What Object Motion Reveals About Shape With Unknown BRDF and Lighting

Manmohan Chandraker
NEC Labs America
manu@nec-labs.com

Dikpal Reddy   Yizhou Wang   Ravi Ramamoorthi
University of California, Berkeley
{dikpal@eecs, yzhwang@me, ravir@eecs}.berkeley.edu

Abstract

We present a theory that addresses the problem of determining shape from the (small or differential) motion of an object with unknown isotropic reflectance, under arbitrary unknown distant illumination, for both orthographic and perspective projection. Our theory imposes fundamental limits on the hardness of surface reconstruction, independent of the method involved. Under orthographic projection, we prove that three differential motions suffice to yield an invariant that relates shape to image derivatives, regardless of BRDF and illumination. Under perspective projection, we show that four differential motions suffice to yield depth and a linear constraint on the surface gradient, with unknown BRDF and lighting. Further, we delineate the topological classes up to which reconstruction may be achieved using the invariants. Finally, we derive a general stratification that relates hardness of shape recovery to scene complexity. Qualitatively, our invariants are homogeneous partial differential equations for simple lighting and inhomogeneous for complex illumination. Quantitatively, our framework shows that the minimal number of motions required to resolve shape is greater for more complex scenes. Prior works that assume brightness constancy, Lambertian BRDF or a known directional light source follow as special cases of our stratification. We illustrate with synthetic and real data how potential reconstruction methods may exploit our framework.

1. Introduction

An open problem in computer vision since early works on optical flow has been to determine the shape of an object with unknown reflectance undergoing differential motion, when observed by a static camera under unknown illumination. This paper presents a theory to solve the problem for both orthographic and perspective camera projections, with arbitrary unknown distant lighting (directional or area).

Unlike traditional approaches to shape recovery from motion like optical flow [6, 10] or multiview stereo [15], our theory does not make physically incorrect assumptions like brightness constancy, or simplifying ones like Lambertian reflectance. In Section 3, we correctly model the dependence of image formation on the bidirectional reflectance distribu-

tion function (BRDF) and illumination, to derive a physically valid differential flow relation. Remarkably, it can be shown that even when the BRDF and illumination are unknown, the differential flow constrains the shape of an object through an invariant relating surface depth to image derivatives.

The form of the invariant depends on the camera projection and the complexity of the illumination (see Table 1 for a summary). For orthographic projections, considered in Section 4, three differential motions suffice and the invariant is a quasilinear partial differential equation (PDE). For perspective projections, we show in Section 5 that surface depth may be directly recovered from four differential motions, with an additional linear PDE constraining the surface normal. The involved PDEs are homogeneous for simple illuminations, but inhomogeneous for complex lighting. Besides characterizing the invariants, in each case, we also study the precise extent to which surface reconstruction may be performed.

Our theory can be considered a generalization of several prior works. Notably, the brightness constancy relation popularly used in optical flow [6, 10] is a special case of our differential flow relation, as are more physically-based studies that relate the motion field to radiometric entities assuming diffuse reflectance [11, 12]. Recent work by Basri and Frolova [1] can be seen as a special case for two views, orthographic projection, Lambertian BRDF and known directional lighting. All of these restrictions are generalized by our theory. We also derive a stratification of shape recovery from differential object motions (Section 6 and Table 1), which demonstrates that the hardness of a reconstruction problem may be qualified by the nature of the underlying BRDF-invariant PDE and quantified by the number of motions required to solve it. Note that the limits imposed by our theory on the hardness of motion field estimation are fundamental ones – they hold true regardless of the actual reconstruction method employed.

In summary, this paper makes the following contributions:

- A theory that relates shape to object motion, for unknown isotropic BRDF and illumination (directional or area), under orthographic and perspective projections.
- Delineation of the precise extent to which shape may be recovered using the proposed invariants.
- A stratification of the hardness of shape recovery from motion, under various imaging conditions (Table 1).
- Generalization of prior works within our stratification.

Camera	Light	BRDF	#Motions	Surface Constraint	Shape Recovery	Theory
Persp.	Brightness Constancy		1	Lin. eqn. (optical flow)	Depth	[6, 10]
Orth.	Known, Dirn.	Lambertian	1	Inhomog. quasilinear PDE	Char. curves	[1], Eqn. (35)
Persp.	Known, Dirn.	Lambertian	1	Inhomog. quasilinear PDE	Char. curves	[11], Eqn. (36)
Orth.	Colocated	Unknown	2	Homog. quasilinear PDE	Level curves	Prop. 2, 3
Orth.	Unknown	Unknown	3	Inhomog. quasilinear PDE	Char. curves	Prop. 2, 4
Persp.	Colocated	Unknown	3	Lin. eqn. + Homog. lin. PDE	Depth + Gradient	Cor. 1
Persp.	Unknown	Unknown	4	Lin. eqn. + Inhomog. lin. PDE	Depth + Gradient	Prop. 5

Table 1. Stratification results from our theory, that establish the hardness of shape recovery under object motion, in relation to the complexity of BRDF and illumination. The qualitative hardness of shape from motion is indicated by the nature of reconstruction invariant and quantified by minimal number of required motions. Our results also generalize several prior works. Note that $k + 1$ images are required to observe k motions.

2. Related Work

Optical flow has traditionally assumed brightness constancy, which is physically inaccurate, even for Lambertian reflectance [11, 17]. While attempts have been made to relax the assumption for certain reflectance models [5, 12], our work provides the first unified theoretical and computational paradigm that relates shape recovery to image derivatives from object motion, with unknown BRDF and illumination.

Within stereo, Simakov et al. propose a dense correspondence measure for Lambertian surfaces that accounts for first order spherical harmonics of environment illumination [16]. Similar to our setup, albeit limited to Lambertian reflectance and directional lighting, passive photometric stereo methods [9, 18] use object motion to reconstruct a dense depth map. Our contributions generalize those methods, since our shape recovery is invariant to BRDF and illumination.

Surface reconstruction for general BRDFs has also been studied for several other situations. Shape recovery from specular flow on mirror surfaces has been analyzed by Canas et al. [2]. Shape from shading is extended to a parametric non-Lambertian BRDF model in [8]. The Helmholtz reciprocity principle is exploited by Zickler et al. to extend stereo to general BRDFs [19]. Sato et al. recover shape under non-Lambertian reflectance using an isometric relationship between change in intensity profiles under light source motion and surface normal differences [14]. In computer graphics, Ramamoorthi et al. have developed a theory of shading variations with first-order illumination changes [13], while Chen and Arvo have studied ray differentials [4].

The key motion cues for shape reconstruction are related to lighting, object or camera. Chandraker et al. [3] answer the question of what (differential) light source motion reveals about object shape, when BRDF is unknown. In similar spirit, this work definitively answers the second question, pertaining to shape recovery from object motion.

3. Shape and Motion for General BRDFs

Setup and notation The camera and lighting in our setup are fixed, while the object moves. The object BRDF is assumed isotropic and homogeneous (or having slow spatial variation), with an unknown functional form. The distant illumination may be directional or environment. Interreflections and shadows are assumed negligible. Let the focal length of the camera be f. The principal point on the image plane is defined as the origin of the 3D coordinate system, with the camera center at $(0, 0, -f)^\top$. Denoting $\beta = f^{-1}$, a 3D point $\mathbf{x} = (x, y, z)^\top$ is imaged at $\mathbf{u} = (u, v)^\top$, where

$$u = x/(1 + \beta z), \qquad v = y/(1 + \beta z). \qquad (1)$$

Differential motion Using the projection equations in (1), the motion field is given by

$$\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \dot{u} \\ \dot{v} \end{bmatrix} = \frac{1}{1 + \beta z} \begin{bmatrix} \dot{x} - \beta u \dot{z} \\ \dot{y} - \beta v \dot{z} \end{bmatrix}. \qquad (2)$$

Consider a small rotation $\mathbf{R} \simeq \mathbf{I} + [\boldsymbol{\omega}]_\times$ and translation $\boldsymbol{\tau} = (\tau_1, \tau_2, \tau_3)^\top$, where $[\boldsymbol{\omega}]_\times$ is the skew-symmetric matrix of $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3)^\top$. Then, $\dot{\mathbf{x}} = \boldsymbol{\omega} \times \mathbf{x} + \boldsymbol{\tau}$ for a point $\mathbf{x}$ on the object. In the perspective case, the motion field is

$$\boldsymbol{\mu} = \left(\alpha_1 + \frac{\alpha_2 + \omega_2 z}{1 + \beta z}, \ \alpha_3 + \frac{\alpha_4 - \omega_1 z}{1 + \beta z} \right)^\top, \qquad (3)$$

where $\alpha_1 = \omega_2 \beta u^2 - \omega_1 \beta uv - \omega_3 v$, $\alpha_2 = \tau_1 - \beta u \tau_3$, $\alpha_3 = -\omega_1 \beta v^2 + \omega_2 \beta uv + \omega_3 u$ and $\alpha_4 = \tau_2 - \beta v \tau_3$. Under orthography, $\beta \to 0$, thus, the motion field is

$$\boldsymbol{\mu} = (\alpha_5 + \omega_2 z, \ \alpha_6 - \omega_1 z)^\top, \qquad (4)$$

where $\alpha_5 = \tau_1 - \omega_3 v$ and $\alpha_6 = \tau_2 + \omega_3 u$.

Differential flow relation Assuming isotropic BRDF ρ, the image intensity of a 3D point $\mathbf{x}$, imaged at pixel $\mathbf{u}$, is

$$I(\mathbf{u}, t) = \sigma(\mathbf{x}) \rho(\mathbf{n}, \mathbf{x}), \qquad (5)$$

where σ is the albedo and $\mathbf{n}$ is the surface normal at the point. The cosine fall-off is absorbed within ρ. The BRDF ρ is usually written as a function of incident and outgoing directions, but for fixed lighting and view, can be seen as a function of surface position and orientation. This is a reasonable image formation model that subsumes traditional ones like Lambertian and allows general isotropic BRDFs modulated by spatially varying albedo. Note that we do not make any assumptions on the functional form of ρ, in fact, our theory will derive invariants that eliminate it.

Considering the *total* derivative on both sides of (5), using the chain rule, we have

$$I_u \dot{u} + I_v \dot{v} + I_t = \sigma \frac{d}{dt} \rho(\mathbf{n}, \mathbf{x}) + \rho \frac{d\sigma}{dt}. \qquad (6)$$

Since σ is intrinsically defined on the surface coordinates, its total derivative vanishes (for a rigorous derivation, please refer to Appendix A). Noting that $\boldsymbol{\mu} = (\dot{u}, \dot{v})^\top$ is the motion field, the above can be rewritten as

$$(\nabla_{\mathbf{u}} I)^\top \boldsymbol{\mu} + I_t = \sigma \left[(\nabla_{\mathbf{n}} \rho)^\top (\boldsymbol{\omega} \times \mathbf{n}) + (\nabla_{\mathbf{x}} \rho)^\top \boldsymbol{\nu} \right], \quad (7)$$

where $\boldsymbol{\nu}$ is the linear velocity and we use $\dot{\mathbf{n}} = \boldsymbol{\omega} \times \mathbf{n}$. Since lighting is distant and BRDF homogeneous (or with slow spatial variation), $\nabla_{\mathbf{x}} \rho$ is negligible. Moreover, using standard vector identities, $(\nabla_{\mathbf{n}} \rho)^\top (\boldsymbol{\omega} \times \mathbf{n}) = (\mathbf{n} \times \nabla_{\mathbf{n}} \rho)^\top \boldsymbol{\omega}$. Denoting $E = \log I$, we note that the albedo can be easily eliminated by dividing out $I(\mathbf{u}, t)$, to yield the *differential flow relation*:

$$\boxed{(\nabla_{\mathbf{u}} E)^\top \boldsymbol{\mu} + E_t = (\mathbf{n} \times \nabla_{\mathbf{n}} \log \rho)^\top \boldsymbol{\omega}.} \quad (8)$$

The differential flow relation in (7) and (8) is a strict generalization of the brightness constancy relation used by the vast majority of prior works on optical flow [6, 10]. Indeed, with a constant BRDF $\rho = 1$, the RHS in (7) or (8) vanishes, which is precisely the brightness constancy assumption. However, note that $\rho = 1$ is physically unrealistic – even the most basic Lambertian assumption is $\rho(\mathbf{n}) = \mathbf{n}^\top \mathbf{s}$, in which case (8) reduces to a well-known relation [12]:

$$(\nabla_{\mathbf{u}} E)^\top \boldsymbol{\mu} + E_t = \frac{(\mathbf{n} \times \mathbf{s})^\top \boldsymbol{\omega}}{\mathbf{n}^\top \mathbf{s}}. \quad (9)$$

In the following, we explore the extent to which the motion field $\boldsymbol{\mu}$ and object shape may be recovered using (8), under both orthographic and perspective image formation. Precisely, we show that it is possible to eliminate all BRDF and lighting effects in an image sequence, leaving a simple relationship between image derivatives, surface depths and normals.

4. Orthographic Projection

In this section, we consider recovery of the shape of an object with unknown BRDF, using a sequence of differential motions. Under orthography, the motion field $\boldsymbol{\mu}$ is given by (4). Denoting $\boldsymbol{\pi} = \mathbf{n} \times \nabla_{\mathbf{n}} \log \rho$, one may rewrite (8) as

$$pz + q = \boldsymbol{\omega}^\top \boldsymbol{\pi}, \quad (10)$$

where, using (4), p and q are known entities given by

$$p = \omega_2 E_u - \omega_1 E_v \quad (11)$$
$$q = \alpha_5 E_u + \alpha_6 E_v + E_t. \quad (12)$$

4.1. Rank-Deficiency in an Image Sequence

For $m \geq 3$, consider a sequence of $m + 1$ images, $E_0, \cdots, E_m$, where E_i is related to E_0 by a known differential motion $\{\boldsymbol{\omega}^i, \boldsymbol{\tau}^i\}$. We assume that the object undergoes general motion, that is, the set of vectors $\boldsymbol{\omega}^i, i = 1, \cdots, m$, span $\mathbb{R}^3$. Then, from (10), we have a set of relations

$$p^i z + q^i = \boldsymbol{\pi}^\top \boldsymbol{\omega}^i, \quad i = 1, \cdots, m. \quad (13)$$

Note that p^i, q^i and $\boldsymbol{\omega}^i$ are known from the images and calibration, while surface depth z and the entity $\boldsymbol{\pi}$ related to normals

and BRDF are unknown. It might appear at a glance that using the above m relations in (13), one may set up a linear system whose each row is given by $[p^i, -\omega_1^i, -\omega_2^i, -\omega_3^i]^\top$, to solve for both z and $\boldsymbol{\pi}$ at every pixel. However, note the form of $p^i = E_u \omega_2^i - E_v \omega_1^i$, which means that the first column in the involved $m \times 4$ linear system is a linear combination of the other three columns. Thus, the linear system is rank deficient (rank 3 in the general case when the set of vectors $\{\boldsymbol{\omega}^i\}, i = 1, \cdots, m$, span $\mathbb{R}^3$), whereby we have:

Proposition 1. *Under orthographic projection, surface depth under unknown BRDF may not be unambiguously recovered using solely motion as the cue.*

4.2. BRDF-Invariant Constraints on Surface

While one may not use (10) directly to obtain depth, we may still exploit the rank deficiency to infer information about the surface depth, as stated by the following:

Proposition 2. *For an object with unknown BRDF, observed under unknown lighting and orthographic camera, three differential motions suffice to yield a BRDF and lighting invariant relation between image derivatives and surface geometry.*

Proof. Consider a sequence of images $E_0, \cdots, E_m$, where $m \geq 3$, such that E_i is related to E_0 by a known differential motion $\{\boldsymbol{\omega}^i, \boldsymbol{\tau}^i\}$ for $i = 1, \cdots, m$ and the vectors $\{\boldsymbol{\omega}^i\}$ span $\mathbb{R}^3$. Let $\mathbf{b}^i = (E_u \omega_2^i - E_v \omega_1^i, -\omega_1^i, -\omega_2^i, -\omega_3^i)^\top$ and $\mathbf{q} = (q^1, \cdots, q^m)^\top$. Then, defining $\mathbf{B} = \left[\mathbf{b}^1, \cdots, \mathbf{b}^m \right]^\top$, we have a system of equations of the form (13) given by $\mathbf{B}(z, \boldsymbol{\pi}^\top)^\top = \mathbf{q}$. From Proposition 1, we have $\text{rank}(\mathbf{B}) = 3$. It may be easily verified that $\text{null}(\mathbf{B}) = (1, -E_v, E_u, 0)^\top$. Thus, we have the parameterized solution

$$(z, \boldsymbol{\pi}^\top)^\top = -\mathbf{B}^+ \mathbf{q} + k(1, -E_v, E_u, 0)^\top, \quad (14)$$

where $\mathbf{B}^+$ is the Moore-Penrose pseudoinverse of $\mathbf{B}$ and k an arbitrary scalar. Define $\boldsymbol{\gamma} = -\mathbf{B}^+ \mathbf{q}$ and $\boldsymbol{\gamma}' = (\gamma_2, \gamma_3, \gamma_4)^\top$. Then, we have the following two relations

$$z = \gamma_1 + k \quad (15)$$
$$\boldsymbol{\pi} = \boldsymbol{\gamma}' + k(-E_v, E_u, 0)^\top. \quad (16)$$

From the definition of $\boldsymbol{\pi}$, we have $\mathbf{n}^\top \boldsymbol{\pi} = 0$. Substituting from the above two relations (with $k = z - \gamma_1$), we get

$$(\lambda_1 + \lambda_2 z)n_1 + (\lambda_3 + \lambda_4 z)n_2 - \gamma_4 n_3 = 0, \quad (17)$$

where $\lambda_1 = -(\gamma_2 + \gamma_1 E_v)$, $\lambda_2 = E_v$, $\lambda_3 = -\gamma_3 + \gamma_1 E_u$ and $\lambda_4 = -E_u$. Noting that $n_1/n_3 = -z_x$ and $n_2/n_3 = -z_y$, we may rewrite (17) as

$$\boxed{(\lambda_1 + \lambda_2 z)z_x + (\lambda_3 + \lambda_4 z)z_y + \gamma_4 = 0,} \quad (18)$$

which is independent of BRDF and lighting. $\square$

Thus, we may directly relate surface depth and gradient to image intensity, even for unknown BRDF and illumination. This is a fundamental constraint that relates object shape to motion, regardless of choice of reconstruction method.

4.3. Surface Depth Estimation

Next, we consider the precise extent to which surface depth may be recovered using Proposition 2. We first consider the simpler case of a colocated source and sensor, where an isotropic BRDF is given by $\rho(\mathbf{n}^\top \mathbf{s})$, for an unknown function ρ. For our choice of coordinate system, $\mathbf{s} = (0, 0, -1)^\top$. Recall that $\boldsymbol{\pi} = \mathbf{n} \times \nabla_\mathbf{n} \log \rho$. It is easily verified that $\pi_3 = 0$, thus, $\gamma_4 = 0$ using (14). The relation in (18) now becomes

$$z_x/z_y = -(\lambda_3 + \lambda_4 z)/(\lambda_1 + \lambda_2 z) \qquad (19)$$

where the λ_i, $i = 1, \cdots, 4$ are defined as before. Now, we are in a position to state the following result:

Proposition 3. *Two or more differential motions of a surface with unknown BRDF, with a colocated source and sensor, yield level curves of surface depth, corresponding to known depths of some (possibly isolated) points on the surface.*

Proof. Define $\mathbf{a} = (\lambda_1 + \lambda_2 z, \lambda_3 + \lambda_4 z)^\top$. Then, from (19),

$$\mathbf{a}^\top \nabla z = 0. \qquad (20)$$

Since ∇z is orthogonal to the level curves of z, the tangent space to the level curves of z is defined by $\mathbf{a}$. Consider a rectifiable curve $\mathcal{C}(x(s), y(s))$ parameterized by the arc length parameter s. The derivative of z along $\mathcal{C}$ is given by

$$\frac{dz}{ds} = \frac{\partial z}{\partial x}\frac{dx}{ds} + \frac{\partial z}{\partial y}\frac{dy}{ds}. \qquad (21)$$

If $\mathcal{C}$ is a level curve of $z(x, y)$, then $dz/ds = 0$ on $\mathcal{C}$. Define $\mathbf{t} = (dx/ds, dy/ds)$. Then, we also have

$$\mathbf{t}^\top \nabla z = 0. \qquad (22)$$

From (20) and (22), it follows that $\mathbf{a}$ and $\mathbf{t}$ are parallel. Thus, $t_2/t_1 = a_2/a_1$, whereby we get

$$dy/dx = (\lambda_3 + \lambda_4 z)/(\lambda_1 + \lambda_2 z). \qquad (23)$$

Along a level curve $z(x, y) = c$, the solution is given by

$$z = c, \qquad \frac{dy}{dx} = \frac{\lambda_3 + \lambda_4 c}{\lambda_1 + \lambda_2 c}. \qquad (24)$$

Given the value of z at any point, the ODE (24) determines all other points on the surface with the same value of z. $\qquad \square$

Thus, (19) allows reconstruction of level curves of the surface, with unknown BRDF, under colocated illumination. Note that (19) is a first-order, homogeneous, quasilinear partial differential equation (PDE). Similarly, we may interpret (18) as a PDE in $z(x, y)$, in particular, it is an inhomogeneous, first-order, quasilinear PDE. This immediately suggests the following surface reconstructibility result in the general case:

Proposition 4. *Three or more differential motions of a surface with unknown BRDF, under unknown illumination, yield characteristic surface curves $\mathcal{C}(x(s), y(s), z(s))$, defined by*

$$\frac{1}{\lambda_1 + \lambda_2 z}\frac{dx}{ds} = \frac{1}{\lambda_3 + \lambda_4 z}\frac{dy}{ds} = \frac{-1}{\gamma_4}\frac{dz}{ds} \qquad (25)$$

corresponding to depths at some (possibly isolated) points.

Proof. Please refer to Appendix B. $\qquad \square$
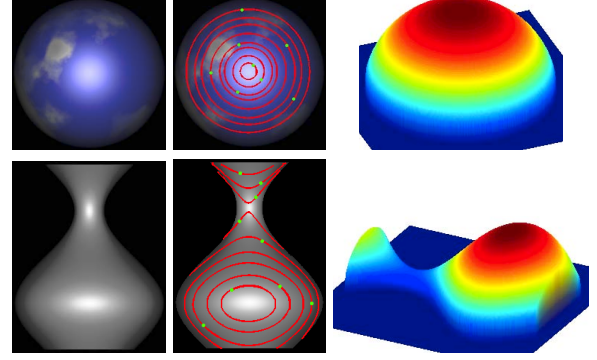
Input (1 of 6)　　(b) Level curves　　　　(c) Depth map

Figure 1. Given orthographic images (left) under five differential motions of a surface with non-Lambertian BRDF under colocated illumination, level curves of the surface are reconstructed using (24) (center). The curves in red represent all points with the same depth as the points marked in green. The surface may be reconstructed by interpolating between the level curves (right).
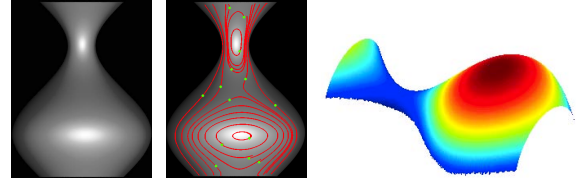
Input (1 of 6)　　Char. curves　　　　Depth map

Figure 2. Given orthographic images (left) under five differential motions of a surface with non-Lambertian BRDF under unknown lighting, characteristic curves of (25) are reconstructed (center). Initial information is provided at green points to compute depths along the corresponding red curve. The surface is reconstructed by interpolating between the characteristic curves (right).

4.4. Surface Reconstruction

Propositions 3 and 4 suggest an approach to surface reconstruction. Given depth z_0 at point $(x_0, y_0)^\top$, for a small step size ds, the relations (24) or (25) yield $(dx, dy, dz)^\top$, such that $(x_0 + dx, y_0 + dy)^\top$ lies on the characteristic curves of (18) through $(x_0, y_0)^\top$, with depth $z_0 + dz$. The process is repeated until the entire characteristic curve is estimated.

Note that dz is zero for the colocated case since characteristic curves correspond to level curves of depth, while it is in general non-zero for the non-colocated case. In practice, initial depths z_0 may be obtained from feature correspondences, or the occluding contour in the non-colocated case, as in [1].

Figure 1 illustrates the characteristic curves recovered for a synthetic sphere and vase, rendered under colocated illumination. Orthographic images are recorded for five differential motions, with arbitrary rotations of approximately $0.5°$ and translations 1 mm. Note the clear presence of specularities. After computing depths along several characteristic curves using the BRDF-invariant relation (23), we interpolate depths between the curves, to recover the entire surface geometry.

Figure 2 shows the same procedure for the non-colocated case, using the BRDF-invariant relation in (25).

5. Perspective Projection

In this section, we relax the assumption of orthography. Surprisingly, we obtain even stronger results in the perspective case, showing that with four or more differential motions with unknown BRDF, we can directly recover surface depth, as well as a linear constraint on the derivatives of the depth. Strictly speaking, our theory is an approximation in the perspective case, since viewing direction may vary over object dimensions, thus, $\nabla_{\mathbf{x}}\rho$ may be non-zero in (7). However, we illustrate in this section that accounting for finite focal length has benefits, as long as the basic assumption is satisfied that object dimensions are small compared to camera and source distance (which ensures that $\nabla_{\mathbf{x}}\rho$ is negligibly small).

5.1. Differential Flow Relation

In the perspective case, one may rewrite (8) as (compare to the linear relation in (10) for the orthographic case),

$$p'\left(\frac{z}{1+\beta z}\right) + r'\left(\frac{1}{1+\beta z}\right) + q' = \boldsymbol{\omega}^{\top}\boldsymbol{\pi}, \qquad (26)$$

where $p' = E_u\omega_2 - E_v\omega_1$, $q' = \alpha_1 E_u + \alpha_3 E_v + E_t$ and $r' = \alpha_2 E_u + \alpha_4 E_v$ are known entities, using (3).

Now, one may derive a theory similar to the orthographic case by treating $z/(1+\beta z)$, $1/(1+\beta z)$ and $\boldsymbol{\pi}$ as independent variables and using the rank deficiency (note the form of p') arising from a sequence of $m \geq 4$ differential motions. We leave the derivations as an exercise for the reader, but note that most of the observations in the preceding section for the orthographic case hold true in the perspective case too, albeit with the requirement of one additional image.

Instead, in the following, we take a closer look at the perspective equations for differential flow, to show that they yield a more comprehensive solution for surface geometry.

5.2. BRDF-Invariant Depth Estimation

We demonstrate that under perspective projection, object motion can completely specify the surface depth, without any initial information:

Proposition 5. *Four or more differential motions of a surface with unknown BRDF, under unknown illumination, suffice to yield under perspective projection:*
 (i) the surface depth
 (ii) a linear constraint on the derivatives of surface depth.

Proof. For $m \geq 4$, let images $E_1, \cdots, E_m$ be related to E_0 by known differential motions $\{\boldsymbol{\omega}^i, \boldsymbol{\tau}^i\}$, where $\boldsymbol{\omega}^i$ span $\mathbb{R}^3$. From (26), we have a sequence of differential flow relations

$$(p'^i + \beta q'^i)z - ((1+\beta z)\boldsymbol{\pi})^{\top}\boldsymbol{\omega}^i + (q'^i + r'^i) = 0, \quad (27)$$

for $i = 1, \cdots, m$. Let $\mathbf{c}^i = [p'^i + \beta q'^i, -\omega_1^i, -\omega_2^i, -\omega_3^i]^{\top}$ be the rows of the $m \times 4$ matrix $\mathbf{C} = \left[\mathbf{c}^1, \cdots, \mathbf{c}^m\right]^{\top}$. Let $\mathbf{q}' = [q'^1, \cdots, q'^m]^{\top}$ and $\mathbf{r}' = [r'^1, \cdots, r'^m]^{\top}$. Then, we may rewrite the system (27) as

$$\mathbf{C}\left[\begin{array}{c} z \\ (1+\beta z)\boldsymbol{\pi} \end{array}\right] = -(\mathbf{q}' + \mathbf{r}') \qquad (28)$$

which yields the solution

$$\left[\begin{array}{c} z \\ (1+\beta z)\boldsymbol{\pi} \end{array}\right] = -\mathbf{C}^{+}(\mathbf{q}' + \mathbf{r}') \qquad (29)$$

where $\mathbf{C}^{+}$ is the Moore-Penrose pseudoinverse of $\mathbf{C}$. Define $\boldsymbol{\epsilon} = -\mathbf{C}^{+}(\mathbf{q}' + \mathbf{r}')$ and $\boldsymbol{\epsilon}' = (\epsilon_2, \epsilon_3, \epsilon_4)^{\top}$. Then, we have

$$z = \epsilon_1, \quad (1+\beta z)\boldsymbol{\pi} = \boldsymbol{\epsilon}'. \qquad (30)$$

By definition, $\boldsymbol{\pi} = \mathbf{n} \times \nabla_{\mathbf{n}} \log \rho$, thus, $\mathbf{n}^{\top}\boldsymbol{\pi} = 0$. We now have two separate relations for depths and normals:

$$z = \epsilon_1 \qquad (31)$$

$$\mathbf{n}^{\top}\boldsymbol{\epsilon}' = 0. \qquad (32)$$

Thus, in the perspective case, one may directly use (31) to recover the surface depth. Further, noting that $n_1/n_3 = -z_x$ and $n_2/n_3 = -z_y$, we may rewrite (32) as

$$\epsilon_2 z_x + \epsilon_3 z_y - \epsilon_4 = 0, \qquad (33)$$

which is a linear constraint on surface depth derivatives. $\square$

Again, in the simpler case of colocated illumination, we observe that $\epsilon_4 = 0$, thus, the minimal imaging requirement is three motions. Further, from (32), the ratio $-\epsilon_2/\epsilon_3$ yields the slope of the gradient, leading to:

Corollary 1. *Three or more differential motions of a surface with unknown BRDF, under unknown illumination, suffice to yield under perspective projection the surface depth and the slope of the gradient.*

Thus, we have shown that in the perspective case, even when BRDF and illumination are unknown, one may derive an invariant that relates shape to object motion, through a linear relation and a linear PDE on the surface depth. Again, we note that this is a fundamental constraint, independent of any particular reconstruction approach.

5.3. Surface Reconstruction

5.3.1 Direct Depth Recovery

As established by Proposition 5, under perspective projection, one may directly recover the surface depth using (31). Figure 3 illustrates this with synthetic data. An object with unknown BRDF is imaged with perspective projection under unknown illumination after undergoing four arbitrary differential motions (approximately $0.5°$ rotation and 1 mm translation each). Note that no prior knowledge of the surface is required in the perspective case, even at isolated points.

5.3.2 Combining Depth and Normal Information

Recall that Proposition 5 in fact supplies an additional constraint on the surface gradient at every point. Thus, one may solve the following linear system that combines the two constraints (31) and (33) on depths and gradients:
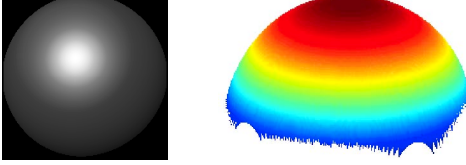
$$\min_{z(x,y)} (z - \epsilon_1)^2 + \lambda(\epsilon_2 z_x + \epsilon_3 z_y - \epsilon_4)^2, \qquad (34)$$

where λ is a relative weighting term. Standard discretization schemes may be used to represent z_x and z_y. Then, the above
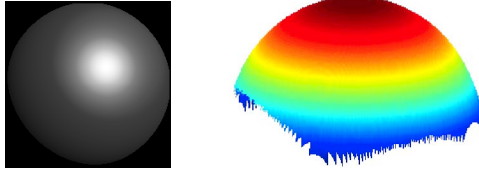
(a) Input image (1 of 5)   (b) Reconstructed surface

Figure 3. (a) One of five images (four motions) under perspective projection, with arbitrary non-Lambertian BRDF and unknown lighting. (b) Depth map estimated using Proposition 5.

(a) Input image (1 of 5)   (b) Reconstructed surface

Figure 4. (a) One of five images (four motions), with arbitrary non-Lambertian BRDF and unknown lighting, under perspective projection. (b) Surface recovery using (34).

is a highly sparse linear system in the depths z, which may be solved using a linear least squares solver.

Incorporating gradient constraints has the effect of regularizing the depth estimation by introducing neighborhood information, which may be advantageous in noisy scenarios. Figure 4 illustrates shape recovery using (34) with synthetic data. The object is imaged under perspective projection after undergoing five random differential motions (approximately $0.5°$ rotation and 1 mm translation each).

6. Stratification of Shape from Motion

Our theory not only shows the possibility of shape recovery under unknown BRDFs and lighting, but also derives the minimum computational and imaging budget required and the precise extent to which surface shape may be recovered. In this section, we discuss how this work establishes a theoretical notion that relates the hardness of surface reconstruction to the scene complexity, which supports our intuitive understanding of an "easy" reconstruction problem versus a "hard" one.

Generalization of [1] For the Lambertian BRDF, under known directional lighting, Basri and Frolova [1] show that shape and image derivatives may be related by a quasilinear PDE. They use special considerations of the two-view setup to arrive at the result. In the context of our theory, under a directional light source $\mathbf{s} = (s_1, s_2, 1)^\top / \sqrt{s_1^2 + s_2^2 + 1}$, we have $\rho(\mathbf{n}) = \mathbf{n}^\top \mathbf{s}$. Then, we may rewrite the basic relation in (8) as (9). For the orthographic case, using (11) and (12), we may again rewrite (9) as:

$$pz + q = \frac{\lambda_1' z_x + \lambda_2' z_y + \lambda_3'}{-s_1 z_x - s_2 z_y + 1}, \quad (35)$$

with known scalars $\lambda_1' = \omega_2 - \omega_3 s_2$, $\lambda_2' = -\omega_1 + \omega_3 s_1$ and $\lambda_3' = -\omega_1 s_2 + \omega_2 s_1$. Note that (35) is a quasilinear PDE, as expected from the result of [1]. It may be verified that the perspective case can also be written as a quasilinear PDE:

$$\frac{(p' + \beta q')z + (q' + r')}{1 + \beta z} = \frac{\lambda_1' z_x + \lambda_2' z_y + \lambda_3'}{-s_1 z_x - s_2 z_y + 1}. \quad (36)$$

Note that the above are obtained as by-products of a theory far more powerful than [1]. In particular, the framework of this paper can also handle general BRDFs, unknown directional or area lighting and various camera projections.

General Stratification The above result and those derived in Sections 4 and 5 suggest a general stratification of shape recovery from object motion, shown in Table 1. Shape may be recovered under Lambertian BRDF and known lighting with one motion, while an unknown BRDF with colocated lighting requires two motions under orthography and three under perspective projection. An unknown BRDF, unknown illumination and perspective projection may be considered an even "harder" reconstruction problem, for which the minimal imaging requirement is four motions. Similarly, simpler illuminations like colocated lighting result in a homogeneous PDE, while more complex illuminations make the PDE inhomogeneous, whose solution is arguably harder.

Thus, the nature of the reconstruction invariant is a qualitative indicator of the hardness of shape from motion. For a given scene complexity, the hardness of reconstruction is quantified by the minimum number of motions specified by our theory for reconstruction to be possible. These conclusions are summarized in Table 1.

7. Experiments on Real Data

In this section, we demonstrate our theory on real data. While it requires committing to a specific acquisition setup, we note that our theory pertains to fundamental limits on reconstruction and will be valid for other choices too.

Setup We use a Canon EOS 5D camera, with focal length 50 mm, placed approximately 0.7 m away from the object with unknown BRDF. Our acquisition environment has no special design to simulate a dark room. The illumination can be arbitrary, in this case, an off-the-shelf light bulb at an unknown position. The object is mounted on a platform whose motion is measured (but not controlled) using a gyroscope. The camera is calibrated and a hand-eye calibration is performed to transform the gyroscope motions into the camera coordinate system. Arbitrary small displacements are manually imparted to the platform and the gyroscope motion is recorded. Typical differential motions are about $0.5 - 1°$ for rotation and $1 - 10$ mm for translation.

For objects with distinctive texture, an alternative is to track correspondences and use structure from motion to determine relative camera poses in the object's frame of reference. Again, we note that the results of our theory will continue to hold for that or any other imaging setup.
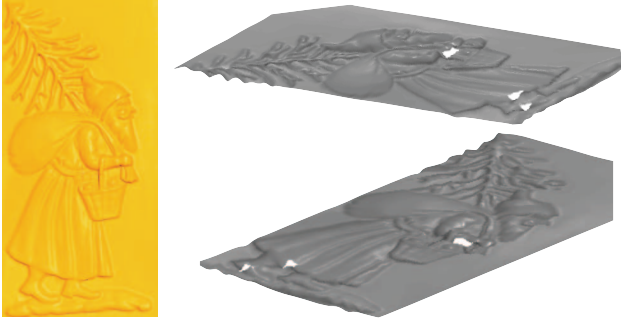
Figure 5. (Left) One of seven input images, related by six differential motions, of a real object with unknown BRDF, acquired under unknown illumination. Note the clearly visible non-Lambertian BRDF effects. (Right) Views of surface reconstructed using the theory of Section 5. Note that our theory correctly accounts for intensity variations even when the BRDF and illumination are arbitrary and unknown, thus, a high level of detail is recovered.

Reconstruction Since the above setup results in a perspective projection, we use the theory of Section 5. Logarithms are evaluated for the input images to eliminate albedo variations. Spatial derivatives are computed using central differencing, with a smoothing filter to suppress high frequency differentiation noise. Temporal derivatives are computed by simply subtracting images smoothed by a Gaussian kernel. The linear system defined by (28) is set up at every pixel and solved as (29). The solution yields the depth map (31), in addition to a linear constraint on the gradient (32).

In Figure 5, we show the reconstruction for a porcelain bas-relief sculpture with fine depth variations, using six differential motions. It is apparent from the clearly visible gloss and specularities in the input images, shown on the left, that the unknown BRDF is non-Lambertian. The reconstructed surface using the solution in (31) is shown on the right. Note the high level of surface detail that is recovered, without any knowledge of BRDF or illumination. This demonstrates the practical utility of our theory which does not assume brightness constancy or simplified forms of the BRDF and illumination, rather it correctly accounts for shading changes through an invariant that eliminates the BRDF and lighting.

In Figure 6, we show a similar reconstruction for a plastic shell. Again, note that the unknown BRDF is clearly non-Lambertian and the lighting is unknown, yet our theory allows surface reconstruction with fine details.

8. Discussion and Future Work

This paper answers the question of what motion reveals about shape, with unknown isotropic BRDF and arbitrary, unknown distant illumination, for orthographic and perspective projections. We derive differential flow invariants that relate image derivatives to shape and exactly characterize the object geometry that can be recovered. This work generalizes traditional notions of brightness constancy or Lambertian BRDFs in the optical flow and multiview stereo literatures.
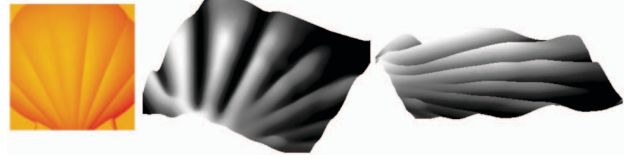

Figure 6. (Left) One of eleven input images, related by ten differential motions, of a real object with unknown BRDF, acquired under unknown illumination. Note the clearly visible non-Lambertian BRDF effects. (Center) Top view of the surface reconstructed using the theory of Section 5. (Right) Side view of the same reconstruction. Correctly accounting for intensity variations, in spite of unknown BRDF and illumination, allows us to recover surface details such as the fine striations between the lobes of the shell.

Our results are not just valid for a particular approach to reconstruction, rather they impose fundamental limits on the hardness of surface reconstruction. In the process, we also present a stratification of shape from motion that relates hardness of reconstruction to scene complexity – qualitatively in terms of the nature of the involved PDE and quantitatively in terms of the minimum number of required motions.

Many of the relations, such as (19), (35) and (36) may be expressed in the form $f(z) = g(\mathbf{n})$. With the availability of depth sensors, it becomes possible to measure $f(z)$, making the optimization problem to solve for only $\mathbf{n}$ easier. An interesting future direction would be to study the accuracy and convergence of alternating minimization approaches to simultaneously estimate depth and normals.

Lighting, object motion and viewpoint change are fundamental cues to understanding shape. This work and previous results have shown that differential motions carry rich information, independent of the lighting and BRDF. This paper developed a general theory to understand the information carried in the motion cue, while prior works like [3] have taken a first step towards considering differential photometric stereo for the lighting cue. An interesting future direction would be to consider small viewpoint changes, as well as a unified framework that combines all the differential cues.

A. General BRDF Differential Flow Relation

Recall that in our setup, the camera and illumination are fixed, while the object is moving. Following optical flow studies [11, 17], we make a distinction between entities directly expressed in terms of intrinsic surface coordinates (such as albedo) and those expressed in 3D coordinates (such as camera direction).

The 3D position vector of a point on the surface at time t is $\mathbf{x}(a, b, t)$, while the corresponding surface normal is $\mathbf{n}(a, b, t)$. Consider a point $\mathbf{u} = (x, y)^\top$ on the image. At time t, it is the image of the point $\mathbf{x}(a, b, t)$. At time $t + \delta t$, it is the image of

$\mathbf{x}(a - \delta a, b - \delta b, t + \delta t)$. Let $\delta\mathbf{x}$ denote that displacement, then the corresponding change in surface normal is

$$\delta\mathbf{n} = \boldsymbol{\omega} \times \mathbf{n}(a - \delta a, b - \delta b, t)\delta t. \tag{37}$$

The isotropic BRDF on the surface is a function of normal and position, denoted by $\rho(\mathbf{n}, \mathbf{x})$. Let the albedo, which is an intrinsic surface property, be $\sigma(a, b)$. Then, the image at pixel $\mathbf{u}$ at time t is:

$$I(\mathbf{u}, t) = \sigma(a, b)\ \rho(\mathbf{n}(a, b, t), \mathbf{x}(a, b, t)). \tag{38}$$

Let $(a', b') = (a - \delta a, b - \delta b)$ and $t' = t + \delta t$. Let the image of $\mathbf{x}(a', b', t)$ be $I(\mathbf{u}', t)$, where $\mathbf{u}' = \mathbf{u} - \delta\mathbf{u}$. Then we have:

$$I(\mathbf{u}, t') = \sigma(a', b')\rho\left(\mathbf{n}(a', b', t'), \mathbf{x}(a', b', t')\right)$$
$$= I(\mathbf{u}', t) + \sigma\left[(\boldsymbol{\omega} \times \mathbf{n})^\top \nabla_\mathbf{n}\rho + \boldsymbol{\nu}^\top \nabla_\mathbf{x}\rho\right]\delta t \tag{39}$$

where the surface entities correspond to (a', b', t) and $\boldsymbol{\nu}$ denotes the velocity of $\mathbf{x}(a', b', t)$. Subtracting $I(\mathbf{u}, t)$ from (39), we have

$$\frac{\partial I}{\partial t}\delta t = I(\mathbf{u}, t') - I(\mathbf{u}, t)$$
$$= -\nabla_\mathbf{u}I\delta\mathbf{u} + \sigma\left[(\boldsymbol{\omega} \times \mathbf{n})^\top \nabla_\mathbf{n}\rho + \boldsymbol{\nu}^\top \nabla_\mathbf{x}\rho\right]\delta t \tag{40}$$

For consistency, we make the transformation $\mathbf{u} \rightarrow \mathbf{u} + \delta\mathbf{u}$ and $\mathbf{a} \rightarrow \mathbf{a} + \delta\mathbf{a}$. Then, the above basic relation can be written as

$$\boxed{(\nabla_\mathbf{u}I)^\top \boldsymbol{\mu} + \frac{\partial I}{\partial t} = \sigma\left[(\nabla_\mathbf{n}\rho)^\top(\boldsymbol{\omega} \times \mathbf{n}) + (\nabla_\mathbf{x}\rho)^\top \boldsymbol{\nu}\right].} \tag{41}$$

where $\boldsymbol{\mu}$ is the velocity of the image point (the motion field), which gives us (7).

B. Proof of Proposition 4

The proof uses standard constructs from PDE theory [7], illustrated here for completeness.

Proof. Consider the PDE in (18), given by

$$(\lambda_1 + \lambda_2 u)u_x + (\lambda_3 + \lambda_4 u)u_y + \gamma_4 = 0, \tag{42}$$

where γ_4 and λ_i, $i = 1, \cdots, 4$, are known functions of (x, y). Section 4.2 establishes that the integral surface of (42), $\mathcal{S} : z = u(x, y)$, is indeed the surface under consideration and the coefficient functions can be obtained from three or more differential motions of the surface. With $\mathbf{a} = (a_1, a_2, a_3)^\top = (\lambda_1 + \lambda_2 u, \lambda_3 + \lambda_4 u, -\gamma_4)^\top$, we rewrite (42) as $\mathbf{a}^\top(\nabla u^\top, -1)^\top = 0$. Then, we note that the integral surface $\mathcal{S} : z = u(x, y)$ is tangent everywhere to the vector field given by $\mathbf{a}$. Consider the curves $\mathcal{C}(x(s), y(s), z(s))$, given by (25), written in terms of a parameter $s \in \mathbf{I} \subset \mathbb{R}$. We note that the curves $\mathcal{C}$, if they exist, have $\mathbf{a}$ as tangent directions. Next, we derive the relationship between $\mathcal{C}$ and $\mathcal{S}$, in particular, we show that if a point $\mathbf{p} = (x_0, y_0, z_0)^\top \in \mathcal{C}$ lies on $\mathcal{S}$, then $\mathcal{C} \subset \mathcal{S}$.

Let there exist $\mathbf{p} = (x_0, y_0, z_0)^\top \in \mathcal{C}$, such that $\mathbf{p} \in \mathcal{S}$, that is,

$$x_0 = x(s_0),\ y_0 = y(s_0),\ z_0 = z(s_0) = u(x_0, y_0). \tag{43}$$

for some parameter value $s = s_0 \in \mathbf{I}$. Next, we define

$$w = w(s) = z(s) - u(x(s), y(s)). \tag{44}$$

Then, it is clear that $w(s)$ is the solution to the initial value problem

$$\frac{dw}{ds} = (u_x, u_y, u_z)^\top \mathbf{a}(x, y, w + u),\quad w(s_0) = 0. \tag{45}$$

Further, we note that $w = 0$ is a particular solution of the above ODE, since $z = u(x, y)$ is a solution to (42). Also, the solution to (45) must be unique. Thus, $z(s) = u(x(s), y(s))$, which establishes that $\mathcal{C} \subset \mathcal{S}$. This completes the proof that the characteristic curves $\mathcal{C}$, given by (25), reside on the surface $\mathcal{S}$. $\square$

References

[1] R. Basri and D. Frolova. A two-frame theory of motion, lighting and shape. In *CVPR*, pages 1–7, 2008.

[2] G. D. Canas, Y. Vasilyev, Y. Adato, T. Zickler, S. J. Gortler, and O. Ben-Shahar. A linear formulation of shape from specular flow. In *ICCV*, pages 191–198, 2009.

[3] M. Chandraker, J. Bai, and R. Ramamoorthi. A theory of differential photometric stereo for unknown isotropic BRDFs. In *CVPR*, pages 2505–2512, 2011.

[4] M. Chen and J. Arvo. Theory and application of specular path perturbation. *ACM ToG*, 19(4):246–278, 2000.

[5] H. W. Haussecker and D. J. Fleet. Computing optical flow with physical models of brightness variation. *PAMI*, 23(6):661–673, 2001.

[6] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[7] F. John. *Partial Differential Equations*. Applied Mathematical Sciences. Springer, 1981.

[8] K. M. Lee and C.-C. J. Kuo. Shape from shading with a generalized reflectance map model. *CVIU*, 67:143–160, 1997.

[9] J. Lim, J. Ho, M.-H. Yang, and D. Kriegman. Passive photometric stereo from motion. In *CVPR*, pages 1635–1642, 2005.

[10] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Image Understanding Workshop*, pages 121–130, 1981.

[11] H.-H. Nagel. On a constraint equation for the estimation of displacement rates in image sequences. *PAMI*, 11(1):13 –30, 1989.

[12] S. Negahdaripour. Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis. *PAMI*, 20(9):961–979, 1998.

[13] R. Ramamoorthi, D.Mahajan, and P. Belhumeur. A first order analysis of lighting, shading and shadows. *ToG*, 26(1), 2007.

[14] I. Sato, T. Okabe, Q. Yu, and Y. Sato. Shape reconstruction based on similarity in radiance changes under varying illumination. In *ICCV*, pages 1–8, 2007.

[15] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multiview stereo reconstruction algorithms. In *CVPR*, pages 519–526, 2006.

[16] D. Simakov, D. Frolova, and R. Basri. Dense shape reconstruction of a moving object under arbitrary, unknown lighting. In *ICCV*, pages 1202–1209, 2003.

[17] A. Verri and T. Poggio. Motion field and optical flow: Qualitative properties. *PAMI*, 11(5):490–498, 1989.

[18] L. Zhang, B. Curless, A. Hertzmann, and S. Seitz. Shape from motion under varying illumination: Unifying structure from motion, photometric stereo and multiview stereo. In *ICCV*, pages 618–625, 2003.

[19] T. Zickler, P. Belhumeur, and D. Kriegman. Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. *IJCV*, 49(2/3):1215–1227, 2002.