A Lighting-Invariant Point Processor for Shading

Kathryn Heal Jialiang Wang Steven J. Gortler Todd Zickler
Harvard University
{kathrynheal@g, jialiangwang@g, sjg@cs, zickler@seas}.harvard.edu

Abstract

Under the conventional diffuse shading model with unknown directional lighting, the set of quadratic surface shapes that are consistent with the spatial derivatives of intensity at a single image point is a two-dimensional algebraic variety embedded in the five-dimensional space of quadratic shapes. We describe the geometry of this variety, and we introduce a concise feedforward model that computes an explicit, differentiable approximation of the variety from the intensity and its derivatives at any single image point. The result is a parallelizable processor that operates at each image point and produces a lighting-invariant descriptor of the continuous set of compatible surface shapes at the point. We describe two applications of this processor: two-shot uncalibrated photometric stereo and quadratic-surface shape from shading.

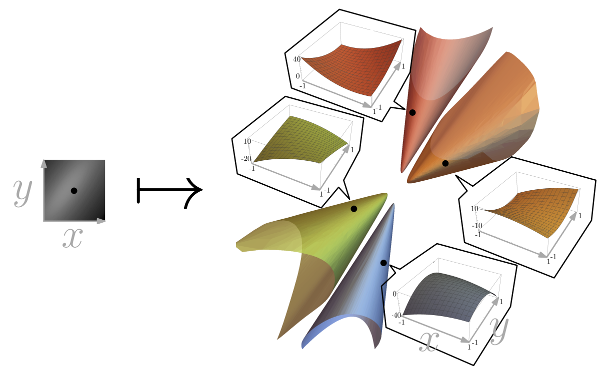

Figure 1. The set of local second order surface shapes $\{(f_x, f_y, f_{xx}, f_{xy}, f_{yy})\}$ that are consistent with the derivatives $\mathbf{I} = (I, I_x, I_y, I_{xx}, I_{xy}, I_{yy})$ at one image point (black circle, left) satisfy three polynomial equations. The zero locus (*i.e.*, variety) is two-dimensional and is visualized here projected to three dimensions (f_{xx}, f_{xy}, f_{yy}). Each element of the variety is a local shape (four are called out) that produces the image derivatives under some light direction. We show that for any non-degenerate $\mathbf{I}$ the two-dimensional variety has four isomorphic components (colored in this example) and can be efficiently approximated by a coupled pair of shallow neural networks.

1. Introduction

The shading variations in an image $I(x, y)$ of a diffuse, curved surface—say, a surface with height function $f(x, y)$—induce a perception of the surface shape. Mimicking this perceptual capability in machines is referred to as recovering "shape from shading."

There exist established techniques for recovering shape from shading in special cases where the strengths and locations of the light sources around the surface are known *a priori*, or are somehow accurately inferred. These techniques can be understood as using a connected two-dimensional array of image "point processors", where each point processor reads the intensity I at a single image point and, based on the known or estimated lighting conditions, calculates an intermediate numerical representation of the set of compatible local shapes at that point, comprising a set of (or probability density over) local surface orientations $\{(f_x, f_y)\}$ at the point. Each of the intermediate per-point orientation sets is ambiguous on its own, but when the array of point processors is connected together—by enforcing surface continuity and by including supplementary visual cues like occluding contours or top-down semantics—one can begin to recover shapes $f(x, y)$.

This has been the dominant paradigm for shape from shading for nearly fifty years [10], but it is far from satisfactory. Despite a half-century of research, it remains sensitive to non-idealities and is rarely deployed without substantial aid from a human annotator who first indicates occluding contours in an image or provides a segmentation of a relevant diffuse surface region. One reason for this fragility is that lighting is typically non-uniform across surfaces, due to self-shadowing and other physical effects. This makes it hard to infer the lighting conditions for each image point, which in turn distorts the per-point orientation sets $\{(f_x, f_y)\}$ upon which reconstruction is based. Moreover, even when lighting *is* uniform across a surface, the veridical location and strength of a scene's dominant light source can be impossible to infer from an image due to inherent mathematical ambiguities [3]. In comparison, monocular human vision seems to perform quite well at perceiving diffusely-shaded shape, at least modulo these ambiguities [12], despite being quite poor at inferring lighting [4].

This paper introduces a point processor for shading that might help address these deficiencies, by providing per-point constraints on shape *without* requiring knowledge of lighting. The input to the processor is a measurement comprising a vector of spatial derivatives of intensity at one point, denoted by $\mathbf{I} := (I, I_x, I_y, I_{xx}, I_{xy}, I_{yy})$, Koenderink's 2-jet [11]. The internal structure of the processor is a coupled pair of shallow neural networks, and the processor's output is a compact representation of a continuous set of compatible local second-order shapes $F(\mathbf{I}) := \{(f_x, f_y, f_{xx}, f_{xy}, f_{yy})\}$ in the form of a parametrized two-dimensional manifold in $\mathbb{R}^5$. The processor provides useful per-point constraints because even though there are many compatible shapes $F(\mathbf{I})$, the overwhelming majority of shapes are ruled out.

Our main contribution is an algebraic analysis of Lambertian shading that provides the foundation for the point processor's internal structure and the format of its output. Specifically, we prove that the set of compatible local second-order shapes $F(\mathbf{I})$ are contained in the zero-set of three polynomial equations, *i.e.*, are contained in a two-dimensional algebraic variety in $\mathbb{R}^5$. We show that special properties of this variety allow it to be represented in explicit form by a function from $\mathbb{R}^2$ to $\mathbb{R}^3$, which in turn can be approximated efficiently by a coupled pair of shallow neural networks.

The most important property of this point processor is that it is "invariant to illumination" in the sense that the output shape-set $F(\mathbf{I})$ always includes the veridical local second-order shape, regardless of how the surface is lit. This means that while a surface lit from different directions will generally induce different measurements $\mathbf{I}$ at a point, and while these different image measurements will in turn produce different shape-sets $F(\mathbf{I})$, all of the predicted shape-sets will include the true second-order shape at that point.

As examples of how the point processor can be used for image analysis, we describe two scenarios in which the intrinsic two-dimensional shape ambiguities $F(\mathbf{I})$ at each point can be reduced to a discrete four-way choice by exploiting additional constraints or information. One scenario is uncalibrated two-shot photometric stereo, where the input is two images of a surface under two unknown light directions. The other is quadratic shape from shading, where the input is a single image of a shape that is quadratic over an extended region. We demonstrate these using synthetic images, leaving the development of robust algorithms and deployment on captured photographs for future work.

Throughout this paper, we assume a frame of reference such that our measurements are graphs of some polynomial function. We represent these local surface height and image values as vectors of their coefficients – applying the Monge-Taylor map – ignoring dependence of f_{xx} on f_x. This is to say that we are not attempting to solve any partial differential equations; instead, we are studying algebraic constraints in a local linear coefficient coordinate space.

2. Background and Related Work

Most approaches to shape from shading rely on a per-point relationship between scalar intensity I and surface orientation (f_x, f_y). If the lighting is from a single direction, for example, then the set of compatible orientations is a right-circular cone with axis equal to the light direction and apex angle proportional to intensity. Similarly, if the lighting is a positive-valued function on the directional two-sphere then the set of compatible orientations is well approximated by a one-dimensional manifold defined by the light function's spherical harmonic coefficients up to third degree [15, 2]. Regardless, any such relation between intensity and surface orientation necessarily requires prior knowledge of, or accurate estimates of, the lighting at every surface point. Despite substantial recent progress [19, 1, 16, 8], including the abilities to accommodate some amounts of non-uniform lighting and non-uniform surface material properties, obtaining useful results continues to require extensive help from a human, who must first label the region that contains a continuous surface and/or indicate the locations of occluding contours.

In contrast, we follow Kunsberg and Zucker [14] by enhancing the per-point analysis to consider not just the intensity and surface orientation at a point, but also higher order derivatives of intensity and shape. This allows eliminating the dependence on lighting entirely, and it suggests the possibility of a different approach where perceptual grouping and shape reconstruction can occur without explicit knowledge of lighting, and perhaps with lighting being (approximately) inferred later, as a by-product of shape perception. In this paper we consider just the first step toward this possibility: the design of the essential point processor.

We are also motivated by the results of Xiong et al. [17], who consider a lighting-invariant local *area* processor instead of a pure point processor, and show that the intensity values in an extended image patch determine the extended quadratic shape up to a discrete four-way choice. This four-way choice leads to the automorphism (i.e. an bijection from a space to itself) group that we describe in Section 4.

Our work is complementary to recent learning-based approaches to monocular depth estimation (*e.g.*, [6]) that aim to exploit diffuse shading and many other bottom-up cues while also exploiting contextual cues in large image datasets. Our goal is to explore alternative front-end architectures and interpretable intermediate representations that can improve the generality and efficiency of such systems in the future.

3. Local Shape Sets as Algebraic Varieties

Our illumination-invariant point processor is inspired by the work of Kunsberg and Zucker [14], who use differential geometry to derive three lighting-invariant rational equations that relate the image 2-jet $\mathbf{I}$ at a point to the surface height derivatives at that point. We take an algebraic-geometric approach instead, which provides an abbreviated derivation of equivalent equations and also reveals the shape-set to be contained in an algebraic variety (*i.e.* in the zero-set of certain polynomial equations) that, as will be seen in Section 4, has useful geometric structure.

3.1. Shading and Surface Models

Our analysis applies to any point in a 2D image. We assign the coordinates $(0, 0)$ to the point of interest and let $I(x, y)$ denote the intensity in a bounded local neighborhood $U \subset \mathbb{R}^2$ of that point. We refer to U as the *receptive field*. In practice it is no larger than is required to robustly compute a discrete approximation to the first and second spatial derivatives of $I(x, y)$ at the origin.

Within the neighborhood U, we assume that the image is the orthographic projection of a curved Lambertian surface, and that the surface can be represented by a height function $f(x, y)$. The surface albedo $\rho \in \mathbb{R}^+$ is assumed to be constant within U. We also assume that the lighting is uniform and directional within U, so that it can be represented by $\mathbf{L} \in \mathbb{R}^{\mathbf{3}}$ with strength $\|\mathbf{L}\|$ and direction $\mathbf{L}/\|\mathbf{L}\|$. Under these assumptions the intensity is

$$I(x, y) = \rho \mathbf{L} \cdot \frac{\mathbf{N}(x, y)}{||\mathbf{N}(x, y)||}, \qquad (x, y) \in U, \tag{1}$$

where $\mathbf{N}(x, y) := (-(\partial f/\partial x)(x, y), -(\partial f/\partial y)(x, y), 1)^T$ is the normal field. Note that we allow for the projection, albedo, and lighting to vary outside of neighborhood U.

We assume that the surface f is locally smooth enough around the point (x, y) that we can ignore any third or higher order derivatives at that point.

$$f(x,y) = f_x x + f_y y + \frac{1}{2}\left(f_{xx}x^2 + 2f_{xy}xy + f_{yy}y^2\right). \tag{2}$$

We refer to $\mathbf{f} := (f_x, f_y, f_{xx}, f_{xy}, f_{yy}) \in \mathbb{R}^5$ as the *local shape* at the point (x, y). We assume that all local shapes are not flat or cylindrical, or more precisely are nondegenerate in this sense:

Definition 1. *A local shape* $\mathbf{f}$ *is* nondegenerate *if* $(f_{xx} + f_{yy})(f_{xx}f_{yy} - f_{xy}^2)(4f_{xy}^2 + (f_{xx} - f_{yy})^2) \neq 0$.

Local shapes can produce many different image intensity patterns depending on the lighting direction. We call the set of all possible image 2-jets generated by any combination of local shape and lighting *realizable*, and we say that a realizable image 2-jet produced by a particular shape is *consistent* with that shape.

Definition 2. *The set of* realizable measurements $\mathcal{I}$ *is the set of vectors* $\boldsymbol{v} \in \mathbb{R}^6$ *for which there exists a light direction* $\boldsymbol{L} \in \mathbb{R}^3$ *and nondegenerate local shape* $\mathbf{f}$ *such that* $\boldsymbol{v} = \mathbf{I}$ *when shape model* (2) *is combined with shading model* (1).

Definition 3. *If for a pair* $(\mathbf{I}, \mathbf{f}) \in \mathcal{I} \times \mathbb{R}^5$ *there exists such an* $\boldsymbol{L}$*, we say that* $\mathbf{I}$ *and* $\mathbf{f}$ *are* consistent*. This means that for some light direction,* $\mathbf{f}$ *is a valid explanation of image measurements* $\mathbf{I}$.

3.2. Sets of Local Shapes

Our immediate goal is to characterize the set of shapes $F(\mathbf{I})$ that are consistent with observation $\mathbf{I}$ for *any* light direction. This set of admissible shapes turns out to be contained in the locus of real solutions to three polynomial equations. An important feature is that the albedo and lighting do not appear in these equations.

Theorem 1. *Assume the shading model of* (1) *and the surface model of* (2)*, and suppose we are given a measurement* $\mathbf{I} \in \mathcal{I}$ *generated by some unknown surface/lighting combination. Define polynomials*

$$\begin{aligned}C_1(\mathbf{f};\mathbf{I}) :=\; & f_x^4 I_{xx} + 2f_x^3 f_{xx} I_x + f_x^2 f_{xy}^2 I + 2f_x^2 f_{xy} f_y I_x + 2f_x^2 f_y^2 I_{xx} \\ & + 2f_x^2 I_{xx} - 2f_x f_{xx} f_{xy} f_y I + 2f_x f_{xx} f_y^2 I_x + 2f_x f_{xx} I_x \\ & + f_{xx}^2 f_y^2 I + f_{xx}^2 I + f_{xy}^2 I + 2f_{xy} f_y^3 I_x + 2f_{xy} f_y I_x \\ & + f_y^4 I_{xx} + 2f_y^2 I_{xx} + I_{xx},\end{aligned} \tag{3}$$

$$\begin{aligned}C_2(\mathbf{f};\mathbf{I}) :=\; & f_x^4 I_{yy} + 2f_x^3 f_{xy} I_y + 2f_x^2 f_y^2 I_{yy} + 2f_x^2 f_y f_{yy} I_y + f_x^2 f_{yy}^2 I \\ & + 2f_x^2 I_{yy} + 2f_x f_{xy} f_y^2 I_y - 2f_x f_{xy} f_y f_{yy} I + 2f_x f_{xy} I_y \\ & + f_{xy}^2 f_y^2 I + f_{xy}^2 I + f_y^4 I_{yy} + 2f_y^3 f_{yy} I_y + 2f_y^2 I_{yy} \\ & + 2f_y f_{yy} I_y + f_{yy}^2 I + I_{yy},\end{aligned} \tag{4}$$

$$\begin{aligned}C_3(\mathbf{f};\mathbf{I}) :=\; & f_x^4 I_{xy} + f_x^3 f_{xx} I_y + f_x^3 f_{xy} I_x + f_x^2 f_{xy} f_y I_y + f_x^2 f_{xy} f_{yy} I \\ & + 2f_x^2 f_y^2 I_{xy} + f_x^2 f_y f_{yy} I_x + 2f_x^2 I_{xy} + f_x f_{xx} f_y^2 I_y \\ & - f_x f_{xx} f_y f_{yy} I + f_x f_{xx} I_y - f_x f_{xy}^2 f_y I \\ & + f_x f_{xy} f_y^2 I_x + f_x f_{xy} I_x + f_{xx} f_{xy} f_y^2 I + f_{xx} f_{xy} I \\ & + f_{xy} f_y^3 I_y + f_{xy} f_y I_y + f_{xy} f_{yy} I + f_y^4 I_{xy} + f_y^3 f_{yy} I_x \\ & + 2f_y^2 I_{xy} + f_y f_{yy} I_x + I_{xy}.\end{aligned} \tag{5}$$

Then, any nondegenerate local shape $\mathbf{f} \in \mathbb{R}^5$ *that is a valid explanation of measurements* $\mathbf{I}$ *will satisfy* $C_i = 0\ \forall i$*. Equivalently, the affine variety* $F := \boldsymbol{V}(C_1, C_2, C_3)$ *contains the set of all shapes* $\mathbf{f}$ *consistent with* $\mathbf{I}$.

Proof sketch. We provide a sketch of the proof here, with details in the supplement. Begin with (1), absorbing albedo ρ into (non-unit length) $\mathbf{L}$. Introduce auxiliary variable w, which plays the role of $1/||\mathbf{N}(x, y)||$. Substitution into (1) yields polynomials $g_1(x, y, w, \mathbf{f}) := I(x, y) - \mathbf{L} \cdot \mathbf{N}(x, y)$ and $g_2(x, y, w, \mathbf{f}) := w^2||\mathbf{N}(x, y)||^2 - 1$. Calculate first and second spatial derivatives of g_1 with respect to x, y, evaluate all polynomials at $(x, y) = (0, 0)$, and re-arrange to eliminate variables $\mathbf{L}$ and w. $\square$

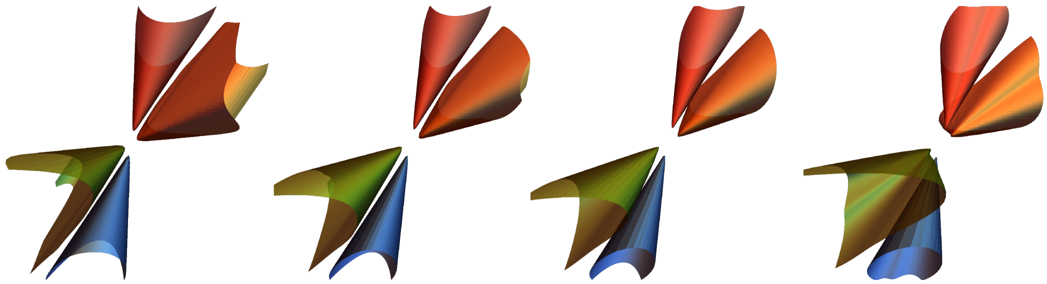

Figure 2. Visualizations of the two-dimensional varieties for different measurements of the form $\mathbf{I} \approx (1-t, \text{-}4.10, \text{-}5.87, \text{-}12.41, \text{-}13.41, \text{-}20.30) + t$. Each variety is projected to the same three dimensions as in Figure 1 and is colored by its isomorphic pieces.

Remark 1. *The real solutions to these equations are identical to those of Corollary 4.2 of [14]; we offer our algebraic derivation as an alternative to the differential-geometric approach presented in that work.*

Theorem 1 states that the set of local shapes that are consistent with a given measurement $\mathbf{I}$ must satisfy a set of three algebraically independent polynomials and thus, by definition, is contained in a real two-dimensional algebraic variety embedded in the five-dimensional shape space. (We use the notation $\mathbf{V}(\cdot)$ to denote the variety corresponding to a set of polynomials. This is essentially their zero locus.) This variety is analogous to the one-dimensional manifold of surface orientations in classical shape from shading, and it provides substantial constraints on local shape, because although there are still infinitely many admissible local shapes, the vast majority of shapes are disqualified.

The variety for a particular measurement $\mathbf{I}$ is visualized in Figure 1, projected from the five-dimensional shape space to a three dimensional space that corresponds to the second-order shape dimensions (f_{xx}, f_{xy}, f_{yy}). Additional examples are in Figure 2, which shows how the varieties change for different measurements.

4. Properties of Local Shape Sets

At this stage we have an implicit description of a shape set $F(\mathbf{I})$ in terms of generating polynomials (3)-(5). For a useful point processor, we want instead an *explicit* representation, as well as an efficient way to calculate (and store) that explicit representation for any particular image 2-jet $\mathbf{I}$. An explicit analytic representation remains out of reach[1], but fortunately the varieties exhibit three properties that make them easy to approximate.

First we show that the variety is equipped with an automorphism group that naturally divides it into four isomorphic pieces, allowing the entire shape set to be represented by any one piece (Section 4.1). We then relate the one piece to a continuous function $\phi_{\mathbf{I}}$ from $\mathbb{R}^2$ to $\mathbb{R}^3$, which implies that the point processor is equivalent to a map from vectors $\mathbf{I} \subset \mathbb{R}^6$ to continuous functions $\phi_{\mathbf{I}} : \mathbb{R}^2 \mapsto \mathbb{R}^3$ (Section 4.2). Finally, we show that any consistent pair of measurement and shape $\mathbf{I}, \mathbf{f}$ can be simultaneously rotated in the image plane without affecting consistency, which allows a lossless compression of the input space $\mathcal{I}$.

As we will see later in Section 5, these three properties enable an efficient point processor in the form of a neural network approximation of the mapping from 2-jets $\mathbf{I}$ to functions $\phi_{\mathbf{I}}$ (see Figure 3). Examples of how this representation can be used for shape from shading are described in Section 6.

4.1. An automorphism group on $F(\mathbf{I})$

The first property follows from the fact that each variety $F(\mathbf{I})$ exhibits two symmetries. These symmetries are useful because they allow each variety $F(\mathbf{I})$ to be partitioned into four isomorphic components and therefore represented more compactly by just a single component. This partition applies everywhere except on what is generically a single pair of points of $F(\mathbf{I})$. Thus while we must technically define this partition over a "punctured" variety (what we will call F_0 below), in practice we can typically ignore this distinction, and may in what follows drop the subscript. The symmetries follow from those described for extended quadratic patches in [17] and can be verified by substitution into (3)-(5).

Observation 1. *There exists a subset $F_+(\mathbf{I}) \subseteq F(\mathbf{I})$ whose orbit under the automorphism group generated by*

$$\begin{aligned} &\rho_1 : (f_x, f_y, f_{xx}, f_{xy}, f_{yy}) \mapsto -(f_x, f_y, f_{xx}, f_{xy}, f_{yy}) \\ &\rho_2 : (f_x, f_y, f_{xx}, f_{xy}, f_{yy}) \mapsto \\ &\frac{1}{\sqrt{4f_{xy}^2 + (f_{xx}-f_{yy})^2}} \begin{pmatrix} f_x f_{xx} - f_x f_{yy} + 2 f_y f_{xy} \\ 2 f_x f_{xy} + f_y f_{yy} - f_y f_{xx} \\ f_{xx}^2 - f_{xx} f_{yy} + 2 f_{xy}^2 \\ f_{xx} f_{xy} + f_{xy} f_{yy} \\ f_{yy}^2 - f_{xx} f_{yy} + 2 f_{xy}^2 \end{pmatrix} \end{aligned} \tag{6}$$

[1] When $\mathbf{I}$ and f_x, f_y are fixed, the solutions of (3)-(5) can be interpreted as the intersection of three quadric hypersurfaces. Algebraic solvers for finding the intersection of three quadric *surfaces* have been proposed [5, 13], but these are computationally intractible for the equations studied here.

is precisely $F_0(\mathbf{I})$*, where*

$$F_0(\mathbf{I}) := F(\mathbf{I}) \backslash \boldsymbol{V}(4f_{xy}^2 + (f_{xx} - f_{yy})^2).$$

Thus for fixed $\mathbf{I}$ and f_x, f_y, there will be zero, two, or four nonzero real solutions to (3)-(5), each of which corresponds to a local shape that is some combination of concave/convex and saddle/spherical. Figure 1 shows an example where the variety's four components are clearly visible, and where the four highlighted surfaces comprise one orbit.

We can choose any of the variety's components to be the representative one. The component that corresponds to shapes with positive curvatures is convenient to characterize, so we choose that one and call it the *positive shape set*.

Definition 4. *We call the semi-algebraic set*

$$F_+ := \{\mathbf{f} \in F_0 : f_{xx} + f_{yy} > 0 \text{ and } f_{xx} f_{yy} - f_{xy}^2 > 0\} \quad (7)$$

the positive shape set*. This subset* $F_+(\mathbf{I})$ *is the set* $F_0(\mathbf{I})$ *modulo the group action of* $\langle \rho_1, \rho_2 \rangle$*.*

It is easily verified for non-planar images that $\mathbf{0} \notin F_+$, that there exist no real fixed points of ρ_2, and that by Definition 1 $4f_{xy}^2 + (f_{xx} - f_{yy})^2 \neq 0$ on $F_+(\mathbf{I})$. Therefore the maps ρ_1, ρ_2 are well-defined on $F_+(\mathbf{I})$.

4.2. $F_+(\mathbf{I})$ is a graph

Our aim is to find a parsimonious representation of the positive shape subsets $F_+(\mathbf{I})$ (and thus of the entire shape set $F(\mathbf{I})$) as well as an efficient way to compute this representation for any particular measurement $\mathbf{I}$. Since $F(\mathbf{I})$ and its subset $F_+(\mathbf{I})$ are determined by $\mathbf{I}$, we may define a map $\Phi : \mathbf{I} \mapsto F_+(\mathbf{I})$.

In order to simplify the map Φ from vectors $\mathbf{I}$ to positive subsets $F_+(\mathbf{I})$, we assume that each (two-dimensional) positive subset can be parametrized by surface orientation (f_x, f_y), so that the map $\Phi(\mathbf{I}) = \{(f_x, f_y, f_{xx}, f_{xy}, f_{yy})\}$ can be decomposed as

$$\Phi(\mathbf{I}) = \{(f_x, f_y, \phi_{\mathbf{I}}(f_x, f_y))\}, \quad (8)$$

with $\phi_{\mathbf{I}} : \mathbb{R}^2 \mapsto \mathbb{R}^3$ a continuous function.

While we frame it here as an assumption, this decomposition may in fact be exact. The Implicit Function Theorem guarantees existence (and uniqueness) of a function $\phi(f_x, f_y) = (f_{xx}, f_{xy}, f_{yy})$ in a local neighborhood of every $\mathbf{f}$ for which the Jacobian of system (C_1, C_2, C_3) is nonsingular. While proving that the Jacobian is *always* nonsingular—that is, non-singular for any $\mathbf{I} \in \mathcal{I}$ and any real (f_x, f_y)—remains an open problem, we conjecture that it is true. Experimentally we have never witnessed a non-singular Jacobian, and we can prove non-singularity in simplified cases like the following.

Example 1. *Consider the case in which the measurements* $\mathbf{I}$ *satisfy* $I_x = I_y = 0$*, i.e. in which the image's normal is parallel to the viewing direction. In this case the determinant of the Jacobian of system* (C_1, C_2, C_3) *is*

$$\det J = \gamma((1 + f_y^2) f_{xx} - 2 f_x f_y f_{xy} + (1 + f_x^2) f_{yy})$$

where $\gamma = -4(f_{xx} f_{yy} - f_{xy}^2)/(1 + f_x^2 + f_y^2)^5$*. This has a real solution only if its discriminant taken with respect to* f_x*,*

$$\operatorname{discr}_{f_x} \det J = \gamma((1 + f_y^2)(f_{xx} f_{yy} - f_{xy}^2) + (f_{xy}^2 + f_{yy}^2)),$$

is strictly positive. On $F_+(\mathbf{I})$*, the term* $f_{xx} f_{yy} - f_{xy}^2 > 0$*, so* $\operatorname{discr}_{f_x} \det J < 0$ *over* $\mathbb{R}$*. This implies that there are no points in* $F_+(\mathbf{I})$ *where the implicit function fails.*

4.3. Isomorphisms induced by rotations

A third property is a rotational symmetry about the local viewing direction that allows us to losslessly compress the input space $\mathcal{I}$. Any local relation that exists between an image $I(x, y)$ and surface $f(x, y)$ must persist for any orthogonal change of basis of their common two-dimensional domain (x, y). We are therefore free to define a local coordinate system that adapts to each measurement $\mathbf{I}$.

One choice is the local coordinate system that aligns with the image gradient (I_x, I_y), using an orthogonal transform that maps I_y to zero and I_x to a non-negative real number. This implies three transformations,

$$T_{\mathbf{I}} := \frac{1}{\sqrt{I_x^2 + I_y^2}} \begin{bmatrix} I_x & I_y \\ -I_y & I_x \end{bmatrix} =: \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}, \quad (9)$$

$$S_{\mathbf{I}} := \begin{bmatrix} G_{11}^2 & 2G_{11}G_{21} & G_{21}^2 \\ G_{11}G_{12} & G_{11}G_{22} + G_{12}G_{21} & G_{21}G_{22} \\ G_{12}^2 & 2G_{12}G_{22} & G_{22}^2 \end{bmatrix}, \quad (10)$$

$$R_{\mathbf{I}} := \begin{bmatrix} 1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & T_{\mathbf{I}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & S_{\mathbf{I}}^{-1} \end{bmatrix}, \quad (11)$$

for which one can verify that $\mathbf{f} \in F(\mathbf{I})$ if and only if $\hat{R}_{\mathbf{I}} \mathbf{f} \in F(R_{\mathbf{I}} \mathbf{I})$, with $\hat{R}_{\mathbf{I}}$ the principal submatrix of $R_{\mathbf{I}}$ obtained by removing its first row and column.

By using these transformations to pre-process each of our point processor's inputs $\mathbf{I}$, and to correspondingly post-process each output shape $\mathbf{f}$, we reduce the effective input space from $\mathcal{I} \subset \mathbb{R}^6$ to $\tilde{\mathcal{I}} \subset \mathbb{R}^4 \times \mathbb{R}_+$. In the supplement, we show that this transformation always maps I_{yy} to a nonpositive value, so $\tilde{\mathcal{I}}$ is actually contained in $\mathbb{R}^3 \times \mathbb{R}_+ \times \mathbb{R}_-$. The size of $\tilde{\mathcal{I}}$ can be reduced further by exploiting the linearity of (3)-(5) in $\mathbf{I}$, which implies that F_+ is invariant under any positive real scaling of $\mathbf{I}$. This means we can additionally restrict $\tilde{\mathcal{I}}$ to the unit sphere S^4 without loss of generality.

Combined, this reduces the effective input space to $\tilde{\mathcal{I}} \subset \mathbb{R}^3 \times \mathbb{R}_+ \times \mathbb{R}_- \cap S^4$. We reap the benefits of this domain simplification when designing and training our neural network.

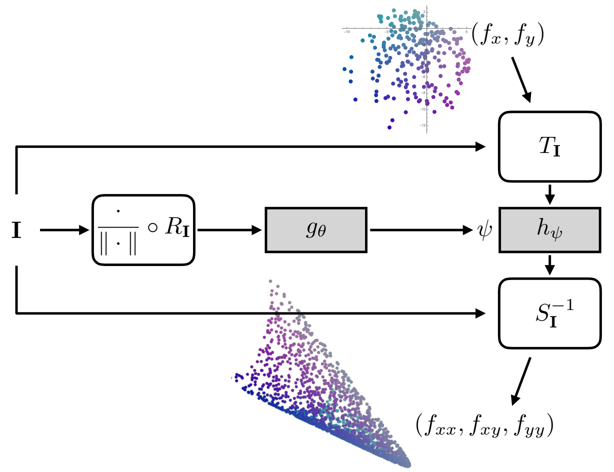

Figure 3. The structure of our two-stage network approximator $\hat{\phi}_{\mathbf{I}}$ for the map from vectors $\mathbf{I}$ to functions $\phi_{\mathbf{I}}$. The right shows orientation domain and output samples for the same $\mathbf{I}$ as in Figure 1.

5. A Neural Network Approximator

Let us ignore for now the pre- and post-processing transformations related to rotations, and consider the task of approximating the mapping from vectors $\mathbf{I} \in \mathcal{I}$ to functions $\phi_{\mathbf{I}}$. One convenient way to do this is to couple a pair of neural networks, with the output of one network providing the weights of the other. That is, we can use

$$\hat{\phi}_{\mathbf{I}}(f_x, f_y) := h\left(f_x, f_y; g_\theta(\mathbf{I})\right), \tag{12}$$

where $g_\theta : \mathbb{R}^6 \mapsto \mathbb{R}^M$ is a (fully-connected, few-layer) neural network with tunable weights $\theta \in \mathbb{R}^N$ and $h_\psi : \mathbb{R}^2 \mapsto \mathbb{R}^3$ is a (fully-connected, single layer) neural network whose weights $\psi \in \mathbb{R}^M$ are provided by the output of g. This means that under the hood, $\hat{\phi}_{\mathbf{I}}$ is a function of θ.

This is convenient because it provides a compact representation that can be efficiently fit to a large set of training samples. We can fit the weights θ by synthetically generating many measurements $\mathbf{I}$ and for each one computing many samples $\mathbf{f}$ from the corresponding semi-algebraic set $F_+(\mathbf{I})$ using Theorem 1 and Observation 1. This produces a set of samples $\{(\mathbf{I}^{(j)}, \mathbf{f}^{(i,j)})\}_{i,j}$ that we can use to solve

$$\theta = \underset{\theta}{\operatorname{argmin}} \sum_j \sum_i \left\| \left(f_{xx}^{(i,j)}, f_{xy}^{(i,j)}, f_{yy}^{(i,j)}\right) - h\left(f_x^{(i,j)}, f_y^{(i,j)}; g_\theta\left(\mathbf{I}^{(j)}\right)\right) \right\|^2 \tag{13}$$

via stochastic gradient descent.

Now, with only small modifications, we can incorporate the rotational transformations of Section 4.3 to make the approximator more efficient and reduce the training burden. This simply requires surrounding the neural network with linear transformation blocks (see Figure 3) that pre-process an input measurement $\mathbf{I}$, and that correspondingly pre-process the orientation domain (f_x, f_y) and post-process the output curvatures (f_{xx}, f_{xy}, f_{yy}) using (9-11). This reduces the domain of network g_θ from $\mathbb{R}^6$ to $\mathbb{R}^3 \times \mathbb{R}_+ \times \mathbb{R}_- \cap S^4$. For example, if the input to block $\frac{\cdot}{\|\cdot\|} \circ R_{\mathbf{I}}$ in Figure 3 is $\mathbf{I} = (I, I_x, I_y, I_{xx}, I_{xy}, I_{yy})$ then its output is

$$(\tilde{I}, \tilde{I}_x, \tilde{I}_{xx}, \tilde{I}_{xy}, \tilde{I}_{yy}) / \|(\tilde{I}, \tilde{I}_x, \tilde{I}_{xx}, \tilde{I}_{xy}, \tilde{I}_{yy})\| \tag{14}$$

with $\tilde{\mathbf{I}} = R_{\mathbf{I}}\mathbf{I}$ (and dropping the now-redundant $\tilde{I}_y = 0$).

5.1. Training Data and Network Architecture

Training requires samples of 2-jets $\mathbf{I}^{(j)} \in \mathcal{I}$ as well as samples of the positive set $F_+(\mathbf{I}^{(j)})$ for each 2-jet. We generate the former by sampling light source directions $\mathbf{L}$ and quadratic patches $\mathbf{f}$ and then applying Eqs. (1) and (2) (and their spatial derivatives) to render 2-jets $\mathbf{I}^{(j)}$. Specifically, we sample light sources uniformly from the subset of S^2 contained in an angular radius of $\pi/4$ from the view direction $(0, 0, 1)$, and we sample surface orientations f_x, f_y uniformly from the unit disk B^2. By Observation 1 it is sufficient to sample positive curvatures, so we sample f_{xx}, f_{xy}, f_{yy} uniformly from a bounded subset of $\mathbb{R}^3 \cap \{f_{xx}f_{yy} - (f_{xy})^2 > 0\} \cap \{f_{xx} + f_{yy} > 0\}$.

To create samples of the positive set $F_+(\mathbf{I}^{(j)})$ for each 2-jet, we first generate a dense set of sample orientations $\{(f_x^{(i)}, f_y^{(i)})\}$ from the unit disk to serve as input to network h_ϕ. Then, for each $\mathbf{I}^{(j)}$ and for each $f_x^{(i)}, f_y^{(i)}$ the corresponding "ground truth" second order shape values $(f_{xy}^{(i,j)}, f_{xy}^{(i,j)}, f_{yy}^{(i,j)})$ are computed by applying a numerical root-finder to (3)-(5). The result is a training set $\{(\mathbf{I}^{(j)}, \mathbf{f}^{(i,j)})\}_{i,j}$.

Numerical root finding can be expensive, but the simplification of the domain of g_θ (see the previous section) in our case reduces the computational burden. Rather than generating enough 2-jets $\mathbf{I}^{(j)}$ to sufficiently sample $\mathcal{I}$, we need only generate enough for the measurements that are pre-processed by block $\frac{\cdot}{\|\cdot\|} \circ R_{\mathbf{I}}$ to sufficiently sample $\tilde{\mathcal{I}}$.

For network $g_\theta : \mathbb{R}^5 \mapsto \mathbb{R}^M$ we use $d_g = 1$ hidden layer with $w_g = 25$ ReLU nodes. For network $h_\psi : \mathbb{R}^2 \mapsto \mathbb{R}^3$ we use one hidden layer with $w_h = 50$ ReLU nodes. The total number of tunable parameters is $N = 6w_g + (d_g - 1)w_g(w_g + 1) + M(w_g + 1) + M$, and once the model is trained, the output description of the shape-set $F(\mathbf{I})$ for any 2-jet $\mathbf{I}$ consists of $M = 3(2w_h + 1)$ rational numbers (the size of vector ψ). The entire shape set at each image point is therefore summarized by only $M = 303$ numbers. Figure 4 visualizes the quality of fit for a representative test measurement $\mathbf{I}$ that was not used during training.

6. Applications

The point processor transforms image values at a single point $\mathbf{I}$ into an intermediate representation of the consistent

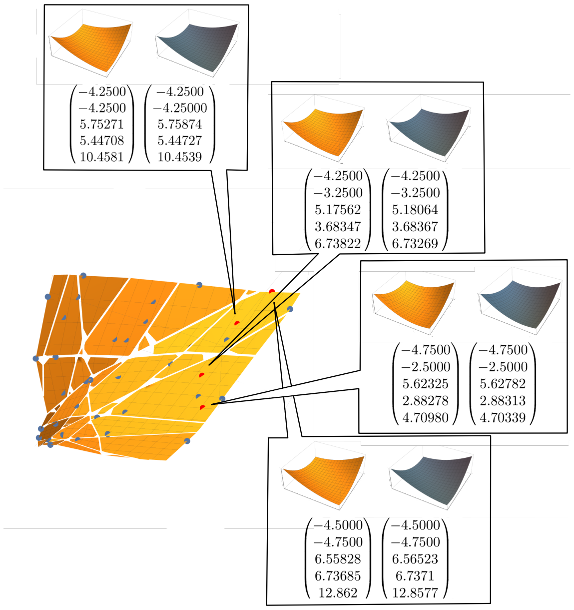

Figure 4. Visualization of the approximator's interpolation error. This figure depicts $\hat{\phi}_{\mathbf{I}}$ for an $\mathbf{I}$ that was randomly chosen from the convex hull of the training data set, but that was not used as a training sample. The inset shows the four randomly-chosen solutions for which our approximation performs worst, *i.e.* those $\mathbf{f}$ that maximize the error $||\mathbf{f} - (f_x, f_y, \hat{\phi}_{\mathbf{I}}(f_x, f_y))||_2^2$.

local shape-set, in the form of a two-dimensional manifold parametrized by surface orientation, $\left(f_x, f_y, \hat{\phi}_{\mathbf{I}}(f_x, f_y)\right)$. To demonstrate how this continuous representation of per-point shapes can be used for image analysis, we consider two simple scenarios. In both cases, the per-point ambiguity is resolved (up to a discrete four-way choice of shapes) by exploiting additional information or assumptions.

Our demonstrations use simple images rendered according to (1) with 1% additive Gaussian noise and 64-bit quantization. We estimate spatial image derivatives using Gaussian derivative filters.

6.1. Uncalibrated two-shot photometric stereo

The per-point ambiguity can be resolved by capturing additional images of the same surface under distinct light directions. When the light directions are unknown this is called uncalibrated photometric stereo [18, 7, 3]. In the traditional formulation, which is based purely on surface orientation (f_x, f_y), it requires at least three images under three distinct lights [9]. Our point processor based on second-order shape provides a similar capability with only two input images instead of three.

Consider two measurements $\mathbf{I}_1, \mathbf{I}_2$ generated at the same point from two (unknown) light sources $\mathbf{L}_1, \mathbf{L}_2$. A simulated example is depicted in the top of Figure 5. The first measurement $\mathbf{I}_1$ limits the shape to being in the set $F_+(\mathbf{I}_1)$, but within this set all shapes are equally likely. Since the set is parametrized by surface orientation $\left(f_x, f_y, \hat{\phi}_{\mathbf{I}_1}(f_x, f_y)\right)$, we can visualize the (uniform) "likelihood" over some reasonably-sized disk of the orientation domain (f_x, f_y). This is shown in the left of Figure 5, with the magenta dot indicating the orientation of the latent true shape $\mathbf{f}^*$ that was used for the simulation.

The second measurement $\mathbf{I}_2$ further restricts the shape to being in the intersection of sets $F_+(\mathbf{I}_1)$ and $F_+(\mathbf{I}_2)$. Thus, we can improve the "likelihood" based on how close each shape is to $F_+(\mathbf{I}_1) \cap F_+(\mathbf{I}_2)$. One way to quantify this is

$$L(f_x, f_y) := \left\| \hat{\phi}_{\mathbf{I}_2}(f_x, f_y) - \hat{\phi}_{\mathbf{I}_1}(f_x, f_y) \right\|^2 \tag{15}$$

for (f_x, f_y) in the disk. For our simulation, this updated two-measurement likelihood is shown on the right in Figure 5, where it provides a successful identification of the true shape.

Recovering the correct per-point shape (up to the four-way choice) by this simple strategy relies on the intersection $F_+(\mathbf{I}_1) \cap F_+(\mathbf{I}_2)$ being a single point, as seems to be the case for our simulation, as shown in the bottom of Figure 5. Our experiments suggest this is typically the case, but analytically characterizing the conditions for uniqueness may be a worthwhile direction for future work. Also, resolving the four-way choice at each point would require making additional surface continuity assumptions, analogous to how "integrability" is used to reduce the inherent global linear ambiguity in traditional three-shot photometric stereo [18].

6.2. Surface continuity

An alternative way to reduce the per-point ambiguity is to design a 2D array of point processors that are connected together by enforcing surface continuity across an extended region of the input image. As a simple example, we consider the scenario in which the entire surface is an extended quadratic function, meaning one that satisfies (2) over the entire image $I(x, y)$ with some "true shape" values $\mathbf{f}^* = (f_x^*, f_y^*, f_{xx}^*, f_{xy}^*, f_{yy}^*)$.

When the surface is known to be an extended quadratic, any single local shape $\mathbf{f} \in F_+(\mathbf{I}_1)$ at one point, say the image origin, immediately predicts a corresponding local shape $\mathbf{f}'$ at every other point (x, y) in the image, via $(f'_{xx}, f'_{xy}, f'_{yy}) = (f_{xx}, f_{xy}, f_{yy})$ and $(f'_x, f'_y) = (f_x, f_y) + A(x, y) \cdot (f_{xx}, f_{xy}, f_{yy})$ with matrix

$$A(x, y) = \begin{bmatrix} x & y & 0 \\ 0 & x & y \end{bmatrix}. \tag{16}$$

As before, we begin with a uniform relative likelihood over the shape set $F_+(\mathbf{I}_1)$ obtained by a single measurement at the origin in an input image of an extended quadratic surface (left of Figure 6). Then given a measurement $\mathbf{I}_2$ at one other point (x_2, y_2), we use that information to update the likelihood

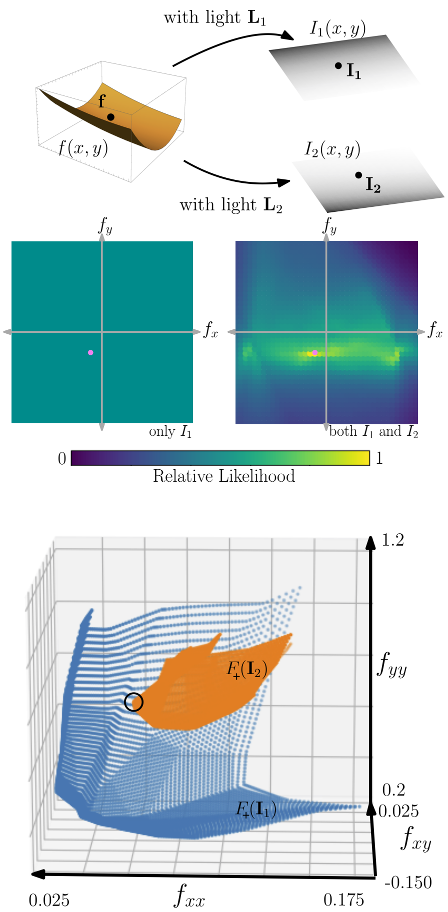

Figure 5. Uncalibrated two-shot photometric stereo. Top row: two simulated images of a surface under different lights, with measurements $\mathbf{I}_1, \mathbf{I}_2$ at the same pixel location. Center row: "Likelihood" of different shapes using only one measurement (left) or both measurements (right), visualized over the orientation domain. Magenta dot indicates true shape used for simulation. Bottom row: Shape sets $F_+(\mathbf{I}_1)$, $F_+(\mathbf{I}_2)$ and their intersection (open circle).

over the first set using (15), but with the term $\hat{\phi}_{\mathbf{I}_2}(f_x, f_y)$ replaced by $\left(\hat{\phi}_{\mathbf{I}_2}(f_x, f_y) + A(x_2, y_2) \cdot \hat{\phi}_{\mathbf{I}_1}(f_x, f_y)\right)$. The updated two-measurement likelihood is shown in the second column of Figure 6.

We continue this process by adding information from additional measurements, $\mathbf{I}_3$ at (x_3, y_3) and $\mathbf{I}_4$ at (x_4, y_4), each time updating the likelihood over the original set $L(f_x, f_y)$ by accumulating the intersection errors between $F_+(\mathbf{I}_i)$ and $F_+(\mathbf{I}_1)$. The evolution of this likelihood for three and four points is shown in Figure 6. We see that the composite likelihood function achieves its global maximum at a shape $\mathbf{f} \in F_+(\mathbf{I})$ that is very close to $\mathbf{f}^*$ modulo the irreconcilable four-way ambiguity. This is consistent with the area-based analysis of Xiong *et al.* [17], which proves the uniqueness of shape reconstruction for extended quadratric patches.

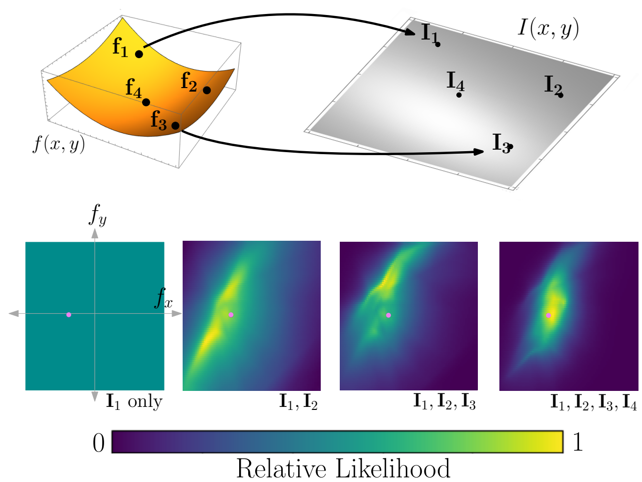

Figure 6. Combining shape information at multiple points on an extended quadratic surface. Given one measurement $\mathbf{I}_1$, all quadratic shapes in $F_+(\mathbf{I}_1)$ are equally likely. This is depicted on the left as a constant relative likelihood over the domain of function $\hat{\phi}_{\mathbf{I}}$. Incorporating measurements $\mathbf{I}_i$ at two or more points modifies the likelihood to have a maximum that is close to the true shape (magenta dot) modulo ρ_1, ρ_2.

7. Conclusion

This paper takes preliminary steps toward a deployable point processor for shading that does not require knowledge of lighting at a point or rely on accurate estimates of that lighting. It suggests a new intermediate representation of the set of consistent second-order shapes at each image point, in the form of an explicit differentiable, parametrized two-dimensional manifold. It also provides two simple examples of how this new intermediate representation can be used for shape analysis. The distinguishing feature of this approach is that it has the potential to enable shape processing to succeed in real-world situations where the lighting varies across surfaces and is therefore difficult or impossible to accurately infer.

The contributions of this paper are primarily theoretical, and turning this research into practice will require substantial progress in several directions. This may include combining multi-scale derivatives, creating spatial regularization schemes that are suitable for piecewise smooth surfaces, extending the approach from local second-order shape to local third-order shape, and exploring the ability of the factored network architecture to represent more general (*e.g.* non-Lambertian) rendering models and to be trained from images instead of algebraic equations.

References

[1] Jonathan T. Barron and Jitendra Malik. Shape, illumination, and reflectance from shading. *IEEE Transactions on Pat-*

tern Analysis and Machine Intelligence (TPAMI), 37(8):1670–1687, 2015. 2

[2] Ronen Basri and David W Jacobs. Lambertian reflectance and linear subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, (2):218–233, 2003. 2

[3] Peter N Belhumeur, David J Kriegman, and Alan L Yuille. The bas-relief ambiguity. *International Journal of Computer Vision (IJCV)*, 35(1):33–44, 1999. 1, 7

[4] Patrick Cavanagh. The artist as neuroscientist. *Nature*, 434:301–307, 2005. 2

[5] Eng-Wee Chionh, Ronald N Goldman, and James R Miller. Using multivariate resultants to find the intersection of three quadric surfaces. *ACM Transactions on Graphics (TOG)*, 10(4):378–400, 1991. 4

[6] David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 2366–2374, 2014. 2

[7] Joel Fan and Lawrence B Wolff. Surface curvature and shape reconstruction from unknown multiple illumination and integrability. *Computer Vision and Image Understanding*, 65(2):347–359, 1997. 7

[8] David A Forsyth. Variable-source shading analysis. *International Journal of Computer Vision (IJCV)*, 91(3):280–302, 2011. 2

[9] Hideki Hayakawa. Photometric stereo under a light source with arbitrary motion. *Journal of the Optical Society of America (JOSA) A*, 11(11):3079–3089, 1994. 7

[10] Berthold KP Horn. Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. Technical Report AITR-232, MIT Artificial Intelligence Laboratory, 1970. 1

[11] Jan J Koenderink and Andrea J van Doorn. Representation of local geometry in the visual system. *Biological cybernetics*, 55(6):367–375, 1987. 2

[12] Jan J Koenderink, Andrea J Van Doorn, Astrid ML Kappers, and James T Todd. Ambiguity and the 'mental eye' in pictorial relief. *Perception*, 30(4):431–448, 2001. 2

[13] Zuzana Kukelova, Jan Heller, and Andrew Fitzgibbon. Efficient intersection of three quadrics and applications in computer vision. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1799–1808, June 2016. 4

[14] Benjamin Kunsberg and Steven W Zucker. How shading constrains surface patches without knowledge of light sources. *SIAM Journal on Imaging Sciences*, 7(2):641–668, 2014. 2, 3, 4

[15] Ravi Ramamoorthi and Pat Hanrahan. An efficient representation for irradiance environment maps. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, pages 497–500. ACM, 2001. 2

[16] Stephan R Richter and Stefan Roth. Discriminative shape from shading in uncalibrated illumination. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1128–1136, 2015. 2

[17] Ying Xiong, Ayan Chakrabarti, Ronen Basri, Steven J Gortler, David W Jacobs, and Todd Zickler. From shading to local shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 37(1):67–79, 2015. 2, 4, 8

[18] A Yuille and D Snow. Shape and albedo from multiple images using integrability. In *Computer Vision and Pattern Recognition (CVPR)*, pages 158–164, 1997. 7

[19] Daniel Zoran, Dilip Krishnan, Jose Bento, and Bill Freeman. Shape and illumination from shading using the generic viewpoint assumption. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 226–234, 2014. 2