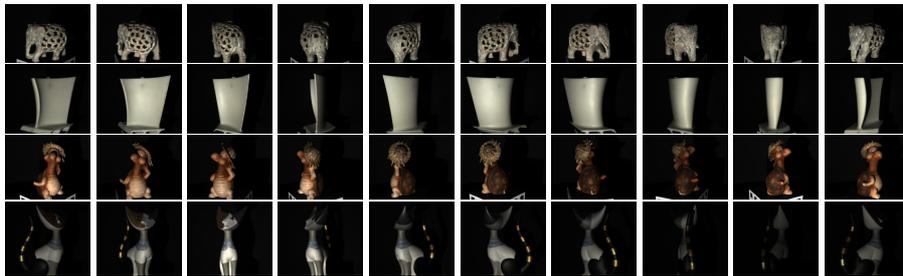


Figure 2: Recorded sequences: Shown are the images recorded for ten rotational views.

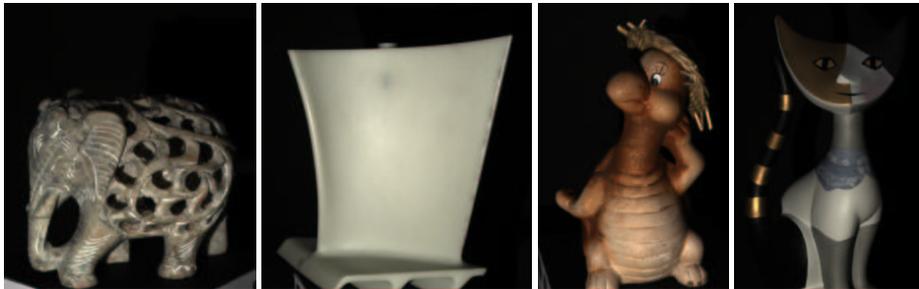


Figure 3: Recorded objects: Shown is one image from each of the recorded sequences.

4.1 Calibration

In order to exploit the relations found in Section 3, we need to calibrate the camera and light-sources and need to determine the motion relative to the object.

- **Camera Calibration:** For the camera calibration we use standard techniques [4, 6] and define object coordinates using a calibration grid.

- **Light Estimation:** Assuming isotropic point light-sources, one can use the calibration sequence also to calculate the positions and intensities of the light-sources. The method used was described in [14] and requires only the images of the calibration sequence to calibrate each light in turn. The resulting locations of the light-sources can then be expressed in the coordinate system of the camera (which remains at rest relative to the lights throughout the sequence).

4.2 Model Generation

Similar to the space carving algorithm [9], the object is represented as a voxel array where each voxel contains the residual value $r(p)$ and an invariant parameter vector $\theta = \rho n$. We used the residual map as the carving criterion. Starting from an initial large cube that contains the object O the segmentation of the object is used to obtain the *visual hull* V . We then exploited several invariant consistency criteria to restrict the object iteratively to the smaller consistency hull C_∞ :

$$\forall i \quad O \subset C_\infty \subset C_{i+1} \subset C_i \subset V \subset \mathbb{R}^3$$

- **Initialisation** ($C_0 = V$): We use *B-spline snakes* [3] on the outline contour to segment each image into two regions: object and background. In combination with the known projection matrices we can therefore calculate the *visual hull* V which is known to contain all of the object O : $O \subset V \subset \mathbb{R}^3$ and serves us as initialisation.
- **Visibility:** Before one can proceed and apply the constraint deriving from the residual map, one needs to determine in which of the images the location in question is observable. Fortunately the problem of determining visibility and shade² for the *unknown* object can be solved in a conservative way: Using the current estimated consistency hull C_i , we can determine for each location $p \in C_i$ whether it would be observed if C_i were the true object. By the containing property we have $O \subset C_i$ and hence if p is not visible in O it will certainly not be visible in the larger hull C_i . Therefore, working with C_i instead of O has merely the effect of restricting the number of observations of which we are sure to be able to make³.
- **Carving Criteria:** Once the set of safe observations is determined, we use the photometric constraints as *carving criteria*: We remove voxel p if:

$$r(p) \geq (m - d) \epsilon_{\max}^2 \quad (4.1)$$

Here $d = 3$ is the number of parameters in the reflectance model and m the number of observations used in the calculation of the residual value $r(p)$. Relation 4.1 is motivated by a χ^2 statistic and a suitable value for ϵ_{\max} can be determined by measuring⁴ the pixel noise Δy and selecting a confidence level.

- **Segmentation Consistency Constraint:** For objects that contain edges or corners, the surface orientation is not well defined and causes photometric inconsistencies which result in large residual errors and can lead to wrong removal of voxels. In

²A point is in shade if it is not *visible* from the position of the light-source.

³Note in particular that this implies that any carving can only take place on the boundary ∂C_i of C_i .

⁴Where we determined the pixel-noise by recording several images of an unchanged scene.

order to extend our method to objects with piece-wise smooth surfaces (e.g. the turbine blade in Figure 3), we protect locations along outside edges in the following way: If the removal of voxel p implies an inconsistency with the observed silhouettes, then p is protected and cannot be removed.

- **Resulting Lambertian Model:** The iterative process stops when all voxels are consistent. The resulting Lambertian model is finally estimated consisting of the boundary of the final consistency hull C_∞ : $M = \partial C_\infty$. Each voxel $p \in M$ has a reflectance parameter θ which is estimated as described in Section 3.

5 Experimental Results

Using the models generated by our method, we produced several views of the original objects displayed in Figure 3. The stone elephant and the turbine blade provide almost homogeneous surfaces. The elephant has a structure of holes and the turbine blade has a concave area. The clay turtle and the porcelain cat are examples for textured objects. The cat shows mild specular reflectance and has some dark coloured areas. Figure 4 shows views of the generated models in novel pose and illumination. While all views are different from the projective projections used in the acquisition, the differences are most obvious in the overhead views of the elephant, turtle and cat. It can be observed that the method gives realistic models even in the case of homogeneous surfaces and under complex occlusion constraints such as the ones provided by the elephant sequence. The concavity of the turbine blade is recovered correctly by our method with an accuracy depending on the curvature (cf. Equation 3.8). The geometry of the reconstructed models is displayed separately in the top row of Figure 4, where we have replaced the estimated albedo by 1. The overall calculations took about four hours per sequence in our present non-optimised implementation on a 900MHz pentium-III processor. The constructed models are available as *VRML* models at: <http://svr-www.eng.cam.ac.uk/~mw232/Research/>

6 Conclusion and Future Work

In this paper we considered a multi-view photometric reconstruction technique that only requires a turn-table, switchable lights and a digital camera. We exploited the fact that surface normal and reflectance are local invariants with respect to lighting and viewpoint. We demonstrated the simultaneous and dense recovery of reflectance and geometry under varying illumination which constitutes a problem for methods based on intensity correlation. All necessary calibration of the camera and the light-sources was solely based on images. We found that the photometric consistency constraint is sufficient to reconstruct curved surfaces and we were able to quantify the obtainable geometric accuracy (Equation 3.8). The practical use of our framework was then explored in space carving, where inconsistent voxels are removed. We were able to demonstrate the successful reconstruction of Lambertian scenes and concluded with images of the re-illuminated model. In our future work, we aim to use an advanced surface model (e.g. level-set methods) and to extend the method by the estimation of more general parametric reflectance models with specular components (e.g. Torrance-Sparrow model).

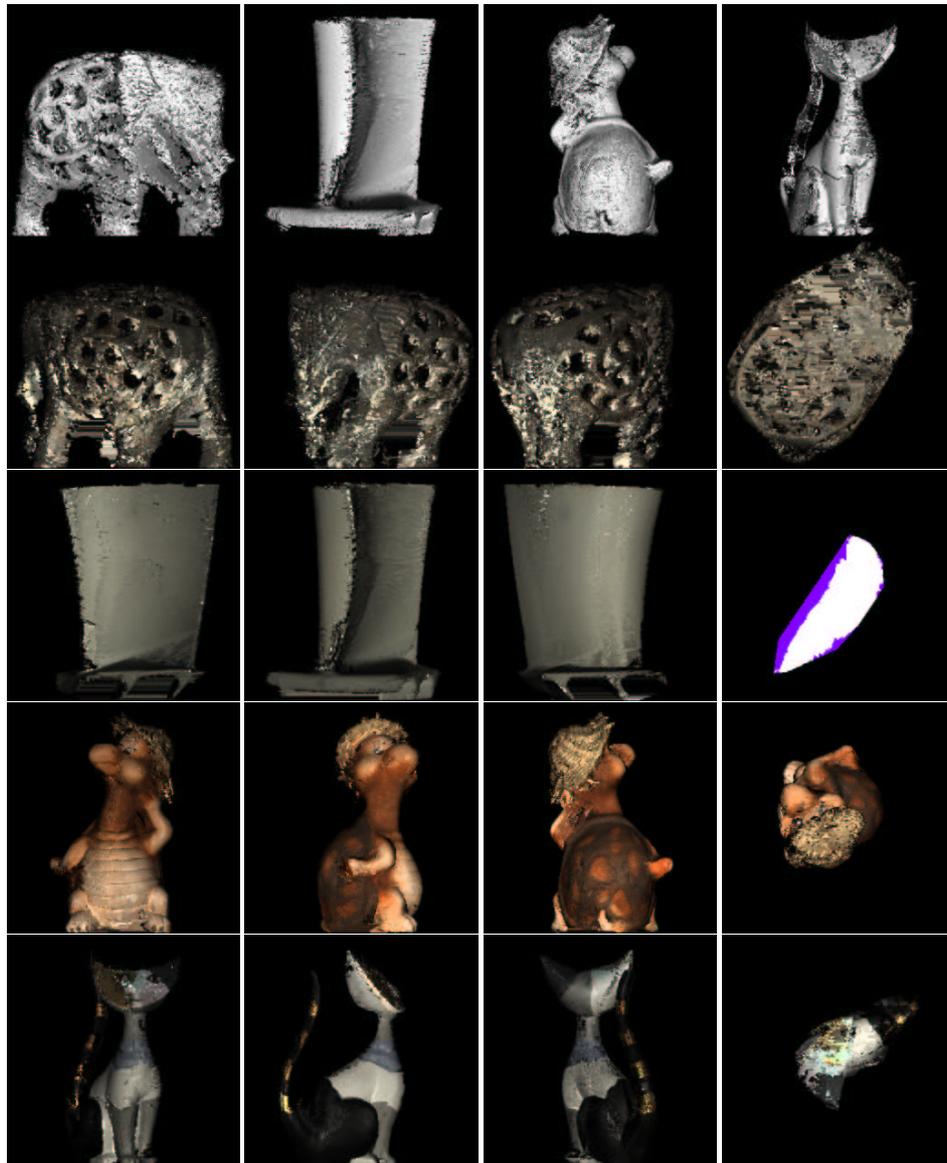


Figure 4: Novel views generated from the estimated models: Shown are orthographic views calculated using a virtual point light-source and a virtual orthographic camera. Images were rendered using the estimated albedo values except images of the top row, where a constant albedo was used. Our method reproduces successfully the structure of the objects even in the presence of complex visibility and shade situations (stone elephant). The reconstructed turbine blade gives an example of a homogeneous object with a concave area. The differences between visual hull and consistency hull are indicated in the horizontal section of the turbine blade. Note that the reconstructions contain all visible details such as the straw hat of the clay turtle. The reconstruction of the porcelain cat is limited by the deviation from Lambert's law but still gives rise to a qualitatively correct model.

References

- [1] A. Blake, A. Zisserman, and G. Knowles. Surface description from stereo and shading. *IVC*, 3(4):183–191, 1985.
- [2] R.L. Carceroni and K.N. Kutulakos. Multi-view scene capture by surfel sampling: From video streams to non-rigid 3d motion, shape and reflectance. In *ICCV01*, pages II: 60–67, 2001.
- [3] R. Cipolla and A. Blake. The dynamic analysis of apparent contours. In *ICCV90*, pages 616–623, 1990.
- [4] O. D. Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, Cambridge, Massachusetts, 1993.
- [5] O. D. Faugeras and R. Keriven. Complete dense stereovision using level set methods. In *ECCV proceedings (1)*, pages 379–393, 1998.
- [6] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, 2000.
- [7] B.K.P. Horn and M.J. Brooks. The variational approach to shape from shading. *Computer Vision, Graphics, and Image Processing*, 33(2):174–208, 1986.
- [8] K. Ikeuchi and B.K.P. Horn. Numerical shape from shading and occluding boundaries. *Artificial Intelligence*, 17:141–184, 1981.
- [9] K.N. Kutulakos and S.M. Seitz. A theory of shape by space carving. In *ICCV99*, pages 307–314, 1999.
- [10] S. Magda, D. Kriegman, T. Zickler, and P. Belhumeur. Beyond Lambert: Reconstructing surfaces with arbitrary BRDFs. In *ICCV01*, pages II: 391–398, 2001.
- [11] K. Nishino, Z. Zhang, and K. Ikeuchi. Determining reflectance parameters and illumination distribution from a sparse set of images for view-dependent image synthesis. In *ICCV01*, pages I: 599–606, 2001.
- [12] Ravi Ramamoorthi and Pat Hanrahan. A signal-processing framework for inverse rendering. In ACM, editor, *SIGGRAPH 2001 Conference Proceedings, August 12–17, 2001, Los Angeles, CA*, pages 117–128, 2001.
- [13] I. Sato, Y. Sato, and K. Ikeuchi. Illumination distribution from shadows. In *CVPR99*, pages I:306–312, 1999.
- [14] M. Weber and R. Cipolla. A practical method for estimation of point light-sources. *proceedings BMVC 2001*, 2:471–480, Sep 2001.
- [15] R.J. Woodham. Gradient and curvature from the photometric-stereo method, including local confidence estimation. *JOSA*, 11:3050–3068, 1994.
- [16] P.L. Worthington and E.R. Hancock. New constraints on data-closeness and needle map consistency for shape-from-shading. *IEEE-PAMI*, 21:1250–1267, 1999.
- [17] A.J. Yezzi and S. Soatto. Stereoscopic segmentation. In *ICCV01*, pages I: 56–66, 2001.