

A Lighting-Invariant Point Processor for Shading

Kathryn Heal

Jialiang Wang

Steven Gortler

Todd Zickler

Harvard University

{kathrynaheal@g, jialiangwang@g, sgj@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 variety embedded in the five-dimensional space of quadratic shapes. We describe the geometry of this variety, and we introduce a fast 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 fast parallelizable processor that operates at each image point and produces a lighting-invariant descriptor of the continuous set of compatible surface shapes at a point. We describe two applications of this processor: two-shot uncalibrated photometric stereo and quadratic-surface shape from shading.

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 are 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

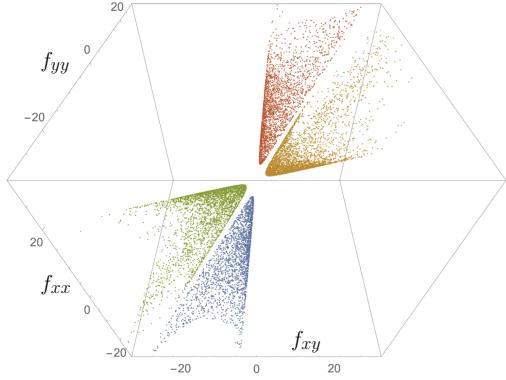


Figure 1. Given a measurement of the image derivatives $(I, I_x, I_y, I_{xx}, I_{xy}, I_{yy})$ at a single point, our processor provides a compact representation for the set of compatible surface derivatives $\{(f_x, f_y, f_{xx}, f_{xy}, f_{yy})\}$ at that point without requiring prior knowledge of lighting. This figure depicts the curvature elements of each compatible surface. Colors indicate four isomorphic components of the set.

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 the surfaces in a scene, 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 because of 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 parameterized 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 $F(\mathbf{I}) := \{(f_x, f_y, f_{xx}, f_{xy}, f_{yy})\}$ is a two-dimensional algebraic variety in \mathbb{R}^5 , and 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 by a coupled pair of 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 represent local surface and image depth maps as vectors of their coefficients, for example 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 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 defined 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 substantial 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 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 group that we describe in Section 3.3.

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 geometry approach instead, which provides an abbreviated derivation of equivalent equations and also reveals the shape-set to be an algebraic variety with useful structure.

Specifically, 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 3.3). We then relate one piece of the shape set 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 3.4). This analysis provides the foundation for a fast neural network approximation of the mapping from vectors \mathbf{I} to functions $\phi_{\mathbf{I}}$, which appears in Section 4.

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 calculate a discrete approximation to the first and second spatial derivatives of $I(x, y)$ at the origin.

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

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

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

We further assume that within neighborhood U the surface has insignificant third-order derivatives, so that

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

We refer to values $\mathbf{f} := (f_x, f_y, f_{xx}, f_{xy}, f_{yy}) \in \mathbb{R}^5$ as the *local shape*, and we assume that all local shapes are nondegenerate according to the following criterion.

Definition 1. A local shape \mathbf{f} is nondegenerate if its mean and gaussian curvatures are both nonzero. This is equivalent to the condition $(f_{xx} + f_{yy})(f_{xx} f_{yy} - f_{xy}^2) \neq 0$.

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*.

Definition 2. The set of realizable measurements \mathcal{I} is the set of vectors $\mathbf{v} \in \mathbb{R}^6$ for which there exists a light direction $\mathbf{L} \in \mathbb{R}^3$ and nondegenerate local shape \mathbf{f} such that $\mathbf{v} = \mathbf{I}(0, 0)$

when shape model (2) is combined with shading model (1). If for a pair $(\mathbf{I}, \mathbf{f}) \in \mathcal{I} \times \mathbb{R}^5$ there exists such an \mathbf{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. Local Shape Set

Our immediate goal is to characterize that 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 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. Suppose we have 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

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

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

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

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 $F := V(C_1, C_2, C_3)$ contains the set of all shapes \mathbf{f} consistent with \mathbf{I} .

Remark 1. The real solutions to these equations are identical to those of Corollary 4.2 of [14].

Proof. For brevity let $a, b, c, d, e = f_x, f_y, f_{xx}, f_{xy}, f_{yy}$. We begin with some derivatives of Lambert's law, and introducing a dummy variable w . Fixing an \mathbf{I} vector, we define polynomials

$$\begin{aligned} s &:= w^2 (a^2 + b^2 + 1) - 1 \\ r_0 &:= w(aL_1 + bL_2 - L_3) + I \\ r_x &:= w^3(-(ac + bd))(aL_1 + bL_2 - L_3) + w(cL_1 + dL_2) + I_x \\ r_y &:= w^3(-(ad + be))(aL_1 + bL_2 - L_3) + w(dL_1 + eL_2) + I_y \\ r_{xx} &:= w^3(3w^2(ac + bd)^2 - c^2 - d^2)(aL_1 + bL_2 - L_3) \end{aligned}$$

$$\begin{aligned}
& -2w^3(ac+bd)(cL_1+dL_2)+I_{xx} \\
r_{xy} &:= 3w^5(ac+bd)(ad+be)(aL_1+bL_2-L_3) \\
& -dw^3(c+e)(aL_1+bL_2-L_3)-w^3(ad+be)(cL_1+dL_2) \\
& -w^3(ac+bd)(dL_1+eL_2)+I_{xy} \\
r_{yy} &:= w^3(3w^2(ad+be)^2-d^2-e^2)(aL_1+bL_2-L_3) \\
& -2w^3(ad+be)(dL_1+eL_2)+I_{yy}
\end{aligned}$$

and we refer to the vectors (r_0, r_x, r_y, r_{xx}) , (r_0, r_x, r_y, r_{xy}) , and (r_0, r_x, r_y, r_{yy}) as \mathbf{r}_1 , \mathbf{r}_2 , and \mathbf{r}_3 , respectively. Each of these seven polynomials is linear in the L_i ; thus, if $\mathbf{L} := (L_1, L_2, L_3, 1)$, we can write this system as $\mathbf{r}_i = A_i \mathbf{L}$, where $A_i = A_i(w, \mathbf{f}, \mathbf{I})$ is a square functional-entered matrix. Then

$$\begin{aligned}
A_i &= \begin{pmatrix} aw & bw & -w & I \\ cw(1-a^2w^2)-abdw^3 & dw(1-b^2w^2)-abcw^3 & w^3(ac+bd) & I_x \\ dw(1-a^2w^2)-abew^3 & ew(1-b^2w^2)-abdw^3 & w^3(ad+be) & I_y \\ \rho_{i1} & \rho_{i2} & \rho_{i3} & \rho_{i4} \end{pmatrix}, \\
\rho_{11} &= w^3(3a^3c^2w^2+6a^2bcdw^2-a(d^2(1-3b^2w^2)+3c^2)-2bcd) \\
\rho_{12} &= w^3(bc^2(3a^2w^2-1)+6ab^2cdw^2-2acd+3b^3d^2w^2-3bd^2) \\
\rho_{13} &= w^3(c^2(1-3a^2w^2)-6abcdw^2+d^2(1-3b^2w^2)) \\
\rho_{21} &= w^3(3a^3d^2w^2+6a^2bdew^2-a(e^2(1-3b^2w^2)+3d^2)-2bde) \\
\rho_{22} &= w^3(bd^2(3a^2w^2-1)+6ab^2dew^2-2ade+3b^3e^2w^2-3be^2) \\
\rho_{23} &= w^3(d^2(1-3a^2w^2)-6abdew^2+e^2(1-3b^2w^2)) \\
\rho_{31} &= w^3(3a^3cdw^2+3a^2bw^2(ce+d^2) \\
& -ad(-3b^2ew^2+3c+e)-b(ce+d^2)) \\
\rho_{32} &= w^3(-bd(-3a^2cw^2+c+3e)+3ab^2w^2(ce+d^2) \\
& -a(ce+d^2)+3b^3dew^2) \\
\rho_{33} &= w^3(c(-3a^2dw^2-3abew^2+d)+d(-3abdw^2-3b^2ew^2+e)) \\
\rho_{14} &= I_{xx} \quad \rho_{24} = I_{yy} \quad \rho_{34} = I_{xy}.
\end{aligned}$$

Notice that for each i , $\det A_i = w^3(d^2-ce)C_i$, and that $s = 0 \implies w \neq 0$. Suppose $\neg(C_i = 0 \forall i)$. That is, $\exists i : C_i \neq 0$. Due to the non-degeneracy assumption and the constraint imposed by s that $w \neq 0$, this is equivalent to $\exists i : w^3(d^2-ce)C_i = \det A_i \neq 0 \iff \exists i : \ker(A_i) = \{\mathbf{0}\}$. This is equivalent to $\exists i : \forall \mathbf{L} \neq \mathbf{0}, A_i \mathbf{L} = \mathbf{r}_i \neq \mathbf{0}$, which is to say that there exists an i which, regardless of \mathbf{L} , will always violate one of the Lambert partial derivative conditions. This implies $\neg((\mathbf{f}, \mathbf{I}) \text{ are consistent})$. The contrapositive of this argument is that under the stated assumptions, (\mathbf{f}, \mathbf{I}) consistent implies that $C_i = 0 \forall i$. \square

This algebraic derivation of the consistent shape-set $F(\mathbf{I})$ is an alternative to the differential-geometric approach presented in [14]. It shows that the set of local shapes that are

consistent with a given measurement \mathbf{I} are contained in a two-dimensional variety embedded in the five-dimensional shape space. It 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.

As elements of $\mathbb{R}[\mathbf{f}]$, polynomials C_1, C_2, C_3 are determined by parameters $\mathbf{I} \in \mathcal{I}$, so the variety $F(\mathbf{I})$ is as well. We may therefore define a map $\Phi : \mathcal{I} \mapsto F(\mathbf{I})$. This leads us to the interpretation of the system as a smooth fiber bundle, with base space \mathcal{I} , fibers $F(\mathbf{I})$, and section Φ .

3.3. Properties of the Local Shape Set

For every \mathbf{I} , the variety $F(\mathbf{I})$ exhibits two symmetries that are useful because they allow each variety $F(\mathbf{I})$ to be (1) partitioned into four isomorphic pieces and (2) represented more compactly by any one piece. The symmetries described below follow from those described for extended quadratic patches in [17], and they can be verified by way of substitution into (3).

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

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

is precisely $F(\mathbf{I})$.

Definition 3. We call the semi-algebraic set

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

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

It is easily verified that for non-planar images, $\mathbf{0} \notin F_+$, that there exist no real fixed points of ρ_2 , and that $4f_{xy}^2 + (f_{xx} - f_{yy})^2 = (f_{xx} + f_{yy})^2 + 4(f_{xy}^2 - f_{xx} f_{yy}) > 0$ on $F_+(\mathbf{I})$. Therefore, maps ρ_1, ρ_2 are well-defined on $F_+(\mathbf{I})$. Observation 1 tells us that for fixed \mathbf{I} and f_x, f_y , there will be zero, two, or four real solutions to (3), each of which corresponds to a surface that is some combination of concave/convex and saddle/spherical. Figure 2 shows an example.

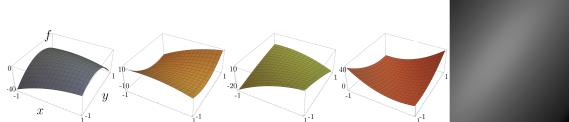


Figure 2. Four quadratic surfaces with a common orientation $(f_x, f_y) \approx (5.61, -4.03)$ at the origin. When each surface is lit from a particular direction it produces the image $I(x, y)$ shown right. The four surfaces make up a single orbit of the automorphism group in Observation 1.

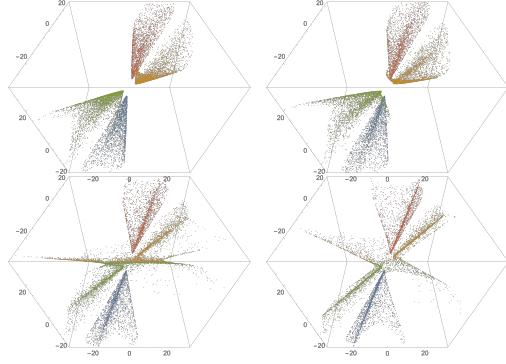


Figure 3. Samples from $\Phi(\mathbf{I})$ for $\mathbf{I} \approx (1-t, -4.10, -5.87, -12.41, -13.41, -20.30) + t$, with $t = 2, 2.5, 3, 3.5$ shown. If \mathcal{I} is the base space of a variety-bundle, we can think of Φ as being a section of that bundle. Thus, this figure depicts fibers of a curve $\mathbf{I}(t) \subset \mathcal{I}$.

3.4. Explicit Representation

When \mathbf{I} and f_x, f_y are fixed, (3) 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.

Once $\mathbf{I} \in \mathcal{I}$ is measured at a particular image point, the polynomials C_1, C_2, C_3 can be understood as elements of the polynomial ring $\mathbb{R}[\mathbf{f}]$ that are determined by parameters $\mathbf{I} \in \mathcal{I}$. It follows that the variety $F(\mathbf{I})$ and its subset $F_+(\mathbf{I})$ are also determined by \mathbf{I} . We may therefore define a map $\Phi : \mathbf{I} \mapsto F_+(\mathbf{I})$ and interpret the system as a smooth fiber bundle, with base space \mathcal{I} , fibers $F_+(\mathbf{I})$, and section Φ .

Our aim is to find a parsimonious representation of the shape subsets $F_+(\mathbf{I})$ (the fibers) and an efficient representation of the map Φ from vectors \mathbf{I} to subsets $F_+(\mathbf{I})$. We have already observed that the subsets are two-dimensional. We further assume that the two-dimensional subsets can be parameterized 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. When \mathbf{I} and (f_x, f_y) are fixed, each of the polynomials C_1, C_2, C_3 is quadric in (f_{xx}, f_{xy}, f_{yy}) ,

so the function $\phi_{\mathbf{I}}$ can be interpreted as an explicit representation of the intersection of three quadrics defined by (f_x, f_y) . 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* non-singular—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 this 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 $I_x = I_y = 0$. In this case the determinant of the Jacobian of system (C_1, C_2, C_3) is

$$\det \text{Jac}C = \gamma((1 + f_y^2)f_{xx} - 2f_x f_y f_{xy} + (1 + f_x^2)f_{yy}) \quad (9)$$

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 ,

$$\text{discr}_{f_x} \det \text{Jac}C = \gamma((1 + f_y^2)(f_{xx}f_{yy} - f_{xy}^2) + (f_{xy}^2 + f_{yy}^2)), \quad (10)$$

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

Next we pursue a parsimonious analytical representation of $\phi_{\mathbf{I}}$, so that we can quickly obtain a representation of the entire shape set $F_+(\mathbf{I})$ for any \mathbf{I} . The next section shows that a simple factored neural network suffices for this.

4. A Neural Network Approximator

One convenient way to approximate the mapping from vectors \mathbf{I} to functions $\phi_{\mathbf{I}}$ is by coupling a pair of neural networks, with the output of one network providing the weights of the other. That is, we can use

$$\hat{\phi}_{\mathbf{I}}(f_x, f_y) := h(f_x, f_y; g(\mathbf{I}; \theta)), \quad (11)$$

where $g_\theta : \mathbb{R}^6 \mapsto \mathbb{R}^M$ is a (fully-connected) neural network with tunable weights $\theta \in \mathbb{R}^N$ and $h_\psi : \mathbb{R}^2 \mapsto \mathbb{R}^3$ is a (fully-connected) 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 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 use to solve

$$\theta = \arg \min_{\theta} \sum_j \sum_i \left\| \begin{pmatrix} f_{xx}^{(i,j)} & f_{xy}^{(i,j)} & f_{yy}^{(i,j)} \end{pmatrix} \right\|$$

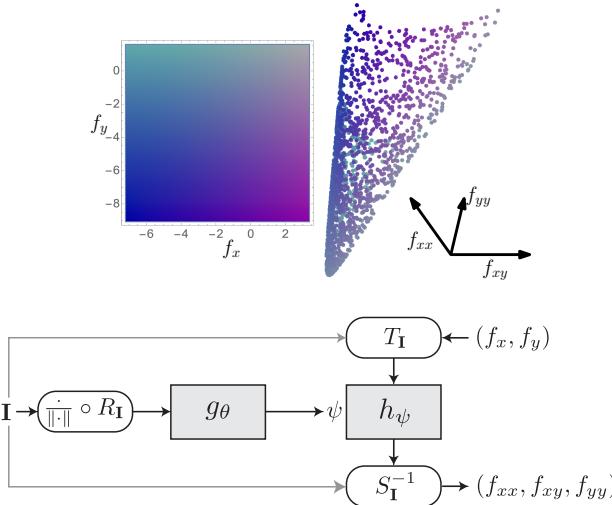


Figure 4. Top: For the same \mathbf{I} as in Figure 1, the correspondence between f_x, f_y and f_{xx}, f_{xy}, f_{yy} on $F_+(\mathbf{I})$. Bottom: The structure of our two-phase network approximator. The shaded blocks represent fully-connected feedforward neural networks.

$$-h\left(f_x^{(i,j)}, f_y^{(i,j)}; g(\mathbf{I}^{(j)}; \theta)\right) \parallel^2 \quad (12)$$

via stochastic gradient descent.

4.1. Rotational symmetry

The approximator can be made even more compact by exploiting rotational symmetry about the viewing direction that is associated with the camera's orthographic projection. Any relation that exists between 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 orientation by mapping I_y to zero and I_x to the positive reals. In terms of implementation, this simply requires surrounding the neural network with linear transformation blocks (see Figure 4) that pre-transform the measurement \mathbf{I} and orientations (f_x, f_y) and then post-transform the resulting curvatures (f_x, f_y, f_z) accordingly using the following transformations:

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

and

$$R_{\mathbf{I}} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & T_{\mathbf{I}} & 0 \\ 0 & 0 & S_{\mathbf{I}}^{-1} \end{bmatrix}.$$

One can verify that $\mathbf{f} \in F(\mathbf{I})$ if and only if $R_{\mathbf{I}}\mathbf{f} \in F(R_{\mathbf{I}}\mathbf{I})$. This relationship allows us to reduce the size of the domain of g_{θ} from \mathbb{R}^6 to $\mathbb{R}^4 \times \mathbb{R}_+$, thus greatly decreasing complexity.

In the supplement, we show that this transformation always maps I_{yy} to a negative value, so the domain of g_{θ} is actually contained in $\mathbb{R}^3 \times \mathbb{R}_+ \times \mathbb{R}_-$.

To reduce complexity a little further without loss of generality, we restrict the domain of g_{θ} to lie in the unit sphere S^4 . The linearity of (3) in \mathbf{I} implies that F_+ is invariant under any positive real scaling of \mathbf{I} . This allows for a vector normalization of $\tilde{\mathbf{I}}$ as a second preprocessing step, and reduces the domain of g_{θ} to one contained in $\mathbb{R}^3 \times \mathbb{R}_+ \times \mathbb{R}_- \cap S^4$. We reap the benefits of these domain simplifications when generating training data for the network.

4.2. Training Data

Synthetic training samples are generated by selecting a “ground truth” light source direction \mathbf{L}^* and quadratic patch \mathbf{f}^* , and applying Eq. (1) to obtain an artificially-rendered image measurement \mathbf{I}^* . The sampling procedure is deterministic and proceeds as follows. Synthetic surface curvatures $f_{xx}^*, f_{xy}^*, f_{yy}^*$ are uniformly sampled from a bounded subset of $\mathbb{R}^3 \cap \{f_{xx}^*f_{yy}^* - (f_{xy}^*)^2 > 0\}$. Synthetic surface orientations f_x^*, f_y^* are chosen approximately uniformly from the unit disk B^2 . Synthetic light source directions are chosen from within a radius of $\pi/4$ from $(0, 0, 1)$ on S^2 .

For every synthetic \mathbf{I}^* vector generated in this way, another more densely sampled set of candidate normal vectors $\{f_x^{**}, f_y^{**}\}$ are selected from the unit disk; this serves as the parameter domain for the training $F(\mathbf{I}^*)$ sets. The corresponding $\{f_{xx}^{**}, f_{xy}^{**}, f_{yy}^{**}\}$ by applying a numerical root-finders to (3). Of course, such numerical solving is a very expensive procedure, and circumventing this solving was part of our original motivation to create a feedforward approximator. Our strategic simplification of the domain of g_{θ} (see the previous section) vastly reduces the number of expensive training samples we need to generate in order to obtain a dense sampling of the domain.

4.3. Network Architecture

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. This means that the number of tunable parameters is $N = 6w_g + (d_g - 1)w_g(w_g + 1) + M(w_g + 1) + M$, and that 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. In our case, this makes $M = 303$ and $N = 8331$. Figure 5 visualizes the quality of fit for a measurement \mathbf{I} that was not used for training.

5. Applications

The point processor transforms image values at a single point \mathbf{I} into an intermediate representation of the consistent local shape-set in the form of a two-dimensional manifold

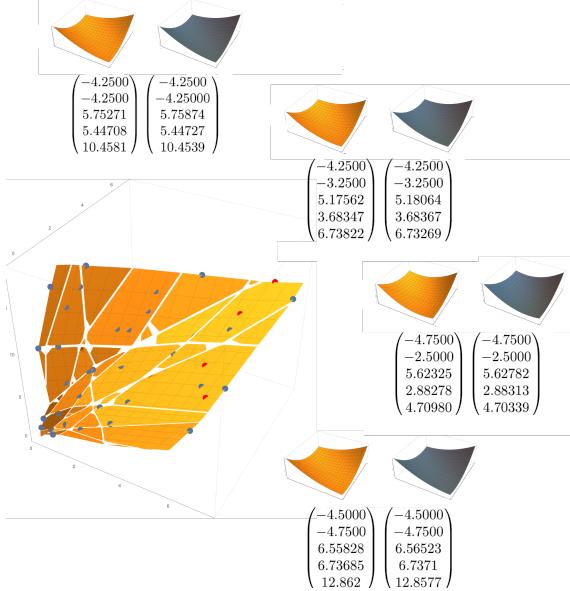


Figure 5. Visualization of the approximator’s interpolation error. This figure depicts $\hat{\phi}_{\mathbf{I}}$ for an \mathbf{I} that was randomly chosen from the convex hull of the training data set, but that was not used as a training sample. The inset shows the four randomly-chosen solutions for which our approximation performs worst, *i.e.* those \mathbf{f} that maximize the error $\|\mathbf{f} - (f_x, f_y, \hat{\phi}_{\mathbf{I}}(f_x, f_y))\|_2^2$.

parameterized by surface orientation, $(f_x, f_y, \hat{\phi}_{\mathbf{I}}(f_x, f_y))$. To demonstrate how this ambiguous 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) by exploiting additional information or assumptions.

5.1. 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 6. 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 parameterized by surface orientation $(f_x, f_y, \hat{\phi}_{\mathbf{I}_1}(f_x, f_y))$, 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 6, 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

$$L(f_x, f_y) := \left\| \hat{\phi}_{\mathbf{I}_2}(f_x, f_y) - \hat{\phi}_{\mathbf{I}_1}(f_x, f_y) \right\|_2^2 \quad (13)$$

for (f_x, f_y) in the disk. For our simulation, this updated two-measurement likelihood is shown on the right Figure 6, 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 6. 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].

5.2. Surface continuity

An alternative way to reduce the per-point ambiguity $F(\mathbf{I})$ 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 of this, 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

$$A(x, y) = \begin{bmatrix} x & y & 0 \\ 0 & x & y \end{bmatrix}. \quad (14)$$

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 7). Then given a measurement \mathbf{I}_2 at one other point (x_2, y_2) , we use that information to update the likelihood over the first set using (13), but with the term $\hat{\phi}_{\mathbf{I}_2}(f_x, f_y)$ replaced by $(\hat{\phi}_{\mathbf{I}_2}(f_x, f_y) + A(x_2, y_2) \cdot \hat{\phi}_{\mathbf{I}_1}(f_x, f_y))$. The updated two-measurement likelihood is shown in the second column of Figure 7.

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

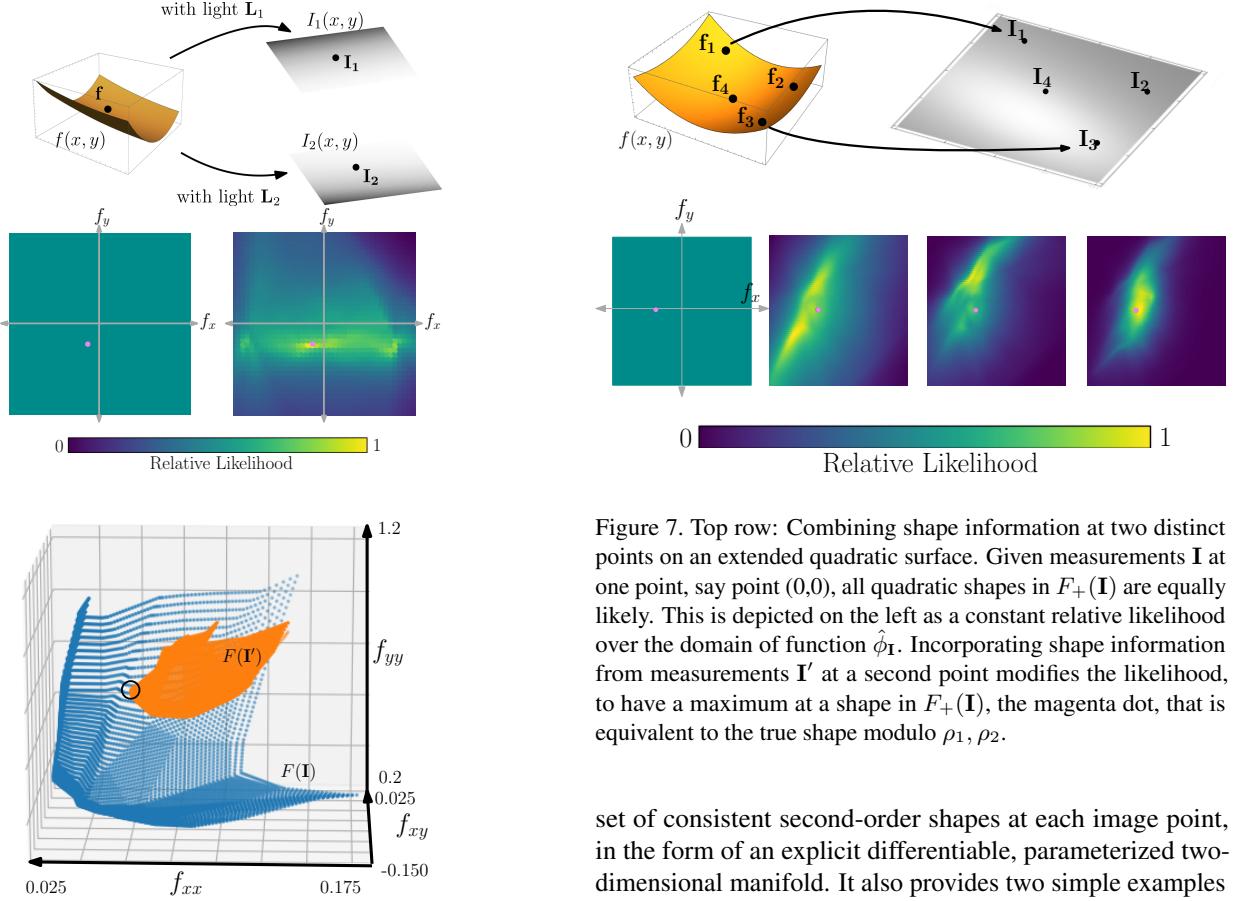


Figure 6. Uncalibrated two-shot photometric stereo. Top row: two simulated images of a surface under different lights, with measurements \mathbf{I}, \mathbf{I}' at the same pixel location. Center row: “Likelihood” of different shapes using only one measurement (left) or both measurements (right), visualized over the orientation domain. Magenta dot indicates true shape used for simulation. Bottom row: Shape sets $F_+(\mathbf{I}), F_+(\mathbf{I}')$ and their intersection (open circle).

time updating the likelihood over the original set $L(f_x, f_y)$ by accumulating the intersection errors between $F_+(\mathbf{I}_i)$ and $F_+(\mathbf{I}_j)$. The evolution of this likelihood for three and four points is shown in Figure 7. 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] that proves the uniqueness of shape reconstruction for extended quadratic patches.

6. 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

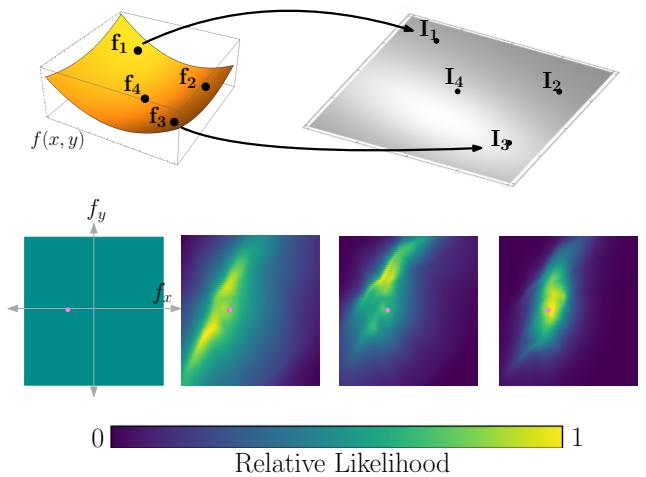


Figure 7. Top row: Combining shape information at two distinct points on an extended quadratic surface. Given measurements \mathbf{I} at one point, say point $(0,0)$, all quadratic shapes in $F_+(\mathbf{I})$ are equally likely. This is depicted on the left as a constant relative likelihood over the domain of function $\hat{\phi}_{\mathbf{I}}$. Incorporating shape information from measurements \mathbf{I}' at a second point modifies the likelihood, to have a maximum at a shape in $F_+(\mathbf{I})$, the magenta dot, that is equivalent to the true shape modulo ρ_1, ρ_2 .

set of consistent second-order shapes at each image point, in the form of an explicit differentiable, parameterized 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 Pattern 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. 1
- [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. 5
- [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. 1
- [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. 5
- [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